Predictive Risk Modelling of Hospital Emergency Readmission, and Temporal Comorbidity Index Modelling Using Machine Learning Methods

Mesgarpour, M.

UNIVERSITY of WESTMINSTER

Predictive Risk Modelling of Hospital Emergency Readmission, and Temporal Comorbidity Index Modelling Using Machine Learning Methods

by

Mohsen Mesgarpour

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Science and Technology
Health and Social Care Modelling Group (HSCMG)

September 2017

Declaration of Authorship

I, MOHSEN MESGARPOUR, declare that this thesis titled, 'Predictive Risk Modelling of Hospital Emergency Readmission and Temporal Comorbidity Index Modelling Using Machine Learning Methods' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
- Parts of this work have been published as:

    1. Mesgarpour, M., Chaussalet, T., & Chahed, S. (2014). A review of dynamic bayesian network techniques with applications in healthcare risk modelling. In OASIcs-OpenAccess Series in Informatics (Vol. 37). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
    2. Mesgarpour, M., Chaussalet, T., & Chahed, S. (2016, June). Risk Modelling Framework for Emergency Hospital Readmission, Using Hospital Episode Statistics Inpatient Data. In Computer-Based Medical Systems (CBMS), 2016 IEEE 29th International Symposium on (pp. 219-224). IEEE.
    3. Mesgarpour, M., Chaussalet, T., Worrall, P., & Chahed, S. (2016, June). Predictive Risk Modelling for Integrated Care: a Structured Review. In Computer-Based Medical Systems (CBMS), 2016 IEEE 29th International Symposium on (pp. 42-47). IEEE.
    4. Mesgarpour, M., Chaussalet, T. & Chahed, S. (In submission process) Ensemble Risk Model of Emergency Readmission (ERMER). International Journal of Medical Informatics.
    5. Mesgarpour, M., Chaussalet, T. & Chahed, S. (In submission process) Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER).
    6. Mesgarpour, M., Chaussalet, T., Worrall, P., & Chahed, S. (In write-up process) Development and Benchmark of Emergency Re/admission for Integrated Care (ERIC) Model. .

Signed: _____

Date: _____

"Finding a good representation of the massive amount of knowledge about the world is hard enough, it is compounded by the need to efficiently extract contextually relevant knowledge depending on the situation."

- Jeff Hawkins

"Often, people who can do, don't because they're afraid of what people that can't do will say about them doing."

- Trevor Noah

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem."

- John Tukey

"Optimism is a strategy for making a better future. Because unless you believe that the future can be better, it's unlikely you will step up and take responsibility for making it so. If you assume that there's no hope, you guarantee that there will be no hope. If you assume that there is an instinct for freedom, there are opportunities to change things; there's a chance you may contribute to making a better world. The choice is yours."

- Noam Chomsky

"Where there is often a question for the doctor if it is ethical to conduct an experiment, but from the broader view-point [it is] a question of whether it is ethical not to conduct experiments."

- William C. Cochran

"All models are wrong, but some models are useful."

- George E.P. Box

"You yourself are your own obstacle, rise above yourself."

- Hafiz

"A fool thinks himself to be wise, but a wise man knows himself to be a fool."

- William Shakespeare

UNIVERSITY of WESTMINSTER

Abstract

Faculty of Science and Technology
Health and Social Care Modelling Group (HSCMG)

Doctor of Philosophy

by Mohsen Mesgarpour

This thesis considers applications of machine learning techniques in hospital emergency readmission and comorbidity risk problems, using healthcare administrative data. The aim is to introduce generic and robust solution approaches that can be applied to different healthcare settings. Existing solution methods and techniques of predictive risk modelling of hospital emergency readmission and comorbidity risk modelling are reviewed. Several modelling approaches, including Logistic Regression, Bayes Point Machine, Random Forest and Deep Neural Network are considered.

Firstly, a framework is proposed for pre-processing hospital administrative data, including data preparation, feature generation and feature selection. Then, the Ensemble Risk Modelling of Hospital Readmission (ERMER) is presented, which is a generative ensemble risk model of hospital readmission model. After that, the Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER) is presented for identifying very sick comorbid patients. A Random Forest and a Deep Neural Network are used to model risks of temporal comorbidity, operations and complications of patients using the T-CARER.

The computational results and benchmarking are presented using real data from Hospital Episode Statistics (HES) with several samples across a ten-year period. The models select features from a large pool of generated features, add temporal dimensions into the models and provide highly accurate and precise models of problems with complex structures. The performances of all the models have been evaluated across different timeframes, sub-populations and samples, as well as previous models.

Acknowledgements

I would like to thank a number of people who helped me throughout my PhD studies. Firstly, I am extremely grateful to my supervisor Prof. Thierry Chaussalet, and Dr. Salma Chahed for their support, valuable guidance and encouragement. I also would like to thank my examiners Dr. Artie Basukoski and Prof. Vincent Augusto, who provided their full supports and constructive feedback.

This work was supported by the Health and Social Care Modelling Group (HSCMG) at the University of Westminster. I sincerely acknowledge Prof. Thierry Chaussalet, Prof. Taj Keshavarz and Dr Andrzej Tarczynski for providing this opportunity, and their help throughout the project.

I am grateful to all my friends who were supportive and made this an enjoyable experience.

Finally, I owe my gratitude to my family for their patience, passionate support and guidance.

Contents

List of Figures

List of Tables

List of Algorithms

List of Symbols

Functions and Operators

Symbol	Definition
$ORDER\, x\, BY\, y_1, ..., y_n$	Order variable x based on variables y_1 to y_n.
$REMOVE\ x$	Remove variable x.
$SET\ x\ TO\ y$	Set variable x to variable y.
$FOR\ \ f(x)\ :\ \ g(x)\ ENDFOR$	Execute statement $g(x)$ as long as condition $f(x)$ is true.
$IF\ f(x),\ THEN\ g(x)\ ENDFOR$	If condition $f(x)$ is true, then run function $g(x)$.
OR	Apply logical *or* operator.
AND	Apply logical *and* operator.
$DATE(x)$	Convert variable x to a $DATE$ type variable.
$f(X),\ WHERE\ g(x)$	Perform statement $f(x)$, where condition $g(x)$ is true.
$FACTORISE_{f(x)}(x)$	Factorise variable x using function $f(x)$.
$CONVERT_{f(x)}(x)$	Convert variable x using function $f(x)$.
$MIN_y(x)$	Calculate the *minimum* value of variable x for subset y.
$MAX_y(x)$	Calculate the *maximum* value of variable x for subset y.
$AVERAGE_y(x)$	Calculate the *mean* of variable x for subset y.
$STD\_DEV_y(x)$	Calculate the standard deviation (*Std-Dev*) of variable x for subset y.
$COUNT_y(x)$	Calculate the *count* of variable x for subset y.
$GO\ TO\ X$	Go to $STEP$ named x in algorithm.
$STEP\ X:$	Define a $STEP$ with name x.
$ACC(x)$	Calculate the *accuracy* of solution x.
$AUC(x)$	Calculate the Area Under Curve (AUC) of Receiver-Operating Characteristic (ROC) for solution x.

$RMSE(x)$	Calculate the Root Mean Square Error ($RMSE$) of solution x.
$SAR(x)$	Calculate the SAR performance indicator of solution x.
$NEXT_x(y)$	Select the *next* variable x from set y.
$RANDOM_x(y)$	Select a *random* variable x from set y.
$VAR(x)$	Calculate the *variance* of variable x.
$BIAS(x)$	Calculate the *bias* of variable x.
$E(x)$	Calculate the *expectation* value of x.
$VAR(x)$	Calculate the *variance* of x.

Roman Letters

Symbol	**Definition**
patient	It represents all records that belong to an individual *patient*.
spell	It represents all records that belong to a *spell* of an individual patient.
episode	It represents an *episode* record of an individual patient.
trial	It represents a *trial* run of the main algorithm.
iteration	It represents an *iteration* run of the sub-algorithm.
ensemble	It represents an *ensemble* model, which is a function of sub-models with their respective weights.
model	It represents a *model*.
sum	It represents a *sum* feature.
min	It represents a *minimum* feature.
max	It represents a *maximum* feature.
σ_x	It represents the *variance* of model x.
E_x	It represents the *expectation* value of variable x.
VAR_x	It represents the *variance* of variable x.
ε	It represents the *error* or the *noise* of a function.

List of Terms

Symbol	Definition
A&E Department	Accident & Emergency (A&E), a.k.a Emergency Department (ED), Emergency Room (ER), Emergency Ward (EW) or Casualty Department, is a medical treatment facility specialising in emergency medicine, acute care of patients who attend without a prior appointment; either by their own means or by that of an ambulance.
ACSC	Ambulatory Care Sensitive Conditions (ACSC) are chronic conditions for which it is possible to prevent acute exacerbations and reduce the need for hospital admission through active management. ACSC are largely preventable, and early identification of patients with ACSC is crucial if their management is to be successful.
CCG	The Primary Care Trusts (PCTs) were part of the NHS in England from 2001 to 2013. PCTs were replaced by the Clinical Commissioning Groups (CCG), as part of the Health and Social Care Act 2012.

CCS	The Clinical Classifications Software (CCS) for ICD-10-CM/PCS is a database and a software tool developed as part of the Healthcare Cost and Utilization Project (HCUP), a US Federal-State-Industry partnership sponsored by the Agency for Healthcare Research and Quality (AHRQ). HCUP databases, tools, and software inform decision-making at the national, State, and community levels. The CCS for ICD-10-CM and PCS is a diagnosis and procedure categorisation scheme that can be employed in many types of projects analysing data on diagnoses and procedures.
SHMI	The Summary Hospital-level Mortality Indicator (SHMI) is a national mortality indicator by Dr Foster that complements the Hospital Standardised Mortality Ratio (HSMR) and is adapted by the NHS.
*CTV*3	Read Codes are a coded thesaurus of clinical terms and have been used in the NHS since 1985. Presently, there are two versions of read codes that are being used by GPs: READ version 2, and Clinical Terms Version 3 (CTV3). They provide the standard vocabulary by which clinicians can record patient findings and procedures. The first release of CTV3 occurred in the late 1990s, and is planned to be replaced by SNOMED-CT before April 2018.
Episode	A consultant episode (hospital provider) is the time a patient spends in the continuous care of one consultant using hospital site or care home bed(s) of one health care provider or, in the case of shared care, in the care of two or more consultants.
Emergency Readmission	The number of finished and unfinished continuous inpatient spells that are emergency admissions within a number of days (e.g. 30-day) of the last, previous discharge from hospital. Emergency readmissions have been proposed as a measure of adverse hospital outcomes for some time.

GP Services	In the UK, General practitioners (GPs) are primary care doctors providing the first point of contact with the NHS for most people in their communities. GPs help patients by trying to identify health problems at an early stage, provide advice and support and promote preventive care.
HRG	Healthcare Resource Groups (HRGs) are standard groupings of clinically similar treatments, which use common levels of healthcare resource. HRGs help organisations to understand their activities, regarding the types of patients they care for and the treatments they undertake.
HES	The Hospital Episode Statistics (HES) warehouse was originally founded in 1987 and is an administrative database that contains all inpatient admissions, outpatient appointments and A&E attendances to NHS hospitals in England. The HES database is a verified and a less detailed version of the SUS, which is used at hospitals.
HSCIC	The Health and Social Care Information Centre (HSCIC) organisation was created as a special health authority in 2005 by a merger of parts of the Department of Health, parts of the NHS Information Authority, and the Prescribing Support Unit. In 2016, it was announced that the HSCIC would be re-branding to *NHS Digital*. On of the main duties of the HSCIC is to maintain NHS healthcare databases, including the HES.
$ICD-10$	ICD-10 is the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list by the World Health Organization (WHO). It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases.
Inpatient Care	Inpatient care is the care of patients whose condition requires admission to a hospital.

$OPCS-4$	OPCS-4, or more formally OPCS Classification of Interventions and Procedures (Version 4), is the procedural classification used by clinical coders within NHS. It is based on the earlier Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (4th revision).
Outpatient Care	Outpatient care is for patients who may need clinical services, although don't necessarily need to be admitted to hospital. Patients may also attend outpatient clinics following a period in hospital for follow-up treatments.
NHS	The National Health Service (NHS) is the name of the public health services of England, Scotland, Wales, and Northern Ireland. They were established together as one of the major social reforms following the Second World War on the founding principles of being comprehensive, universal and free at the point of delivery.
ONS	The Office for National Statistics (ONS) is the UK's largest independent producer of official statistics and its recognised national statistical institute. It is responsible for collecting and publishing statistics related to the economy, population and society at national, regional and local levels.
PbR	Payment by Results (PbR) provides a transparent, rules-based national tariff system, used to determine the reimbursement of NHS funded care in England. It rewards efficiency, supports patient choice and diversity and encourages activity for sustainable waiting time reductions.
PCT	The Primary Care Trusts (PCTs) were subcommittees of discrete health authorities and part of the NHS in England from 2001 to 2013. Afterwards, PCTs were replaced by the Clinical Commissioning Groups (CCG), as part of the Health and Social Care Act 2012.

QoF — The Quality and Outcomes Framework (QOF) is part of the General Medical Services (GMS) contract for GPs and was introduced in 2004. The QOF rewards practices for the provision of *quality care* and helps to fund further improvements in the delivery of clinical care.

Read Code V2 — Presently, Read codes are the standard clinical terminology system used by GPs in the UK. It supports detailed clinical encoding of multiple patient phenomena including: occupation; social circumstances; ethnicity and religion; clinical signs, symptoms and observations; laboratory tests and results; diagnoses; diagnostic, therapeutic or surgical procedures performed; and a variety of administrative items. The first release of 5-Byte READ released prior to 1991, and it is planned to be replaced by SNOMED-CT before April 2018.

Spell — A hospital provider spell is the total continuous stay of a patient using a hospital bed on premises controlled by a health care provider during which medical care is the responsibility of one or more consultants, or the patient is receiving care under one or more nursing episodes or midwife episodes in a ward.

$SNOMED-CT$ — The SNOMED-CT (Systematized Nomenclature of Medicine - Clinical Terms) is the most comprehensive and precise clinical health terminology product, owned and distributed by SNOMED International. The SNOMED-CT has been developed collaboratively worldwide and is now accepted as a common global language for health terms. The SNOMED CT must be adopted by all GPs and in systems used by GP service providers, before April 2018.

SUS	The HES database is derived from the Secondary Uses Service (SUS) database, which is supplied by the Commissioning Data Sets (CDS) directly from hospitals. The SUS database is a collection of administrative data that contains all inpatient admissions, outpatient appointments and A&E attendances to National Health Service (NHS) hospitals in England.

Glossary

A&E Accident & Emergency. 1, 3, 19, 23, 38, 53–56, 62, 64, 65, 76, 79, 102, 105, 125, 136, 145, 147, 148, 151

ABS Agent-Based Simulation. 157

ACB Amber Care Bundle. 159

ACC Accuracy. 93

ACG Adjusted Clinical Group. 4, 21, 67, 113, 143

ACP Advance Care Planning. 158, 159

ACS Acute Coronary Syndrome. 158

ACSC Ambulatory Care Sensitive Condition. 16, 17, 20, 137, 152, 169–171, 252, 253

ADG Aggregated Diagnosis Group. 4, 21, 67

AE Accident & Emergency. 147

AHRQ Agency for Healthcare Research & Quality. 20, 22, 70, 168

AMI Acute Myocardial Infarction. 39, 153

ANN Artificial Neural Network. 6, 7, 18, 33, 35, 40, 42, 47–51, 118, 145

APC Admitted Patient Care. 164

ASA-PS American Society of Anesthesiologists Physical Status. 22

AUC Area Under the Curve. 29, 90, 93, 96, 97, 104, 114, 142, 143, 145–154, 160

BETS Bayesian-Estimated Tools for Survival. 160

BMA Bayes Model Averaging. 35

BMC Bayesian Model Combination. 35

BN Bayesian Network. 32, 33, 45, 160–162

BOC Bayes Optimal Classifier. 35, 36

BPM Bayes Point Machine. 9, 31, 33, 45, 46, 63, 83, 86, 87, 89–91, 104, 108, 109, 129, 131, 135, 138, 145

C-PHD Coxian Phase-type Distribution. 4, 26, 62, 89, 145, 146, 151, 153, 156

CART Classification & Regression Tree. 39, 117

CCG Clinical Commissioning Group. 3

CCI Charlson Comorbidity Index. 4, 10, 20, 21, 28, 67–70, 113, 114, 121, 123, 124, 145, 168

CCS Clinical Classifications Software. 115, 125, 168

CDS Commissioning Data Sets. 55

CHD Coronary Heart Disease. 110, 124

CHF Congestive Heart Failure. 110, 124

CMG Case-Mix Group. 142–154

CMS Centers for Medicare & Medicaid Services. 18, 144, 147

CMS-HCC The Centers for Medicare & Medicaid Services (CMS) Hierarchical Condition Categories (HCC). 21, 22, 144

COPD Chronic Obstructive Pulmonary Disease. 38, 110, 124, 144, 150, 156

CPM Combined Predictive Model. 3, 10, 17, 18, 38, 62–66, 90, 103, 104, 110, 151

CPRD Clinical Practice Research Datalink. 18

CPU Central Processing Unit. 109, 116, 131

CSV Comma Separated Value. 132

CTV3 Clinical Terms Version 3. 54, 77

DBN Dynamic Bayesian Network. 45, 161, 162

DCLNs Discharge Community Link Nurses. 158, 159

DCNN Deep Convolutional Neural Network. 139, 140

DES Discrete Event Simulation. 157

DNN Deep Neural Network. 6, 31, 33, 34, 49, 50, 80, 81, 117, 118, 126, 131, 139, 140

DNN-HMM Deep Neural Network - Hidden Markov model. 139

DoH Department of Health. 3, 7, 16–18, 25, 55, 64

DRG Diagnosis-Related Group. 16, 67, 70, 113

DT Decision Tree. 33

EARLI Emergency Admission Risk Likelihood Index. 19

ECI Elixhauser Comorbidity Index. 4, 10, 20, 67–70, 82, 113, 114, 121, 123, 124

EDC Expanded Diagnosis Cluster. 4, 67

EHR Electronic Health Record. 23

ELCA End of Life Care for All. 158

EM Expectation Maximisation. 36, 44

EMIS Egton Medical Information Systems. 54, 77

EMSE Estimated Mean Squared Error. 99

EoL End of Life. 14, 32, 38, 148, 158, 161

EP Expectation Propagation. 27, 44, 46, 104, 108, 109

ERMER Ensemble Risk Model of Emergency Admissions. v, 9–11, 90, 93, 96, 98, 103, 104, 108, 111, 116, 126, 127, 129–131, 135, 137–139, 145

EU European Union. 15

EuroSCORE European System for Cardiac Operative Risk Evaluation. 22

FPR False Positive Rate. 93

GAM Generalised Additive Model. 38, 39

GBRT Gradient Boosted Regression Trees. 86

GLM Generalised Linear Model. 36, 50, 116, 149, 151, 155

GLMM Generalised Linear Mixed Model. 155, 156

GP General Practice. 3, 17–19, 23, 38, 53, 54, 56, 62, 64–66, 76, 77, 79, 105, 136, 142, 146–148, 150, 151, 164

GPU Graphical Processing Unit. 116, 131

GRACE Global Registry of Acute Coronary Events. 158

GSF Gold Standards Framework. 158, 159

HDFS Hadoop Distributed File System. 72

HES Hospital Episode Statistics. v, 1, 3, 7, 10, 11, 17–19, 22, 25, 38, 53–58, 61–64, 68, 69, 74, 76, 77, 90, 121, 136, 138, 145, 146, 148–152, 160, 168, 171, 172

HF Heart Failure. 40, 149, 150, 153

HLM Hierarchical Linear Model. 155

HMM Hidden Markov Model. 5, 45, 46, 156

HRG Healthcare Resource Groups. 16, 21, 22, 64, 74, 76, 79, 91, 104, 155, 171, 253

HSCIC Health & Social Care Information Centre. 7, 25, 55, 61, 62, 68, 69, 72, 77, 79, 105, 114, 132, 164, 168

HSCIC-CCI HSCIC version of the CCI. 68, 69, 79, 91, 105, 114, 123, 124, 203

ICD International Classification of Diseases. 67, 70, 125

ICD-10 10th revision of the International Statistical Classification of Diseases (ICD). 24, 70, 115

IMD Index of Multiple Deprivation. 76, 80

IP Inpatient. 104, 147

LACE Stands for: length of stay (L); acuity of the admission (A); co-morbidity of the patient (C); & emergency department use (E). 19, 40, 42, 148, 149

LCP Liverpool Care Path. 158, 159

LMM Linear Mixed Model. 155

LoS Length-of-Stay. 7, 15, 21, 28, 32, 38, 70, 105, 113, 114, 138, 142, 143, 145–150, 153, 155, 156, 160, 161

LR Logistic Regression. 4, 18, 19, 22, 26, 31, 33, 35, 37–39, 42, 43, 46, 51, 62, 64, 65, 70, 89, 109, 114, 116, 119, 123–126, 129, 131, 135, 143, 145–155

LSTM Long Short-Term Memory. 48, 50

LTC Long Term Conditions. 6

MACSS Multipurpose Australian Comorbidity Scoring System. 21, 123

MAP Maximum a Posteriori. 27, 44

MARS Multivariate Adaptive Regression Splines. 38, 39

MC Markov Chain. 44, 162

MCMC Markov Chain Monte Carlo. 44, 108

MHMM Mixed Hidden Markov Model. 156

MHSDS Mental health Services Data Set. 76

MLE Maximum Likelihood Estimate. 37, 43

MM Markov Model. 156

MSE Mean Squared Error. 99, 101, 102, 124

MxE Mixture of Experts. 36, 37, 39

NEoLCP NHS End of Life Care Programme. 158

NHS National Health Service. 1–3, 7, 8, 11, 14–18, 20, 22, 25, 38, 54, 55, 61, 63, 64, 69, 77, 136, 148, 150, 151, 155, 158, 164

NRI Net Reclassification Improvement. 29

NRLS National Reporting & Learning System. 76, 164, 168

ODS Organisation Data Service. 77

OECD Organisation for Economic Co-operation & Development. 15

ONS Office of National Statistics. 7, 55, 69, 164

OP Outpatient. 147

OPCS Office of Population Censuses & Surveys (OPCS) Classification of Interventions & Procedures. 22, 115

OPCS-4 Office of Population Censuses & Surveys (OPCS) Classification of Interventions & Procedures, version 4. 115

OR Operational Risk. 32

OS Operating System. 131

PARAMO PARAllel predictive MOdeling. 2

PARR Patients at Risk of Re-hospitalisation. 3, 10, 17, 18, 57, 62–65, 79, 90, 91, 103, 104, 110, 148, 152

PARR-30 Patients at Risk of Re-admission within 30 days. 3, 18

PbR Payment by Results. 16, 55, 76, 77

PCA Principal Component Analysis. 7, 26, 80, 82

PCCC Palliative Care Coordination Centre. 158, 159

PCSC Primary Care Sensitive Conditions. 16

PCT Primary Care Trust. 3, 17, 19, 38, 54, 64–66

PHD Phase-Type Distribution. 156, 160

POSSUM Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. 22

PPC Preferred Priorities for Care. 158, 159

PPV Positive Predictive Value. 143, 145–148, 150–153

PRISM Predictive Risk Stratification Model. 17

PROM Private Major Score Procedure database. 22

QOF Quality & Outcomes Framework. 16, 164

RAM Random-Access Memory. 109, 116, 131

ReLU Rectified Linear Unit. 47, 48, 118

RF Random Forest. 18, 26, 31, 33, 36, 39, 40, 50, 86, 87, 91, 114, 116, 117, 119, 123–126, 129, 131, 135, 145

RLR Randomized Logistic Regression. 86

RMSE Root Mean Square Error. 93, 98, 101

ROC Receiver-Operating Characteristic. 10, 29, 38, 90, 93, 103, 104, 109, 110, 114, 119, 121, 123, 125, 160

RRT Rapid Response Team. 158, 159

SAR Squared Error, Accuracy, & Receiver-Operating Characteristic Area. 93

SD System Dynamics. 157

SEM SUS Extract Mart. 55

SHMI Summary Hospital-level Mortality Indicator. 169

SNOMED-CT Systematized Nomenclature of Medicine - Clinical Terms. 54, 77

SORT Surgical Outcome Risk Tool. 22

SPARRA Scottish Patients at Risk of Readmission & Admission. 17, 148

SPARRA-MH Scottish Patients at Risk of Readmission & Admission for Mental Health. 150

SRS Surgical Risk Scale. 22

ST Survival Tree. 160

SUS Secondary Use Service. 1, 11, 53–55, 61, 62, 64, 65, 76, 77, 136, 142, 145, 147

SVM Support Vector Machine. 6, 7, 18, 31, 33, 35, 40–42, 45, 46, 50, 83, 87, 91, 108, 145

SVM-RFE Support Vector Machine Recursive Feature Extraction. 42, 83, 86, 87

T-CARER Temporal-Comorbidity Adjusted Risk of Emergency Readmission. v, 10, 11, 114–116, 119, 121, 123, 124, 126, 127, 129–132, 135, 138, 139

TN True Negative. 93

TP True Positive. 93

TPR True Positive Rate. 93, 123

UK United Kingdom. 3, 15, 17, 21, 22, 25, 38, 46, 53, 64, 142, 145–153

UML Unified Modelling Language. 129

US United State. 21, 22, 42

USA United State of America. 35, 53, 142–145, 147, 149–151, 153, 154

WDNN Wide & Deep Neural Network. 31, 34, 50, 51, 114, 118, 119, 123–125, 129, 131, 135

Dedicated to my granddad and to my parents.

Chapter 1

Introduction

The cost of care is increasing at a rate that is unaffordable in the current economy. This is mainly due to the impact of the accumulated rise in the ageing population, population growth, deprivations, age-related or long-term comorbidities, emergency admissions, increased expectations, and the cost of treatments and technologies (Lewis et al., 2011, DH, 2013b, NHS, 2013c, Strandberg et al., 2011). The current system is unsustainable and unfair, and available financial options to support people in meeting care costs are limited.

In this research, our focus is on three main areas: developing a healthcare pre-processing framework for hospital data, producing a predictive risk model of emergency re-admission to hospital, and development of a temporal comorbidity index. In the following sub-sections, initially, a brief background about the England healthcare data is provided. After that, the importance of the pre-processing and feature generation is highlighted. Afterwards, the summary of emergency readmission and comorbidity index are provided. Then, gaps and motivations, aims and objectives, and the main contributions of this research are explained. Finally, the outline of the thesis is provided.

1.1 Hospital Administrative Data

In the England's National Health Service (NHS), patients' interactions with hospital services are recorded on statutorily defined datasets, known as Secondary Uses Service (SUS). The SUS data is cleaned and combined on a national basis to create Hospital Episode Statistics (HES) data. The HES contains administrative hospital data for all inpatient, outpatient and Accident and Emergency (A&E) admissions in England. The

databases hold admissions, clinical, utilisation and demographics details in the format of *episodes* and *spells*. The *spell* refers to a continuous period of care, which includes one or more *episodes* of care activities (HSCIC, 2016d).

1.2 Pre-processing and Generating Features

Finding representations that capture the structure of the input data, is of particular importance. As Jeff Hawkins, founder of the Redwood Center for Theoretical Neuroscience, said once "Finding a good representation of the massive amount of knowledge about the world is hard enough, it is compounded by the need to efficiently extract contextually relevant knowledge depending on the situation" (Hawkins and Blakeslee, 2007).

The existing studies (Section 2.2) have obtained features using three main approaches:

- Referring to self-judgement or an obscure clinical experience;
- Justifying by a vague and basic exploratory analysis of a few features;
- Using features in the previous studies.

Although disregarding prior probabilities and distribution patterns are clear indications of a bad modelling practice, no framework or systematic way of data pre-processing and feature generation have been found in the existing studies. However, there are a number of frameworks available, like the PARAMO framework proposed by Ng et al. (2014), that focus on facilitating large-scale modelling endeavour to speed-up research workflow.

1.3 Prediction of Emergency Readmission to Hospitals

The National Health Service (NHS) spends an estimated £11 billion per-year on emergency admissions in England (Lewis et al., 2011). According to the Nuffield Trust report in 2012 (Nuffield-Trust, 2012), about 8% of discharged patients are readmitted within 30 days, costing an estimated £2.2 billion a year. Based on the retrospective study by Clarke et al. (2012), about half of the 30-day emergency readmissions were potentially preventable between 2004 and 2010.

Four major risks contribute to the increase in emergency (or unplanned) readmissions to hospitals (Lewis et al., 2011, HSCIC, 2013a): ageing population and frailty (Caley and Sidhu, 2011), patients with long-term conditions (DH, 2012), premature discharge and unpredictable accident and emergency (Clarke et al., 2012). While discharging patients provides a way of freeing beds in healthcare systems, premature discharge can still increase the risk of emergency readmissions. Often hospital admission or readmission can be avoided by providing adequate care (Bardsley et al., 2012), therefore identification of high-risk patients can help clinicians to prevent avoidable readmissions.

Predictive risk models can help patients and carers obtain appropriate support services in clinical decision-making. In addition, such models can improve care quality and reduce the costs of inappropriate admissions to hospital and A&E. In 2005, the UK Department of Health (DoH) commissioned the Patients at Risk of Re-hospitalisation (PARR) (Billings et al., 2006a, Lewis, 2011) algorithm and *PARR++* software for Primary Care Trusts (PCTs)[1] (Fund, 2016a, Lewis et al., 2011). The aim of the PARR model was to identify individuals at high-risk of emergency readmission to the hospital within a year based on the inpatient data from the HES database. After that, to address the need for identifying the patient risk along a continuum, in 2006 the DoH released the Combined Predictive Model (CPM), which was based on the General Practice (GP) and the HES data (DH, 2006).

In 2011, the DoH commissioned an upgrade to the PARR and the CPM models (DH, 2011a, Nuffield-Trust, 2012). The Patients at Risk of Readmission within 30 days (PARR-30) model was developed as an upgrade, to be run by acute hospitals. The PARR-30 model was based on a broad range of measures used in the PARR (Billings et al., 2012), but features more restricted due to restriction in recording within 30-days.

After the controversies of the 2012 Heath and Social Care Act (Timmins, 2013), the care system moved towards developing new models of integrated care. The NHS's strategic 5-year forward view (NHS, 2014) outlines that commissioners, the NHS and other providers will co-design the services based on a model of integrated care that targets specific cohorts, with their own exemplars, potential benefits, risks and transition cost.

[1]The Primary Care Trusts (PCTs) were subcommittees of discrete health authorities and part of the NHS in England from 2001 to 2013. Afterwards, PCTs were replaced by the Clinical Commissioning Groups (CCGs), as part of the Health and Social Care Act 2012.

1.4 Comorbidity Risk Index

There is increasing evidence that the quantification of high-risk operations and procedures with adequate adjustment can greatly improve the quality of readmission models. There have been two streams of work on risk scoring comorbidities to estimate future resource utilisation, readmission and mortality.

One stream of research looks at the odds ratio of major diagnoses groups and therefore is highly reliant on the whole population statistics. Another weakness of such model stems from crudely summing up the risk score for comorbidities, which are based on the most recent admission of the patients. One major model is the Charlson Comorbidity Index (CCI), which relies on twenty-two comorbidity groups (Charlson et al., 1987).

Another stream uses a diagnoses classification approach or a case-mix model based on similarity, type of care, likelihood or duration, which are usually very complex, specialised to highly particular settings and population characteristics. One popular model is Elixhauser Comorbidity Index (ECI), which is using patients' diagnoses and an Aggregated Diagnosis Groups (ADGs) classifier (Elixhauser et al., 1998, AHRQ, 2016b). Another well-established method is the John Hopkin's (Weiner and Abrams, 2011) Adjusted Clinical Groups (ACGs), which uses the ADGs to encapsulates 32 diagnoses groups, and their aggregations called Expanded Diagnosis Clusters (EDCs).

1.5 Motivation

Firstly, in recent years many predictive risk models of emergency readmission have emerged. However, many studies in the literature simplify the model, and the selected features are mainly based on previous models, personal experience or a very shallow exploratory analysis.

Secondly, most of the existing predictive risk models (Section 2.2) that used hospital administrative data were based on Logistic Regression (LR) or Coxian Phase-type Distribution (C-PHD) models. Although they are uncomplicated and powerful, they are bounded by algorithms' shortfalls, restricted assumptions and limited parameters.

The main shortcomings of these models can be summarised as:

- Oversimplifying complex correlations;
- Not accounting for small probabilities in an appropriate way;

- Not updating the beliefs (prior probabilities) based on the environmental variables or changes in the policies;
- Do not adjust the model for important factors;
- Not using an iterative or parallel mechanism to compare different models and settings (Fenton and Neil, 2012, Ng et al., 2014).

After the breakdown of financial markets at 2008, Rodriguez (2011) wrote, "predictive modelling, the process by which a model is created or chosen to try to best predict the probability of an outcome has lost credibility as a forecasting tool". This is either due to the modeller's expertise or knowledge, or the lack of resources. In below, the main common causes of predictive modelling failures are outlined:

- Inadequate data pre-processing approaches;
- Incomplete model validation methods;
- Unjustified extrapolations;
- Over-fitted models (Kuhn and Johnson, 2013).

In healthcare risk modelling, there have been many successful implementations of machine learning methods, but, there are limited numbers of literature that applied a Bayesian approach or specialised sub-models of cohorts.

The combination of machine learning and forecasting helps to recognise subtle changes in patients flow and behavioural patterns (Weiner and Abrams, 2011). Moreover, generative models have the capability to move beyond associations between predictors and the outputs and recognise hidden structures in the data by encoding the prior probabilities into the model. Examples of generative methods are Bayesian approaches, mixture models and generative models using hidden units or latent stochastic variables, like Hidden Markov Models (HMMs) and Generative Adversarial Networks (Goodfellow et al., 2016, Koller and Friedman, 2009). Furthermore, Ensemble methods in machine learning are a powerful type of approach, which use a finite set of weaker models and an algorithm to combine and optimise the performance of the Ensemble model (Chapter 4).

Finally, the majority of comorbidity risk models only consider the most recent admission and the first few diagnoses of the patient to rank the comorbidity risk of the patient. But, very sick and comorbid patients usually have multiple medical conditions

and operations, or procedure with complex existing conditions. There are four major areas that comorbidity index models can improve. Firstly, to make the risk score relevant to different environments, an approach must be used to model complex correlation between variables and states. Secondly, to better distinguish the short and long-term conditions (LTCs), the temporal dimension may be included in the form of life-table (Singer and Willett, 2003) or in the form of a polynomial weight function (Bee Dagum and Bianconcini, 2016). Thirdly, population stratification is a major factor in the prevalence of medical conditions and must be adjusted. Fourthly, major correlated factors to diagnoses must be included directly (observable) or indirectly (latent) to improve the risk estimates, including secondary diagnoses, operations, procedures and complications.

Consequently, there are three main areas that can enhance the performance of the predictive risk models. Firstly, designing a generic framework for data pre-processing and feature generation is highly desirable, in order to be able to include a large pool of features in the analyses. Secondly, developing a robust decision support tool for modelling emergency readmission is beneficial, which can add the prior probability of patients characteristics into the model. Thirdly, comorbidity index is an extremely significant factor in predictive risk models and has a high potential for further improvement. Presently, comorbidity index models have four major weakness areas: robustness, temporal dimensions (length-of-stay and delta-time between admissions), population stratification and adjusting for associated risk factors.

Moreover, in the healthcare risk modelling research area, there have been many successful implementations of machine learning methods (Bottaci et al., 1997, Green et al., 2006, Lee et al., 2012, Nilsson et al., 2006, Peelen et al., 2010, Song et al., 2004). But, there are a few numbers of literature that used a Bayesian approach or a Deep Neural Network (DNN) to address emergency hospital readmission and comorbidity index modelling problems (Álvaro-Meca et al., 2012, Cui et al., 2015, Demir and Chaussalet, 2011, Gupta et al., 2014, Helm et al., 2015, Huws et al., 2008).

1.5.1 Scepticism

There is always machine learning scepticism in reaction against the hype or failures of inappropriate modelling approaches. Bottle et al. (2014) stated that advanced machine learning methods, particularly Artificial Neural Network (ANN) and Support Vector Machines (SVMs), did not offer noticeably better predictions for readmission risk compared to Linear Regression, and they were relatively harder to implement. Based on

the reported details in this study, standard approaches for ANN and SVM modelling are mainly used. There are four main concerns regarding the applied machine learning methods in this study:

- The inputted features into the utilised Principal Component Analysis (PCA) algorithm may have missed some main influencing features; therefore, derived features could be fictional and consequently have a hit on the performance of the applied SVM algorithm (Yang et al., 2005b);
- Using inappropriate input features, small training set, unsuitable network design or poor initial solution may have negatively impacted on the performance and the convergence of the ANN (Matignon, 2005);
- The temporal dimension may have been included into machine learning models using weighted observations or including prior probabilities into the modelling.
- The linearity and normality assumptions are not necessarily true for the variables. The model performance could be improved with the non-parametric assumption or heavy-tailed priors for the variables, which are usually more robust to outliers (Congdon, 2010).

1.6 Aims and Objective

The ultimate goal of this research is to produce solutions that can be used as a systematic methodology for data processing, and predictive modelling solutions that improve patients' life quality. In this context, it is important to get a better understanding of individual patient pathways and evaluate the solution based on patient-centred outcome parameters, such as readmissions, comorbidities, utilisation risks, health risks and Length-of-Stay (LoS) (Adeyemi et al., 2013, Weiner and Abrams, 2011).

Moreover, the model development is based on the hospital administrative data (HES database) without any intervention. In addition, supplementary data from the NHS, the DoH, the HSCIC and the ONS are used in a number of models, in order to confirm variable distributions and re-categorise variables.

The main objectives of this research are as follows:

- Healthcare pre-processing framework:

  - Developing a comprehensive framework to clear and prepare hospital;
  - Defining a feature generation process for creation of a pool of potential features.

- Hospital readmission model:
  - Designing a robust generative approach to predict the risk of hospital emergency readmission within a year.
  - Using an Ensemble method to increase overall precision and improve risk predictions for the main sub-cohort of patients.

- Comorbidity risk model:
  - Proposing a comorbidity index to identify very sick patients that are in risk of emergency admission in future (which is highly correlated to increased length-of-stay).
  - Including temporal aspects of patients comorbidities and associated risks into the comorbidity risk model.
  - Trying to capture complex abstract layers of patients health status into the comorbidity risk model.

- Benchmarking:
  - Testing the applicability of modelling approaches across different cohorts and time-frames using the NHS hospital inpatient data.
  - Comparing the performances of the designed models against the recent models from several analytical angles.

- Reproducibility:
  - Producing open-source and easy-to-use software toolkits to allow users to apply modelling solutions and incrementally develop the solutions.
  - Providing clear documentation of toolkits and produce step-by-step guideline of modelling.

1.7 Contribution

In recent years, many patients risk modelling approaches, in the form of hospital emergency readmission models and comorbidity risk indices, have been emerged as a result

of collaborations between academia and local healthcare providers. However, many studies in the literature use extremely similar features and very simplistic modelling techniques and have moderate performances. There is a pitfall of model oversimplification and relying too much on outdated researchers, which affects the precision of designed solutions.

Healthcare modelling using administrative data leads itself to complex feature generation and extensive model designing and tuning. Some studies use self-judgement or vague exploratory analysis to produce features, and the majority other simply use a version of features based on the previous studies, without any systematic feature generation. Also, there is a very poor integration of temporal dimensions into comorbidity risk models, which leads to heavily biased risk scores. Therefore, an adaptable feature pre-processing and generation, as well as a robust modelling technique can have major benefits for continuously changing healthcare systems.

In view of the above discussions, the key contributions of this thesis are as follows:

- In this research, a generic healthcare pre-processing framework has been developed for healthcare data. The framework defines a systematic way to sample, clean and treat input data, create *super-spells*, generate and select features. The feature selection includes steps for filtering stationary features, visual exploratory analyses, feature transformation, filtering correlated features and feature importance ranking. The developed framework has been proved to be highly effective in practice, and it has led to the generation of highly significant features in our modelling, including some new significant features that have not appeared in the previous studies.

- A predictive risk model for the 1-year emergency readmission to the England's hospitals has been developed using a generative method. The chosen generative method is a version of Bayes Point Machine (BPM) (Herbrich et al., 2001, Minka, 2001b), which is a nonlinear Bayesian classifier and uses quadratic programming to optimise the classification's hyperplanes using support vectors and margin. It has been shown that the inclusion of the prior probabilities of the features can increase stability and precision (positive predictive value) of the models.

- An Ensemble of generated BPM models of emergency readmission has been developed, which is based on a collection of sub-models that are conditioned on different population characteristics. The proposed model, Ensemble Risk Model of Emergency Admissions (ERMER), utilises a weighted average ranking method to optimise the weights of sub-classifier using a bidirectional hill-climbing heuristic. However, the ensemble of specialised sub-models for prediction of patients

risks has not been addressed with existing studies. The novelty of our solution lies in the intuitive adaptation of an Ensemble modelling with a generative approach for prediction of patients' risks.

- A Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER) has been produced that is robust, temporal (incorporate length-of-stay and delta-time between admissions), and adjust for population stratification and major comorbidity-related risk factors, including operations and complications. The produced solution has sufficient generality to be extended to other healthcare settings and internationally.

- The developed solutions are trained, tested and cross-validated across several samples from 10-year of inpatient records (1999-2010) from the HES inpatient database. The performance profiling has been done from several aspects using a wide range of methods, in order to produce clear insight into the models' strength and bias. All the produced models are stable with high prediction accuracies and precisions.

- The models are benchmarked against the solutions that are currently in use. The readmission models are compared against the CPM (DH, 2006, Paton et al., 2014), the PARR (Billings et al., 2006a), and Billings et al. (2013) (a.k.a. CPM update) models using the reported performance statistics. Moreover, the co-morbidity risk models are benchmarked against two major models, the Charlson Comorbidity Index (CCI) (Charlson et al., 1987) and the Elixhauser Comorbidity Index (ECI) (Elixhauser et al., 1998, AHRQ, 2016b) using a meta-analysis of multiple benchmarking surveys. The comparison of prediction performances and comorbidity-wise analyses strongly indicates that the ERMER and the T-CARER models produce significant improvements in terms of precision, accuracy and can greatly enhance the decision support systems in healthcare. The ERMER had ROCs between 0.76-0.77 and the T-CARER had ROCs between 0.73-0.80 using different settings and datasets.

- Generic, open-source and easy-to-use software toolkits have been developed to model the emergency readmission and the comorbidity risk, which are based on the developments of the ERMER and the T-CARER. They consist of procedures in MySQL, Python, Bash and C#, as well as third-party libraries, which are controlled via user-friendly Jupyter Notebooks. Moreover, the usefulness of the developed tools is going be exposed to the academic community and healthcare researchers, which will potentially lead to a wider adaptation of more sophisticated methods of patient risks modelling by healthcare sectors.

- The presented solutions are transferable to other problems across healthcare domain, due to the general applicability of patient risk scoring techniques and the generality of the developed toolkits. Moreover, the research is expected to contribute to the academic communities, including the *operational research* and health and social care modellers, in order to implement more effective decision support systems and risk indices.

1.8 Thesis Overview

The remainder of this thesis is organised as follows. Chapter 2 provides a background on emergency readmission and risk scoring modelling. Then, complexity levels in data quality, feature generation, modelling and validation are presented in Chapter 3. Moreover, the modelling approaches that are used throughout this research are explained in Chapter 4. Next, the three main phases of the analyses are defined in Chapter 6, including the considered benchmarking models. Chapter 5 contains a brief overview of the healthcare data, NHS administrative data and the description of the extracted samples. Thereafter, the first phase of the project, the healthcare pre-processing framework is presented in Chapter 7, which is based on the HES and the SUS, but has a generic structure. Chapter 8 presents the Predictive Risk Modelling of Hospital Readmission (ERMER). Then, the Temporal-Comorbidity Adjusted Risk Emergency Readmission (T-CARER) is presented in Chapter 9. Chapter 10, describes the open-source toolkits developed for applying the ERMER and the T-CARER solutions. Finally, Chapter 11 contains the conclusion and the future work.

Chapter 2

Background

For many healthcare providers and purchasers identification of high-risk events has been a major concern. According to Lewis et al. (2011), there are three major sources of risk to the healthcare system:

- Ageing population and frailty;
- The increasing number of people with long-term conditions;
- The rising rate of emergency admissions to hospitals.

Firstly, a major concern in healthcare organisations throughout the world is about coping with an ageing population (Caley and Sidhu, 2011). Survey results (Chitnis et al., 2012, Leadbeater and Garber, 2010) show that many people would prefer to die with appropriate care support at home rather than at a hospital, and yet the number of death in hospitals can reach to 65% if there is no appropriate policy in place.

Also, the average cost of hospital care is higher than the social care for older and terminally ill patients. And, the costs of care in the final phases before death are very high in hospitals. Looking further ahead, it is projected that people aged over 85 to almost double by 2030, with an additional 600,000 of the ageing population to need significant care (FCS, 2011). While a quarter of people aged over 65 will need to spend very little in care over their life, half can expect the cost of up to £20,000, and one in 10 can expect the cost of £100,000 (FCS, 2011). According to a recent Nuffield Trust report (Georghiou et al., 2012), the average cost of social care increases with the age of the patient. However, the cost of social care stays cheaper than hospital care for age below 85. Based on gender, the intersection point of the hospital and the social care

costs for male patients are close to age 90 and the female patients are approximately at age 80.

Furthermore, Figure 2.1 presents the risk segmentation for a typical population according to Kaiser pyramid. Although it demonstrates that moderate to high-risk patients represent a very small percentage of the population, their future utilisation is extremely high compared to the majority of the population (Lewis et al., 2011).

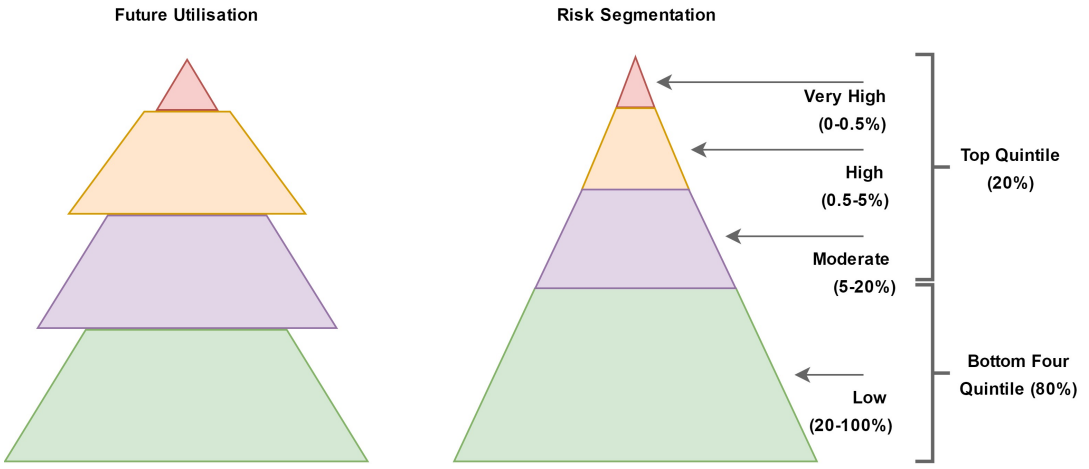

FIGURE 2.1: Risk segmentation and future utilisation

Sometimes a hospital admission can be avoided by residential setting substitution or social care. According to the analysis by Bardsley et al. (2012) on a wide population in England, the use of social care may prevent the need for hospital care. The End-of-Life (EoL) research help patients to get appropriate support services towards EoL by better management of resources and patients. The ambition of the NHS is to increase the number of people who die in their usual place of residence to 60%. This baseline in 2007 was 38%, and with the EoL practices that were in place, it reached to 42% in 2012 (NHS, 2012, 2013d).

Secondly, the ageing population and changes in lifestyles mean an increase in the number of people with long-term conditions or comorbidities. For instance, there have been significant rises in chronic kidney disease, diabetes and cancer between 2006 and 2011 (DH, 2012). Also, it has been predicted that people with comorbidity conditions to rise from 1.9 million in 2009 to 2.9 million in 2018 (Fund, 2013).

Moreover, the time that is spent in poor general health, a limiting chronic health or disability, can be attributed to frailty in some cases. Frailty refers to the condition of being weak and delicate, and it mainly develops as a result of ageing. It is associated with the state of high vulnerability and decreased the ability to sustain homoeostasis,

which is correlated with high risk of adverse outcome including falls, delirium, immobility and disability, incontinence and susceptibility to medications and their side effects (Eeles et al., 2012, BGS, 2014, 2015, Walston et al., 2006). In the UK, the life expectancy is 17.8 years on average for a 65 years old male, of which about 43.3% is in poor general health, and about 41.6% is with a limiting chronic health condition or disability. Similarly, on average a 65-year-old female has a life expectancy of 20.4 years, of which 43.1% will be likely in poor general health and 45.1% with a limiting chronic health condition or disability (NICE, 2016, ONS, 2012).

Thirdly, the rise in the rate of emergency admission to hospitals is another contributing element. Discharging patients is a primary way of providing free beds in healthcare sectors. But if the estimated risks by healthcare administrators and decision support systems are not correct, it may lead to readmission of patients. Patient-flow modelling solutions, like Length-of-Stay (LoS), enable managers to better understand the operational and clinical functions (Adeyemi et al., 2013). The LoS modelling includes capturing the flow of patients from admission to discharge. The flow is through a number of conceptual (virtual) phases that patients go through. The predictive models of LoS use the time spent in phases, in addition to the clinical data and the demographic data, to identify events.

In the following subsections, first, the preventable emergency admissions are defined and discussed. Then, the emergency readmission predictive modelling and the comorbidity risk index modelling are summarised.

2.1 Preventable Emergency Admissions

According to a recent report by the Organisation for Economic Co-operation and Development (OECD), the healthcare spending have fallen in the half of the European Union (EU) countries in real terms[1], including the UK between 2009 and 2012 after about forty years (OECD, 2014). The spending in real terms per-capita was increasing by an average 4.9% per-year over the previous decade in the UK, until 1.3% decline between 2009 and 2012. In general, these declines were due to cuts in workforce and salaries, reductions in fees and pharmaceutical prices and increase in patient co-payments.

Many countries are developing strategies to reduce down avoidable hospital care (OECD, 2014, Nolte and McKee, 2008). Over the last decade, the National Health Service (NHS) in England has been transformed through efficiency savings measures, such

[1]The healthcare expenditure in real terms is the spending after adjustment for inflation.

as the payment reform and quality improvement measures like marginal rate tariffs (Charlesworth et al., 2014). However, increasing demands for emergency admission still remain a major issue. A well-performing healthcare system must be able to provide necessary policies (NICE, 2016) for preventive care. In below, four major policies that are directly related to emergency readmissions are highlighted, and their impacts are discussed.

Firstly, there is sound evidence that the quality of care at the primary care level can reduce down potentially avoidable admissions. One approach is to use admissions of patients with the Ambulatory Care Sensitive Conditions (ACSCs)[2] as a general indicator for optimality assessment of primary care, community services and outpatient care (Ansari et al., 2006, Billings et al., 1993, Purdy et al., 2009a,b). At present, twenty-seven ACSCs are specified in the NHS Outcomes Framework (Bardsley et al., 2013, Blunt, 2013b) as markers of improved health outcomes (Section 2.3.1).

Moreover, ACSCs are identified by experts and do not usually take into account the population and the quality of care. Consequently, they can be misleading and may reduce down the predictive model's accuracy. Therefore, the rate of admission for care sensitive conditions may be adjusted by the characteristics of local population, such as age, deprivations, morbidity levels, area of residence, ethnicity, environmental factors, prevalence rates of diagnosed and undiagnosed based using the Quality and Outcomes Framework (QOF), QOF patient experience and QOF clinical quality of care (Calderón-Larrañaga et al., 2011, HSCIC, 2014c, Purdy et al., 2009b, Sanderson and Dixon, 2000, Tian et al., 2012).

Secondly, the Payment by Results (PbR) strategy is based on the Healthcare Resource Group (HRG) and is likely to be the heart of payment system for the coming decade. Most countries in Europe use a similar system known as case-based payment, which is based on the Diagnosis-Related Group (DRG) Fetter et al. (1980), Mistichelli (1984) and several other metrics (e.g. metrics based on assessments of demands and supplies) to calculate fixed annual budget. Since the introduction of the PbR in the NHS, about a decade ago, the general evaluations have been positive. But, there is no robust evidence on its long-term impact or its health system efficiency (Busse et al., 2011, Charlesworth et al., 2014, Quentin et al., 2011).

Thirdly, the marginal rate was originally introduced to discourage unnecessary emergency admissions. But, according to the recent report by the DoH (DH, 2014), the introduction of the national 30% marginal rate tariff to limit the incentive for increased emergency admission did not meet the costs, and demand still continues to increase.

[2]The Ambulatory Care Sensitive Conditions (ACSCs) is also known as the primary care sensitive condition (PCSCs)

Also, the 2011/12 operating framework proposed that the cost of 30-day readmission should not be reimbursed (DH, 2010); but, the enforcement of this penalty varies across the country (Dowler, 2011).

Finally, it is an ongoing challenge to monitor and manage the financial sustainability of NHS bodies. For instance, lack of transparent reports of income and expenditure can lead to wrong financial performance benchmarks and incorrect efficiency saving achievements (DH, 2014). In the recent report by the DoH (Marshall et al., 2014, DH, 2014) the need for more transparent reports of income and expenditure is planned for NHS Monitor, the NHS England and the NHS Trust Development Authority.

Considering the long-term view, health providers need to address the underlying causes of ACSCs. Purdy et al. (2009b) suggested the following evidence-based interventions for avoidable admissions:

- Implementing disease management and supporting for self-management for patients with long-term conditions;
- Encouraging patient lifestyle change with behavioural change programmes, like telephone health coaching;
- Providing easy access to the urgent care;
- Increasing continuity of care with General Practitioners (GPs);
- Ensuring effective local primary care arrangements.

2.2 Emergency Readmission Prediction Modelling

The UK's DoH in 2005 commissioned to develop the PARR (Billings et al., 2006a, Lewis, 2011) algorithm and the *PARR++* software for PCTs (Fund, 2016a, Lewis et al., 2011). The aim of PARR was to identify individuals at high-risk of emergency re-admission to hospital within a year using the HES's inpatient data. The developed PARR model was very similar to the SPARRA (NHS, 2011) and the PRISM (Dialog, 2008, Hutchings et al., 2013) for Scotland and NHS Wales. Then, the Combined Predictive Model (CPM) was released in 2006 by the DoH from combined general practices (GPs) and HES databases, in order to address the need of identifying the patient risk along the continuum, (DH, 2006).

Thereafter, the DoH commissioned an upgrade in 2011 to the PARR++ and the CPM models of the NHS England (DH, 2011a, Nuffield-Trust, 2012). Patients at Risk of Readmission within 30 days (PARR-30) was developed as the upgrade to be run by acute hospitals. The PARR-30 was based on a broad range of measures used in the PARR (Billings et al., 2012). Next, Billings et al. (2013) released an update to the CPM model in 2013, using a different set of features, with more clear report of performance statistics. The data setting was very similar to the CPM, but each of the sub-models presented moderately good performance and better than the CPM.

A comprehensive literature review regarding the prediction modelling of hospital emergency admission has been carried out, and the full summary is presented in Appendix A.1.1. All the commercial tools except John Hopkin's models are closed source, and it is very hard to evaluate their performance independently. In below, five other important emergency admission models are summarised.

In a recent research by Bottle et al. (2014) several modelling approaches were applied to predict cohort-specific emergency readmission, in addition to other cohort-specific models for prediction of mortality, return to theatre, comorbidity index and outpatient non-attendance. The research used a large extract from the HES and several modelling approaches, including LR, Random Forest (RF), Artificial Neural Network (ANN) and Support Vector Machines (SVM). In general, models have very modest performance across different cohorts using different modelling approaches, including models with time-weighted polynomials features; but, the mortality models have modestly high performance.

The *QAdmissions* score was developed by Hippisley-Cox and Coupland (2013) using two data sources: *QResearch* validation cohort and the Clinical Practice Research Datalink (CPRD) database, to predict the risk of emergency admission within 2-years. The proposed model incorporated thirty features from GPs and HES with the collaboration of 405 general practices across England. The performance of the model using GP data only is moderate and the overall model is moderately high. Also, the model is specialised to the *QResearch* database, and it is very hard to adapt it to other healthcare systems.

The CMS Model, the Hospital-Wide All-Cause Unplanned Readmission Measure, was introduced by YNHHSC/CORE (2012, 2015) to predict 30-day readmission using Medicare data from two datasets between 2007 to 2010. The model was developed using the LR method for a cohort of patients aged above 65. The model adjusts for case-mix differences and service-mix differences, based on a very wide range of inpatient features. Initially, five sub-models based on different speciality cohorts were designed,

and then they were combined. However, the performances of all models are very low or moderate across samples, cohorts and performance tests.

Shortly after the release of the LACE (van Walraven et al., 2010), the LACE+ was developed by van Walraven et al. (2010, 2012) to predict 30-day death or emergency readmission, based on the administrative data from the Ontario population. The LACE+ feature space consists of patient age, sex, discharge method, emergency diagnoses and procedures, alternative cares and count of admission methods, which were selected using a stepwise LR. The performance of the model was moderately high; however, the selected population size was fairly small. Also, one of the most significant features in the model, the *Case-Mix Group* feature, can only be calculated by the Canadian Institute for Health Information data.

In the study by Lyon et al. (2007), hospital emergency readmission to inpatient within 12-month was modelled using the LR model. The Emergency Admission Risk Likelihood Index (EARLI) was designed using inpatient, outpatient and A&E data from the HES database, and GP records, in addition to the mortality records. The HES records from 2002 to 2003 was linked with mortality records in conjunction with data collected from seventeen PCTs, using questionnaire from the patient over 75 years of age. The performance of the developed solution was very moderate and the dataset size was very small.

2.3 Risk Scoring

Risk scoring, in the healthcare domain, relates to a systematic and effective method of identifying risks and predisposing factors that might give rise to a specific event or allow for partial classification. Examples of the application of risk scoring include: identifying patients who are at risk of a heart attack, have unmet needs, represent complex cases or are socially isolated. Furthermore, risk scoring is a useful aid in efficiently identifying and isolating cohorts of patients for which an intervention will be made or for the purposes of stratification and more general analysis of a given patient population.

2.3.1 Comorbidity Risk Index Modelling

Adjustment for comorbidity is common in clinical outcome risk adjustment. The two most common measures (Austin et al., 2012, Baldwin et al., 2006, Khuu et al., 2015, Kuo and Lai, 2010, Lieffers et al., 2011, Sharabiani et al., 2012) are the Charlson Comorbidity Index (CCI) (Charlson et al., 1994, 1987) and the Elixhauser Comorbidity Index (ECI) (Elixhauser et al., 1998), which are used for predicting admission and mortality. The CCI and the ECI calculate the frequency of some comorbidity categories, weight them based on the proportion of expected admission or mortality and then linearly sum them up. There have been revised versions of the CCIs (Deyo et al., 1992, D'Hoore et al., 1993, 1996, Ghali et al., 1996, Quan et al., 2005, Romano et al., 1993) and the ECI (van Walraven et al., 2009), including the most recent adaptation of the CCI (Aylin et al., 2010, DFI, 2010, HSCIC, 2014d) and Bottle et al. (2014) and the actively maintained ECI by the AHRQ (AHRQ, 2016b). In addition, Gagne et al. (2011) introduced a combined version of CCI and ECI indices, and demonstrated that the combined scoring can boost the performance, especially for short-term prediction of mortality and resource usage.

Moreover, the acute conditions in the England NHS can be categorised into three subgroups (HSCIC, 2016k): Ambulatory Care Sensitive Conditions (ACSCs), vaccine-preventable conditions and conditions that usually do not require hospital admission. The ACSCs (Blunt, 2013a, Fund, 2016c, ACI, 2015) are seen as potentially avoidable and are highly correlated to multiple admissions over time and quality of care (HSCIC, 2016k). The use of ACSCs has had some success in order to hold commissioners to account and reduce the emergency admission (Bardsley et al., 2013, OECD, 2014, QualityWatch, 2016).

At present, twenty-seven ACSCs are used in the NHS Outcomes Framework (Bardsley et al., 2013, Blunt, 2013b) as markers of improved health outcomes. Between 2001 and 2013, the patterns of change over time for each ACSC across all the deprivation levels were similar. The standardised rates of admission per 100,000 population for conditions in acute group[3], chronic[4] and other[5] groups changed by +0.49%, -0.03% and +0.47%, respectively.

[3]The acute group of ACSC: Acute conditions, Cellulitis, Dehydration, Dental conditions, Ear, nose and throat infections, Gangrene, Gastroenteritis, Nutritional deficiencies, Pelvic inflammatory disease, Perforated/bleeding ulcer, Urinary tract infection/ pyelonephritis.

[4]The chronic group of ACSC: Chronic conditions, Angina, Asthma, Chronic obstructive pulmonary disease, Congestive heart failure, Convulsions and epilepsy, Diabetes complications, Hypertension, Iron deficiency anaemia.

[5]The other and vaccine-preventable group of ACSC: Influenza, Pneumonia, Tuberculosis, Other vaccine-preventable.

Also, there have been other attempts to classify conditions, such as John Hopkin's Aggregated Diagnosis Groups (ADGs) (JHU, 2014) and Selection of Multipurpose Australian Comorbidity Scoring System (MACSS) (Holman et al., 2005). The ADG clustering method is part of the John Hopkin's ACG system and defines thirty-two clusters of diagnoses. It is used to draw five aspects of morbidity: duration, severity, diagnostic certainty, aetiology and speciality of care. Moreover, the MACSS selected 102 comorbid conditions based on readmission, Length-of-Stay (LoS) or mortality predictability. Based on the validation results on a large population in Australia, MACSS significantly outperformed the CCI.

An alternative approach to comorbidity scoring is to use a cost function, like the UK's HRG (HSCIC, 2016a), and the US's Centers for Medicare and Medicaid Services Hierarchical Condition Categories (CMS-HCC) (Kautter et al., 2014, CMS, 2016b). Commercial implementations of such approaches exist in John Hopkin's ACG system (JHU, 2014) and Verisk Health's DxCG Risk Analytics (Verisk Health, 2016). Also, it has been demonstrated (Billings et al., 2012, Li et al., 2010) that use of cost functions, such as HRG and CMS-HCC, can improve the performance of comorbidity models.

On the other hand, the use of comorbidity scoring in predictive models is sometimes criticised. Firstly, unrepresentative versions of the comorbidity scoring, like the CCI, are being used widely, even though their accuracies have been shown to be sensitive to the time-frame and population settings (Bottle and Aylin, 2011, Bottle et al., 2014, Quan et al., 2005). Also, the coding accuracy of diagnoses, cost groups and validation techniques are another set of important factors (Bardsley et al., 2013, Bottle et al., 2013, Hurst and Williams, 2012). Other criticisms (Bottle and Aylin, 2011, D'Hoore et al., 1996, Quan et al., 2005, Romano et al., 1993) of such scoring methods relate to using very small validation sets and not adjusting for key factors, such as age, gender, deprivations, LoS and temporal patterns.

2.3.2 Operations and Procedures

There is increasing evidence that the quantification of high-risk operations and procedures with adequate adjustment can significantly improve the quality of mortality and readmission models (Aylin et al., 2013, Finks et al., 2011, Jhanji et al., 2008, Symons et al., 2013). Yet, unlike comorbidity, there is no generic risk model for operations and procedures (Barnett and Moonesinghe, 2011, Mehta et al., 2016, Moonesinghe et al., 2013, Rix and Bates, 2007), and the categorisation is typically carried out using

clinical groups. In the UK, the NHS uses Office of Population Censuses and Surveys (OPCS) Classification of Interventions and Procedures (HSCIC, 2016j). And, in the US, the Public Health Service uses the AHRQ's procedure categorisation scheme (AHRQ, 2016a).

Nonetheless, there have been several attempts to define a scoring mechanism for patients with specific conditions, such as the Royal College of Surgeons Charlson Score (Armitage and Van der Meulen, 2010), EuroSCORE (Nashef et al., 1999) and the model is developed by Aktuerk et al. (2016) using HES. It has been shown that such tools can potentially increase measurement effectiveness of a patients' general risk and the risks associated with complications (Keltie et al., 2014).

In contrast to using HRG and CMS-HCC classifications surrounding cost, alternative approaches that are more focused on operations and procedures include (Mehta et al., 2016, Pearse et al., 2006): the Bupa's Operative Severity Score (Bupa, 2016) and the Surgical Outcome Risk Tool (SORT) (NCEPOD, 2011, Protopapa et al., 2014). However, they are constrained by their limited data collection range and very narrow population cohorts.

The Bupa's Operative Severity Score was developed by the UK's largest private medical insurer using the Bupa private major score procedure database (PROM). It provides a range of information about treatment options including benefits, risks, burden and likelihood of success, which has been proven to be successful in producing more information about risks of readmission, mortality as well as the effect of interventions. But, there is a concern about its accuracy, because of the discrete way of scoring risks (Devlin and Appleby, 2010, Protopapa et al., 2014).

Moreover, the SORT was developed by Protopapa et al. (2014) to predict the preoperative risk of 30-day mortality after non-cardiac surgery, using cases from 326 NHS hospitals in England, Wales and Northern Ireland. The model uses LR with forty-five features to predict the risk, including American Society of Anesthesiologists Physical Status (ASA-PS) grade, the urgency of surgery, high-risk surgical speciality, surgical severity, cancer and age. The model performance is comparable (Moonesinghe et al., 2013, Protopapa, 2016) with leading preoperative risk models: the Portsmouth POSSUM (Prytherch et al., 1998) and the Surgical Risk Scale (SRS). The model incorporates a manageable set of features and has a moderately high performance in overall, but more external validation of the model is necessary to prove its resilience.

In the following chapter, the main complexity levels in data extraction and analysis stages are outlined.

Chapter 3

Complexity Levels of Models

Generally, data-driven approach needs to filter results based on statistical significance, importance and novelty, in order to identify significant correlations from electronic health records (EHRs). A true data interpretation needs development and implementation of guidelines and clinical models to allow unambiguous representation of clinical meaning (Jensen et al., 2012).

There are considerable challenges in comparing the predictive models across international boundaries (Lewis et al., 2011, Mihaylova et al., 2011). Four distinctive aspects in the analyses of research studies are: data, produced features, modelling, and performance. In each area, there are several layers of complexity, which influence the quality and interpretability of the models. In the following subsections, these four aspects are presented.

3.1 Data

Three layers of complexity can be defined for the data aspect: availability, quality and pre-processing. In terms of availability, three main areas are presented in below:

- Selected populations: may be subjective to ethical approvals and the concession of patients, and can be limited to terms of use.
- Important variables: may not be available or have different methods of collection.
- Linked sources: may be required to connect databases to query variables from different healthcare sectors, such as inpatient, outpatient, A&E GP, and national statistical registry like mortality.

Furthermore, quality of obtained data varies across several dimensions and is generally very hard to quantify and compare (Bardsley et al., 2013, Bottle et al., 2013):

- Variables: may have different accuracies and recording policies (HSCIC, 2013b, NHS, 2013a, of Physicians, 2006):
  - Events chronology: can have inconsistencies in recording or usage, such as the delays in mortality recording, or risk scores that rely only on the most recent care status and undermine the long-term and chronic conditions.
  - Coding orders: may not be consistent with coding manuals and the chronological order of the patient codes are not usually similar. Therefore, relying only on primary diagnoses or operations codes may not be relevant for very sick patients with complex comorbidities.
  - Unclassifiable events: can be recorded based on different practices. In the ICD-10, abnormal signs and symptoms are known as the R codes. The R codes may be correctly identified in the following *episodes* or *spells*.
  - Local factors: can have profound effects on administration and clinical variables. For instance, regions or hospitals can follow different policies or be constrained under different conditions.
  - Time lags: can be present between care setting changes and can affect the performance or operational applicability of models.
- Time-frames: may have a significant correlation to the national and local care policies and regulations. It is usually possible to keep track of major high-level policies, but minor national and local changes are either unobservable or very hard to track their enforcements across organisations (HSCIC, 2016i)
- Institution characteristics: can largely vary in terms of type, staffing, quality, accidents, resources and registered patients.

Moreover, the data pre-processing stage is one of the main time-consuming stages when dealing with new data sources. After exploratory analysis of the raw data, there are a number of issues that may arise:

- Missing and invalid data: may be excluded or treated by the imputation of values, removing records or replacing. For instance, multiple imputation methods may be used to explicitly impute variables with a substitute. An alternative approach is to calculate a log-likelihood function to estimate the maximum likelihood of variables (Fitzmaurice et al., 2008, O'Kelly and Ratitch, 2014).

- Conversions and aggregations: may be used to convert variables and temporal events to a format that can be processed by the model (Kuhn and Johnson, 2013). For instance, re-categorisation of discrete and continuous variables can be an effective approach to reduce sparsity and to better capture non-linear relationships. Also, it may help training algorithms to converge faster.

Moreover, there are usually some care aspects that are not observable in research problems, which can introduce a degree of bias in the model. For instance, it was speculated (Bottle et al., 2014) that a lot of variations for readmission can be due to the method of delivery of care, which cannot be quantified using the HES database. In addition, extra information, such as physiological and pathology records, might boost the performance of models.

3.1.1 Policies

Any study that is collecting measurements over time is vulnerable to errors caused by the way data is collected and maintained. The majority of literature did not investigate changes in policies, care services and facilities, and the major country-wide policies were mainly mentioned (Bottle et al., 2013). In order to identify shifts in data recording and changes in system's structure two different approaches can be pursued: partially derive the patterns of changes using exploratory analyses, or alternatively record all the changes using reported references.

For the UK healthcare systems, these changes are not recorded in a centralised database, but they may be extracted manually from large sets of document repositories. The HSCIC keeps records of any methodological changes (HSCIC, 2016i), and Department of Health (DoH), which is responsible for the NHS policies, keep records of national policies (DH, 2016), and the local changes only kept by the local authorities.

3.2 Producing Features

The producing features aspect is one of the most time-consuming part of many analysis problems. Five main layers of difficulties are presented in the following:

- Correlated or causal features: must be linked, combined or removed. For instance, a hierarchical or a graphical algorithm may be used to formulate correlated features into the model (Koller and Friedman, 2009, Kuhn and Johnson, 2013).
- Transformations: may be applied to address skewness, kurtosis, sparseness or other theoretical considerations (Kuhn and Johnson, 2013, Mihaylova et al., 2011, Walpole et al., 2014).
- Feature generation methods: may be designed to derive features from raw variables. In addition, specialised risk scoring models, such as comorbidity index, may be used to adjust for more complex features.
- Feature construction: may be used to extract more significant features using a dimension reduction technique, such as filter, wrapper and embedded methods, including Kernel Principal Component Analysis (PCA) (Izenman, 2008, Schölkopf et al., 1997).
- Feature selection: may be used to select the most important features. Methods like randomised Logistic Regression LR and Random Forest (RF) may be used to rank features importance (Yang et al., 2005b).

3.3 Modelling Techniques

The majority of recent emergency admission modelling approaches are limited to LR and Coxian Phase-type Distribution (C-PHD). Although these solutions are simple and powerful, they have limited powers (Appendix A.1.1). In below, the main complexity layers, related to modelling techniques, are presented:

- Assumptions: can excessively simplify the problem or fitting approximation. An example is an inductive bias that is embodied in parametric models, like Linear Regression (Mihaylova et al., 2011, Murphy, 2012), or normality of random errors in regression (Walpole et al., 2014).
- Temporal dimensions: may be introduced to adjust for pre-existing conditions and resource usage. Although well-established methodologies exist, temporal analysis of *longitudinal* healthcare data is still in its early stage (Bellazzi et al., 2011, Jensen et al., 2012).
- Prior probabilities: may be included using an approximation by direct (generative) or indirect (frequentist) inference. Ignoring prior probability of a conditional

event (i.e. base rate fallacy) and inaccurate distribution of a feature can invalidate a model (Murphy, 2012).

- Sampling: may be used to produce samples that can represent the population and are more manageable for the derivation of the fit. If the Normal distribution is assumed despite skewness, excess zeros, multimodality or heavy tails, then the sample should be large enough to guarantee near normality (Mihaylova et al., 2011). If normality is not assumed, the pre-processing and modelling stages must deal with the problem by using methods, such as bivariate zero-inflated Binomial Regression for count data with excess zeros (Wang, 2003) and transformation or skewed distribution for skewed data (Mihaylova et al., 2011).

- Adjustment: may need to include other associated factors. However, use of case-mix adjustment does not necessarily remove all the variability (case-mix fallacy) Lilford et al. (2004).

- Error estimations: must be used as a tool to minimise errors appears in measurements or those that are introduced as extraneous noise. Unlike systematic noises, which are associated with measurement methods, any extraneous noise in the predictors can lead to poor performance. Extraneous noise can originate from measurement type, the inclusion of non-informative predictors and high level of noise in the *response variable* (Kuhn and Johnson, 2013).

- Efficient search methods: may be used in the form of approximation to model fittings. For instance, probabilistic and Bayesian inference methods usually require approximation of expectation, computing marginalisation and normalising constants using deterministic methods (e.g. Maximum a Posteriori (MAP), Laplace's method, Expectation Propagation (EP) (Minka, 2001b), Variational Bayes (Hinton and Van Camp, 1993)) or stochastic (Monte Carlo) (Koller and Friedman, 2009, MacKay, 2003, Neal, 1993, Paquet, 2008, Robert and Casella, 2004).

Furthermore, an appropriate modelling approach must be chosen to account for censored observations (partially known observations) or competing risks (Cook, 2007, Pencina et al., 2008, Steyerberg et al., 2010). Firstly, in survival analysis, a common approach to deal with censored data is to use life-table, which can also be applied to *longitudinal* predictive models in healthcare (Singer and Willett, 2003). Moreover, one of the competing risks is the clinical intervention, which may be triggered for the high-risk patients by clinicians and is not always observable to healthcare models. Examples of clinical interventions are: invasive and non-invasive procedures, cognitive interventions, and social and community cares.

3.3.1 Comorbidity Risk

Capturing high-risk patients using diagnoses can be hard due to coding inaccuracies, incomplete coding of transferred patients, and non-unified coding of comorbidities. Therefore, comorbidity grouping and risk scoring are usually used to determine the very sick patients and those who are in high-risk of mortality, emergency admission or high resource utilisation.

But, comorbidity scores that are outdated (Appendix A.1.1) (Brilleman et al., 2014, Carey et al., 2013, Holman et al., 2005, Mosley, 2013) or performance indicators that are invalidated have been controversial (Fischer et al., 2011). For instance, old versions of the Charlson Comorbidity Index (CCI) (Charlson et al., 1987) are still being used (Bottle and Aylin, 2011, Quan et al., 2005), despite their inaccuracy for the studied years. Following is a list of other criticisms of similar models, such as Charlson (Bottle and Aylin, 2011, D'Hoore et al., 1996, Quan et al., 2005, Romano et al., 1993):

- Training on small cohorts, and neglecting the cohort-specific diagnoses prevalence;
- Using additive risk model to calculate risk for different medical conditions;
- Ignoring important factors, such as emergency admission, complexities, operations and Length-of-Stay (LoS);
- Ignoring temporal patterns and not adjusting for long-term conditions and previous health and care status.

Moreover, the inclusion of comorbidity indices in the models are beneficial, but disregarding groups and population that are being compared can introduce bias, which is known as constant risk fallacy (Nicholl, 2007). An example is stratifying (separating populations) into different risk groups based on demographics or care status using separate models, hierarchical models or case-mix models.

3.4 Testing, Validating and Benchmarking

The final hurdle is interpretation and comparison of tests, validations and benchmarks:

- Sampling: must be used to extract population-representative samples for testing, cross-validation, validation and benchmarking.

- Performance statistics: must be applied using different methodologies to validate the effectiveness of the model from several aspects.
- Sub-populations: may be defined to examine strengths and weaknesses of the model for different groups of patients.

The curve of true positive versus false positive, a.k.a. the Receiver-Operating Characteristic (ROC) curve, is a popular tool for benchmarking binary classification. One advantage of the ROC is that it can visualise the performance of a classifier disregarding the class distributions or error cost (Fawcett, 2006). The Area Under the Curve (AUC) of ROC (c-statistics) provides an average sensitivity score weighted by derivative of the specificity (Cook, 2007, Pencina et al., 2008, Steyerberg et al., 2010).

However, only using the AUC of ROC is not sufficient for comparison. Naive calculation of the ROC curve for survival analysis can often lead to misestimating, because of inattention to censored observations or competing risks. Moreover, when two ROC curves cross over, then curves can become superior for some ranges only and consequently, the AUC of ROC interpretation will become subjective (Cook, 2007, Hand, 2009, Pencina et al., 2008, Steyerberg et al., 2010, Webb, 2003).

Alternatively, the performance statistics for several desirable cut-off thresholds may be selected for comparison (e.g. 50% and 75% cut-off points and the top 5% of patients in risk). The thresholds might be specified based on expert judgements, or instead, models may be optimised using a cost or profit function (e.g. intervention cost) to strike a balance between specificity[1] and sensitivity[2]. Also, they may be accompanied by other measures, including sensitivity, precision[3], specificity, F1-score[4], and log-loss[5].

In addition, other comparison approaches might be required in benchmarking, such as risk-band (e.g. 20 risk-bands from 0%-100%) and top risk-segment (e.g. top 1000 high-risks) performance comparisons. Another approach is adopting a reclassification metric, like Net Reclassification Improvement (NRI) (Pencina et al., 2008), to investigate the role of independent variables in the performance, with some precautions (Pepe et al., 2014). Another alternative metric is the invariant AUC method that was proposed by Hand (2009), which is based on a common belief distribution for all test samples.

In the next chapter, the main modelling methods that are used throughout the development phases are described.

[1] Specificity, a.k.a. true negative rate.
[2] Sensitivity, a.k.a. recall or true positive rate.
[3] Precision, a.k.a. positive predictive value.
[4] F1-score (a.k.a. F-measure) is the weighted average of the precision and sensitivity
[5] Log Loss represents the accuracy of a classifier by penalising false classifications.

Chapter 4

Modelling Approaches

In this chapter, the algorithms and modelling approaches that are used throughout the research are described. Firstly, the predictive modelling and risk adjustment are explained. Then, seven modelling approaches are specified in detail, including *Transfer Learning*, Ensemble learning, Logistic Regression (LR), Decision Trees, Random Forest (RF), Support Vector Machine (SVM), Bayesian Methods, Bayes Point Machine (BPM), Deep Neural Network (DNN), and Wide and Deep Neural Network (WDNN).

4.1 Predictive Risk Modelling

In healthcare sectors, many types of scoring systems are used to support clinical decisions, such as Glasgow Coma Scale for patients with brain injuries. However, statistical and stochastic models are needed to estimate the risks according to changes in care and environmental variables (Stedman, 2010).

Data mining techniques can help to predict risks, in healthcare problems, improve the health status of high-risk patients and consequently make overall savings. There are various predictive risk models in the literature, and each can forecast a small range of healthcare and social care outcomes. They differ in terms of the predicted time range, variables, data sources and modelling approaches (Lewis et al., 2011).

There are two major branches of risk modelling in healthcare: predictive modelling and risk adjustment. The predictive modelling is frequently used for high-risk member *case finding*, like finding patients with high-risk of readmission and predicting costs and utilisation. While, the risk adjustment is a normalisation technique for comparison

purposes, such as classifying patients by potential risk level for the purpose of insurance providers' reimbursement (Holmes and Jain, 2012, Lewis et al., 2011).

Moreover, the identification of emergent risk can be categorised into modelling of three main aspects: stratification, clinical profiles and resource utilisation profiles. Also, in the modelling of the events, the time dimensions can be designed as time-to-event models (Appendix A.1.4) or as risk score models.

4.1.1 Predictive Modelling

Predictive modelling is directly associated with machine learning[1], pattern recognition and data mining. The practice of predictive modelling defines the process of development of models that their prediction accuracy can be understood and quantified (Kuhn and Johnson, 2013). Geisser (1993) defines predictive modelling as "the process by which a model is created or chosen to try to best predict the probability of an outcome."

Physicians are interested in evaluating and forecasting adverse events that may provoke mortality or longer hospital stay for the patient, and assign a quantity to the patient risk profile (Cornalba, 2009).Regarding risk impact, healthcare risk analysis can be categorised into two categories: Operational Risks (ORs) and Clinical Risks (Kohn et al., 2000). The predictive modelling of ORs in healthcare modelling, such as emergency readmission (Appendix A.1.1), Length-of-Stay (LoS) and End-of-Life (EoL) (Appendix A.1.3) modelling, varies across systems and often lacks robustness and generalisation.

A popular OR analysis approach in financial modelling problems is loss-event risk modelling using Bayesian Networks (BNs) (Fenton and Neil, 2012). The BN approach is an ideal choice since it is great at identifying common causes of failures that affect the whole trading process. This type of OR analysis tries to quantify the OR that affects the system as a whole, to identify the routes and causes. Another advantage of this approach is that it enables stress testing the system to determine the effects. Two examples of OR analysis in loss-event risk modelling are the identification of rogue trading and stress testing financial markets.

[1]Note that predictive modelling is different notion from Predictive or Supervised Learning approach in machine learning.

4.1.2 Risk Adjustment

Although the medical advances have contributed to the improvement of life expectancy, they have little to do with life expectancy and much more to do with life quality. Risk adjustment methods are either used directly by health insurers for selection of good (profitable) risks from an insurer pool or indirectly by designing insurance products. The models are often based on a linear utility function framework (Newhouse, 1996), and the objective is to minimise the outcome (i.e. risk) (Culyer and Newhouse, 2000, Culyer et al., 2012).

4.2 Modelling Techniques

Since the late 1980s, machine learning methods have been used in extending the statistical analysis for making inferences from data, and there are a lot to be done in the area of automated methods for learning and forecasting in healthcare.

Based on the knowledge of interest, BN, Artificial Neural Network (ANN), Decision Tree (DT) and kernel methods, like Support Vector Machine (SVM) and Gaussian Processes, are often used in healthcare data mining problems (Bardsley, 2012, Kansagara et al., 2011, Lewis et al., 2011, ACI, 2014, DH, 2011a, Paton et al., 2014). Other approaches in machine learning can be found in the work of Bishop and Nasrabadi (2006).

Predictive models vary in terms of prediction time-window (*time-horizon*), selected population, input variables, algorithm design and benchmarking methods. A list of good practices is proposed by Sinha et al. (2013), which covers a number of issues. However, not all studies (Lewis et al., 2011) clearly specify the details of analyses, including publicly and privately funded projects (Appendix A.1.1).

This chapter is divided into six sections. Firstly, a brief introduction to *Transfer Learning* is provided. Then, Ensemble learning is discussed in detail. Afterwards, regression modelling and Logistic Regression (LR) approaches are briefly summarised. Next, major Decision Tree modelling algorithms are outlined, including the Random Forest (RF). Moreover, a recap of the SVM models is presented. Furthermore, Bayesian approaches are reviewed. After that, an abstract introduction to the Deep Neural Networks (DNNs) is provided. Then the Bayes Point Machines (BPM) modelling approach is defined. Later, Deep learning approach and Wide and Deep Neural Network

(WDNN) model are described extensively. Finally, outlines of some other major modelling techniques are given in Appendix A.1.2.

4.2.1 Transfer Learning

The *Transfer Learning* (Woodworth and Thorndike, 1901) is a wide area of research in machine learning (Pratt et al., 1993) that focuses on improvement of the learning through the transfer of knowledge from sub-models or inputs that are learnt. The *Transfer Learning* refers to methods that harness and adapt models to a specific new predictive task at hand. The *Transfer Learning* is also known as *Multi-Task Learning* (Caruana, 1998) or *Learning to Learn* (Thrun and Pratt, 1997), and it refers to fitting many related models to get better performance.

Transfer Learning methodologies can help to use forecasting and predictive modelling techniques to provide a systematic methodology of analysis for similar cases with a smaller number of visible parameters. This may also be extended to perform a semi-supervised machine learning modelling such as active learning (semi-supervised machine learning) or latent feature modelling (Ghahramani et al., 2007) for use in complex, real-world settings (Graham et al., 2011, Horvitz, 2010, Koller and Friedman, 2009).

The main application of the *Transfer Learning* is in domain adaptation. Examples of domain adaptation problems are spam filtering, news analysis and many other personalised classifiers, or models that transfer the learnt features to another problem.

In this chapter, Ensemble learning is discussed, which is a partial subset of *Multi-Task Learning*. In addition, DNNs are briefly overviewed. Because of recent breakthroughs in graphics hardware (Oh and Jung, 2004), accelerated computing (Weber et al., 2011) and backpropagation optimisation (Hinton, 2007) allowed DNNs to become one of the most powerful tools in the *Transfer Learning* domain (Yosinski et al., 2014).

4.2.2 Ensemble Learning

The Ensemble learning approaches (Dasarathy and Sheela, 1979, Hansen and Salamon, 1990, Schapire, 1990) are used in statistics and machine learning techniques to combine multiple learning algorithms to achieve a better performance. Ensemble methods have been applied or integrated within a wide range of modelling techniques, including

ANNs, Decision Trees, and unsupervised learning scenarios, like anomaly detection. Some of the common Ensemble algorithms are Bagging, Boosting and various Bayesian methods (Murphy, 2012, Rokach, 2010, Sammut and Webb, 2011, Sewell, 2008, Zhou, 2012).

Firstly, the Bagging method (Breiman, 1996) stands for bootstrap aggregating, and it combines classifications of randomly generated training sets to decrease the error and improve the classification. Firstly, the algorithm uses bootstrap distribution for generating different base learners (Efron and Tibshirani, 1994). Then, it applies a popular combination method, known as Voting, in order to aggregate the output of learners. For example, Smedira et al. (2013) used a Bagging method to enhance the stability of the multivariate analysis of a non-proportional hazard hospital readmission model. The Bagging approach helped to increase the stability of the model, to be able to analyse the association between readmission, resource use and mortality. However, the studied population was very small and isolated, and the presented performance benchmark was subjective.

The Boosting method (Schapire and Freund, 2012) can reduce the variance of probability estimates, by averaging together many estimates. In another word, the models in the Ensemble modelling space try to correct weaker ones by focusing on the mistaken cases. AdaBoost method (Freund et al., 1996) is an extension of Boosting with many variations (e.g. M1, M2 and R algorithms), which allows it to be implemented on multi-class problems and regression problems. For instance, Turgeman and May (2016) applied a boosted Decision Tree in combination with an SVM algorithm to model hospital readmission. The model was tested on a dataset from veteran hospitals in a city in the USA. The model performed considerably better than other basic models, including LR, SVM and Decision Tree. But, the applied optimisation approach had a moderate performance.

Bayesian methods, like Bayes Optimal Classifier (BOC), Bayes Model Averaging (BMA) and Bayesian Model Combination (BMC) can be used to include hypotheses from the hypothesis space and the associated prior probabilities. For instance, Monteith et al. (2011) demonstrated that BMC provides a theoretical basis for soft-selecting from a space of Ensemble models. The model was applied to a machine learning dataset, and it was shown that BMC could outperform BMA, Bagging, and Boosting, in terms of prediction accuracy.

However, two major disadvantages of Ensemble methods are moderately high computing resource usage, and difficulty in interpretability. Firstly, the computing resources have been improved significantly in the past decade, and an Ensemble model with a

moderate number of sub-models can run very quickly with comparable prediction performance. Secondly, there are various post-processing techniques that can be applied to interpret the models. The partial dependence plot (Goldstein et al., 2015) and importance rankings (Breiman et al., 1984, Chen and Lin, 2006) of features are two generic approaches that can be used to interpret a black-box method (Murphy, 2012).

In the following subsection, a number of approaches for combining and selecting sub-models in Ensemble modelling are discussed.

4.2.2.1 Combination Methods

In the final level of Ensemble modelling, a combination method must be applied to include the estimated probability of all the sub-models in Ensemble modelling space. The popular methods for combining a set of models are Voting, Stacking, Sum, Median, Mean, Product, Mixture of Experts (MxE) and finally using weighting in combination with other methods (Murphy, 2012, Sammut and Webb, 2011, Sewell, 2008, Zhou, 2012).

Firstly, Voting algorithms use a selection approach, like majority, soft averaging and weighted combination of estimates to combine models. Voting algorithms are applied in Bagging and Boosting algorithms and many classification algorithms, such as the Random Forest (RF).

Moreover, the Stacking method (Wolpert, 1992) (a.k.a. Stacking Generalisation) uses the produced estimated probability from a combination of sub-models as an additional input to the main prediction model.

Furthermore, the weighting approach is used in Ensemble methods, like Weighted-Average, Weighted Voting and Bayesian methods, like BOC. The weights are usually derived using an approximation technique, like Expectation Maximisation (EM), to optimise a performance indicator.

Moreover, the MxE algorithm (Jacobs et al., 1991, Jordan and Jacobs, 1994) generates a group of sub-classifiers (i.e. Experts) whose outputs are combined and inputted into a Generalised Linear Model (GLM). The inputted classifiers to the GLM are weighted by a gating function using a method like EM. The MxE is particularly useful when the feature space is heterogeneous, and classifiers on different parts of the space provide more informative and synthetic estimates.

For instance, Liu et al. (2014) proposed an Ensemble model based on MxE to predict risk scores for acute cardiac complications. The developed MxE predictive model incorporates multiple sources of features and the weights of experts are defined using a hybrid method. The model was developed using a small sample of cardiac patients in the Singapore, and the performance of the model was fairly high based on a small population in Singapore.

4.2.3 Logistic Regression (LR)

Before 1980, almost all learning methods were learnt linear surfaces. Linear Regression modelling methods, such as the Logistic Regression (LR) and mixed models have been applied extensively in previous literature in social science and healthcare modelling.

The LR (Cox, 1958, Walker and Duncan, 1967) method is similar to Linear Regression methods, but it has been developed for binary linear classification. For the LR, the observed variable has Bernoulli distribution (Uspensky, 1937) instead of Gaussian (Feller, 2008), and the estimated *response variable* is passed through a Sigmoid function (i.e. Logistic or Logit) to squash the estimates between zero and one. Moreover, to fit the LR, there is a wide range of estimation and optimisation algorithms. One of the popular methods is the Maximum Likelihood Estimation (MLE), which is the same as minimising cross-entropy. A method like the MLE suffers from overfitting and is sensitive to sparse features. The algorithm of a basic logistic regression model can be represented as the following (Eq. 4.1):

$$\hat{f}(x) = \frac{1}{1 + e^{(a+bx)}} \tag{4.1}$$

, where $\hat{f}$ is the prediction of the dependent variable for a vector of data points x. $\frac{1}{1+e^t}$ represents a Sigmoid function, and a and b are the coefficients and the error term is implicit.

To overcome overfitting, *L2* regularisation (a.k.a. weight decay) may be applied to sparse models with a large number of features, and *L1* regularisation may be applied to sparse models with a small number of features. Regularisation in statistics is an effective approach to favour simpler models (Blumer et al., 1987), which can work very well with a large amount of data to reduce overfitting (Halevy et al., 2009). However, when the dataset is small or more personalised results are required, then more complex approaches are needed. For instance, a *Multi-Task Learning* or an Ensemble modelling

methods may be used to include multiple classifiers (Section 4.2.1) and create more specialised or personalised solutions.

The Linear Regression modelling is a well-understood approach with a very broad range of algorithms. A brief summary of other major regression algorithm is provided in Appendix A.1.2. For example, Demir et al. (2009) presented a predictive model for emergency readmission to hospital using HES database. It evaluated the use of LR with a simple transition model to incorporate patients' history of readmission and other covariates. The research focused on Chronic Obstructive Pulmonary Disease (COPD) patients that are admitted to the England's hospitals during a 7-year period. Factors, including demographics, admission events and Length-of-Stay (LoS) were included, and ultimately the performance was compared using ROC. The research demonstrated that use of an only administrative database and a simple phase-type distribution could effectively predict the risk; however, it was designed for a very specific cohort.

Wennberg et al. (2006) developed the LR to predict hospital emergency readmission, similar to other NHS's models (Lewis, 2011, DH, 2006, Nuffield-Trust, 2012). The developed model, Combined Predictive Model (CPM), takes advantage of variables from inpatient, outpatient, A&E and GP from five PCTs. However, very little performance statistics were reported for the model. The update of CPM (Billings et al., 2013) was published in 2013 and reported a modestly high ROC for the model by including data from all the four care sectors, but it had a very weak true positive rate (sensitivity).

Howell et al. (2009) used a multivariate LR to predict hospital readmission within 12-month for patients with chronic medical conditions, using Queensland Hospital data in Australia. The model includes demographics, socioeconomic status, geographic remoteness, comorbidities and previous care utilisation. The model has a modest performance and very narrow focus.

Demir (2014) presented a comparison between Decision Trees, LR, GAM and MARS for predicting hospital readmission for a PCT in the UK. The benchmark shows that for this particular population and very narrow specification, the LR came first and others came very close.

In addition, one application of regression modelling is in the pathway modelling (i.e. the factors that arise from heterogeneity amongst patients) of the EoL (Appendix A.1.3) and frailty function modelling. For instance, a multinomial Logit model was developed by Adeyemi and Chaussalet (2009) for modelling COPD patients' pathway. In the model, the patient frailties were regarded as mixed effect type, and the random effects distributions were modelled based on patient pathways. The model was successful in identifying the high probability pathways for survival and cost objective functions, but

must be tested on other cohorts of patients.

4.2.4 Decision Trees

In 1980's, Decision Trees allowed efficient learning of nonlinear decision surfaces. Decision Trees in predictive modelling are defined by recursively partitioning the *input space*, and defining a sub-model in each resulting region of *input space* (Murphy, 2012, Zhou, 2012). There are many algorithms for Decision Tree, with specific criteria for building and training, including C4.5 by Quinlan (1986) and its commercial successor C5.0 (Quinlan, 2014), Classification And Regression Tree (CART) by Breiman et al. (1984) and Multivariate Adaptive Regression Splines (MARS) by Friedman (1991).

Decision Trees are popular, because of the ease of interpretability, ability to handle discrete and continuous features, insensitiveness to the monotone transformation of features, automated feature importance ranking, robustness to outliers and scalability in terms of number features and observations.

One of the main disadvantages of Decision Trees is that a small change in the distribution of top features in the tree can have a large effect on the model. The algorithms that use greedy search to find optimal tree usually have lower accuracy than any other kind of algorithm, which uses a more sophisticated optimisation algorithm, such as the hierarchical MxE (Murphy, 2012, Zhou, 2012).

For example, Austin (2007) compares CART, Generalised Additive Model (GAM), MARS and Logistic Regression (LR) for prediction of mortality after Acute Myocardial Infraction (AMI) hospitalisation, using Ontario hospital data. All the models demonstrated very close prediction performance, except the CART model, which suffered due to its algorithmic inability to incorporate complex or non-piecewise linear relationships.

In the following subsection, Random Forest (RF) classifier is outlined, which is an ensemble of Decision Trees, and provides a solution to the high variance in the estimations.

4.2.4.1 Random Forest (RF)

The Random Forest (RF) is an Ensemble Decision Tree, which was first introduced by Breiman (2001), and is based on the CART algorithm (Breiman et al., 1984) and the Bagging Ensemble method (Breiman, 1996). To reduce the correlation between the

classifiers, Breiman (2001) algorithm implements a technique to decorrelate the base learning trees based on a random feature selection.

Moreover, the Breiman RF is sensitive to highly correlated features (i.e. correlation bias), and the scale or the number of categories of features. Although it can produce very good fit for the data, the RF feature importance predictions must be treated with caution (Strobl et al., 2007, Toloşi and Lengauer, 2011). The *cForest* algorithm proposed by Hothorn et al. (2006), is an alternative to the original RF. It is based on conditional inference trees and reduces selection bias with a much higher computation burden. Therefore, if the original RF is going to be used, the input features must be pre-processed by a correlation analysis and a feature transformation approach, in order to guarantee unbiased and reliable feature importance ranking (Figure 4.1).

For a given training set with $\{(X_i, Y_i)\}_{i=1}^{n}$ as input features and *response variables*, the Bagging part of the RF model can be represented as the following (Eq. 4.2).

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^{B} f_b(x') \tag{4.2}$$

, where Bagging has been carried out for B number of times, and random sampling with replacement at each iteration. The $f_b(x')$ represents a trained tree for unseen sample x'. In addition, in the random forest at each candidate split in trees, a random subset of the features is used.

For instance, Zheng et al. (2015) benchmarks the LACE score (van Walraven et al., 2010). The study used a RF, a particle swarm optimisation based SVM and a Radial basis function ANN, for predicting hospital readmission using a small sample of heart failure (HF) patients. The presented statistics indicate that the SVM outperformed the rest, but with a very steep computation cost. Additionally, the RF was in the second place with high accuracy and sensitivity.

Moreover, RF is one of the most accurate learning algorithm, which can efficiently handle missing observations well. However, the main disadvantage of RF, as many other techniques, is its algorithmic weakness in dealing with noisy classification problems. Also, when the number of observations is lower than the number of features or order of problem's convolutional structures, RF can over-fit and under-perform.

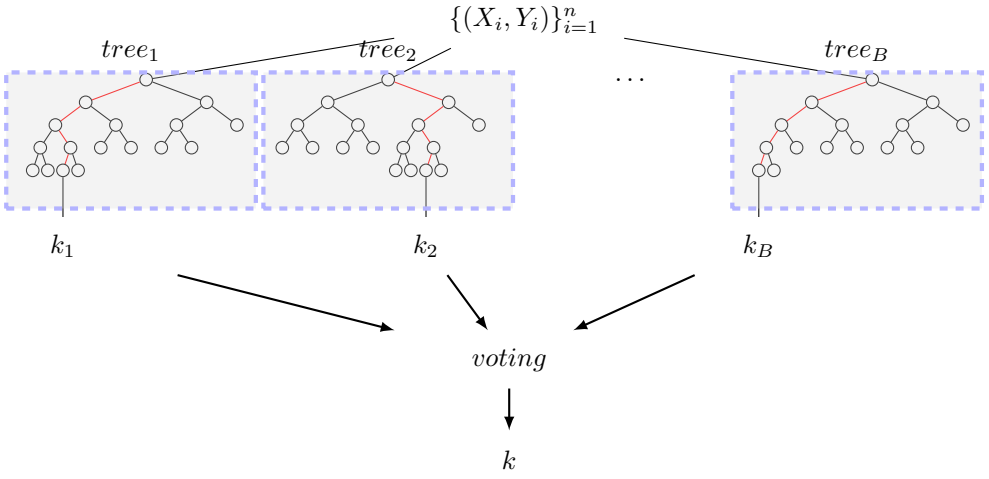

FIGURE 4.1: An abstract representation of the Random Forest

4.2.5 Support Vector Machine (SVM)

In 1990's and 2000's, efficient learning algorithms for non-linear functions based on computational learning theory were developed, including efficient separability of non-linear regions using kernel functions, quadratic optimisations and better optimisation algorithms rather than greedy search.

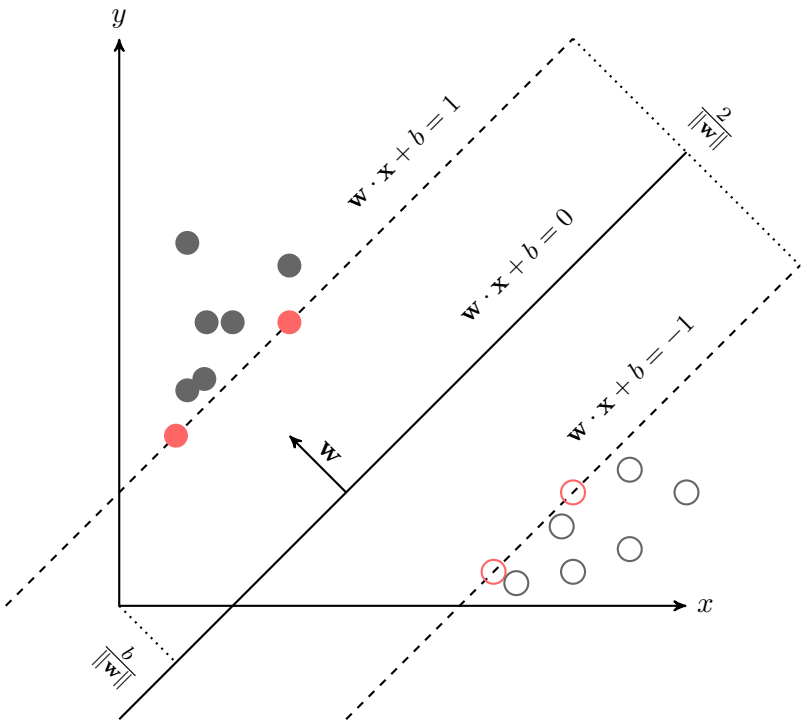

FIGURE 4.2: Separating hyperplane in the Support Vector Machine (SVM)

The Support Vector Machine (SVM) algorithm searches for an optimal hyperplane for linearly separating patterns and it extends to patterns that are not linearly separable.

SVM can take advantage of Kernel Trick (Hofmann et al., 2008) to transform input data to be mapped into a new space (kernel space). The support vector refers to the data points that lie closest to the decision surface (hyperplane). There are many possible ways to find a separating hyperplane (Figure 4.2) and the SVM tries to find an optimum solution, unlike a method such as Artificial Neural Network (ANN). A hyperplane may be represented as the following (Eq. 4.3):

$$f(x) = \sum_{i=1}^{n} y_i \alpha_i K(x, x_i) + b \tag{4.3}$$

, where x_i represents the input data from a sample of size n, and y_i is the *response variable* of input i. $K(x, x_i)$ is a kernel function, which a non-linear mapping function from the input space, and α is a Lagrange multiplier, which is non-zero if input data belongs to the support vector. The following demonstrate the decision function ($sgn(f(x))$) using the Kernel Trick (Eq. 4.4):

$$sgn(f(x)) = sgn(wx + b) = sgn(\sum_{i=1}^{n} y_i \alpha_i K(x, x_i) + b) \tag{4.4}$$

, where sgn is the *Signum* function that extracts the sign of the estimate where $y_i \in \{-1, 1\}$.

For instance, Yu et al. (2015) modelled a 30-day readmission model with adjustment for hospitals, and then compared a linear SVM, a non-linear SVM and a Cox regression against the LACE score. The model used Medicare inpatients 65+ years from the general US population, with a moderately small dataset. All the proposed models performed better, in comparison with the LACE. However, the two more complex models, the non-linear SVM and the Cox regression, failed to perform better than the linear SVM.

Moreover, to evaluate the importance of features, Guyon et al. (2002) proposed Support Vector Machine Recursive Feature Extraction (SVM-RFE) with w_i^2 as the ranking criterion. The SVM-RFE method uses a weighted sum of support vectors to rank features by importance.

The advantage of the SVM methods against Linear Regression like LR is that it can classify non-linearly classifiable problems. However, the major downside of the SVM is its proportional complexity to the number of input data. But, the number of features has very little effect on the complexity of the optimisation.

4.2.6 Bayesian Methods

There are two main approaches for incorporating stochastic models into the statistical modelling: discriminative (conditional distribution model) and generative (joint probability model). Firstly, the discriminative modelling does not make any assumption about the prior distribution and only includes the conditional probability. Therefore, discriminative modelling is also known as the frequentist approach, and linear classifiers, like LR, are examples of it. On the other hand, Bayes modelling (Bayes and Price, 1763) methods are known as generative, and they include the prior (marginal) distribution of the evidence data to the discriminative model (Jordan, 2002).

Under the frequentist approach, unknown parameters are considered to have fixed but unknown values (i.e. fixed priors). The unknown parameters might be calculated by maximisation of total marginal likelihood, or be estimated using methods, such as Maximum Likelihood Estimate (MLE) or numerical integral approximations like Laplace approximation. Moreover, the frequentist inference of the posterior probability distribution can be interpreted as procedures that guarantee long-run frequency. This inference is derived from procedures that guarantee probability within a random confidence interval (Berger et al., 2006, Gill, 2014, Koller and Friedman, 2009).

In contrast, a Bayesian approach considers all parameters to be random variables (i.e. functions of the data), and the data is used to update the prior probabilities of these parameters. In this approach, the Bayesian inference of posterior probability distribution is used for setting and updating beliefs. The computation methods of Bayesian priors can be categorised into three distinct groups (Berger et al., 2006, Gelman et al., 2013, Gill, 2014, Koller and Friedman, 2009, Lunn et al., 2012, Press, 2009):

- Subjective Bayesians: uses personal degrees of belief. It is applying informative priors, based on historical data or underlying theory.
- Objective Bayesians: uses non-informative priors (a.k.a. objective, diffuse, flat or reference priors). It is applying prior distributions that are formally expressing ignorance (vague information); but, have well-defined posterior probability distributions[2]:
  - Using conjugate priors to approximate uninformative priors.
  - Using prior distributions that can span the range of likelihoods, such as flat prior or Gaussian prior.

[2]It is especially useful for complex problems with many parameters that have very little amount of information about the data. However, for many problems, this approach can be misguided or have no clear choice of prior distributions and inference approximation.

  - Defining priors that are transformation invariant based on Jeffreys' Prior (Jeffreys, 1946).

- Empirical Bayesian: Using data to estimate the prior.

Before doing the Bayesian inference, it is useful for comparison to ignore some information and do a crude estimation of the missing data. However, ultimately inference of missing data should be included as part of the model. The Bayesian inference approximation can be categorised into two distinct groups: deterministic and stochastic. In below, some of the well-known inference approaches are highlighted (Barber, 2012, Gelman et al., 2013, Koller and Friedman, 2009):

- Deterministic approximation:
  - Laplace approximation: finds the Gaussian approximation to a probability density, which is based on the second-order Taylor approximation of the log posterior around the Maximum-a-Posteriori (MAP) (Bishop and Nasrabadi, 2006).
  - Expectation Propagation (EP): is an iterative approach for choosing the best approximation from within some tractable class of distributions (Minka, 2001c).
  - Loopy belief propagation: is a dynamic programming approach, which calculates the marginal distributions for unobserved nodes, conditional on any observed nodes (Murphy et al., 1999).
  - Expectation Maximisation (EM): is an iterative method for finding the maximum likelihood or MAP estimates, where the model depends on unobservable variables (Bailey et al., 1994).
  - Variational Bayesian methods: is an extension of EM algorithm from MAP, and it finds a set of optimal parameter values based on a set of interlocked equations (Bernardo et al., 2003).
- Stochastic Approximation:
  - Direct simulation: can be used for simulation of simple models, and it is often easy to draw from the posterior distribution (e.g. rejection sampling, univariate sampling, and multivariate sampling).
  - Markov Chain (MC) simulation, a.k.a. Markov Chain Monte Carlo (MCMC): is a general method for drawing sequential samples, which distribution of

sampled draws depends on the last value drawn. Gibbs sampler and Metropolis sampling are two examples of the Markov-based sampling algorithm (Gilks, 2005, Koller and Friedman, 2009).

Moreover, in Bayesian Network (BN) modelling, template-based representations are used to produce a single compact model that can represent properties of system dynamics and to produce distribution over different trajectories (e.g. Dynamic Bayesian Network) or to produce a distribution over different worlds (e.g. Genetics networks). To be able to reason about non-static situations, Dynamic Bayesian Network (DBNs) (Dean and Kanazawa, 1989) are used to represent nodes with *system states*. The *system states* are either considered as stationary time-slices (homogeneous or invariant), like Markov Models, or they are regarded as the state observation model, like Hidden Markov Models (HMMs). In state observation models, the states are variant and evolve on their own separately from the observations (Koller and Friedman, 2009). Five principal methods have identified to model a Time-Varying DBN and are presented in Appendix A.1.5.

In the following subsection, Bayes Point Machines (BPM) is summarised, which is a generative approach for non-linear classification. Moreover, before the model development stage, a list of suitable Bayesian libraries are produced for the purpose of this research, that is presented in Appendix A.3.

4.2.6.1 Bayes Point Machine (BPM)

The Bayes Point Machine (BPM) (Herbrich et al., 2001, Minka, 2001b) is a type of nonlinear classification algorithm that identifies an average classifier known as a Bayes point in a version space. A version space can be defined as a set of hypotheses, each of which is an approximation of the main hypothesis class. Similar to SVMs, BPMs are more geometrically motivated and are aimed to find a hyperplane with optimal margins between classes. In contrast, logistic regression maximises the probability of data by optimising the distance of each point to the decision boundary.

The soft margin SVM can be thought of as an approximation to the BPM (Herbrich et al., 2001). SVMs (Vapnik and Vapnik, 1998) use a mapping to indirectly transform data into higher dimensional space using a kernel function. Then, they use quadratic programming to optimise the classification's hyperplanes using support vectors and margins. However, SVMs are only efficient for a symmetric version space and its complexity is characterised by the number of support vectors.

On the other hand, BPMs sample the Bayesian posterior (Eq. 4.5) for a nonlinear classification in a kernel space. Then, they approximate the centre of the version space, which is a set of consistent hypothesis, and the effective size is determined from the training sample. BPMs minimise the generalisation error over a set of hypotheses according to a prior probability, instead of maximising the classification boundary margin explicitly, as SVMs do. The predictive distribution can be thought of as a linear discriminant function, which is assumed to have the following parametric density:

$$p(y|x,w) = p(y|s = w^T x) \tag{4.5}$$

, where w is the weight or latent parameter vector, x is the fully observed feature vector, and s is the score function. BPMs use the kernel trick (Hofmann et al., 2008) to find an optimised w. The centre mass of the version space is approximated using an average of the weight vectors while minimising the average generalisation error. The derived scores are subject to additive Gaussian noise (ε) to allow for measurement or *labelling* errors (Eq. 4.6).

$$\begin{aligned} &p(y|s,\varepsilon) = (ys + \varepsilon > 0)1 \\ &\quad \text{, with } p(\varepsilon) = N(\varepsilon|0,1) \ \wedge \ 1(\alpha > 0) = \begin{cases} 1 & if\ \alpha > 0 \\ 0 & if\ \alpha \leq 0 \end{cases} \end{aligned} \tag{4.6}$$

In this research, Microsoft's Infer.Net library (Research, 2016) was used to construct the BPM model (Figure 4.3). The applied algorithm uses the original version of the BPM, with two main modifications. Firstly, it uses a mixture of Gamma-Gamma, a heavy-tailed prior probability distribution for the precision of weights and features. Secondly, it applies Expectation Propagation (EP) message passing to infer posterior probabilities, which has been demonstrated (Minka, 2001a,b) in Gaussian Mixture problems to be better than approximation techniques.

Therefore, the applied BPM is invariant to parameter rescaling or shifting, unlike LR or SVM methods. Moreover, active Bayesian training can allow continuous updates of the model and account for changes in the prior probabilities. Furthermore, the BPM can efficiently handle a relatively larger number of features.

For instance, (Tan et al., 2008) applied a BPM model in combination with a Hidden Markov Model (HMM) to analyse immunological data of Asthma patients from a hospital in the UK. The research has provided a basic proof of concept for the analysis of large-scale immunological datasets. But, the study provided very little detail about

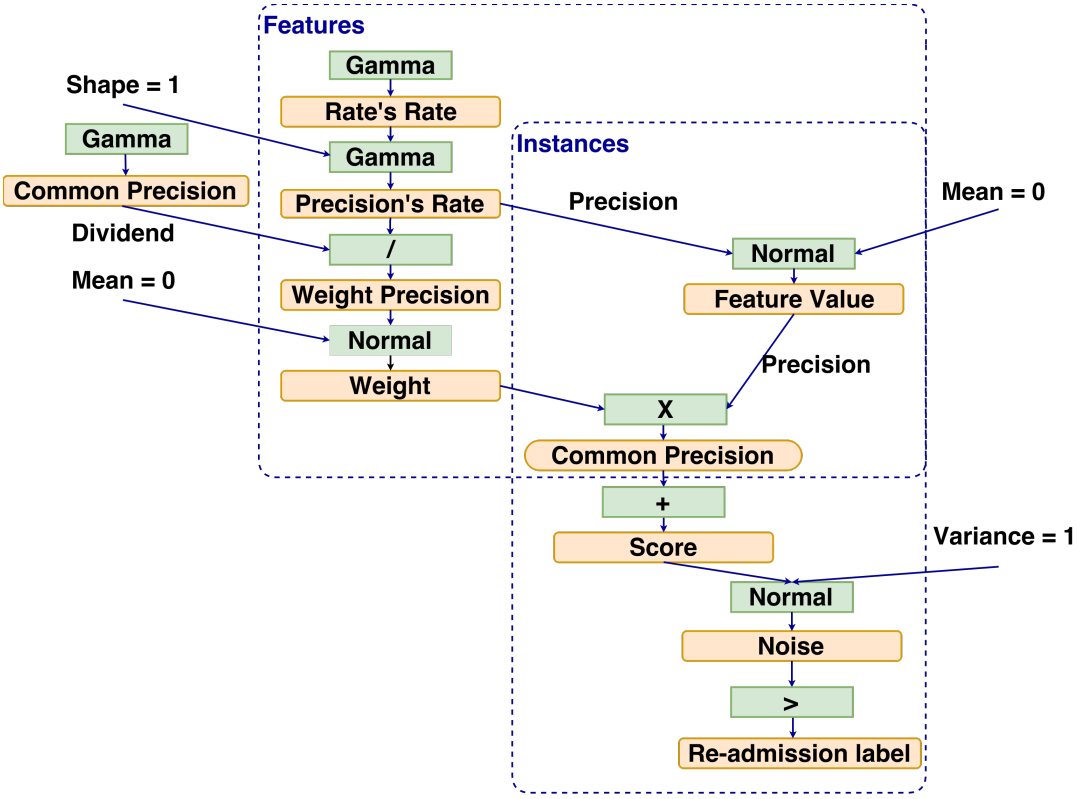

FIGURE 4.3: Infer.Net implementation of Bayes Point Machine (BPM)

the performance statistics and the dataset characteristics.

4.2.7 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) (McCulloch and Pitts, 1943) are based on a collection of artificial neurons (i.e. neural units), and are inspired by our understanding of our biological brain's axons. An ANN can be defined using three parameters: interaction patterns between layers, weights of the interactions and activation functions. Unlike brain neurons that can connect to any nearby neurons, an ANN consists of discrete layers, connections and direction of data propagation. Each neuron computes using an activation function (a.k.a. rectifier) (Hahnloser et al., 2000), with a minimum limit on the output before the data propagation.

Moreover, activation functions provide a form of linear or non-linear rectifiers to allow faster and effective training by efficiently activating the neurons. There is a wide range of activation functions, including Linear, Tanh, Sigmoid, Softmax, Softplus, Softsign, ReLU, ReLU6, LeakyReLU and PReLU. The Rectified Linear Unit (ReLU) was used in

this research, because of its effective approximation technique for classification problem. The ReLU is defined in Eq. 4.7, where $f(x)$ is the rectifier for input signal x.

$$f(x) = max(0, x) \tag{4.7}$$

Furthermore, ANN algorithms use an optimisation function to minimise a defined cost function and a method to stimulate neurons in the network. The cost function (Eq. 4.8) is a measure of model's goodness based on input data, and depends on ANNs weights, biases, inputs and outputs (for observation i), but independent of the activation values

$$C(W, B, S^i, E^i) \tag{4.8}$$

The signals in training stage can flow either in feed-forward (one-way) mode, recurrent mode (a.k.a feedback or interactive) or recurrent coupled with Long Short-Term Memory (LSTM) (Demuth et al., 2014). Because of the computational complexity in the training, most of ANNs use a from Gradient Descent (a.k.a. Steepest Descent) with backpropagation, to perform optimisation (Gron, 2017, Ruder, 2016). Some popular versions of Gradient Descent are highlighted in below:

- Gradient Descent: moves into the direction of the decreasing gradient in its search, based on a manually configured learning rate (a.k.a. learning step).
- Batch Gradient Descent: is similar to the generic algorithm, but in each step, it calculates gradient on changes in features (partial derivative). Therefore, it uses the whole training set in each step.
- Stochastic Gradient Descent: is similar to the Batch Gradient Descent, but it picks only one random instance to calculate the gradient.
- Mini-Batch Gradient Descent: is similar to the Stochastic Gradient Descent, but it picks a few random instance to calculate the gradient.
- Momentum (Sutton, 1986): accelerates Stochastic Gradient Descent by adding a momentum to the descent, when gradient does not change direction.
- Nesterov accelerated gradient (Nesterov and Nemirovskii, 1994): improves the Momentum algorithm by adapting the learning rate in anticipation of the current descent.
- Adagrad (Duchi et al., 2011): improves the Nesterov accelerated gradient by adapting the learning rate to each parameter importance.

- Adadelta (Zeiler, 2012): is a modified version of the Adagrad, which decreases learning rate with a smaller rate.

In this research, the Adadelta (Zeiler, 2012) Gradient Descent was applied, because it eliminates the need to manually adjust the learning rate, and the learning rate does not decay to zero like the Adagrad. Gradient Descent methods generally have some degree of weakness toward falling into *local minima* trap. Also, the input features that are not scaled can have big negative impact on the convergence and the learning rate. The gradient in Adadelta model is defined in Eq. (4.9):

$$\begin{aligned} g_0 &= 0 \\ g_t &= (1-\gamma) f'(\theta_t)^2 + \gamma g_{t-1} \end{aligned} \tag{4.9}$$

, where g_t represents expected gradient in iteration t, where γ is the decay term. Moreover, $f'(\theta_t)$ is the derivative of the loss, with respect to parameters (θ) at time t.

In the following subsection, a brief introduction to Deep Neural Network (DNN) is provided.

4.2.7.1 Deep Neural Network (DNN)

Firstly, Deep learning (a.k.a. hierarchical learning) refers to a class of algorithms in machine learning that attempts to learn multiple levels of complexities that correspond to different abstract levels. It uses many layers of non-linear *learning representation* to transform raw input data to a format that can be effectively exploited (Deng et al., 2014).

Furthermore, Deep Neural Network (DNN) (Bengio et al., 2009, Schmidhuber, 2015) is a class of ANNs with multiple hidden layers, which allows modelling more complex non-linear problems. DNNs act like ANNs, but with better ability to model non-linear models with more complex and effective representation of features in each layer.

Researchers in the area of DNN modelling try to create more effective methods that can implement more abstract layers. But it does not mean that more layers and neurons in DNNs would translate to a stronger model. Generally, hidden layers are not needed if a problem is linearly separable, and neither an ANN. For non-linear problems, one hidden layer is usually enough for most problems. Also, input neurons may be optimised after training using a pruning method to filter out insignificant input nodes. Also, there

are iterative pruning methods that can be run during the training phase to perform *selective pruning* by gradually reducing neurons, or *incremental pruning* by gradually increasing the number of hidden neurons, to simplify the model (Heaton, 2008).

For instance, Pham et al. (2016) presented DeepCare model which predicts illness states using a Dynamic DNN algorithm with LSTM. The model applied on hospital data from 2002 to 2013 for two chronic conditions: diabetes (11,000 patients) and mental health conditions (52,000 patients). The full model includes features like diagnoses codes, interventions, *episodes* and irregular timing, hypertension and tobacco use as inputs with input node and network layer trimmings using *L2* regularisation. The model applies a weight function on the illness state to make the model temporal, and a forget gate in the network to control the decay of this weight vector. The DeepCare model performed significantly better compared to SVM and RF and a simple DNN methods. However, the model used a very little number of features, and the comparisons were made only using two performance indicators: F-score and precision.

In the following section, Deep learning and Wide and Deep Neural Network (WDNN) are summarised, which takes advantage of latent variables, and deep feed-forward network with memorisation and generalisation.

4.2.7.2 Wide and Deep Neural Network (WDNN)

For the purpose of this research, a Wide and Deep Neural Network (WDNN) method was implemented, which combines benefits of memorisation and generalisation. The WDNN was introduced by Cheng et al. (2016). The WDNN consists of two parts: wide model, and deep model (Appendix A.7.2).

Firstly, the wide part of the ANN consists of a wide linear model for highly sparse features (random features, which are only rarely active). Secondly, the wide part that includes groups of crossed features (a.k.a. interaction terms). For each group of the crossed features, each level of one feature occurs in combination with each level of the other features. The Generalised Linear Model (GLM) (Eq. 4.10) and the cross-product transformation (Eq. 4.11) are defined as the following:

$$y = w^T x + b \tag{4.10}$$

$$\phi_k(x) = \prod_{i=1}^{d} x_i^{c_{ki}} \quad c_{ki} \in \{0, 1\} \tag{4.11}$$

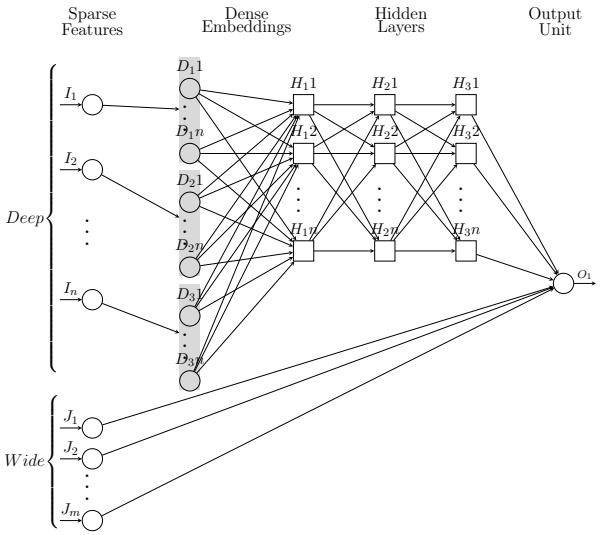

FIGURE 4.4: Abstract graph of the Wide and Deep Neural Network (WDNN)

, where y is the prediction, x is a vector of features of d features, w is model parameters and b is the bias. The $\phi_k(x)$ is the k-th transformation for vector of features x. Secondly, the deep part of ANN composed of hidden layers of feed-forward Neural Network with an embedding layer and several hidden layers for any other variable (Abadi et al., 2016). Each hidden layer performs the following operation (Eq. 4.12).

$$a^{(l+1)} = f(W^{(l)}a^{(l)} + b^{(l)}) \tag{4.12}$$

, where $W^{(l)}$, $a^{(l)}$ and $b^{(l)}$ represent weights, activations and bias for layer l, respectively. Finally, the WDNN for the LR problem (Y) can be formulated as the following (Eq. 4.13):

$$p(Y = 1|x) = \sigma(w_{wide}^T[x, \phi(x)] + w_{deep}^T a^{(l_f)} + b) \tag{4.13}$$

, where $\sigma(.)$ is the sigmoid function, $\phi(x)$ is the cross-product transformations of x features and $w_.$ are the weights.

The WDNN is an attractive modelling choice since it can generalise better the unseen features, using dense embedding in the deep part of ANN. Also, it can memorise feature interactions, using cross-product of features in the wide part of the ANN.

In the next chapter, the main healthcare data sources are defined, and the bespoke dataset that is used in this research is described.

Chapter 5

Data

In this chapter, the data sources are described and the extracts and samples are defined. In the following sections, firstly, an introduction is provided to healthcare administrative databases. Then, the Hospital Episode Statistics (HES) is summarised. After that, Secondary Uses Service (SUS) is outlined. Next, the emergency admission types are precisely defined. Afterwards, the extracted samples and created sub-samples are specified. Finally, a discussion about the data quality is supplemented with exploratory analyses.

5.1 Healthcare Administrative Databases

Administrative databases are used in performance monitoring healthcare systems in the UK, the USA and other countries. Healthcare data, such as hospital admissions, Accident and Emergency (A&E), outpatient attendance and General Practice (GP) data, are used in predictive modelling problems (Billings et al., 2013, David et al., 2006, Dialog, 2008, Kansagara et al., 2011, Lewis et al., 2011).

Although clinical databases, like primary care and Nuffield Trust databases, are a compliment to administrative databases. But, they are expensive and usually are not free, and therefore with limited applicability. According to a study, cost per-record for clinical data can range from £10 to £60 per-record compared to £1 per-record for the HES database (Raftery et al., 2005). Aylin et al. (2007) compared a set of predictive mortality risk models using administrative data only, and demonstrated that the creative use of such data can be useful for performance monitoring and is a complement to the clinical data.

Moreover, another challenge in using health databases is in dealing with different clinical coding systems (e.g. use of *Read Codes Version 2*, CTV3 and SNOMED-CT by GPs) and under-reporting of diagnostic variables (Billings et al., 2013, NHS, 2016d). For instance, SystmOne and its rival EMIS are software systems that are used by the majority of GPs in England. Although such systems can capture a wide range of record types, the way they organise and encode the patient data is not standardised and varies (TPP, 2016b, Morrison et al., 2012).

5.1.1 Hospital Episode Statistics (HES)

The Hospital Episode Statistics (HES) warehouse was originally founded in 1987 and is an administrative database that contains all inpatient admissions, outpatient appointments and A&E attendances to National Health Service (NHS) hospitals in England. In additions, the HES database is a verified and less detailed version of the SUS, which is used at hospitals. Moreover, HES data covers all the NHS trusts, including acute cares, Primary Care Trusts (PCTs) and mental health trusts. The secondary copy of the HES database is an anatomised or a pseudonymised version of the original database and can be used by researchers for non-clinical purposes. At present, the HES database holds records since 1998, and the records are being updated on monthly basis (HSCIC, 2016d).

The HES's data columns (variables) can be categorised into the following business definitions (HSCIC, 2014a):

- Clinical classifications: include details of any diagnosis, procedures and interventions, like $DIAG\_NN$ that keeps patient diagnoses codes.
- *Episodes* and *spells*: contain details of units of care provided and details of entire stay of the patient, such as $SPELLDUR$ that holds the calculation of the *spell* duration.
- Patient classifications: record details of admissions, like $CLASSPAT$ that represent the patient classification.
- Unfinished records: represent records that have open status. *Episode* status variable ($EPISTAT$) is used to flag the status of treatment.
- Outpatient appointments: hold details of the admitted patients who do not stay overnight, and records are separate from other admission types.

- A&E attendances: keep details of admitted patients to A&Es, and it is separate from other types of admissions.

Moreover, the databases hold health status and care details in format of *episodes* and *spells*. *Spell* refers to a continuous period of care, which includes *episodes* of care activities (HSCIC, 2016d). However, hospital transfers and same-day admissions sometimes are recorded as separate *spells*. Therefore, depending on the problem in hand, *super-spells* might be considered, which broaden the definition of *spells* to combine more related *episodes* of care, in order to reduce the bias.

In this study, no linkage was made to the ONS's mortality data, but the flag for in-hospital death from the HES was used to filter patients. The $DISMETH$ (Discharge Method) and $DISDATE$ (Discharge Date) variables were used to label in-hospital deaths (HSCIC, 2016d).

5.1.2 Secondary Uses Service (SUS)

The HES database is derived from the Secondary Uses Service (SUS) database, which is supplied by the Commissioning Data Sets (CDS) directly from hospitals. Compared to the HES, SUS is missing some derived variables and a series of data verifications and corrections procedures. Moreover, there are some detailed administrative, clinical and cost variables which are not passed to the HES. The SUS engine extracts two set of databases: the SUS Extract Mart (SEM) and the Payment by Results (PbR).

Firstly, the SEM data is directly used by the NHS providers and commissioners. However, SEM is much more difficult to use for analysis, compared to the PbR, because it is in raw format and very weakly validated.

Moreover, the PbR contains derived items, including very detailed tariffs, and goes through consistent data verifications. The SUS for the PbR extraction is validated according to the HSCIC guidelines and the Department of Health (DoH) policies. Therefore, PbR is a more suitable option for analysis and has the capability to run special purpose extracts to support critical care models. In addition, the PbR provides extracts based on *spells*, as well as *episodes*, with exclusion of some fields and removal of incomplete *episodes* (NHS, 2013b, 2016c,e).

5.2 Emergency Admission

The emergency admission is a method of admission ($ADMIMETH$), and by definition is not predictable and happens at short notice. It can have one of the following attributes in the HES database:

- State 21: A&E or dental casualty department;
- State 22: Immediate admission requested by a GP;
- State 23: Bed bureau;
- State 24: A consultant clinic;
- State 25: Admission via a mental health crisis resolution team, or domiciliary visit by consultant;
- State 2A: Transfer from another A&E;
- State 2B: Transfer from another hospital;
- State 2C: Baby born at home as intended;
- State 2D: Other emergency means;
- State 28: Other means.

Moreover, maternity may be considered as an emergency admission method, which is typically included in previous models:

- State 31: Admitted ante-partum;
- State 32: Admitted post-partum.

The two other types of admissions, which are not considered as an emergency, are as the following:

- Elective admissions: waiting list, booked and planned;
- Others and not applicable.

Furthermore, patient classification variable ($CLASSPAT$) should also be set to value one, to indicate that the patient was not admitted electively.

5.3 Population Samples

In this research, only inpatient data from the HES database is used. The bespoke extract of the HES database includes records from April 1995 to April 2010. The inpatient table consists of 206,528,432 *episodes*. This excludes records invalid ($NULL$ value), patient identification ($HESID$) or admission date ($ADMIDATE$). In total, it adds up to 39,403 *episodes* with $NULL$ value $ADMIDATE$ and 11,212,871 *episodes* with $NULL$ value $HESID$. Similar to the PARR model, each sample covers about 20% of unique patients in England that are admitted within the *trigger-year* of the selected time-frame (Table 5.1).

TABLE 5.1: Selected samples from the HES Inpatient database

Samples	Timeframe	Population size		Sample size		Filtered patients		
		Episodes	Patients	Episodes	Patients	Total	No prior spell	No post spell
Sample-1	1999/04 - 2004/03	18,885,777	7,206,133	6,347,067	1,441,227	1,157,873	492,458	148,950
Sample-2	2004/04 - 2009/03	31,731,488	8,104,748	11,394,152	1,615,347	1,410,923	395,522	110,961
Sample-3	2000/04 - 2005/03	32,217,541	7,370,830	6,449,169	1,474,166	1,324,712	671,919	194,097

Analogous to the PARR, the data was divided into about three years of *prior-history*, one year of *trigger-event* and one year of *prediction-period*. For instance, Figure 5.1 demonstrates how *Sample*-1 was divided into three periods.

TABLE 5.2: Combinations of the selected test and train samples

Samples	Training Sub-sample	Testing Sub-sample
Sample-1	$train_{sample\text{-}1}$	$test_{sample\text{-}1}$
Sample-2	$train_{sample\text{-}2}$	$test_{sample\text{-}2}$
Sample-3	$train_{sample\text{-}3}$	$test_{sample\text{-}3}$
Sample-1-train-half-2-test-half	$train_{sample\text{-}1}$	$test_{sample\text{-}2}$
Sample-1-train-half-3-test-half	$train_{sample\text{-}1}$	$test_{sample\text{-}3}$

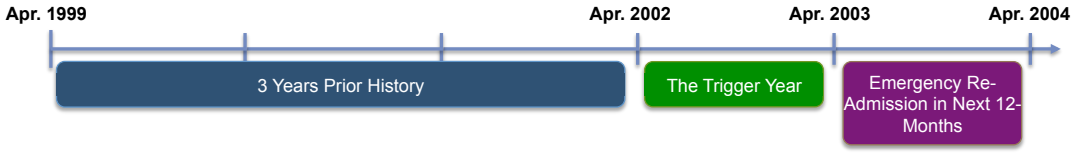

FIGURE 5.1: Data time-frame for the *PARR* model

Finally, the selected main samples from each time-frame were divided into two sub-samples for training and testing (Table 5.2). No validation sub-sample was created since different modelling algorithms were not going to be compared against each other. Each sub-sample includes 50% of the main sample, and train and test sub-samples have about 10% of all the unique patients that were admitted during in the *trigger-year*.

5.3.1 Terminologies

Finally, a number of terminologies are used through the document, which are related to the emergency admission prediction, and are outlined in below:

- *Trigger-period* (*trigger-year*): Only the patients who have an admission during *Trigger-period* are included in our data extracts. In our analysis, it is the fourth year of a selected time-frame.
- *Trigger-event* (*trigger-admission*): In our research, the *Trigger-event* is considered as the first qualified admission of a patient and is drawn from the *trigger-period*. Based on this admission, the *prior-history* periods and the *Future-admission* are calculated.
- *Prior-history* (*prior-period*, *prior-trigger* or *prior-admissions*): The prior care records before the *trigger-event* are specified as the *prior-period*. In this study, the *Prior-history* is the 3-year period prior to the *trigger-event*.
- *Future-admission* (*target*, *label* or *response variable*): The *Future-admission* refers to the emergency admission outcome that is going to be predicted. It happens within a time-window from the *trigger-event*.
- *Prediction-period* (*time-horizon* or *post-trigger*): The period after the *trigger-event* of the patient is defined as the *time-horizon*. For instance, the *Prediction-period* can be 30-day or 365-day after the discharge date ($DISDATE$) for the *trigger-event*.

5.4 Data Quality

Every healthcare administrative database has usually some degrees of quality issues, and the HES database is no exception. In Appendix A.5, some detailed statistics are provided for the populations and samples, to identify potential issues in the data before progressing with pre-processing.

Firstly, Appendix A.5.1 demonstrates that a large portion of *spells* is missing before April 1997, and there is a little record after April 2010 in the bespoke HES extract.

Moreover, in-hospital death shows a seasonal pattern, it has a peak in the fourth quarter (October to January), and in-hospital deaths gradually decreases with a very slow-rate.

Furthermore, as it is expected, the number of *episodes* gradually increases per year, but its rate is moderately higher for unique patients from about 2003 onwards.

In addition, Appendix A.5.2 and Appendix A.5.3 present descriptive statistics for the selected features in the pre-processing stage. In Chapter 7, a framework is presented that can deal with major data quality issues.

In the following chapter, the definitions of the problems that are studied in this research are presented. In addition, the benchmarking models that used are briefly described.

Chapter 6

Problem Definition

In this research three major pieces of work have been carried out. Firstly, a framework is designed for pre-processing the Hospital Episode Statistics (HES) and the Secondary Uses Service (SUS). Then, a model for hospital emergency readmission is provided. Finally, a temporal comorbidity risk index is developed.

Furthermore, for all the developed solutions, a bespoke extract of the HES was used which covers admissions from April 1995 to April 2010. Three different samples, each covering a 5-year period, were selected to train, test, validate and benchmark the models.

Moreover, studying interventions, control groups and stratified classes can provide an opportunity to study factors that can affect the implementation of risk models (Davies et al., 2015, Eggli et al., 2014, Hutchings et al., 2013). However, the effects of interventions were not included in the scope of this research, because the NHS has no centralised service to record the interventions.

In the following sections, firstly, the specification of the healthcare pre-processing framework is defined. After that, the hospital emergency readmission problem is explained, including the benchmarking models. Then, the comorbidity index problem is outlined, and a brief overview of the benchmarking models is provided.

6.1 Phase I: Healthcare Pre-Processing Framework

The HES database is a less detailed and verified version of the SUS, and it is maintained by the Health and Social Care Information Centre (HSCIC) (HSCIC, 2016f). Although

the quality of the data improves after a set of standard cleaning and validation by the HSCIC, the data still needs to be pre-processed before it can be used for modelling (Section 1.2).

Moreover, patient records include a very large set of variables, which are very sparse and are not collected on a regular basis. Also, many of the patients have censored (i.e. partially observable) observations from left and right, which may introduce a level of bias in the analyses. Consequently, a systematic way must be developed to generate features, to be able to capture as mush as features possible with temporal dimensions that are consumable by statistical methods.

Furthermore, for engineering features, all the models in the literature (Appendix A.1.1) rely either on previous studies, clinicians' experts or a very shallow exploratory analysis. And, our literature review did not find a framework for cleaning the HES or the SUS data, and neither it could identify any framework for generating healthcare features.

The aim of this phase is to design a set of steps, to pre-process the HES and the SUS data, and treat invalid data by removing or imputation. Also, a systematic set of procedures must be defined to generate features, discretise or re-categorise features, aggregate temporal dimensions and rank importance of features.

6.2 Phase II: Modelling Hospital Emergency Readmission

Developing and implementing a robust decision support tool for admitted patients is critical. Predictive risk models can help patients and carers obtain appropriate support services in clinical decision-making. In addition, such models can improve care quality, and reduce the costs of inappropriate admissions to hospital or Accident and Emergency (A&E) (Section 2.2).

In 2005, the Patients at Risk of Re-hospitalisation (PARR) (Billings et al., 2006a, Lewis, 2011) was commissioned to identify patients at high risk of emergency readmission to inpatient within a year. After that, in 2006, the Combined Predictive Model (CPM) was commissioned, to use the General Practice (GP), inpatient, outpatient and A&E data, in order to predict the risk of 1-year emergency admission (DH, 2006). Later in 2013, an update of the CPM model was published with new features and more detailed performance statistics (Billings et al., 2013).

Most existing decision support tools that are based on hospital administrative data, use Logistic Regression (LR) or Coxian Phase-type Distribution (C-PHD) models (Adeyemi

et al., 2013, Bardsley, 2012, Bottle et al., 2014, Kansagara et al., 2011, Lewis et al., 2011, ACI, 2014, DH, 2011a, Paton et al., 2014). Although these models are simple and popular, they have limited powers, because of algorithmic shortfalls, restricted assumptions and weak variable selection strategies. In the area of healthcare risk modelling research, there have been many successful implementations of advanced machine learning methods (Green et al., 2006, Lee et al., 2012, Nilsson et al., 2006, Peelen et al., 2010, Song et al., 2004). However, few studies used a Bayesian approach to address emergency hospital readmission problems (Álvaro-Meca et al., 2012, Cui et al., 2015, Demir and Chaussalet, 2011, Gupta et al., 2014, Helm et al., 2015, Huws et al., 2008).

The aim of this phase of research is to develop an Ensemble generative risk model, to predict emergency readmission within a year to the England's hospitals. Firstly, based on the designed healthcare pre-processing framework, the features can be cleaned and generated, filtered and ranked from a bespoke extract of the HES. After that, sub-models specialised for sub-population must be trained using a Bayes Point Machine (BPM) approach. Afterwards, an optimised Ensemble model of these sub-models must be generated. Moreover, the proposed models can be trained, tested, validated and benchmarked using different samples and cohorts against previous models. Finally, the source code of the produced solution is desirable to be released with sufficient level of documentations.

6.2.1 Benchmarking Models

In this part of the research, performances of the PARR, the CPM and the CPM update were used as a benchmark, since these tools use the HES data and are being used by commissioners across the NHS England. These decision support tools help clinicians and commissioners to rank and group patients based on anticipated intervention intensity, including case management, disease management, supported care, prevention and wellness promotion.

Moreover, none of the chosen benchmarking models was reconstructible due to lack of documentation and data accessibility. Therefore, the benchmarking models were not applied to our data, and instead the reported perfromance statistics were used.

In the following subsections, firstly, the PARR (Billings et al., 2006a, Lewis, 2011) model is described. Then, the CPM (DH, 2006) is summarised. After that, the CPM update that was developed by Billings et al. (2013) is outlined. Finally, a summary of the benchmarked models' settings is provided for more clarity.

6.2.1.1 Patients at Risk of Re-hospitalisation (PARR)

The UK's Department of Health (DoH) commissioned the Patients at Risk of Re-hospitalisation (PARR) model in 2005 (Billings et al., 2006a, Lewis, 2011). The characteristics that made the PARR model an ideal benchmarking model are presented in below:

- Modelling hospital emergency readmission of inpatient patients within 12-month period;
- Being applied by the NHS England for inpatients before discharge, in order to identify if an extra intervention may be required;
- Using only inpatient administrative records from the HES database;
- Narrowing down the cohort to patients with 75+ years of age that have a *reference* condition. The *reference* conditions include some of the frequent Healthcare Resource Groups (HRGs), which are identified at the time of the study.

At the modelling stage, the PARR had two branches of models that were fitted using a LR: the PARR-1 and the PARR-2 (Billings et al., 2006b). Unlike the PARR-2, the PARR-1 limits the emergency admissions at the *trigger-event* to have a *reference* condition (such as diabetes, congestive heart diseases and sickle cell disease).

6.2.1.2 Combined Predictive Model (CPM)

After the PARR development, the Combined Predictive Model (CPM) was commissioned to include care records from the primary care and the secondary care sectors (DH, 2006). The main characteristics of the CPM are as the following:

- Modelling the 12-month emergency readmission to hospital;
- Being applied by the NHS England for patients in the primary and the secondary care to identify high-risk patients for a period of close follow-up;
- Using inpatient, outpatient A&E from the SUS database, and records from GP systems.

The first CPM model that was developed in 2006, included data from 2002 to 2005 from five Primary Care Trusts (PCTs). The population was divided equally into training

and testing, and the final year of data was used as the *prediction-period* for emergency readmission. Similar to the PARR, the CPM used a LR model, but with features from different care sectors, which allowed it to be implemented in primary and secondary care settings. Regarding the performance, the CPM could identify more positive high-risk patients with a minimum of two years historical data from the SUS and GPs databases. Also, it showed that linking four different data sources and handling missing data can be very challenging, and a large portion of the development needs to be invested in data preparation, linking, validation and correction.

6.2.1.3 CPM Update

There are other modelling efforts that focused on hospital emergency readmission to inpatient within 12-month, such as the solution provided by Billings et al. (2013). The main characteristics of the Billings et al. (2013) study can be summarised as follows:

- Modelling hospital emergency readmission of inpatient patients within 12-month;
- Using inpatient, outpatient and A&E from the SUS database, in addition to the GP records.

The developed model was based on patient aged 18-95 with admissions from 2007-2010. It included GP records from five PCTs in combination with the SUS records extracted from hospitals. The performance of the model was noticeably higher than the PARR and the CPM using the LR method. Moreover, this model is referred to as the CPM update, the updated CPM model or the Billings et al. (2013) model throughout this document.

6.2.1.4 Summary of Models Settings

The CPM, the PARR and the CPM update used different data sources and settings. Therefore, in the benchmarking stage of our modelling, the statistics are presented for three sub-populations (Table 8.1), to minimise the differences. The main differences in the modelling settings of the studies are presented in below:

- The PARR-1 model:

  - Includes patients who only had an emergency admission during the *trigger-period*;
  - Removes invalid patient identification numbers;
  - Removes invalid or missing admission or discharge dates;
  - Removes invalid or missing admission classification;
  - Removes deceased patients at the *trigger-event*;
  - Removes patients aged below 65;
  - Applies other data quality measures, which are not specified in the public documentations.

- The CPM model:
  - Includes patients who only had any type of admission during the *trigger-period*;
  - Removes invalid patient identification numbers;
  - Removes invalid or missing gender;
  - Removes invalid or missing age;
  - Removes deceased patients at the *trigger-event*;
  - Limits model to just five PCTs;
  - Applies other data quality measures, which are not specified in the public documentations.

- The updated CPM model:
  - Includes patients who only had any type of admission during the *trigger-period*;
  - Includes patients who only were registered with a GP;
  - Includes patients that were aged 18-95;
  - Limits model to just five PCTs;
  - Applies other data quality measures, which are not specified in the public documentations.

In contrast, the emergency readmission models that are developed during this study have the following settings:

- Include patients who only had an emergency admission during the *trigger-period*;
- Remove invalid patient identification numbers;

- Remove invalid or missing admission or discharge dates;
- Remove invalid or missing admission classification;
- Remove deceased patients at the *trigger-event*;
- Remove patient aged below one;
- Impute invalid or missing variables based on other variables or *episodes*;
- Categorise invalid or missing variables as a separate state.

6.3 Phase III: Modelling Comorbidity Index

There is increasing evidence that the quantification of high-risk operations and procedures, with adequate adjustment, can greatly improve the quality of readmission models. There have been two streams of work on risk scoring comorbidities to estimate future resource utilisation, emergency admission and mortality (Section 2.3.1).

The first stream of researches looks at the odds ratio of major diagnoses groups and therefore is highly reliant on the whole population statistics. Another weakness of such models stems from crudely summing up the risk score for comorbidities, which are based on the most recent admission of the patients. A popular example is the Charlson Comorbidity Index (CCI) (Charlson et al., 1987), which relies on twenty-two comorbidity groups.

The second stream of models uses a diagnoses classification approach based on similarities, type of care, likelihood or duration, which is usually very complex and specialised to highly particular settings. One prominent method is the Elixhauser Comorbidity Index (ECI) (Elixhauser et al., 1998, AHRQ, 2016b), which relies on thirty comorbidity groups. Unlike the CCI, the ECI is using Diagnosis-related Groups (DRG), which was first developed by Fetter et al. (1980), Mistichelli (1984) and is based on ICD diagnoses, procedures, age, sex, discharge status, complications and comorbidities. Another well-established method is the John Hopkin's (Weiner and Abrams, 2011) Adjusted Clinical Groups (ACG), which is a commercial tool. It encapsulates 32 diagnoses groups, known as Aggregated Diagnosis Groups (ADGs), and their aggregations called Expanded Diagnosis Clusters (EDCs).

Based on the machine learning pipeline that was developed in the prior stage of our research (Section 6.1 and Section 6.2), comorbidity index is an extremely significant

factor and has a high potential for further improvement. Presently, comorbidity risk indices have four major weakness areas: robustness, temporal dimensions, population stratification and associated factors to comorbidities and complications.

In this phase of research, we aim to improve on these four major areas. Firstly, to make the risk score relevant to different environments, an approach must be used to model complex correlations between variables and states. Secondly, to better distinguish the short- and long-term conditions (i.e. *prior-admissions*, Length-of-Stay, and delta-time between admissions), the temporal dimension may be included in the form of life-table or a polynomial weight function. Thirdly, population stratification is a major factor in the prevalence of medical conditions, and therefore must be adjusted. Fourthly, major correlated factors to diagnoses may be included directly or indirectly (as latent) to improve the risk estimates, including secondary diagnoses, operations, procedures and complications.

The aim of this phase of our study was to develop a generic temporal comorbidity risk index to predict the risk of emergency admission within 30-day and 365-day periods. Firstly, the input data from the HES may be processed using the healthcare pre-processing framework. After that, the comorbidity risk may be modelled using the temporal history of diagnoses, operations, complexities and other major correlated factors. The modelling approach must be generic enough to be applicable to different healthcare setting and input data sources. Afterwards, the models can be trained, tested, validated and benchmarked using multiple samples and cohorts against previous models. Finally, a user-friendly toolkit must be designed to allow wider adaptation and validation by research communities.

6.3.1 Benchmarking Models

Since the introduction of the CCI and the ECI models, researchers across the world have tried to translate and to implement versions of the comorbidity indices that are more representative of their population. Austin et al. (2012), Baldwin et al. (2006), Khuu et al. (2015), Kuo and Lai (2010), Lieffers et al. (2011), Sharabiani et al. (2012) provided systematic reviews of some major comorbidity indices, and some have compared the risk model for 30-day emergency admission in addition to mortality.

In this phase, the developed comorbidity index models were benchmarked against our implementation of the HSCIC adaptation of the CCI (HSCIC-CCI), and the reported performance statistics of CCIs and ECIs. In the following subsections, the CCI and

the ECI models are outlined.

6.3.1.1 Charlson Comorbidity Index (CCI)

Initially the Charlson Comorbidity Index (CCI) was developed by Charlson et al. (1987), using a cohort of 559 medical patients from New York Hospital-Cornell Medical Center for a 1-month period in 1984, with 1-year follow-up. The aim of the model was to derive a comorbidity index to predict mortality within a year. The CCI created thirty diagnoses groups within nineteen major groups, and then calculated the Cox proportional hazard to derive the adjusted weights for the diagnoses groups. After that, it rounded the weights and calculated the comorbidity risk for patients by summing up the weights. The CCI scores the nineteen diagnoses groups with various weightings, which results in minimum possible risk score of zero and maximum of thirty-one. Finally, the model was benchmarked against the Kaplan and Feinstein methods (Kaplan and Feinstein, 1974), which was developed for a cohort of diabetes, using cumulative survival plot.

The CCI introduced an easy to use but very basic way to predict comorbidity risk of 1-year mortality. One of the recent translation of the CCI is the HSCIC version of the CCI (HSCIC-CCI), that is developed by Dr Foster Unit (Aylin et al., 2010, Bottle et al., 2011) and adapted by the NHS England (HSCIC, 2014d, 2015, 2016c). The HSCIC-CCI includes seventeen main groups of diagnoses, and the diagnoses weights are continuously being reviewed by the HSCIC. The HSCIC-CCI, with modified diagnoses groups and new weightings, has a maximum score of hundred and twenty-nine in theory. Also, this version of comorbidity index excludes stillbirths and cancer code anomalies. Firstly, stillbirth is removed, because it is not included in the HES-ONS link. Moreover, when both cancer and metastatic cancer are present in the diagnoses of a patient at the same time, it is set to zero to avoid miss-classification and lower model bias.

In this study, our comorbidity models are compared directly against the HSCIC-CCI, which was implemented using our data. Also, the models were compared indirectly against the survey benchmarks, because they reviewed multiple versions of comorbidity indices with the inclusion of emergency admission predictability power (Bottle and Aylin, 2011, Bottle et al., 2014, Holman et al., 2005, Mehta et al., 2016).

6.3.1.2 Elixhauser Comorbidity Index (ECI)

Firstly, Elixhauser et al. (1998) developed Elixhauser Comorbidity Index (ECI) to predict the Length-of-Stay (LoS), in-hospital mortality and hospital charges. The study used thirty refined diagnoses groups with ordinary least square regression and a LR to predict the contribution of comorbidities. Unlike the CCI, the ECI model also adjusted for demographics, financial incentives and clinical differences (including age, gender, race, insurance, type of admission, operations and complications). The population study was drawn from California hospitals inpatient data, which contained about 1.8 million admissions from 439 hospitals after excluding under 18 years old, non-emergency admissions, maternal, and discharges to long-term care. The study presented a method to rank the risk of comorbidity of patients for heterogeneous, as well as homogeneous population sub-groups (DesHarnais, 1990).

Moreover, an ECI adaptation (AHRQ, 2016b) is sponsored and actively maintained by the Agency for Healthcare Research and Quality (AHRQ). The inputs to the model are ICD-10 diagnoses, and the generated Diagnosis-related Groups (DRGs). The model classifies hospital cases into about 470 cases, using ICD diagnoses, procedures, age, sex, discharge status, complications and comorbidities Fetter et al. (1980), Mistichelli (1984). Li et al. (2008), Southern et al. (2004) published benchmarks of Charlson and Elixhauser indices and provided more insight into their performances and general characteristics.

In the next chapter, the healthcare pre-processing framework is presented as the first phase of research.

Chapter 7

Phase I: Healthcare Pre-Processing Framework

The developed healthcare pre-processing framework consists of four generic steps of data pre-processing and is part of the development toolkits (Chapter 10). Initially, the data was ingested and explored. Next, a set of procedures was carried out to remove invalid variables and data records, and then the features that have inconsistencies were treated. Thereafter, a pool of features was generated in a systematic way. Finally, the features were filtered and ranked by their importance, and then top features were selected.

Figure 10.1 outlines the pre-processing steps, using a process-flow diagram. In the following sections, these four steps are outlined.

7.1 Step I: Data Management

In the data management step, a database was configured, and a bespoke extract of the data sources was obtained. Then, an exploratory analysis of the features was carried out, in order to visually observe their distributions properties, like quantiles, most frequent values and invalid values. Thereafter, the variables were filtered out based on their relevance and quality. Afterwards, cleaning, validation and re-categorisation rules for the next steps are designed.

In this research, a MySQL database engine was used for our analyses (Oracle, 2016). The MySQL database is an open source and highly scalable relational database management system with a strong development community. Although it was feasible to choose

a distributed computing solution, like MySQL Cluster[1] or Hadoop Distributed File System (HDFS)[2] with Apache Spark interface[3], our healthcare data was well-manageable on a single node. Since our MySQL queries, indexes and keys were designed very carefully, the migration from MySQL database to a more advanced engine, like MySQL Cluster, would be very smooth.

Moreover, the features in the database have been explored in the *R* environment (R Foundation, 2016), mainly using *R* base packages. In addition, continuous features with complex patterns have been investigated using more specialised *Cognostics* tools (Izenman, 2008), including *Tableau* software (Tableau, 2016), *ggplot* library (Wickham, 2016) and *Trelliscope* library (DeltaRho, 2016).

Thereafter, the duplicate variables were removed based on their level of quality. And, unusable features with low-quality, inactive and irrelevant variables were excluded.

Finally, a set of cleaning, validation and re-categorisation rules was designed based on our exploratory analyses and the HSCIC documentations (HSCIC, 2014a).

7.2 Step II: Data Preparation

The data pre-processing step can be divided into four main sub-steps (Figure 7.1). Firstly, a sample was acquired from the main database, and a number of basic cleaning rules have been applied (Figure 7.2). Thereafter, the re-factorisation, including discretisation and re-categorisation, and imputation of features were done simultaneously (Figure 7.3). Finally, *spells* with same admission date and their *episodes* were grouped into *super-spells* (Figure 7.4).

7.2.1 Sample and Remove Invalid Records

In this sub-step (Figure 7.2), firstly, about 20% of unique patients were selected from all admissions within the *trigger-year*.

[1] MySQL Cluster is a distributed database engine that has high availability and high throughput with low latency.

[2] Hadoop Distributed File System (HDFS) is a distributed file-system that stores data in computer clusters built from commodity hardware.

[3] Apache Spark is an open-source cluster computing framework, that provides an interface for programming entire clusters with implicit data parallelism and fault-tolerance.

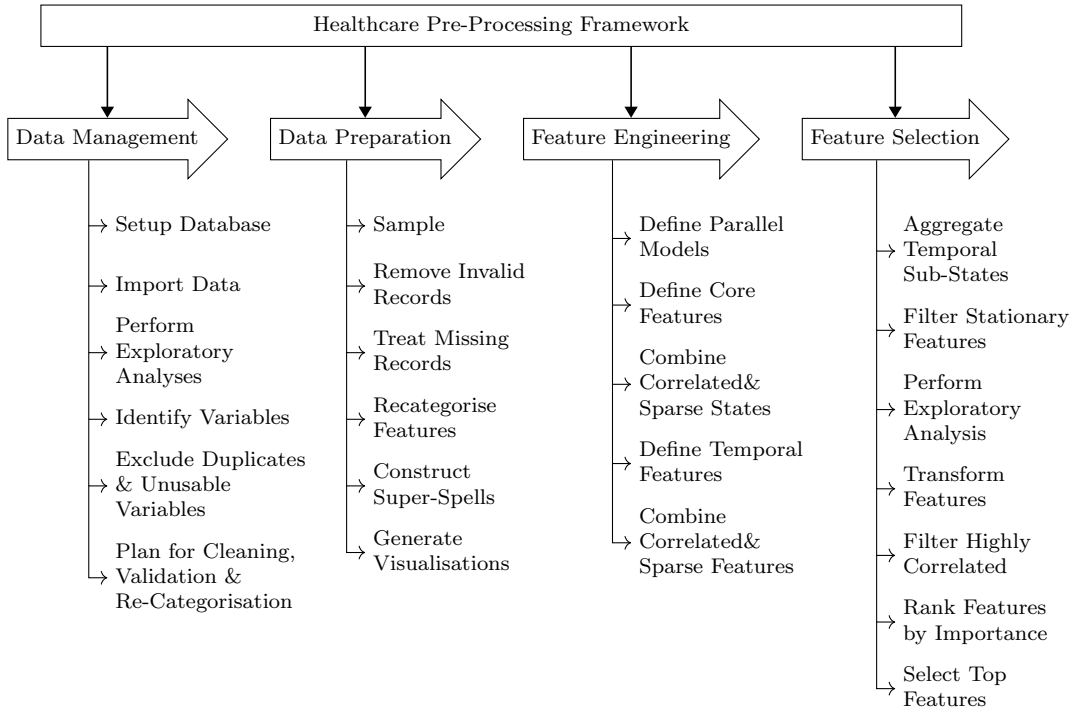

FIGURE 7.1: Healthcare pre-processing framework: Process-flow diagram

Since the inpatient records are *episode-wise*, they were sorted by patient identification ($HESID$), admission date ($ADMIDATE$), *episode* start-date ($EPISTART$), *episode* order ($EPIORDER$), *episode* end-date ($EPIEND$) and *episode* key ($EPIKEY$).

Next, invalid records and excluded populations were removed. Initially, infants less than 1-year old were dropped. Then, patients with invalid ($NULL$ value) patient identification ($HESID$) were removed. Thereafter, unknown admission dates ($ADMIDATE$) were removed. Later, if a patient died at the *trigger-event* or during the *prior-period*, the records were filtered out. Also, if a patient did not have emergency admission within the *trigger-year*, the records were removed.

7.2.2 Cleaning and Treatment

Initially at this sub-step (Figure 7.3), invalid discharge date ($DISDATE$) and *episode* end-date ($EPIEND$) were cleared (turned to $NULL$).

Furthermore, discharge date ($DISDATE$) and *episode* end-date ($EPIEND$) were consolidated when one was missing, as part of the imputation procedures. The main reasons of missing $DISDATE$ or $EPIEND$ are: regular day case *episodes*; not being the last *episode* in the *spell*; or not known.

Thereafter, sex ($GENDER$) and ethnicity ($ETHNOS$) were imputed based on all the patients' records. Next, the Healthcare Resource Groups (HRGs) values ($HRGLATE$ and $HRGLATE35$) were set to their minimum in each *spell*. In the HES database, there are six main HRGs categories, numbered two to seven, where two represents the lowest resource use, and seven represents the highest resource use (NHS, 2013e).

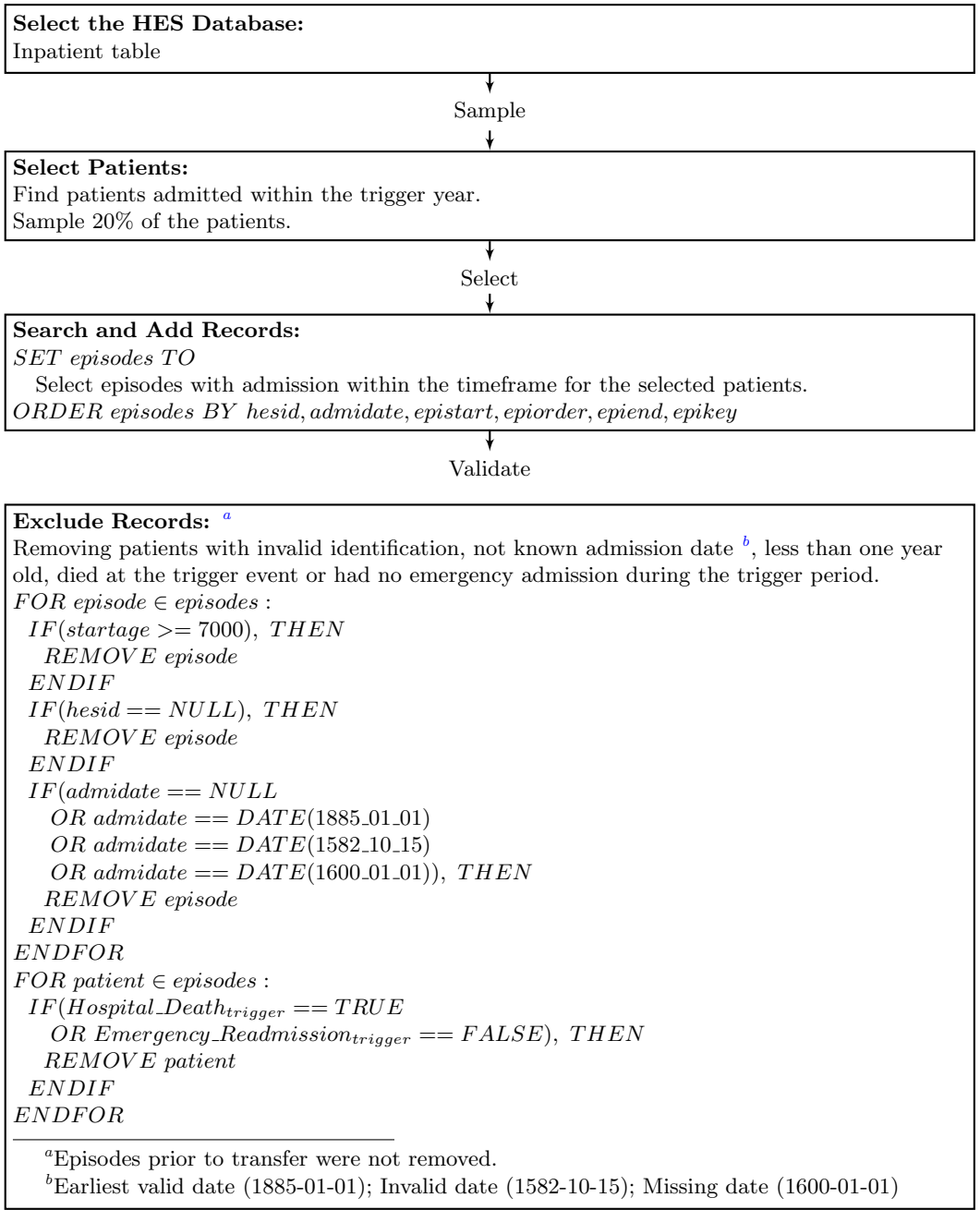

FIGURE 7.2: Healthcare pre-processing framework - Step I: sampling sub-step

Select Sub-samples:
Train and test subsamples

↓

'NULL Invalid Dates':[a]

$FOR\ var \in \{disdate, epiend\}:$
$IF(var == NULL$
$OR\ var == DATE(1885\_01\_01)$
$OR\ var == DATE(1582\_10\_15)$
$OR\ var == DATE(1600\_01\_01)),\ THEN$
$var = NULL$
$ENDIF$
$ENDFOR$

[a]Earliest valid date: 1885-01-01; Invalid date: 1582-10-15; Missing date: 1600-01-01

↓

Imputations and Removals:[a]

$disdate_{spell} = MAX_{spell}(disdate)$
$IF(epiend == NULL\ AND\ disdate \neq NULL),\ THEN$
$epiend_{spell} = MAX_{spell}(disdate)$
$ENDIF$
$IF(disdate == NULL\ AND\ epiend \neq NULL),\ THEN$
$disdate_{spell} = MAX_{spell}(epiend)$
$ENDIF$
$IF(MAX_{spell}(disdate) < MAX_{spell}(admidate)),\ THEN$
$disdate_{spell} = admidate$
$ENDIF$
$epiend = disdate - admidate$
$IF(Any_{spell}(dismeth == 4)), then$
$dismeth_{spell} = 4$
$ENDIF$
$gender_{patient} = MAX_{patient}(COUNT_{patient}(gender))$
$ethnos_{patient} = MAX_{patient}(COUNT_{patient}(ethnos))$
$hrglate_{spell} = MIN_{spell}(hrglate)$
$hrglate35_{spell} = MIN_{spell}(hrglate35)$

[a]By default *disdate* must be NULL, except for transfers and the final episode.

↓

Factorisations and Imputations:
Factorise and convert variables using the rules presented in (Appendix A.4.3).
$genderRecoded = FACTORISE_{genderRecoded}(gendere_{patient})$
$ethnosRecoded_{patient} = FACTORISE_{ethnosRecoded}(ethnos_{patient})$
$imd04rkRecoded_{patient} = FACTORISE_{imd04rkRecoded}(imd04rk_{patient})$
$ageRecoded_{patient} = FACTORISE_{ageRecoded}(startAge_{patient}, endAge_{patient})$
$admimethRecoded_{spell} = FACTORISE_{admimethRecoded}(admimeth_{spell})$
$classpatRecoded_{spell} = FACTORISE_{classpatRecoded}(classpat_{spell})$
$epidurRecoded_{spell} = FACTORISE_{epidurRecoded}(epidur_{spell})$
$admisorcRecoded_{spell} = FACTORISE_{admisorcRecoded}(admisorc_{spell})$
$intmanigRecoded_{spell} = FACTORISE_{intmanigRecoded}(intmanig_{spell})$
$orgCluster_{spell} = FACTORISE_{orgCluster}(procode3_{spell})$
$rotreatRecoded_{spell} = FACTORISE_{rotreatRecoded}(rotreat_{spell})$
$hrg_{sample_{spell}} = CONVERT_{hrglate35}(hrglate_{spell})$

FIGURE 7.3: Healthcare pre-processing framework: data cleanings and treatments step

Finally, re-factorisations (Appendix A.4.3) were applied on a set of features, including:

- Sex ($GENDER$);
- Ethnicity ($ETHNOS$);
- Index of Multiple Deprivation (*IMD*) overall rank ($IMD04RK$);
- Age (AGE);
- Admission method ($ADMIMETH$);
- Patient classification ($CLASSPAT$);
- *episode* duration ($EPIDUR$);
- Admission source ($ADMISORC$);
- Intended management for patient classification ($INTMANIG$);
- NRLS (NRLS, 2010b) organisation clusters (derived from $PROCODE3$);
- Region of treatment ($ROTREAT$).

Next, conversion of the HRGs to the latest version (HRG v3.5) were carried out ($HRGLATE$), to make it consistent across our selected samples (HSCIC, 2016b).

7.2.2.1 Other Care Sectors

The aim of this subsection is to highlight only major relevant cleaning rules, that have been encountered throughout our researches.

Other data sources that may be considered to be included in healthcare modelling are:

- Inpatient, outpatient and Accident and Emergency (A&E) data from the Secondary Uses Service - Payment by Results (SUS-PbRs).
- Diagnoses, drugs, encounters and lab data from General Practises (GPs).
- Outpatient and A&E data from the HES.
- Community and social care data from local authorities;
- Mental health data from the Mental health Services Data Set (MHSDS).

Firstly, the use of SUS-PbR is very similar to the HES, but additional procedures must be carried out to remove invalid data records, duplicate and missing variables (HSCIC, 2016e). The HSCIC performs some additional removals during the HES core processing, and must be carried out for the SUS-PbR. These additional removal rules are as follows:

- The records that cannot be attributed to a valid Organisation Data Service (ODS) provider.
- The records with invalid organisation code ($PROCODE5$), old query date (*Query Date*), unattended patients (*Attended Or Did Not Attend*), excluded records (*Exclusion Reason*), unfinished *spells* (*Finished Indicator*) and *spells* not included in the PbR calculations (*Spell In PbR/Not In PbR*).
- The duplicate records that can be captured using the published HSCIC guideline (HSCIC, 2016e).
- The records that fall out of the reporting period. These must be determined for each time-frame and data source independently, due variety of data recording quality.

Moreover, pre-processing GP data are usually very challenging as there is a very scatter adaptions of standard clinical terminology systems across GPs, as well as inconsistent disease coding practices throughout time by GPs. At the moment, there is a mix-adaptation of *Read Codes Version 2*, CTV3 and SNOMED-CT, in addition to the EMIS National Codes that are added to supplement the *Read Codes Version 2* (NHS, 2016d, TPP, 2016a).

More details about cleaning rules must be enquired directly from data providers, as there is very little public documentation for analyses of these data sources.

7.2.3 Grouping to Super-Spells

In this sub-step (Figure 7.4), initially, *episodes* were ordered by patient identification ($HESID$), *episode* start-date ($EPISTART$), *episode* order ($EPIORDER$), *episode* end-date ($EPIEND$) and *episode* key ($EPIKEY$). Then, each patient record was ordered by admission date ($ADMIDATE$) and discharge date ($DISDATE$), and later it is determined if *spells* could be categorised as a transfer.

Afterwards, the same day *spells* were combined into one *spell*, with exception of regular admissions. The defined rules are based on the NHS publications which specify

an approach to construct a type of *super-spell* (HSCIC, 2014e). The *super-spell* term refers to an entire period of care of a patient, which can include multiple *episodes* and *spells* of care at multiple hospitals (DFI, 2014).

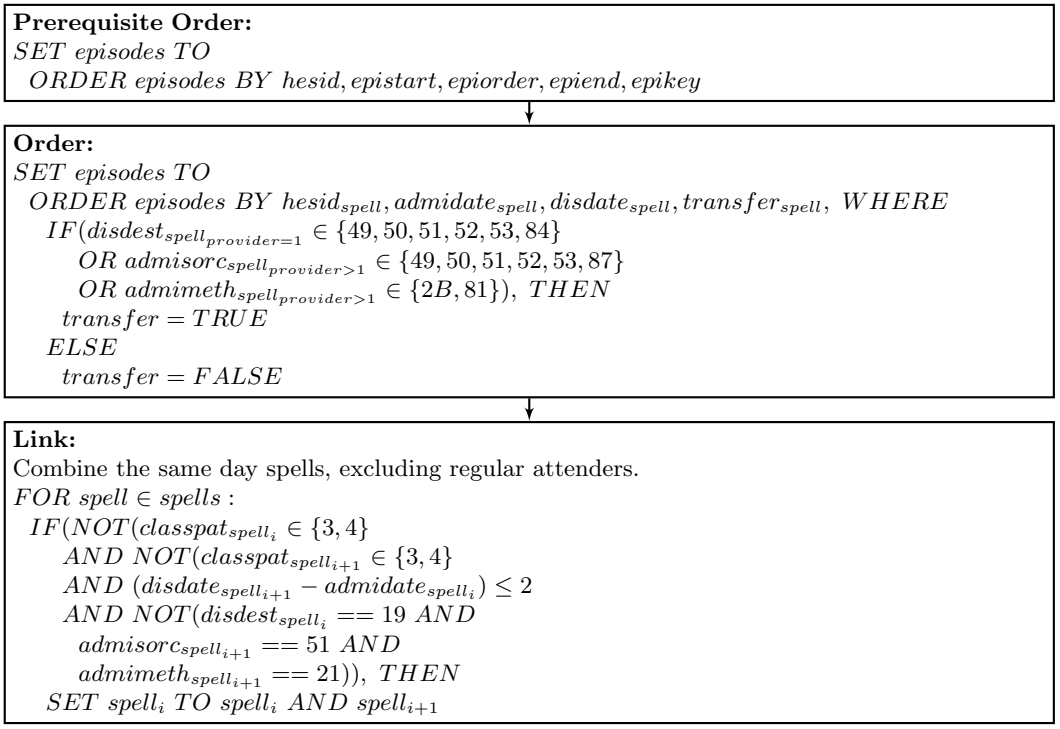

FIGURE 7.4: Healthcare pre-processing framework: the super-spells declaration step

7.3 Step III: Feature Engineering

In this step, the features pool was generated through several sub-processes. Initially, parallel models can be defined if there are more than one care sector or cohort to be modelled. Thereafter, four sub-steps were defined to construct a pool of features with minimal levels of sparsity in features and with efficient generation procedures.

The feature pool specification is defined in the following subsections, and the detailed design of the process-flow is described in Section 10.3.

7.3.1 Parallel Models

Different parallel sub-models can be defined at this instance, in order to split-up the modelling efforts, in-line with *Multi-Task Learning* (Section 4.2.1) or Ensemble learning (Section 4.2.2) approaches. For instance, parallel sub-models can be defined to predict risk of emergency admission from GP, inpatient, outpatient, A&E, social and community care. In addition, more parallel models can be defined to target a very specific cohort of patients, like cancer patients or patients in risk of frailty.

Because, there is only one data source and one principal population cohort (England patients with inpatient admission), this sub-step was not applicable for this research.

7.3.2 Core Features

In total, 738 summary features were generated, which can be categorised into three main groups:

- Administrative:
  - Admission: patient classifications; number of *episodes* and *spells*; admission, readmission and discharge times; and sources and methods of admissions and discharges.
  - Bed days: durations of *spells*; and preoperative and post-operative durations.
  - Geographical: providers' codes; and regions of registrations and treatments.
  - Hospital: operation codes.
  - Identifications: patient identification; and admission time-frame.
  - Speciality: specialities of consultants; and palliative cares.
  - Waiting time: admissions waiting time.
- Clinical:
  - Diagnosis: HSCIC version of the Charlson Comorbidity Index (HSCIC-CCI) (Aylin et al., 2010, Bottle et al., 2011, HSCIC, 2014d, 2015, 2016c); and PARR's Healthcare Resource Groups (HRGs) *reference* conditions.
  - Operation: number of operations; and categories of operations.
- Patient:

– Demographic: age; the Index of Multiple Deprivation (*IMD*); ethnicity; and gender.

All features categorisations are presented in Appendix A.4 and their mathematical definitions are presented in Appendix A.4.3.

In addition to the generated features, two groups of features were constructed for performance assessment and grouping of the predicted outcomes: *prior-trigger* and *post-trigger* features.

7.3.3 Combine States

It is very likely that discrete features are very sparse, including the recategorised features. Therefore, in this sub-step another round of explorative analyses can be carried out to observe the distribution of features states (dummy features or terms), and identify features states that can be combined.

An alternative approach would be to use an unsupervised approach, like k-means clustering and Principal Component Analysis (PCAs), to find optimal clusters of states that may improve the effectiveness of a feature. Although when the number of features is high, this approach might increase the development phase, because the generated clusters must be interpreted and validated (Bishop and Nasrabadi, 2006, Coates and Ng, 2012).

Another option is to use a simple latent model for selected features to create a namespace that includes possible interaction terms (Murphy, 2012). Initially, interactions can be created explicitly for selected terms (e.g. a function of age and gender) or instead all possible interactions can be created for all sub-groups of features using a quadratic or cubic function. Then a latent modelling approach, like non-negative matrix factorisation, can be used to characterise the correlations between features, with the assumption that the feature lives on a low dimensional linear manifold.

Also, it is possible to effectively include many feature states using a multilayer learning architectures, like Deep Neural Network (DNN) or using a data fusion approach, like Bayesian approaches. Firstly, in deep architectures (Section 4.2.7.1), compactly represented functions can sufficiently be represented using a very large and deep architecture; providing that enough data is supplemented and an appropriate architecture is designed (Bengio et al., 2009). Also, we can use a Bayesian framework to introduce

several sources of data for variables that are noisy versions of the true values (Koller and Friedman, 2009).

For instance, amplification of a large group of features (or states) with complex interdependencies using DNN may be accomplished by convolutional models of feature (Lee et al., 2009), embedding layers (Esteban et al., 2015), groups of crossed features (a.k.a. interaction terms) (Cheng et al., 2016), or latent features, in order to increase their discriminative power (Abadi et al., 2016).

7.3.4 Temporal Features

Healthcare administrative data are severely unbalanced regarding the amount of *longitudinal* (panel) data per patient and their distribution over the years. Statistical methods are not equipped to handle these types of unbalances directly. Therefore, the life-table approach, which is also used in survival analysis modelling, was used for modelling time-to-event (Appendix A.1.4), to keep track of temporal events with minimal bias and sparsity (Singer and Willett, 2003). Based on the previous studies and the initial statistical analyses, four levels of temporal features were generated: 0-30, 30-90, 90-365 and 365-730 days (Bardsley et al., 2013, Billings et al., 2006a, 2012, 2013, NHS, 2011, Mullins et al., 2006).

Moreover, following four main groups of features were initially generated from the inpatient database: *cross-sectional*, *trigger-event*, and *longitudinal* features. Generated features were stored into four separate groups (database tables) (Appendix A.4):

- *cross-sectional*: *prior-history* features (565 features)
  - prior-3-years (219 features)
  - prior-90-Days (173 features)
  - prior-12-months (173 features)
- *trigger-event* (173 features)
- *prediction-period* (173 features)
- *longitudinal*: Monthly temporal levels (151 features)

The *cross-sectional* groups include features for each patient prior to admission at the *trigger-event*, which is a mix of demographic, admission and clinical features. Similarly, the *trigger-event* group consists of features at the *trigger-event* for each patient.

The *longitudinal* features generated for the whole time-frame, using monthly temporal levels.

7.3.5 Combine Features

At this stage of analysis, there might be highly sparse and correlated features, which can be combined, in order to increase their overall impact. There are two main motivations in combining the features: dimension reduction of related characteristics; and reducing sparsity. It is especially useful when there are multiple indicators of a patient attributes, like diagnoses, consultation and drugs. Encoding a feature combination method allows to gather all the evidence from multiple sources of information.

For instance, indicator about frailty or respiratory conditions can be gathered from the type of consultancy, diagnoses, operations and complexities. And, it can also be beneficial to stratify them based on demographics and other indicators. This type of approach is usually more exploited, when a risk score is being developed to capture a very particular high-risk group of patients, like Elixhauser Comorbidity Index (ECI) (Section 2.3).

Similar to the other combination sub-step (Section 7.3.3) related features can be combined using an unsupervised technique, a deep architecture or a data fusion method.

7.4 Step IV: Feature Selection

Since numbers of generated features were very high, a feature reduction strategy was needed. At this stage, implementation of a classification, a dimensionality reduction method, like PCA (Hotelling, 1933, Pearson, 1901), or a type of Ridge Regression (Hoerl and Kennard, 1970) can not guarantee a good fit.

Initially, highly stationary features were dropped. Then, features that were highly linearly correlated were excluded. Afterwards, the features with very small predictability for emergency readmission were excluded. These steps are outlined in Figure 7.5 and explained in the following subsections.

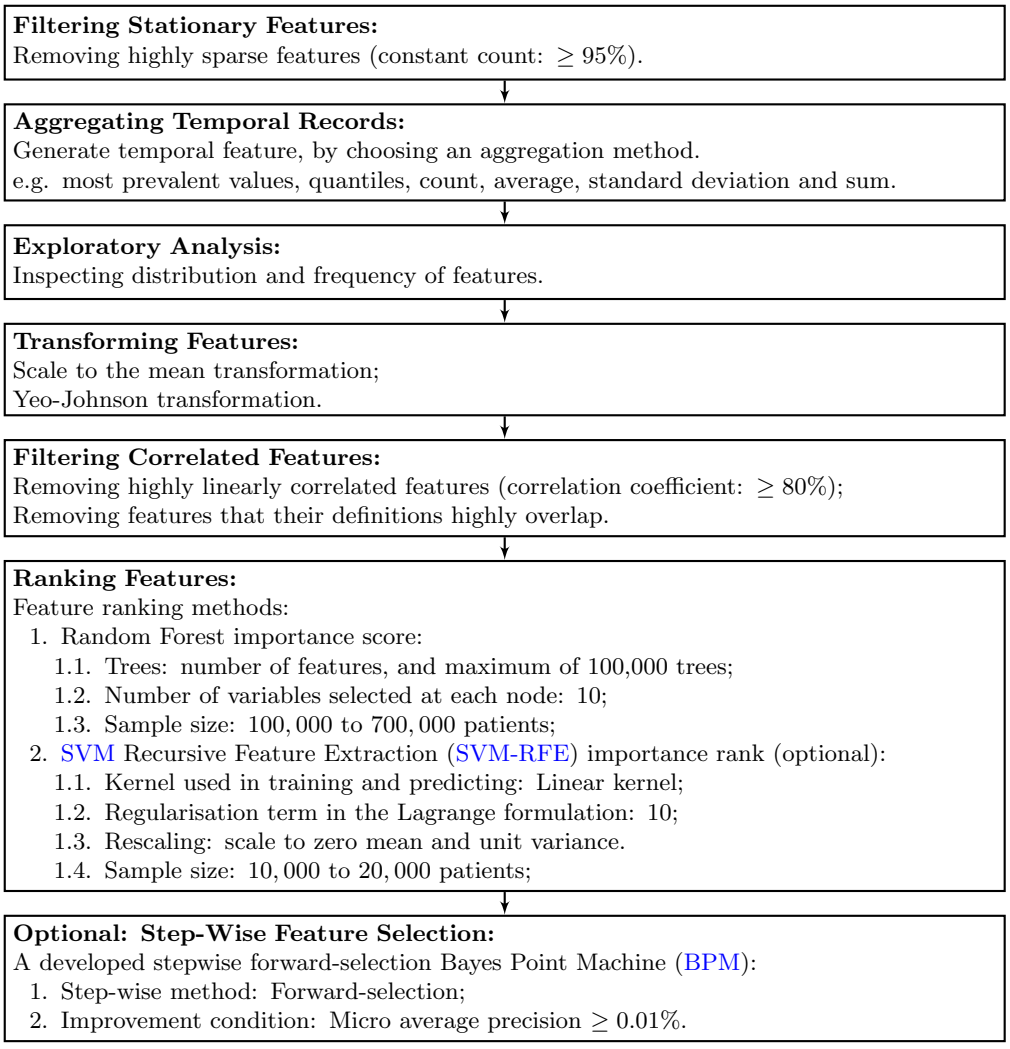

FIGURE 7.5: Healthcare pre-processing framework: the feature selection step

7.4.1 Filtering Stationary Features

Firstly, in order to reduce sparseness and invalid features, highly stationary ones were withdrawn (Kuhn and Johnson, 2013). The features that had constant counts less than or equal a threshold were filtered out, to exclude highly constants and near-zero variances. In this study, the threshold was set to $\geq 95\%$, because the linear correlations to 1-year emergency admission were $\leq 50\%$. The near-zero variance rules are presented in below:

- Frequency ratio: The frequency of the most prevalent value over the second most frequent value to be greater than a threshold;

- Percent of unique values: The number of unique values divided by the total number of samples to be greater than the threshold.

7.4.2 Aggregating Temporal Records

Finally, the sub-states of each temporal feature[4] must be aggregated using a function that can capture the temporal state of patient appropriately. In this research, a number of aggregation functions have been implemented, including most prevalent values, quantiles, count, average, standard deviation and sum.

This step can lead to one or more sub-features, but the sub-features must be checked to have low intercorrelations before including into the model.

A more complex approach for aggregating the sub-states is to use a static or dynamic weighted polynomial. The weighted polynomial combines the sub-states by weighting the time dimension. For instance, if a polynomial weighting with prevalence ranking is combined, it only outputs the sub-states that have been more prevalent recently. However, if there is a very little number of sub-states, observations or patients, polynomial weighting might introduce more bias into our model.

7.4.3 Exploratory Analyses

After filtering features, the distribution and frequency of the features were investigated, in order to identify any anomalies. Descriptive summaries were produced for the continuous features to investigate their distributions' properties. In addition, descriptive summaries were produced for the discrete features to identify the most frequent categories and missing counts.

7.4.4 Transforming Features

There is a wide range of transformation methods that can be applied to features to improve normality of their distributions and equalise variances. Although transformation methods may improve model convergence, increase features effectiveness, deal

[4]For instance, feature x can have multiple sub-states for different *episodes* of care during the 0-365-days temporal state: $substates_{x,\_365\_days} = \{1, 20, 100, 118, 213\} \ \wedge \ substates_{x,\_365\_days} \in feature_x$

with outliers or meet some modelling assumption, but they must be used with caution (Osborne, 2010, Ohara and Kotze, 2010).

The first step is to test distributions of features. There are several ways to identify if a feature deviates significantly from the Normal distribution: visual inspection, and statistical tests. Examples of basic statistical tests, which can be easily applied, are skewness, Kurtosis, Q-Q plot[5] and P-P plots[6]. Another type of statistical approach is inferential tests, like Kolmogorov-Smirnov test, Lilliefors corrected Kolmogorov-Smirnov test, Anderson-Darling test, Shapiro-Wilk test, Cramer-von Mises test, DAgostino skewness test, Anscombe-Glynn kurtosis test, DAgostino-Pearson omnibus test, and the Jarque-Bera test (Ghasemi et al., 2012, Salkind, 2006).

The next step is to decide on the transformation function. When the error structure of our data is simple, a transformation can be very useful to improve the model performance. Examples of popular transformation methods are scale to mean, Square Root, Log, Arcsine, Multiplicative Inverse (reciprocal), Box-Cox and Yeo-Johnson transformations. However, transformations do not necessarily guarantee to stabilise the variance or to make a better linear model. Also, transformations make model interpretation harder, and can negatively impact the relationship between correlated features in the model.

In this study, two feature transformations have been applied to the continuous features: the scale to mean and the Yeo-Johnson (Yeo and Johnson, 2000). Both methods can be used to transform the data, to improve normality. Firstly, the scale to mean method was applied to standardise features by removing the mean and scaling to the unit variance. Moreover, the Yeo-Johnson method is from the family of *power transformations* and is very similar to the box-cox (Sakia, 1992), which can be applied to both negative and positive numbers. The Yeo-Johnson power transformation family is defined in below (Eq. 7.1).

$$\Psi(\lambda, y) = \begin{cases} ((y+1)^{\lambda} - 1)/\lambda & if\ \lambda \neq 0, y \geq 0 \\ log(y+1) & if\ \lambda = 0, y \geq 0 \\ -[(-y+1)^{2-\lambda} - 1)]/(2-y) & if\ \lambda \neq 2, y < 0 \\ -log(-y+1) & if\ \lambda = 2, y < 0 \end{cases} \tag{7.1}$$

[5]The Q-Q Plot compares the quantiles of a data distribution against the quantiles of a standard theoretical distribution.

[6]The P-P plot compares the empirical distribution of a data against a specified theoretical distribution.

, where y is a list on strictly *power positive* numbers, and λ is the *power* parameter.

7.4.5 Filtering Correlated Features

Then, features with high linear correlation to 1-year readmission were excluded. The Pearson correlation coefficient (Pearson, 1895) was calculated for all the pair of variables to measure linear dependence between them (Eq. 7.2). The linear correlation coefficient was set to $\geq 80\%$, because the applied algorithms including the rankings are sensitive to correlated features. The Pearson correlation coefficient is presented in below (Eq. 7.2).

$$\rho(X, Y) = \frac{cov(X, Y)}{\sigma_X * \sigma_Y} \tag{7.2}$$

, where *cov* represents the covariance, σ is the standard deviation of a feature and the features are represented by X and Y symbols.

7.4.6 Ranking and Selecting Features

In this sub-step, features were ranked using a selection of methods, and were chosen based on literature review (Section 4.2 and Section 2.2) and several tests and trials. The considered methods include Breiman (2001) version of Random Forest (RF), Gradient Boosted Regression Trees (GBRTs) by Friedman (2001), Randomized Logistic Regression (RLR) with $L1$ regularisation (Xing et al., 2001), Guyon et al. (2002) version of Support Vector Machine Recursive Feature Extraction (SVM-RFE) and Herbrich et al. (2001) version of stepwise Bayes Point Machine (BPM).

In this study, the RF importance score was implemented for all models. The SVM-RFE importance ranking and the stepwise BPM have been applied for the emergency readmission modelling. In all the models, excluding the stepwise BPM, six trials have been executed, with three different settings for each ranking method. Then, the features were inspected and selected manually using the average ranks. Thereafter, the stepwise BPM model for the emergency readmission models was run. Based on the average of importance, initially the prior-3-years *cross-sectional* features were inputted to the stepwise BPM, then other features were added.

Firstly, the applied RF (Section 4.2.4.1) importance score uses the Breiman algorithm with gradient boosted regression trees, Gini index, balanced samples, the minimum leaf split size of hundred, and minimum leaf size of fifty. The number of trees in the forest was set to the maximum of 100,000 trees. The applied RF is a non-linear method and is an implementation of Breiman's algorithm (Breiman, 2001), which applies significance test criteria (Hothorn et al., 2010, R-Project, 2010). It performs recursive univariate splitting and selects covariates based on the significance test. The significance test approach, unlike the maximising information, does not suffer a systematic tendency towards covariates with many possible splits or many missing values. However, highly similar features and linearly correlated features were excluded in the prior step, since the applied algorithm is sensitive to correlated features. Although the applied RF algorithm is not very data dependent, the results can vary largely by the choice of hyper-parameters (number of variables selected at each node and number of trees) (Díaz-Uriarte and De Andres, 2006, Gromski et al., 2014, Verikas et al., 2011).

Moreover, the applied SVM-RFE (Section 4.2.5) importance ranking was used with the sample sizes of 10,000 to 20,000, because of the high computational cost of SVM algorithms. The SVM-RFE algorithm proposed by Guyon (Guyon et al., 2002) was applied to rank features recursively using SVM. The SVM-RFE returns ranking of the features by training a SVM with a linear kernel and removing the features with the smallest ranking criterion.

Furthermore, the stepwise BPM (Section 4.2.6) model was run using a stepwise approach with backward elimination and forward selection, which was developed explicitly for the purpose of this study. The stepwise algorithm was defined to use the micro-average precision $\geq 0.01\%$, as the selection criteria in each step.

7.5 Concluding Remarks

What is already known?

- The pre-processing of healthcare administrative data is a very challenging task, due to inconsistent data recording practices and incomplete data validation processes by data warehouse administrators.

- A wide range of administrative data is recorded by healthcare sectors for patients. But, present healthcare models mainly lack a systematic feature generation technique, which can effectively search for potential features and assess their importance.
- In literature, no pre-processing framework have been proposed, that can deal with the pre-processing or the feature generation in healthcare analytics.

What this phase of research adds?

- A generic and novel healthcare pre-processing framework have been proposed that systematically define analyses steps and approaches, that may be carried out to manage meta-data, prepare data, engineer features and select top factors.
- The data management and preparation step are customised to England primary and secondary care data sources, but the steps are generic and applicable to many healthcare systems.
- The feature engineering is one of the main time-consuming steps in many healthcare modelling problems, and this framework provides a universal methodological approach to generate and select features.

In conclusion, the proposed healthcare pre-processing framework defines four main steps of data management and analyses. Initially, a number of steps were defined to ingest data and manage the metadata. Next, several procedures were specified to clean and validate the data. After that, a feature pool was generated using a generic and systematic way. Finally, the redundant features were filtered out, and then features sorted by importance and top factors were selected.

Finally, the toolkits that are presented in Chapter 10 provide a set of mechanisms to carry out part of this framework. It is a comprehensive, user-friendly and free software package, and is released for public use and incremental development.

In the next chapter, a predictive model for hospital emergency readmission is presented.

Chapter 8

Phase II: Predictive Risk Modelling of Hospital Readmission (ERMER)

About half of hospital readmissions can be avoided with preventive interventions, and healthcare commissioners are looking for more powerful predictive risk tools to target the high-risk patients. Developing decision support tools for identification of patients' emergency readmission risk is an important area of research (Section 2.2). Because, it remains unclear how to design features and develop predictive models that can adjust continuously to a fast-changing healthcare system and population characteristics. The objective of this phase of study was to develop a generic Ensemble of Bayesian emergency readmission risk models to better adjust for prior probabilities and various population characteristics.

Most existing decision support tools, that are based on hospital administrative data, use Logistic Regression (LR) or Coxian Phase-type Distribution models (C-PHDs), and have very limited capability (Section 6.2). This phase of research developed an Ensemble generative risk model of emergency readmission within a year to the England's hospitals. The Machine Learning Ensemble method is a powerful technique (Section 4.2.2), which uses a finite set of weaker models and an algorithm to combine and optimise the performance of the Ensemble model.

Based on the defined healthcare pre-processing framework introduced in the *Phase-I* of research (Chapter 7), features were generated, filtered and ranked. Thereafter, a number of sub-models based on population characteristics were trained using a Bayes Point Machine (BPM) approach (Section 4.2.6). Afterwards, an optimised Ensemble

model of these sub-models was generated. The developed Ensemble Risk Model of Emergency Admissions (ERMER) was trained and tested using three time-frames: 1999-2004, 2000-05 and 2004-09, each of which includes 20% of patients admitted within the *trigger-year* (Section 5.3). In addition, a development toolkit is supplemented to ease the validation and adaptation of the ERMER (Chapter 10).

Furthermore, the benchmarking comparisons were made for different time-frames, sub-populations, risk cut-offs, risk bands, and top risk segments. The ERMER has 71.6%-73.9% precision, 88.3%-91.7% specificity, and 42.1%-49.2% sensitivity across different time-frames. Also, the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) of the proposed model is 75.9%-77.1%.

The proposed decision support tool performed considerably better than the previous modelling approaches, and it was robust and stable with high precision. Moreover, the healthcare pre-processing framework and the Ensemble Bayesian approach allowed the ERMER to continuously be adjusted to new significant features, different population characteristics and changes in the care system.

The chapter is structured as follows. Firstly, we describe the data and then define the process of selecting an optimised number of features that are highly significant. After that, the applied BPM algorithm is presented and the Ensemble model is specified. Then, we discuss the results of training, testing and validation, as well as the benchmarking against CPM (DH, 2006), PARR (Billings et al., 2006a), and Billings et al. (2013) models (Section 6.2). Finally, concluding remarks are presented, and then the specification of the ERMER development toolkit is provided separately in Chapter 10.

8.1 Data and Features

The healthcare pre-processing framework (Chapter 7) has been applied, in order to sample, clean and treat input data, combine episodes, and engineer features. Firstly, the inpatient table of the HES database is selected. Then, the patients with admission within the *trigger-year* were identified. Afterwards, 20% of patients were selected, and all their records were extracted and sorted based on the patient and admission date. Thereafter, the records were validated and filtered. Finally, the selected sample was divided into two sub-samples for training and testing. Section 7.1 and Section 7.2 outlines the metadata management, sampling and validation stages, which were completed as part of the pre-processing stage but not presented here (HSCIC, 2011, 2014d,e).

TABLE 8.1: ERMER: The sub-population settings for the benchmarking

Sub-population	Model Setting
Sub_PARR-2-Settings	Population setting for the PARR-2 model (Billings et al., 2006a, Lewis, 2011): Age: 65+; Trigger admission: Emergency admission.
Pop_IPAEOPGP	Population setting for the Billings et al. (2013) models: Age: 18 to 95; Trigger admission: Emergency admission.
Pop_Any-Acute	All the population for the selected sample: Trigger admission: Emergency admission.

In the data management and preparation (Section 7.1 and Section 7.2), firstly, the extracted data was sorted by patients and the order of *episodes*. Next, invalid records were excluded. Thereafter, several corrections and imputations were carried out on dates, Healthcare Resource Groups (HRGs) and demographics. After that, some of the continuous features were converted into discrete with other features to better capture non-linear interactions. And, some of the discrete features were re-categorised into bigger groups to reduce sparseness and overfitting risk. Finally, *episodes* of care were grouped into *super-spells*, and then a feature pool was generated (Section 7.3).

Furthermore, Kernel classifiers, such as the BPM and the Support Vector Machine (SVM), are usually resistant to over-fitting, because of an implementation of weight regularisation (Cawley and Talbot, 2007, 2010). However, since numbers of generated features were very high, features were selected based on two different methods (Section 7.4): a Random Forest (RF) importance score and a SVM importance ranking.

Finally, a stepwise BPM procedure was developed using the forward-selection approach, to evaluate significance of the top features. Based on the average of importance, initially three years *cross-sectional* features were included, then other features were added. A summary of the features' definitions and their importance rankings are provided in Appendix A.6.1.

TABLE 8.2: ERMER: Main categories of all the initially defined features

Category	Sub-category
Administrative	Admission: patient classification; number of episodes and spells; admission, readmission and discharge times; source and methods of admission and discharge. Bed days: duration of spells; preoperative and post-operative duration. Geographical: provider code; region of treatment. ID: patient identification, and admission timeframe number. Speciality: speciality of consultant; palliative cares. Waiting time: admission waiting time.
Clinical	Diagnosis: Charlson comorbidity groups; Elixhauser comorbidity groups; frequent categories of diagnoses; HSCIC-CCI (Aylin et al., 2010, Bottle et al., 2011, HSCIC, 2016c); PARR's HRGs *reference* conditions, using version 3.5. Operation: operation groups; number of operations; frequent categories of operations.
Patient	Demographic: age; deprivations; ethnicity; gender.

For the purpose of testing, validation and benchmarking, five different combinations of training and testing sub-samples are considered (Table 5.2). In addition, after running

the model, results are filtered based on three different sub-populations for benchmarking purpose (Table 8.1).

8.2 Model

Firstly, one main model (*cond_main*) and four conditional sub-models were specified, with significantly diverse populations that represent unique clinical and behavioural categories (Figure 8.1). The conditional sub-models includes: prior 12-month acute *spells* (*Cond_Prior-Acute-12-month*), prior 12-month operation (*Cond_Prior-Oper-12-month*), prior *spells* (*Cond_Prior-Spells*) and age 65+ (*Cond_Age-65p*).

TABLE 8.3: ERMER: The conditional sub-models based on sub-populations

Model	Condition
Cond_Main	Main Model: NO condition
Cond_Prior-Spells[a]	Prior spells: 0, or > 0
Cond_Prior-Acute-12-month	Prior emergency within 12-month: 0, or > 0
Cond_Prior-Oper-12-month	Prior operation within 12-month: 0, or > 0
Cond_Age-65p[a]	Age: < 65, or ≥ 65

[a] *Cond_Prior-Spells* and *Cond_Age-65p* sub-models are only applied for the *Pop_Any-Acute* and *Pop_Any-Acute-NO-Mental* modelling group.

Afterwards, they were trained and tested across the sub-sample combinations (Table 5.2). Considering that the filtered features are more relevant for the main model, the sub-models have very different performances, but features weights and overall performances are stable across all tests and validations.

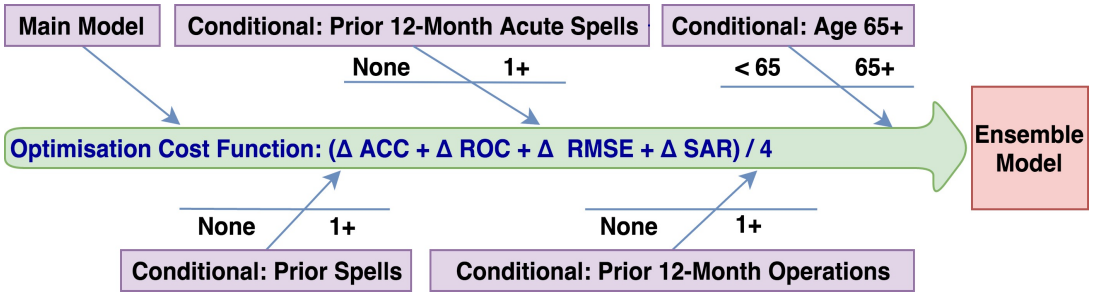

FIGURE 8.1: ERMER: Abstract representation of the proposed Ensemble model

Then, we decided to use an Ensemble model, to improve the performance of the decision support system. Three main challenges in our Ensemble modelling were: method of constructing sub-classifiers, weighting the classifier, and defining a cost function and optimising it. Based on the background research and multiple trials, a weighted average ranking method was constructed, in addition to a heuristic method, to optimise the weights of sub-classifiers (Murphy, 2012, Rokach, 2010, Sammut and Webb, 2011, Sewell, 2008, Zhou, 2012).

In another word, the ERMER partitions the data instance space, based on some populations similarities (sub-models). Then, it uses *data envelop analysis* methodology (Charnes and Cooper, 1984) to assign weights to different classifiers (Rokach, 2005). In this research, we refer to this weight function as the cost function, because we applied a search technique to optimise the weights that are assigned to each sub-model.

The cost function for the optimisation was defined as a normalised combination of four performance metrics (Alvarez, 2011, Brown, 2011, Fukunaga, 2013):

- ACC (Accuracy): The fraction of the sum of True Positives (TPs) and True Negatives (TNs) from the total.
- AUC of ROC: The Area Under the Curve of ROC curve, which y-axis is the True Positive Rate (TPR) and x-axis is the False Positive Rate (FPR).
- RMSE (Root Mean Square Error): Root of averaged of squared difference of the predicted outcomes ($\hat{Y}$) from the actual outcomes (Y).
- SAR (Squared Error, Accuracy, and ROC Area): a robust metric for performance.

The applied Ensemble algorithm (Algorithm 1) uses a bidirectional hill-climbing algorithm with a greedy initial solution set ($models_{ensemble}$) to generate an optimised Ensemble model from the sub-models. Firstly, it generates an initial solution based on the main model, and one other sub-model with the highest AUC of ROC. Then, a bidirectional hill-climbing (Russell and Norvig, 2002) heuristic was applied to optimise the average of the four performance metrics, through iterations, trials ($trials$) and across samples ($samples$).

The hill-climbing method is a greedy sequential search with forward and backward passes, where the *learning rate* for each performance metric can be tuned manually prior to the execution. The *learning rate* in the algorithm (Algorithm 1) defined using $alpha_{ensemble_{min}}$ for the performance indicators (Caruana et al., 2004, Fukunaga, 2013, Opitz and Maclin, 1999).

The sub-models in the Ensemble heuristic are selected using a Bagging Ensemble (selection with replacement). Then, the sub-models are combined using a *mean* combiner, which is the approximate posterior probability based on the weighted average of the risk scores, without any additional training. When the first run of the algorithm with the defined iterations, trails and across samples is finished; then, the second run is executed using the best solutions of the first round, with less sensitive limits and thresholds.

Algorithm 1 The ensemble modelling algorithm (part 1).

Require: ▷ Set samples, modelling groups and sub-models

$samples \leftarrow$ {"*sample*-1","*sample*-2","*sample*-3",
"*sample*-1-*train*-2-*test*","*sample*-1-*train*-3-*test*"}
$groups \leftarrow$ {"*Pop_Any-Acute*","*Pop_Any-Acute-NO-Mental*"}
$models \leftarrow$ {"*Cond_Main*","*Cond_Prior_Spells*",
"*Cond_Prior_Acute_12_month*","*Cond_Prior_Oper_12_month*",
"*Cond_Age_65p*"} ▷ Set limits on weight of each sub-model
$weight_{ensemble_{sum_{max}}} \leftarrow 2$
$weight_{ensemble_{sum_{min}}} \leftarrow 20$ ▷ Second iteration of the algorithm: 300
$weight_{ensemble_{max}} \leftarrow 15$ ▷ Second iteration of the algorithm: 150 ▷ Set limit on iterations of searches
$search_{trials_{max}} \leftarrow 40$ ▷ Second iteration of the algorithm: 20
$search_{iterations_{max}} \leftarrow 150$ ▷ Second iteration of the algorithm: 150 ▷ Set the thresholds for model selection step and the predicted probability cut-off
$alpha_{ensemble_{min}} \leftarrow 0.0005$ ▷ Second iteration of the algorithm: 0.0001
$alpha_{model_{min}} \leftarrow 0.50$
▷ Other notations
TP_x : True positive of model x with cut-off point $alpha_{model_{min}}$
FP_x : False positive of model x with cut-off point $alpha_{model_{min}}$
FN_x : False negative of model x with cut-off point $alpha_{model_{min}}$
TPR_x : True positive rate of model x with cut-off point $alpha_{model_{min}}$
FPR_x : False positive rate of model x with cut-off point $alpha_{model_{min}}$

▷ The ensemble modelling algorithm:
procedure EnsembleModels($models$)
    $models_{ensemble} \leftarrow$ InitialSolution($models$) ▷ Set the initial greedy solution
    $selected_{model} \leftarrow$ MainSearch($models_{ensemble}$) ▷ Run the main heuristic search

procedure InitialSolution($models$)
    $model \in models$
    $model_{ensemble} \in models$
    $model_{max_{AUC}} \leftarrow MAX_{model_{AUC}}(models),\ WHERE\ model \neq$ "*Cond_Main*"
    return {"*Cond_Main*"} $\cup\ model_{max_{AUC}}$

procedure ACC(x)
    return $\frac{TP_x+TN_x}{TP_x+TN_x+FP_x+FN_x}$

procedure AUC(x)
    return $\int_{-\infty}^{+\infty} TPR_x FPR'_x dx$

procedure RMSE(x)
    return $\sqrt{\frac{\sum_{i=1}^{n}(\hat{Y}_x - Y_x)}{n}}$

procedure SAR(x)
    return $\frac{ACC_x+ROC_x+(1-RMSE_x))}{3}$

Algorithm 2 The ensemble modelling algorithm (part 2).

```
1: procedure MainSearch(models_ensemble)
2:     models_selected = {}
3:     for all s ∈ samples do                                   ▷ Run for each sample
4:         for all g ∈ groups do                                ▷ Run for each modelling group
5:             for t ← 1, search_trials_max do                  ▷ Run trials
6:                 for all model_ensemble ∈ models_ensemble do  ▷ Run for each initial solution
7:                     acc_0 ← auc_0 ← rmse_0 ← sar_0 ← 1
8:                     for i ← 1, search_iterations_max do      ▷ Run iterations
9:                         acc_i ← ACC(x)
10:                        auc_i ← AUC(x)
11:                        rmse_i ← RMSE(x)
12:                        sar_i ← SAR(x)
13:                        improvement ← (acc_i − acc_{i−1} >= alpha_ensemble_min) * 1
                               +(auc_i − auc_{i−1} >= alpha_ensemble_min) * 1
                               +(rmse_i − rmse_{i−1} >= alpha_ensemble_min) * 1
                               +(sar_i − sar_{i−1} >= alpha_ensemble_min) * 1
14:                        degradation ← (acc_i − acc_{i−1} < −alpha_ensemble_min) * 1
                               +(auc_i − auc_{i−1} < −alpha_ensemble_min) * 1
                               +(rmse_i − rmse_{i−1} < −alpha_ensemble_min) * 1
                               +(sar_i − sar_{i−1} < −alpha_ensemble_min) * 1
15:                        if i == 1 then                       ▷ Select a step
16:                            backwardStep ← True
17:                        else
18:                            if backwardStep == True then
19:                                if degradation < 0.5 then    ▷ Forward
20:                                    model_ensemble ← model_ensemble ∪ selected_model
21:                                else                         ▷ Switch
22:                                    backwardStep ← False
23:                                    switchStep ← True
24:                            else
25:                                if backwardStep == False then
26:                                    if improvement ≥ 0.5 then ▷ Switch
27:                                        switchStep ← True
28:                                        backwardStep ← True
29:                        if backwardStep == True then         ▷ Backward selection
30:                            if switchStep == True then
31:                                counter ← 0
32:                            model_ensemble ← model_ensemble \ model_ensemble_{counter+1}
33:                        if backwardStep == False then        ▷ Forward selection
34:                            model ∈ models
35:                            model_ensemble ← model_ensemble ∪ model
36:                    models_selected ← models_selected ∪ model_ensemble   ▷ Add selected model
37:    return models_selected
```

$$
\begin{aligned}
model_{ensemble} = & Mean\{Cond\_Main + Cond\_Age\text{-}65p_0 + \\
& 9\ Cond\_Age\text{-}65p_1 + 4\ Cond\_Prior\text{-}Oper\text{-}12\text{-}month_0 + \\
& 2\ Cond\_Prior\text{-}Oper\text{-}12\text{-}month_1\}.
\end{aligned} \tag{8.1}
$$

Finally, the best performing Ensemble model, with the minimum number of unique sub-models is selected (Blumer et al., 1987). The optimised Ensemble Risk Models of

Emergency Admissions (ERMER) based on our datasets is defined in Eq. (8.1). In this equation, sub-model subscripts, like $Age\text{-}65p_0$, represent the conditional state, and the coefficients are the weights in the Ensemble *mean* combiner.

8.3 Results

Four stages of performance checks were performed across test sub-samples to assess the goodness of fit. Firstly, a learning-curve plot of training errors versus the number of training points for sub-models was generated, using training sub-samples. The learning-curve is a function of the number of training points and the prediction accuracy rate, and it allows investigating the effect of sample sizes on the performance of models (Murphy, 2012, Nordhausen, 2009). Figure 8.2 demonstrates that training sub-sample size greater than 40,000 patients contributes very little to the sub-model performance.

TABLE 8.4: ERMER: The top significant features in the submodels

Feature	Calculation [a]
Sum of number of operations within 90 days & at the trigger.	$Count_{spell}(Unique_{spell}(opertn\_nn_{episode}))$
Count of recoded main speciality of state 'Maternity' in the past & at the trigger.	$mainspef_{spell} \in \{501, 560, 610\}$
Count of recoded main speciality of state 'Gynaecology' in the past & at the trigger.	$mainspef_{spell} \in \{502\}$
Count of recoded main speciality of state 'General' in the past & at the trigger.	$mainspef_{spell} \in \{300, 600, 620\}$
Having recoded gender of state 'Female'.	$sex_{patient} == 2$
Age of patient at the trigger.	$startage_{spell}$
Average of post-operative durations at the trigger.	$posopdur_{spell}$
Count of the acute admission method between 12 to 36 months, & within 90 days.	$Count_{spell}(admimeth_{spell} \in \{21, 22, 23, 24, 25, 2A, 2B, 2C, 2D, 28, 31, 32, 81, 82, 83, 84, 89, 98\})$
Average of spells durations in the past & at the trigger.	$Mean_{spell}(Max_{episode}(epidur))$
Average of gaps between admissions in the past.	$admidate_{spell_i} - dismeth_{spell_{i-1}}$
Having recoded ethnicity of state 'NA'.	$ethnos_{patient} \in \{S, 8, L, G\}$
Average value of the Charlson Index in the past.	$Mean(CharlsonIndex_{Dr\_Foster\_CCI}(diag\_nn_{spell}))$

[a] Refer the the HES dictionary for the raw variables' definition (HSCIC, 2010).

Moreover, the effects of complexity levels were investigated for the main model (*Cond_Main*) using F-score versus the number of features, using training sub-samples. The plot of the effects of complexity levels shows how the stepwise addition of top features changes the prediction performance of a model. Figure 8.3 shows that adding up to 18 features (Table 8.4), from the top significant features, improves the model's performance significantly. However, the gains then become very small (on average 0.005 change in AUC percentage). More detailed specifications of the complexity levels are provided in Section 8.3.1. Moreover, the presented learning-curve and the complexity plots are for *Sample-1*, although the results are consistent across all other samples and sub-samples.

Thereafter, the convergences of the sub-models were tested using an iterative fitting, using train sub-samples, in order to assess over-fitting and variations in convergence.

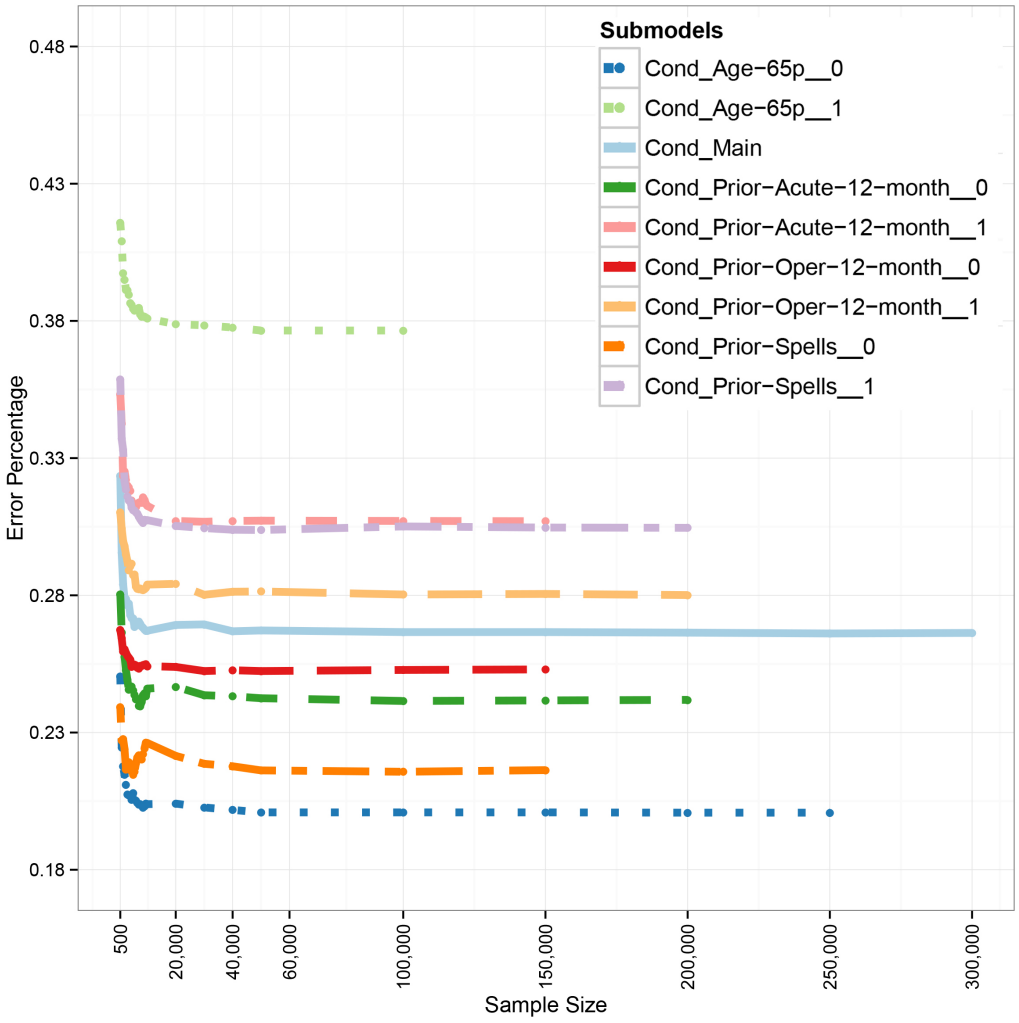

FIGURE 8.2: ERMER: Learning curve plot of sub-models, micro-average error vs. number of training points (*Sample-1*)

Figure 8.4 shows that after the first few iterations, all sub-models converge quickly after about forty iterations, and the weights differences become very small.

Furthermore, a *5-fold* cross-validation (Murphy, 2012) algorithm was implemented for all three test sub-samples (Table 5.2). The *K-fold* cross-validation splits each test sub-sample into five equal-sized random samples. Then, *K - 1* folds are used for training and the *K-th* fold is used for validation. Finally, the *K-fold* performance output is generated after the cross-validation cycled through all *K* combinations of splits.

Figure 8.5 exhibits very small standard deviations in accuracy, mean of negative log-probability and AUC, for the sub-models' cross-validations.

Finally, the profiling was done using three test sub-samples, based on population characteristics and performance indicators (Table 8.7 and Table 8.8). Table A.61 and Table

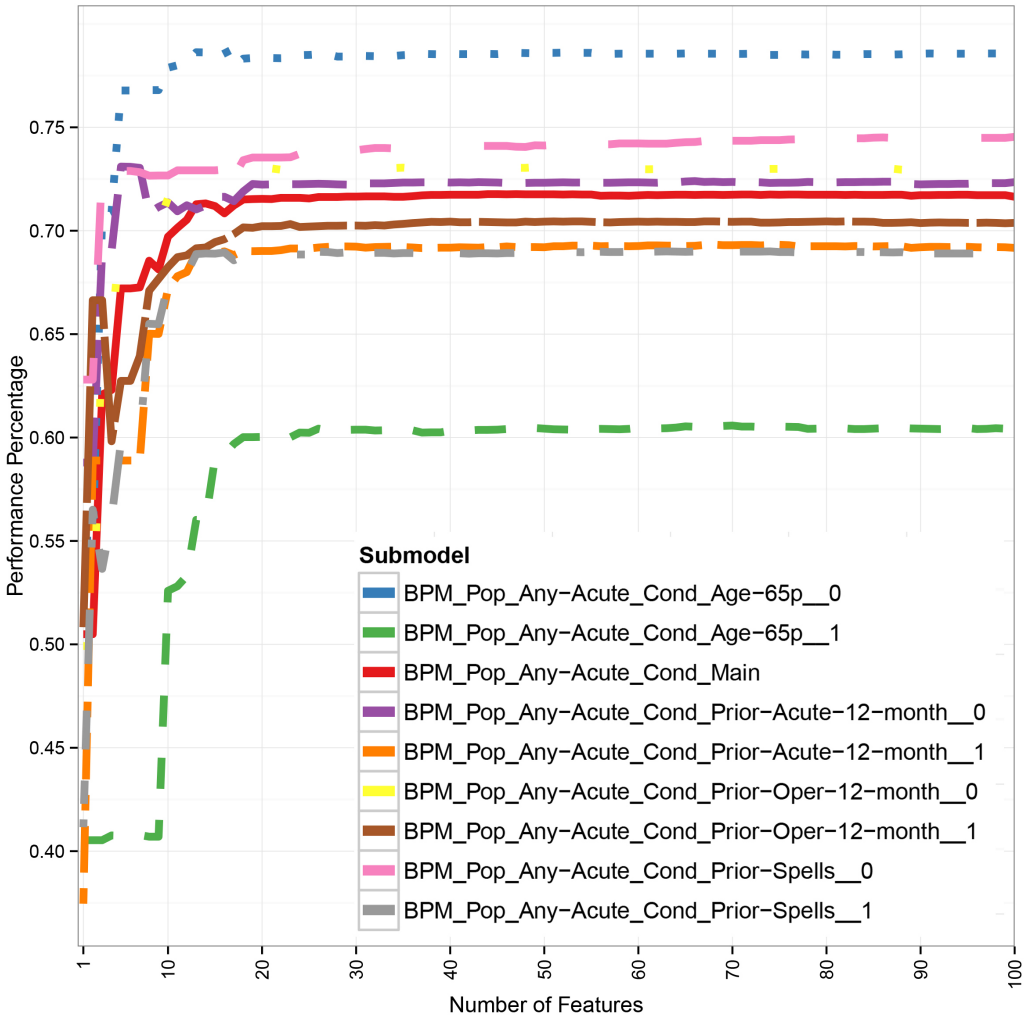

FIGURE 8.3: ERMER: Complexity analysis of sub-models, the F-score vs. number of features (*Sample-1*)

A.62 demonstrate the weights of the features for each sub-model, as well as the features' definition, encoded category and temporal state.

In the following subsection, more insight is provided into the complexity analysis measures, including the RMSE that is used in the Ensemble Cost function. Then, in the next section, the benchmark against previous models is discussed.

8.3.1 Effects of complexity

The Root Mean Squared Error (RMSE) is one of the performance indicator in the ERMER cost function for the Ensemble method. Figure 8.6 demonstrates the number of indicators that capture complexity of the sub-models across different subpopulations

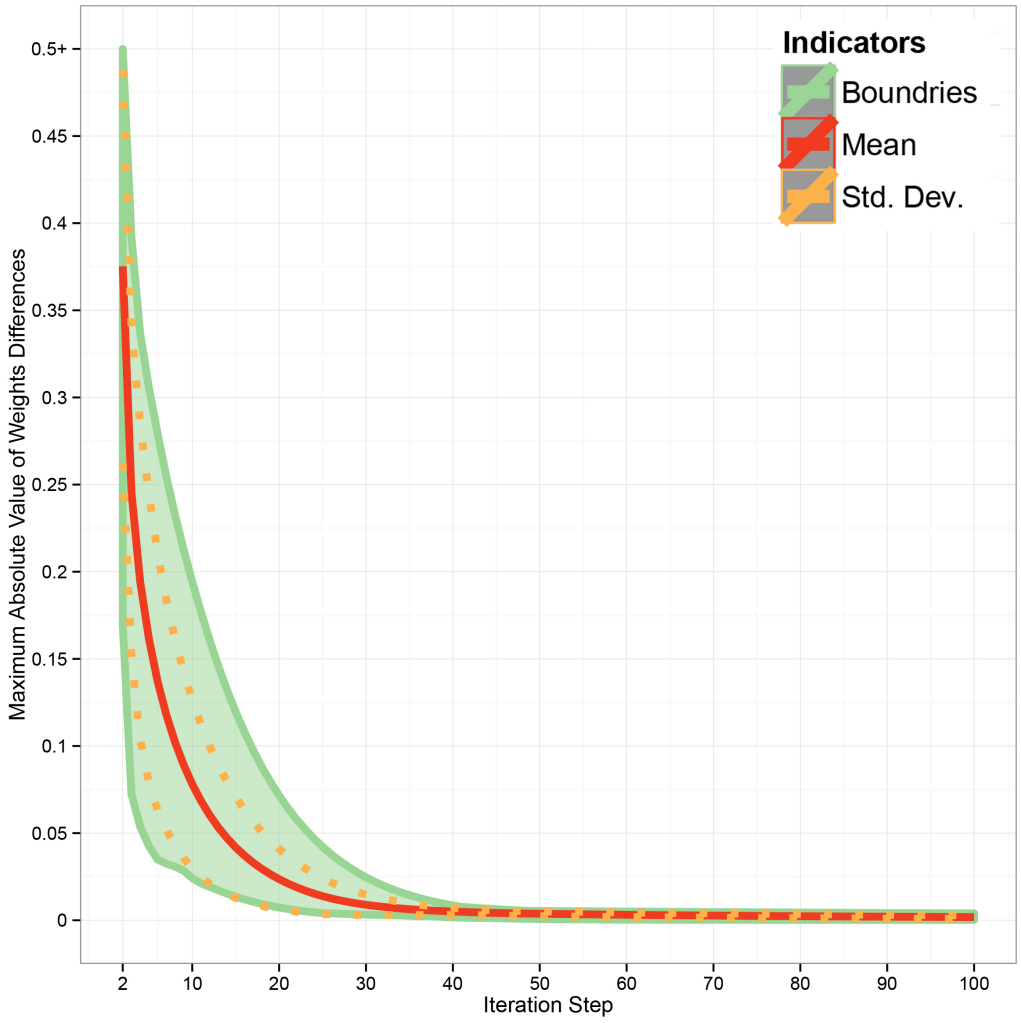

FIGURE 8.4: ERMER: Average and range of convergences (all samples)

and data-frames. Two types of errors that were included in the analyses are the Mean Squared Error (MSE) and expected prediction error, a.k.a. the Expected Mean Squared Error (EMSE). With this assumption, data points are statistically independent and residuals (ε) with mean of zero (E_ε) and a constant variance (VAR_ε). The expected prediction error ($E_{MSE(x)}$) for training sub-sample τ at x can be decomposed as follows (James et al., 2013, Kuhn and Johnson, 2013, Nordhausen, 2009):

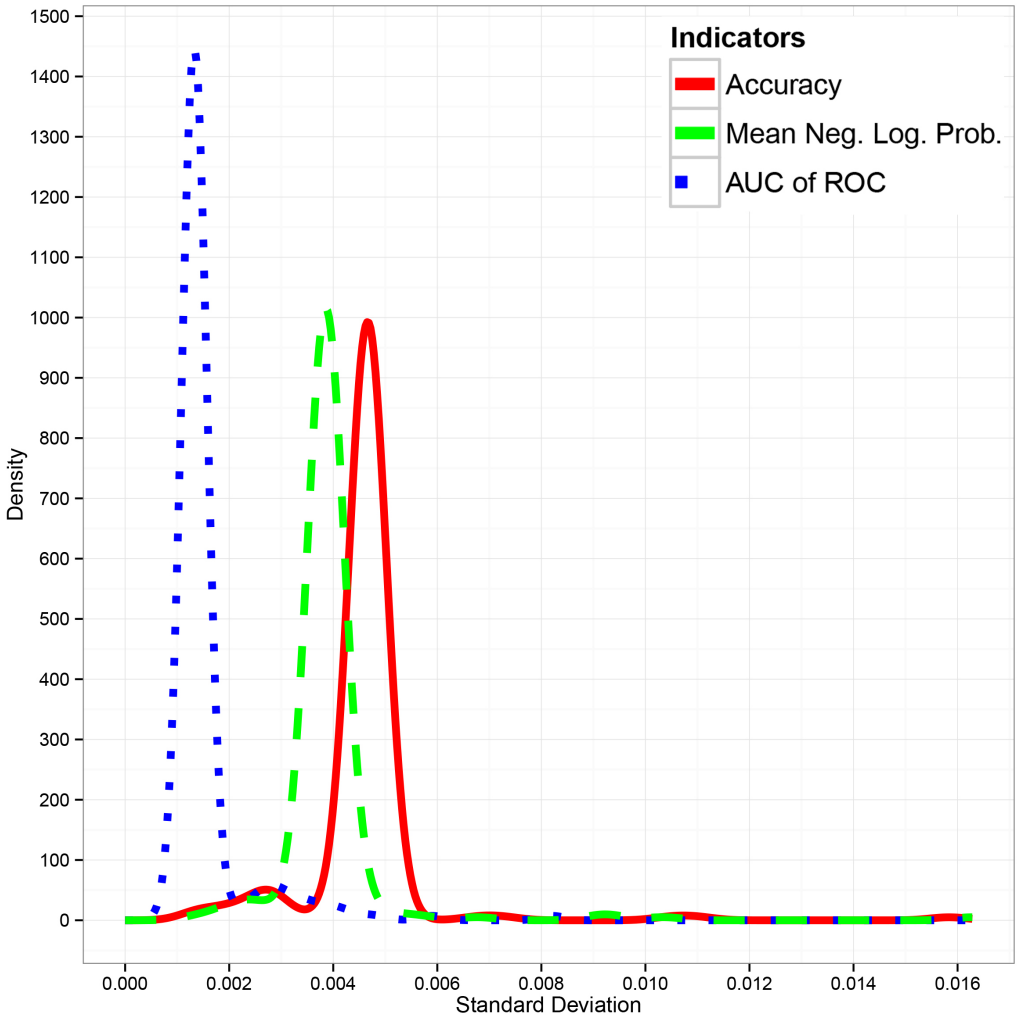

FIGURE 8.5: ERMER: Distributions of the standard deviation of the cross-validations' performance indicators (all samples)

$$E_{MSE(x)} = E_\tau[(Y - \hat{Y}_x)^2] = VAR_\tau(\hat{Y}_x) + BIAS^2_\tau(\hat{Y}_x) + \sigma^2_\varepsilon \tag{8.2a}$$

$$s.t.\ Y = f(x) + \varepsilon \tag{8.2b}$$

$$VAR_\tau(\hat{Y}_x) = E_\tau[(\hat{Y}_x - E_\tau(Y_x)^2] \tag{8.2c}$$

$$BIAS_\tau(\hat{Y}_x) = E_\tau[\hat{Y}_x] - f(x) \tag{8.2d}$$

$$\varepsilon \sim \{E_\varepsilon = 0 \ \wedge \ VAR_\varepsilon = \sigma_\varepsilon\} \tag{8.2e}$$

$$x \in \{x_1, x_2, ..., x_n\} \ : \ x_i \perp x_j \tag{8.2f}$$

, where the function $\hat{Y}_x$ approximates function Y_x, to be as close as possible to f_x. σ^2_ε is the initial variance of the *target* around mean, $VAR_\tau(\hat{Y}_x)$ is the variance of the estimated *target* about its mean, and $BIAS^2_\tau(\hat{Y}_x)$ is the amount that the average

estimate varies from true mean. The combination of the first two parts is also known as the reducible noise, and the last part (σ_ε^2) is known as the irreducible noise, because Y is a function of ε.

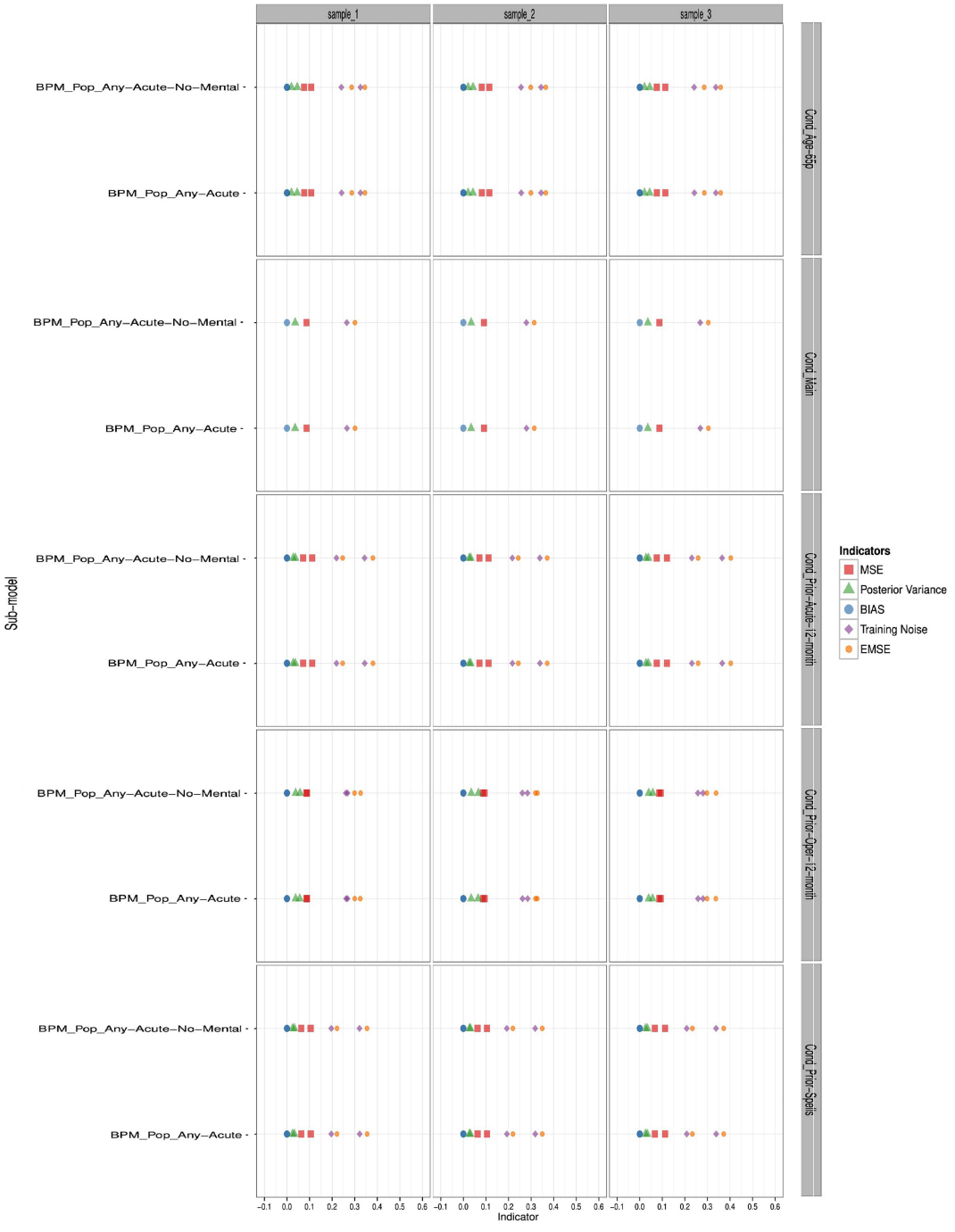

FIGURE 8.6: ERMER: Error, bias and variance plots for all sub-models (all samples)

Moreover, the MSE is a performance indicator, which sometimes is being used as part of a regularisation (penalisation) function to control complexity of fit or certain type of smoothness behaviours. The Root of MSE (RMSE), which is used as performance indicator in our Ensemble model generation, gives more weight to points further away.

The MSE algorithm can be broken down into bias-variance decomposition (Nordhausen, 2009):

$$MSE(x) = E_\tau[(Y - \hat{Y}_x)^2] \tag{8.3a}$$

$$MSE(x) = E_\tau[\hat{Y}_x - E_\tau(\hat{Y}_x)]^2 + (E_\tau[\hat{Y}_x] - Y)^2 \tag{8.3b}$$

$$MSE(X) = VAR_\tau(\hat{Y}_x) + BIAS^2_\tau(\hat{Y}_x) \tag{8.3c}$$

TABLE 8.5: ERMER: The performance statistics for different sub-populations and risk cut-offs (all samples)

Statistic	**Sub_PARR-2-Settings** [a]			**Sub_IPAEOPGP** [b]			**Sub_Any-Acute** [c]		
Threshold	**0.50**	**0.60**	**0.70**	**0.50**	**0.60**	**0.70**	**0.50**	**0.60**	**0.70**
Train: train sub-sample of *Sample-1*; Test: test sub-sample of *Sample-1*									
True & False Positive (TP+FP)	19,646	7,946	2,991	51,422	30,361	14,719	52,842	31,260	15,231
True Positive (TP)	11,962	5,512	2,291	36,966	24,051	12,432	37,979	24,759	12,878
Sensitivity (True Positive Rate)	0.390	0.180	0.075	0.478	0.311	0.161	0.461	0.300	0.156
Specificity (True Negative Rate)	0.805	0.938	0.982	0.887	0.950	0.982	0.900	0.956	0.984
Precision (Positive Predictive Value)	0.609	0.694	0.766	0.719	0.792	0.845	0.719	0.792	0.846
Emer. admi. post 12 m. per TP	1.242	1.600	2.105	1.581	1.857	2.146	1.586	1.863	2.158
Emer. admi. prior 12 m. per TP	0.462	0.607	0.740	0.351	0.365	0.368	0.351	0.364	0.367
Emer. admi. prior 13-24 m. per TP	0.401	0.532	0.646	0.319	0.336	0.327	0.318	0.335	0.326
Emer. admi. prior 25-36 m. per TP	0.006	0.007	0.009	0.004	0.004	0.005	0.004	0.004	0.005
AUC of ROC	0.661			0.767			0.771		
Total number of patients	70,147			204,672			231,755		
Train: train sub-sample of *Sample-2*; Test: test sub-sample of *Sample-2*									
True & False Positive (TP+FP)	25,972	11,121	4,212	61,229	34,292	15,745	62,910	35,230	16,177
True Positive (TP)	15,916	7,577	3,169	43,858	26,920	13,180	45,032	27,611	13,539
Sensitivity (True Positive Rate)	0.470	0.224	0.094	0.503	0.309	0.151	0.492	0.302	0.148
Specificity (True Negative Rate)	0.745	0.910	0.974	0.873	0.946	0.981	0.883	0.950	0.983
Precision (Positive Predictive Value)	0.613	0.681	0.752	0.716	0.785	0.837	0.716	0.784	0.837
Emer. admi. post 12 m. per TP	1.296	1.604	2.051	1.623	1.925	2.272	1.624	1.922	2.270
Emer. admi. prior 12 m. per TP	0.452	0.591	0.723	0.365	0.403	0.441	0.365	0.402	0.440
Emer. admi. prior 13-24 m. per TP	0.388	0.507	0.635	0.327	0.361	0.395	0.327	0.360	0.393
Emer. admi. prior 25-36 m. per TP	0.007	0.009	0.010	0.005	0.006	0.007	0.005	0.006	0.007
AUC of ROC	0.663			0.756			0.759		
Total number of patients	73,315			224,001			243,712		
Train: train sub-sample of *Sample-1*; Test: test sub-sample of *Sample-3*									
True & False Positive (TP+FP)	22,351	8,351	2,896	60,515	35,642	18,487	62,213	36,753	19,117
True Positive (TP)	14,003	5,942	2,337	44,730	28,783	16,114	45,950	29,654	16,678
Sensitivity (True Positive Rate)	0.340	0.144	0.057	0.438	0.282	0.158	0.421	0.272	0.153
Specificity (True Negative Rate)	0.834	0.952	0.989	0.905	0.959	0.986	0.917	0.964	0.988
Precision (Positive Predictive Value)	0.627	0.712	0.807	0.739	0.808	0.872	0.739	0.807	0.872
Emer. admi. post 12 m. per TP	1.311	1.730	2.361	1.646	1.913	2.163	1.655	1.926	2.186
Emer. admi. prior 12 m. per TP	0.522	0.684	0.805	0.364	0.348	0.304	0.364	0.347	0.304
Emer. admi. prior 13-24 m. per TP	0.435	0.565	0.667	0.322	0.306	0.260	0.321	0.305	0.259
Emer. admi. prior 25-36 m. per TP	0.005	0.006	0.010	0.004	0.004	0.004	0.004	0.004	0.004
AUC of ROC	0.658			0.767			0.771		
Total number of patients	91,369			268,575			304,888		

[a] Population setting for the PARR-2 model: age: 65+; Trigger admission: Emergency.
[b] Population setting for the Billings et al. (2013) model: Age: 18-95; Trigger admission: Emergency.
[c] All the population for the selected sample: Trigger admission: Emergency admission.

8.4 Benchmarks

Admission risk models are limited by the characteristics of the selected sub-population and data quality issues, such as missing diagnoses for outpatients and A&E patients (Billings et al., 2013), delayed death registration (ONS, 2014b), and the number of registered or consented patients. Also, models developed by researchers usually have different settings and assumptions; hence, comparisons of models in literature become more subjective.

The developed ERMER model is benchmarked against the CPM (DH, 2006, Paton et al., 2014), the PARR (Billings et al., 2006a) and Billings et al. (2013) models using their reported performance statistics.

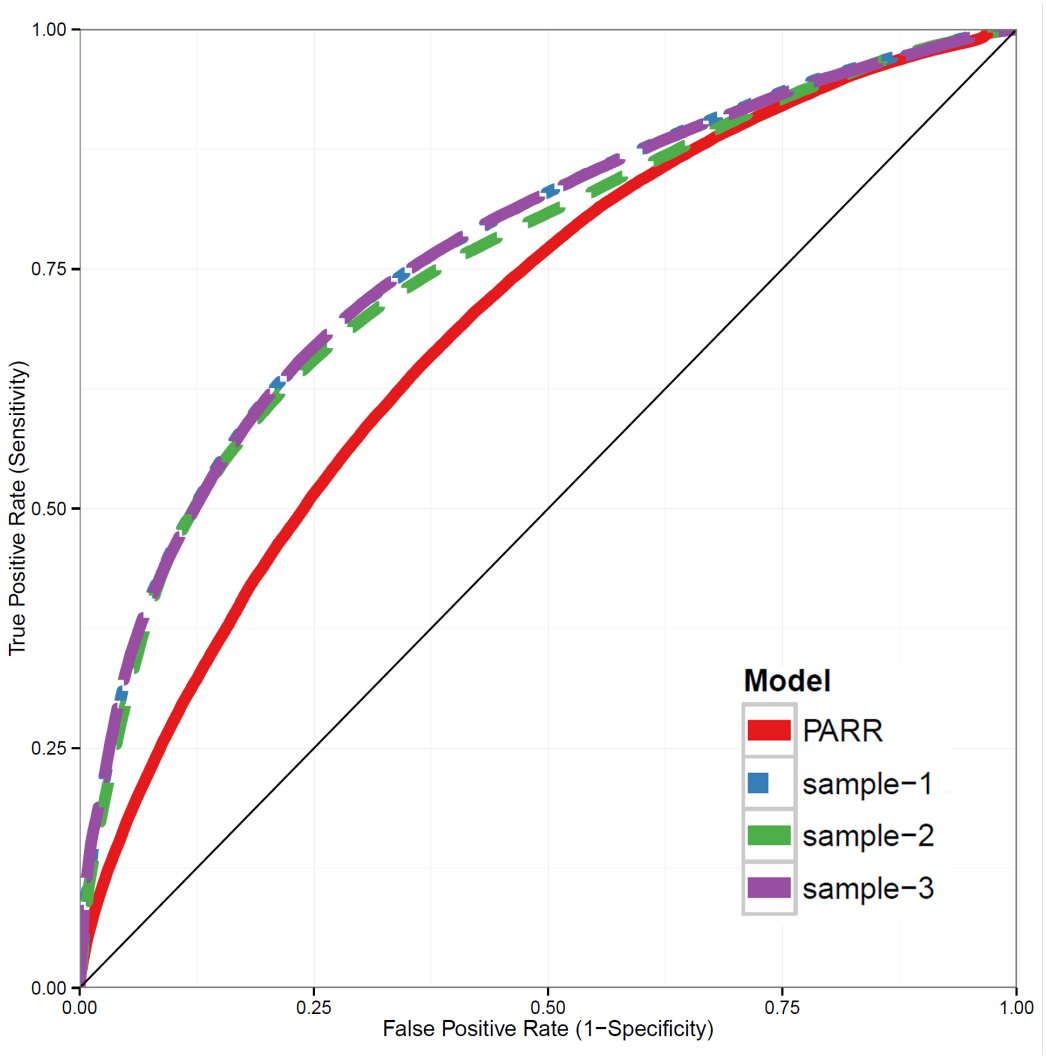

FIGURE 8.7: ERMER: ROC of model against the PARR model (all samples)

For the testing, validation and benchmarking phase, three data settings were considered: *Sample-1*'s train and test sub-samples, *Sample-2*'s train and test sub-samples, and finally a rolling window setting with *Sample-1*'s train sub-sample and *Sample-3*'s test sub-sample (Table 5.2). The rolling window is configured as a 1-year gap between time-frames, to better assess the stability of the model over time. In addition, three different sub-populations were selected from the outputted test results ($Sub\_PARR-2-Settings$, $Sub\_IPAEOPGP$ and $Sub\_Any-Acute$), for better comparison against the benchmarking models.

In comparisons, numerical summaries beyond the ROC and abstract statistical summaries must be used to avoid misinterpretation (Cook, 2007, Pencina et al., 2008,

Steyerberg et al., 2010). In addition to the ROC (Figure 8.7), the profiling is presented using three forms of presentations: the summary statistics for different risk cut-offs (Table 8.5) against the previous models (Table 8.6), the summary statistics for 20 risk bands (Table 8.7) and the profile of top risk segments (Table 8.8).

TABLE 8.6: The reported performance statistics for the benchmarking models (PARR, CPM and CPM update)

Statistic	**PARR**			**CPM**	**Billings-13 (IP)**	**Billings-13 (IPAEOPGP)**[a]
Threshold[b]	**0.50**	**0.60**	**0.70**	**0.50**	**0.50**	**0.50**
True & False Positive (TP+FP)	17,455	4,810	2,011	NR	8,743	10,545
True Positive (TP)	NR[c]	NR	NR	NR	4,627	5,669
Sensitivity (True Positive Rate)	0.543	0.178	0.081	NR	0.049	0.060
Specificity (True Negative Rate)	0.722	0.950	0.986	NR	NR	NR
Precision (Positive Predictive Value)	0.653	0.774	0.843	0.538	0.529	0.538
Emer. admi. post 12 m. per TP[d]	1.47	2.23	3.0	NR	NR	NR
Emer. admi. prior 12 m. per TP	2.22	3.43	4.59	NR	NR	NR
Emer. admi. prior 13-24 m. per TP	0.93	1.84	2.80	NR	NR	NR
Emer. admi. prior 25-36 m. per TP	0.73	1.48	2.25	NR	NR	NR
AUC of ROC	0.69			0.780	0.73	0.78
Total number of patients	42,778			281,617	1,836,099	1,836,099

[a] The Billings et. al. (2013) model with inpatient (IP), A&E (AE), outpatient (OP) and GP data. [b] The threshold on the predicted risk. [c] Not reported (NR). [d] Average number of emergency readmission of the truly positively predicted patients.

The ERMER model made a considerable improvement to the previous models. For instance, according to Table 8.5, the ERMER model with sub-population $Sub\_Any-Acute$ has precision 0.719 and AUC of 0.771 with *Sample-1* as the test set, compared to 0.529, 0.73 for the Billings et al. (2013) model with inpatient (IP) data.

8.5 Discussions

In this study, a set of significant features was initially developed using a framework. Then, several predictive models were trained based on different sub-populations. The defined sub-models were fitted using a Bayes Point Machine (BPM) algorithm, with Gamma priors and the Expectation Propagation (EP) message passing for the inference of the posterior. Furthermore, an optimised Ensemble of five sub-models was produced based on age group sub-models, 1-year prior operation sub-models and the general model.

Thereafter, the developed decision support tool, the Ensemble Risk Model of Emergency Admissions (ERMER), was benchmarked against the PARR, the CPM and Billings et al. (2013) models, with very similar settings. The proposed model outperforms other models for any-emergency readmissions and the sub-population of 18-95 year-old patients. The ROC of any-emergency readmission is 0.759-0.771, compared with the PARR, which is 0.69 with an age restriction (65+) and an HRG restriction (*reference* conditions). In addition, the performance is very close to the CPM and Billings et al. (2013) models, which predicts any-emergency admissions using much

more care data (inpatient, outpatient, A&E and GP data).

8.5.1 Data

Firstly, the feature preparation is the most time-consuming part of many analyses. There are three main layers of difficulties in the preparation of features: correlations, re-categorisations and selections (Mihaylova et al., 2011, Walpole et al., 2014, Yang et al., 2005a). In this study, the variables were generated and selected based on the healthcare pre-processing framework developed in the *Phase-I* of our research (Chapter 7). Based on this framework a large pool of variables was generated and filtered based on a set of defined criteria. Then, these features were ranked and a selection of top features were inputted into the model.

Capturing high-risk patients using diagnoses can be difficult owing to variate coding practices, under-reporting of diagnostic variables, incomplete coding of transferred patients and comorbidities' complexities (Billings et al., 2013, Bottle et al., 2011, Reimer et al., 2016). Therefore, only high-level of diagnoses groups were included, and the remaining detailed codes were aggregated.

In this study, a recent version of the HSCIC Charlson Comorbidity Index (HSCIC-CCI) was used, which is actively maintained by the HSCIC and Dr Foster unit (Aylin et al., 2010, Bottle et al., 2011). Comorbidity scoring is usually used to distinguish the conditions present on admission from complications. But, poor coding and disregarding the effects of population characteristics can introduce bias (constant risk fallacy) (Fischer et al., 2011, Nicholl, 2007). Other criticisms of scoring originate from choosing small cohorts, using additive risk models of different medical conditions, ignoring important factors, such as the Length-of-Stay (LoS) and the presence of different valid principal diagnoses across different cohorts (Bottle and Aylin, 2011, Quan et al., 2005).

Moreover, left-censored and right-censored observations introduce bias in the features and predicted risk estimates (Singer and Willett, 2003). According to Table 5.1, about 8%-15% of patients do not have any admissions after the *trigger-event*. In addition, about 28%-51% of patients do not have any other *prior-admissions* before the *trigger-event*.

Finally, it has been speculated that many of the variations in readmission can be due to the delivery of the care method, which cannot be quantified using an administrative database only (Billings et al., 2013, Bottle et al., 2014, DH, 2006).

TABLE 8.7: ERMER: the risk bands statistics (all samples)

	PARR-2-Settings[a]						IPAEOPGP[b]						Any-Acute[c]					
Band	**TP+FP**	**TP**	**Preci.**	**Sens.**	**Avg.**[d]	**C.I.**[e]	**TP+FP**	**TP**	**Preci.**	**Sens.**	**Avg.**	**C.I.**	**TP+FP**	**TP**	**Preci.**	**Sens.**	**Avg.**	**C.I.**
Train: train sub-sample of *Sample-1*; Test: test sub-sample of *Sample-1*																		
1	14	0	0.000	0.000	0.00	0.00, 0.00	2,101	140	0.067	1.000	6.66	0.06, 0.08	3,797	240	0.063	1.000	6.30	0.06, 0.07
2	103	5	0.049	1.000	4.85	0.00, 0.10	8,065	945	0.117	0.871	11.68	0.11, 0.12	12,435	1,411	0.113	0.855	11.36	0.11, 0.12
3	522	48	0.092	0.906	9.19	0.07, 0.12	14,916	2,046	0.137	0.653	13.70	0.13, 0.14	20,067	2,675	0.133	0.618	13.33	0.13, 0.14
4	2,329	369	0.158	0.874	15.8	0.14, 0.17	15,054	2,643	0.176	0.458	17.55	0.17, 0.18	20,237	3,457	0.171	0.444	17.08	0.17, 0.18
5	3,404	742	0.218	0.637	21.79	0.20, 0.23	20,850	3,979	0.191	0.408	19.09	0.19, 0.20	24,368	4,613	0.189	0.372	18.92	0.18, 0.19
6	6,356	1,832	0.288	0.611	28.80	0.28, 0.30	20,969	4,585	0.219	0.320	21.87	0.21, 0.22	23,313	5,075	0.218	0.290	21.77	0.21, 0.22
7	7,681	2,618	0.341	0.466	34.09	0.33, 0.35	21,445	5,593	0.261	0.281	26.09	0.25, 0.27	23,063	5,968	0.259	0.255	25.87	0.25, 0.26
8	9,604	3,705	0.386	0.398	38.57	0.38, 0.40	18,623	6,271	0.337	0.239	33.64	0.33, 0.34	19,461	6,523	0.335	0.218	33.49	0.33, 0.34
9	11,501	5,080	0.442	0.353	44.18	0.43, 0.45	17,265	7,216	0.418	0.216	41.77	0.41, 0.43	17,827	7,425	0.417	0.199	41.66	0.41, 0.42
10	8,987	4,310	0.480	0.230	47.95	0.47, 0.49	13,962	6,896	0.494	0.171	49.38	0.49, 0.50	14,345	7,068	0.493	0.159	49.27	0.48, 0.50
11	6,913	3,713	0.537	0.166	53.66	0.53, 0.55	10,921	6,160	0.564	0.133	56.38	0.55, 0.57	11,191	6,313	0.564	0.124	56.38	0.55, 0.57
12	4,787	2,737	0.572	0.109	57.21	0.56, 0.59	10,140	6,755	0.666	0.127	66.57	0.66, 0.67	10,391	6,907	0.665	0.120	66.50	0.66, 0.67
13	3,076	1,948	0.633	0.072	63.32	0.62, 0.65	10,109	7,426	0.735	0.122	73.43	0.73, 0.74	10,357	7,585	0.732	0.116	73.21	0.72, 0.74
14	1,879	1,273	0.677	0.045	67.80	0.66, 0.70	5,533	4,193	0.758	0.065	75.81	0.75, 0.77	5,672	4,296	0.757	0.062	75.74	0.75, 0.77
15	1,116	800	0.717	0.027	71.68	0.69, 0.74	4,301	3,423	0.796	0.050	79.58	0.78, 0.81	4,424	3,517	0.795	0.048	79.49	0.78, 0.81
16	721	547	0.759	0.018	75.86	0.73, 0.79	2,975	2,447	0.823	0.035	82.31	0.81, 0.84	3,089	2,549	0.825	0.034	82.55	0.81, 0.84
17	460	364	0.791	0.012	79.13	0.75, 0.83	4,595	4,076	0.887	0.054	88.72	0.88, 0.90	4,757	4,223	0.888	0.053	88.77	0.88, 0.90
18	306	240	0.784	0.008	78.43	0.74, 0.83	1,697	1,475	0.869	0.019	86.91	0.85, 0.89	1,769	1,542	0.872	0.019	87.22	0.86, 0.89
19	199	167	0.839	0.005	83.92	0.79, 0.89	597	509	0.853	0.007	85.26	0.82, 0.88	619	527	0.851	0.006	85.13	0.82, 0.88
20	189	173	0.915	0.006	91.53	0.87, 0.95	554	502	0.906	0.006	90.61	0.88, 0.93	573	520	0.908	0.006	90.75	0.88, 0.93
N	70,147	30,671	0.609	0.390	43.72	0.43, 0.44	204,672	77,280	0.719	0.478	37.75	0.38, 0.38	231,755	82,434	0.719	0.461	35.56	0.35, 0.36

[a] The performance of the model for the sub-population *Sub_PARR-2-Settings*. [b] The performance of the model for the sub-population *Sub_IPAEOPGP*.
[c] The performance of the model for the sub-population *Sub_Any-Acute*. [d] The average of number of readmitted patients. [e] The confidence interval for the average of number of readmitted patients using the bootstrapped central estimate with 95% CI.

TABLE 8.8: ERMER: The top risk segments profile of the predicted patient (all samples)

Risk Seg.[a]	Model	Sub-population	Min Risk[b]	Asthma [c]	COPD [d]	Depres. [e]	Diab. [f]	Hyper. [g]	Cancer [h]	CHD [i]	CHF [j]	Avg. Age[k]	Avg. LoS[l]	5-9 Meds[m]	10+ Meds[n]
Train: train sub-sample of *Sample-1*; Test: test sub-sample of *Sample-1*															
		PARR-2-Settings	0.576	16.69	35.68	41.94	23.49	53.20	19.65	50.93	39.82	80.80	11.06	NA	NA
10,000		IPAEOPGP	0.759	11.54	12.19	12.24	8.25	20.71	6.51	14.88	10.34	39.68	4.49	NA	NA
		Any-Acute	0.766	11.25	11.52	11.57	7.93	19.74	6.24	14.01	9.72	38.61	4.39	NA	NA
		PARR-2-Settings	0.647	20.84	44.14	45.10	26.00	56.28	21.24	57.00	45.10	80.33	11.32	NA	NA
5,000		IPAEOPGP	0.817	15.80	15.78	15.28	10.14	25.20	7.52	18.58	12.96	42.36	4.91	NA	NA
		Any-Acute	0.818	15.84	15.72	15.32	10.38	25.12	7.60	18.48	12.92	41.99	4.93	NA	NA
		PARR-2-Settings	0.815	31.40	59.10	50.70	26.90	61.70	22.90	66.40	53.30	78.95	10.04	NA	NA
1,000	ERMER	IPAEOPGP	0.910	33.40	35.70	30.50	21.90	39.70	14.30	38.60	26.30	53.38	6.98	NA	NA
		Any-Acute	0.912	33.20	34.80	29.90	21.80	39.00	14.40	37.50	25.40	52.21	6.85	NA	NA
		PARR-2-Settings	0.881	37.40	67.60	52.00	26.40	63.20	25.20	69.60	55.20	77.98	9.35	NA	NA
500		IPAEOPGP	0.957	38.20	38.60	34.80	25.20	42.80	14.20	43.20	27.40	54.49	7.43	NA	NA
		Any-Acute	0.958	37.80	37.60	33.80	25.00	41.20	14.00	41.80	26.80	52.95	7.37	NA	NA
		PARR-2-Settings	0.933	36.40	70.00	53.60	27.60	63.20	25.60	69.60	53.60	77.34	9.57	NA	NA
250		IPAEOPGP	0.985	40.80	39.20	36.40	27.20	42.80	11.60	42.00	29.60	53.88	7.99	NA	NA
		Any-Acute	0.986	40.40	38.80	36.40	27.60	42.40	11.60	41.20	28.80	52.76	7.89	NA	NA

[a] The top predicted risk segment. [b] The minimum predicted risk in the segment.
[c] The percentage of patients with a history of Asthma diagnosis (ICD-10: J45-J46). [d] The percentage of patients with a history of Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-10: J20, J41-J44, J47). [e] The percentage of patients with a history of Depression diagnosis (ICD-10: I10-I15).
[f] The percentage of patients with a history of Diabetes diagnosis (ICD-10: E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9, E10.2-E10.5, E10.7, E11.2-E11.5, E11.7, E12.2-E12.5, E12.7, E13.2-E13.5, E13.7, E14.2-E14.5, E14.7). [g] The percentage of patients with a history of Hypertension diagnosis (ICD-10: I10-I15, I27, I6, I87.0, I87, I97, K76.6, H35.0, R03, O13, O14, O16, O10, G93.2, H40.0, P292, P293). [h] The percentage of patients with a history of Cancer diagnosis (ICD-10: C00-D49). [i] The percentage of patients with a history of Coronary Heart Disease (CHD) diagnosis (ICD-10: I20-I25). [j] The percentage of patients with a history of Congestive Heart Failure (CHF) diagnosis (ICD-10: I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0). [k] The average age of patients at the trigger event. [l] The average length of stay of patient at the trigger event. [m] The percentage of patients with 5-9 medication prescription. [n] The percentage of patients with 10+ medication prescription.

8.5.2 Model

Scepticism against using advanced machine learning techniques in healthcare modelling has been repeatedly highlighted in literature, due its hype, bad practices and lack of transparency (Section 1.5.1).

In general, accuracy and efficiency of a Bayesian model depend on five main design choices: the representation of features, fitness algorithm, inference approximation, assignment and update of prior probabilities and the framework of system states.

Firstly, the features were carefully generated, selected and ranked before generating the models. The initial prototype models, without the feature selection strategies mentioned earlier, have shown very high sensitivity to intercorrelations, sparsity and noisy features. As a result, these caused non-convergence, weight decay and performance degradation.

Moreover, in comparison with the SVM, the BPM method is demonstrated (Herbrich et al., 2001) that can provide a better solution for an asymmetric version space, greater ability to handle large datasets and may produce a smoother decision boundary more efficiently.

Furthermore, Microsoft's version of the BPM algorithm (Research, 2016) uses EP message passing, which in Gaussian Mixture problems is demonstrated (Minka, 2001a,b) to be better than approximation techniques, such as the Markov Chain Monte Carlo (MCMC), Laplace and Variational Bayes techniques. The EP does not guarantee convergence, but in many cases it does. Especially if the features are not highly interdependent, to become trapped in a region of local optima.

Finally, the choice of prior probability distributions of the weights and features can have a significant impact on the robustness of the algorithm. The applied algorithm uses a heavy-tailed prior, which is more robust towards outliers of the weight distributions. Also, the incremental Bayesian training of the ERMER allows it to incorporate the effects of changes in prior distributions.

8.5.3 Results

All the sub-models are stable in the convergence and the cross-validation tests. Moreover, the features are initially selected based on the main model's population. And, the weights are very similar, proportionally, for all sub-models owing to very similar feature

distributions, except for two: the sub-model with no prior *spell* ($Cond\_Prior\text{-}Spells_0$) and the sub-model with no prior operation or procedure ($Cond\_Prior\text{-}Oper\text{-}12\text{-}month_0$).

Firstly, the learning curve plots for the BPM sub-models are presented in Figure 8.2. The learning curve plots for each modelling group was generated for *Sample*-1 with fixed test sub-sample size. Each plot demonstrates the micro-average error of the sub-models against different training sub-sample sizes. In all the modelling groups and sub-models, the micro-average error reduces by 0.02 to 0.04 for training sub-sample size from 500 up to $10,000$, and then the error changes less than < 0.005 for training sub-sample size from $10,000$ up to $200,000$. Therefore, this is an indication that training sub-sample sizes $> 10,000$ do not provide a huge benefit to the performance of our models.

Furthermore, the applied BPM algorithm can handle a large number of features and a moderately large number of observations in comparison to the LR. On average, it takes about two to eight minutes[1] to train a sub-model with 100 features.

Moreover, for this modelling approach, 5 random folds were used, and 100 iterations were performed for training (similar to the model fitting stage). In Figure 8.5, the distribution of standard deviation of cross-validation's performance indicators are presented.

Finally, a way to find the generalisation error can be to test models with fresh and independent samples from the same source (Shalizi, 2015). The dataset splitting method was carried out at first, and different sub-samples were used for testing purpose. The detailed results are provided in Appendix A.6.

For the BPM models, the training needs to reach a uniform consistency that can bound the probability of error. The EP algorithm that Infer.Net used can not guarantee convergence, and if there is a conflicting solution, it may lead to non-convergence (Minka, 2001b, 2016, Research, 2016). Average and range of convergence for all sub-models, and distributions of cross-validation performance indicators are presented in Figure 8.4 and Figure 8.5.

Also, the models' performances are consistently high across all the samples. The performance of the main sub-models improves the ROC (Figure 8.7), sensitivity, specificity and precision percentage by 2.83, 0.50, 1.26 and 2.83, respectively (Table 8.5).

Furthermore, the populations of readmitted patients are very low; therefore, the samples are significantly unbalanced in terms of the dependent variable distribution. The main models have 3.0-4.5 times less readmitted patients, and sub-models have 1-10

[1] Windows 10 machine with Intel i7 2 GHz quad-core CPU and 8 GB 1600 MHz RAM.

times less readmitted patients compared to non-readmissions. Therefore, based on the sensitivity, precision and the ROC, our models can more confidently identify low-risk patients, and avoid unnecessary interventions.

In addition, the Ensemble model improves the overall performance compared to the individual sub-models. The ROC and precision statistics of the any-acute model increase by 2.83 and 7.16, respectively, and sensitivity decreases in consequence. The detailed performance indicators of sub-models are presented in Appendix A.6.3.

Moreover, the features were selected based on the main model, which considers all the emergency admission population. Therefore, the PARR sub-population underperforms. However, compared with the PARR model, the predicted high-risk patients have less number of *prior-admissions* for all the sub-populations, which makes it considerably harder to predict.

In addition, based on the population profile of the top 1000 risk segments (Table 8.8), the model (*Any-Acute*) predicts more patients with chronic obstructive pulmonary disease (COPD), depression, diabetes, coronary heart disease (CHD), congestive heart failure (CHF) and smaller average-age as high-risk, than the CPM and the PARR models did. On the other hand, cancer that is highly predictable and manageable has a smaller share among the high-risk patients.

Finally, because sensitivity and precision vary across risk scores, and the costs of interventions or readmissions are not zero, it is better to define a profit function. However, owing to the lack of necessary variables for mapping the costs, this was not considered.

8.6 Concluding Remarks

What is already known?

- Avoidable emergency hospital admission can be an indicator of suboptimal care quality.
- Identification of high-risk patients for intervention can substantially improve care quality and reduce costs.
- Designing features and developing predictive models that can adjust continuously to a fast-changing healthcare system and population characteristics are very challenging.

What this phase of research adds?

- The optimised Ensemble model of sub-population was proved to significantly improve the risk model.
- The combination of using a nonlinear Bayesian model and applying a healthcare pre-processing framework to generate a pool of features and to select significant features can effectively create a highly adaptable predictive model.
- The Ensemble of generative models is a new effective way to predicts patients with harder predictability, such as patients with chronic conditions and patients with fewer prior hospitalisation records.

In conclusion, ERMER provides a generic approach in modelling readmission emphasising on robustness and feature discovery. Moreover, based on a large number of iterations for performance assessment across different settings, the ERMER maintained its high discriminatory performance. Consequently, ERMER can bring a significant improvement to the current decision support system in use, increase care quality and reduce the costs.

Finally, the ERMER toolkit is presented in Chapter 10. It is a generic, user-friendly and open-source software package, and is released for public use and incremental development.

In the next chapter, the final phase of the research is presented, which is dedicated to the development of a generic temporal comorbidity risk model.

Chapter 9

Phase III: Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER)

Patients' comorbidities, operations and complications can be associated with reduced long-term survival probability and increased healthcare utilisation. The aim of this research phase was to produce an adjusted case-mix model of comorbidity risk and develop a user-friendly software tool to encourage public adaptation and incremental development.

It has been shown in healthcare research, that demographics, temporal dimensions, Length-of-Stay (LoS) and time between admissions, can noticeably improve the statistical measures related to comorbidities.

In previous literature, there have been two streams of work on risk scoring comorbidities to estimate future resource utilisation, emergency admission and mortality. Firstly, one stream of research looks at the odds ratio of major diagnoses groups, like Charlson Comorbidity Index (CCI) which rely on twenty-two comorbidity groups (Charlson et al., 1987). The second stream uses a case-mix model or a diagnoses classification approach based on similarities, type of care, likelihood or duration, like Elixhauser Comorbidity Index (ECI) (Elixhauser et al., 1998, AHRQ, 2016b), Diagnosis-related Groups (DRGs) (Fetter et al., 1980, Mistichelli, 1984) and John Hopkin's (Weiner and Abrams, 2011) Adjusted Clinical Groups (ACGs).

Furthermore, comorbidity risk models are constrained by the population and sample characteristics, data quality (e.g. missing diagnoses or delayed death registration) and modelling approaches. There is a wide range of literature that focuses on modification

and benchmarking comorbidity risk indices, using different datasets, cohorts, diagnoses groups, complexity types, LoS and claims. The models' prediction targets varies and include: 1-year in-hospital or general mortality, and 7- and 30-day emergency admissions (Austin et al., 2012, Bottle et al., 2014, Gagne et al., 2011, Holman et al., 2005, Januel et al., 2011, Mehta et al., 2016, Sharabiani et al., 2012). Moreover, there have been many attempts at scoring surgical outcomes and complications (Section 2.3.2), which are affected by comorbidity (Armitage and Van der Meulen, 2010, Mehta et al., 2016), but they lack generality and high dependency to extra clinical data.

There has been very little research in the area of temporal comorbidity risk scores (Wang et al., 2009a), and the majority of temporal models (Appendix A.1.1) in the literature focus on survival analysis aspect of comorbidity indices. Unlike temporal models of mortality, temporal emergency admissions risk scores are more difficult to model, optimise and implement, due to complex characteristics of comorbidities and care utilisation over time and their relations to emergency admissions.

The proposed model in this phase, incorporates temporal aspects, operations and procedures groups, demographics, and admission details, as well as diagnoses groups. The research resulted in the development of Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER) model using routinely collected hospital inpatient data. The T-CARER model is published publicly as an interactive IPython Notebook, with generic inputs, features and population settings for general purpose use.

Moreover, several stages of analysis have been carried out to test and benchmark the T-CARER. Firstly, two data-frames across a 10-year period (1999-2010) were selected. Afterwards, three different modelling approaches were developed: a Logistic Regression (LR), a Random Forest (RF), and a Wide and Deep Neural Network (WDNN). Then, the models were benchmarked against the HSCIC implementation of the CCI (HSCIC-CCI) (Aylin et al., 2010, DFI, 2013), and the reported performance of CCI and ECI implementations (Bottle and Aylin, 2011, Bottle et al., 2014, Holman et al., 2005, Mehta et al., 2016).

The WDNN and the RF methods outperform in terms of the Area Under the Curve (AUC) of Receiver-Operating Characteristic (ROC) against the LR, as well as HSCIC-CCI, CCI and ECI models. For 30- and 365-day emergency admissions, ROCs of different modelling approaches were from 0.772%-0.804% for the two sampled time-frames.

The WDNN method produced predictions with high precision, and the RF method outperformed regarding micro-average of F1-score. The precisions were 0.582%-0.639%,

and the micro-average of F1-score were 0.730%-0.790% for the best modelling methods across different sampled time-frames.

This chapter is structured as follows. Firstly, the data and pre-processing stages are summarised. Furthermore, the T-CARER and modelling methods are described. Afterwards, the results are presented, and then the benchmarking comparisons are provided. Finally, concluding remarks are presented. Moreover, the description of the T-CARER development toolkit is provided separately in Chapter 10.

9.1 Data

After the data ingestion step (Section 7.1), four stages of data pre-processing were carried out (Section 7.2), including removals, imputations and re-categorisation[1] of some discrete and continuous variables to reduce sparsity and better capture non-linear relationships (Table 5.1).

Based on previous studies and the initial statistical analyses, four levels of temporal features were generated: 0-30, 30-90, 90-365 and 365-730 days. These four levels capture part of the temporal aspect of comorbidities, in addition to the delta-time between admissions (*GapDays*) and the length-of-stay (*epidur*) features that include temporal metadata. Furthermore, in the modelling stage, we applied several techniques to capture the complex temporal patterns of patients comorbidities.

For the diagnoses, a clinical grouper, known as the Clinical Classifications Software (CCS), was used to better capture comorbidities' patterns. The CCS categorises the ICD-10 diagnoses and operations into a number of categories that are clinically meaningful (Elixhauser and Steiner, 2006, AHRQ, 2016a). In this study, operations and procedures are categorised using the main categories of the OPCS-4[2].

After exploratory analyses of the variables, three major related risk factors were defined and re-categorised: demographics, admission and clinical (Table 9.1). Next, a feature pool was generated similar to the feature engineering step of our pre-processing framework Section 7.3. Also, the temporal features were summarised in each temporal level based on several aggregation functions, including prevalence, count and average. This step increased the number of features by more than fifty folds.

[1] Re-categorisation is also known as recoding, grouping, classifying or aggregating

[2] The OPCS has hierarchical coding structure and the first character in the code represents the main category of operation or procedure.

TABLE 9.1: T-CARER: The main defined features

Main Feature	Definition
gender	Gender of patient (Female, Male, Others)
ethnos	Ethnicity of patient (Bangladeshi, Black African, Black Caribbean, Black Other, Chinese, Indian, Pakistani, White, Others).
imd04rk	The Index of Multiple Deprivation (IMD) overall ranking of income (22.5%), employment (22.5%), health deprivation and disability (13.5%), education & skills (13.5%), barriers to housing & services (9.3%), crime (9.3%), & living environment (9.3%).
ageTrigger	Age of patient at the trigger event. Categorisation bins: {10-, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90+}.
gapDays (temporal)	Delta-times from discharge to the trigger admission.
epidur (temporal)	Spell durations.
preopdur (temporal)	Pre-operative durations.
posopdur (temporal)	Post-operative durations.
operOPCSL1 (temporal)	The level-1 categories (25 groups) of operating procedure codes (OPCS), the national standard (HSCIC, 2014b) version 4.0 (¿4,000 codes).
diagCCS (temporal)	The level-1 categories (302 groups) of Clinical Classifications Software (CCS) for ICD-10 (AHRQ, 2016b) diagnoses (69,800 codes).
admimeth (temporal)	The level-1 categories (3 groups) of admission method (20 codes): {Elective, Emergency, Others}.
mainspef (temporal)	The level-1 categories (33 groups) of the main specialities of the consultants (86 codes), based on the exploratory analysis.

Then, the feature selection step has been carried out (Section 7.4). The recoding states of the features were filtered out based on their linear cross-correlation, as well as frequency and sparseness (percentage of distinct and the ratio of the most common value to the second most common). Afterwards, features were sorted using the average importance score, which was produced using the Breiman RF method after six trials and three different decision tree generation settings (Breiman, 2001).

9.2 Model

The aim of this research phase is to model emergency readmission using a minimal number of generic features that can be used for short and long-term predictions and have a high correlation to comorbidity risk. In this phase of research, emergency readmission refers to the emergency admission in future without any condition on the *trigger-event*, unlike the ERMER model, which enforced the emergency admission condition for both the *trigger-event* and the *future-admission*.

Based on our literature reviews, three training methods have been considered for the T-CARER, in order to closely access the algorithms strengths and weaknesses[3].

The first algorithm is a LR with *L1* regularisation (value of 1.0), using *liblinear* optimisation algorithm (Fan et al., 2008) with a maximum of hundred iterations and a warm-up period (scikit learn, 2016). The LR method is a linear regression model and is a special case of Generalised Linear Model (GLM), which assumes the model error has a standard logistic distribution. Moreover, the addition of *L1* regularisation to LR

[3]CPU: Intel i7-7700K 4.2 GHz; GPU: NVIDIA Titan X 1.5GHz, 12GB RAM; Memory: Samsung SM951 512GB, PCI-E v3 on Intel Z270 chipset; RAM: 4x16GB Corsair DDR4 2666 MHz C15 XMP 2.0; Libraries: TensorFlow (1.0), Cuda (8.0), SciKit-Learn (0.18) and SciPy (0.18).

TABLE 9.2: T-CARER: Top groups of diagnoses, operations and consultant specialities

CCS Categories of the ICD-10 Diagnoses:	Spondylosis; disc disorders; other back problems
Abdominal pain	Thyroid disorders
Administrative/social admission	Urinary tract infections
Alcohol-related disorders	**Level-1 Categories of OPCS-4 for Operations:**
Allergic reactions	Arteries & Veins
Asthma	Bones & Joints of Skull & Spine
Cardiac dysrhythmias	Diagnostics & Tests
Cataract	Female Genital Tract
Chronic obstructive pulmonary & bronchiectasis	Heart
Complication of device; implant or graft	Lower Digestive Tract
Congestive heart failure; nonhypertensive	Lower Female Genital Tract
Coronary atherosclerosis & other heart disease	Male Genital Organs
Deficiency & other anemia	Mental Health
Delirium dementia & amnestic & other cognitives	Miscellaneous Operations
Diabetes mellitus without complication	Nervous System
Disorders of lipid metabolism	Other Abdominal Organs
Esophageal disorders	Other Bones & Joints
Essential hypertension	Others
External cause codes	Respiratory Tract
Fetal distress & abnormal forces of labor	Skin
Fracture of upper limb	Soft Tissue
Genitourinary symptoms & ill-defined conditions	Upper Digestive Tract
Normal pregnancy and/or delivery	Upper Female Genital Tract
OB-related trauma to perineum & vulva	Urinary
Osteoarthritis	**Bespoke Categories of Consultant Specialities:**
Other & unspecified benign neoplasm	A&E
Other aftercare	Cardiothoracic
Other birth complications; mother's puerperium	Ear, nose & throat
Other complications of pregnancy	Gastroenterology
Other connective tissue disease	General
Other female genital disorders	General Surgery
Other gastrointestinal disorders	Geriatric
Other injuries & conditions due to external causes	Gynaecology
Other lower respiratory disease	Haemotology
Other nervous system disorders	Maternity
Other suspected screening (excl. mental & infectious)	Ophthalmology
Other skin disorders	Others
Other upper respiratory disease	Paediatrics
Others	Plastic
Phlebitis; thrombophlebitis & thromboembolism	Psychiatry
Residual codes; unclassified	Urology
Skin & subcutaneous tissue infections	

allows the model to select a simpler model when there are moderate number of features with high sparsity (Section 4.2.3).

Secondly, we used a RF method using the Breiman (2001) algorithm with Gradient Boosted Regression trees, Gini index, fifty trees in the forest with balanced *labels*, the minimum split size of hundred, and minimum leaf size of fifty (scikit learn, 2016). The RF method (Section 4.2.4.1) is an Ensemble Decision Tree, which was initially introduced by Breiman (2001), and is based on the CART algorithm and the Bagging Ensemble method. However, the Breiman RF is sensitive to highly correlated features and the scale or categories of features (Strobl et al., 2007, Tolo̧si and Lengauer, 2011).

Thirdly, a Deep Neural Network (DNN) was designed based on the Wide and Deep

Neural Network (WDNN) algorithm, that was introduced by Cheng et al. (2016). DNNs are a class of Artificial Neural Networks (ANNs) with multiple hidden layers (Section 4.2.7.2), which allow modelling more complex non-linear problems with more effective representation of features in each layer. The WDNN is a DNN which combines benefits of memorisation and generalisation. The WDNN consists of two parts: the wide model and the deep model (Figure 4.4).

The wide part of the model consists of a wide linear model for highly sparse features (random features that are rarely active). The wide part may also include groups of crossed features (a.k.a. interaction terms). Inside of a group of crossed features, each level of one feature occurs in combination with each level of other features. On the other hand, the deep part of the model composed of hidden layers of the feed-forward artificial neural network (ANN) and may also include an embedding layer to convert categorical inputs into low-dimensional and dense real-valued vectors (Abadi et al., 2016).

TABLE 9.3: T-CARER: The input layer of the Wide and Deep Neural Network

Sub-Model	Feature Type	Features
Wide	Categorical	*ageTrigger* (17 states), *epidur* (6 states), *ethnos* (11 states), *gapDays* (6 states), *gender* (2 states), & *imd04rk* (11 states)
	Crossed (memorised)	*gender* ≈ *ethnos* (80 cross states), *imd04rk* ≈ *gender* (200 cross states), *imd04rk* ≈ *ethnos* (400 cross states), & *imd04rk* ≈ *ageTrigger* (400 cross states).
Deep	Embedded	*ageTrigger* (5 states), *ethnos* (3 states), *gender* (2 states), *imd04rk* (5 states), *epidur* (3 states), & *gapDays*. (3 states)
	Continuous	All the selected categories of *admimeth*, *diagCC*, *gapDays*, *mainspef*, *operOPCSL1*, *posopdur*, & *preopdur*.

In our study, the WDNN model applies Adadelta optimiser (Duchi et al., 2011) for the gradients of the deep part, and the Rectified Linear Unit (ReLU) activation function was applied to each layer of the ANN (LeCun et al., 2015). Moreover, the WDNN model was developed after several stages of ad-hoc benchmarking to reach an optimal setting for hyper-parameters. The first aim was to minimise the loss function in learning iterations, to avoid weight decay and to assure convergence. The second objective was to maximise the layers and neurons under computing resources constraints, to increase stability and minimise resource utilisation. Also, an implicit optimisation was carrying out in the background by the Adadelta optimiser to configure the learning rate dynamically.

Furthermore, because of the huge size of the WDNN, the designed *tensors* (Google, 2016) was trained in batches of 2,000 observations per-step, for 40,000 iterations. The training of each model using our hardware and software setups took about 12 hours (with regular storage of the trained model). The outline of the nodes are presented in Table 9.3, and an abstract representation of the designed model with the TensorFlow is presented in Appendix A.7.2. The wide part of the model consists of twenty-two

categorical features (1-17 states) and four crossed columns (80-400 states). The deep part of the model includes fourteen embedding features (3-5 states), 286 features continuous features (one-dimensional) and three hidden layers of neurons. The defined hidden layers one to three were fully interconnected and were configured as 24,000, 12,000 and 6,000 nodes, respectively.

TABLE 9.4: T-CARER: The performance statistics (all samples)

Time Horizon	**30-day**						**365-day**					
Method	**RFC**[a]		**LR**[b]		**WDNN**[c]		**RFC**		**LR**		**WDNN**	
Sample	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Sample: Sample-1 (Train, Test)												
ROC	0.827	0.804	0.778	0.772	0.817	0.796	0.789	0.780	0.760	0.759	0.795	0.787
Precision	0.180	0.180	0.530	0.520	0.641	0.617	0.430	0.430	0.690	0.690	0.644	0.631
Sensitivity [d]	0.760	0.730	0.070	0.070	0.104	0.098	0.710	0.700	0.260	0.270	0.382	0.374
F1 [e]	0.300	0.280	0.130	0.130	0.178	0.170	0.540	0.530	0.380	0.380	0.480	0.470
Macro F1	0.790	0.790	0.900	0.900	0.900	0.900	0.740	0.730	0.770	0.770	0.790	0.790
Accuracy	0.728	0.724	0.926	0.925	0.928	0.928	0.718	0.713	0.802	0.802	0.808	0.805
Log-Loss	9.392	9.538	2.571	2.576	2.476	2.496	9.746	9.914	6.840	6.835	6.636	6.748
Brier-Score	0.168	0.171	0.061	0.061	0.059	0.060	0.186	0.188	0.144	0.144	0.137	0.139
TP+FP [f]	43,494	43,466	43,494	43,466	43,494	43,466	134,101	133,901	134,101	133,901	134,101	133,901
Total	578,936	578,937	578,936	578,937	578,936	578,937	578,936	578,937	578,936	578,937	578,936	578,937
Sample: Sample-2 (Train, Test)												
ROC	0.766	0.743	0.718	0.715	0.759	0.735	0.791	0.785	0.765	0.766	0.793	0.772
Precision	0.340	0.320	0.580	0.570	0.600	0.582	0.610	0.610	0.690	0.690	0.651	0.639
Sensitivity	0.590	0.550	0.110	0.120	0.207	0.198	0.690	0.690	0.460	0.460	0.585	0.573
F1	0.430	0.400	0.190	0.190	0.308	0.295	0.650	0.650	0.550	0.550	0.616	0.604
Macro F1	0.790	0.780	0.810	0.810	0.830	0.830	0.720	0.720	0.700	0.700	0.720	0.720
Accuracy	0.770	0.756	0.857	0.855	0.862	0.859	0.722	0.717	0.720	0.720	0.728	0.719
Log-Loss	7.955	8.416	4.931	5.011	4.738	4.878	9.616	9.775	9.672	9.668	9.397	9.719
Brier-Score	0.194	0.197	0.112	0.114	0.107	0.110	0.187	0.190	0.185	0.185	0.176	0.184
TP+FP	47,487	48,207	47,487	48,207	47,487	48,207	120,285	120,838	120,285	120,838	120,285	120,838
Total	322,300	322,301	322,300	322,301	322,300	322,301	322,300	322,301	322,300	322,301	322,300	322,301

[a] Random forest classification (RFC). [b] Logistic regression (LR). [c] Wide and deep neural network (WDNN).
[d] Recall or true positive rate. [e] F1-score (F1). [f] True and false positives.

9.3 Results and Benchmarks

Firstly, the three T-CARER models: LR, RF and WDNN, have been benchmarked across samples and the two prediction *target* variables: 30- and 365-day emergency admissions (Table 9.4). In overall, the WDNN and the RF models provide a better fit for the 30- and 365-day emergency readmission problems. For the 365-day, the WDNN produces a marginally better Receiver-Operating Characteristic (ROC) compared to the RF, and significantly better than the LR (Figure 9.1). Also, the WDNN models have very strong precision (positive predictive value), accuracy and micro-average F1-score. On the other hand, the RF models have very high sensitivity (True Positive Rate) and F1-score (Figure 9.2).

TABLE 9.5: T-CARER: The profile of main comorbidity groups for the 365-day model using Random Forest method (all samples)

Main Comorbidity Groups	Population Profile				T-CARER Profile							HSCIC-CCI Profile				Comparisons	
Diagnoses Group [a]	Prior [b]	Male	Age [c]	LoS [d]	Total	Sens. (0.5) [e]	F1 (.5) [f]	TP (.5) [g]	TN (.5) [h]	TP (.7)	TN (.7)	CCI 1-3 [i]	CCI 1-3 (TP)	CCI 4+	CCI 4+ (TP)	Delta Score (.5, 4+) [j]	Delta Score (.7, 4+)
Sample: Sample-1 (Test)																	
Hypertension (HT)	29,207	12,311	22	9	89004	0.296	0.6	27,954	11,017	7,833	51,964	7,079	2,380	23,022	8,962	-8,962 (-10.1%)	10,926 (12.3%)
Depression	21,635	9,925	16	8	69154	0.311	0.611	22,864	7,143	5,849	41,670	6,356	2,130	18,168	7,166	-7,166 (-10.4%)	9,379 (13.6%)
CHD [k]	20,849	11,669	16	8	57550	0.222	0.504	13,540	7,616	6,146	30,555	3,871	1,547	19,360	7,776	-7,776 (-13.5%)	7,762 (13.5%)
Cancer	20,475	9,888	24	7	80579	0.479	0.768	42,240	4,699	3,921	56,183	2,054	708	14,643	5,589	-5,589 (-6.9%)	6,479 (8.0%)
Asthma	10,196	3,576	43	6	32718	0.356	0.67	12,405	4,272	2,489	20,033	559	267	18,073	5,124	-5,124 (-15.7%)	10,752 (32.9%)
Diabetes	11,249	5,799	20	8	31673	0.259	0.56	8,730	4,067	2,941	17,483	14,307	4,545	8,014	3,400	-3,400 (-10.7%)	11,435 (36.1%)
COPD [l]	9,144	4,729	14	9	19892	0.119	0.351	2,482	4,621	3,017	7,731	542	303	10,935	4,839	-4,839 (-24.3%)	3,318 (16.7%)
CHF [m]	9,248	4,466	15	10	20838	0.097	0.29	2,098	4,468	3,457	8,133	1,083	559	11,385	4,937	-4,937 (-23.7%)	3,515 (16.9%)
Prior 30-day non-emergency	781	310	43	9	2203	0.291	0.61	676	374	238	1,184	71	35	272	112	-112 (-5.1%)	-42 (-1.9%)
Prior 30-day emergency	112,570	39,530	46	6	360657	0.331	0.64	126,308	44,469	23,167	224,920	11,348	4,223	48,512	18,579	-18,579 (-5.2%)	13,891 (3.9%)
Sample: Sample-2 (Test)																	
Hypertension (HT)	40,163	16,555	23	7	85422	0.232	0.574	21,913	11,178	3,845	41,414	7,908	3,425	30,744	16,271	-16,271 (-19.0%)	15,111 (17.7%)
Depression	32,312	14,583	17	8	69956	0.224	0.554	17,409	7,711	3,270	34,374	7,481	3,230	27,170	14,424	-14,424 (-20.6%)	13,727 (19.6%)
CHD	21,714	11,662	18	7	42427	0.171	0.481	8,037	5,819	2,322	18,391	3,372	1,725	20,758	11,267	-11,267 (-26.6%)	8,816 (20.8%)
Cancer	15,732	6,965	25	7	33143	0.227	0.567	8,408	4,198	1,601	15,810	1,602	757	12,386	6,783	-6,783 (-20.5%)	4,847 (14.6%)
Asthma	14,124	4,562	46	6	31962	0.312	0.682	11,086	4,956	1,400	16,438	805	424	18,387	7,715	-7,715 (-24.1%)	9,653 (30.2%)
Diabetes	13,006	6,482	21	8	27138	0.221	0.56	6,649	3,489	1,357	12,775	11,730	4,834	10,368	5,794	-5,794 (-21.3%)	10,113 (37.3%)
COPD	10,717	5,439	16	7	18912	0.123	0.41	2,507	4,126	1,481	6,714	722	436	12,365	7,007	-7,007 (-37.0%)	4,163 (22.0%)
CHF	9,686	4,700	14	9	16361	0.095	0.348	1,670	3,371	1,343	5,332	1,004	598	10,722	6,470	-6,470 (-39.5%)	3,315 (20.3%)
Prior 30-day non-emergency	755	283	41	6	1394	0.201	0.567	309	311	76	563	59	35	343	200	-200 (-14.3%)	91 (6.5%)
Prior 30-day emergency	120,838	39,590	48	5	322301	0.411	0.764	147,815	39,168	9,021	192,442	12,321	5,052	55,456	26,535	-26,535 (-8.2%)	27,169 (8.4%)

[a] The Charlson Comorbidity Index (CCI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age. [d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1 score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point. [h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+. [k] Coronary heart disease (CHD). [l] Chronic obstructive pulmonary disease (COPD). [m] Congestive heart failure (CHF).

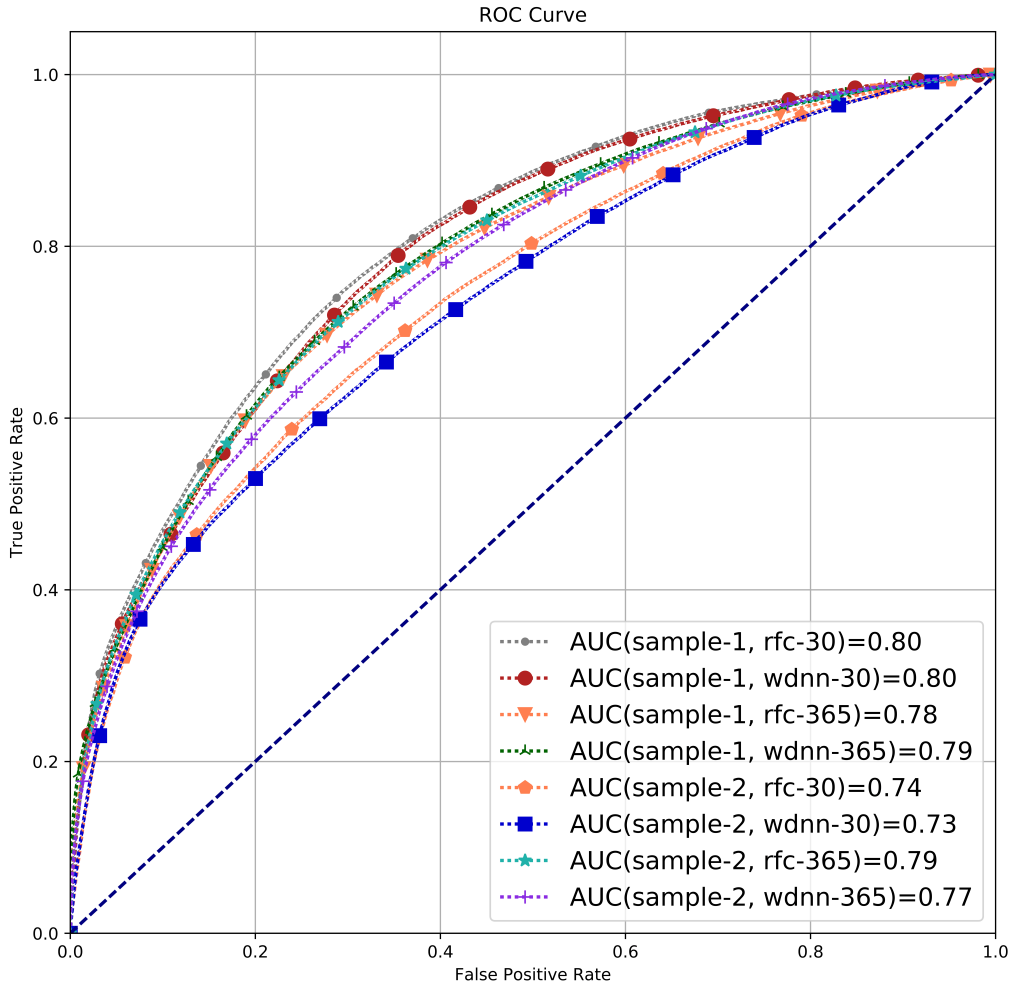

FIGURE 9.1: T-CARER: ROC plot of RF and WDNN Models (all samples)

Moreover, based on the published comparison tests of CCIs and ECIs in literature, T-CARER performs considerably better for 30-day emergency admission. For the emergency admission within a year, no previous benchmarking studies of CCIs or ECIs calculated this, due to constraints on data collection, poor prediction power of models or different research priorities.

Three previous studies have been selected that include benchmarks of various versions of CCIs and ECIs for the emergency admission problem. Firstly, Mehta et al. (2016) reported ROC of 0.70-0.766 for CCIs and ECIs using the Texas Medicare data (2006-2011). Furthermore, Bottle and Aylin (2011), Bottle et al. (2014) benchmarked CCIs using the England's HES data (2007-2009) and produced ROC of 0.57-0.79. Moreover,

TABLE 9.6: T-CARER: The profile of main comorbidity groups for the 365-day model using Wide and Deep Neural Network method (all samples)

Main Comorbidity Groups	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Sample: Sample-1 (Test)																	
Hypertension (HT)	29,207	12,311	22	9	89004	0.109	0.47	11,613	51,212	4,630	58,409	7,079	2,380	23,022	8,962	-8,962 (-10.1%)	17,371 (19.5%)
Depression	21,635	9,925	16	8	69154	0.089	0.419	7,426	41,098	2,004	46,547	6,356	2,130	18,168	7,166	-7,166 (-10.4%)	14,256 (20.6%)
CHD[k]	20,849	11,669	16	8	57550	0.117	0.455	8,238	29,601	2,454	35,477	3,871	1,547	19,360	7,776	-7,776 (-13.5%)	12,684 (22.0%)
Cancer	20,475	9,888	24	7	80579	0.051	0.332	4,949	55,706	1,564	59,401	2,054	708	14,643	5,589	-5,589 (-6.9%)	9,697 (12.0%)
Asthma	10,196	3,576	43	6	32718	0.116	0.51	4,447	19,742	2,112	21,886	559	267	18,073	5,124	-5,124 (-15.7%)	12,605 (38.5%)
Diabetes	11,249	5,799	20	8	31673	0.12	0.468	4,585	16,663	1,433	19,766	14,307	4,545	8,014	3,400	-3,400 (-10.7%)	13,718 (43.3%)
COPD[l]	9,144	4,729	14	9	19892	0.218	0.577	5,197	7,071	1,912	9,895	542	303	10,935	4,839	-4,839 (-24.3%)	5,482 (27.6%)
CHF[m]	9,248	4,466	15	10	20838	0.183	0.531	4,656	7,961	1,474	10,849	1,083	559	11,385	4,937	-4,937 (-23.7%)	6,231 (29.9%)
Prior 30-day non-emergency	781	310	43	9	2203	0.146	0.532	380	1,154	183	1,357	71	35	272	112	-112 (-5.1%)	131 (5.9%)
Prior 30-day emergency	112,570	39,530	46	6	360657	0.117	0.519	49,548	219,157	29,322	241,351	11,348	4,223	48,512	18,579	-18,579 (-5.2%)	30,322 (8.4%)
Sample: Sample-2 (Test)																	
Hypertension (HT)	40,163	16,555	23	7	85422	0.235	0.598	23,876	29,461	9,893	41,592	7,908	3,425	30,744	16,271	-16,271 (-19.0%)	15,289 (17.9%)
Depression	32,312	14,583	17	8	69956	0.22	0.575	18,473	24,165	6,952	34,521	7,481	3,230	27,170	14,424	-14,424 (-20.6%)	13,874 (19.8%)
CHD	21,714	11,662	18	7	42427	0.257	0.601	13,125	11,881	5,359	18,430	3,372	1,725	20,758	11,267	-11,267 (-26.6%)	8,855 (20.9%)
Cancer	15,732	6,965	25	7	33143	0.236	0.592	9,345	10,891	3,648	15,894	1,602	757	12,386	6,783	-6,783 (-20.5%)	4,931 (14.9%)
Asthma	14,124	4,562	46	6	31962	0.245	0.634	9,062	12,444	4,266	16,608	805	424	18,387	7,715	-7,715 (-24.1%)	9,823 (30.7%)
Diabetes	13,006	6,482	21	8	27138	0.24	0.598	7,772	8,931	3,181	12,846	11,730	4,834	10,368	5,794	-5,794 (-21.3%)	10,184 (37.5%)
COPD	10,717	5,439	16	7	18912	0.338	0.67	7,481	4,077	3,520	6,927	722	436	12,365	7,007	-7,007 (-37.0%)	4,376 (23.1%)
CHF	9,686	4,700	14	9	16361	0.345	0.672	6,675	3,165	3,107	5,428	1,004	598	10,722	6,470	-6,470 (-39.5%)	3,411 (20.8%)
Prior 30-day non-emergency	755	283	41	6	1394	0.336	0.688	540	365	273	564	59	35	343	200	-200 (-14.3%)	92 (6.6%)
Prior 30-day emergency	120,838	39,590	48	5	322301	0.185	0.604	69,254	162,356	34,938	193,319	12,321	5,052	55,456	26,535	-26,535 (-8.2%)	28,046 (8.7%)

[a] The Charlson Comorbidity Index (CCI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age. [d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1 score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point. [h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+. [k] Coronary heart disease (CHD). [l] Chronic obstructive pulmonary disease (COPD). [m] Congestive heart failure (CHF).

Holman et al. (2005) reported ROC of 0.61-0.77 for CCIs, ECIs and Multipurpose Australian Comorbidity Scoring System (MACSS) models, based on data from hospitals in Western Australia (1989-1996).

Furthermore, T-CARER models are compared against our implementation of the HSCIC-CCI across all the CCI and the ECI diagnoses categories. Appendix A.7.3 indicates that the T-CARER performs significantly better than the HSCIC-CCI for all the 46 diagnoses categories (The 2009-10 versions) regarding the True Positive Rate (TPR). The WDNN models with the cut-off of 0.70 outperform against the HSCIC-CCI score of greater than four. Also, the RF model, with the cut-off of 0.50, can beat the accuracy of HSCIC-CCI score of greater than zero for the majority of diagnoses.

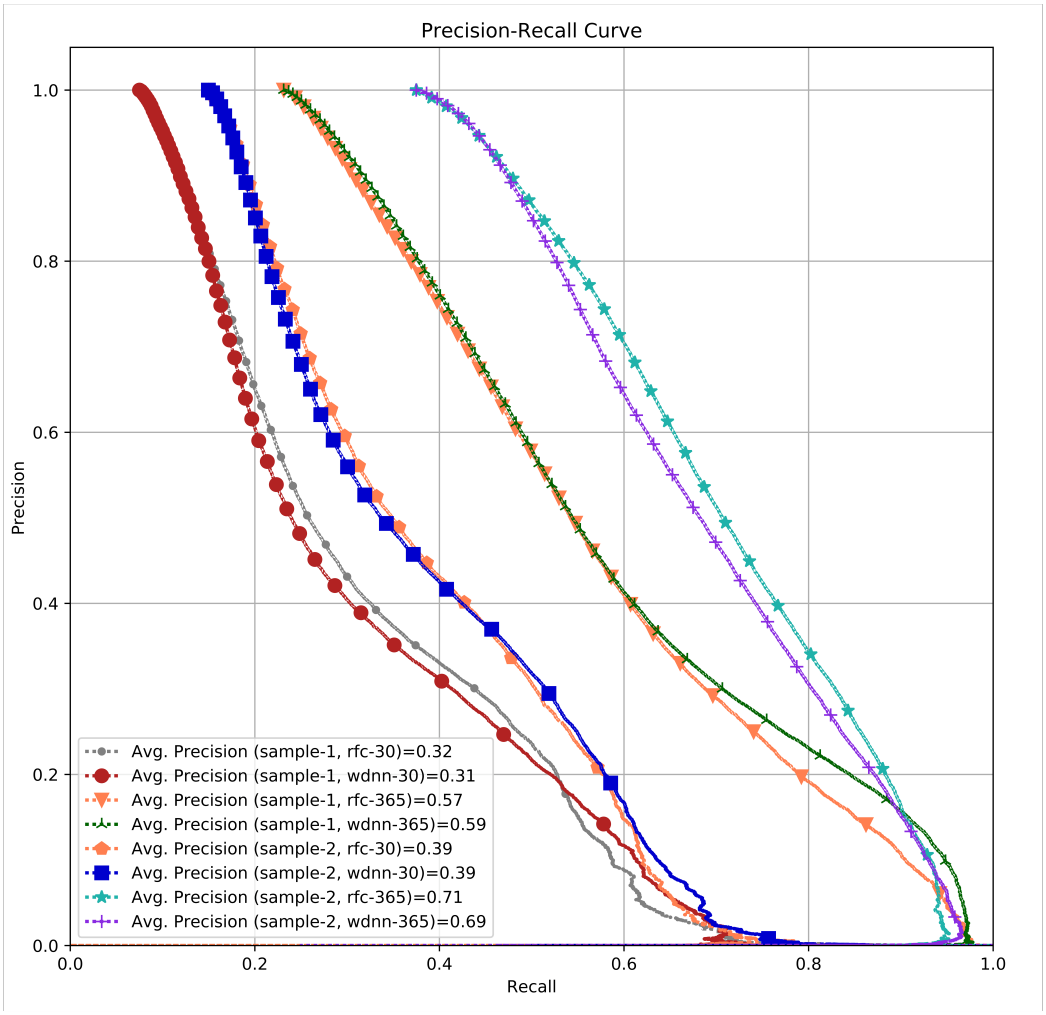

FIGURE 9.2: T-CARER: precision-recall curves of RF and WDNN Models (all samples)

Also, the performance of the emergency admission models using only HSCIC-CCI are very poor, therefore are not presented here. For instance, constructed LR and RF 365-day emergency admission models using only the HSCIC-CCI, have ROC of 0.53-0.58

across the samples.

Moreover, the two top T-CARER models are compared against the HSCIC-CCI based on eight main comorbidity groups: hypertension, depression, Coronary Heart Disease (CHD), asthma, diabetes, cancer, Chronic Obstructive Pulmonary Disease (COPD) and Congestive Heart Failure (CHF). Table 9.5 and Table 9.6 demonstrate that for all the main comorbidity categories, the T-CARER models outperform against the HSCIC-CCI.

Finally, a *10-fold* cross-validation (Murphy, 2012) algorithm was run for the LR and the RF using the two test sub-samples (Table 5.2). The cross-validation tests were stable and consistent, with negative Mean Square Error (MSE) variance of -0.7-2.9. The applied *K-fold* cross-validation splits each sample into ten equal-sized random samples. Then, *K - 1* folds are used for training and 1-fold is used for validation. Finally, the *K-fold* cycles through all combinations to generate performance outputs.

9.4 Discussions

We compared the performance of the T-CARER against commonly used comorbidity index models using different samples and population cohorts across a ten year period. Our analyses of the T-CARER and the HSCIC-CCI for different diagnoses categories demonstrated that our model performed best in the majority of comorbidity groups, and in overall T-CARER models show better results against previous surveys of CCIs and ECIs.

Furthermore, the progression of patients comorbidities over time and patterns of care utilisation can have great impacts on the performance of comorbidity models, and it is important that modelling algorithms are equipped to capture temporal changes and interactions of correlated factors. The T-CARER's performances for prediction of 30- and 365-day emergency readmissions indicate that it can overtake conventional risk scoring methods with more flexibility on the features and customisations. Also, our study shows that boosting algorithms, like RF, and deep learning models, like the WDNN, can learn better multiple levels of comorbidities complexities.

In the best-case scenario, a comorbidity score can perform only as well as the included diagnoses categories and their correlated factors (Austin et al., 2015). The deployment of the healthcare pre-processing framework that was proposed in *Phase-I* (Chapter 7), helped to systematically perform the data pre-processing and feature engineering of

the comorbidity risk scoring. Furthermore, the CCS allowed to categorise the ICD-10 diagnoses into a manageable number of clinically meaningful categories. The applied CCS *clinical grouper* made it simpler to understand patterns of diagnoses and easily add a wider range of comorbidity groups (Elixhauser and Steiner, 2006, 2013, AHRQ, 2016a).

Benchmarking comorbidity scores can be very useful as it offers more insight into strength and weaknesses of models. Our benchmarking demonstrates that the RF modelling method may lead to a low level of positive predictive value, but high sensitivity. On the other hand, the designed deep learning model (the WDNN) can produce models with high precision and weak sensitivity. In overall, the micro-average of F1-score for the WDNN models is greater across samples and prediction targets, but with high training cost. However, the implemented LR models can only train estimators that have weaker overall performance and higher bias.

In summary, the designed temporal case-mix risk models outperform against major previous models, with superior precision, F1-score and ROC. The developed risk index can help in monitoring temporal comorbidities of patients, and potentially reduce down the cost of inappropriate hospital and A&E admissions.

9.5 Concluding Remarks

What is already known?

- Identification of avoidable emergency admission remains a troublesome problem for healthcare.
- Providing intervention before and after discharge for high-risk patients can significantly improve care quality and reduce costs.
- Development of a risk index calculator that can adjust very precisely patients demographics and temporal conditions is very challenging, due to the complexity of comorbidities and patients health status over time.
- The majority of in-use comorbidity risk indices use population odds-ratio to weight diagnoses groups, and then apply a crude sum of present conditions to calculate the comorbidity risk index.

What this phase of research adds?

- The temporal model of comorbidities, operations and complexities was proved to notably improve the comorbidity risk model.
- Adjustment for demographics and admission type was significantly influential.
- The DNN method and the RF method with Boosting were provided very good fits, in comparison with LR.
- Inputting a pool of features, including comorbidity groups, operations and complexities, into the feature selection can lead to the discovery of new important risk factors.

This study has sought to identify an approach to score commodities by the inclusion of diverse categories of diagnoses, operations, and complexities. The T-CARER performs consistently across tests and validations, and it outperforms against Charlson and Elixhauser indices which are widely used for prediction of comorbidity risks.

In the following chapter, the produced development toolkits for the ERMER and the T-CARER are presented, which can be applied to many different healthcare settings. They are generic, user-friendly and open-source software packages, and are released for public use and incremental development.

Chapter 10

Development Toolkits

Jupyter is an interactive computing interface, and it supports interactive data visualisation. Jupyter provides a browser-based language shell (a.k.a Read-Eval-Print Loop) and has many Jupyter-compatible kernel languages, like Python, R, Julia, Matlab/Octave, SAS, JavaScript and Scala (IPython, 2016, Jupyter, 2016).

Furthermore, Jupyter provides an ideal lab notebook for saving computational workflow. The IPython Notebook, that is Jupyter Notebook with Python Kernel, is increasingly being used by scientists who need to keep a detailed record of analytical processes.

In this project, the developed models can be controlled and configured through an IPython Notebook. IPython Notebooks (Figure 10.2) were produced to implement the workflows of all the analyses steps of the Ensemble Risk Model of Emergency Admissions (ERMER) and the Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER).

The process-flow diagram of the whole data ingestion, feature generation and modelling processes are presented in Figure 10.1. The process-flow diagram includes six main logical steps: the data management, the data preparation, the feature engineering, the feature selection, the model development and the deployment. Firstly, the first four steps are included in the healthcare pre-processing framework (Chapter 7). And, the developed IPython Notebooks apply the feature selection and the model development steps, which need more configurations and interactive design. The data preparation and feature generation steps will be released in near future as a separate package, with generic design for inclusion of multiple healthcare data sources. Finally, the deployment step may be applied by engineers after developing a modelling solution using the development toolkits.

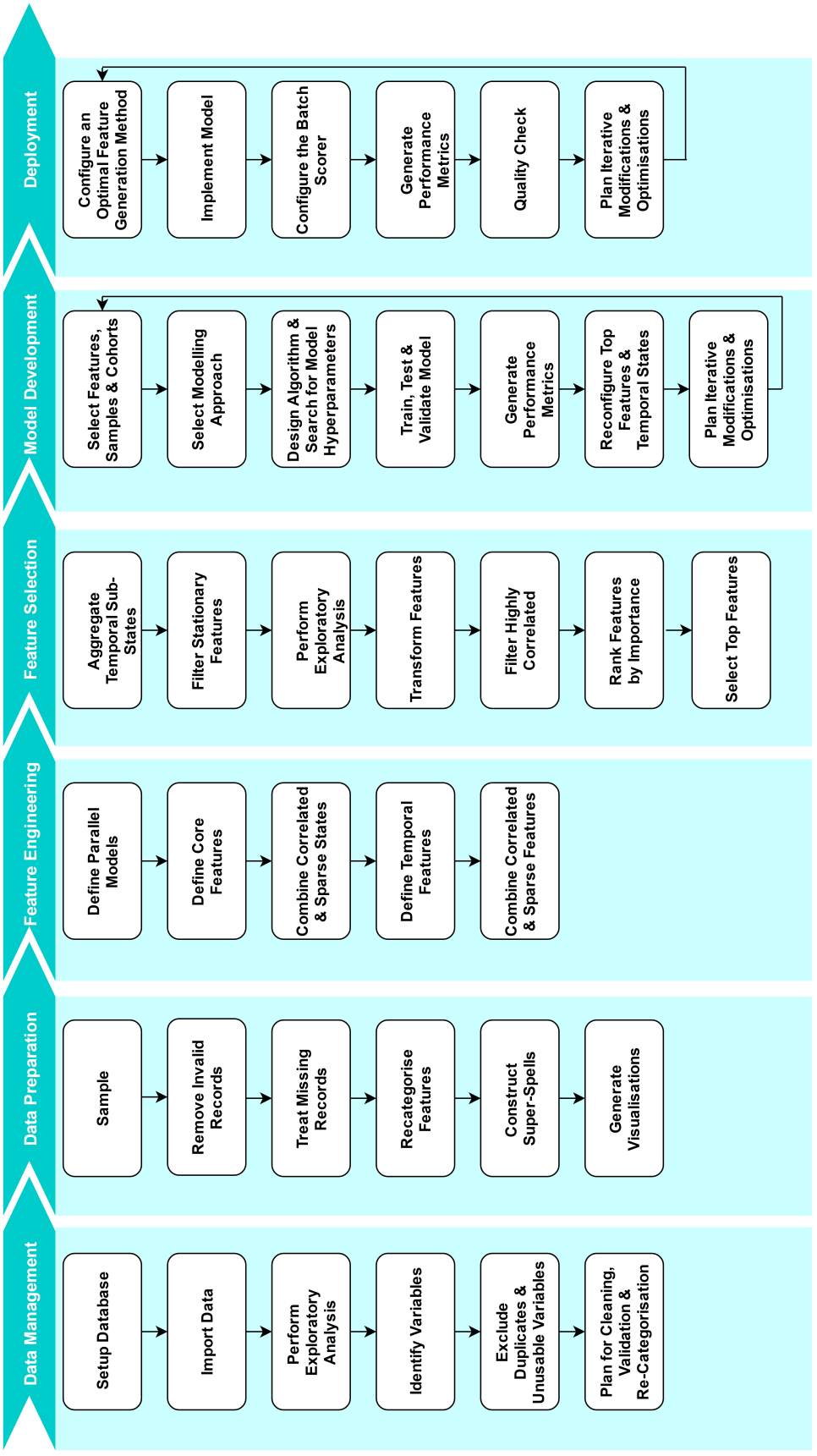

FIGURE 10.1: Process-flow diagram of the development phases

The defined workflow script in the IPython Notebooks is calling third-party libraries and the developed sub-packages to handle input and outputs, pre-process features, train and test models. The Notebooks call procedures that are from different language environments, including Python, MySQL, Bash and C#.

Moreover, a set of generic Python sub-package were developed to facilitate the reading and writing interface, statistical functions, feature parsing, and configuration modules. The Unified Modelling Language (UML) representation of the sub-packages and their classes are presented in Appendix A.8.1.

FIGURE 10.2: Toolkits: Screenshot of a developed Jupyter Notebook

Moreover, basic modelling approaches, like Logistic Regression (LR) and Random Forest (RF), are implemented in these developed Python sub-packages. But, the applied advanced modelling techniques, the Bayes Point Machine (BPM) and the Wide and Deep Neural Network (WDNN), have been integrated independently from these sub-packages and Notebooks.

Furthermore, the abstract structures of the main modelling Notebooks for the ERMER and the T-CARER have very similar specifications and they are presented in Figure 10.3. The defined steps in Notebooks are pre-configured and can be re-configured down to the every finest detail by the user.

In the following sections, firstly, the objectives of the toolkits are stated. Then, the software and hardware prerequisites and deployments of the ERMER and the T-CARER toolkits are discussed. Afterwards, the feature generation process workflow is discussed, which will be added to the public release. Next, the toolkits inputs and outputs types are described. After that, the pre-processing and the modelling steps are explained.

Finally, the toolkits developments and applications are discussed.

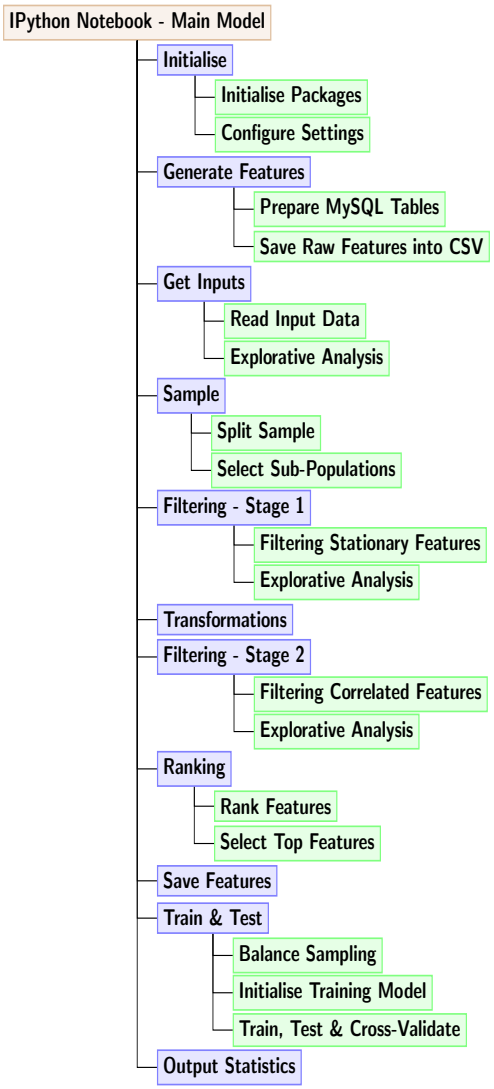

FIGURE 10.3: Toolkits: Abstract structure of the main Jupyter IPython Notebook

10.1 Objectives

The developments of the ERMER and the T-CARER are intended to move toward the key objectives of adaptation of open tools, transparency and reproducibility in healthcare modelling. The developed solutions provide generic, user-friendly and transparent tools to model the risk of hospital readmission and the comorbidity risk index of patients.

The IPython Notebooks provide a perfect environment, in which the users can explore the tools, tune the settings and try different features and modelling techniques. And, due to nature of the problems, designing a complex Graphical User Interface with limited functionality and vague modelling process may limit the usage and adaptation of the tools.

10.2 Prerequisites and Deployment

The ERMER and the T-CARER models are using high-performance statistical software packages, which are highly optimised and scalable and are actively maintained. In addition to the installation of Python (Ver. 3.5, 64bit) and the *Jupyter* library, a number of other libraries and dependencies must be installed and configured, before using the toolkits.

Firstly, the developed toolkits use scientific Python packages, including *Pandas*, *SciPy* and *NumPy* libraries, in order to implement the pre-processing stages and exploratory analyses. Moreover, the toolkits call statistical modelling functions in the *SciKit-Learn* library, to apply basic modelling techniques, like LR and Breiman (2001) RF. Furthermore, the input data are read from a configured MySQL server using *SQLAlchemy* library, via calling queries and predefined MySQL procedures. Finally, the optional summary plots are generated using *ggplot* and *Matplotlib* libraries.

However, there are two main Operating System (OS) related dependencies for the ERMER and the T-CARER toolkits. Firstly, the ERMER package is dependent on the Microsoft Infer.Net library (Research, 2016) if the BPM is chosen to be the modelling method of choice. Therefore, it may only be run on a Windows OS that has Infer.Net installed. Also, the applied WDNN in the T-CARER uses the TensorFlow (Google, 2016) installation (Ver. 1.0, 64bit). The WDNN model has some dependencies for the unofficial TensorFlow *Contrib* sub-packages, which may only be able to run on Linux based distributions, like Fedora, Ubuntu and Mac OSs.

Furthermore, it is recommended to chose a machine that has a minimum of 50GB of free disk memory and 32GB of physical RAM for sample sizes less than three millions, to satisfy the requirements for in-memory processing. Also, use of a GPU and a CPU that are suitable for Deep Neural Network (DNN), like XEON family of INTEL CPU and the NVIDIA's Pascal GPU architecture.

Finally, the tools are fully documented and are made available on-line. The toolkits are hosted by the GitHub (`https://github.com/mesgarpour`) and are licensed under

Apache License, Version 2.0 (Apache, 2016). The granted rights under the Apache License, Version 2.0, covers copy, modify and distribute in source and binary forms. The present release of the toolkits is considered as Version 1.0 and are open to third party contributions. Also, the IPython Notebooks may be hosted locally, and after meeting the prerequisites, they can be configured to work with custom input data, settings and models.

10.3 Data Preparation and Feature Engineering

In this part of the development, the data preparation and the feature engineering steps have been applied, which will be included in the next public version as a seperate toolkit (Figure 10.1). This step is equivalent to the *Step II* and *Step III* (Section 7.2 and Section 7.3) of the healthcare pre-processing framework (Chapter 7).

The data preparation step includes custom data cleaning and treatments, and the feature engineering includes core feature generation and temporal feature generation. The feature engineering step has a generic design with this in mind that multiple healthcare data sources may be provided and wide range of temporal features might be needed.

Figure 10.4 and Figure 10.6 demonstrate the process-flow diagram of the data preparation and the feature engineering steps, that are applied using MySQL procedures and Python modules, with the expansion of possible input data sources. The main steps can be controlled using an IPython Notebook and a number of configuration files.

The designed mapping tables are released as part of the development toolkits, except some of the HSCIC data that have a very protective licensing.

10.4 Inputs and Outputs

The T-CARER uses a generic input layer to read the input data, which is configured using a Comma Separated Value (CSV). Based on the input configuration file, the input columns are selected from the specified MySQL tables and the temporal features are generated using an aggregation function, such as average, min, max, count of prevalence states and count of other states. Figure 10.5 presents a snapshot of the configuration file.

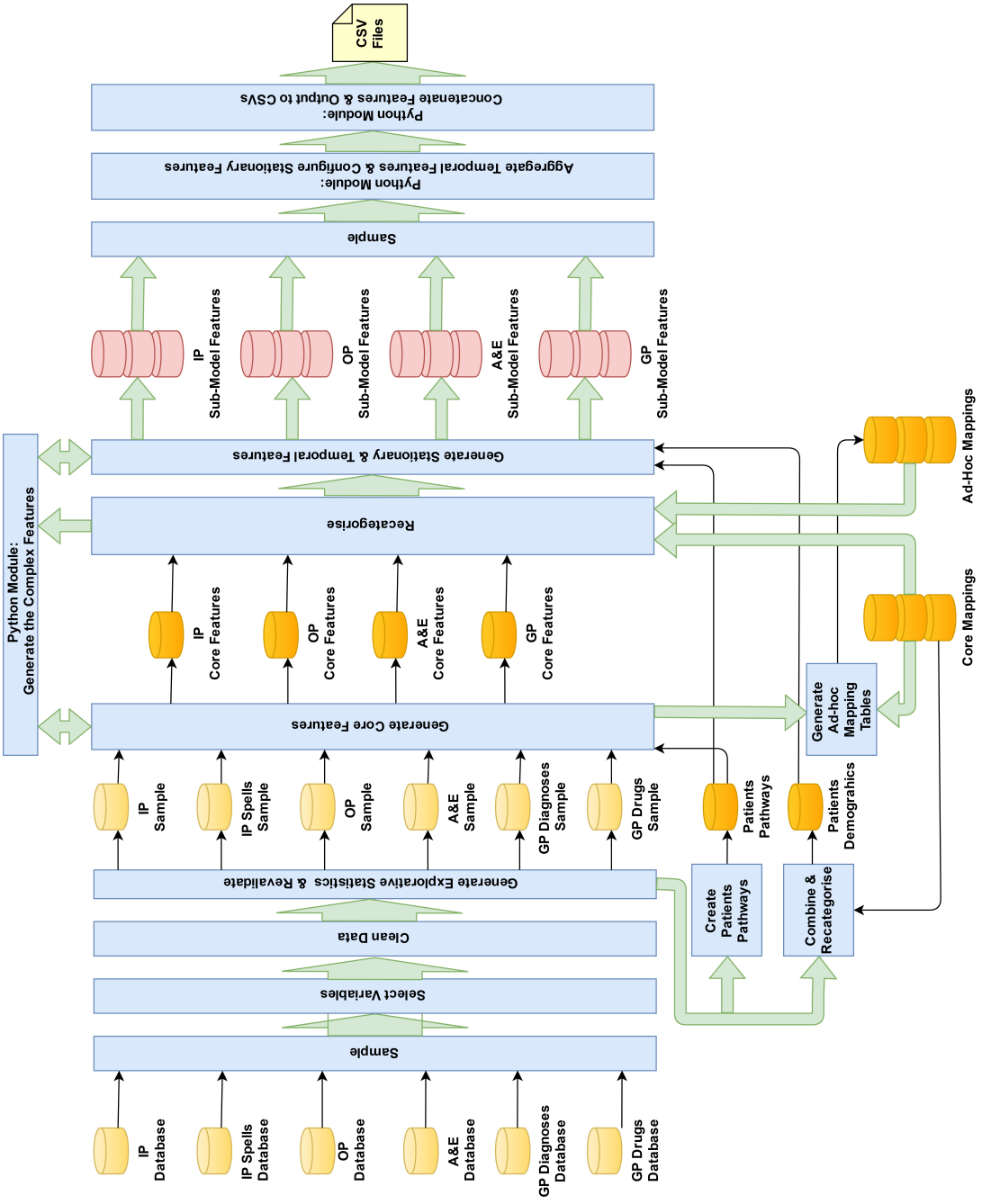

FIGURE 10.4: Process-flow diagram of the feature generation

	A	B	C	D	E	F	G	H	I	J	K
1	Table_Reference_Name	Table_History_Name	MySQL_Table	Variable_Name	Variable_Temporal	Variable_Aggregation	Variable_Type_Original	Variable_Type	Variable_dType	Variable_Max_States	Selected
2	heslp	heslp	tcarer_features_heslp_heslp	localID	0		INT	ID	i4		0
3	heslp	heslp	tcarer_features_heslp_heslp	patientID	0		STR	ID	U32		1
4	heslp	heslp	tcarer_features_heslp_heslp	Admidate	0		STR	ID	U10		0
5	heslp	heslp	tcarer_features_heslp_heslp	label30	0		INT	TARGET	i1		1
6	heslp	heslp	tcarer_features_heslp_heslp	label365	0		INT	TARGET	i1		1
7	heslp	heslp	tcarer_features_heslp_heslp	gender	0		INT	CATEGORICAL	i1	1	1
8	heslp	heslp	tcarer_features_heslp_heslp	ethnos	0		INT	CATEGORICAL	i2	0,1,2,3,4,5,6,7,8,9,10	1
9	heslp	heslp	tcarer_features_heslp_heslp	imd04rk	0		INT	CATEGORICAL	i2	0,1,2,3,4,5,6,7,8,9,10	1
10	heslp	heslp	tcarer_features_heslp_heslp	ageTrigger	0		INT	CATEGORICAL	i4	10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90	1
11	heslp	heslp	tcarer_features_heslp_heslp	gapDays_0t30d	1	others_cnt, avg	INT	CONTINUOUS, CONTINUOUS	i4, f4		1
12	heslp	heslp	tcarer_features_heslp_heslp	gapDays_30t90d	1	others_cnt, avg	INT	CONTINUOUS, CONTINUOUS	i4, f4		1
13	heslp	heslp	tcarer_features_heslp_heslp	gapDays_90t180d	1	others_cnt, avg	INT	CONTINUOUS, CONTINUOUS	i4, f4		1
	heslp	heslp	tcarer_features_heslp_heslp	gapDays_180t365d	1	others_cnt, avg	INT	CONTINUOUS	i4, f4		1

FIGURE 10.5: Toolkits: Screenshot of the inputs configuration file

Firstly, all the major steps in IPython Notebooks produce outputs in form of statistical outputs, configurations outputs or features backups. Also, an additional Notebook was added for more detailed analyses of the models' performances and benchmarking, to reduce the main modelling Notebook's complexity.

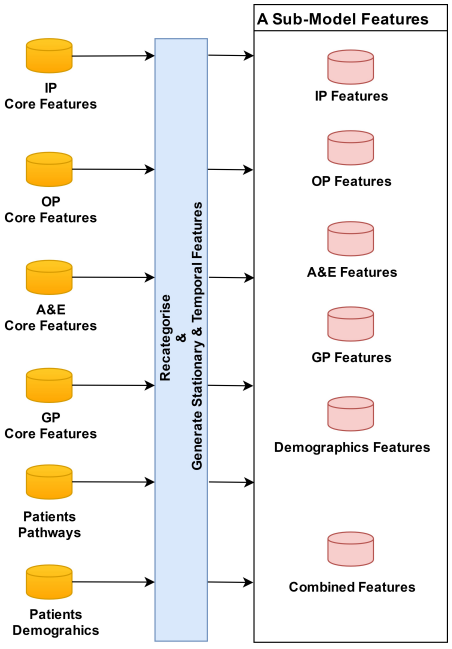

FIGURE 10.6: Detailed process-flow diagram of the feature generation step

10.5 Pre-Processing and Modelling Techniques

According to the proposed healthcare pre-processing framework in this thesis, the feature generation is called via the IPython Notebooks. This stage is highly resource-intensive; therefore the features and settings may be saved throughout the workflow.

Moreover, the filtering stationary features step and the filtering correlated features (Section 7.4) are partially automated. Because, the number of input features can be high and it can be burdensome for the user to specify them manually in each run. However, it is strongly recommended to manually review the features to make sure that the right features are removed. Figure 10.7 presents a screenshot of the filtering stationary features step.

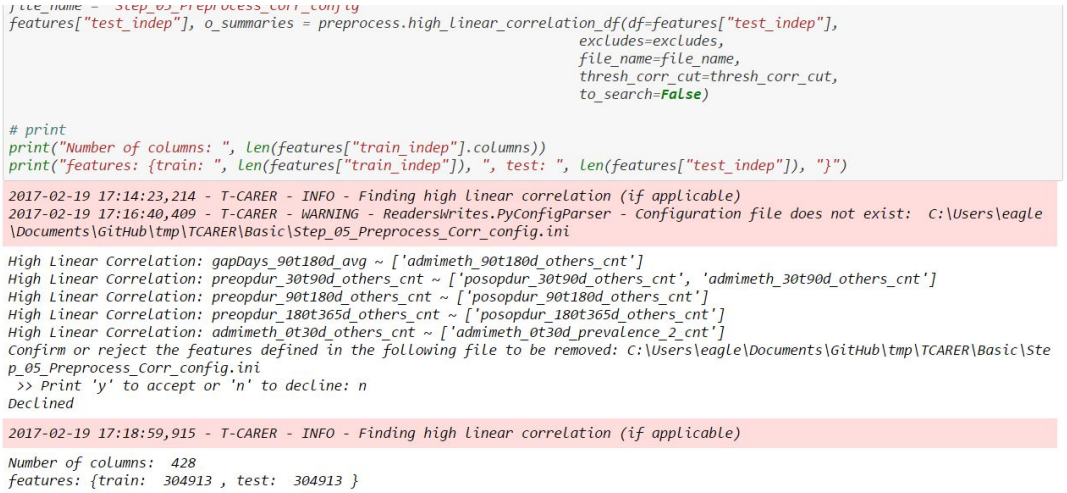

FIGURE 10.7: Toolkits: Screenshot of the filtering stationary features step

Similarly, in the feature ranking step, the list of ranked features can be approved, before progressing with the analysis. These two feature removal steps are only triggered after the features list is confirmed by the user!

Finally, the basic modelling approaches in the modelling stage, including the LR and the RF, are presented in the main Notebooks and can be configured quickly. The advanced models, BPM and WDNN, are implemented in separate Notebook, due to their complex configurations.

10.6 Discussions

The version 1.0 of the ERMER and the T-CARER toolkits are now released and can be applied and customised to any type of healthcare setting or data source. Also, there is a number of mapping tables provided as part of the release, which is used for re-categorising features.

One of the most challenging tasks during the development of the feature pre-processing was efficient feature generation and processing. In part, this was due to complex nature

of healthcare data, but also the fact that very little public research was available about the prepossessing the NHS data with a clear and detailed specification.

Moreover, another major challenge in the development was the development of mapping tables, to reduce sparsity and improve fitness. For instance, the design of effective diagnoses or cost grouping can be considered as the most important step in patient risk modelling, due to their high correlations and high levels of complexities. And, the design of features that have adequate precision with optimally low sparsity are very complex, when the number of feature categories is very high, population sample sizes are moderate, and prevalence of categorising varies across several dimensions.

Finally, there is a plan to release an extension of the feature pre-processing that can be fully implemented on HES, Secondary Uses Service (SUS) and General Practice (GP) data. In this separate extension, the hospital features will include all three sectors, inpatient, outpatient, and Accident and Emergency (A&E). In addition, the mapping tables for feature re-categorisation are going to be included, to allow the generation of features that have lower sparsity and higher significance, but are clinically meaningful.

In the following chapter, the concluding remarks and feature works are highlighted.

Chapter 11

Concluding Remarks

In this chapter, firstly, a brief overview of the thesis is provided. Then, the future work and extensions are highlighted.

11.1 Conclusion

In this thesis, we have investigated several important problems regarding the identification of patients risks. The principal motivation of this research was to provide a framework for analysing administrative healthcare data to generate significant features that are correlated to patients health and care status, and then to model the high layers of risks complexities using robust techniques. Because, at present, no other framework available for pre-processing healthcare data, and current predictive models for patients risk are very simplistic and mainly fail to learn the significant complex patterns in health and care status.

Moreover, hospital readmissions are rising, due to growth in long-term comorbidities, the ageing population, premature discharges and accidents. It has been estimated that about half of the Ambulatory Care Sensitive Conditions (ACSCs) can be predicted and may be avoided by adequate interventions. The present models of hospital emergency readmission and comorbidity risks have moderate performances and use very similar features and modelling techniques.

This thesis looked at three sub-problems in the area of healthcare modelling. Firstly, a healthcare pre-processing framework was developed to prepare data, generate a pool of features and select important features. Then, an Ensemble Risk Model of Emergency Admissions (ERMER) was developed as a decision support tool, to help clinicians and

commissioners to identify risks of patients. Next, a Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER) was designed to identify patients' risks of comorbidities and complexities, with more accuracy and higher confidence.

Firstly, the proposed healthcare pre-processing framework was used to sample, clean and treat input data. Then, it creates *super-spells* out of related *episodes*. After that, it systematically generates a pool of features, transforms, filters correlated features, ranks feature importance and select top features. The proposed healthcare pre-processing framework has been proven to be effective in prediction models of readmission and comorbidity risks, and it has potential to be used in other areas of healthcare modelling.

Secondly, the ERMER was developed using an Ensemble of Bayes Point Machine (BPM) models. The sub-models in the Ensemble were generated using a collection of different cohorts, including prior *spells*, prior emergency admissions, prior operations and age limits. Then, the ERMER used a hill-climbing heuristic to optimise the weighted average rank of predicted estimates using several performance criteria. Introducing prior probabilities and using a collection of weaker sub-models have been demonstrated to be effective in the production of highly stable readmission models with strong confidence and accuracy.

Finally, the T-CARER implements a comorbidity risk model with inclusion of temporal dimensions: Length-of-Stay (LoS) and delta-time between admissions. Also, in addition to comorbidity groups, T-CARER adds population stratification, consultant specialities, operations and complications. The offered solution introduces a generic method for generating a pool of features out of re-categorised and temporal features, in order to create a customised comorbidity risk index.

Towards meeting our objectives, several extracts of the Hospital Episode Statistics (HES) within a 10-year timeframe have been obtained, to train, test and cross-validate the models. And, all the proposed models were benchmarked against previous models from several aspects, including different population cohorts, time-frames, fitting algorithms and risk segments. The benchmarks of the ERMER and the T-CARER using multiple comparison criteria have shown significant improvements against previous models, in terms of precision, accuracy and stability.

Finally, the proposed solutions are implemented in the form of user-friendly toolkits, using Jupyter IPython Notebook. The toolkits use a wide range of high-performance computing packages to process input data, generate features, and train and test models. Moreover, the developed IPython Notebooks provide an ideal environment for researchers to model a custom predictive model with great flexibility in feature generation and applying modelling algorithms.

In conclusion, the produced solutions in this research are transferable to other healthcare environments, due to the general applicability of the framework and modelling approaches. Moreover, the research outcomes are expected to contribute to the academic and healthcare providers communities. Furthermore, development of customised prediction models using the ERMER and the T-CARER can significantly save time and research costs in healthcare, without compromising the quality of models.

11.2 Extensions and Future Work

In this thesis, we have studied the pre-processing healthcare data, the emergency readmission modelling and the comorbidity risk modelling problems, using hospital administrative data. The developed solutions look at the general patterns in the data, in contrast to micro-processes and pathways of individual patients. And, there is a great potential in the modelling patients risk using approaches, like process mining techniques (Mans et al., 2015), discrete event simulation (Marshall et al., 2015) and unsupervised clustering (Burgel et al., 2014). Such approaches allow defining more complex clusters of comorbidities and other complexities, which were not possible to derive in this research.

Moreover, with the advancement in computing and Deep Neural Network (DNN) learning algorithms, there is a great potential in design and development of *Transfer Learning* approaches (Section 4.2.1) that can adapt to different healthcare settings. Recently, there have been some advancements in the area of Deep Convolutional Neural Network (DCNN) (Abadi et al., 2016, Pan and Yang, 2010), and it is of economic and academic interest to produce models that can adapt to different settings and objectives. Also, *Transfer Learning* approaches allow to discover processes across different settings and to define parallel models for patients that apply generic and specialised algorithms (Kuhn and Johnson, 2013).

Furthermore, there are other approaches for modelling the temporal aspects that can potentially have good prediction performance. Methods, like DNN (Dean and Kanazawa, 1989, Koller and Friedman, 2009) and Deep Neural Network Hidden Markov (DNN-HMM) (Dahl et al., 2012) allow to encode stationary and dynamic time-slices and unobserved system states.

In below, an outline of the future work directions is presented:

- Integrating a process mining approach into the healthcare pre-processing framework;
- Designing a DCNN to predict patients risks of comorbidities and emergency readmission.
- Developing a latent model using a Bayesian or a DNN, to model the temporal and hidden aspects of the models.
- Producing a heuristic to create an Ensemble readmission model that can take advantage of a wider set of modelling algorithms.

Appendix A

Appendices

A.1 Background Research

A.1.1 Major Emergency Readmission Models

TABLE A.1: Studies investigating factors related to hospital readmission (part 2: commercial solutions)

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
MARA (Milliman Advanced Risk Adjuster): DxAdjuster, RxAdjuster & CxAdjuster (Milliman, 2016)	Milliman, UK	NR[g]	NR	NR	NR	NR	NR	NR	NR
HealthNumerics-RISC; Episode Risk Groups (ERG); Pharmacy Risk Groups (PRG); EpisodeTreatment Groups (ETG); Impact Pro; Natural History of Disease (NHD) (OPTUM, 2014, 2016b,c), App. (Kasteridis et al., 2015)	Optum, UnitedHealth Group, International	12-m admi.	NR	All	SUS, GP, social care & mental health	NR	NR	– Acute (12-m): AUC 0.84 – Acute (3-m): AUC 0.85 – Acute + GP (12-m): AUC 0.85 – Acute + GP (3-m): AUC 0.86	CMG, demographics, DS, LoS, psychological & social care
SF Health Surveys (Maruish, 2011, OPTUM, 2016d), App. (Scoggins and Patrick, 2009)	Optum, UnitedHealth Group, International	NR	NR	Including age 18+	Questionnaires: – SF-36v2: 36 questions & measures health – SF-12v2: 12 questions & similar to SF-36v2 – SF-8: 8 questions & similar to SF-36v2 – PIQ-6: impact of pain on health	NR	NR	NR	Physical & psychological
DxCG Risk Analytics (Verisk Health, 2016), App. (Freund et al., 2010, 2013)	Verisk Health, USA	12-m admi.	DCG	All	Hospital: Inpatient, Outpatient	NR	NR	NR	Demographics, DCG & Pharmacy

Continued on next page

Table A.1 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
Probability of Readmission (Pra); PraPlus (JHU, 2016c), App. (Allaudeen et al., 2011, Bowles and Cater, 2003, Doñate-Martínez et al., 2014, Novotny and Anderson, 2008, Pacala et al., 1995, 1997, Sidorov and Shull, 2002)	University of Minnesota & Johns Hopkins University, USA	12-m admi. (Doñate-Martínez et al., 2014), 30-d admi. (Allaudeen et al., 2011)	Regr.	Including age 65+	Questionnaires: – Pra: 8 questions & predict usage of health services – PraPlus: Pra + 8 questions & predict health needs – Spain: using Pra questionnaire in 3 health departments (Doñate-Martínez et al., 2014) – UCSF Medical Center (Allaudeen et al., 2011)	– NA & 500 patients (Doñate-Martínez et al., 2014) – NA & 164 patients (Allaudeen et al., 2011)	– 2-y: 2008-10 (Doñate-Martínez et al., 2014) – 5 weeks: 2008-08 (Allaudeen et al., 2011)	– Sen. 0.54; Spe. 0.81; PPV 0.30; AUC 0.67 (Doñate-Martínez et al., 2014) – Sen. 0.31; Spe. 0.72; AUC 0.56; PPV 0.38 (Allaudeen et al., 2011)	Admissions, CMG, demographics, physical & social care,
The Johns Hopkins Adjusted Clinical Groups (ACG) System: Care Management; Population health (Lemke, 2013, JHU, 2014, 2016a,b), App. (Sollis, 2016)	Johns Hopkins University, USA	30-d admi.	LR	All	USA: hospital (IMS)	– NR (JHU, 2014) – 270,020 & 272,050 patients (Lemke, 2013)	– 2-y: 2009-10 (JHU, 2014) – 3-y: 2009-12 (Lemke, 2013)	– ¿0.80: AUC 0.73; PPV:0.16 (JHU, 2014) – Sen. 0.06; Spe. 0.99; AUC 0.75; PPV:0.52 (Lemke, 2013)	Admissions, CMG, demographics, DCG, DS, LoS, medical, operations, psychological & social care
Health Risk Assessments (HRA) Framework (CDC, 2016)	Centers for Disease Control & Prevention, USA	Risk assessment	Framework	Including Medicare	NR	NR	NR	NR	Clinical, demographics, physical & psychological
3M Potentially Preventable Readmissions (PPR) Grouping Software; 3M Potentially Preventable Complications (PPC) Grouping Software; 3M Population-focused Preventables Software (3M, 2016)	3M, International	7-, 15-, 30-d admi.	NR	NR	NR	NR	NR	NR	NR

Continued on next page

Table A.1 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables** [f]
LexisNexis Risk Solutions: Risk Navigator Performance; Population Health Monitor (LexisNexis, 2016a,b,c)	LexisNexis, USA	Risk assessment	NR	NR	NR	NR	NR	NR	Administrative, medical claims, pharmacy & Public records data
Diagnosis-Related Group (DRG) (CMS, 2016a,c, NTIS, 2016)	CMS, USA	Grouping	Grouping	All	NR	NR	NR	NR	Diagnoses
Health & Human Services - Hierarchical Condition Categories (HHS-HCC); CMS-HCC (Kautter et al., 2014, CMS, 2016b)	Centers for Medicare & Medicaid Services (CMS), USA	Medical expenditure	Reg	Including under 65 & disabled	USA: hospital & claims (MarketScan)	NR	NR	NR	CMG, DCG, demographics, diagnoses, DS, pharmacy & psychological
Care Pathways framework (Health Dialog, 2016a,b)	Health Dialog, Rite Aid, USA	12-m	NR	Including cardiovascular disease, diabetes, chronic kidney disease, Asthma, COPD, musculoskeletal (low back, hip, knee, shoulder), cancer & mental health	NR	NR	NR	NR	Census data, demographics, financial data & geographical data

[a] (App.) Applied cases. [b] (Admi.) admissions; (AMI) Acute Myocardial Infarction; (CCI) Charlson Comorbidity Index; (Cond.) conditions; (COPD) Chronic Obstructive Pulmonary Disease; (EoL) End of Life; (HF) Heart Failure; (HR) Hip Replacement; (KR) Knee Replacement. [c] (-d) -day; (-m) -month; (-y) -year. [d] (BPM) Bayes Point Machine; (Cond.) Conditional; (C-PHD) Coxian Phase-type Distribution; (DCG) Diagnostic Cost Grouping; (FP) Fractional Polynomials; (GLM) Generalised Linear Models; (LR) Logistic Regression; (ANN) Artificial Neural Network; (PR) Piosson Regression; (Regr.) Linear Regression; (RF) Random Forest; (SVM) Support Vector Machine; (TM) Transition Model. [e] (AUC) C-Statistic; (D-stat) D-Statistic; (IDI) Integrated Discrimination Improvement; (PPV) Positive Predictive Value; (Sen.) Sensitivity; (Spe.) Specificity; (X^2): Chi-Square Statistic. [f] (CMG) Case-Mix Group; (DS) Diagnosis Scoring; (LoS) Length of Stay. [g] (NR) Not Reported.

TABLE A.2: Studies investigating factors related to hospital readmission (part 2: open researches)

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
Ensemble Risk Model of Emergency Readmission (ERMER) (Mesgarpour et al., 2016)	University of Westminster, UK	12-m admi.	Ensemble, BPM, SVM & RF	Including age 1+	England: hospital (HES) Inpatient	– 578,936 & 578,936 patients – 705,461 & 705,461 patients – 662,356 & 662,356 patients	– 5-y: 1999-04 – 5-y: 2004-09 – 5-y: 2000-05	Sen. 0.42 to 0.49; Spe. 0.88 to 0.92; AUC 0.76 to 0.77; PPV 0.72 to 0.74	Admissions, demographics, DS, LoS, operations & specialities
Cronin et al. (2014)	Massachusetts General Hospital, Boston, USA	30-d admi.	C-PHD & LR	Excluding chemotherapy, radiation, dialysis & obstetrics	Massachusetts General Hospital, USA: Inpatient, Outpatient, emergency, laboratory, billing & medications	36,462 & 9,325 patients	– 1 year: 2012-13 – 2-m validation: 2013	AUC 0.53 to 0.70	Admissions, DS, pharmacy, psychological & social care
Bottle et al. (2014)	National Institute for Health Research & Dr. Foster Unit, UK	– Readmission: 7-, 30-, 90-, 182- & 365-d – Mortality: 30-d – Outpatient non-attendance – Return to theatre: 90-d – Other cond.: 7- & 28-d	LR, SVM, ANN & RF	Including condition specific	England: – Hospital (HES & SUS): Inpatient, Outpatient & A&E – Mortality	– CCI: all Inpatient – HR Procedure: 260,370 patients in total – KR Procedure: 286,590 patients in total – Other cond.: NR	– 1 year: CCI for admi. & mortality: 2008-09 – 5-y: CCI for admi. & mortality with weights on time: 2008-13 – 6-y: HR & KR Operations: 2007-13 – 4-y: Others cond.: 2009-13	– Readmission: AUC 0.58 to 0.68 – Mortality: AUC 0.68 to 0.86	Admissions, clinical, CMG, demographics, deprivation, LoS & psychological

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
Hospital Admission Risk Prediction (HARP) (Ontario, 2013a,b)	Health Quality Ontario, Canada	1-m & 15-m admi.	LR	Including age 18+	Ontario & Manitoba, Canada: hospital	191,321 & 191,627 episodes	4-y: 2008-12	– 1-m: Sen. 0.75; Spe. 0.46 to 0.50; AUC 0.66 to 0.68; PPV 0.17 to 0.18 – 15-m: Sen. 0.68 to 0.70; Spe. 0.58 to 0.59; AUC 0.69 to 0.70; PPV 0.51 to 0.52	Admissions, CMG, demographics, DS, LoS & resource intensity level
Hospital Admission Risk Program (HARP) Victoria (SGV, 2002, 2011, 2016)	Victorian Government Department of Human Services, Australia	Measure inputs, processes & outputs of services	Framework	All	State Government of Victoria: hospital	NA	NA	NA	Admissions, CMG, demographics, LoS, quality of life, physical & psychological
QAdmissions score (Hippisley-Cox and Coupland, 2013)	University of Nottingham, North East London Commissioning & National School for Primary Care Research UK	24-m admi.	C-PHD & FP	Including age 18 to 100	QResearch practices, England: – Hospital (HES) – GP (QResearch database & Clinical Practice Research DataLink)	2,849,381 & (1,340,622 & 2,475,360) patients	2-y: 2010-12	– HES-GP: AUC 0.77 to 0.78; D-stat 1.69 to 1.76 – GP: AUC 0.76 to 0.77; D-stat 1.58 to 1.65	Admissions, clinical, CMG, demographics, deprivation, pharmacy & psychological

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables** [f]
Billings et al. (2013)	Nuffield Trust & Department of Health, UK	12-m admi.	LR	Including age 18 to 95	5 Primary Care Trusts (PCTs), England: Newham, Cornwall, Kent, Croydon, Redbridge: – Hospital (SUS): Inpatient, Outpatient & A&E – GP	1,836,099 & 1,836,099 patients	3-y: 2007-10	– IP data: Sen. 0.05; AUC 0.73; PPV 0.53 – IP+AE data: Sen. 0.05; AUC 0.74; PPV 0.53 – IP+AE+OP data: Sen. 0.05; AUC 0.75; PPV 0.52 – IP+AE+OP+GP Data: Sen. 0.06; AUC 0.78; PPV 0.54	Admissions, clinical, CMG, demographics, deprivation, pharmacy, psychological, social care & speciality
Shulan et al. (2013)	Department of Veterans Affairs USA	30-d admi.	LR	All	Veterans Healthcare Network Upstate New York: hospital	4,359 & 4,359 patients	13-m: 2011-12	AUC 0.80	Admissions, DCG, demographics, DS, Geographical, insurance & LoS
CMS Model - Hospital-Wide All-Cause Unplanned Readmission Measure (YNHHSC/CORE, 2012, 2015)	CMS, USA	30-d admi.	LR	Including age 65+; Excluding insufficient data, against medical advice, cancer treatment, not typically cared for in short-stay acute care hospitals & PPS-exempt cancer hospitals	USA: Hospitals (Medicare) Inpatient	7,957,901 patients in total (2007-08)	– 2-y: 2007-09 – 2-y & 1 year: 2007-10	AUC 0.66	CMG, demographics & operations

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
LACE+ (van Walraven et al., 2012)	Canadian Institutes of Health Research, Physicians' Services Incorporated Foundation & University of Ottawa, Canada	30-d admi.	LR	Including age 18+; Excluding psychiatric, obstetric & no healthcare coverage	Ontario's hospital, Canada: – Discharge Abstract Database – Ontario Mental Health Reporting System – National Ambulatory Care Reporting System – Registered Patient Database	250,000, 250,000 patients	6-y: 2003-09	AUC 0.76 to 0.77	Admissions, CMG, demographics, DS & LoS
Patients at Risk of Re-hospitalisation within the next 30 days (PARR-30) (Billings et al., 2012)	Nuffield Trust & Department of Health, UK	30-d admi.	LR	All	England: hospital (HES) Inpatient	576,868 patients & Bootstrapping	3-y: 2006-09	Sen. 0.05; Spe. 0.99; AUC 0.70; PPV 0.59	Admissions, CMG, demographics, deprivation, geographical
Nairn Case Finder (Baker et al., 2012, NHS, 2010)	NHS Highland, Scotland, UK	12-m	Cond. PR	Including care-home, proactively care & with complex, palliative or EoL care needs	1 GP practice Scotland: – Hospital: Inpatient, Outpatient – GP	96 patients	3-y: 2006-08	AUC 0.79	Admissions, CMG, demographics, DS & psychological
Scottish Patients at Risk of Readmission & Admission (SPARRA) - version 3 (NHS, 2011)	NHS Scotland, UK	12-m admi.	LR	Including age 16+; Excluding psychiatric Inpatient only	Scotland: – Hospital: Inpatient, Outpatient & A&E – Community Health Partnerships – Psychiatric Inpatient – GP	3,506,796 patients in total	4-y: 2006-10	– Admi.: Sen. 0.10; PPV 0.60 – Bed days: Sen. 0.20	Admissions, CMG, demographics, deprivation, LoS, pharmacy & psychological

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables** [f]
Hammill et al. (2011)	American Heart Association National Center, USA	30-d admi. & mortality	GLM & Cond. GLM	Including age 65+ & HF & 2+ days LoS	Association's Get With The Guidelines-Heart Failure registry from 344 hospitals, USA: – Hospital (Medicare): Inpatient – American Heart Association: HF	36,267 patients & Bootstrapping	3-y: 2004-07	– Mortality: AUC 0.71 to 0.76 – Admi.: AUC 0.59 to 0.61	CMG, clinical, demographics & psychological
Jen et al. (2011)	Dr. Foster Unit, National Institute of Health Research, UK Clinical Research Collaboration, UK	28-d admi. & mortality	LR	Including 78 diagnosis groups & 126 operations	England: hospital (HES) Inpatient	NR & Bootstrapping	10-y: 1996-06	– Mortality: AUC 0.50 to 0.98 – Admi.: AUC 0.55 to 0.78	Admissions, CMG, demographics, deprivation, DS & operations
Elders Risk Assessment Index (ERA) (Crane et al., 2010)	Mayo Clinic, Rochester, USA	24-m admi.	LR	Including age 60+; Excluding skilled nursing care	Rochester, USA: Primary Care Internal Medicine	12,650 patients & Bootstrapping	4-y: 2003-07	– Hospital visits: AUC 0.70 – Emergency room visits: AUC 0.64 – Combined: AUC 0.68	Admissions, demographics & LoS
LACE: stands for: length of stay (L); acuity of the admission (A); co-morbidity of the patient (C); & emergency department use (E) (van Walraven et al., 2010)	Canadian Institutes of Health Research, Physicians' Services Incorporated Foundation & University of Ottawa, Canada	30-d admi.	LR & FP	Including age 18+ & patients of medical & surgical services	11 hospitals in 5 cities, Ontario, Canada: – Discharge Abstract Database – National Ambulatory Care Reporting System – Registered Patient Database – Questionare	4,812 & (4,812 & 1,000,000) patients	3-y: 2004-08	AUC 0.68	Admissions, DS & LoS

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
Amarasingham et al. (2010)	University of Texas & National Institutes of Health, USA	30-d admi. & mortality	LR	Including HF	Parkland Memorial Hospital & admi. records to 136 hospitals, Dallas, USA: – Hospital – Mortality	1,372 patients & Bootstrapping	18-m: 2007-08	– Mortality: AUC 0.86; IDI 0.10 – Admi.: AUC 0.72; IDI 0.11	Demographics, insurance & deprivation
Hasan et al. (2010)	Agency for Healthcare Research & Quality & Department of Health & Human Services , USA	30-d admi.	LR	Including age 18+; Excluding patients under the care of their GP	6 large academic medical centers across the USA: – Questionnaire – Hospital – Mortality	7,287 & 3,659 patients	2-y: 2001-03	AUC 0.61	Demographics, insurance & LoS
Scottish Patients at Risk of Readmission & Admission for Mental Health (SPARRA-MH) (NHS, 2009)	NHS Scotland, UK	12-m admi.	LR	Including age 15+; Including psychiatric hospitals	Scotland: psychiatric Inpatient	36,500 & NR patients	4-y: 2003-07	AUC 0.74; PPV 0.56 to 0.75	Admissions, CMG, demographics, LoS & psychological
Howell et al. (2009)	Queensland Health & University of Queensland, Brisbane, Australia	12-m admi.	LR	Including 28 chronic medical cond.	Queensland Hospital: hospital Inpatient	13,207 & 4,492 patients	2-y: 2005-07	Sen. 0.44; Spe. 0.78; AUC 0.65	Admissions, CMG, demographics, deprivation & psychological
Demir et al. (2009)	University of Westminster, UK	38-d admi.	TM & LR	Including COPD	England: hospital (HES)	307,394 & NR patients	7-y: 1997-04	AUC 0.73	Admissions, demographics, deprivation & LoS
Predict Emergency admissions Over the Next year (PEONY) (Donnan et al., 2008)	NHS Scotland, UK	12-m admi.	LR	Including age 40+	Tayside, Scotland: – Hospital (Scottish Morbidity Record) – GP	90,522 & 90,879 patients	8-y: 1996-04	Sen. 0.08; Spe. 0.99; AUC 0.79; PPV 0.59	Demographics, deprivation & LoS

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
Predictive Risk Stratification Model (PRISM) (Dialog, 2008), App. (Hutchings et al., 2013, TRUST, 2016)	Health Dialog, Wales, UK	12-m admi.	LR (Dialog, 2008) GLM, C-PHD (Hutchings et al., 2013)	Including patient with 39+-m registration	Wales: – Hospital: Inpatient, Outpatient – GP – Questionnaire (PRISMATIC research) (Hutchings et al., 2013)	149,038 & 149,038 patients	3-y: 2004-07	¿0.80: Sen. 0.49; PPV 0.16 (Dialog, 2008)	CMG, demographics, deprivation, pharmacy & psychological
Silverstein et al. (2008)	Baylor Health Care System, Dallas, USA	30-d admi.	LR	Including age 65+	7 acute care hospitals, Dallas, USA: Inpatient	19,528 & 9,764 patients	2-y: 2002-04	AUC 0.65	CMG, demographics, insurance & operations
Billings and Mijanovich (2007)	New York Community Trust & the United Hospital Fund, USA	12-m admi.	LR	Inluding eligible Medicaid disabled patients & disabled adult patients with serious mental condition	New York City, USA: – Hospital (Medicaid) – Supplemental Security Income of disabled adult	35,000 & 35,000 patients	4-y: 2000-04	– Medical admi. (Medicaid, mentally ill): PPV 0.60, 0.19 – Mental illness (Medicaid, mentally ill): PPV 0.50, 0.56 – Substance abuse (Medicaid, mentally ill): PPV 0.13, 0.18	Admissions, CMG, demographics, operations, psychological, social care & speciality
Emergency Admission Risk Likelihood Index (EARLI) (Lyon et al., 2007)	NHS Executive North West Research & Development, UK	12-m admi.	LR	Including age 75+; Excluding unsuitable patients (terminal cancer & mental illness)	1 PCT, England: – Questionnaire – Hospital (HES) – Mortality	3,032 patients & Bootstrapping	18-m: 2002-03	Sen. 0.64; Spe. 0.64, AUC 0.69; PPV 0.35	Admissions, CMG & Physical
Combined Predictive Model (CPM) (Paton et al., 2014, Wennberg et al., 2006), App. (HI, 2013)	Department of Health, UK	12-m admi.	LR	All	5 PCTs, England: – Hospital (HES): Inpatient, Outpatient, A&E; – GP	280,000 & 280,000 patients	3-y: 2002-05	AUC 0.78, PPV: 0.59	Admission, CMG, demographics, medical, pharmacy & psychological

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
Bottle et al. (2006)	Dr Foster Unit, UK	12-m admi.	LR	Including ACSC	England: – Hospital (HES): Inpatient – Mortality	1,373,754 & 1,373,754 patients	5-y: 1999-04	– Top 250,000 high risks: – All: Sen. 0.27; Spe. 0.93; AUC 0.72; PPV 0.30 – Excluding death: Sen. 0.27; Spe. 0.92; AUC 0.70; PPV 0.23 – ACSC: Sen. 0.86, Spe. 0.45; AUC 0.75; PPV 0.20	Admissions, demographics, deprivation, DS & speciality
Patients at Risk of Re-hospitalisation (PARR) (Billings et al., 2006a,c)	Department of Health, UK	12-m admi.	LR	Including age 65+ & 32 most common cond.	England: hospital (HES) Inpatient	42,778 & 42,778 patients	5-y: 1999-04	Sen. 0.54; Spe. 0.72; AUC 0.68; PPV 0.65	Admissions, CMG, DCG, demographics, psychological & speciality
Halfon et al. (2006)	University of Lausanne, Switzerland	30-d admi.	PR	Excluding healthy newborns, residents outside of Switzerland & elective surgical stays that could usually be performed as day surgery.	12 out of 221 acute cares, Swiss: hospital	65,740 & 66,069 patients	1 year: 2000-01	– Nonclinical Model: AUC 0.67; X^2 0.84 – Charlson-based Model: AUC 0.69; X^2 0.18 – SQLape-based Model: AUC 0.72; X^2 0.70	Admissions, CMG, demographics, DS, operations & psychological

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables**[f]
Holman et al. (2005)	National Health & Medical Research Council & Western Australia Department of Health, Australia	30-d admi. & LoS & 12-m mortality	C-PHD, LR, FP & Regr.	Including asthma, AMI, breast cancer mastectomy, transurethral prostatectomy & major depressive disorder	Western Australia: hospital	1,118,989 patients & subsampling	7-y: 1989-96	– Admi: LR AUC 0.64 to 0.77; Cox D-stat 341 to 5715 – LoS: regression R^2 0.13 to 0.33 – Mortality: LR AUC 0.81 to 90; Cox D-stat 184 to 1760	CMG & demographics
Coleman et al. (2004)	American Federation for Aging Research, USA	30-d admi.	LR	Including age 65+; Excluding long-term care & hospice care patients	Medicare, USA: – Hospital (Medicare) – Questionnaire	700 & 704 patients	2-y: 1997-99	– Admin.: Sen. 0.95; Spe. 0.39; AUC 0.77 – Admin. + self-reported: Sen. 0.98; Spe. 0.45; AUC 0.83	Admissions, CMG, demographics, DS, insurance, LoS, physical & social care
Morrissey et al. (2003)	Queen's University Belfast, Northern Ireland, UK	12-m admi.	LR	Including age 65+, & general medicine wards emergency admi.	Antrim Area hospital, Northen Ireland: Questionnaire	487 & 732 patients	8-m: 1997-98	Sen. 0.60; Spe. 0.79; AUC 0.77	Admissions, CMG, demographics, pharmacy & physical
Krumholz et al. (2000)	Yale University, USA	6-m admi.	C-PHD	Including age 65+ & HF; Excluding severe aortic stenosis, severe mitral stenosis & HF caused by a medical illness	9 acute care, Connecticut, USA: hospital (Medicare)	1,129 & 1,047 patients	2-y: 1994-96	– Admi. & Mortality: PPV 0.65 – HF Admi: PPV 0.31 – Admi: 0.59	CMG, demographics, LoS, medical, operations & physical

Continued on next page

Table A.2 – *Continued from previous page*

Authors		Objectives	Modelling	Service	Data			Results	
Tool[a]	**Funders / Org.**	**Outcome**[bc]	**Approach**[d]	**Aspect**[b]	**Sources**	**Size: train & test)**[b]	**Timeframe**[bc]	**Performance**[ec]	**Variables** [f]
Community Assessment Risk Screen (CARS) (Shelton et al., 2000), App. (Doñate-Martínez et al., 2014)	Carle Clinic Association & University of Wisconsin-Madison, USA	12-m admi.	LR	Including age 65+, 6-m prior admi., lived alone, no caregiver, 4+ medications, difficulty in walking, limitations in activities of daily living, memory difficulties, incontinent of urine or stool & multiple illnesses or disabilities	Carle Clinic site, Urbana, USA: – Hospital (Medicare) – Questionnaire	411 & 1,054 patients	2-y: 1993-95	AUC 0.74	Admissions, CMG, demographics, pharmacy, physical & psychological

[a] (App.) Applied cases. [b] (Admi.) admissions; (AMI) Acute Myocardial Infarction; (CCI) Charlson Comorbidity Index; (Cond.) conditions; (COPD) Chronic Obstructive Pulmonary Disease; (EoL) End of Life; (HF) Heart Failure; (HR) Hip Replacement; (KR) Knee Replacement. [c] (-d) -day; (-m) -month; (-y) -year. [d] (BPM) Bayes Point Machine; (Cond.) Conditional; (C-PHD) Coxian Phase-type Distribution; (DCG) Diagnostic Cost Grouping; (FP) Fractional Polynomials; (GLM) Generalised Linear Models; (LR) Logistic Regression; (ANN) Artificial Neural Network; (PR) Piosson Regression; (Regr.) Linear Regression; (RF) Random Forest; (SVM) Support Vector Machine; (TM) Transition Model. [e] (AUC) C-Statistic; (D-stat) D-Statistic; (IDI) Integrated Discrimination Improvement; (PPV) Positive Predictive Value; (Sen.) Sensitivity; (Spe.) Specificity; (X^2): Chi-Square Statistic. [f] (CMG) Case-Mix Group; (DS) Diagnosis Scoring; (LoS) Length of Stay. [g] (NR) Not Reported.

A.1.2 Other Modelling Approaches

There are five major modelling methodologies presented in the previous healthcare predictive modelling studies: simulation, formula-based, statistical, probabilistic and queueing. Methods like formula-based and queueing theory are not practical for this research. Formula-based models derived from empirical data based on observed patterns in variables, like crowding, are generally poor regarding predictability power. Moreover, queueing-theory modelling has poor predictability performance and has limitation due to assumptions, which in a dynamic system this must be relaxed and tested by data from the actual system (e.g. stationary arrival time) (Wiler et al., 2011).

In the following, firstly, the main regression methods in the modelling are outlined. Next, the Markov modelling is presented. Afterwards, the pros and cons of applying simulation modelling on healthcare problems are discussed.

A.1.2.1 Regression Modelling

Regression methods such as LR, LMMs[1], GLM(GLMs), Generalised Linear Mixed Models (GLMMs) have been popular methodologies for modelling pathways and correlations (Garson, 2012).

The LMM is a broad name for Hierarchical Linear Models (HLMs) and Multilevel Models, and it is used in the analysis of variance correlation, regression and factor analysis. The LMM is suitable for modelling problems with dependent observations with correlated errors. The LMM supports analysis of three types of variables: *random effects*, *hierarchical effects* and *repeated effects*. The GLMM is an extension of LMM, which supports a variety of link functions. The fundamental importance of the GLMM is that it supports continuous and ordinal features with non-normal distributions (Garson, 2012).

Kulinskaya et al. (2005) provided a comparison between GLM method against a robust method such as truncated maximum likelihood for a LoS problem. The comparison was carried out on the Nervous System classification of Health Resource Groups (HRG) of the NHS. Although the robust model produce a better fit regarding variance, the differences between the two models were not significantly high in overall. Also, the GLM model did outperform the robust models for a subset of factors.

[1]Mixed Models are also known as Mixed Effects Models, Random Coefficient Regression Models, Multilevel Models or Covariance Components Models.

Adeyemi et al. (2013) presented a GLMM for detecting stage-wise transitions in patient pathways modelling with excluding the clinical flow pathway. The solution modelled the serial independence of the readmission using continuous ratio logit model. The continuous ratio logit model was used on Chronic Obstructive Pulmonary Disease (COPD) patients, to compare the categorisation factors for frailties. The method was effective in detecting the most critical threshold for readmissions.

The advantage of mixed modelling in regression is that it can account for the uncertainty in models and small evidence data. But, the major shortcomings of the mentioned regression methods are the linearity assumption. Moreover, the regression methods ignore the prior distributions and are very dependent on the subject of design; therefore, it is very hard to generalise or re-use them on similar problems.

A.1.2.2 Markov Modelling

In stochastic state-space modelling, Markov Model (MM) (Norris, 1998, Ross, 1993) is one of the most powerful tool. Markov modelling is simplistic approach; however, it becomes very complex, when there are a large number of states, or multiple events are modelled.

In the area of survival analysis, the PHD modelling (Neuts, 1974, 1981) is a popular approach for modelling systems with state-space and latent parameters and is a way of modelling Markov stochastic process. Coxian phase-type distributions (C-PHD) (Cox, 1955) is a special type of PHD with an initial and an absorbing state, which avoids over parametrisation of the model (Fackrell, 2009, Marshall and Zenga, 2009).

Altman (2007) presented an extension to the HMM called Mixed HMM which accounts for two sources of heterogeneity: time constant unit-specific effects and serial correlation. Maruotti (2011) provides two case studies for analysis of longitudinal data using MHMM. The time-series that were studied are patent data and financial econometrics; however, there is no library available for it and the estimation step is complex particularly for parametric models.

Length-of-Stay (LoS) models of is another research area which is highly associated to hospital readmission (Kelly et al., 2012). Xie et al. (2005) develops a MM to predict LoS in continuous time for elderly patients. The proposed continuous time MM uses home care to model movement of elderly in residential, nursing homes and discharge to community or hospital, with two possible states: short-term and long-term stays. The model provides a good fit to the data and can capture the movement of patients

between care facilities.

A.1.2.3 Simulation Modelling

There are three main approaches in simulation modelling in healthcare: Discrete Event Simulation (DES), System Dynamics (SD) and Agent-Based Simulation (ABS) modelling. Simulation modelling techniques provide a better understanding of the interactions and flows. However, most reported studies are limited by the parameters bounds, problem domains, lack of scalability and reusability. The reason behind these shortcomings is partly because of models complexities and the amount of data they depend on. Also, another reason stems from the weak economic and political support of projects.

DES modelling techniques generally have been used for planning healthcare services, economic modelling and disease intervention. Moreover, SD modelling techniques mainly have been used for policy evaluation, economic modelling, system and infrastructure modelling. Finally, the applications of ABS modelling is not yet widespread, because it is a newly developed methodology, and modelling the agents is highly complex. ABS has the potential to model the quality of care and used to study the scale and granularity of system behaviour. The DES and SD can be considered as a compliment to each other, since DES models look at the detailed level, and SD uses the aggregated level. Therefore, depending on the problems, these two approaches alone can impose some limitations in model development (Gunal and Pidd, 2010, Kanagarajah et al., 2008, Katsaliaki and Mustafee, 2010).

A.1.3 End of Life Care Frameworks and Approaches

There are a number of frameworks (Table A.1 and A.2) that have been introduced to the practices in England that take systematic approaches towards End of Life (EoL) care by utilising information such as administrative triggers, clinical triggers, phases of illness and phases of care (DH, 2008, 2009). Each of the frameworks is particular to a sub-problem with different predictability powers; therefore, they are used in different phases of care by different users. In Table A.1, a list of general EoL frameworks is presented, and the aim of the main approaches are used in these frameworks are listed in Table A.2.

Moreover, the predictive modelling powers of frameworks are limited and are variable across systems. For example, Garson (2012) compared Gold Standards Framework (GSF), which is a palliative care prognostic tool against Global Registry of Acute Coronary Events (GRACE) score to assess their predictive powers. This assessment was carried out for 172 patients with Acute Coronary Syndrome (ACS) with admission period of over eight weeks. The study concluded that GRACE and GSF criteria can identify many of the ACS patients. The main bias of this research is that only a small number of patients was studied with caring physician and also it excluded cardiologists.

Table A.3: Major end of life care frameworks

Framework	Source	Tools/Approaches
NHS End of Life Care Programme (NEoLCP) (Henry and Fenner, 2007, DH, 2008)	NHS (2004 - 2007)	• Preferred Priorities for Care (PPC) • Liverpool Care Path (LCP) • Gold Standards Framework (GSF) • Advance Care Planning (ACP)
National End of Life Care strategy (Henry and Fenner, 2007, DH, 2008, 2013a)	NHS (2008 - 2012)	• Preferred Priorities for Care (PPC) • Liverpool Care Path (LCP) • Gold Standards Framework (GSF) • Advance Care Planning(ACP) • Holistic Needs Assessment • NHS End of Life Care Programme (NEoLCP) Support and Fact Sheets • The Route to Success Series • End of Life Care for All (ELCA) • Guidance for Commissioners
Marie Curie Cancer Care - Delivering Choice Programme (Payne et al., 2008)	Marie Curie Cancer Care (2004)	• Rapid Response Team (RRT) • Discharge Community Link Nurses (DCLNs) • Palliative Care Coordination Centre (PCCC)

TABLE A.4: Major end of life care approaches

Framework/Approach	Usage	timeframe
GSF (Hansford and Meehan, 2005, Lea, 2013, Shaw et al., 2010)	GSF optimises primary palliative care; GSF is a stepwise approach to enhance communication, coordination, control of symptoms, continuity, continued learning, carer support and care of the dying.	6 - 12 month
LCP (Lea, 2013)	LCP provides alerts, guidance and a structured record for doctors, nurses and multidisciplinary team that are inexpert in palliative care.	0 - 14 days
PPC (Curie, 2010, DH, 2011b)	PPC facilitates individual choice in relation to end of life care.	6 month
ACP (Curie, 2010, DH, 2007)	ACP is a process if discussion between an individual and their care provider regardless of discipline.	6 month
RRT (Thomas et al., 2007)	RRTs are designated groups of clinicians that deliver critical care expertise to patients outside of a critical care unit.	6 month
DCLNs (Addicott and Dewar, 2008)	DCLNs facilitates discharge by co-ordinating packages of home care and community healthcare; DCLNs provides support and consultancy to professionals and patients;	6 month
PCCC (Addicott and Dewar, 2008)	PCCC is an administrative centre for packages of care; PCCCs administrates based on the assessments by the district nurses.	6 month
Amber Care Bundle (ACB) (Lea, 2013)	A simple approach used in hospital when there is uncertainty on patient's recovery.	1 week to 9 month

A.1.4 Time To Event Modelling

Survival Tree (ST) (Davis and Anderson, 1989) is traditionally used for modelling time to occurrence of an event, such as time to recovery, stimulus-response time and time to death. ST modelling is used for finding the probability of time-to-event rather than the probability of an event. Also, the time dimension may be replaced by other parameters, such as assessing breaking strength of a concrete block or the level attained in some court trials (Crowder, 2012). Garg et al. (2011) implemented a Phase-type Distribution (PHD) modelling and mixed distribution modelling approach in constructing STs for clustering patients' LoS at hospital. The proposed approaches were tested on stroke-related patients from the HES database. The presented ST models allowed the representation of heterogeneous pathways which is not possible using the traditional ST. Both modelling approaches demonstrated that could improve the likelihood functions of the clustering method. Furthermore, Garg et al. (2009, 2010, 2012) presented discrete and continuous PHD models with non-homogeneous Markov states.

Forsberg et al. (2012) developed a number of BNs referred to as Bayesian-Estimated Tools for Survival (BETS) model. The model is capable of estimating the survival likelihood after surgery based on 84 demographic and clinical attributes with varying missing data. Regarding robustness, the Area Under Curve (AUC) of ROC test was 79% for different patient populations. There are some limitations to the model capability. Firstly, it is limited to patients who were under orthopaedic surgery for their skeletal metastatic disease. Secondly, a homogeneous population was used for training and validation of the model. Finally, the covariates that were used in the modelling are very specific to the case study and the collected data.

A.1.5 Time-Varying Dynamic Bayesian Network

In a predictive modelling problem, such as Length-of-Stay (LoS) or End-of-Life (EoL), the features are mainly inhomogeneous, because the processes in the models are either non-stationary (e.g. length of illness or treatment stages) or the events are sparse (e.g. morbidity conditions or patient states). The unobservable and inhomogeneous properties of models cause the momentum of the system dynamic to change across temporal access.

Generally, it is not statistically tractable to consider all of the variances for every time-point, because of the complexity in the inference and the lack of enough training evidence. Also, it is not possible to segment the time, since the model characteristics are unknown for each segment. There are five main approaches to model a Bayesian Network (BN) that is time-varying Dynamic Bayesian Network (DBN). The approaches are highlighted in the following and the summary of the studies is presented in (Table A.5).

Firstly, a basic indirect approach is to transform time in order to make the process homogeneous. A naive approach is to use a time interpolation technique. Instead of using direct time transformation, a method like Kalman filter can be used, which is a Linear Dynamical System technique and is based on an autoregressive function to estimate a value at a time-point Cook and Lawless (2013), Wang et al. (2009b). Xu et al. Xu et al. (2007) proposed a state space model based on Kalman filter to estimate mean and variance for equally and unequally spaced longitudinal count data with serial correlation. The model applied to Epileptic Seizure and Primary Care Visits Data, and with high-number of observations, the model produced comparable results to those by a numerical approach.

Moreover, another indirect approach is to re-weight the likelihoods at each time-point using a particle-based approach, such as feed-forward and sparse Kalman filtering. Since DBN is a generative model, it often works better for sparse models, because of its assumption about the underlying probabilities. However, it needs to be applied with extreme care, since inappropriate sampling technique can rapidly slide the weights to zero or make the model assumptions and prior probabilities incorrect Koller and Friedman (2009).

Furthermore, another approach is network rewiring which is also known as time-evolving graphical models Robinson and Hartemink (2010). It is a feasible option for large-scale time-varying networks. Time-evolving graphical models have been recently used in designing large networks in biological and social studies Ahmed and Xing

(2009), Guo et al. (2007), Zhou et al. (2008), with the objective to find unobserved network topologies or to rewire network under different conditions (e.g. edge stability and transitivity). For instance, recently a new modelling approach known as temporal Exponential Random Graph Model has been proposed for modelling networks evolving over discrete time-steps with Monte Carlo Markov Chain based or convex optimisation algorithms for posterior inference Ahmed and Xing (2009), Guo et al. (2007), Hanneke et al. (2010).

Another approach is conditional BN modelling, which is also known as multilevel or hierarchical BN model and is popular in the literature. In BN modelling, a time-varying framework on top of a Markov Chain (MC) technique can be used to model multilevel time properties. A Cox phase-type model is used for modelling durations on top of a Hidden Semi-Markov Model by Duong et al. (2009) for human activity recognition. In this research, an extension added to the Coxian Hidden Semi-Markov Model, which incorporates both duration and hierarchical modelling. Also, a DBN framework is proposed by Lappenschaar et al. Lappenschaar et al. (2013b) for modelling non-stationary events in multi-morbidity modelling. This PM of the interactions between heart failure and diabetes mellitus could closely resemble the PM techniques which use multilevel Linear Mixed Model. Moreover, a framework is designed by Lappenschaar et al. Lappenschaar et al. (2013a) for formulating a Linear Mixed Model into a BN using a Logistic Regression function.

Finally, the Linear Dynamical Systems are useful temporal models, which represent one or more real-valued variables that evolve linearly over time, with some Gaussian noise Koller and Friedman (2009). There are two categories of the Linear Dynamical Systems methods in the modelling of time-varying DBN: Switching Linear Dynamic System and Time-Varying Autoregression.

Switching Linear Dynamic Systems have been studied extensively for piecewise modelling of linear systems Wang et al. (2011). Based on the Switching Linear Dynamic Systems modelling, time-varying observations Ghahramani and Hinton (2000) and time-varying duration Blake et al. (1999), Pavlovic et al. (2000) can be formulated using a latent MC. But, the MC method which is a piecewise stationary, does not have a very general application in learning and inference, and a time-varying linear regression can be used instead. For instance, a time-varying DBN has been introduced by Song et al. Song et al. (2009), which aggregates observations of adjacent time points by a kernel re-weighting function.

Time-Varying Autoregression models are another type of the Linear Dynamic Systems Wang et al. (2011) models, which focus on non-stationary models with a fixed structure.

TABLE A.5: The summary of the studies in Time-Varying Dynamic Bayesian Network area

Approach	Method(s)	Study	Domain	Findings/Outcomes
Time transformation	Autoregression and Kalman filter	Xu et al. (2007)	Healthcare events	The likelihood estimation approach performs better than the numerical integration approach
Time Evolving Graphical Network	Prevailing networks	Robinson and Hartemink (2010)	Generic non-stationary data	Demonstrating the feasibility
	Scalable inference for time-evolving networks	Ahmed and Xing (2009)	Biological systems to social science	Having asymptotically value-consistent under fixed model dimension
Conditional BN Modelling	A multilevel BN	Lappenschaar et al. (2013a)	Multimorbidity condition prediction	Providing more insight into interaction of multiple diseases
	A multilevel BN	Lappenschaar et al. (2013b)	The course of a medical condition	An informative clinical decision making tool
	Semi-Markov model, Coxian and HMM	Duong et al. (2009)	Recognition of human activities of daily living	Having high and comparable accuracy
Switching Linear Dynamic Systems	Using hidden variables for network changes	Wang et al. (2011)	Camera tracking	Being successful for both simulated non-stationary data and video sequences
Time Varying Autoregression	Kalman filter	Johnson and Sakoulis (2008)	Equity market prediction	Performing as well as the Capital Asset Pricing Model benchmark, despite of using non-traditional pricing measures

Time-Varying Autoregression models have been applied to a wide range of research applications, such as predictive modelling of equity market Johnson and Sakoulis (2008), inferring time-varying data from gene expression Perrin et al. (2003), Rao et al. (2007) and modelling non-Gaussian autoregression Gencaga et al. (2010).

A.2 Data Sources

TABLE A.6: Potential external data sources

Data Source	Org.	Start Year	Features
Admitted Patient Care (APC): Monthly statistics (HSCIC, 2016f,h)	HSCIC, ONS	2002	A wide range of summary information about patients admitted to NHS hospitals.
National Reporting and Learning System (NRLS): Patient safety (HSCIC, 2016m, NHS, 2016f, NRLS, 2010b, 2014)	NRLS	2003	The organisation level data for each NHS organisation on patient safety incidents in England and Wales, which are grouped by cluster. Most incidents are submitted electronically from local risk management systems.
Avoidable Mortality (ONS, 2014a)	ONS	2001	Statistics on avoidable mortality (i.e. deaths caused by certain conditions that should not occur in the presence effective health care or health interventions).
Hospital and community healthcare staff (ONS, 2013)	ONS	2001	Estimates of the usually resident population and changes made to the population estimates over time.
Quality and Outcomes Framework (QOF): Recorded prevalence (Fund, 2011, HSCIC, 2014c, 2016l, NHS, 2016a)	NHS	2004	A voluntary annual reward and incentive programme for all GP surgeries in England, detailing practice achievement, exceptions and recorded prevalence.

TABLE A.7: Impractical external data sources

Data Source	Organisation	Reason
Patient Safety (NHS, 2016f)	NRLS	Slowly adaptation; different reporting practices (NRLS, 2010a)
Number of GP per demographic groups (HSCIC, 2016g)	HSCIC	Very small correlation to readmission (Laudicella et al., 2013)
Distance of Patient from the hospital (HSCIC, 2016h)	HSCIC	Very small correlation to readmission (Laudicella et al., 2013)

A.3 Software Tools

TABLE A.8: Considered software tools for Bayesian modelling

Tool	Developers	License	Source	Language	Interface
AgenaRisk (Agena, 2016)	Agena	Commercial	Closed	Java	GUI [2]
Bayesian Network Tool BNT for MATLAB (Mathworks, 2016, Murphy, 2016)	Kevin Murph	GNU GPL [3]	Opened	Matlab	Matlab
Infer.Net (Research, 2016)	Microsoft Research	Non-Commercial	Opened	C#	C#, F#
Bayes Server (Server, 2016)	Bayes Server Ltd	Commercial	Closed	C#	GUI, C#
SAMIAM (UCLA, 2016)	Automated Reasoning Group, UCLA	GNU	Closed	JAVA	GUI
Bayesian Network tools in Java (BNJ) (Hsu, 2016)	Laboratory for Knowledge Discovery in Databases, Kansas State University	GNU GPL	Closed	Java	Java
GeNIe & SMILE (DSL, 2016)	Decision Systems Laboratory, University of Pittsburgh	As-Is	Closed	C++	GUI, C++
OpenBUGS (Fund, 2016b, Unit, 2016)	OpenBUGS community	GNU GPL	Opened	Component Pascal	GUI
GMTK (Bilmes, 2006, of Washington, 2016)	University of Washington	OSL [4]	Opened	C++	GMTKL, a simple textual language

A.4 Features

A.4.1 Main Categories of Correlated Variables

TABLE A.9: Main categories of identified Features

Category	Examples
Case mix group of diagnoses	Patterns of Morbidity with Adjusted Clinical Groups (ACGs), chronic condition with Expanded Diagnosis Clusters (EDCs) (JHU, 2014), Agency for Healthcare Research & Quality's (AHRQ) diagnosis categorisation scheme (AHRQ, 2016a), ACS conditions (AHRQ, 2001, NHS, 2016b), frequent comorbidities & chronic conditions.
Case mix group of operations & procedures	National Clinical Coding Standards (OPCS) classification (HSCIC, 2016j), & AHRQ's procedure categorisation (AHRQ, 2016a).
Clinical indicators, treatments, medications & compliance	Lab test results, & prescribed medications.
Demographics	Age, race, gender, living arrangements, level of education & marital status.
Deprivations	Index of Multiple Deprivation (IMD) (DCLG, 2012), which includes: income, employment, health & disability, education, skills & training, barriers to housing & services, living environment, & crime.
Geographical location of patient & care provider	Area of residence, Type & location of hospital (HSCIC, 2016h, NRLS, 2014) & population estimates of local authorities (ONS, 2016).
Insurance & medical claims	Grouping clinically similar treatments with Healthcare Resource Groups (HRGs) (HSCIC, 2016a).
Legal status	Formally detained or subject to guardianship under mental health act.
Physical condition	Functional physical activities, specific comorbidities & general health.
Psychological health, emotional state & social functioning	Health of the Nation Outcome Scales (HoNOS). ADLs, Instrumental ADLs, Social isolation, loneliness, presence of aggressive, disturbed or psychotic behaviour, physical activities, life satisfaction, anxiety & depression, quality of life, self-reported health status, & general mental health.
Risk indices	A version of Charlson comorbidity index (Charlson et al., 1987), Elixhauser comorbidity index (Elixhauser et al., 1998, AHRQ, 2016b), & Bupa Operative Severity Score (Bupa, 2016).
Social care status	Skilled nursing facility, rehab, hospice & palliative care.
Times, types, consultations, sources, waiting & length-of-stays for admissions or discharges	Using clinically homogeneous units that describe complete episodes of care using Optum Episode Treatment Groups (ETGs) (OPTUM, 2014, 2016a), the number of emergency admissions in different time frames, types and the number of specialities, & Previous hospitalisation.

A.4.2 Main Features

TABLE A.10: ERMER: Definition of all the main features

Id.	Group	Category	Sub-category	Feature Name	Definition
1	Admin.	Admission	Classification	classpatRecoded_freq	The recoded HES's classpat (patient classification) of an episode. [a]
2	Admin.	Admission	Count	episodeAdmission_freq	The number of episodes of a patient, within the selected timeframe.
3	Admin.	Admission	Count	observations	The number of episodes of a patient.
4	Admin.	Admission	Count	spell	The spell number of a patient.
5	Admin.	Admission	Count	spellAdmission_freq	The number of spells of a patient, within the selected timeframe.
6	Admin.	Admission	Date	admidate	The HES's admidate (date of admission) of a spell. [a]
7	Admin.	Admission	Date	firstAdmidate	The first HES's admidate of a patient. [a]
8	Admin.	Admission	Date	lastAdmidate	The last HES's admidate of a patient. [a]
9	Admin.	Admission	Discharge Method	dismeth	The HES's dismeth (method of discharge) of a spell. [a]
10	Admin.	Admission	Discharge Method	dismeth_deadAlive	The hospital death based on the HES's dismeth value of a spell. [a]
11	Admin.	Admission	General	readmiGap	The gap between the current spell's admission to the next spell.
12	Admin.	Admission	Method	admimethRecoded_freq	The recoded HES's admimeth (method of admission) of an episode. [a]
13	Admin.	Admission	Source	admisorcRecoded_freq	The recoded HES's admisorc (source of admission) of an episode. [a]
14	Admin.	Admission	Source	intmanigRecoded	The recoded HES's intmanig (intended management) of an episode. [a]
15	Admin.	Bed Days	Count	epidur	The HES's epidur (episode duration) of a spell. [a]
16	Admin.	Bed Days	Count	epidurRecoded	The recoded HES's epidur of a spell. [a]
17	Admin.	Bed Days	Operation	posopdur	The HES's posopdur (post-operative duration) of a spell. [a]
18	Admin.	Bed Days	Operation	preopdur	The HES's preopdur (pre-operative duration) of a spell. [a]
19	Admin.	Geographical	Provider	procode3	The first three characters of the HES's procode (Provider code) of an episode. The procode is managed by the National Administrative Codes Service. [a]
20	Admin.	Geographical	Provider	rotreatRecoded	The recoded HES's rotreat of an episode (based on the coding from 1996 to present). [a]
21	Admin.	Hospital	Provider	orgCluster	The organisation cluster of an episode, based on the NRLS classification of procode3. [c]
22	Admin.	Hospital	Provider	protype	The HES's protype (provider type) of an episode. [a]
23	Admin.	ID	Patient	hesid	The HES's hesid (patient identifier) of an episode. [a]
24	Admin.	ID	Time	timeframe	The month of the year of an episode, based on the HES's admidate. [a]
25	Admin.	Speciality	General	mainspef	The list of HES's mainspef_nn (main speciality) of an episode. [a]
26	Admin.	Speciality	General	mainspefRecoded	The list of recoded HES's mainspef_nn (main speciality) of an episode (refer to its corresponding table for the mapping). [a]
27	Admin.	Speciality	Palliative	palliativeCare	The palliative care flag of an episode, based on the diagnosis code. [f]
28	Admin.	Speciality	Palliative	palliativeMedicine	The palliative medicine flag of an episode, based on the treatment function code. [f]
29	Admin.	Waiting Time	General	elecdur	The HES's elecdur (waiting time) of an episode. [a]
30	Clinical	Diagnosis	Diagnosis	charlsonIndex	The CCI of a spell, based on HSCIC CCI. [def]
31	Clinical	Diagnosis	Diagnosis	diag2Diag20	The list of secondary diagnosis codes of an episode, based on the HES's diag_nn (diagnosis codes). [a]
32	Clinical	Diagnosis	Diagnosis	mainDiag	The AHRQ CCS category of the spell's primary diagnosis based on the HES's diag_nn & the guideline provided by the HSCIC. [ae]

Continued on next page

Id.	Group	Category	Sub-category	Feature Name	Definition
33	Clinical	Diagnosis	Diagnosis	shmiIndex	The SHMI index of a spell. [def]
34	Clinical	Diagnosis	Blood & Blood-Forming Organs & Immune Mechanism	diagCci_39_coagulopathy_freq	The count of coagulopathy conditions. [i]
35	Clinical	Diagnosis	Blood & Blood-Forming Organs & Immune Mechanism	diagCci_43_bloodLoss_freq	The count of blood loss anemia conditions. [i]
36	Clinical	Diagnosis	Blood & Blood-Forming Organs & Immune Mechanism	diagCci_44_anemia_freq	The count of deficiency anemia conditions. [i]
37	Clinical	Diagnosis	Blood & Blood-Forming Organs & Immune Mechanism	diagOther_3_blood_freq	The count of thrombocytopenia & thrombocytosis & elevated white blood cell count conditions. [i]
38	Clinical	Diagnosis	Blood & Blood-Forming Organs & Immune Mechanism	diagOther_4_chronic_g_freq	The count of blood (ACSC category 'g') [g]conditions. [i]
39	Clinical	Diagnosis	Blood & Blood-Forming Organs & Immune Mechanism	diagOther_6_diagSickle_freq	The count of Sickle cell conditions. [i]
40	Clinical	Diagnosis	Circulatory System	diagCat_1_freq	The count of rheumatic fever or rheumatic heart conditions. [i]
41	Clinical	Diagnosis	Circulatory System	diagCat_10_freq	The count of other unspecified disorders of the circulatory system. [i]
42	Clinical	Diagnosis	Circulatory System	diagCat_11_freq	The count of congenital cardiovascular defects
43	Clinical	Diagnosis	Circulatory System	diagCat_2_freq	The count of hypertensive conditions. [i]
44	Clinical	Diagnosis	Circulatory System	diagCat_3_freq	The count of ischemic (coronary) heart conditions. [i]
45	Clinical	Diagnosis	Circulatory System	diagCat_4_freq	The count of pulmonary heart & pulmonary circulation conditions. [i]
46	Clinical	Diagnosis	Circulatory System	diagCat_5_freq	The count of other forms of heart conditions. [i]
47	Clinical	Diagnosis	Circulatory System	diagCat_6_freq	The count of cerebrovascular (stroke) conditions. [i]
48	Clinical	Diagnosis	Circulatory System	diagCat_7_freq	The count of atherosclerosis conditions. [i]
49	Clinical	Diagnosis	Circulatory System	diagCat_8_freq	The count of other arteries, arterioles & capillaries conditions. [i]
50	Clinical	Diagnosis	Circulatory System	diagCat_9_freq	The count of other veins, lymphatics & lymph nodes conditions. [i]
51	Clinical	Diagnosis	Circulatory System	diagCci_01_myocardial_freq	The count of myocardial infarction conditions. [i]
52	Clinical	Diagnosis	Circulatory System	diagCci_02_chf_freq	The count of congestive heart failure conditions. [i]
53	Clinical	Diagnosis	Circulatory System	diagCci_03_pvd_freq	The count of peripheral vascular conditions. [i]
54	Clinical	Diagnosis	Circulatory System	diagCci_13_renal_freq	The count of renal conditions. [i]
55	Clinical	Diagnosis	Circulatory System	diagCci_15_liverSevere_freq	The count of moderate or severe liver conditions. [i]
56	Clinical	Diagnosis	Circulatory System	diagCci_19_cardiac_freq	The count of cardiac arrhythmias conditions. [i]
57	Clinical	Diagnosis	Circulatory System	diagCci_21_pulmonary_freq	The count of pulmonary circulation disorders. [i]
58	Clinical	Diagnosis	Circulatory System	diagCci_22_vascular_freq	The count of peripheral vascular disorders. [i]
59	Clinical	Diagnosis	Circulatory System	diagCci_23_hypertensionNotComplicated_freq	The count of hypertension with no complication conditions. [i]
60	Clinical	Diagnosis	Circulatory System	diagCci_27_pulmonaryChronic_freq	The count of chronic pulmonary conditions. [i]
61	Clinical	Diagnosis	Circulatory System	diagCci_31_renal_freq	The count of renal failure conditions. [i]
62	Clinical	Diagnosis	Circulatory System	diagOther_4_chronic_c_freq	The count of cardiovascular (ACSC category 'c') [g]conditions. [i]
63	Clinical	Diagnosis	Circulatory System	diagOther_4_chronic_f_freq	The count of hypertensive heart & renal with heart failure (ACSC category 'f') [g]conditions. [i]
64	Clinical	Diagnosis	Circulatory System	diagOther_4_chronic_h_freq	The count of cardiovascular (ACSC category 'h') [g]conditions. [i]
65	Clinical	Diagnosis	Circulatory System	diagRisk_3_blood_extra_freq	The count of hypertension conditions. [i]
66	Clinical	Diagnosis	Circulatory System	diagRisk_3_blood_freq	The count of pulmonary hypertension, chronic venous hypertension, post procedural hypertension & ocular hypertension conditions. [i]
67	Clinical	Diagnosis	Digestive System	diagCci_08_ulcer_freq	The count of peptic ulcer conditions. [i]

Continued on next page

Id.	Group	Category	Sub-category	Feature Name	Definition
68	Clinical	Diagnosis	Digestive System	diagCci_09_liverMild_freq	The count of mild liver conditions. [i]
69	Clinical	Diagnosis	Digestive System	diagCci_17_aids_freq	The count of AIDS/HIV conditions. [i]
70	Clinical	Diagnosis	Digestive System	diagCci_32_liver_freq	The count of liver conditions. [i]
71	Clinical	Diagnosis	Digestive System	diagCci_33_ulcerNotBleeding_freq	The count of peptic ulcer with no bleeding conditions. [i]
72	Clinical	Diagnosis	Digestive System	diagCci_45_alcohol_freq	The count of alcohol abuse conditions. [i]
73	Clinical	Diagnosis	Digestive System	diagMorbid_8_periodontitis_freq	The count of periodontitis conditions. [i]
74	Clinical	Diagnosis	Digestive System	diagRisk_7_kidney_freq	The count of liver conditions. [i]
75	Clinical	Diagnosis	Digestive System	diagOther_4_chronic_a_freq	The count of Hepatitis B infections (ACSC category 'a') [g] [i]
76	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagCci_10_diabetesNotChronic_freq	The count of diabetes with no chronic complication conditions. [i]
77	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagCci_11_diabetesChronic_freq	The count of diabetes with chronic complication conditions. [i]
78	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagCci_28_diabetesNotComplicated_freq	The count of diabetes with no complicated conditions. [i]
79	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagCci_29_diabetesComplicated_freq	The count of diabetes with complication conditions. [i]
80	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagCci_30_hypothyroidism_freq	The count of hypothyroidism conditions. [i]
81	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagCci_40_obesity_freq	The count of overweight & obesity conditions. [i]
82	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagCci_41_weightLoss_freq	The count of weight loss conditions. [i]
83	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagCci_42_fluidDisorder_freq	The count of fluid & electrolyte disorders. [i]
84	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagOther_1_ulcers_freq	The count of diabetic ulcer conditions. [i]
85	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagOther_4_chronic_d_freq	The count of diabetes conditions (ACSC category 'd') [g] [i]
86	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagRisk_2_Cholesterol_freq	The count of disorders of lipidemias conditions. [i]
87	Clinical	Diagnosis	Endocrine, Nutritional & Metabolic	diagRisk_6_Metabolic_freq	The count of metabolic syndrome conditions. [i]
88	Clinical	Diagnosis	External Causes of Morbidity	diagRisk_10_externalMorbidity_freq	The count of external causes of morbidity conditions. [i]
89	Clinical	Diagnosis	Genitourinary System	diagMorbid_2_kidney_freq	The count of kidney conditions. [i]
90	Clinical	Diagnosis	Genitourinary System	diagMorbid_4_erectile_freq	The count of male erectile dysfunction conditions. [i]
91	Clinical	Diagnosis	Infectious & Parasitic	diagCci_20_valvular_freq	The count of valvular conditions. [i]
92	Clinical	Diagnosis	Injury, Poisoning & External Causes	diagMorbid_10_vascularOper_freq	The count of heart or lungs transplant status & aftercare. [i]
93	Clinical	Diagnosis	Injury, Poisoning & External Causes	diagRisk_5_diabetes_extra_freq	The count of complications due to insulin pump malfunction & secondary diabetes mellitus due to pancreatectomy. [i]
94	Clinical	Diagnosis	Injury, Poisoning & External Causes	diagRisk_9_external_freq	The count of injury, poisoning & certain other consequences of external causes or complications, which not elsewhere classified. [i]
95	Clinical	Diagnosis	Mental, Behavioral & Neurodevelopmental	diagCci_05_dementia_freq	The count of dementia conditions. [i]
96	Clinical	Diagnosis	Mental, Behavioral & Neurodevelopmental	diagCci_18_depression_freq	The count of depression conditions. [i]
97	Clinical	Diagnosis	Mental, Behavioral & Neurodevelopmental	diagCci_34_psychoses_freq	The count of psychoses conditions. [i]
98	Clinical	Diagnosis	Mental, Behavioral & Neurodevelopmental	diagCci_46_drug_freq	The count of drug abuse conditions. [i]
99	Clinical	Diagnosis	Mental, Behavioral & Neurodevelopmental	diagOther_5_alcohol_freq	The count of alcohol conditions. [i]
100	Clinical	Diagnosis	Mental, Behavioral & Neurodevelopmental	diagOther_8_mental_freq	The count of mental conditions. [i]

Continued on next page

Id.	Group	Category	Sub-category	Feature Name	Definition
101	Clinical	Diagnosis	Musculoskeletal System & Connective Tissue	diagCci_07_rheumatic_freq	The count of rheumatic conditions. [i]
102	Clinical	Diagnosis	Musculoskeletal System & Connective Tissue	diagMorbid_6_rheumatoid_freq	The count of rheumatoid arthritis conditions. [i]
103	Clinical	Diagnosis	Neoplasms	diagCci_14_malignancy_freq	The count of malignancy conditions, including lymphoma & leukemia, except malignant neoplasm of skin. [i]
104	Clinical	Diagnosis	Neoplasms	diagCci_16_tumorSec_freq	The count of metastatic solid tumor conditions. [i]
105	Clinical	Diagnosis	Neoplasms	diagCci_35_lymphoma_freq	The count of lymphoma conditions. [i]
106	Clinical	Diagnosis	Neoplasms	diagCci_36_cancerSec_freq	The count of metastatic cancer conditions. [i]
107	Clinical	Diagnosis	Neoplasms	diagCci_37_tumorNotSec_freq	The count of solid tumour without metastasis conditions. [i]
108	Clinical	Diagnosis	Neoplasms	diagOther_7_cancer_freq	The count of neoplasm conditions. [i]
109	Clinical	Diagnosis	Nervous System	diagCci_04_cerebrovascular_freq	The count of cerebrovascular conditions. [i]
110	Clinical	Diagnosis	Nervous System	diagCci_12_hemiplegia_freq	The count of hemiplegia or paraplegia conditions. [i]
111	Clinical	Diagnosis	Nervous System	diagCci_25_paralysis_freq	The count of paralysis conditions. [i]
112	Clinical	Diagnosis	Nervous System	diagCci_26_neuroOther_freq	The count of other neurological disorders. [i]
113	Clinical	Diagnosis	Nervous System	diagMorbid_3_sleep_freq	The count of obstructive sleep apnea
114	Clinical	Diagnosis	Nervous System	diagOther_4_chronic_i_freq	The count of mental & behavioural disorders or neurological disorders (ACSC category 'i') [g] [i]
115	Clinical	Diagnosis	Not Elsewhere Classified	diagRisk_4_Glucose_freq	The count of elevated blood glucose level conditions. [i]
116	Clinical	Diagnosis	Other Factors	diagMorbid_9_vascularRadi_freq	The count of radiation exposure conditions. [i]
117	Clinical	Diagnosis	Other Factors	diagRisk_8_smoke_freq	The count of exposure to tobacco smoke conditions. [i]
118	Clinical	Diagnosis	Palliative	palliativeCare_freq	The count of radiotherapy conditions. [i]
119	Clinical	Diagnosis	Reference	reference_freq	The count of the reference [h]conditions in the HRG record.
120	Clinical	Diagnosis	Respiratory System	diagCci_06_cpd_freq	The count of chronic pulmonary conditions. [i]
121	Clinical	Diagnosis	Respiratory System	diagMorbid_1_Influenza_freq	The count of influenza a pneumonia conditions. [i]
122	Clinical	Diagnosis	Respiratory System	diagOther_4_chronic_b_freq	The count of respiratory conditions (ACSC category 'b') [g] [i]
123	Clinical	Diagnosis	Respiratory System	diagOther_4_chronic_e_freq	The count of respiratory conditions (ACSC category 'e') [g] [i]
124	Clinical	Diagnosis	Skin & Subcutaneous	diagCci_38_rheumatoid_freq	The count of collagen vascular or rheumatoid arthritis conditions. [b]
125	Clinical	Diagnosis	Skin and Subcutaneous	diagMorbid_5_psoriasis_freq	The count of psoriasis conditions. [i]
126	Clinical	Diagnosis	Skin & Subcutaneous	diagMorbid_7_lupus_freq	The count of lupus erythematosus conditions. [i]
127	Clinical	Operation	Count	episodeOpertn_freq	The number of operations of an episode, based on the HES's opertn (operation codes). [a]
128	Clinical	Operation	Count	spellOpertn_freq	The number of operations of a spell, based on the HES's opertn (operation codes). [a]
129	Clinical	Operation	Circulatory System	oper_2_heart_freq	The count of heart operations. [i]
130	Clinical	Operation	Endocrine, Nutritional & Metabolic	oper_1_obesity_freq	The count of obesity operations. [i]
131	Clinical	Operation	Genitourinary System	oper_3_urinary_freq	The count of urinary operations. [i]
132	Clinical	Operation	Other Factors	oper_4_radio_freq	The count of radiotherapy related operations. [i]
133	patient	Demographic	Age	triggerStartAge	The HES's startage (age at start of episode) of a patient at the trigger year. [a]
134	patient	Demographic	Age	triggerStartAgeRecoded	The HES's startage (age at start of episode) of a patient at the trigger year. [a]

Continued on next page

Id.	Group	Category	Sub-category	Feature Name	Definition
135	patient	Demographic	Deprivation	imd04rk	The HES's imd04k (the Index of Multiple Deprivation Overall Rank) of a patient. [a]
136	patient	Demographic	Deprivation	imd04rkRecoded	The recoded HES's imd04k (the Index of Multiple Deprivation Overall Rank) of a patient. [a]
137	patient	Demographic	Ethnicity	ethnos	The HES's ethnos (ethnic category) of a patient. [a]
138	patient	Demographic	Ethnicity	ethnosRecoded	The recoded HES's ethnos (ethnic category) of a patient. [a]
139	patient	Demographic	Gender	genderRecoded	The recoded HES's sex (sex of patient) of a patient. [a]

[a] It is based on the HSCIC definition of the variable. [b] It is based on the ICD-10 coding system. [c] It is based on the organisation clustering provided by the National Reporting & Learning System (NRLS). [d] It is based on the HSCIC version of the Charlson Comorbidity Index (CCI). [e] It is based on the Summary Hospital-level Mortality Indicator (SHMI) documentation by the Health & Social Care Information Centre(HSCIC). [f] It is based on the Hospital Standardised Mortality Ratio (HSMR) mortality indicators by the Dr. Foster Intelligence [g] It is based on the chronic Ambulatory Care Sensitive Conditions (ACSC) categorisation, which provided by Clinical Commissioning Group (CCG). [h] It is based on the reference Healthcare Resource Groups (HRG) conditions that are specified in the PARR modellings. [i] It is based on the OPCS-4 coding system.

A.4.3 Features Calculations

TABLE A.11: Mathematical definition of the main features included in the feature pool

Formula	Sub-Features
1. classpatRecoded_freq $IF(classpat_{episode} \in \{NULL, 9\})\ THEN$ $classpatRecoded\_freq_{episode} = "NA"$ $ELSE\ IF(classpat_{episode} \in \{1, 5, 8\})\ THEN$ $classpatRecoded\_freq_{episode} = "Ordinary"$ $ELSE\ IF(classpat_{episode} \in \{2, 3, 4\})\ THEN$ $classpatRecoded\_freq_{episode} = "Regular"\ ENDIF$	NA
2. episodeAdmission_freq $episodeAdmission\_freq_{patient} = COUNT_{episode}(episode_{patient}), WHERE$ $TIME_{start} \leq episode_{time} \leq TIME_{end}$	NA
3. observations $episodes_{patient} = COUNT_{episode}(episode_{spell})$	NA
4. spell $spell_{patient}$	NA
5. spellAdmission_freq $spellAdmission\_freq_{patient} = COUNT_{spell}(episode_{patient}), WHERE$ $TIME_{start} \leq spell_{time} \leq TIME_{end}$	NA
6. admidate $admidate_{spell}$	NA
7. firstAdmidate $firstAdmidate_{patient} = MIN(admidate_{spell})$	NA
8. lastAdmidate $lastAdmidate_{patient} = MAX(admidate_{spell})$	NA
9. dismeth $dismeth_{spell}$	NA
10. dismeth_deadAlive $IF(dismeth_{spell} == 4)\ THEN$ $hospitalDeath_{spell} = TRUE$ $ELSE\ hospitalDeath_{spell} = FALSE\ ENDIF$	NA
11. readmiGap $readmiGap_{spell} = COUNT_{day}(admidate_{spell_1} - dismeth_{spell_{i-1}})$	NA
12. admimethRecoded_freq $IF(admimeth_{episode} = NULL)\ THEN$ $admimethRecoded\_freq_{episode} = "NA"$ $ELSE\ IF(admimeth_{episode} \in \{11, 12, 13\})\ THEN$ $admimethRecoded\_freq_{episode} = "Elective"$ $ELSE\ IF(admimeth_{episode} = 99)\ THEN$ $admimethRecoded\_freq_{episode} = "Unknown"$ $ELSE\ IF(admimeth_{episode} \in \{21, 22, 23, 24, 25, "2A", "2B", "2C", "2D", 28, 31,$ $32, 81, 82, 83, 84, 89, 98\})\ THEN$ $admimethRecoded\_freq_{episode} = "Acute"\ ENDIF$	NA
13. admisorcRecoded_freq	

Continued on next page

Formula	Sub-Features
$IF(admisorc_{episode} \in \{NULL, 98, 99\})\ THEN$	NA
$admisorcRecoded\_freq_{episode} = "NA"$	
$ELSE\ IF(admisorc_{episode} \in \{19, 29\})\ THEN$	
$admisorcRecoded\_freq_{episode} = "Residential"$	
$ELSE\ IF(admisorc_{episode} \in \{54, 65, 66, 69, 85, 86, 88\})\ THEN$	
$admisorcRecoded\_freq_{episode} = "Transferred\ from\ Residential\ Care"$	
$ELSE\ IF(admisorc_{episode} \in \{30, 37, 38, 39, 48, 49, 50, 51, 53, 87, 89\})\ THEN$	
$admisorcRecoded\_freq_{episode} = "Transferred\ from\ Others"$	
$ELSE\ IF(admisorc_{episode} \in \{79, 52\})\ THEN$	
$admisorcRecoded\_freq_{episode} = "Maternity"\ ENDIF$	
14. intmanigRecoded	
$IF(intmanig_{episode} \in \{NULL, 9\})\ THEN$	NA
$intmanigRecoded_{episode} = "NA"$	
$ELSE\ IF(intmanig_{episode} \in \{1, 2, 5, 8\})\ THEN$	
$intmanigRecoded_{episode} = "Ordinary"$	
$ELSE\ IF(intmanig_{episode} \in \{3, 4\})\ THEN$	
$intmanigRecoded_{episode} = "Regular"\ ENDIF$	
15. epidur	
$epidur_{spell}$	$epidur\_avgAvg_{patient} = AVERAGE(AVERAGE(epidur_{episode})_{spell})$ $epidur\_maxAvg_{patient} = MAX(AVERAGE(epidur_{episode})_{spell})$ $epidur\_maxStdev_{patient} = MAX(STD\_DEV(epidur_{episode})_{spell})$
16. epidurRecoded	
$IF(epidur_{spell} \in \{NULL, -1\})\ THEN$	$epidurRecoded\_avgAvg_{patient} = AVERAGE(AVERAGE(epidurRecoded_{episode})_{spell})$ $epidurRecoded\_maxAvg_{patient} = MAX(AVERAGE(epidurRecoded_{episode})_{spell})$ $epidurRecoded\_maxStdev_{patient} = MAX(STD\_DEV(epidurRecoded_{episode})_{spell})$ $ENDIF$
$epidurRecoded_{spell} = "NA"$	
$ELSE\ IF(epidur_{spell} = 0)\ THEN$	
$epidurRecoded_{spell} = 0$	
$ELSE\ IF(epidur_{spell} = 1)\ THEN$	
$epidurRecoded_{spell} = 1$	
$ELSE\ IF(epidur_{spell} = 2)\ THEN$	
$epidurRecoded_{spell} = 2$	
$ELSE\ IF(epidur_{spell} = 3)\ THEN$	
$epidurRecoded_{spell} = 3$	
$ELSE\ IF(epidur_{spell} = 4)\ THEN$	
$epidurRecoded_{spell} = 4$	
$ELSE\ IF(epidur_{spell} = 5)\ THEN$	
$epidurRecoded_{spell} = 5$	
$ELSE\ IF(epidur_{spell} = 6)\ THEN$	
$epidurRecoded_{spell} = 6$	
$ELSE\ IF(epidur_{spell} = 7)\ THEN$	
$epidurRecoded_{spell} = 7$	
$ELSE\ IF(7 < epidur_{spell} \leq 30)\ THEN$	
$epidurRecoded_{spell} = 8$	
$ELSE\ IF(30 < epidur_{spell} \leq 180)\ THEN$	
$epidurRecoded_{spell} = 9$	
$ELSE\ IF(epidur_{spell} > 180)\ THEN$	
$epidurRecoded_{spell} = 10$	
17. posopdur	

Continued on next page

Formula	Sub-Features
$posopdur_{spell}$	$posopdur\_avg_{patient} = AVERAGE(posopdur_{spell})$ $posopdur\_stdev_{patient} = STD\_DEV(posopdur_{spell})$
18. preopdur $preopdur_{spell}$	$preopdur\_avg_{patient} = AVERAGE(preopdur_{spell})$ $preopdur\_stdev_{patient} = STD\_DEV(preopdur_{spell})$
19. procode3 $procode3_{episode}$	$procode3\_freq1_{patient} = FRQUENT(procode3_{episode})$
20. rotreatRecoded $IF(rotreat_{episode} \in NULL, Y)\ THEN$ $rotreatRecoded_{episode} = "NA"$ $ELSE\ IF(rotreat_{episode} = "Y01")\ THEN$ $rotreatRecoded_{episode} = "Northern\ \wedge\ Yorkshire"$ $ELSE\ IF(rotreat_{episode} = "Y02")\ THEN$ $rotreatRecoded_{episode} = "Trent"$ $ELSE\ IF(rotreat_{episode} = "Y07")\ THEN$ $rotreatRecoded_{episode} = "West\ Midlands"$ $ELSE\ IF(rotreat_{episode} = "Y08")\ THEN$ $rotreatRecoded_{episode} = "North\ West"$ $ELSE\ IF(rotreat_{episode} = "Y09")\ THEN$ $rotreatRecoded_{episode} = "Eastern"$ $ELSE\ IF(rotreat_{episode} = "Y10")\ THEN$ $rotreatRecoded_{episode} = "London"$ $ELSE\ IF(rotreat_{episode} = "Y11")\ THEN$ $rotreatRecoded_{episode} = "South\ East"$ $ELSE\ IF(rotreat_{episode} \in \{"Y12", "Y06"\})\ THEN$ $rotreatRecoded_{episode} = "South\ West\ \wedge\ South\ \wedge\ West\ (old\ coding)"$ $ELSE\ IF(rotreat_{episode} \in \{"Y03", "Y04", "Y05"\})\ THEN$ $rotreatRecoded_{episode} = "Others"$	$rotreatRecoded\_freq1_{patient} = FRQUENT(rotreatRecoded_{episode})$ $ENDIF$
21. episodeOpertn_freq $opertn_{episode} = opertn_{episode}$	NA
22. spellOpertn_freq $opertn_{spell} = \sum_{episode_{spell}} opertn_{episode}$	NA
23. orgCluster $orgCluster_{episode} =$ $COUNT_{cluster}(DICTIONARY_{NRLS\_Cluster}(procode3, cluster)),$ $WHERE\ cluster \in \{"NULL", "Acute\ teaching\ trust",$ $"Acute\ specialist\ trust\ (including\ acute\ specialist\ (children))",$ $"Large\ acute\ trust", "Medium\ acute\ trust",$ $"Small\ acute\ trust", "Ambulance", "Mental\ health",$ $"PCO:\ No\ inpatient\ provision", "PCO:\ Inpatient\ provision"\}$	$orgCluster\_freq1_{patient} = FREQUENT(orgCluster_{episode})$
24. protype $protype_{episode}$	$protype\_freq1_{patient} = FREQUENT(protype_{episode})$
25. hesid $hesid_{episode}$	NA
26. timeframe	

Continued on next page

Formula	Sub-Features
$timeframe_{episode} = DATE_{month}(admidate_{episode})$	$timeframeTrigger_{patient} = timeframe_{trigger}$
27. mainspef	
$mainspef_{nn_{episode}}$	$mainspef\_freq1_{patient} = FREQUENT(mainspef_{episode})$ $mainspef\_uniques\_freq_{patient} = FREQUENT(mainspef_{episode})$
28. mainspefRecoded	
$mainspefRecoded_{nn_{episode}} = COUNT(group(DICTIONARY_{mainspefRecoded}(mainspef, group))$	$mainspefRecoded\_freq1_{patient} = FREQUENT(mainspefRecoded_{episode})$
29. palliativeCare	
$IF(ANY_{nn=1}^{20}(diag_{nn_{episode}} = Z515))\ THEN$ $palliativeCare_{episode} = TRUE\ ENDIF$	NA
30. palliativeMedicine	
$IF(tretspef_{episode} = 315))\ THEN$ $palliativeMedicine_{spell} = TRUE\ ENDIF$	NA
31. elecdur	
$elecdur_{episode}$	$elecdur\_avg_{patient} = AVERAGE(elecdur_{spell})$ $elecdur\_freq_{patient} = COUNT_{day}(elecdur_{spell})$ $elecdur\_nulls\_freq_{patient} = COUNT_{NA}(elecdur_{spell})$ $elecdur\_stdev_{patient} = STD\_DEV(elecdur_{spell})$
32. charlsonIndex	
$episode_{spell_{selected}} = episode_{spell}, WHERE$ $IF(COUNT(episode_{spell}) > 1)\ THEN$ $episode_{spell_{order}} = 2$ $ELSE\ episode_{spell_{order}} = 1$ $charlsonIndex_{spell} = \sum_{nn=2}^{20} =$ $COUNT(weight(DICTIONARY_{Foster\_CCI}($ $diag_{nn_{episode_{spell_{selected}}}}, weight))$ $IF(15, 11 \subseteq ANY_{nn=2}^{20}(diag_{nn_{episode_{spell_{selected}}}})\ THEN$ $charlsonIndex_{spell} = charlsonIndex_{spell} - 8$ $IF(charlsonIndex_{spell} < 0)\ THEN\ charlsonIndex = 0$	$charlsonIndex\_avg_{patient} = AVERAGE(charlsonIndex_{spell})$ $charlsonIndex\_max_{patient} = MAX(charlsonIndex_{spell})$ $charlsonIndex\_stdev_{patient} = STD\_DEV(charlsonIndex_{spell})$ $charlsonIndex\_zerofreq_{patient} = COUNT_{NA}(charlsonIndex_{spell})$
33. diag2Diag20	
$diag_{nn_{episode}}, WHERE\ nn \in 2, 3, ..., 20$	NA
34. mainDiag	
$mainDiag_{spell} = COUNT(category(DICTIONARY_{AHRQ\_CCS}($ $diag_{episodenn}, category)), WHERE$ $IF(diag_{episode_1} == "R\ CODE"\ \wedge\ diag_{episode_2} \neq "R\ CODE")\ THEN$ $nn = 2$ $ELSE\ nn = 1$	NA
35. shmiIndex	

Continued on next page

Formula	Sub-Features
$IF(charlsonIndex_{spell} = 0)\ THEN$ $shmiIndex_{spell} = 0$ $ELSE\ IF(0 < charlsonIndex_{spell} \leq 5)\ THEN$ $shmiIndex_{spell} = 1$ $ELSE\ IF(charlsonIndex_{spell} > 5)\ THEN$ $shmiIndex_{spell} = 2$	$shmiIndex\_avg_{patient} =$ $AVERAGE(shmiIndex_{spell})$ $shmiIndex\_max_{patient} =$ $MAX(shmiIndex_{spell})$ $shmiIndex\_stdev_{patient} =$ $STD\_DEV(shmiIndex_{spell})$ $shmiIndex\_zero\_freq_{patient} =$ $COUNT_{zero}(shmiIndex_{spell})$ $ENDIF$
36. diagCci_39_coagulopathy_freq $diagCci\_39\_coagulopathy\_freq_{spell} =$ $COUNT(diag\_nn_{episode} \in \{"D65", ..., "D68.x", "D69.1", "D69.3", ..., "D69.6"\})$	NA
37. diagCci_43_bloodLoss_freq $diagCci\_43\_bloodLoss\_freq_{spell} =$ $COUNT(diag\_nn_{episode} \in \{"D50.0"\})$	NA
38. diagCci_44_anemia_freq $diagCci\_44\_anemia\_freq_{spell} =$ $COUNT(diag\_nn_{episode} \in \{"D50.8", "D50.9", "D51.x", ..., "D53.x",$ $"F10", "E52", "G62.1", "I42.6"\})$	NA
39. diagOther_3_blood_freq $diagOther\_3\_blood\_freq_{spell} =$ $COUNT(diag\_nn_{episode} \in \{"D47.3", "D96.3", "D72.82"\})$	NA
40. diagOther_4_chronic_g_freq $diagOther\_4\_chronic\_g\_freq_{spell} =$ $COUNT(diag\_nn_{episode} \in \{"D50.1", "D50.8", "D50.9", "D51", "D52"\})$	NA
41. diagOther_6_diagSickle_freq $diagOther\_6\_diagSickle\_freq_{spell} =$ $COUNT(diag\_nn_{episode} \in \{"D57"\})$	NA
42. diagCat_1_freq $diagCat\_1\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I00", ..., "I09"\})$	NA
43. diagCat_10_freq $diagCat\_10\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I95", ..., "I99"\})$	NA
44. diagCat_11_freq $diagCat\_11\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"Q20", ..., "Q28"\})$	NA
45. diagCat_2_freq $diagCat\_2\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I10", ..., "I15"\})$	NA
46. diagCat_3_freq $diagCat\_3\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I20", ..., "I25"\})$	NA
47. diagCat_4_freq $diagCat\_4\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I26", ..., "I28"\})$	NA
48. diagCat_5_freq $diagCat\_5\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I30", ..., "I52"\})$	NA
49. diagCat_6_freq $diagCat\_6\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I60", ..., "I69"\})$	NA
50. diagCat_7_freq $diagCat\_7\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I70"\})$	NA
51. diagCat_8_freq $diagCat\_8\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I71", ..., "I79"\})$	NA
52. diagCat_9_freq $diagCat\_9\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I80", ..., "I89"\})$	NA

Continued on next page

Formula	Sub-Features
53. diagCci_01_myocardial_freq $diagCci\_01\_myocardial\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I21.x", "I22.x", "I25.2"\})$	NA
54. diagCci_02_chf_freq $diagCci\_02\_chf\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I09.9", "I11.0", "I13.0", "I13.2", "I25.5", "I42.0", "I42.5", ..., "I42.9", "I43.x", "I50.x", "P29.0"\})$	NA
55. diagCci_03_pvd_freq $diagCci\_03\_pvd\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I70.x", "I71.x", "I73.1", "I73.8", "I73.9", "I77.1", "I79.0", "I79.2", "K55.1", "K55.8", "K55.9", "Z95.8", "Z95.9"\})$	NA
56. diagCci_13_renal_freq $diagCci\_13\_renal\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I12.0", "I13.1", "N03.2", ..., "N03.7", "N05.2", ..., "N05.7", "N18.x", "N19.x", "N25.0", "Z49.0", ..., "Z49.2", "Z94.0", "Z99.2"\})$	NA
57. diagCci_15_liverSevere_freq $diagCci\_15\_liverSevere\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I85.0", "I85.9", "I86.4", "I98.2", "K70.4", "K71.1", "K72.1", "K72.9", "K76.5", "K76.6", "K76.7"\})$	NA
58. diagCci_19_cardiac_freq $diagCci\_19\_cardiac\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I44.1", ..., "I44.3", "I45.6", "I45.9", "I47.x", ..., "I49.x", "R00.0", "R00.1", "R00.8", "T82.1", "Z45.0", "Z95.0"\})$	NA
59. diagCci_21_pulmonary_freq $diagCci\_21\_pulmonary\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I26.x", "I27.x", "I28.0", "I28.8", "I28.9"\})$	NA
60. diagCci_22_vascular_freq $diagCci\_22\_vascular\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I70.x", "I71.x", "I73.1", "I73.8", "I73.9", "I77.1", "I79.0", "I79.2", "K55.1", "K55.8", "K55.9", "Z95.8", "Z95.9"\})$	NA
61. diagCci_23_hypertensionNotComplicated_freq $diagCci\_23\_hypertensionNotComplicated\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I10.x"\})$	NA
62. diagCci_27_pulmonaryChronic_freq $diagCci\_27\_pulmonaryChronic\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I27.8", "I27.9", "J40.x", ..., "J47.x", "J60.x", ..., "J67.x", "J68.4", "J70.1", "J70.3"\})$	NA
63. diagCci_31_renal_freq $diagCci\_31\_renal\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I12.0", "I13.1", "N18.x", "N19.x", "N25.0", "Z49.0", ..., "Z49.2", "Z94.0", "Z99.2"\})$	NA
64. diagOther_4_chronic_c_freq $diagOther\_4\_chronic\_c\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I11.0", "I13.0", "I50", "J81X"\} \wedge oper\_nn_{episode} \notin \{"K0", "K1", "K2", "K3", "K4", "K50", "K52", "K55", "K56", "K57", "K60", "K61", "K66", "K67", "K68", "K69", "K71"\})$	NA
65. diagOther_4_chronic_f_freq $diagOther\_4\_chronic\_f\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I20", "I25"\} \wedge oper\_nn_{episode} \notin \{"A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "O", "P", "Q", "R", "S", "T", "V", "W", "X0", "X1", "X2", "X4", "X5"\})$	NA
66. diagOther_4_chronic_h_freq $diagOther\_4\_chronic\_h\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I10.X", "I11.9"\} \wedge oper\_nn_{episode} \notin \{"K0", "K1", "K2", "K3", "K4", "K50", "K52", "K55", "K56", "K57", "K60", "K61", "K66", "K67", "K68", "K69", "K71"\})$	NA
67. diagRisk_3_blood_extra_freq $diagRisk\_3\_blood\_extra\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I10", "I11", "I12", "I13", "I14", "I15", "I27.0", "I27.2", "I6", "I87.0", "I87.30", "I87.31", "I87.32", "I87.33", "I87.39", "I97.3", "K76.6", "H35.0", "I10", "R03", "O13", "O14", "O16", "O10", "G93.2", "H40.05", "P29.2", "P29.3"\})$	NA

Continued on next page

Formula	Sub-Features
68. diagRisk_3_blood_freq $diagRisk\_3\_blood\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I27", "I87", "I97", "H40.0"\})$	NA
69. diagCci_08_ulcer_freq $diagCci\_08\_ulcer\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"K25.x", ..., "K28.x"\})$	NA
70. diagCci_09_liverMild_freq $diagCci\_09\_liverMild\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"B18.x", "K70.0", ..., "K70.3", "K70.9", "K71.3", ..., "K71.5", "K71.7", "K73.x", "K74.x", "K76.0", "K76.2", ..., "K76.4", "K76.8", "K76.9", "Z94.4"\})$	NA
71. diagCci_17_aids_freq $diagCci\_17\_aids\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"B20.x", ..., "B22.x", "B24.x"\})$	NA
72. diagCci_32_liver_freq $diagCci\_32\_liver\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"B18.x", "I85.x", "I86.4", "I98.2", "K70.x", "K71.1", "K71.3", ..., "K71.5", "K71.7", "K72.x", ..., "K74.x", "K76.0", "K76.2", ..., "K76.9", "Z94.4"\})$	NA
73. diagCci_33_ulcerNotBleeding_freq $diagCci\_33\_ulcerNotBleeding\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"K25.7", "K25.9", "K26.7", "K26.9", "K27.7", "K27.9", "K28.7", "K28.9"\})$	NA
74. diagCci_45_alcohol_freq $diagCci\_45\_alcohol\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"K29.2", "K70.0", "K70.3", "K70.9", "T51.x", "Z50.2", "Z71.4", "Z72.1"\})$	NA
75. diagMorbid_8_periodontitis_freq $diagMorbid\_8\_periodontitis\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"K04.4", "K04.5", "K05.2", "K05.3"\})$	NA
76. diagRisk_7_kidney_freq $diagOther\_2\_liver\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"K70", ..., "K77", "I85"\})$	NA
77. diagOther_4_chronic_a_freq $diagOther\_4\_chronic\_a\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"B18.0", "B18.1"\} \ \wedge \ oper\_nn_{episode} \notin \{"D57"\})$	NA
78. diagCci_10_diabetesNotChronic_freq $diagCci\_10\_diabetesNotChronic\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E10.0", "E10.1", "E10.6", "E10.8", "E10.9", "E11.0", "E11.1", "E11.6", "E11.8", "E11.9", "E12.0", "E12.1", "E12.6", "E12.8", "E12.9", "E13.0", "E13.1", "E13.6", "E13.8", "E13.9", "E14.0", "E14.1", "E14.6", "E14.8", "E14.9"\})$	NA
79. diagCci_11_diabetesChronic_freq $diagCci\_11\_diabetesChronic\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E10.2", ..., "E10.5", "E10.7", "E11.2", ..., "E11.5", "E11.7", "E12.2", ..., "E12.5", "E12.7", "E13.2", ..., "E13.5", "E13.7", "E14.2", ..., "E14.5", "E14.7"\})$	NA
80. diagCci_28_diabetesNotComplicated_freq $diagCci\_28\_diabetesNotComplicated\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E10.0", "E10.1", "E10.9", "E11.0", "E11.1", "E11.9", "E12.0", "E12.1", "E12.9", "E13.0", "E13.1", "E13.9", "E14.0", "E14.1", "E14.9"\})$	NA
81. diagCci_29_diabetesComplicated_freq $diagCci\_29\_diabetesComplicated\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E10.2", ..., "E10.8", "E11.2", ..., "E11.8", "E12.2", ..., "E12.8", "E13.2", ..., "E13.8", "E14.2", ..., "E14.8"\})$	NA
82. diagCci_30_hypothyroidism_freq $diagCci\_30\_hypothyroidism\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E00.x", ..., "E03.x", "E89.0"\})$	NA
83. diagCci_40_obesity_freq $diagCci\_40\_obesity\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E66.x"\})$	NA

Continued on next page

Formula	Sub-Features
84. diagCci_41_weightLoss_freq $diagCci\_41\_weightLoss\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E40.x", ..., "E46.x", "R63.4", "R64"\})$	NA
85. diagCci_42_fluidDisorder_freq $diagCci\_42\_fluidDisorder\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E22.2", "E86.x", "E87.x"\})$	NA
86. diagOther_1_ulcers_freq $diagOther\_1\_ulcers\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E08.6", "E09.6", "E10.6", "E11.6", "E13.6", "Z86.3"\})$	NA
87. diagOther_4_chronic_d_freq $diagOther\_4\_chronic\_d\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E10", ..., "E14"\})$	NA
88. diagRisk_2_Cholesterol_freq $diagRisk\_2\_Cholesterol\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E78", "Z13.220"\})$	NA
89. diagRisk_6_Metabolic_freq $diagRisk\_6\_Metabolic\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"E88.81"\})$	NA
90. diagRisk_10_externalMorbidity_freq $diagRisk\_10\_externalMorbidity\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"V", "Y"\})$	NA
91. diagMorbid_2_kidney_freq $diagMorbid\_2\_kidney\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"N17", ..., "N19", "N18"\}\ OR\ \{"N18" \wedge "Z94.0"\}\ OR\ \{"N18" \wedge "I12"\})$	NA
92. diagMorbid_4_erectile_freq $diagMorbid\_4\_erectile\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"N52"\})$	NA
93. diagCci_20_valvular_freq $diagCci\_20\_valvular\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"A52.0", "I05.x", ..., "I08.x", "I09.1", "I09.8", "I34.x", ..., "I39.x", "Q23.0", ..., "Q23.3", "Z95.2", ..., "Z95.4"\})$	NA
94. diagMorbid_10_vascularOper_freq $diagMorbid\_10\_vascularOper\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"T86.2", "Z94.3", "Z94.1", "Z48.21", "Z48.280"\})$	NA
95. diagRisk_5_diabetes_extra_freq $diagRisk\_5\_diabetes\_extra\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"T85.6", "T85.6", "T38.3x1", "T38.3x6"\}\ OR\ \{"E89.1" \wedge "Z90.41"\})$	NA
96. diagRisk_9_external_freq $diagRisk\_9\_external\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"S", "T"\})$	NA
97. diagCci_05_dementia_freq $diagCci\_05\_dementia\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"F00.x", ..., "F03.x", "F05.1", "G30.x", "G31.1"\})$	NA
98. diagCci_18_depression_freq $diagCci\_18\_depression\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"F20.4", "F31.3", ..., "F31.5", "F32.x", "F33.x", "F34.1", "F41.2", "F43.2"\})$	NA
99. diagCci_34_psychoses_freq $diagCci\_34\_psychoses\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"F20.x", "F22.x", ..., "F25.x", "F28.x", "F29.x", "F30.2", "F31.2", "F31.5"\})$	NA
100. diagCci_46_drug_freq $diagCci\_46\_drug\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"F11.x", ..., "F16.x", "F18.x", "F19.x", "Z71.5", "Z72.2"\})$	NA
101. diagOther_5_alcohol_freq $diagOther\_5\_alcohol\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"F10"\})$	NA
102. diagOther_8_mental_freq $diagOther\_8\_mental\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"F01", ..., "F19"\})$	NA

Continued on next page

Formula	Sub-Features
103. diagCci_07_rheumatic_freq $diagCci\_07\_rheumatic\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"M05.x", "M06.x", "M31.5", "M32.x", ..., "M34.x", "M35.1", "M35.3", "M36.0"\})$	NA
104. diagMorbid_6_rheumatoid_freq $diagMorbid\_6\_rheumatoid\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"M05", ..., "M06"\})$	NA
105. diagCci_14_malignancy_freq $diagCci\_14\_malignancy\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"C00.x", ..., "C26.x", "C30.x", ..., "C34.x", "C37.x", ..., "C41.x", "C43.x", "C45.x", ..., "C58.x", "C60.x", ..., "C76.x", "C81.x", ..., "C85.x", "C88.x", "C90.x", ..., "C97.x"\})$	NA
106. diagCci_16_tumorSec_freq $diagCci\_16\_tumorSec\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"C77.x", ..., "C80.x"\})$	NA
107. diagCci_35_lymphoma_freq $diagCci\_35\_lymphoma\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"C81.x", ..., "C85.x", "C88.x", "C96.x", "C90.0", "C90.2"\})$	NA
108. diagCci_36_cancerSec_freq $diagCci\_36\_cancerSec\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"C77.x", ..., "C80.x"\})$	NA
109. diagCci_37_tumorNotSec_freq $diagCci\_37\_tumorNotSec\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"C00.x", ..., "C26.x", "C30.x", ..., "C34.x", "C37.x", ..., "C41.x", "C43.x", "C45.x", ..., "C58.x", "C60.x", ..., "C76.x", "C97.x"\})$	NA
110. diagOther_7_cancer_freq $diagOther\_7\_cancer\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"C00", ..., "D4"\})$	NA
111. diagCci_04_cerebrovascular_freq $diagCci\_04\_cerebrovascular\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"G45.x", "G46.x", "H34.0", "I60.x", ..., "I69.x"\})$	NA
112. diagCci_12_hemiplegia_freq $diagCci\_12\_hemiplegia\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"G04.1", "G11.4", "G80.1", "G80.2", "G81.x", "G82.x", "G83.0", ..., "G83.4", "G83.9"\})$	NA
113. diagCci_25_paralysis_freq $diagCci\_25\_paralysis\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"G04.1", "G11.4", "G80.1", "G80.2", "G81.x", "G82.x", "G83.0", ..., "G83.4", "G83.9"\})$	NA
114. diagCci_26_neuroOther_freq $diagCci\_26\_neuroOther\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"G10.x", ..., "G13.x", "G20.x", ..., "G22.x", "G25.4", "G25.5", "G31.2", "G31.8", "G31.9", "G32.x", "G35.x", ..., "G37.x", "G40.x", "G41.x", "G93.1", "G93.4", "R47.0", "R56.x"\})$	NA
115. diagMorbid_3_sleep_freq $diagMorbid\_3\_sleep\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"G47.33", "P28.3"\})$	NA
116. diagOther_4_chronic_i_freq $diagOther\_4\_chronic\_i\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"G40", "G41", "F00", "F01", "F02", "F03", "I48X"\})$	NA
117. diagRisk_4_Glucose_freq $diagRisk\_4\_Glucose\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"R73"\})$	NA
118. diagMorbid_9_vascularRadi_freq $diagMorbid\_9\_vascularRadi\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"Z51", "W88", "Y36.5", "Y37.5", "Y38.5", "Y63"\})$	NA
119. diagRisk_8_smoke_freq $diagRisk\_8\_smoke\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"Z77", "P968", "Z878", "Z573", "F17", "Z72.0"\})$	NA
120. palliativeCare_freq $palliativeCare\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"Z515"\})$	NA

Continued on next page

Formula	Sub-Features
121. reference_freq $reference\_freq_{spell} = hrg\_n.n_{episode} \in \{"A18", "A29", "D16", "D17", "D20", "D21", "D26", "D33", "D99", "E18", "E19", "E22", "E29", "E33", "E99", "G25", "H25", "K11", "K99", "P02", "P23", "P25", "Q17", "S04", "S05", "S06", "T01"\}$	NA
122. diagCci_06_cpd_freq $diagCci\_06\_cpd\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"I27.8", "I27.9", "J40.x", ..., "J47.x", "J60.x", ..., "J67.x", "J68.4", "J70.1", "J70.3"\})$	NA
123. diagMorbid_1_Influenza_freq $diagMorbid\_1\_Influenza\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"J09", ..., "J18"\})$	NA
124. diagOther_4_chronic_b_freq $diagOther\_4\_chronic\_b\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"J45", "J46X"\})$	NA
125. diagOther_4_chronic_e_freq $diagOther\_4\_chronic\_e\_freq_{spell} = COUNT(mainDiag_{episode} \in \{"J20", "J41", "J42X", "J43", "J44", "J47X", "J20"\}) + COUNT(diag2Diag20_{episode} \in \{"J41", "J42", "J43", "J44", "J47"\})$	NA
126. diagCci_38_rheumatoid_freq $diagCci\_38\_rheumatoid\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"L94.0", "L94.1", "L94.3", "M05.x", "M06.x", "M08.x", "M12.0", "M12.3", "M30.x", "M31.0", ..., "M31.3", "M32.x", ..., "M35.x", "M45.x", "M46.1", "M46.8", "M46.9"\})$	NA
127. diagMorbid_5_psoriasis_freq $diagMorbid\_5\_psoriasis\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"L40"\})$	NA
128. diagMorbid_7_lupus_freq $diagMorbid\_7\_lupus\_freq_{spell} = COUNT(diag\_nn_{episode} \in \{"L93", "M32"\})$	NA
129. oper_2_heart_freq $oper\_2\_heart\_freq_{spell} = COUNT(oper\_nn_{episode} \in \{"G28", "G30", "G32", "G61"\})$	NA
130. oper_1_obesity_freq $oper\_1\_obesity\_freq_{spell} = COUNT(oper\_nn_{episode} \in \{"K", "L"\})$	NA
131. oper_3_urinary_freq $oper\_3\_urinary\_freq_{spell} = COUNT(oper\_nn_{episode} \in \{"M"\})$	NA
132. oper_4_radio_freq $oper\_4\_radio\_freq_{spell} = COUNT(oper\_nn_{episode} \in \{"X65", "X67", "X68"\})$	NA
133. triggerStartAge $triggerStartAge_{patient} = MIN(startage_{patient_{trigger}})$	NA
134. triggerStartAgeRecoded	

Continued on next page

Formula	Sub-Features
$IF(triggerStartAge_{patient} = NULL)\ THEN$	NA
$triggerStartAgeRecoded_{patient} = 0$	
$ELSE\ IF(1 \leq triggerStartAge_{patient})\ THEN$	
$triggerStartAgeRecoded_{patient} = 1$	
$ELSE\ IF(1 \leq triggerStartAge_{patient} < 4)\ THEN$	
$triggerStartAgeRecoded_{patient} = 2$	
$ELSE\ IF(4 \leq triggerStartAge_{patient} < 9)\ THEN$	
$triggerStartAgeRecoded_{patient} = 3$	
$ELSE\ IF(9 \leq triggerStartAge_{patient} < 14)\ THEN$	
$triggerStartAgeRecoded_{patient} = 4$	
$ELSE\ IF(14 \leq triggerStartAge_{patient} < 19)\ THEN$	
$triggerStartAgeRecoded_{patient} = 5$	
$ELSE\ IF(19 \leq triggerStartAge_{patient} < 24)\ THEN$	
$triggerStartAgeRecoded_{patient} = 6$	
$ELSE\ IF(24 \leq triggerStartAge_{patient} < 29)\ THEN$	
$triggerStartAgeRecoded_{patient} = 7$	
$ELSE\ IF(29 \leq triggerStartAge_{patient} < 34)\ THEN$	
$triggerStartAgeRecoded_{patient} = 8$	
$ELSE\ IF(34 \leq triggerStartAge_{patient} < 39)\ THEN$	
$triggerStartAgeRecoded_{patient} = 9$	
$ELSE\ IF(39 \leq triggerStartAge_{patient} < 44)\ THEN$	
$triggerStartAgeRecoded_{patient} = 10$	
$ELSE\ IF(44 \leq triggerStartAge_{patient} < 49)\ THEN$	
$triggerStartAgeRecoded_{patient} = 11$	
$ELSE\ IF(49 \leq triggerStartAge_{patient} < 54)\ THEN$	
$triggerStartAgeRecoded_{patient} = 12$	
$ELSE\ IF(54 \leq triggerStartAge_{patient} < 59)\ THEN$	
$triggerStartAgeRecoded_{patient} = 13$	
$ELSE\ IF(59 \leq triggerStartAge_{patient} < 64)\ THEN$	
$triggerStartAgeRecoded_{patient} = 14$	
$ELSE\ IF(64 \leq triggerStartAge_{patient} < 69)\ THEN$	
$triggerStartAgeRecoded_{patient} = 15$	
$ELSE\ IF(69 \leq triggerStartAge_{patient} < 74)\ THEN$	
$triggerStartAgeRecoded_{patient} = 16$	
$ELSE\ IF(74 \leq triggerStartAge_{patient} < 79)\ THEN$	
$triggerStartAgeRecoded_{patient} = 17$	
$ELSE\ IF(79 \leq triggerStartAge_{patient} < 84)\ THEN$	
$triggerStartAgeRecoded_{patient} = 18$	
$ELSE\ IF(84 \leq triggerStartAge_{patient} < 89)\ THEN$	
$triggerStartAgeRecoded_{patient} = 19$	
$ELSE\ IF(89 \leq triggerStartAge_{patient} < 120)\ THEN$	
$triggerStartAgeRecoded_{patient} = 20\ ENDIF$	
135. imd04rk	
$imd04rk_{patient} = AVERAGE(imd04rk_{patient_{trigger}})$	NA
136. imd04rkRecoded	

Continued on next page

Formula	Sub-Features
$IF(imd04rk_{patient} = NULL)\ THEN$	NA
$imd04rkRecoded_{patient} = 0$	
$ELSE\ IF(1 \leq imd04rk_{patient} < 3,249)\ THEN$	
$imd04rkRecoded_{patient} = 1$	
$ELSE\ IF(3,249 \leq imd04rk_{patient} < 6,497)\ THEN$	
$imd04rkRecoded_{patient} = 2$	
$ELSE\ IF(6,497 \leq md04rk_{patient} < 9,746)\ THEN$	
$imd04rkRecoded_{patient} = 3$	
$ELSE\ IF(9,746 \leq imd04rk_{patient} < 12,994)\ THEN$	
$imd04rkRecoded_{patient} = 4$	
$ELSE\ IF(12,994 \leq imd04rk_{patient} < 16,242)\ THEN$	
$imd04rkRecoded_{patient} = 5$	
$ELSE\ IF(16,242 \leq md04rk_{patient} < 19,490)\ THEN$	
$imd04rkRecoded_{patient} = 6$	
$ELSE\ IF(19,490 \leq imd04rk_{patient} < 22,738)\ THEN$	
$imd04rkRecoded_{patient} = 7$	
$ELSE\ IF(22,738 \leq imd04rk_{patient} < 25,987)\ THEN$	
$imd04rkRecoded_{patient} = 8$	
$ELSE\ IF(25,987 \leq md04rk_{patient} < 29,235)\ THEN$	
$imd04rkRecoded_{patient} = 9$	
$ELSE\ IF(29,235 \leq imd04rk_{patient} \leq 32,482)\ THEN$	
$imd04rkRecoded_{patient} = 10\ ENDIF$	
137. ethnos	
$ethnos_{patient} = REQUENT(ethnos_{patient_{trigger}})$	NA
138. ethnosRecoded	
$IF(ethnos_{patient} \in \{NULL, "X", "Z", "9"\})\ THEN$	NA
$ethnosRecoded_{patient} = "NA"$	
$ELSE\ IF(ethnos_{patient} \in \{"A", "0", "B", "F", "C"\})\ THEN$	
$ethnosRecoded_{patient} = "White"$	
$ELSE\ IF(ethnos_{patient} \in \{"H", "4"\})\ THEN$	
$ethnosRecoded_{patient} = "Indian"$	
$ELSE\ IF(ethnos_{patient} \in \{"J", "5"\})\ THEN$	
$ethnosRecoded_{patient} = "Pakestani"$	
$ELSE\ IF(ethnos_{patient} \in \{"K", "6"\})\ THEN$	
$ethnosRecoded_{patient} = "Bangeladeshi"$	
$ELSE\ IF(ethnos_{patient} \in \{"M", "1"\})\ THEN$	
$ethnosRecoded_{patient} = "Caribbean"$	
$ELSE\ IF(ethnos_{patient} \in \{"N", "2"\})\ THEN$	
$ethnosRecoded_{patient} = "African"$	
$ELSE\ IF(ethnos_{patient} \in \{"P", "3", "D", "E"\})\ THEN$	
$ethnosRecoded_{patient} = "Other\ Black"$	
$ELSE\ IF(ethnos_{patient} \in \{"R", "7"\})\ THEN$	
$ethnosRecoded_{patient} = "Chinese"$	
$ELSE\ IF(ethnos_{patient} \in \{"S", "8", "L", "G"\})\ THEN$	
$ethnosRecoded_{patient} = "Others"\ ENDIF$	
139. genderRecoded	
$IF(FREQUENT(sex_{patient_{trigger}} \in \{NULL, 0\})\ THEN$	NA
$genderRecoded_{patient} = "NA"$	
$ELSE\ IF(FREQUENT(sex_{patient_{trigger}} = 1)\ THEN$	
$genderRecoded_{patient} = "Male"$	
$ELSE\ IF(FREQUENT(sex_{patient_{trigger}} = 2)\ THEN$	
$genderRecoded_{patient} = "Female"\ ENDIF$	

TABLE A.12: Definitions of the *mainspef* and the *mainspefRecoded* features

#	mainspef	Definition	mainspefRecoded
0	900	Community medicine	Other
0	901	Occupational medicine	Other
0	902	Community health services	Other
0	903	Public health medicine	Other
0	904	Public health denta	Other
0	950	Nursing episode	Other
0	960	Allied health professional episode	Other
1	180	Accident & emergency (A&E)	A&E
2	190	Anaesthetics	Anaesthetics & pain management
2	191	Pain management	Anaesthetics & pain management
3	170	Cardiothoracic surgery	Cardiothoracic
3	320	Cardiology	Cardiothoracic
3	321	Paediatric cardiology	Cardiothoracic
4	141	Restorative dentistry	Dental
4	142	Paediatric dentistry	Dental
4	149	Surgical dentistry	Dental
4	450	Dental medicine	Dental
4	601	General Dental Practice	Dental
5	330	Dermatology	Dermatology
6	120	Ear, nose & throat (ENT)	Ear, nose & throat (ENT)
7	302	Endocrinology	Endocrinology
8	301	Gastroenterology	Gastroenterology
9	300	General medicine	General
9	600	General Medical Practice	General
9	620	General practice other than maternity	General
10	100	General surgery	General Surgery
11	430	Geriatric medicine	Geriatric
12	502	Gynaecology	Gynaecology
13	303	Clinical haemotology	Haemotology
14	350	Infectious diseases	Infectious diseases
15	501	Obstetrics	Maternity
15	560	Midwifery	Maternity
15	610	General practice with maternity function	Maternity
16	360	Genito-urinary medicine	Nephrology & Urinary
16	361	Nephrology	Nephrology & Urinary
17	150	Neurosurgery	Neurology
17	400	Neurology	Neurology
17	401	Clinical neuro-physiology	Neurology
17	421	Paediatric neurology	Neurology
18	370	Medical oncology	Oncology
19	130	Ophthalmology	Ophthalmology
20	140	Oral surgery	Oral & Maxillofacial Surgery
20	145	Oral & maxillo facial surgery	Oral & Maxillofacial Surgery
21	143	Orthodontics	Other Medicine & Surgical
21	146	Endodontics	Other Medicine & Surgical
21	147	Periodontics	Other Medicine & Surgical
21	148	Prosthodontics	Other Medicine & Surgical
21	192	Critical care medicine	Other Medicine & Surgical
21	199	Non-UK Provider	Other Medicine & Surgical
21	304	Clinical physiology	Other Medicine & Surgical
21	310	Audiological medicine	Other Medicine & Surgical
21	311	Clinical genetics	Other Medicine & Surgical
21	312	Clinical cytogenics & molecular genetics	Other Medicine & Surgical
21	313	Clinical immunology & allergy	Other Medicine & Surgical
21	352	Tropical medicine	Other Medicine & Surgical
21	460	Medical opthalmology	Other Medicine & Surgical
21	499	Non-UK Provider	Other Medicine & Surgical
22	420	Paediatrics	Paediatrics
23	171	Paediatric surgery	Paediatrics Surgery
24	315	Palliative medicine	Palliative
25	820	General pathology	Pathology
25	821	Blood transfusion	Pathology
25	822	Chemical pathology	Pathology
25	823	Haematology	Pathology
25	824	Histopathology	Pathology
25	830	Immunopathology	Pathology
25	831	Medical microbiology	Pathology
25	832	Neuropathology	Pathology
26	305	Clinical pharmacology	Pharmacology & Medicine
27	110	Trauma & orthopaedics	Plastic
27	160	Plastic surgery	Plastic
28	700	Learning disability	Psychiatry
28	710	Adult mental illness	Psychiatry
28	711	Child & adolescent psychiatry	Psychiatry
28	712	Forensic psychiatry	Psychiatry
28	713	Psychotherapy	Psychiatry
28	715	Old age psychiatry	Psychiatry
29	371	Nuclear medicine	Radiology
29	800	Clinical oncology/ Radiotherapy	Radiology
29	810	Radiology	Radiology
30	314	Rehabilitation	Rehabilitation
31	340	Respiratory medicine	Respiratory
32	410	Rheumatology	Rheumatology
33	101	Urology	Urology

A.5 Exploratory Analysis

A.5.1 Descriptive Statistics of the HES Inpatient Table

TABLE A.13: Descriptive statistics of the HES Inpatient (full population)

Timeframe	**Episodes**[a]	**Unique Patients**[b]	**In-Hospital Deaths**[c]
[before - 95.01)	120,265	26,329	2,917
[95.01 - 95.04)	7,038	1,101	222
[95.04 - 95.07)	7,454	1,183	207
[95.07 - 96.10)	8,737	1,298	258
[96.10 - 96.01)	14,652	1,663	247
[96.01 - 96.04)	163,215	3,476	278
[96.04 - 96.07)	2,749,577	5,344	328
[96.07 - 97.10)	2,805,893	4,685	375
[97.10 - 97.01)	2,848,432	12,006	732
[97.01 - 97.04)	2,837,168	126,700	10,035
[97.04 - 97.07)	2,918,779	2,134,594	58,341
[97.07 - 98.10)	2,912,508	2,112,901	56,377
[98.10 - 98.01)	2,863,725	2,074,988	63,032
[98.01 - 98.04)	2,960,848	2,127,869	65,400
[98.04 - 98.07)	2,905,382	2,112,059	60,701
[98.07 - 99.10)	3,023,064	2,192,427	58,100
[99.10 - 99.01)	3,055,677	2,205,082	71,889
[99.01 - 99.04)	3,095,404	2,220,246	67,654
[99.04 - 99.07)	3,105,967	2,148,992	58,264
[99.07 - 00.10)	3,167,735	2,197,861	56,792
[00.10 - 00.01)	3,166,087	2,182,087	72,721
[00.01 - 00.04)	3,267,136	2,246,435	68,114
[00.04 - 00.07)	3,187,241	2,173,315	58,830
[00.07 - 01.10)	3,193,730	2,190,678	56,398
[01.10 - 01.01)	3,200,789	2,187,832	64,373
[01.01 - 01.04)	3,275,679	2,221,350	66,503
[01.04 - 01.07)	3,195,150	2,155,444	61,521
[01.07 - 02.10)	3,224,230	2,177,022	57,873
[02.10 - 02.01)	3,258,326	2,187,605	65,905
[02.01 - 02.04)	3,274,550	2,184,154	66,058
[02.04 - 02.07)	3,273,544	2,188,117	61,305
[02.07 - 03.10)	3,356,944	2,234,744	60,014
[03.10 - 03.01)	3,345,833	2,217,512	67,731
[03.01 - 03.04)	3,444,811	2,277,501	67,537
[03.04 - 03.07)	3,393,718	2,228,996	61,763
[03.07 - 04.10)	3,510,433	2,298,963	60,607
[04.10 - 04.01)	3,537,263	2,305,662	71,777
[04.01 - 04.04)	3,633,673	2,357,917	66,489
[04.04 - 04.07)	3,557,005	2,305,457	60,168
[04.07 - 05.10)	3,639,798	2,348,126	58,991
[05.10 - 05.01)	3,664,560	2,350,715	67,211
[05.01 - 05.04)	3,666,341	2,332,739	71,462
[05.04 - 05.07)	3,775,273	2,405,817	60,702
[05.07 - 06.10)	3,801,861	2,410,653	56,957
[06.10 - 06.01)	3,854,788	2,435,704	64,967
[06.01 - 06.04)	3,905,841	2,448,819	68,630
[06.04 - 06.07)	3,860,587	2,423,167	59,473
[06.07 - 07.10)	3,917,355	2,447,565	55,326
[07.10 - 07.01)	3,967,287	2,468,216	61,543
[07.01 - 07.04)	4,040,330	2,524,416	64,785
[07.04 - 07.07)	4,022,303	2,507,277	56,723
[07.07 - 08.10)	4,068,534	2,542,374	53,861
[08.10 - 08.01)	4,149,038	2,595,681	62,965
[08.01 - 08.04)	4,220,524	2,637,602	63,051
[08.04 - 08.07)	4,287,498	2,661,918	55,475
[08.07 - 09.10)	4,325,488	2,676,844	52,907
[09.10 - 09.01)	4,412,952	2,725,331	66,767
[09.01 - 09.04)	4,400,573	2,724,586	61,343
[09.04 - 09.07)	4,432,390	2,716,139	52,922
[09.07 - 10.10)	4,489,317	2,742,665	51,281
[10.10 - 10.01)	4,617,801	2,807,007	60,772
[10.01 - 10.04)	4,564,225	2,787,305	59,529
[10.04 - 10.07)	4,583,220	2,782,770	51,651
[10.07 - 10.10)	2,953,483	1,934,450	25,775
Total	206,489,029	127,469,481	3,312,905

[a] The number of patient excludes records with NULLs $HESID$.
[b] The number of episodes excludes records with NULLs $ADMIDATE$.
[c] The patients with $dismeth$ value of 4.

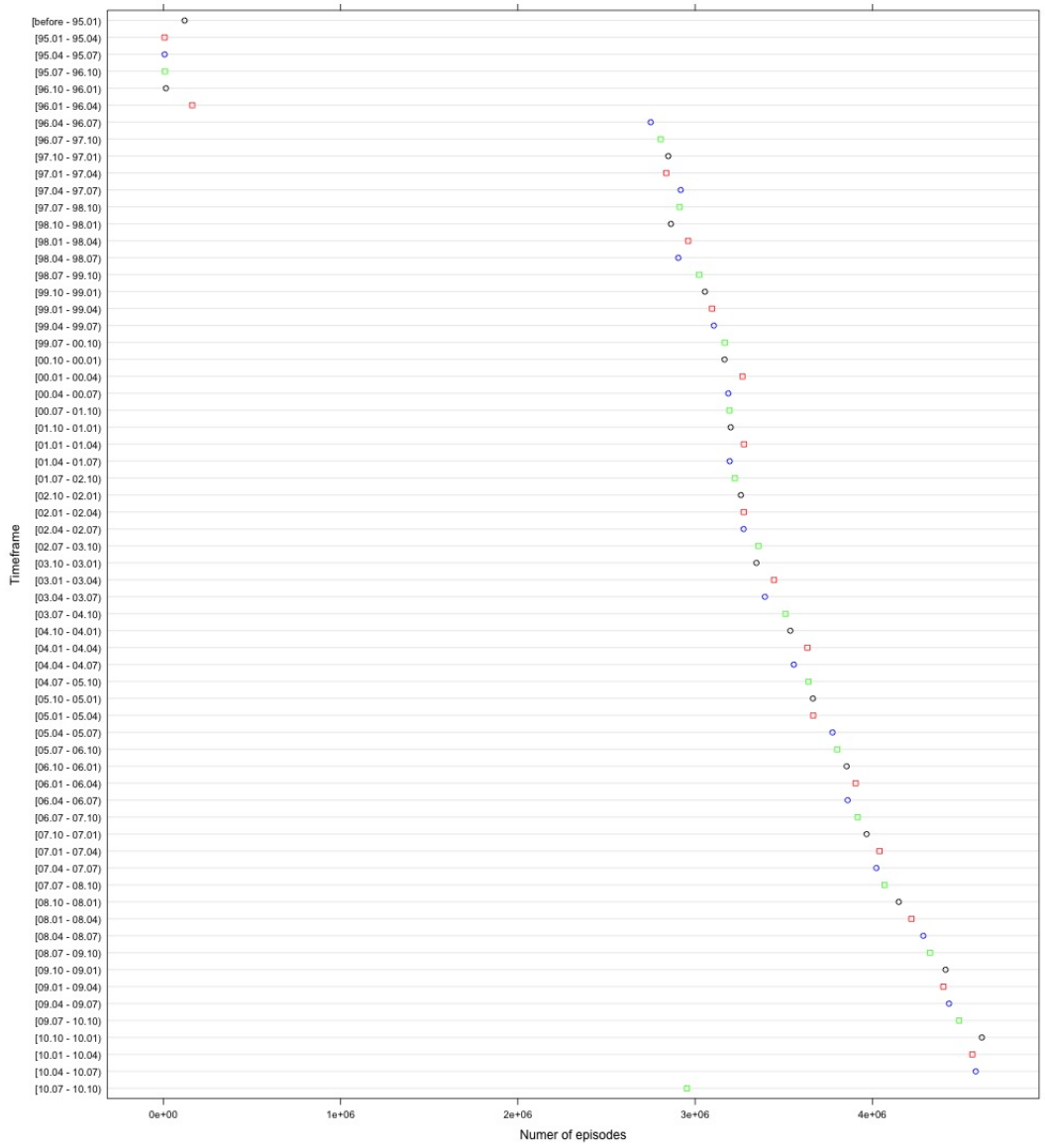

FIGURE A.1: The year quarters (colour coded) vs number of episodes for the HES Inpatient (full population, part 1)

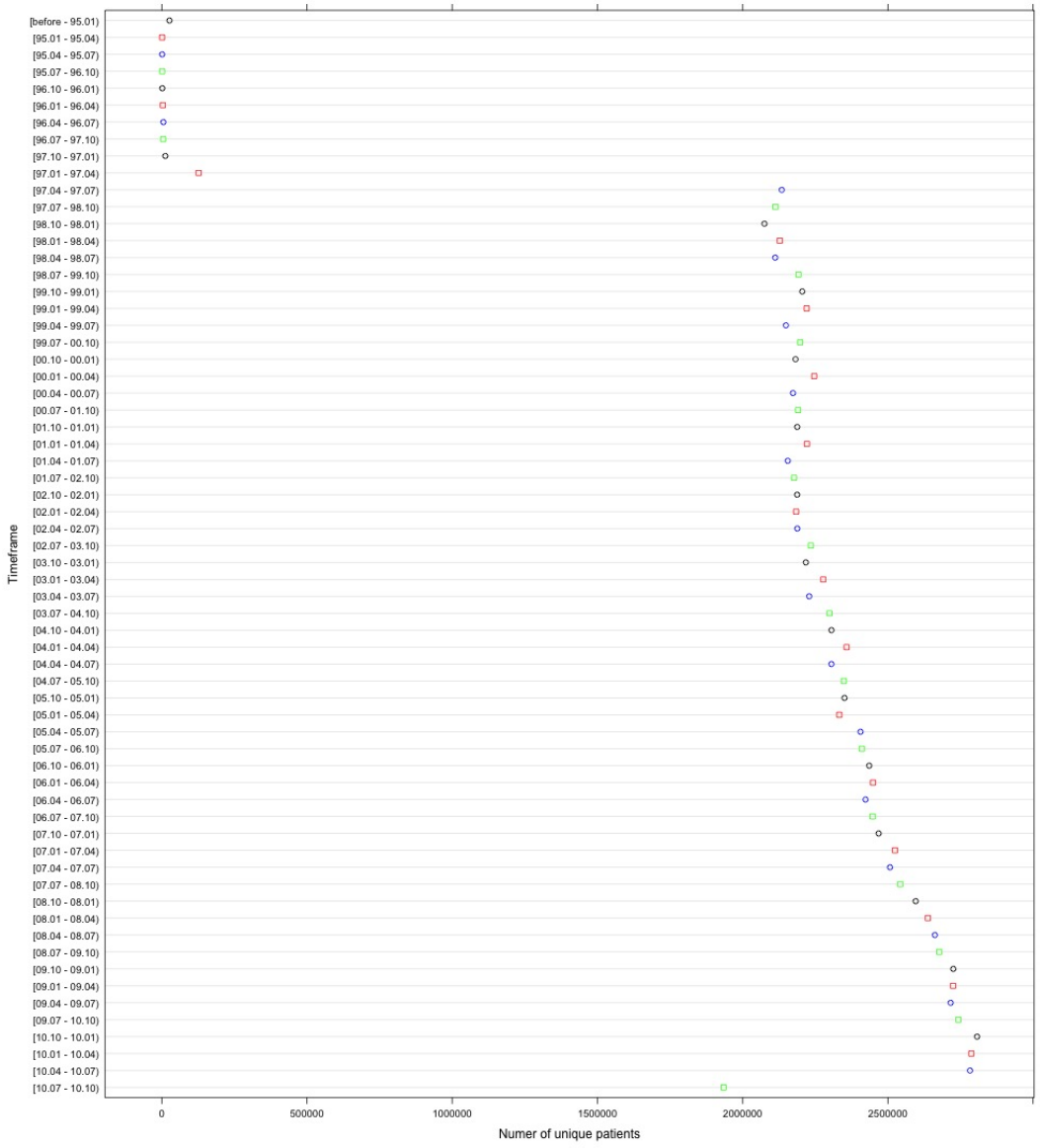

FIGURE A.2: The year quarters (colour coded) vs number of episodes for the HES Inpatient (full population, part 2)

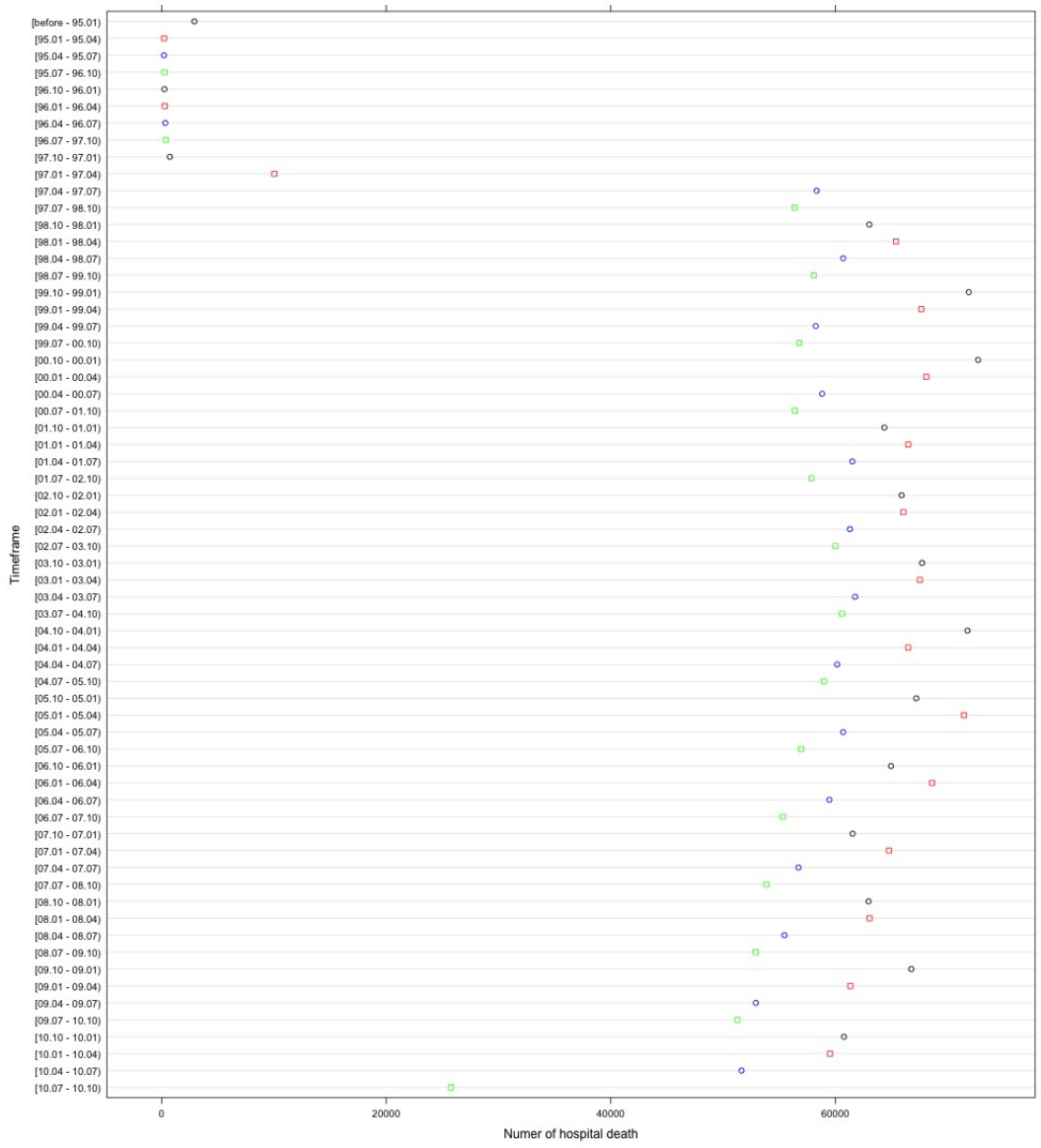

FIGURE A.3: The year quarters (colour coded) vs number of episodes for the HES Inpatient (full population, part 3)

A.5.2 Descriptive Statistics of the Samples

A.5.2.1 Sample-1

TABLE A.14: Descriptive statistics of the continuous variables (*sample-1*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Conditional: Cond_Main;
Sample: Sample-1;
Test sub-sample size: 231,755

Feature [a]	**Mean** [b]	**Stnd. Devi.** [c]	**Min** [d]	**Q. 1** [e]	**Q. 2** [f]	**Q. 3** [g]	**Max** [h]	**Zero or NA (%)** [i]
readmiGap_avg__trigger	1.00	0.00	1.00	1.00	1.00	1.00	1.00	0 (0.00%)
mainspefRecoded_freq1__trigger_15	0.14	0.35	0.00	0.00	0.00	0.00	1.00	198,227 (85.53%)
mainspefRecoded_freq1__trigger_12	0.10	0.30	0.00	0.00	0.00	0.00	1.00	208,100 (89.79%)
mainspefRecoded_freq1__trigger_09	0.25	0.43	0.00	0.00	0.00	1.00	1.00	173,377 (74.81%)
mainspefRecoded_freq1__trigger_11	0.08	0.27	0.00	0.00	0.00	0.00	1.00	213,954 (92.32%)
mainspefRecoded_freq1__trigger_10	0.10	0.30	0.00	0.00	0.00	0.00	1.00	208,447 (89.94%)
mainspefRecoded_freq1__trigger_27	0.09	0.29	0.00	0.00	0.00	0.00	1.00	210,262 (90.73%)
mainspefRecoded_freq1__trigger_01	0.02	0.13	0.00	0.00	0.00	0.00	1.00	227,609 (98.21%)
mainspefRecoded_freq1__trigger_28	0.03	0.16	0.00	0.00	0.00	0.00	1.00	225,610 (97.35%)
mainspefRecoded_freq1__trigger_03	0.03	0.16	0.00	0.00	0.00	0.00	1.00	225,365 (97.24%)
mainspefRecoded_freq1__trigger_19	0.01	0.08	0.00	0.00	0.00	0.00	1.00	230,440 (99.43%)
mainspefRecoded_freq1__trigger_33	0.02	0.13	0.00	0.00	0.00	0.00	1.00	227,729 (98.26%)
mainspefRecoded_freq1__trigger_08	0.01	0.12	0.00	0.00	0.00	0.00	1.00	228,607 (98.64%)
mainspefRecoded_freq1__trigger_06	0.01	0.12	0.00	0.00	0.00	0.00	1.00	228,517 (98.60%)
mainspefRecoded_freq1__trigger_31	0.01	0.10	0.00	0.00	0.00	0.00	1.00	229,418 (98.99%)
s_spellAdmiMeth_acute_freq__delta	1.33	0.98	1.00	1.00	1.00	1.00	48.00	114,857 (49.56%)
s_spellAdmiMeth_acute_freq__365days__delta	-0.28	0.84	-41.00	0.00	0.00	0.00	0.00	212,427 (91.66%)
spellOpertn_freq__90days__delta	1.38	3.33	-1.00	0.00	1.00	2.00	395.00	145,093 (62.61%)
posopdur_avg__trigger	1.61	7.21	0.00	0.00	0.00	1.00	1,103.00	166,767 (71.96%)
epidurRecoded_avg	2.99	2.58	0.00	1.00	2.00	4.50	10.00	27,736 (11.97%)
charlsonIndex_avg	1.75	4.14	0.00	0.00	0.00	0.00	84.00	180,405 (77.84%)
readmiGap_avg	0.55	0.50	0.00	0.00	1.00	1.00	1.00	103,415 (44.62%)
epidur_maxAvg__trigger	7.09	29.31	0.00	1.00	2.00	6.00	2,989.00	46,039 (19.87%)
s_spellAdmiMeth_acute_freq__90days__delta	1.09	3.05	-1.00	0.00	0.00	2.00	128.00	147,942 (63.84%)
posopdur_avg	2.23	7.69	0.00	0.00	0.00	2.00	1,315.00	119,334 (51.49%)
charlsonIndex_max__365days	1.83	4.49	0.00	0.00	0.00	0.00	84.00	206,514 (89.11%)
preopdur_avg__trigger	0.54	3.43	0.00	0.00	0.00	0.00	518.00	195,244 (84.25%)
reference_freq__90days__sum	0.50	1.53	0.00	0.00	0.00	0.00	123.00	201,403 (86.90%)

Continued on next page

Feature [a]	**Mean** [b]	**Stnd. Devi.** [c]	**Min** [d]	**Q. 1** [e]	**Q. 2** [f]	**Q. 3** [g]	**Max** [h]	**Zero or NA (%)** [i]
s_spellAdmiMeth_elective_freq__delta	0.18	1.01	0.00	0.00	0.00	0.00	144.00	219,162 (94.57%)
preopdur_avg	0.69	3.33	0.00	0.00	0.00	0.50	518.00	158,721 (68.49%)
posopdur_avg__365days	2.17	9.29	0.00	0.00	0.33	2.00	1,397.00	171,432 (73.97%)
diagRisk_9_external_freq__90days__sum	0.45	1.21	0.00	0.00	0.00	1.00	117.00	199,737 (86.18%)
intmanigRecoded_other_freq__90days	0.56	1.42	0.00	0.00	0.00	1.00	87.00	193,573 (83.52%)
mainspef_uniques_freq__trigger	1.09	0.31	0.00	1.00	1.00	1.00	7.00	50 (0.02%)
epidur_maxAvg__365days	12.05	238.18	0.00	0.50	2.00	5.00	17,866.00	138,017 (59.55%)
diagCat_3_freq__90days__sum	0.60	2.12	0.00	0.00	0.00	0.00	77.00	210,719 (90.92%)
diagCci_06_cpd_freq__trigger	0.10	0.40	0.00	0.00	0.00	0.00	9.00	214,359 (92.49%)
diagCat_5_freq__90days__sum	0.43	1.48	0.00	0.00	0.00	0.00	96.00	212,053 (91.50%)
diagRisk_3_blood_freq__90days__sum	0.58	1.63	0.00	0.00	0.00	0.00	126.00	202,635 (87.44%)
spellOpertn_freq__365days__delta	-0.19	0.86	-114.00	0.00	0.00	0.00	0.00	217,039 (93.65%)
preopdur_avg__365days	0.63	3.06	0.00	0.00	0.00	0.50	405.50	193,126 (83.33%)
diagOther_4_chronic_e_freq__90days__sum	0.33	1.73	0.00	0.00	0.00	0.00	96.00	222,676 (96.08%)
orgCluster_freq1__trigger_1	0.18	0.38	0.00	0.00	0.00	0.00	1.00	190,018 (81.99%)
orgCluster_freq1__trigger_3	0.34	0.47	0.00	0.00	0.00	1.00	1.00	153,815 (66.37%)
orgCluster_freq1__trigger_5	0.11	0.31	0.00	0.00	0.00	0.00	1.00	207,242 (89.42%)
orgCluster_freq1__trigger_4	0.25	0.43	0.00	0.00	0.00	0.00	1.00	174,137 (75.14%)
rotreatRecoded_freq1__trigger_6	0.14	0.35	0.00	0.00	0.00	0.00	1.00	198,591 (85.69%)
rotreatRecoded_freq1__trigger_8	0.10	0.30	0.00	0.00	0.00	0.00	1.00	208,965 (90.17%)
rotreatRecoded_freq1__trigger_7	0.15	0.36	0.00	0.00	0.00	0.00	1.00	196,496 (84.79%)
rotreatRecoded_freq1__trigger_4	0.15	0.36	0.00	0.00	0.00	0.00	1.00	197,284 (85.13%)
rotreatRecoded_freq1__trigger_5	0.09	0.28	0.00	0.00	0.00	0.00	1.00	211,224 (91.14%)
rotreatRecoded_freq1__trigger_3	0.11	0.31	0.00	0.00	0.00	0.00	1.00	205,945 (88.86%)
rotreatRecoded_freq1__trigger_1	0.14	0.35	0.00	0.00	0.00	0.00	1.00	198,870 (85.81%)
rotreatRecoded_freq1__trigger_2	0.11	0.31	0.00	0.00	0.00	0.00	1.00	207,050 (89.34%)
diagOther_4_chronic_d_freq__90days__sum	0.36	1.80	0.00	0.00	0.00	0.00	89.00	220,883 (95.31%)
diagOther_8_mental_freq__90days__sum	0.20	1.07	0.00	0.00	0.00	0.00	48.00	221,716 (95.67%)
diagOther_4_chronic_b_freq__90days__sum	0.24	1.25	0.00	0.00	0.00	0.00	77.00	219,849 (94.86%)
s_spellAdmiMeth_elective_freq__90days__delta	1.33	3.73	-1.00	0.00	1.00	2.00	395.00	158,628 (68.45%)
diagOther_4_chronic_i_freq__90days__sum	0.37	1.63	0.00	0.00	0.00	0.00	221.00	215,528 (93.00%)
diagCci_44_anemia_freq__90days	0.11	0.82	0.00	0.00	0.00	0.00	53.00	225,842 (97.45%)
oper_2_heart_freq__90days	0.13	0.54	0.00	0.00	0.00	0.00	29.00	221,283 (95.48%)
admisorcRecoded_other_freq	0.11	0.46	0.00	0.00	0.00	0.00	43.00	213,702 (92.21%)
diagOther_7_cancer_freq__90days	0.28	1.98	0.00	0.00	0.00	0.00	112.00	220,027 (94.94%)

Continued on next page

Feature [a]	Mean [b]	Stnd. Devi. [c]	Min [d]	Q. 1 [e]	Q. 2 [f]	Q. 3 [g]	Max [h]	Zero or NA (%) [i]
protype_freq1_9	0.02	0.14	0.00	0.00	0.00	0.00	1.00	227,368 (98.11%)
protype_freq1__trigger_9	0.02	0.13	0.00	0.00	0.00	0.00	1.00	227,918 (98.34%)
diagCci_19_cardiac_freq__90days__sum	0.29	1.24	0.00	0.00	0.00	0.00	96.00	217,947 (94.04%)
diagCci_02_chf_freq__90days	0.12	0.76	0.00	0.00	0.00	0.00	55.00	225,283 (97.21%)
elecdur_elective_nulls_freq	0.12	1.50	0.00	0.00	0.00	0.00	393.00	218,051 (94.09%)
oper_3_urinary_freq__90days	0.14	0.75	0.00	0.00	0.00	0.00	51.00	222,055 (95.81%)
diagCci_26_neuroOther_freq__90days	0.16	1.32	0.00	0.00	0.00	0.00	220.00	225,785 (97.42%)
diagCci_14_malignancy_freq__90days	0.18	1.78	0.00	0.00	0.00	0.00	112.00	226,797 (97.86%)
diagCci_18_depression_freq__90days	0.08	0.66	0.00	0.00	0.00	0.00	127.00	227,240 (98.05%)
diagRisk_8_smoke_freq__90days	0.07	0.52	0.00	0.00	0.00	0.00	37.00	226,903 (97.91%)
diagRisk_10_externalMorbidity_freq__90days	0.10	0.99	0.00	0.00	0.00	0.00	274.00	223,390 (96.39%)
diagMorbid_1_Influenza_freq__90days	0.04	0.30	0.00	0.00	0.00	0.00	12.00	228,366 (98.54%)
diagRisk_2_Cholesterol_freq__90days	0.06	0.48	0.00	0.00	0.00	0.00	50.00	228,253 (98.49%)
diagRisk_7_kidney_freq__90days	0.10	2.84	0.00	0.00	0.00	0.00	403.00	229,331 (98.95%)
diagCci_04_cerebrovascular_freq__90days	0.10	0.68	0.00	0.00	0.00	0.00	31.00	226,052 (97.54%)
diagCat_9_freq__90days	0.06	0.40	0.00	0.00	0.00	0.00	31.00	226,830 (97.87%)
spellOpertn_freq__trigger	0.35	0.48	0.00	0.00	0.00	1.00	1.00	149,970 (64.71%)

[a] Feature: Name of the defined feature.
[b] Mean: The mean value of the feature.
[c] Stnd. Devi.: The standard deviation of the feature.
[d] Min: The minimum of the feature.
[e] Q. 1: The value of the first quantile.
[f] Q. 2: The value of the second quantile.
[g] Q. 3: The value of the third quantile.
[h] Max: The maximum of the feature.
[i] Zero or NA (%): The frequency and percentage of zeros or NAs.

TABLE A.15: Descriptive statistics of the discrete variables (*sample-1*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Conditional: Cond_Main;
Sample: Sample-1;
Test sub-sample size: 231,755

Feature [a]	1st Freq. value [b]	1st Freq. (%) [c]	2nd Freq. value [d]	2nd Freq. (%) [e]	Zero or NA (%) [f]
future365_s_spellAdmiMeth_emergency_freq_bool	1	82,434 (35.57%)	0	149,321 (64.43%)	149,321 (64.43%)
mainspefRecoded_15	25	1 (0.00%)	26	1 (0.00%)	190,388 (82.15%)
mainspefRecoded_12	25	1 (0.00%)	27	1 (0.00%)	192,050 (82.87%)
mainspefRecoded_09	49	1 (0.00%)	51	1 (0.00%)	141,647 (61.12%)
mainspefRecoded_11	26	1 (0.00%)	30	1 (0.00%)	205,687 (88.75%)

Continued on next page

Feature [a]	**1st Freq. value** [b]	**1st Freq. (%)** [c]	**2nd Freq. value** [d]	**2nd Freq. (%)** [e]	**Zero or NA (%)** [f]
mainspefRecoded_10	29	1 (0.00%)	33	1 (0.00%)	181,564 (78.34%)
mainspefRecoded_27	15	1 (0.00%)	19	1 (0.00%)	192,258 (82.96%)
mainspefRecoded_01	12	1 (0.00%)	15	1 (0.00%)	223,542 (96.46%)
mainspefRecoded_28	36	1 (0.00%)	37	1 (0.00%)	221,672 (95.65%)
mainspefRecoded_03	20	1 (0.00%)	22	1 (0.00%)	217,199 (93.72%)
mainspefRecoded_19	13	1 (0.00%)	16	1 (0.00%)	220,079 (94.96%)
mainspefRecoded_33	21	1 (0.00%)	25	1 (0.00%)	218,885 (94.45%)
mainspefRecoded_08	16	2 (0.00%)	20	2 (0.00%)	221,289 (95.48%)
mainspefRecoded_06	13	1 (0.00%)	14	1 (0.00%)	222,155 (95.86%)
mainspefRecoded_31	12	1 (0.00%)	18	1 (0.00%)	226,922 (97.91%)
triggerStartAge	21	1,132 (0.49%)	2	5,395 (2.33%)	0 (0.00%)
gender_2	0	89,367 (38.56%)	1	142,388 (61.44%)	89,367 (38.56%)
ethnosRecoded_0	1	52,998 (22.87%)	0	178,757 (77.13%)	178,757 (77.13%)
ethnosRecoded_1	0	74,293 (32.06%)	1	157,462 (67.94%)	74,293 (32.06%)
imd04rkRecoded	0	4,425 (1.91%)	10	15,310 (6.61%)	4,425 (1.91%)

[a] Feature: Name of the defined feature.
[b] 1st Freq. value: The value of the first most frequent value.
[c] 1st Freq. (%): The frequency and percentage of the first most frequent value.
[d] 2nd Freq. value: The value of the second most frequent value.
[e] 2nd Freq. (%): The frequency and percentage of the second most frequent value.
[f] Zero or NA (%): The frequency and percentage of zeros or NAs.

A.5.2.2 Sample-2

TABLE A.16: Descriptive statistics of the continuous variables (*sample-2*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Conditional: Cond_Main;
Sample: Sample-2;
Test sub-sample size: 243,712

Feature	Mean	Stnd. Devi.	Min	Q. 1	Q. 2	Q. 3	Max	Zero or NA (%)
readmiGap_avg__trigger	1.00	0.00	1.00	1.00	1.00	1.00	1.00	0 (0.00%)
mainspefRecoded_freq1__trigger_15	0.15	0.35	0.00	0.00	0.00	0.00	1.00	208,080 (85.38%)
mainspefRecoded_freq1__trigger_12	0.10	0.30	0.00	0.00	0.00	0.00	1.00	220,231 (90.37%)
mainspefRecoded_freq1__trigger_09	0.21	0.41	0.00	0.00	0.00	0.00	1.00	191,420 (78.54%)
mainspefRecoded_freq1__trigger_11	0.07	0.26	0.00	0.00	0.00	0.00	1.00	225,592 (92.56%)
mainspefRecoded_freq1__trigger_10	0.09	0.29	0.00	0.00	0.00	0.00	1.00	221,150 (90.74%)
mainspefRecoded_freq1__trigger_27	0.08	0.27	0.00	0.00	0.00	0.00	1.00	225,139 (92.38%)
mainspefRecoded_freq1__trigger_01	0.09	0.28	0.00	0.00	0.00	0.00	1.00	222,266 (91.20%)
mainspefRecoded_freq1__trigger_28	0.01	0.12	0.00	0.00	0.00	0.00	1.00	240,209 (98.56%)
mainspefRecoded_freq1__trigger_03	0.04	0.19	0.00	0.00	0.00	0.00	1.00	234,553 (96.24%)
mainspefRecoded_freq1__trigger_19	0.00	0.06	0.00	0.00	0.00	0.00	1.00	242,790 (99.62%)
mainspefRecoded_freq1__trigger_33	0.02	0.12	0.00	0.00	0.00	0.00	1.00	239,933 (98.45%)
mainspefRecoded_freq1__trigger_08	0.02	0.14	0.00	0.00	0.00	0.00	1.00	238,640 (97.92%)
mainspefRecoded_freq1__trigger_06	0.01	0.11	0.00	0.00	0.00	0.00	1.00	240,609 (98.73%)
mainspefRecoded_freq1__trigger_31	0.02	0.15	0.00	0.00	0.00	0.00	1.00	238,013 (97.66%)
s_spellAdmiMeth_acute_freq__delta	1.34	1.02	1.00	1.00	1.00	1.00	43.00	81,054 (33.26%)
s_spellAdmiMeth_acute_freq__365days__delta	-0.28	0.87	-36.00	0.00	0.00	0.00	0.00	217,356 (89.19%)
spellOpertn_freq__90days__delta	2.24	8.86	-1.00	0.00	1.00	3.00	1,264.00	114,597 (47.02%)
posopdur_avg__trigger	1.81	7.12	0.00	0.00	0.00	1.00	388.00	168,557 (69.16%)
epidurRecoded_avg	2.34	2.16	0.00	0.75	1.75	3.43	10.00	32,721 (13.43%)
charlsonIndex_avg	2.28	4.64	0.00	0.00	0.00	4.00	103.00	172,695 (70.86%)
readmiGap_avg	0.70	0.46	0.00	0.00	1.00	1.00	1.00	73,536 (30.17%)
epidur_maxAvg__trigger	4.79	16.66	0.00	0.00	1.00	4.00	1,125.00	70,135 (28.78%)
s_spellAdmiMeth_acute_freq__90days__delta	1.93	5.29	-1.00	0.00	1.00	3.00	1,050.00	121,256 (49.75%)
posopdur_avg	2.29	6.90	0.00	0.00	0.75	2.00	974.00	97,987 (40.21%)
charlsonIndex_max__365days	2.30	5.15	0.00	0.00	0.00	3.00	94.00	201,079 (82.51%)
preopdur_avg__trigger	0.61	2.94	0.00	0.00	0.00	0.00	452.00	197,116 (80.88%)
reference_freq__90days__sum	0.54	3.73	0.00	0.00	0.00	0.00	1,241.00	205,652 (84.38%)
s_spellAdmiMeth_elective_freq__delta	0.18	1.36	0.00	0.00	0.00	0.00	160.00	227,060 (93.17%)
preopdur_avg	0.73	3.04	0.00	0.00	0.00	0.67	452.00	140,606 (57.69%)

Continued on next page

Feature	Mean	Stnd. Devi.	Min	Q. 1	Q. 2	Q. 3	Max	Zero or NA (%)
posopdur_avg__365days	1.99	6.97	0.00	0.00	0.67	2.00	1,095.00	147,231 (60.41%)
diagRisk_9_external_freq__90days__sum	0.66	2.22	0.00	0.00	0.00	1.00	565.00	189,268 (77.66%)
intmanigRecoded_other_freq__90days	0.27	1.63	0.00	0.00	0.00	0.00	503.00	218,756 (89.76%)
mainspef_uniques_freq__trigger	1.13	0.38	0.00	1.00	1.00	1.00	7.00	242 (0.10%)
epidur_maxAvg__365days	6.57	117.85	0.00	0.50	1.67	4.00	14,906.00	109,894 (45.09%)
diagCat_3_freq__90days__sum	0.85	3.65	0.00	0.00	0.00	0.00	522.00	215,088 (88.25%)
diagCci_06_cpd_freq__trigger	0.13	0.47	0.00	0.00	0.00	0.00	9.00	220,743 (90.58%)
diagCat_5_freq__90days__sum	0.60	2.07	0.00	0.00	0.00	0.00	146.00	214,611 (88.06%)
diagRisk_3_blood_freq__90days__sum	1.15	5.30	0.00	0.00	0.00	1.00	1,080.00	189,201 (77.63%)
spellOpertn_freq__365days__delta	-0.21	1.20	-124.00	0.00	0.00	0.00	0.00	222,013 (91.10%)
preopdur_avg__365days	0.60	2.96	0.00	0.00	0.00	0.50	366.00	178,417 (73.21%)
diagOther_4_chronic_e_freq__90days__sum	0.46	2.49	0.00	0.00	0.00	0.00	148.00	230,886 (94.74%)
orgCluster_freq1__trigger_1	0.22	0.41	0.00	0.00	0.00	0.00	1.00	190,975 (78.36%)
orgCluster_freq1__trigger_3	0.35	0.48	0.00	0.00	0.00	1.00	1.00	158,425 (65.01%)
orgCluster_freq1__trigger_5	0.11	0.31	0.00	0.00	0.00	0.00	1.00	216,738 (88.93%)
orgCluster_freq1__trigger_4	0.27	0.44	0.00	0.00	0.00	1.00	1.00	178,182 (73.11%)
rotreatRecoded_freq1__trigger_6	0.16	0.36	0.00	0.00	0.00	0.00	1.00	205,874 (84.47%)
rotreatRecoded_freq1__trigger_8	0.10	0.30	0.00	0.00	0.00	0.00	1.00	219,975 (90.26%)
rotreatRecoded_freq1__trigger_7	0.16	0.36	0.00	0.00	0.00	0.00	1.00	205,521 (84.33%)
rotreatRecoded_freq1__trigger_4	0.15	0.36	0.00	0.00	0.00	0.00	1.00	206,636 (84.79%)
rotreatRecoded_freq1__trigger_5	0.10	0.29	0.00	0.00	0.00	0.00	1.00	220,390 (90.43%)
rotreatRecoded_freq1__trigger_3	0.11	0.31	0.00	0.00	0.00	0.00	1.00	217,296 (89.16%)
rotreatRecoded_freq1__trigger_1	0.14	0.34	0.00	0.00	0.00	0.00	1.00	210,606 (86.42%)
rotreatRecoded_freq1__trigger_2	0.10	0.30	0.00	0.00	0.00	0.00	1.00	219,686 (90.14%)
diagOther_4_chronic_d_freq__90days__sum	0.58	2.84	0.00	0.00	0.00	0.00	271.00	226,239 (92.83%)
diagOther_8_mental_freq__90days__sum	0.34	1.77	0.00	0.00	0.00	0.00	125.00	225,466 (92.51%)
diagOther_4_chronic_b_freq__90days__sum	0.40	2.97	0.00	0.00	0.00	0.00	924.00	222,651 (91.36%)
s_spellAdmiMeth_elective_freq__90days__delta	2.10	9.13	-1.00	0.00	1.00	2.00	1,092.00	130,912 (53.72%)
diagOther_4_chronic_i_freq__90days__sum	0.58	2.35	0.00	0.00	0.00	0.00	181.00	218,325 (89.58%)
diagCci_44_anemia_freq__90days	0.20	1.48	0.00	0.00	0.00	0.00	119.00	232,957 (95.59%)
oper_2_heart_freq__90days	0.21	0.74	0.00	0.00	0.00	0.00	49.00	223,873 (91.86%)
admisorcRecoded_other_freq	0.18	0.72	0.00	0.00	0.00	0.00	127.00	213,963 (87.79%)
diagOther_7_cancer_freq__90days	0.39	2.39	0.00	0.00	0.00	0.00	184.00	223,215 (91.59%)
protype_freq1_9	0.78	0.41	0.00	1.00	1.00	1.00	1.00	52,920 (21.71%)
protype_freq1__trigger_9	0.63	0.48	0.00	0.00	1.00	1.00	1.00	90,351 (37.07%)

Continued on next page

Feature	Mean	Stnd. Devi.	Min	Q. 1	Q. 2	Q. 3	Max	Zero or NA (%)
diagCci_19_cardiac_freq__90days__sum	0.46	1.83	0.00	0.00	0.00	0.00	119.00	221,671 (90.96%)
diagCci_02_chf_freq__90days	0.15	1.01	0.00	0.00	0.00	0.00	117.00	235,573 (96.66%)
elecdur_elective_nulls_freq	0.23	3.38	0.00	0.00	0.00	0.00	579.00	219,972 (90.26%)
oper_3_urinary_freq__90days	0.21	1.07	0.00	0.00	0.00	0.00	53.00	227,330 (93.28%)
diagCci_26_neuroOther_freq__90days	0.24	1.96	0.00	0.00	0.00	0.00	180.00	234,646 (96.28%)
diagCci_14_malignancy_freq__90days	0.25	2.17	0.00	0.00	0.00	0.00	184.00	235,517 (96.64%)
diagCci_18_depression_freq__90days	0.12	1.04	0.00	0.00	0.00	0.00	248.00	235,588 (96.67%)
diagRisk_8_smoke_freq__90days	0.18	0.99	0.00	0.00	0.00	0.00	96.00	230,769 (94.69%)
diagRisk_10_externalMorbidity_freq__90days	0.15	0.63	0.00	0.00	0.00	0.00	66.00	227,028 (93.15%)
diagMorbid_1_Influenza_freq__90days	0.07	0.44	0.00	0.00	0.00	0.00	14.00	237,185 (97.32%)
diagRisk_2_Cholesterol_freq__90days	0.16	1.17	0.00	0.00	0.00	0.00	316.00	233,027 (95.62%)
diagRisk_7_kidney_freq__90days	0.26	8.64	0.00	0.00	0.00	0.00	1,100.00	238,491 (97.86%)
diagCci_04_cerebrovascular_freq__90days	0.15	1.35	0.00	0.00	0.00	0.00	381.00	234,758 (96.33%)
diagCat_9_freq__90days	0.10	0.61	0.00	0.00	0.00	0.00	68.00	234,065 (96.04%)
spellOpertn_freq__trigger	0.41	0.49	0.00	0.00	0.00	1.00	1.00	144,058 (59.11%)

TABLE A.17: Descriptive statistics of the discrete variables (*sample-2*))

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Conditional: Cond_Main;
Sample: Sample-2;
Test sub-sample size: 243,712

Feature	1st Frequent - value	1st Frequent - Freq. (%)	2nd Frequent - value	2nd Frequent - Freq. (%)	Zero or NA (%)
future365_s_spellAdmiMeth_emergency_freq_bool	1	91,517 (37.55%)	0	152,195 (62.45%)	152,195 (62.45%)
mainspefRecoded_15	35	1 (0.00%)	39	1 (0.00%)	192,388 (78.94%)
mainspefRecoded_12	34	1 (0.00%)	35	1 (0.00%)	193,074 (79.22%)
mainspefRecoded_09	64	1 (0.00%)	67	1 (0.00%)	140,699 (57.73%)
mainspefRecoded_11	25	1 (0.00%)	28	1 (0.00%)	211,596 (86.82%)
mainspefRecoded_10	34	1 (0.00%)	37	1 (0.00%)	174,328 (71.53%)
mainspefRecoded_27	25	1 (0.00%)	32	1 (0.00%)	190,232 (78.06%)
mainspefRecoded_01	24	1 (0.00%)	26	1 (0.00%)	206,598 (84.77%)
mainspefRecoded_28	31	1 (0.00%)	41	1 (0.00%)	233,989 (96.01%)
mainspefRecoded_03	26	1 (0.00%)	30	1 (0.00%)	217,050 (89.06%)
mainspefRecoded_19	15	1 (0.00%)	20	1 (0.00%)	223,566 (91.73%)

Continued on next page

Feature	1st Frequent - value	1st Frequent - Freq. (%)	2nd Frequent - value	2nd Frequent - Freq. (%)	Zero or NA (%)
mainspefRecoded_33	35	1 (0.00%)	36	1 (0.00%)	224,095 (91.95%)
mainspefRecoded_08	18	1 (0.00%)	23	1 (0.00%)	222,642 (91.35%)
mainspefRecoded_06	13	1 (0.00%)	14	1 (0.00%)	227,980 (93.54%)
mainspefRecoded_31	19	1 (0.00%)	27	1 (0.00%)	231,485 (94.98%)
triggerStartAge	21	657 (0.27%)	2	1,781 (0.73%)	0 (0.00%)
gender_2	0	92,625 (38.01%)	1	151,087 (61.99%)	92,625 (38.01%)
ethnosRecoded_0	1	19,568 (8.03%)	0	224,144 (91.97%)	224,144 (91.97%)
ethnosRecoded_1	0	51,397 (21.09%)	1	192,315 (78.91%)	51,397 (21.09%)
imd04rkRecoded	0	4,749 (1.95%)	10	14,863 (6.10%)	4,749 (1.95%)

A.5.2.3 Sample-3

TABLE A.18: Descriptive statistics of the continuous variables (*sample-3*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Conditional: Cond_Main;
Sample: Sample-3;
Test sub-sample size: 304,888

Feature	Mean	Stnd. Devi.	Min	Q. 1	Q. 2	Q. 3	Max	Zero or NA (%)
readmiGap_avg__trigger	1.00	0.00	1.00	1.00	1.00	1.00	1.00	0 (0.00%)
mainspefRecoded_freq1__trigger_15	0.15	0.35	0.00	0.00	0.00	0.00	1.00	260,347 (85.39%)
mainspefRecoded_freq1__trigger_12	0.09	0.29	0.00	0.00	0.00	0.00	1.00	276,450 (90.67%)
mainspefRecoded_freq1__trigger_09	0.25	0.44	0.00	0.00	0.00	1.00	1.00	227,198 (74.52%)
mainspefRecoded_freq1__trigger_11	0.07	0.26	0.00	0.00	0.00	0.00	1.00	282,281 (92.59%)
mainspefRecoded_freq1__trigger_10	0.10	0.30	0.00	0.00	0.00	0.00	1.00	274,676 (90.09%)
mainspefRecoded_freq1__trigger_27	0.09	0.29	0.00	0.00	0.00	0.00	1.00	277,468 (91.01%)
mainspefRecoded_freq1__trigger_01	0.03	0.17	0.00	0.00	0.00	0.00	1.00	295,674 (96.98%)
mainspefRecoded_freq1__trigger_28	0.02	0.15	0.00	0.00	0.00	0.00	1.00	297,874 (97.70%)
mainspefRecoded_freq1__trigger_03	0.03	0.17	0.00	0.00	0.00	0.00	1.00	296,309 (97.19%)
mainspefRecoded_freq1__trigger_19	0.01	0.07	0.00	0.00	0.00	0.00	1.00	303,246 (99.46%)
mainspefRecoded_freq1__trigger_33	0.02	0.13	0.00	0.00	0.00	0.00	1.00	299,594 (98.26%)
mainspefRecoded_freq1__trigger_08	0.01	0.12	0.00	0.00	0.00	0.00	1.00	300,694 (98.62%)
mainspefRecoded_freq1__trigger_06	0.01	0.12	0.00	0.00	0.00	0.00	1.00	300,608 (98.60%)
mainspefRecoded_freq1__trigger_31	0.01	0.10	0.00	0.00	0.00	0.00	1.00	301,790 (98.98%)
s_spellAdmiMeth_acute_freq__delta	1.37	1.06	1.00	1.00	1.00	1.00	45.00	182,103 (59.73%)
s_spellAdmiMeth_acute_freq__365days__delta	-0.31	0.90	-35.00	0.00	0.00	0.00	0.00	282,671 (92.71%)
spellOpertn_freq__90days__delta	1.05	3.29	-1.00	0.00	1.00	1.00	420.00	216,492 (71.01%)
posopdur_avg__trigger	1.49	7.48	0.00	0.00	0.00	1.00	1,522.00	222,130 (72.86%)
epidurRecoded_avg	2.96	2.67	0.00	1.00	2.00	4.60	10.00	44,351 (14.55%)
charlsonIndex_avg	1.71	4.19	0.00	0.00	0.00	0.00	105.00	240,103 (78.75%)
readmiGap_avg	0.47	0.50	0.00	0.00	0.00	1.00	1.00	161,375 (52.93%)
epidur_maxAvg__trigger	6.53	26.64	0.00	1.00	2.00	5.00	2,534.00	64,924 (21.29%)
s_spellAdmiMeth_acute_freq__90days__delta	0.74	2.37	-1.00	-1.00	0.00	1.00	114.00	216,789 (71.10%)
posopdur_avg	2.02	8.62	0.00	0.00	0.00	1.75	1,581.00	176,085 (57.75%)
charlsonIndex_max__365days	1.82	4.52	0.00	0.00	0.00	0.00	81.00	278,319 (91.29%)
preopdur_avg__trigger	0.51	3.33	0.00	0.00	0.00	0.00	394.00	258,247 (84.70%)
reference_freq__90days__sum	0.46	1.23	0.00	0.00	0.00	0.00	45.00	270,192 (88.62%)
s_spellAdmiMeth_elective_freq__delta	0.19	1.06	0.00	0.00	0.00	0.00	138.00	290,405 (95.25%)
preopdur_avg	0.67	3.71	0.00	0.00	0.00	0.33	672.00	223,418 (73.28%)

Continued on next page

Feature	Mean	Stnd. Devi.	Min	Q. 1	Q. 2	Q. 3	Max	Zero or NA (%)
posopdur_avg__365days	2.02	8.86	0.00	0.00	0.00	2.00	1,581.00	250,693 (82.22%)
diagRisk_9_external_freq__90days__sum	0.39	0.98	0.00	0.00	0.00	0.00	37.00	272,445 (89.36%)
intmanigRecoded_other_freq__90days	0.31	0.90	0.00	0.00	0.00	0.00	41.00	276,185 (90.59%)
mainspef_uniques_freq__trigger	1.10	0.32	0.00	1.00	1.00	1.00	5.00	125 (0.04%)
epidur_maxAvg__365days	6.23	28.37	0.00	0.40	2.00	5.00	3,567.00	210,668 (69.10%)
diagCat_3_freq__90days__sum	0.55	1.83	0.00	0.00	0.00	0.00	74.00	281,257 (92.25%)
diagCci_06_cpd_freq__trigger	0.11	0.41	0.00	0.00	0.00	0.00	9.00	280,777 (92.09%)
diagCat_5_freq__90days__sum	0.42	1.71	0.00	0.00	0.00	0.00	386.00	282,768 (92.74%)
diagRisk_3_blood_freq__90days__sum	0.60	2.19	0.00	0.00	0.00	0.00	438.00	271,498 (89.05%)
spellOpertn_freq__365days__delta	-0.21	0.90	-118.00	0.00	0.00	0.00	0.00	287,896 (94.43%)
preopdur_avg__365days	0.62	4.44	0.00	0.00	0.00	0.50	730.00	270,914 (88.86%)
diagOther_4_chronic_e_freq__90days__sum	0.34	1.65	0.00	0.00	0.00	0.00	60.00	294,166 (96.48%)
orgCluster_freq1__trigger_1	0.18	0.39	0.00	0.00	0.00	0.00	1.00	249,397 (81.80%)
orgCluster_freq1__trigger_3	0.35	0.48	0.00	0.00	0.00	1.00	1.00	197,841 (64.89%)
orgCluster_freq1__trigger_5	0.11	0.31	0.00	0.00	0.00	0.00	1.00	272,709 (89.45%)
orgCluster_freq1__trigger_4	0.26	0.44	0.00	0.00	0.00	1.00	1.00	224,737 (73.71%)
rotreatRecoded_freq1__trigger_6	0.15	0.35	0.00	0.00	0.00	0.00	1.00	260,452 (85.43%)
rotreatRecoded_freq1__trigger_8	0.10	0.30	0.00	0.00	0.00	0.00	1.00	274,503 (90.03%)
rotreatRecoded_freq1__trigger_7	0.15	0.36	0.00	0.00	0.00	0.00	1.00	257,673 (84.51%)
rotreatRecoded_freq1__trigger_4	0.15	0.36	0.00	0.00	0.00	0.00	1.00	259,273 (85.04%)
rotreatRecoded_freq1__trigger_5	0.10	0.30	0.00	0.00	0.00	0.00	1.00	275,356 (90.31%)
rotreatRecoded_freq1__trigger_3	0.11	0.31	0.00	0.00	0.00	0.00	1.00	271,138 (88.93%)
rotreatRecoded_freq1__trigger_1	0.14	0.35	0.00	0.00	0.00	0.00	1.00	262,141 (85.98%)
rotreatRecoded_freq1__trigger_2	0.10	0.30	0.00	0.00	0.00	0.00	1.00	273,682 (89.76%)
diagOther_4_chronic_d_freq__90days__sum	0.36	1.60	0.00	0.00	0.00	0.00	89.00	291,871 (95.73%)
diagOther_8_mental_freq__90days__sum	0.21	0.99	0.00	0.00	0.00	0.00	42.00	293,164 (96.15%)
diagOther_4_chronic_b_freq__90days__sum	0.23	1.06	0.00	0.00	0.00	0.00	72.00	291,625 (95.65%)
s_spellAdmiMeth_elective_freq__90days__delta	1.05	3.43	-1.00	0.00	1.00	1.00	417.00	231,690 (75.99%)
diagOther_4_chronic_i_freq__90days__sum	0.37	1.69	0.00	0.00	0.00	0.00	386.00	285,976 (93.80%)
diagCci_44_anemia_freq__90days	0.11	0.75	0.00	0.00	0.00	0.00	39.00	298,514 (97.91%)
oper_2_heart_freq__90days	0.10	0.47	0.00	0.00	0.00	0.00	44.00	294,784 (96.69%)
admisorcRecoded_other_freq	0.08	0.37	0.00	0.00	0.00	0.00	26.00	284,946 (93.46%)
diagOther_7_cancer_freq__90days	0.25	1.82	0.00	0.00	0.00	0.00	130.00	293,346 (96.21%)
protype_freq1_9	0.99	0.11	0.00	1.00	1.00	1.00	1.00	3,796 (1.25%)
protype_freq1__trigger_9	0.99	0.11	0.00	1.00	1.00	1.00	1.00	4,055 (1.33%)

Continued on next page

Feature	Mean	Stnd. Devi.	Min	Q. 1	Q. 2	Q. 3	Max	Zero or NA (%)
diagCci_19_cardiac_freq__90days__sum	0.30	1.59	0.00	0.00	0.00	0.00	386.00	289,068 (94.81%)
diagCci_02_chf_freq__90days	0.11	0.71	0.00	0.00	0.00	0.00	54.00	298,127 (97.78%)
elecdur_elective_nulls_freq	0.09	1.28	0.00	0.00	0.00	0.00	320.00	291,479 (95.60%)
oper_3_urinary_freq__90days	0.11	0.58	0.00	0.00	0.00	0.00	31.00	295,865 (97.04%)
diagCci_26_neuroOther_freq__90days	0.14	0.95	0.00	0.00	0.00	0.00	61.00	298,212 (97.81%)
diagCci_14_malignancy_freq__90days	0.17	1.69	0.00	0.00	0.00	0.00	130.00	299,644 (98.28%)
diagCci_18_depression_freq__90days	0.06	0.47	0.00	0.00	0.00	0.00	24.00	300,428 (98.54%)
diagRisk_8_smoke_freq__90days	0.08	0.54	0.00	0.00	0.00	0.00	50.00	298,734 (97.98%)
diagRisk_10_externalMorbidity_freq__90days	0.08	0.56	0.00	0.00	0.00	0.00	146.00	297,117 (97.45%)
diagMorbid_1_Influenza_freq__90days	0.04	0.29	0.00	0.00	0.00	0.00	11.00	301,405 (98.86%)
diagRisk_2_Cholesterol_freq__90days	0.06	0.47	0.00	0.00	0.00	0.00	31.00	300,515 (98.57%)
diagRisk_7_kidney_freq__90days	0.10	2.80	0.00	0.00	0.00	0.00	388.00	301,843 (99.00%)
diagCci_04_cerebrovascular_freq__90days	0.09	0.67	0.00	0.00	0.00	0.00	127.00	299,073 (98.09%)
diagCat_9_freq__90days	0.05	0.44	0.00	0.00	0.00	0.00	67.00	300,322 (98.50%)
spellOpertn_freq__trigger	0.34	0.47	0.00	0.00	0.00	1.00	1.00	200,954 (65.91%)

TABLE A.19: Descriptive statistics of the discrete variables (*sample-3*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Conditional: Cond_Main;
Sample: Sample-3;
Test sub-sample size: 304,888

Feature	1st Frequent - value	1st Frequent - Freq. (%)	2nd Frequent - value	2nd Frequent - Freq. (%)	Zero or NA (%)
future365_s_spellAdmiMeth_emergency_freq_bool	1	109,045 (35.77%)	0	195,843 (64.23%)	195,843 (64.23%)
mainspefRecoded_15	22	1 (0.00%)	23	1 (0.00%)	253,385 (83.11%)
mainspefRecoded_12	20	1 (0.00%)	21	1 (0.00%)	260,674 (85.50%)
mainspefRecoded_09	37	1 (0.00%)	49	1 (0.00%)	191,162 (62.70%)
mainspefRecoded_11	19	1 (0.00%)	25	1 (0.00%)	273,361 (89.66%)
mainspefRecoded_10	28	1 (0.00%)	31	1 (0.00%)	247,875 (81.30%)
mainspefRecoded_27	13	1 (0.00%)	17	1 (0.00%)	259,718 (85.18%)
mainspefRecoded_01	20	1 (0.00%)	23	1 (0.00%)	291,065 (95.47%)
mainspefRecoded_28	27	1 (0.00%)	35	1 (0.00%)	294,023 (96.44%)
mainspefRecoded_03	16	1 (0.00%)	21	1 (0.00%)	287,075 (94.16%)
mainspefRecoded_19	14	1 (0.00%)	15	1 (0.00%)	292,860 (96.05%)

Continued on next page

Feature	1st Frequent - value	1st Frequent - Freq. (%)	2nd Frequent - value	2nd Frequent - Freq. (%)	Zero or NA (%)
mainspefRecoded_33	20	1 (0.00%)	22	1 (0.00%)	291,069 (95.47%)
mainspefRecoded_08	20	1 (0.00%)	24	1 (0.00%)	292,512 (95.94%)
mainspefRecoded_06	10	1 (0.00%)	15	1 (0.00%)	295,175 (96.81%)
mainspefRecoded_31	19	1 (0.00%)	20	1 (0.00%)	299,055 (98.09%)
triggerStartAge	21	1,052 (0.35%)	3	8,361 (2.74%)	0 (0.00%)
gender_2	0	118,670 (38.92%)	1	186,218 (61.08%)	118,670 (38.92%)
ethnosRecoded_0	1	64,965 (21.31%)	0	239,923 (78.69%)	239,923 (78.69%)
ethnosRecoded_1	0	93,555 (30.69%)	1	211,333 (69.31%)	93,555 (30.69%)
imd04rkRecoded	0	5,654 (1.85%)	10	20,742 (6.80%)	5,654 (1.85%)

A.5.3 Distribution Plots of HSCIC-CCI for Diagnoses Groups

In this section, the box-plot statistics for the Charlson and Elixhauser diagnoses categories are plotted. Each plot represents one diagnosis category in the Charlson or Elixhauser comorbidity indices. The plots visually represent the bias of the HSCIC Charlson Comorbidity Index (HSCIC-CCI).

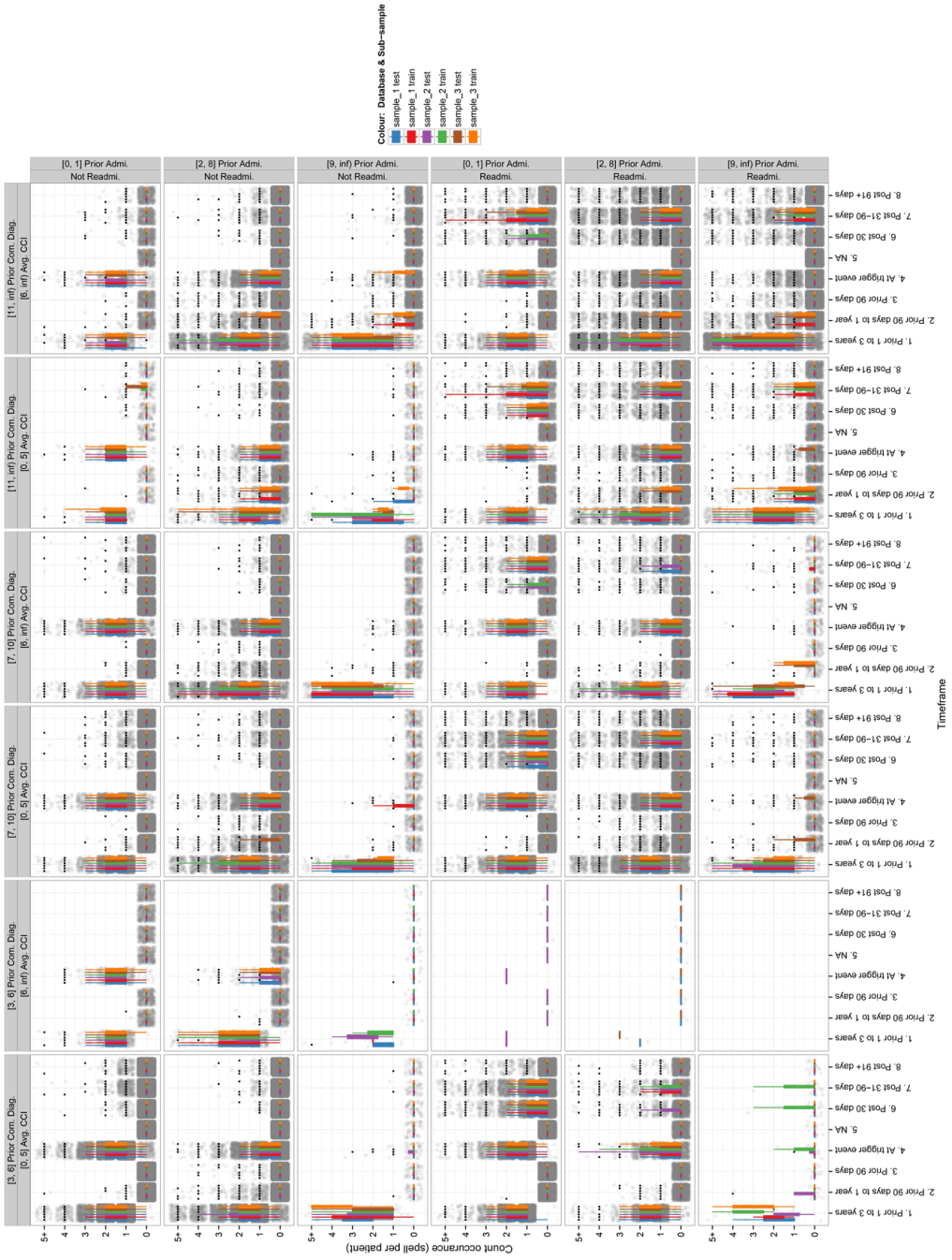

FIGURE A.4: The box-plot statistics of myocardial infarction (all samples)

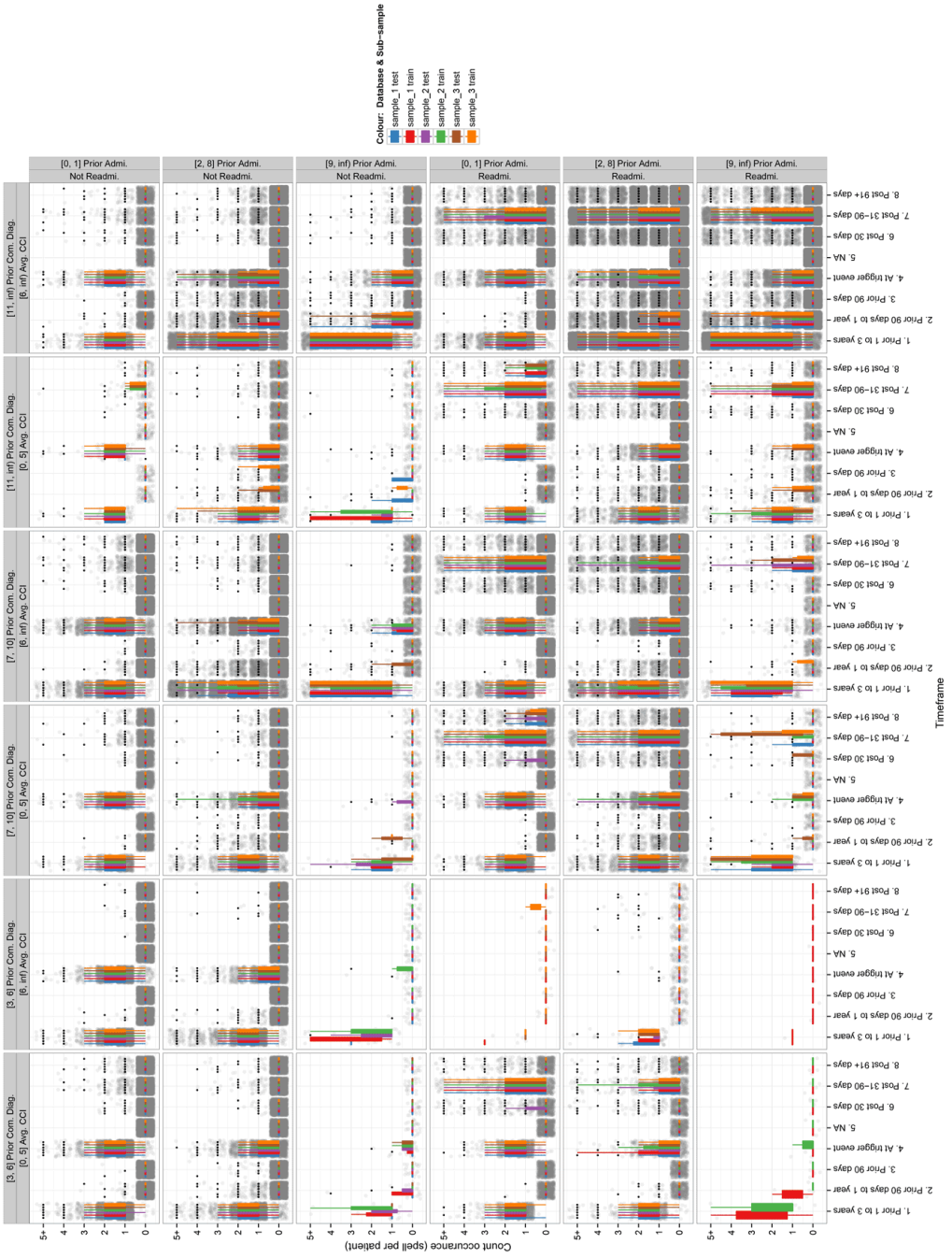

FIGURE A.5: The box-plot statistics of congestive heart failure (all samples)

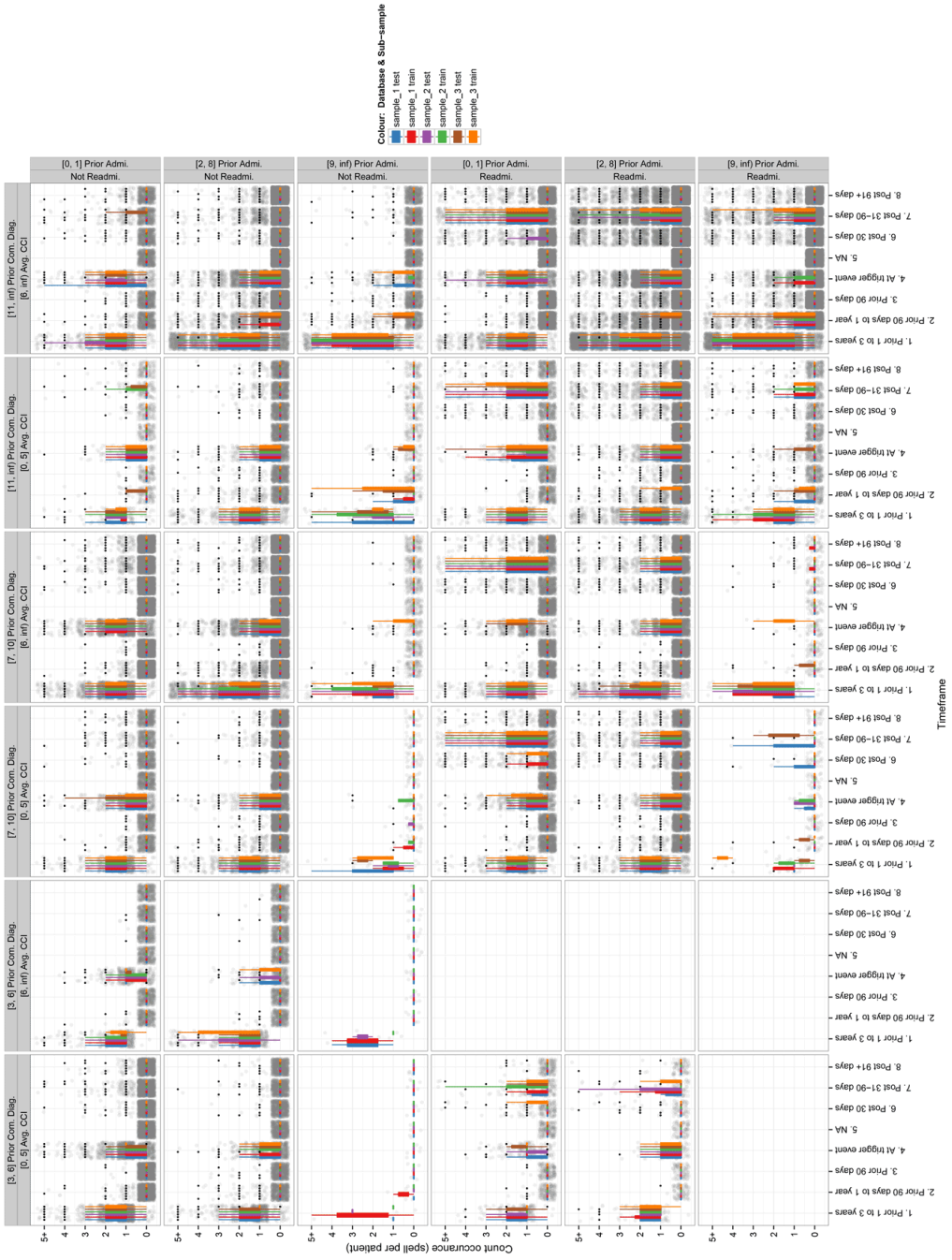

FIGURE A.6: The box-plot statistics of peripheral vascular disease (all samples)

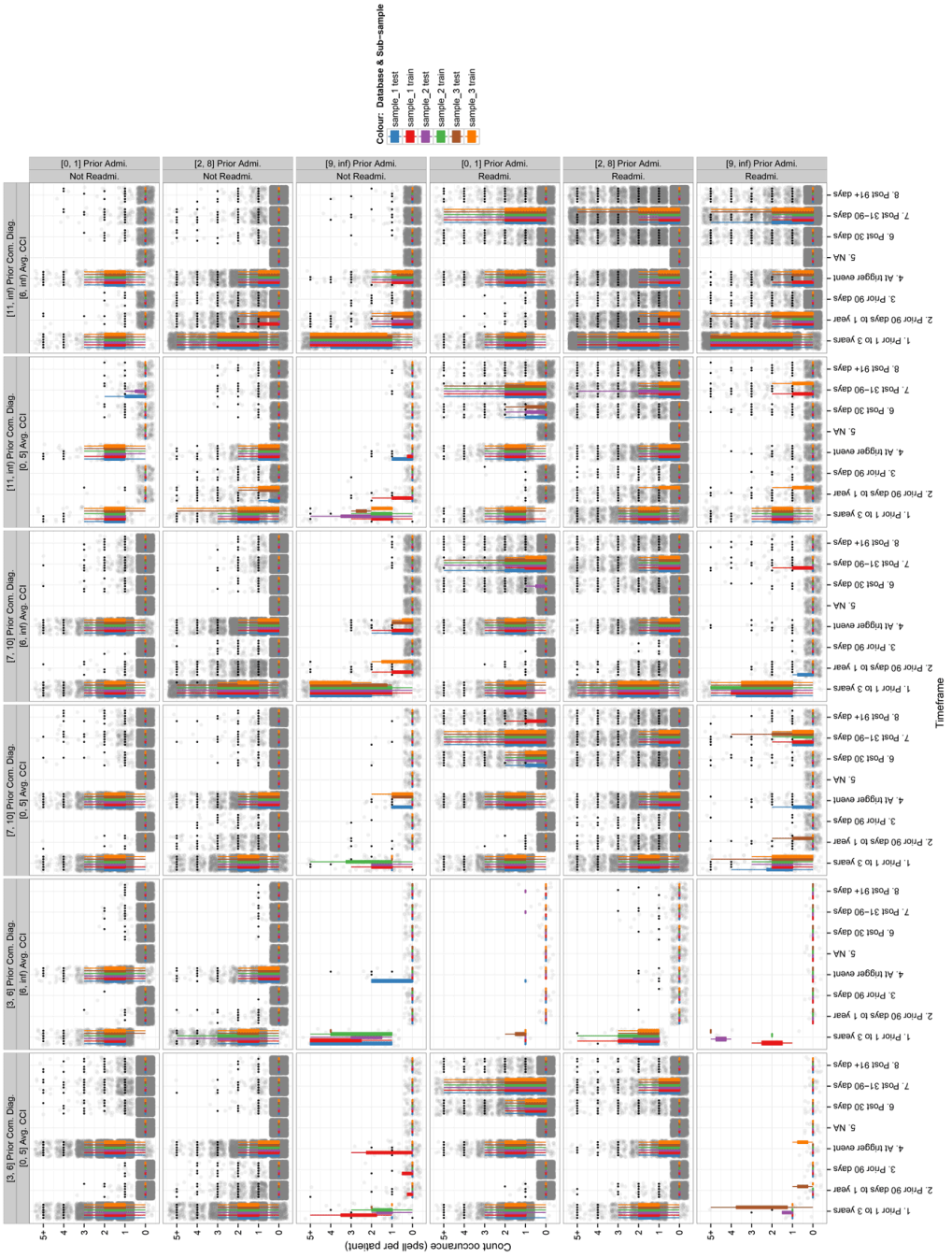

FIGURE A.7: The box-plot statistics of cerebrovascular disease (all samples)

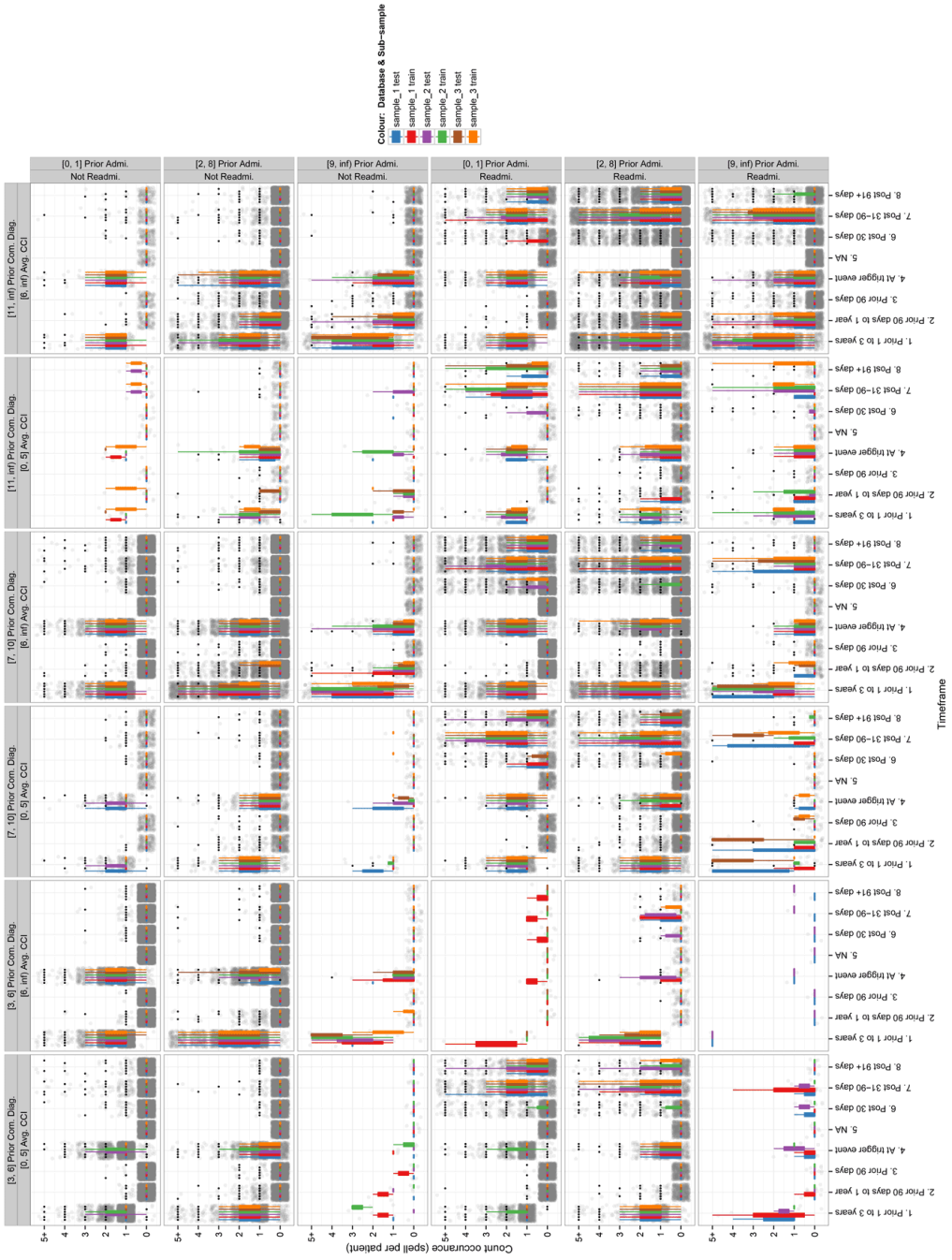

FIGURE A.8: The box-plot statistics of dementia (all samples)

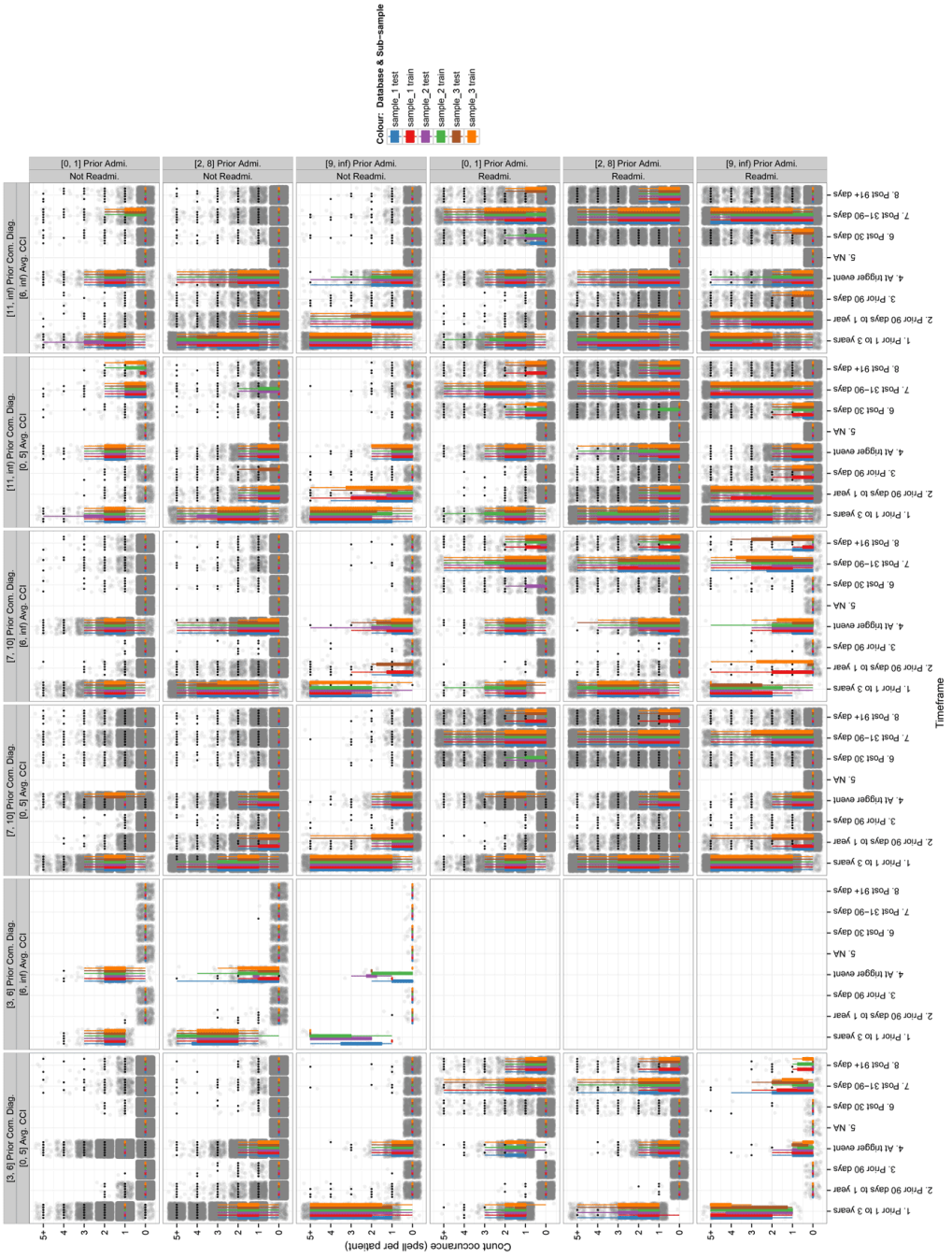

FIGURE A.9: The box-plot statistics of chronic pulmonary disease (all samples)

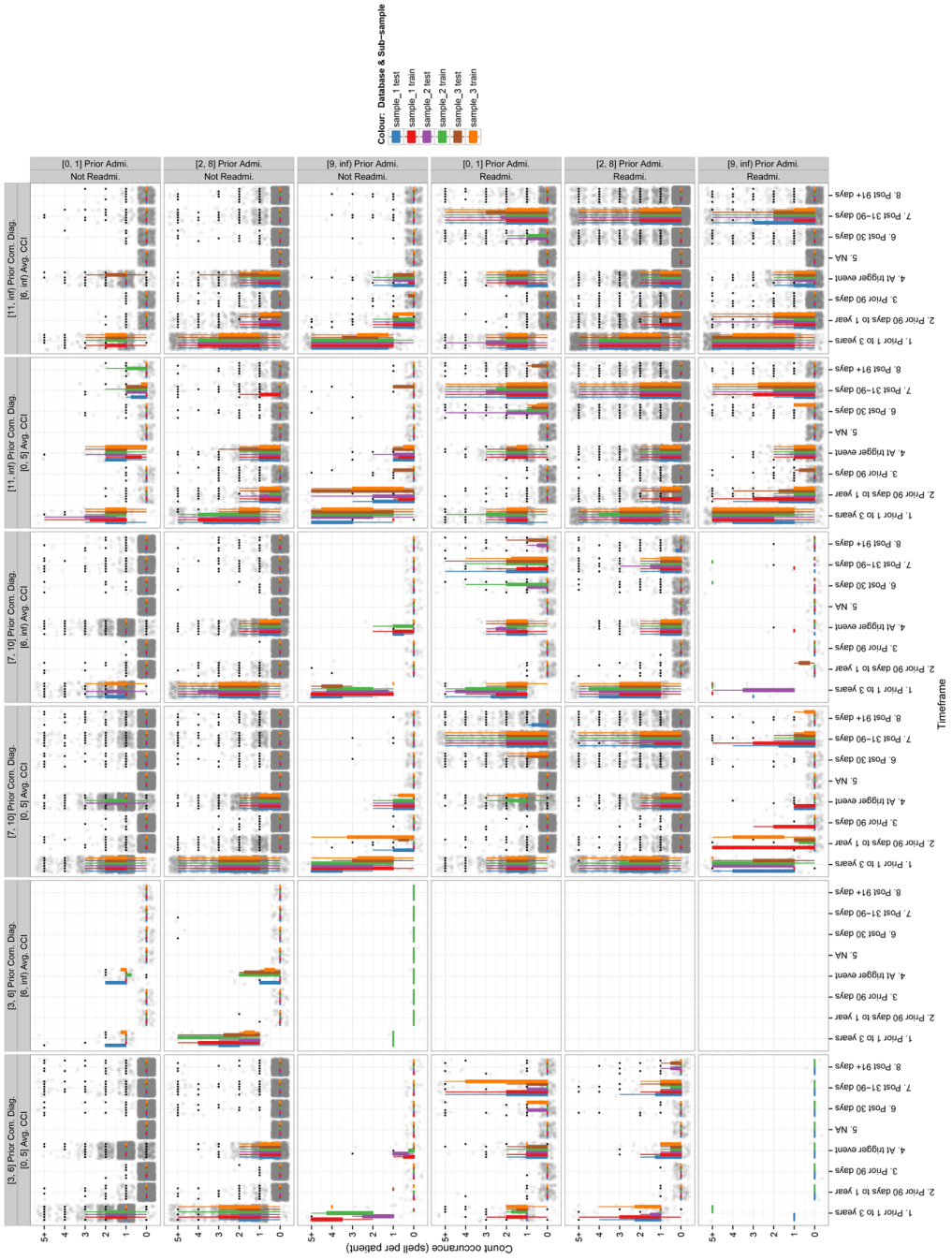

FIGURE A.10: The box-plot statistics of rheumatic disease (all samples)

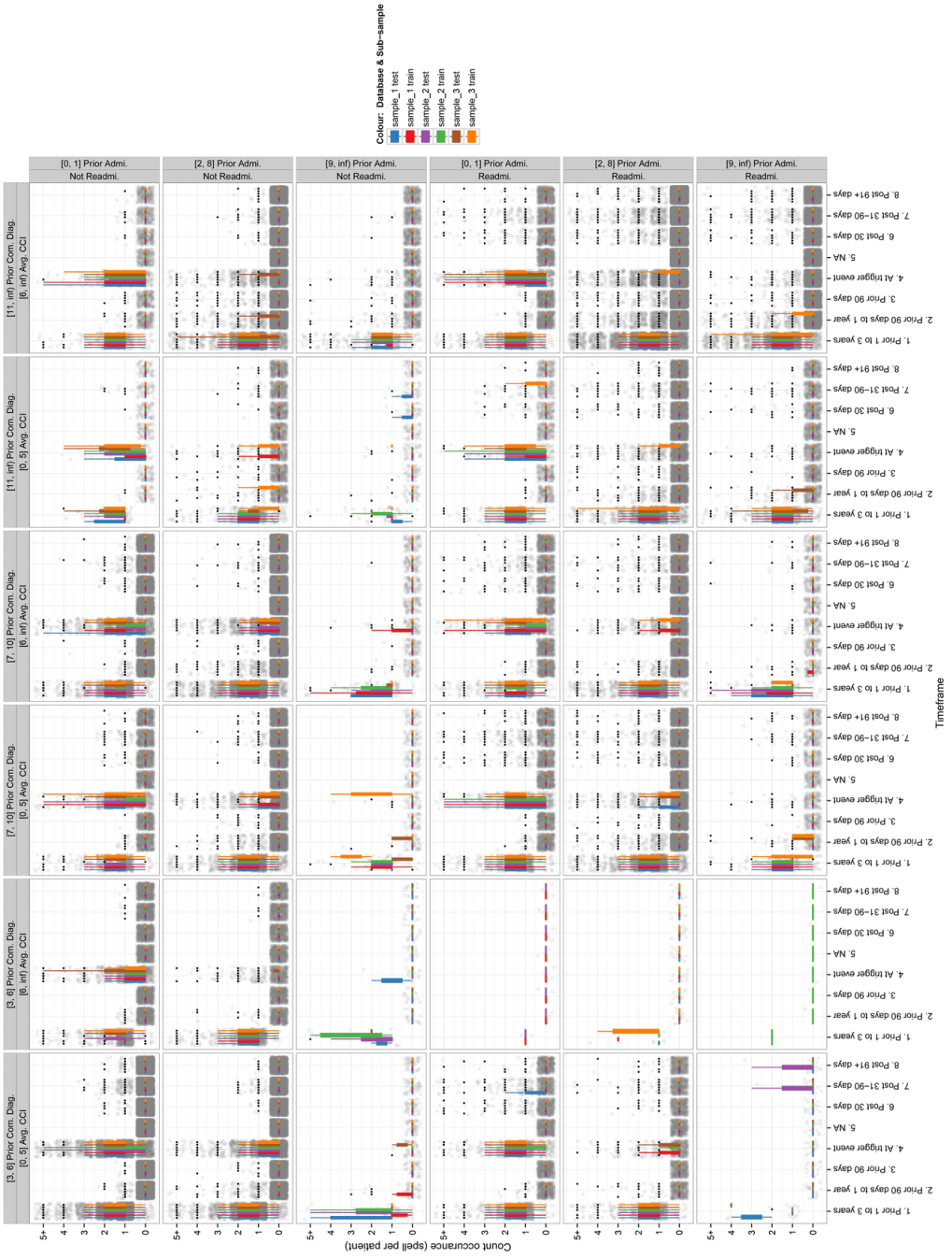

FIGURE A.11: The box-plot statistics of peptic ulcer disease (all samples)

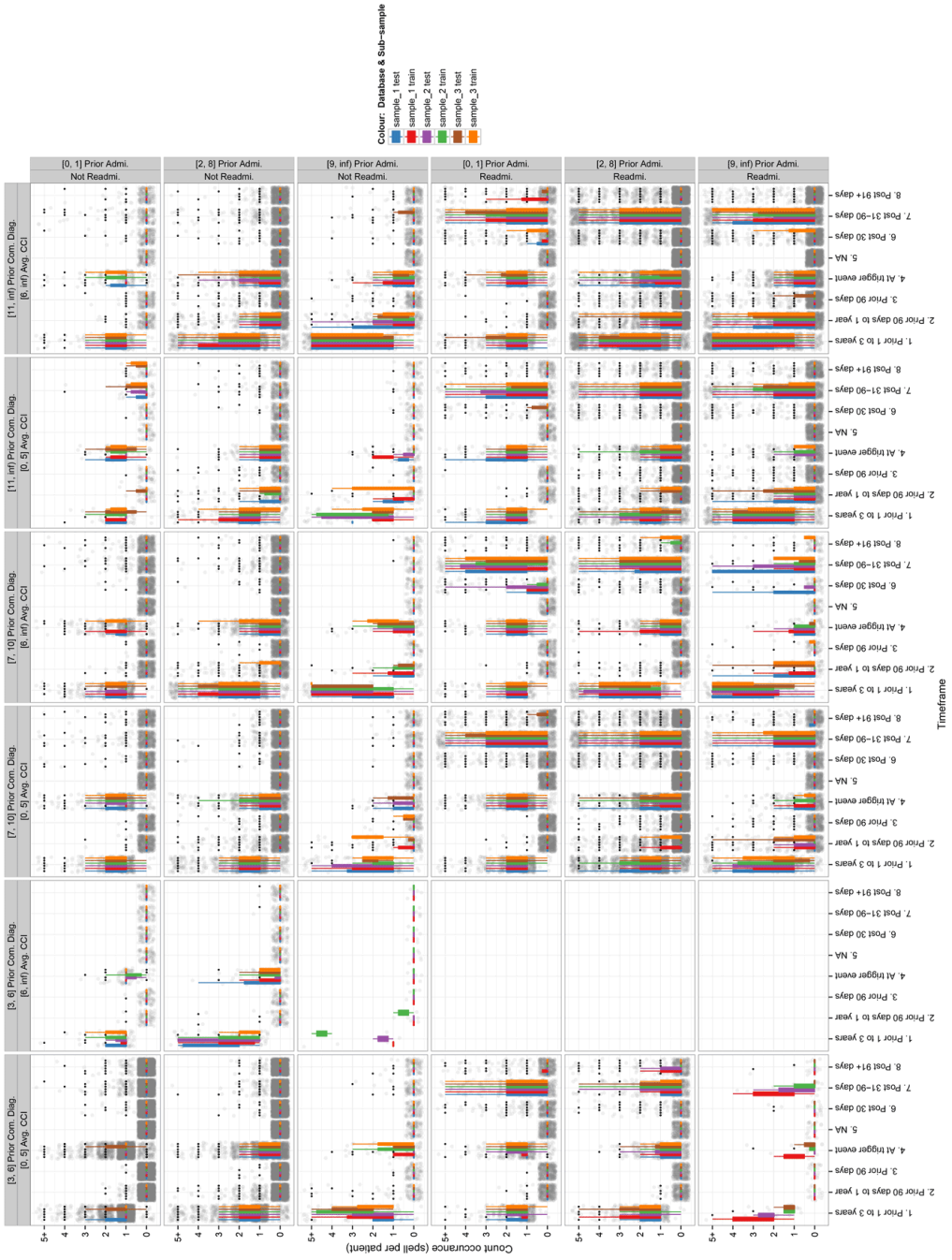

FIGURE A.12: The box-plot statistics of mild liver disease (all samples)

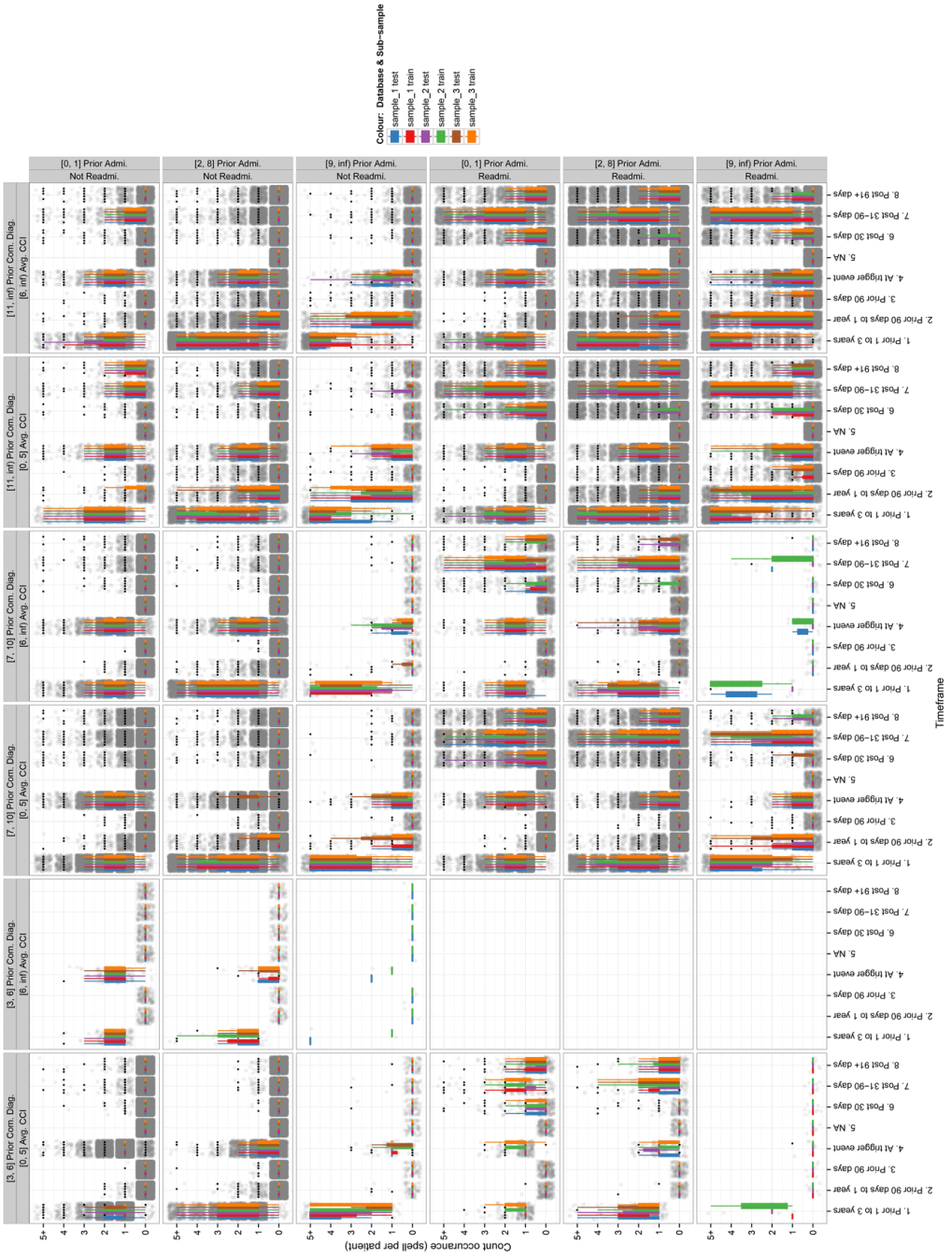

FIGURE A.13: The box-plot statistics of diabetes without chronic complication (all samples)

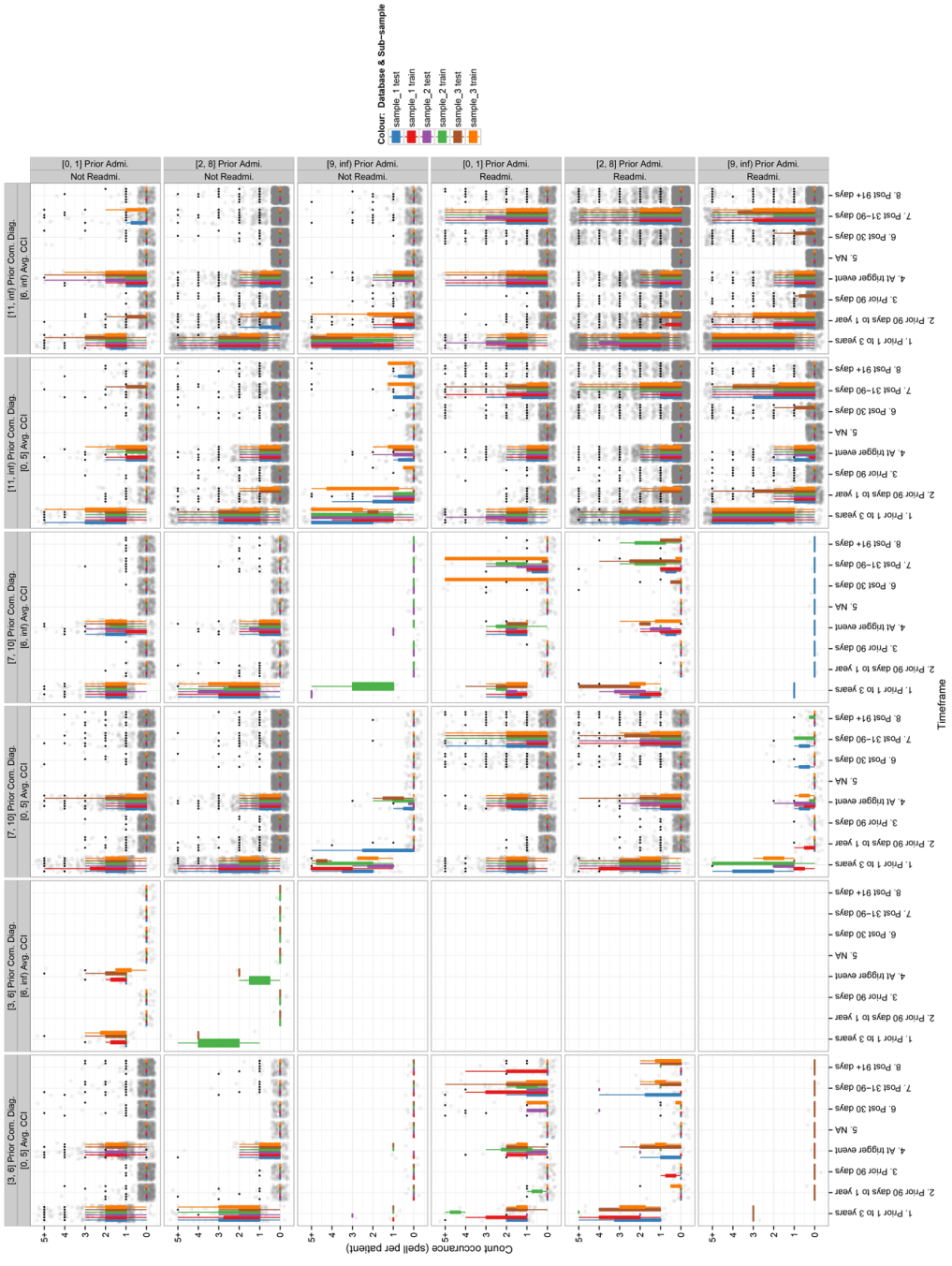

FIGURE A.14: The box-plot statistics of diabetes with chronic complication (all samples)

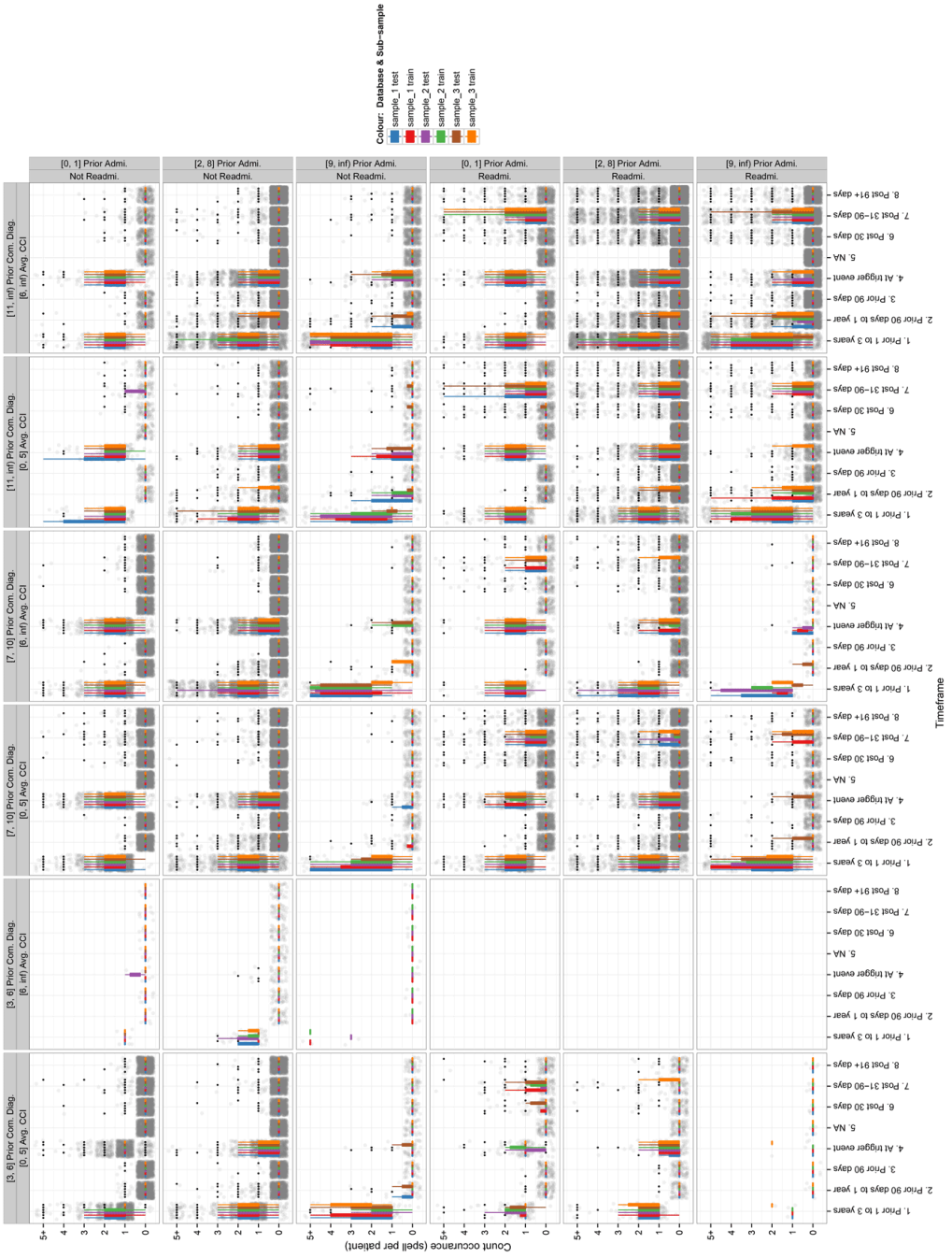

FIGURE A.15: The box-plot statistics of hemiplegia or paraplegia (all samples)

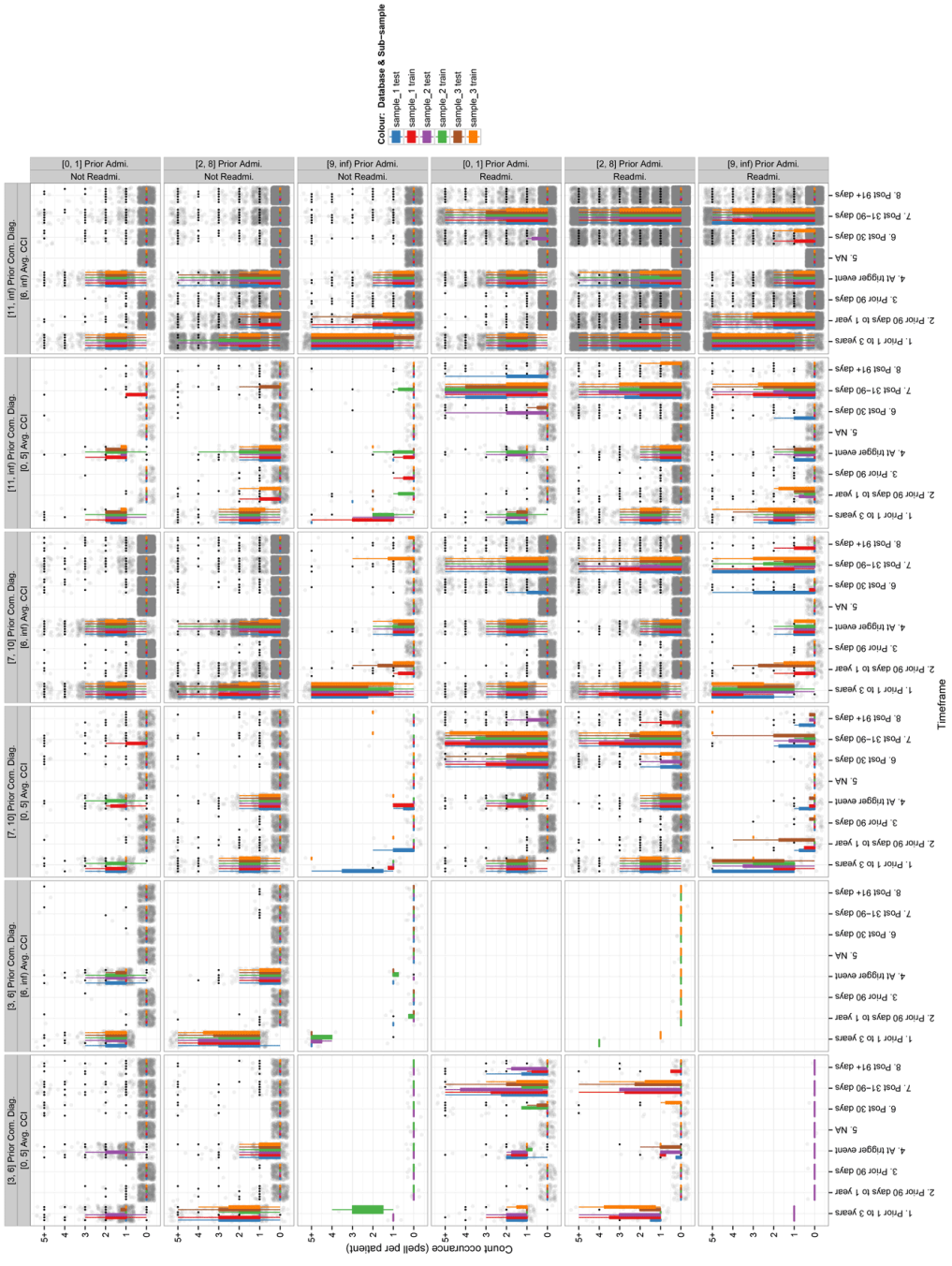

FIGURE A.16: The box-plot statistics of renal disease (all samples)

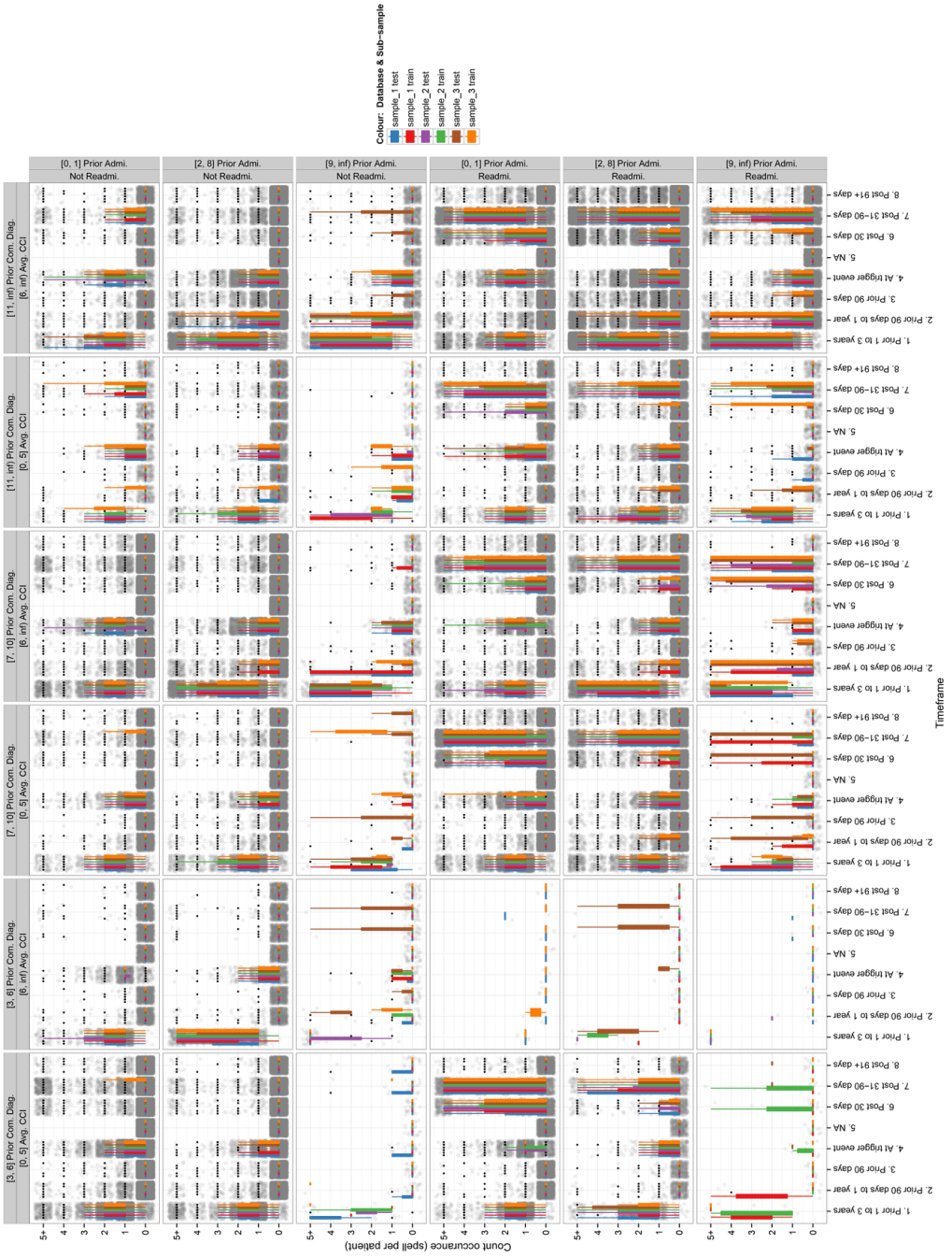

FIGURE A.17: The box-plot statistics of any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin (all samples)

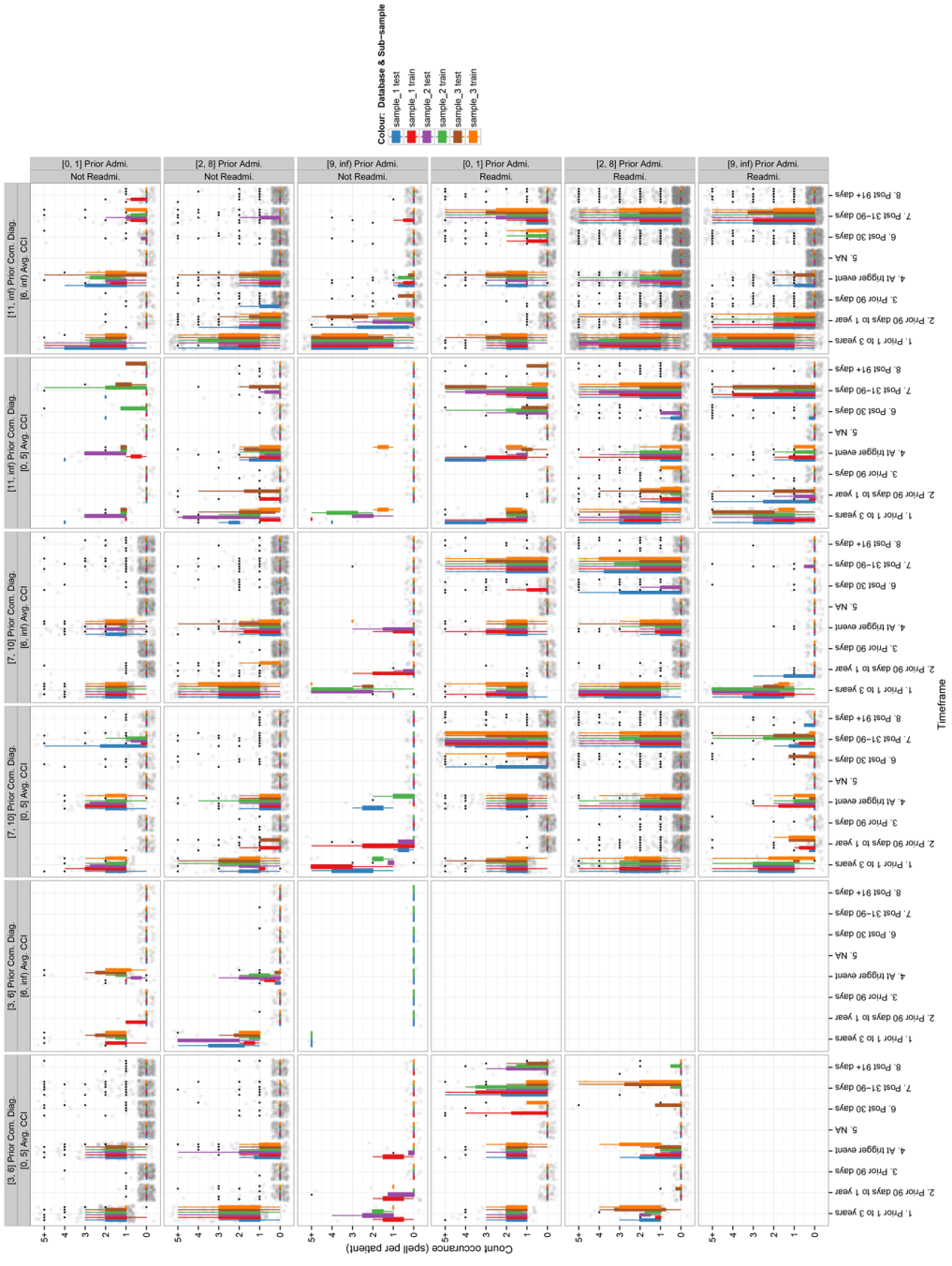

FIGURE A.18: The box-plot statistics of moderate or severe liver disease (all samples)

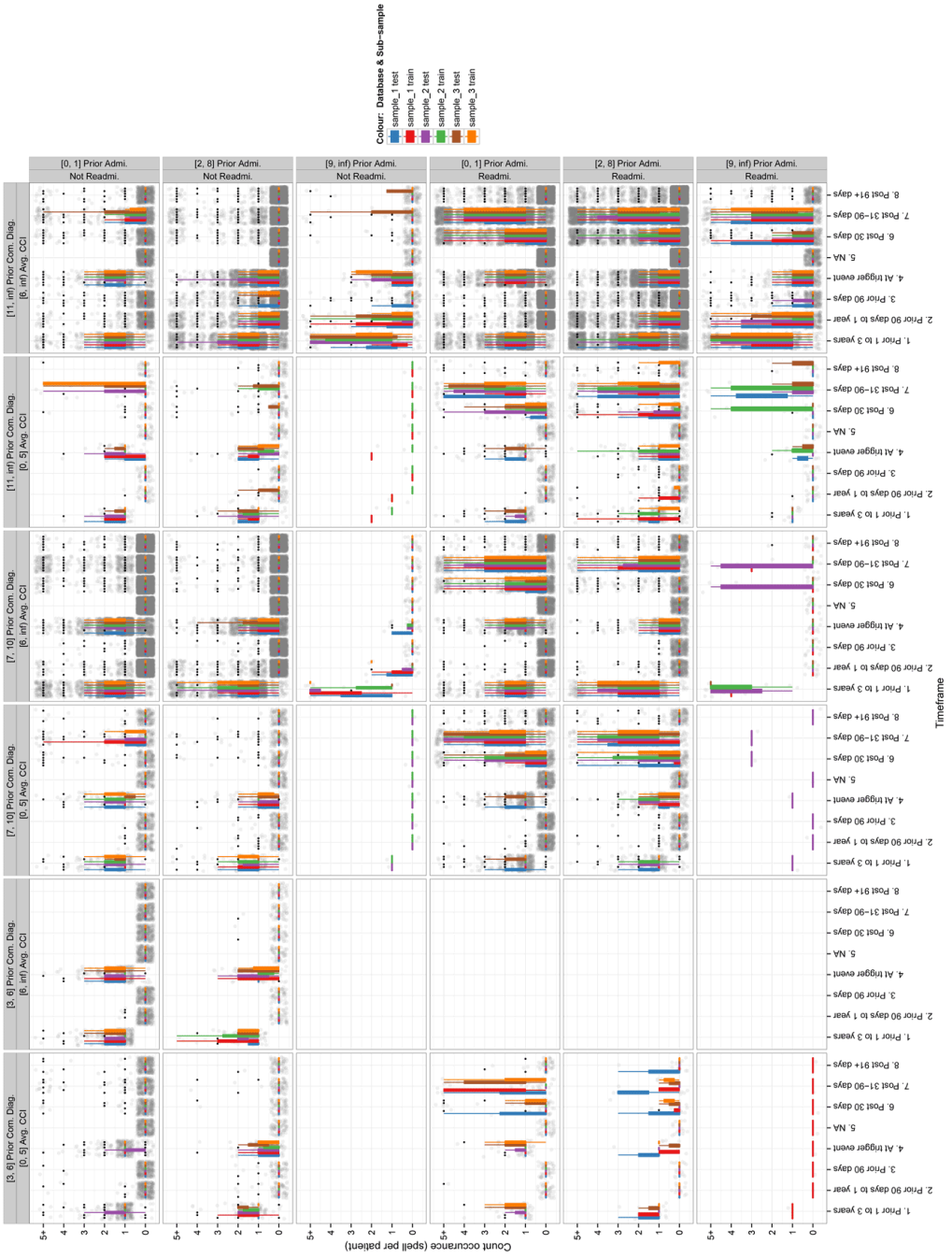

FIGURE A.19: The box-plot statistics of metastatic solid tumor (all samples)

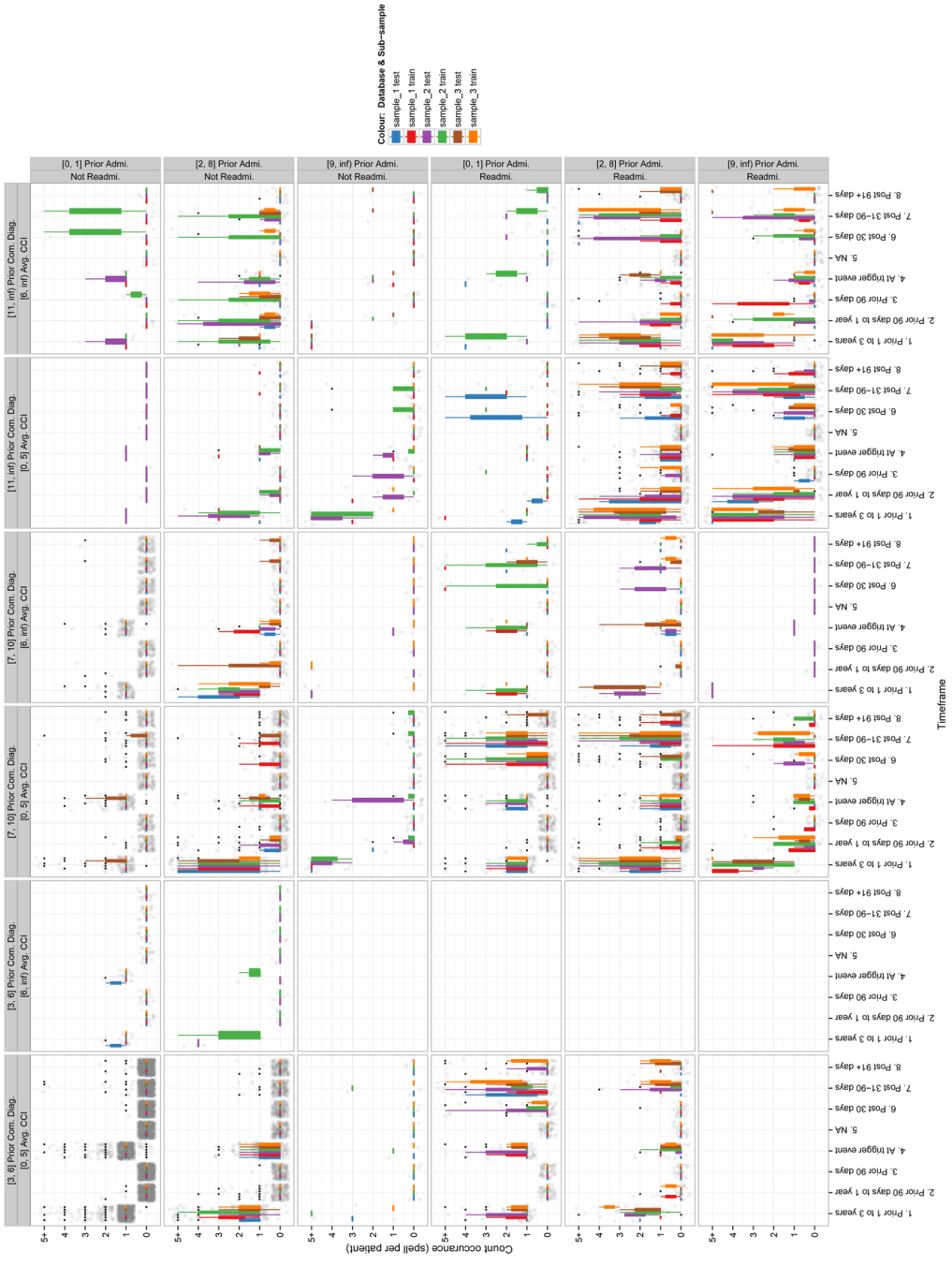

FIGURE A.20: The box-plot statistics of AIDS/HIV (all samples)

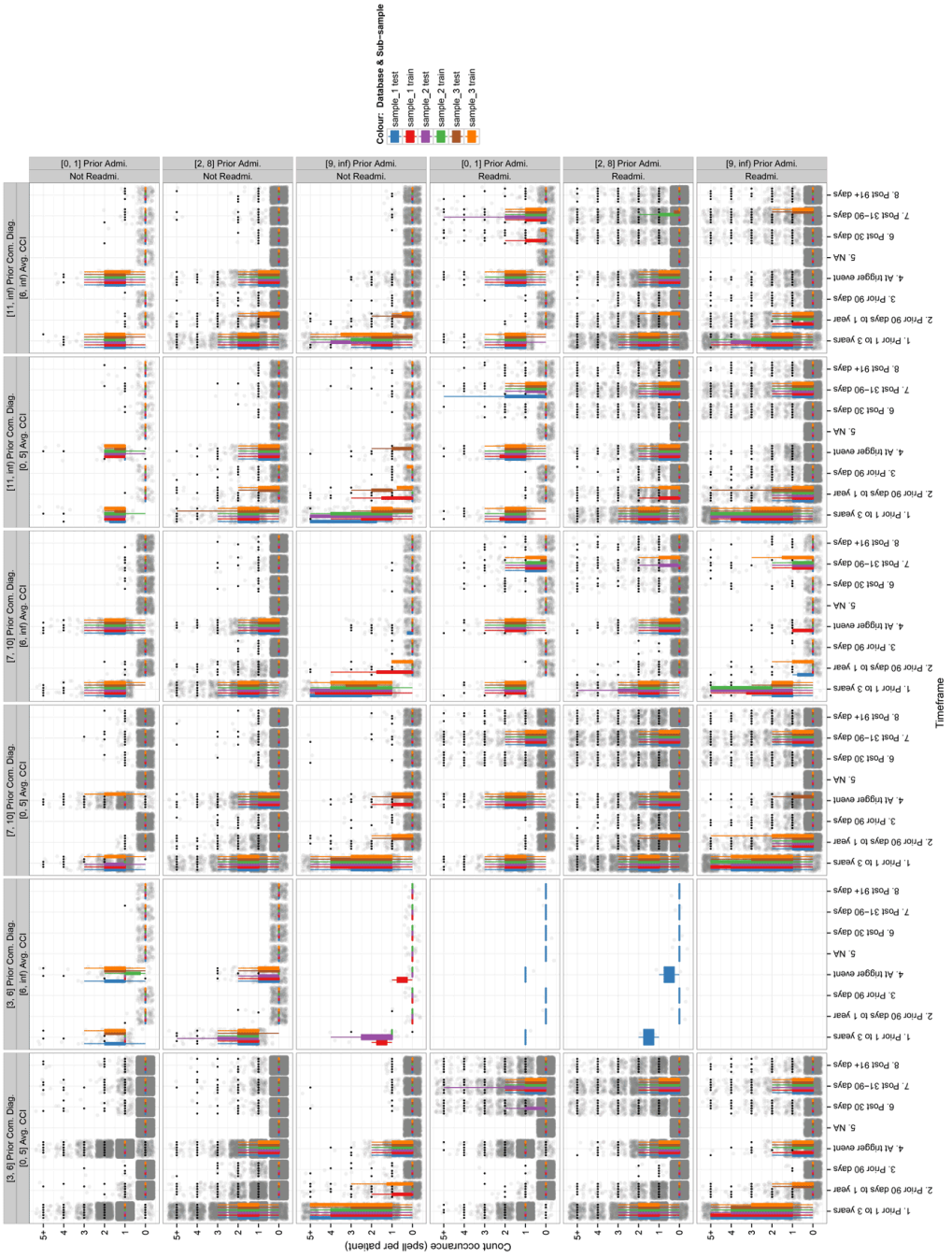

FIGURE A.21: The box-plot statistics of depression (all samples)

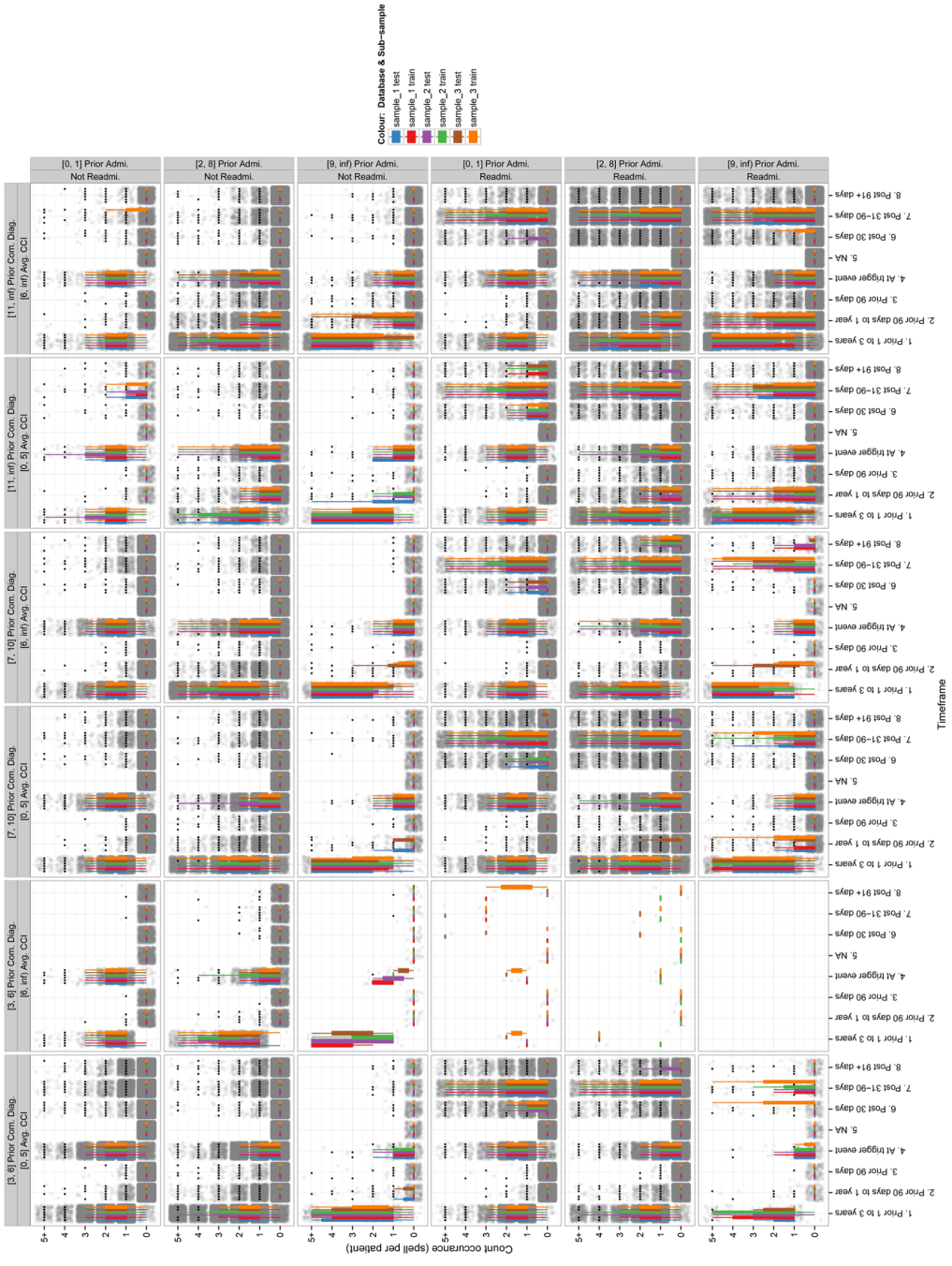

FIGURE A.22: The box-plot statistics of cardiac arrhythmias (all samples)

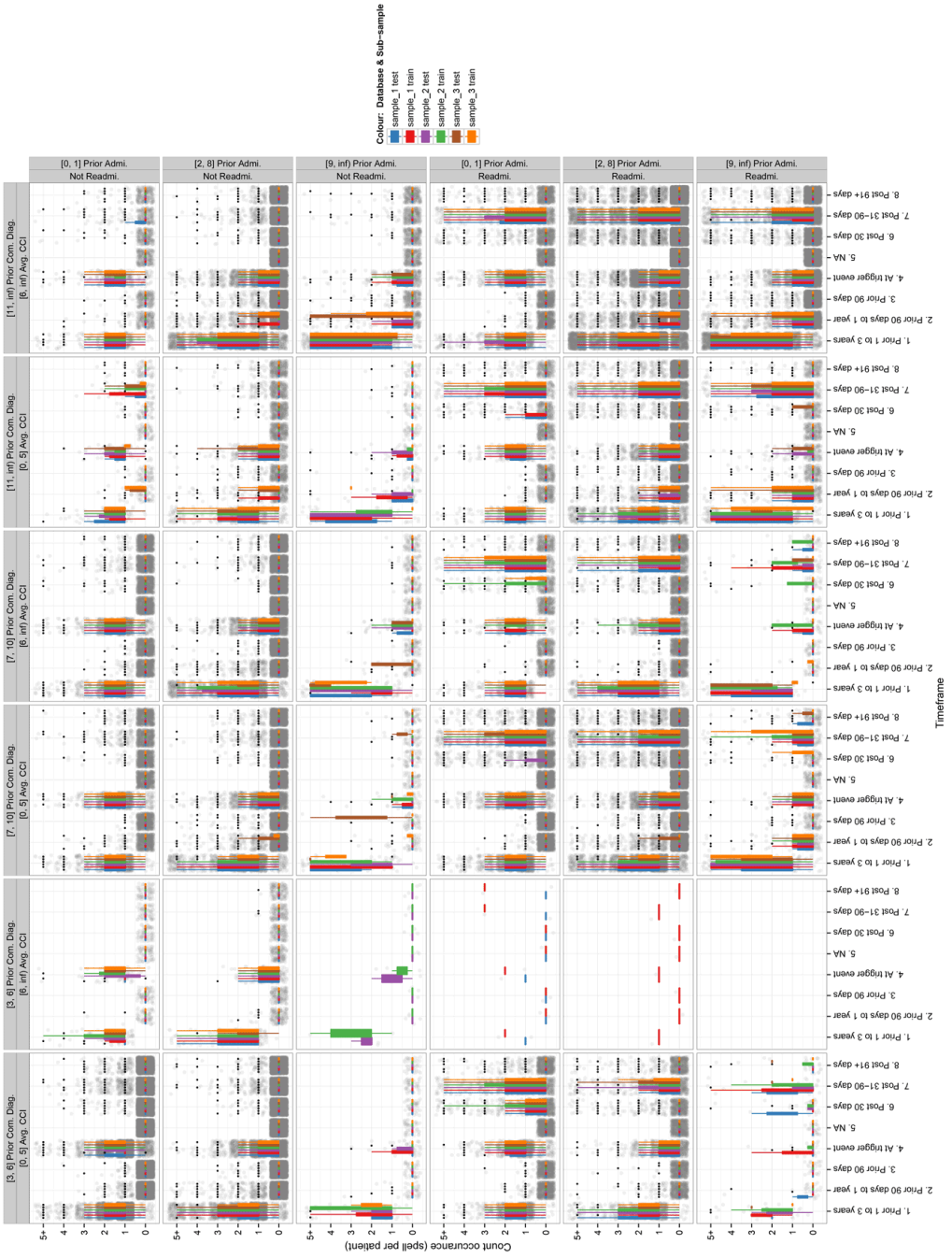

FIGURE A.23: The box-plot statistics of valvular disease (all samples)

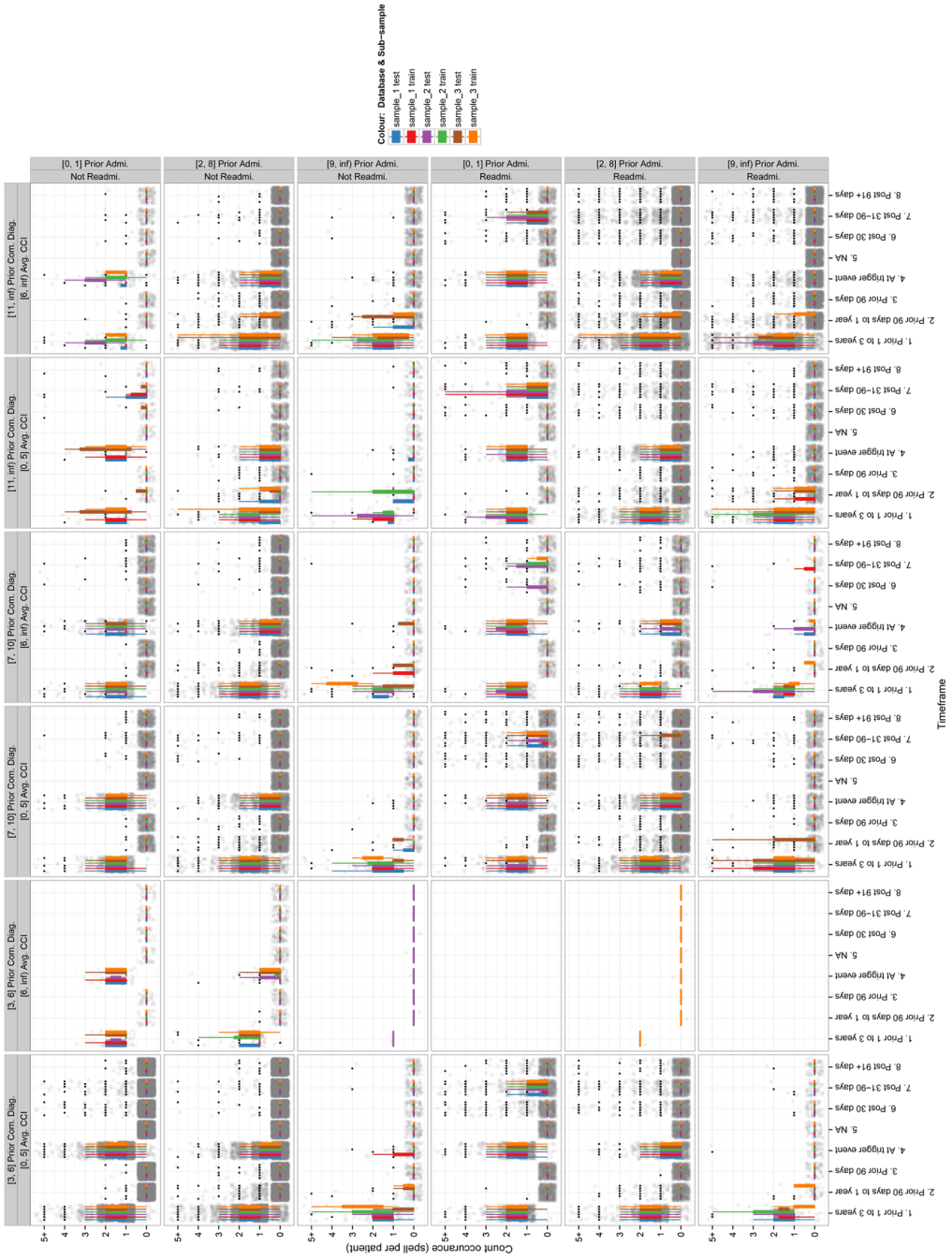

FIGURE A.24: The box-plot statistics of pulmonary circulation disorders (all samples)

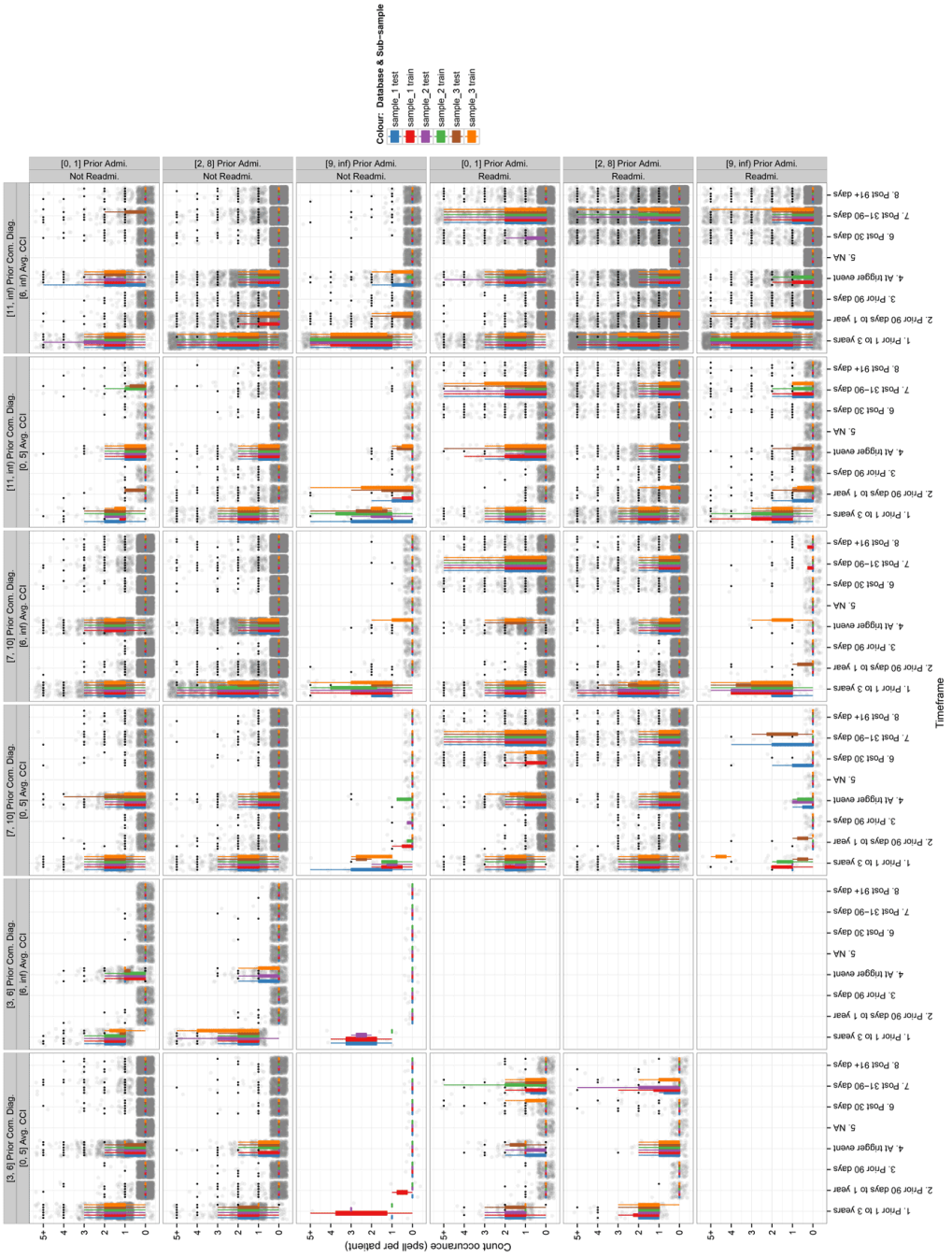

FIGURE A.25: The box-plot statistics of peripheral vascular disorders (all samples)

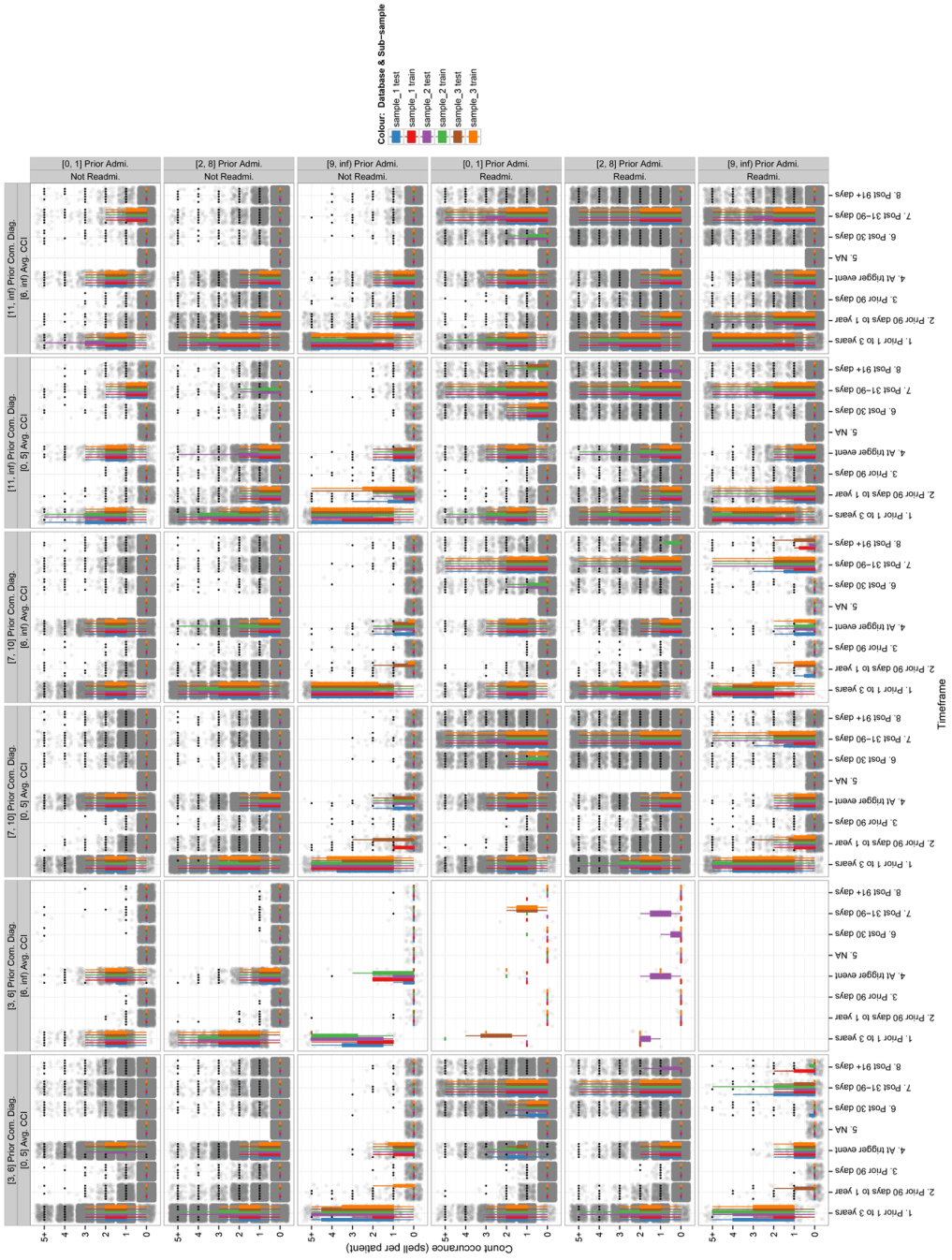

FIGURE A.26: The box-plot statistics of hypertension - uncomplicated (all samples)

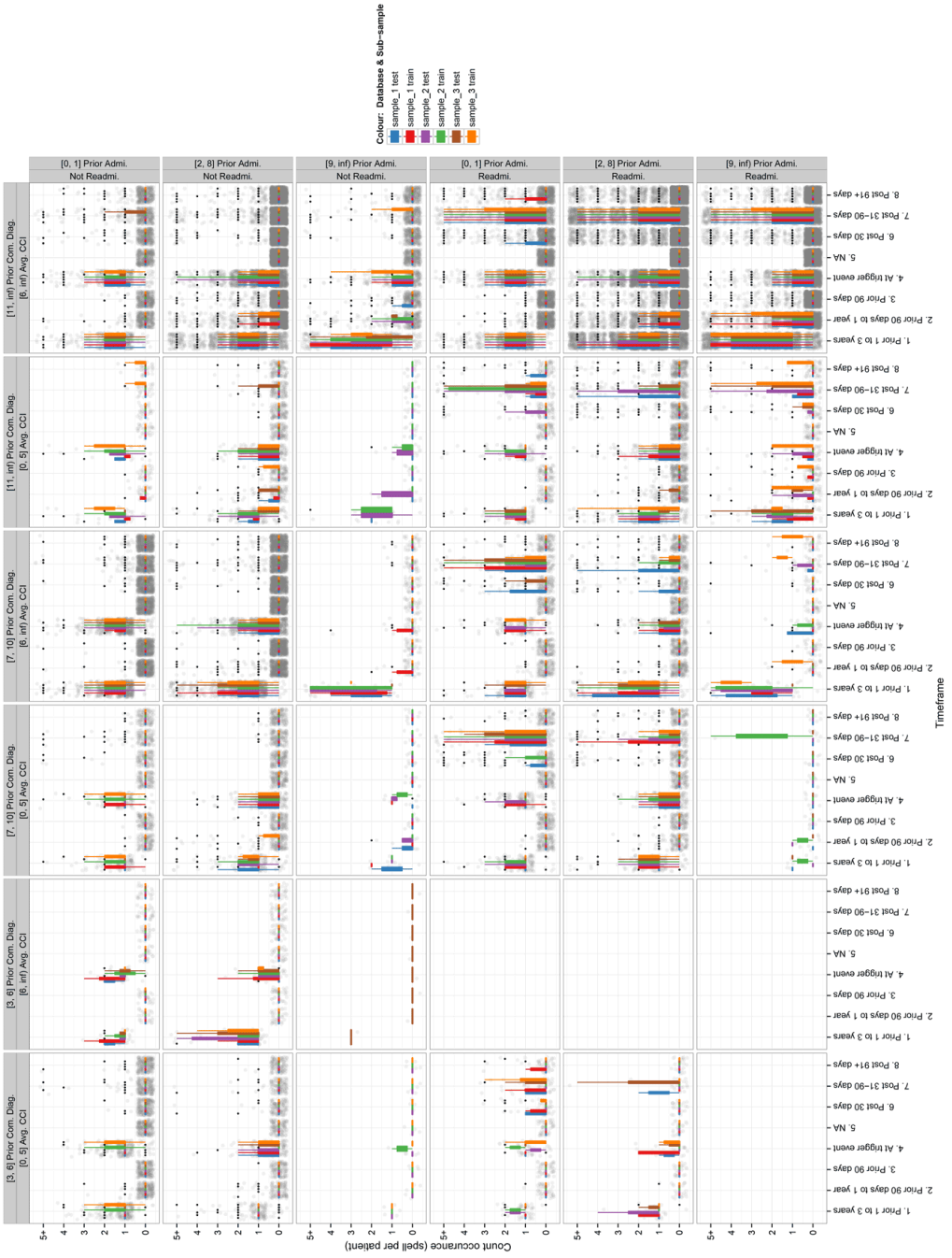

FIGURE A.27: The box-plot statistics of hypertension - complicated (all samples)

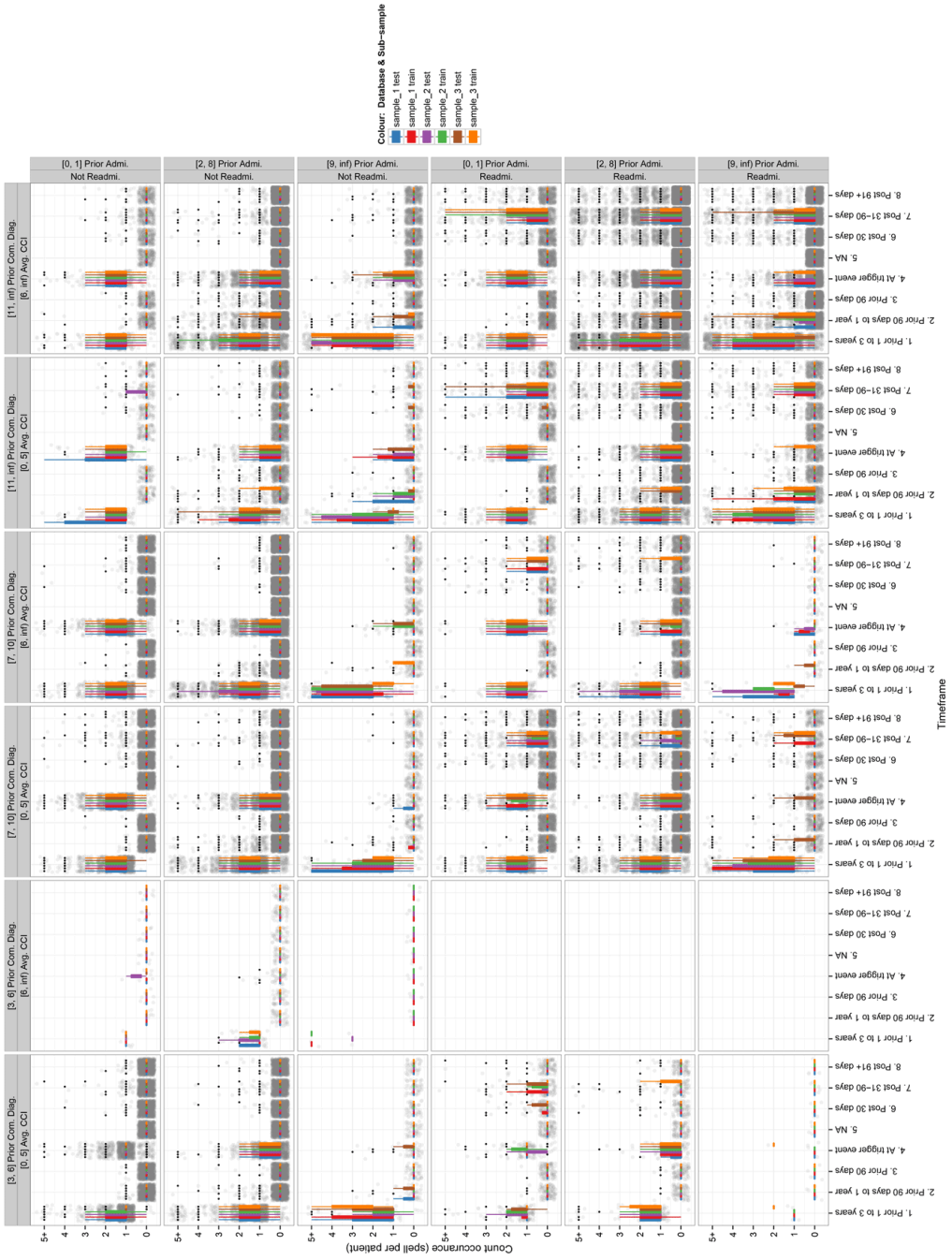

FIGURE A.28: The box-plot statistics of paralysis (all samples)

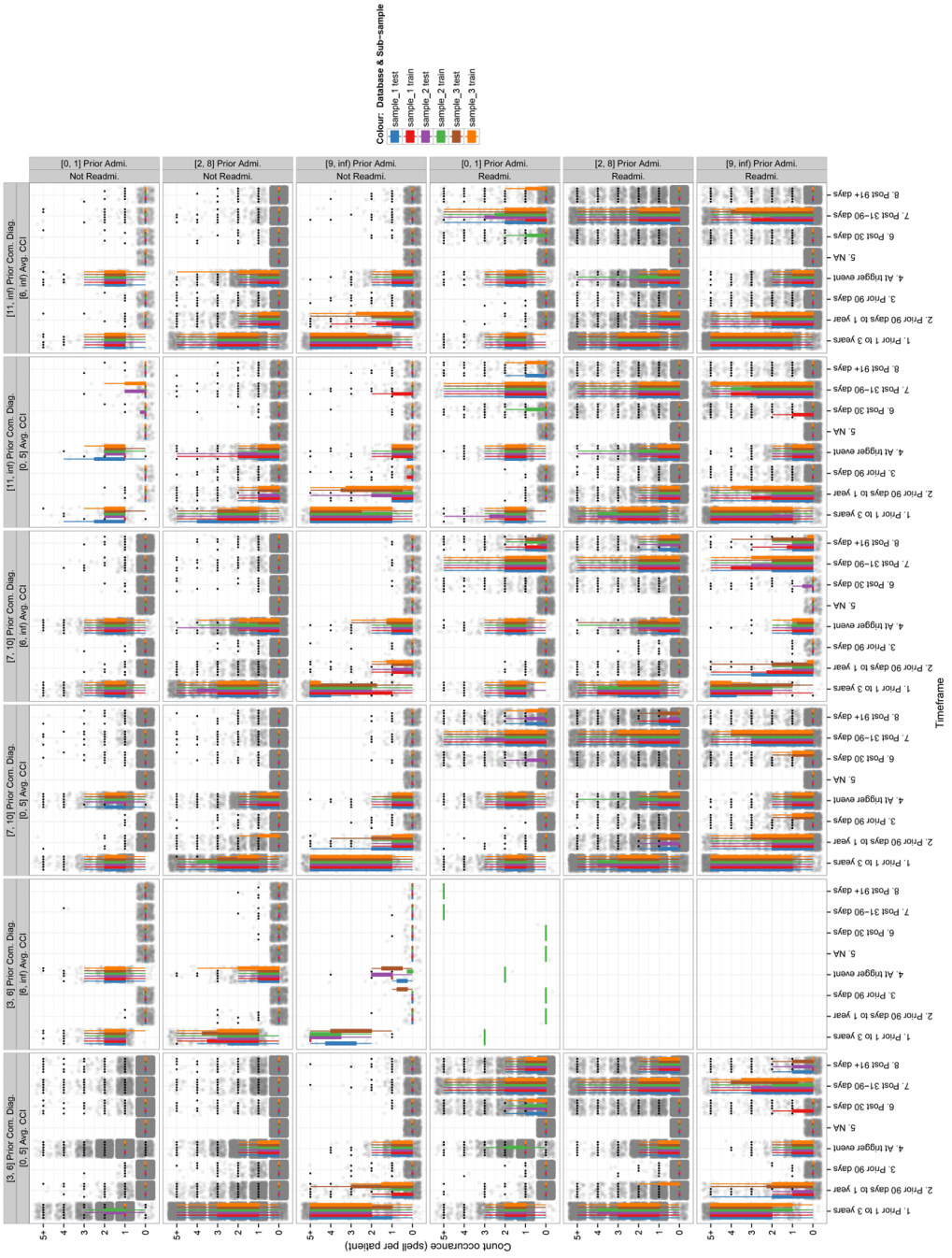

FIGURE A.29: The box-plot statistics of other neurological disorders (all samples)

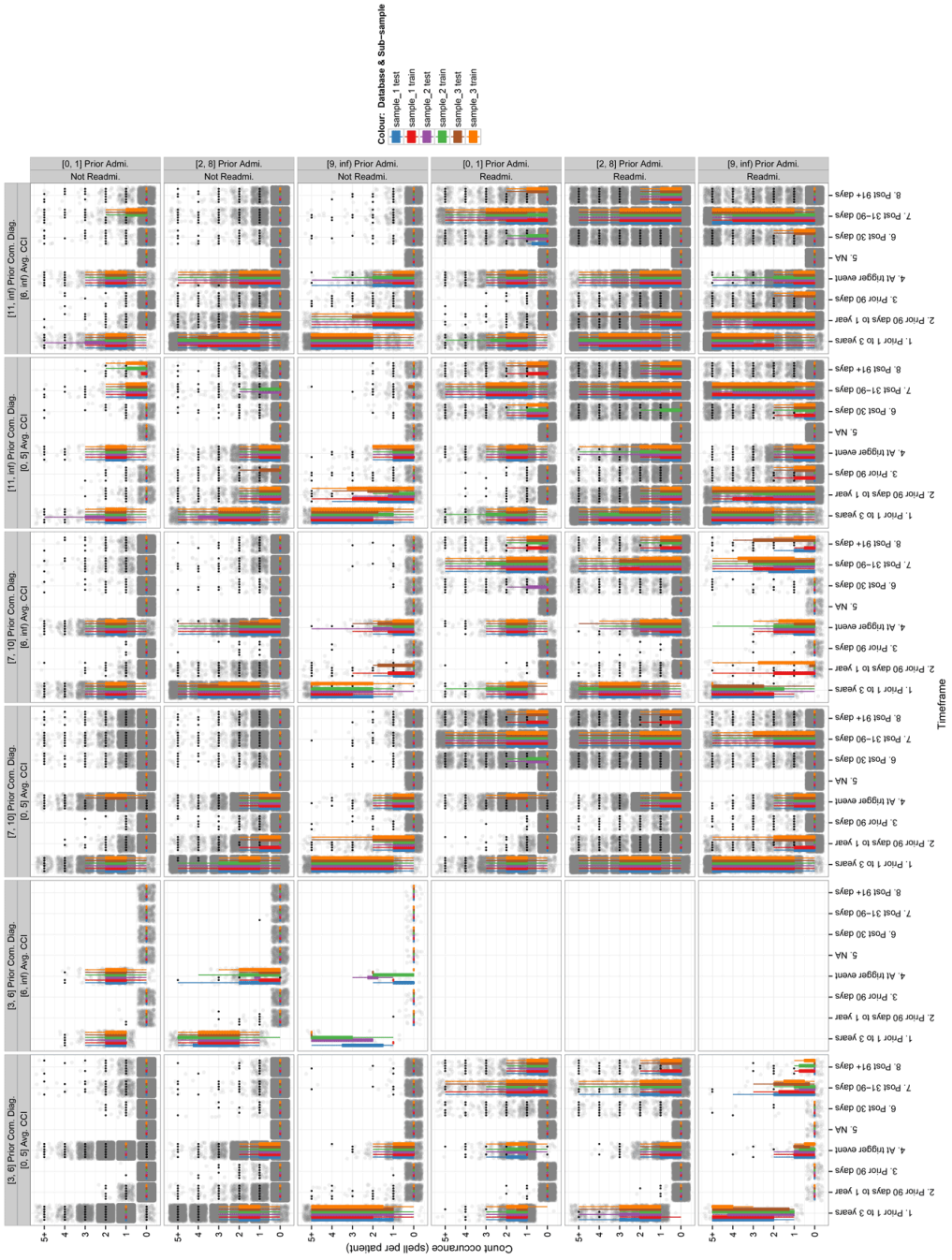

FIGURE A.30: The box-plot statistics of chronic pulmonary disease (all samples)

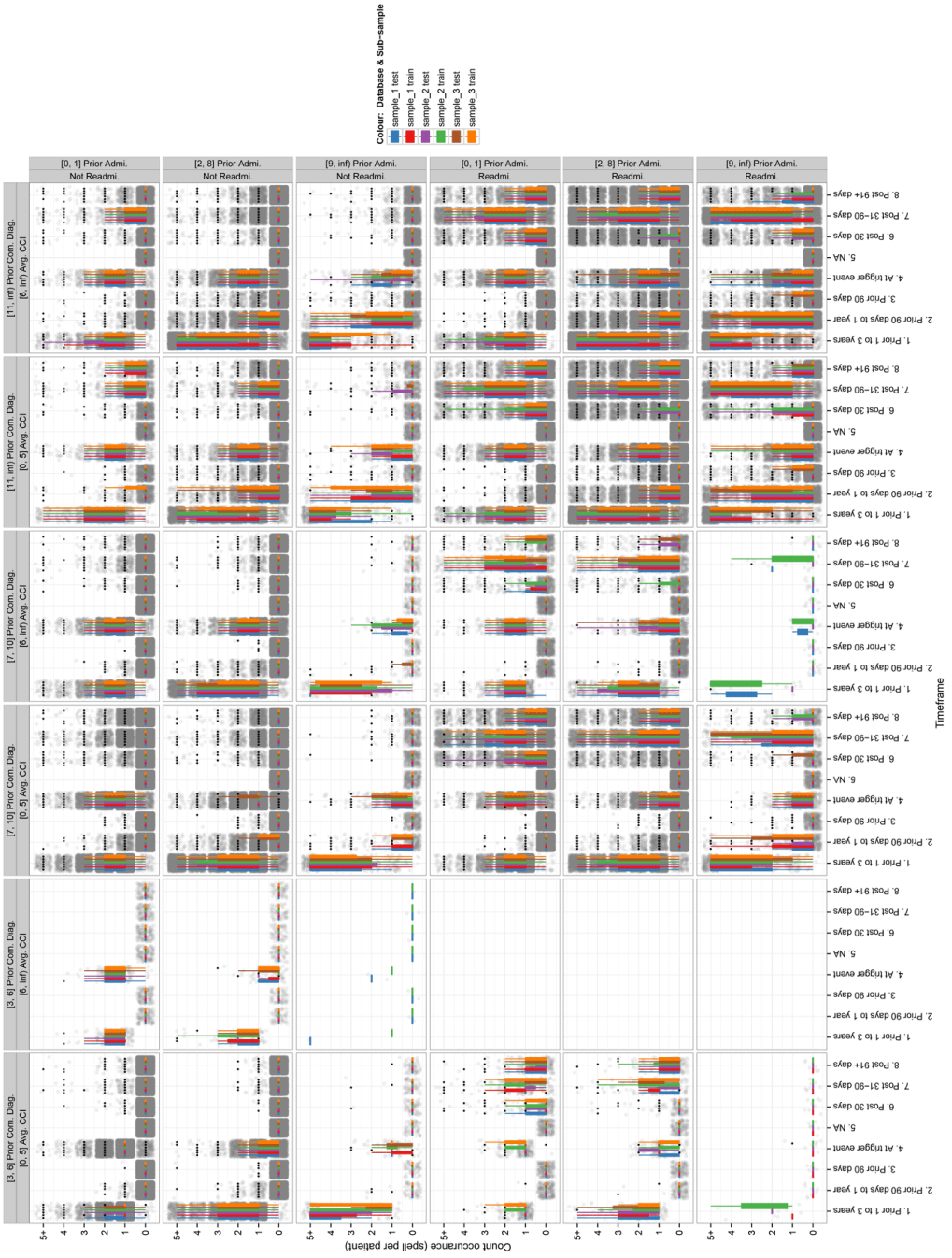

FIGURE A.31: The box-plot statistics of diabetes - uncomplicated (all samples)

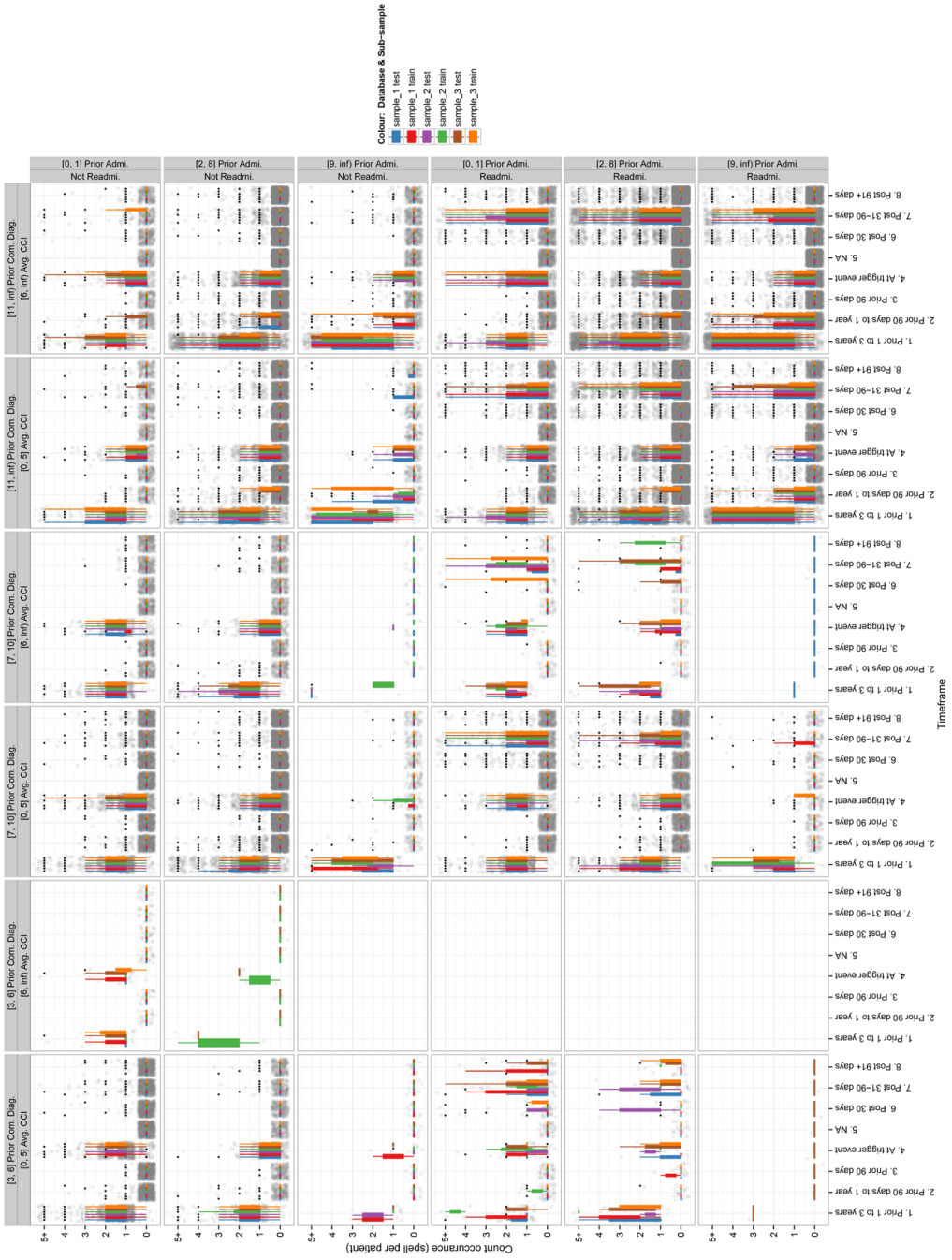

FIGURE A.32: The box-plot statistics of diabetes - complicated (all samples)

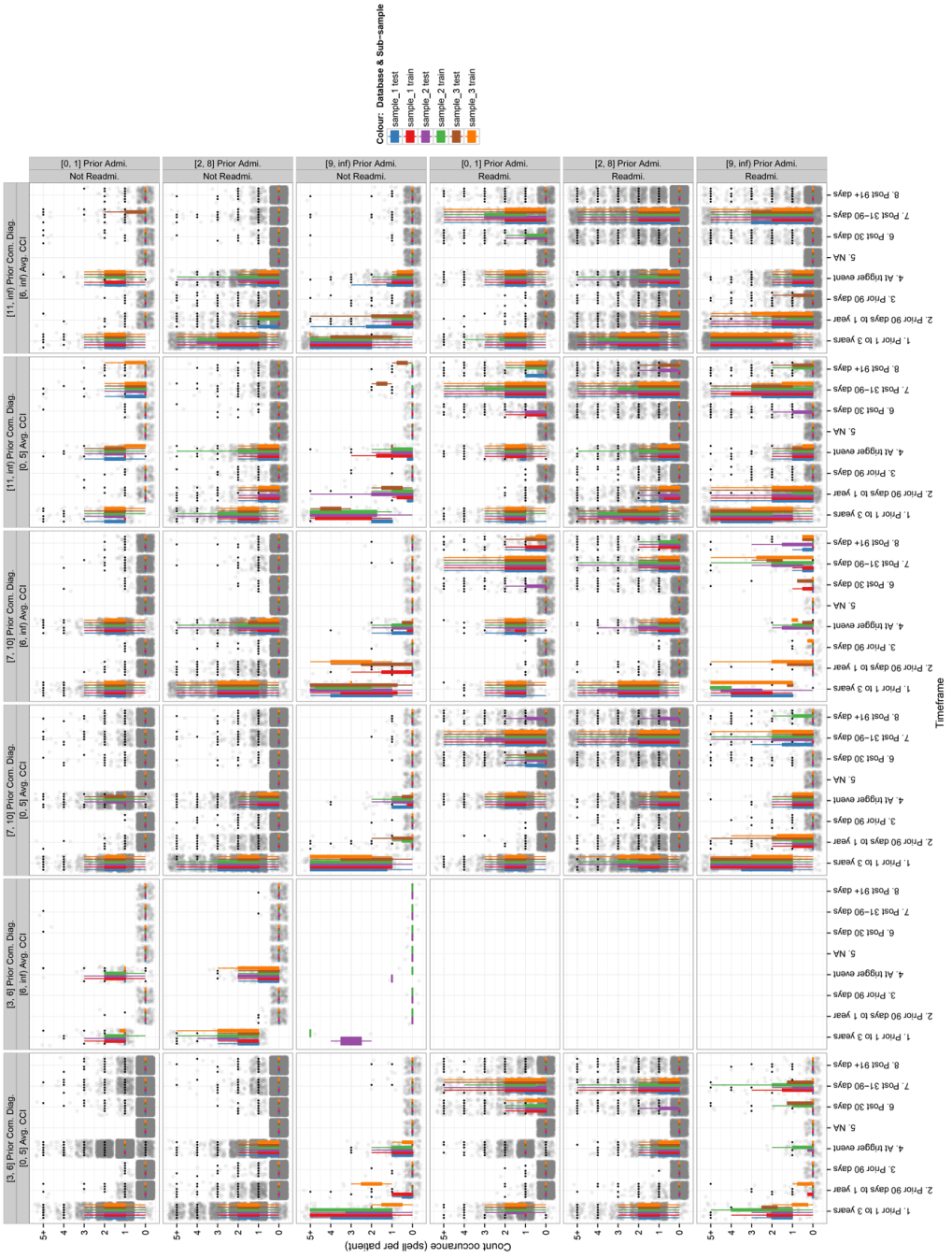

FIGURE A.33: The box-plot statistics of hypothyroidism (all samples)

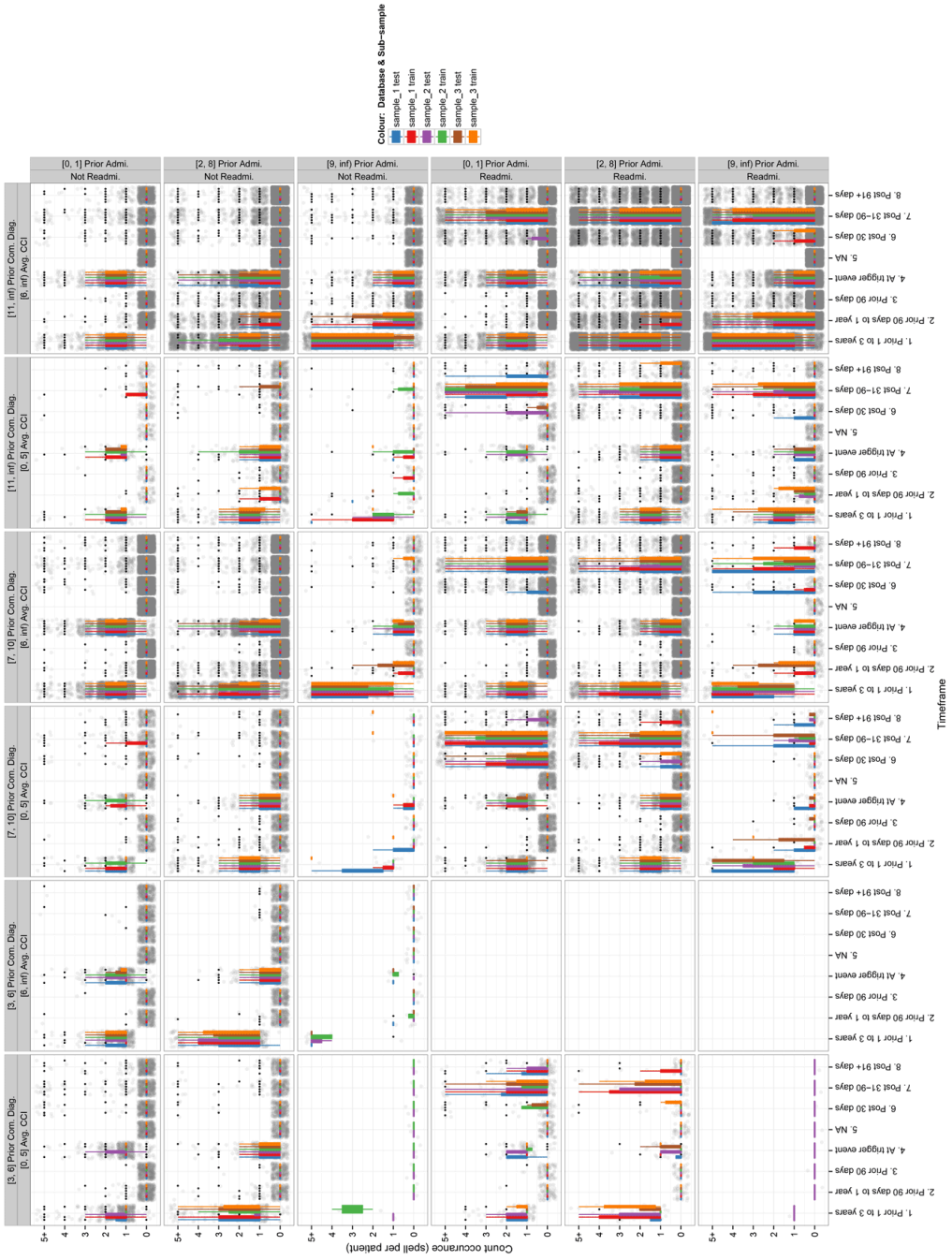

FIGURE A.34: The box-plot statistics of renal failure (all samples)

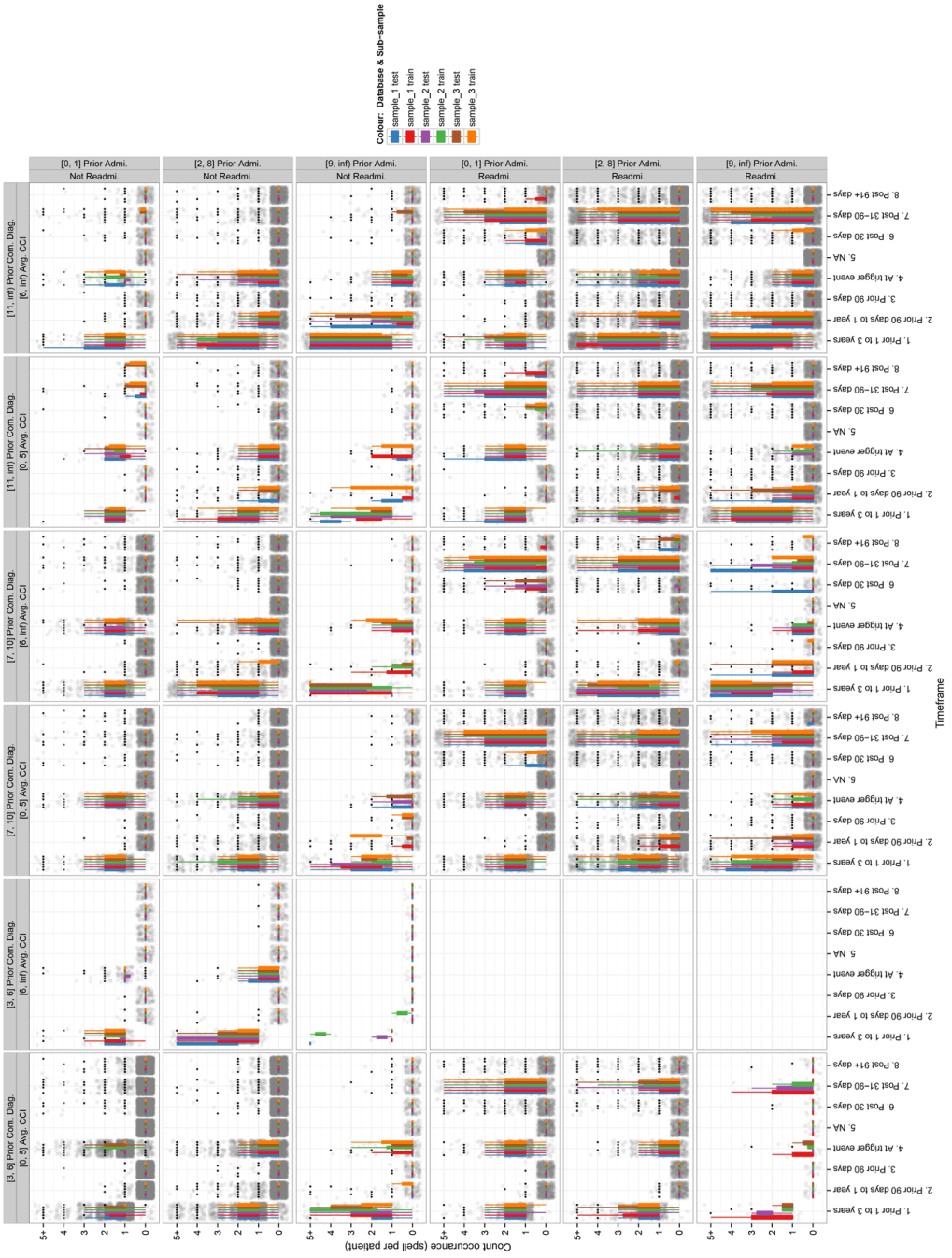

FIGURE A.35: The box-plot statistics of liver disease (all samples)

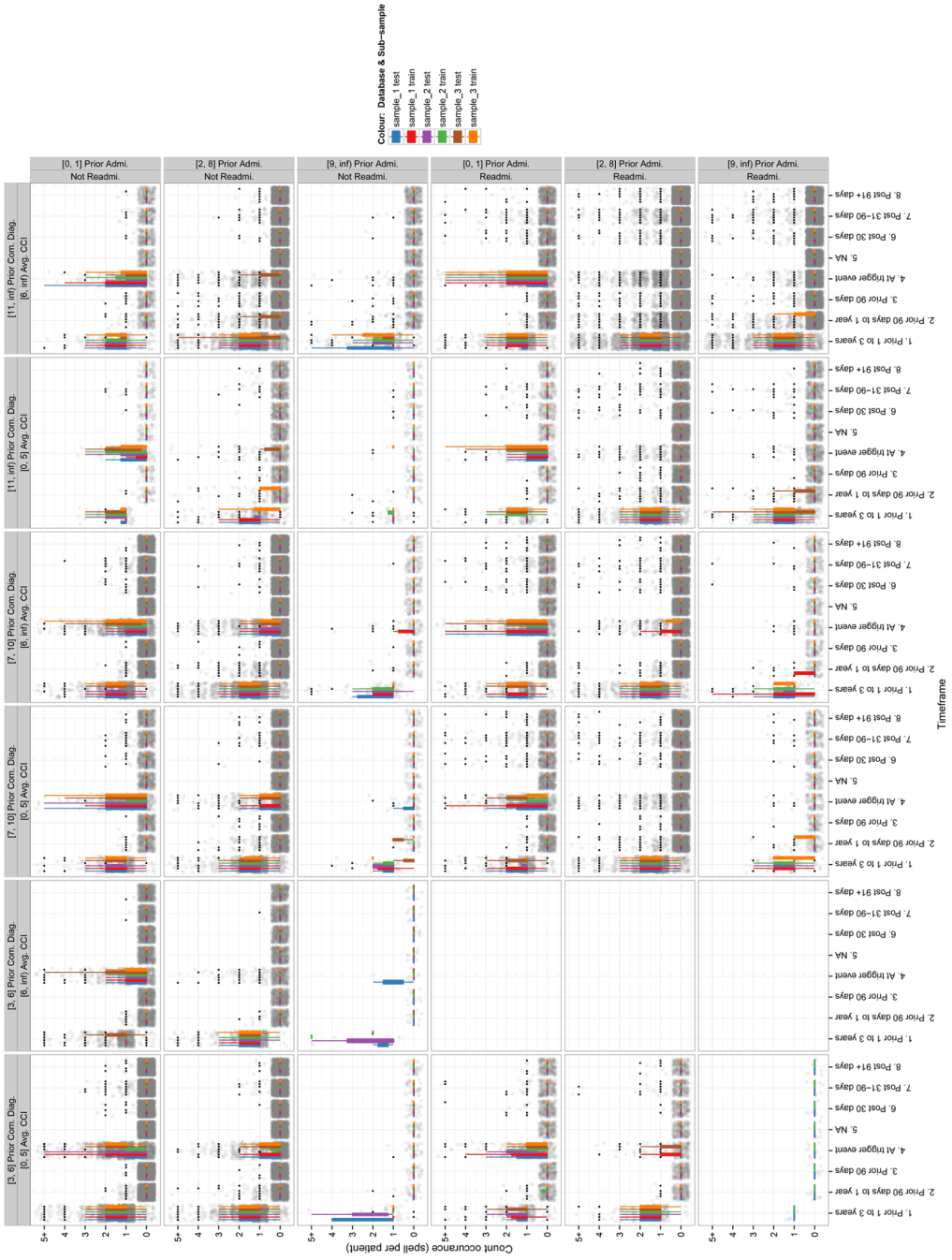

FIGURE A.36: The box-plot statistics of peptic ulcer disease excluding bleeding (all samples)

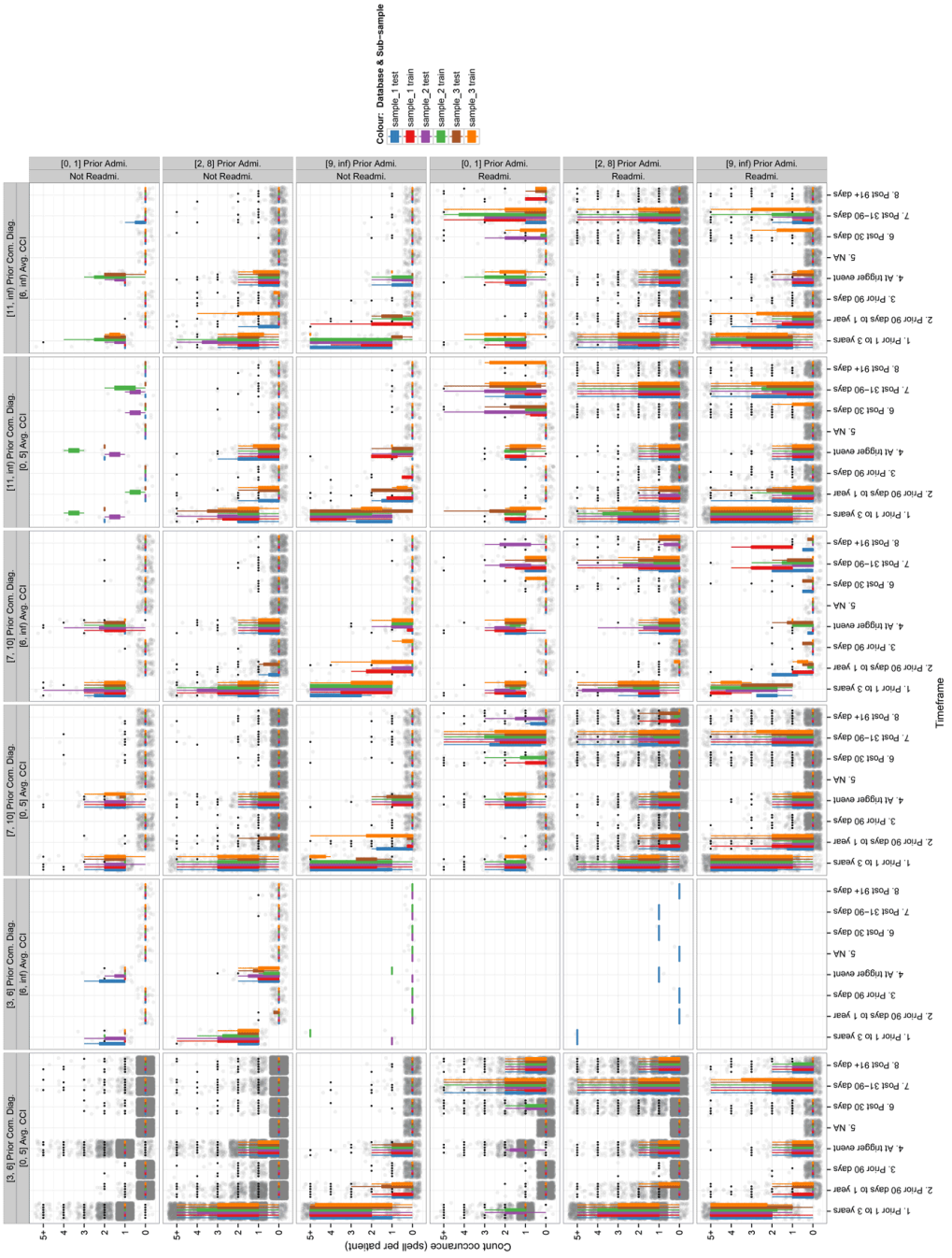

FIGURE A.37: The box-plot statistics of psychoses (all samples)

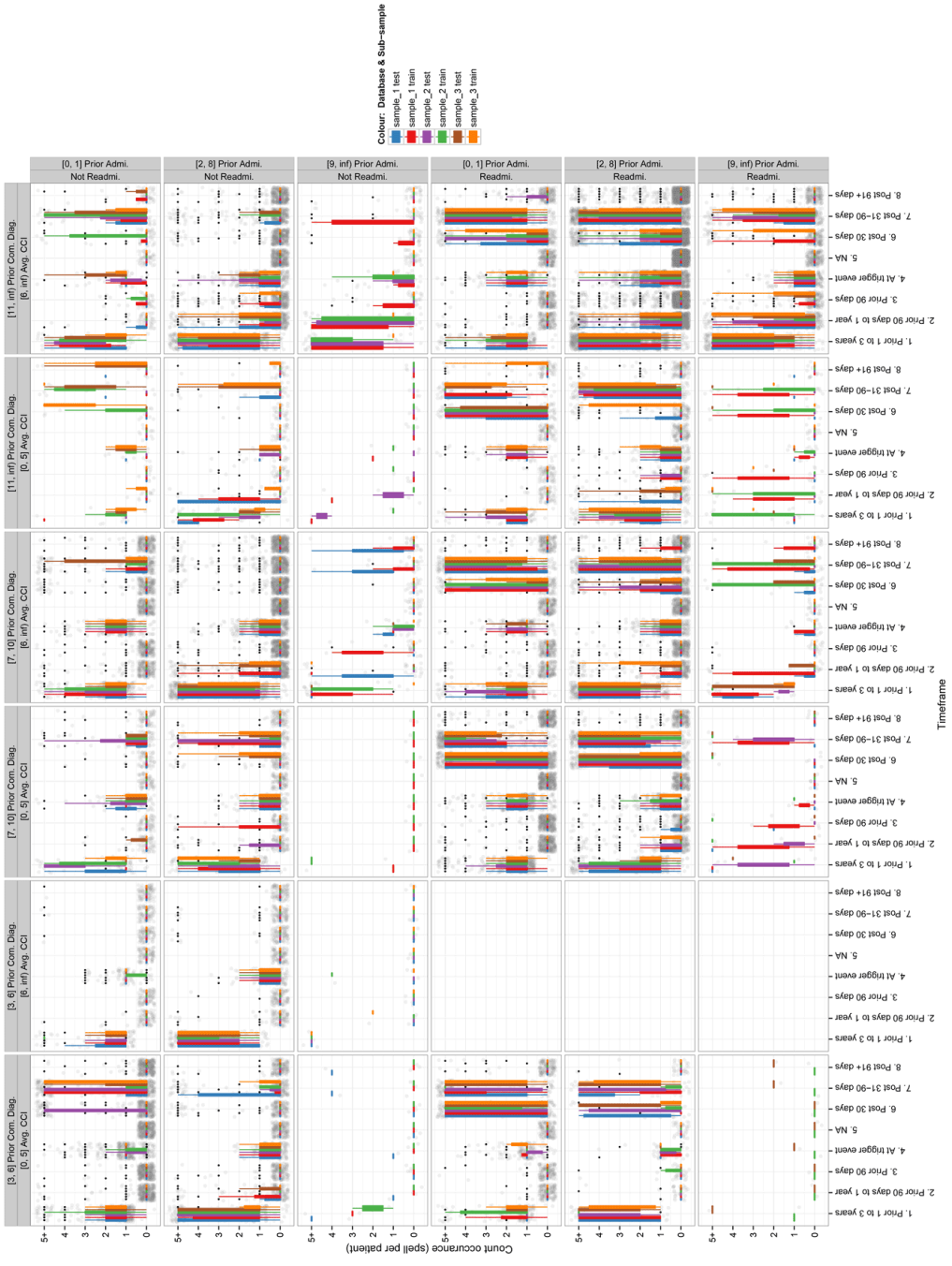

FIGURE A.38: The box-plot statistics of lymphoma. (all samples)

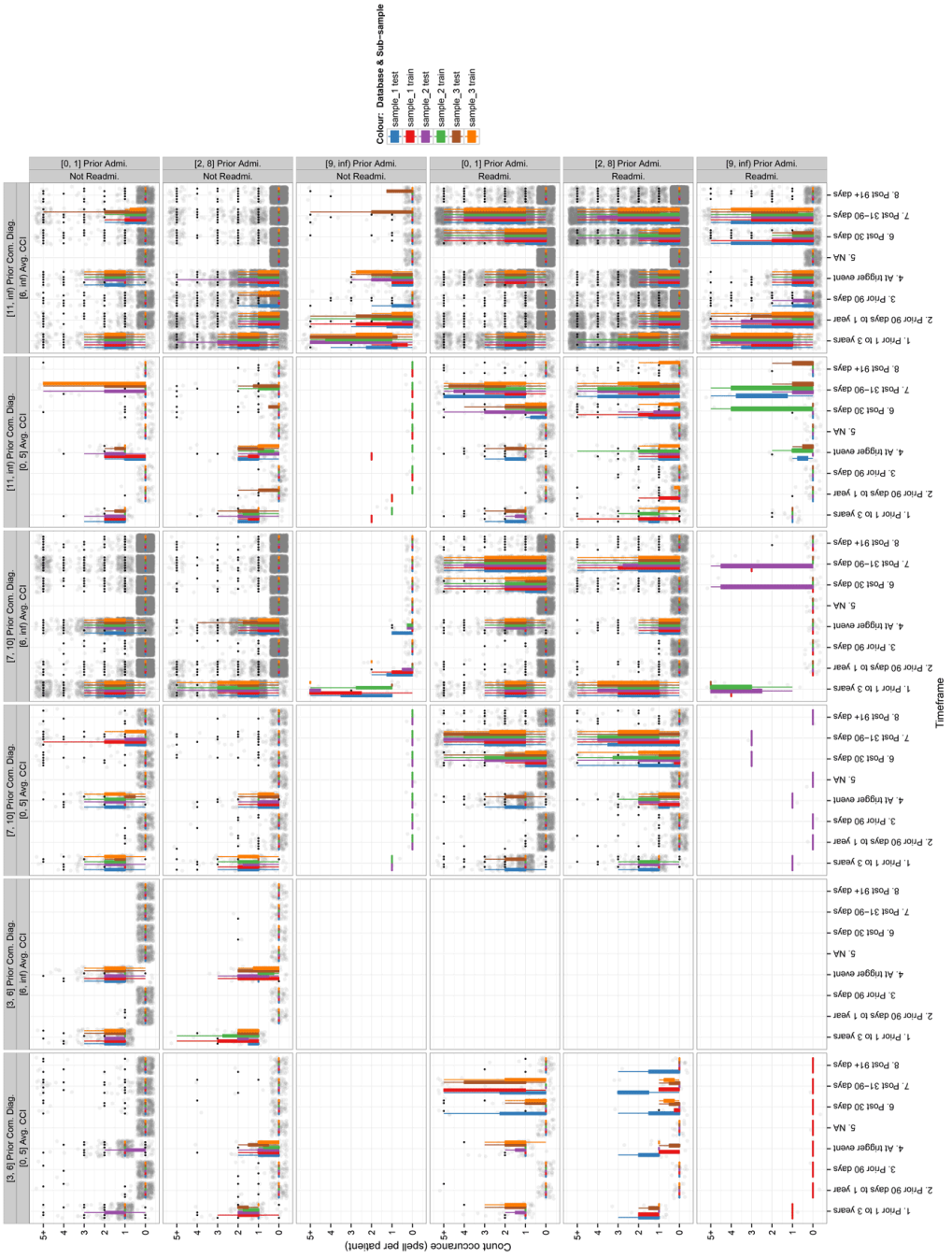

FIGURE A.39: The box-plot statistics of metastatic cancer (all samples)

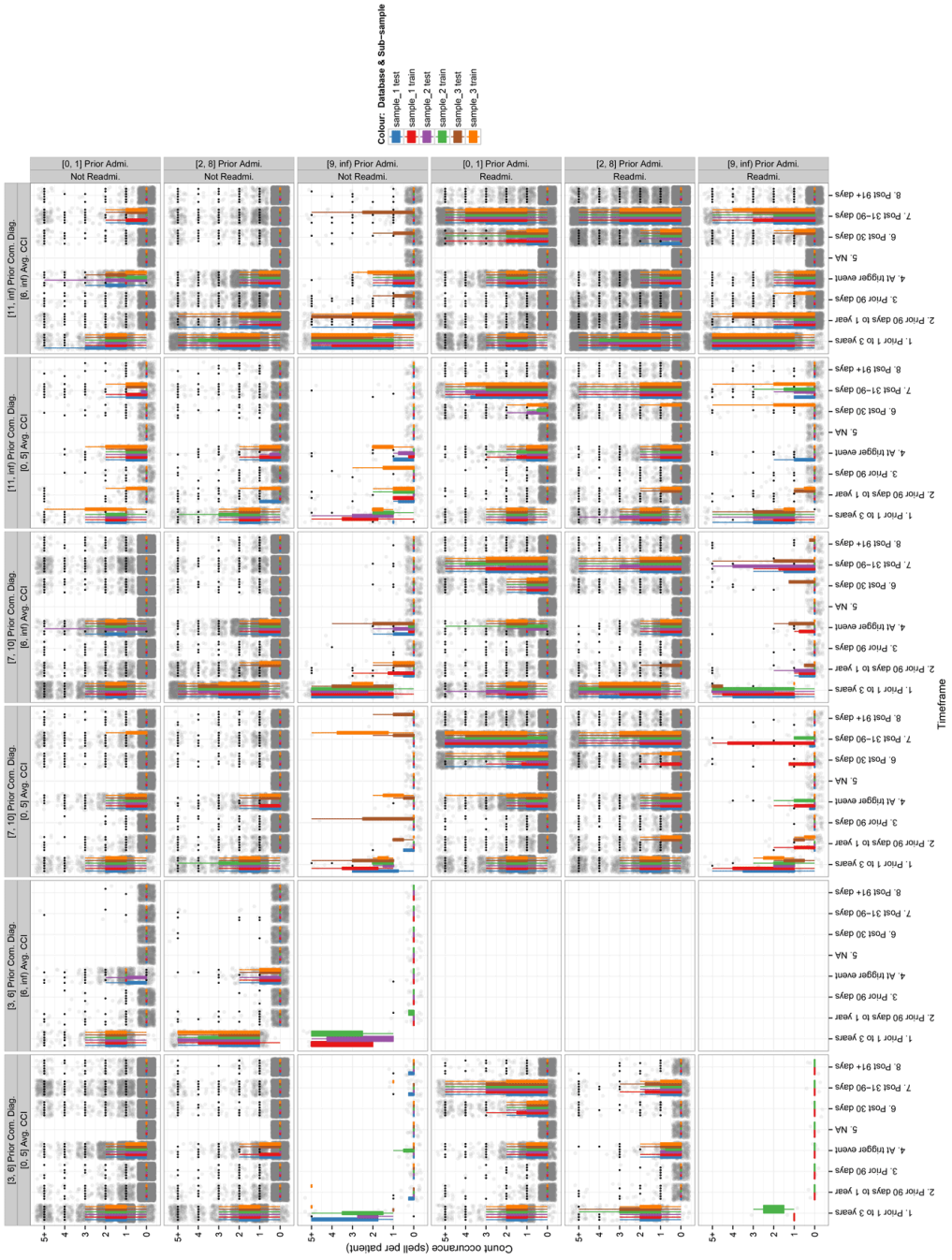

FIGURE A.40: The box-plot statistics of solid tumor without metastasis (all samples)

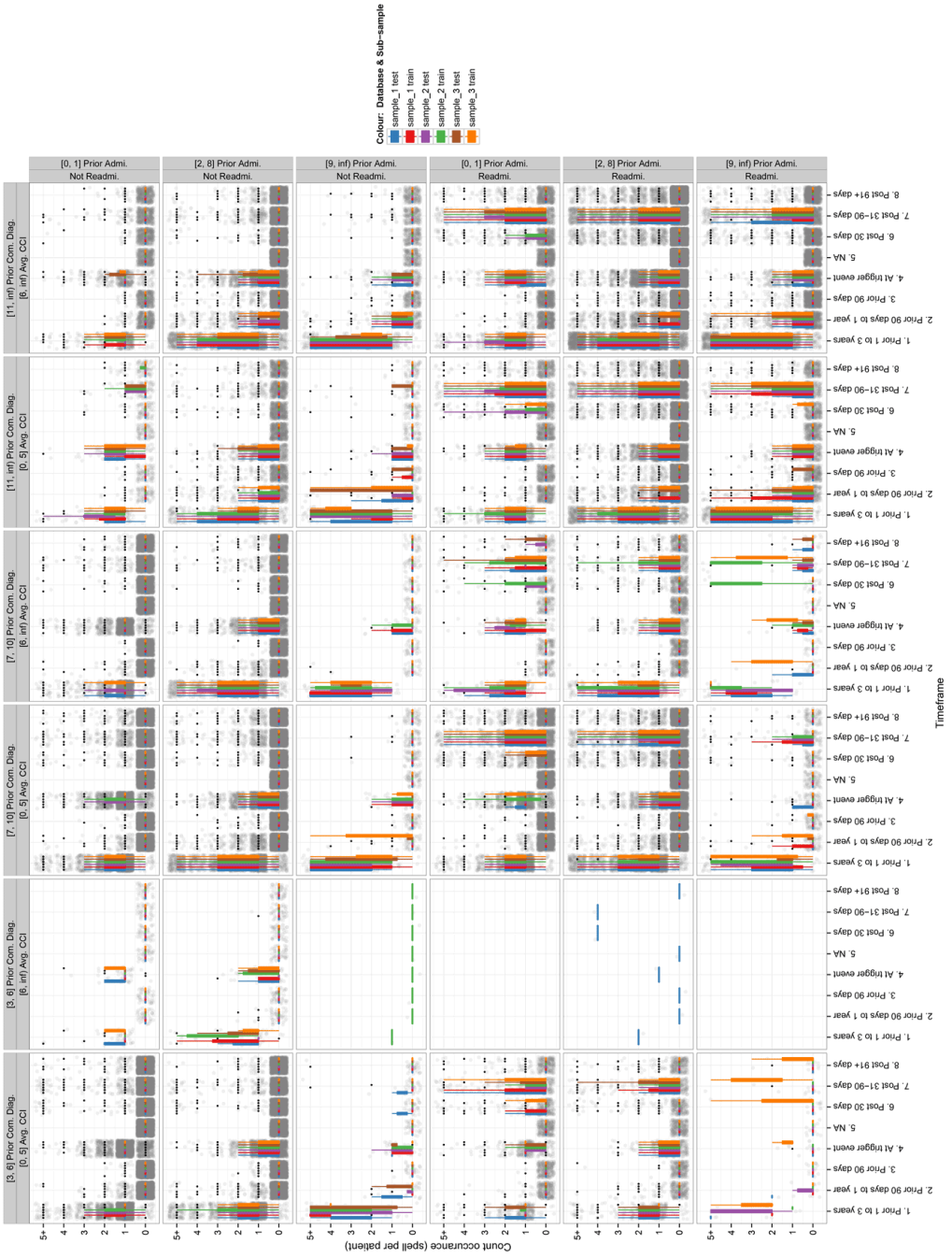

FIGURE A.41: The box-plot statistics of rheumatoid arthritis/ collagen vascular diseases (all samples)

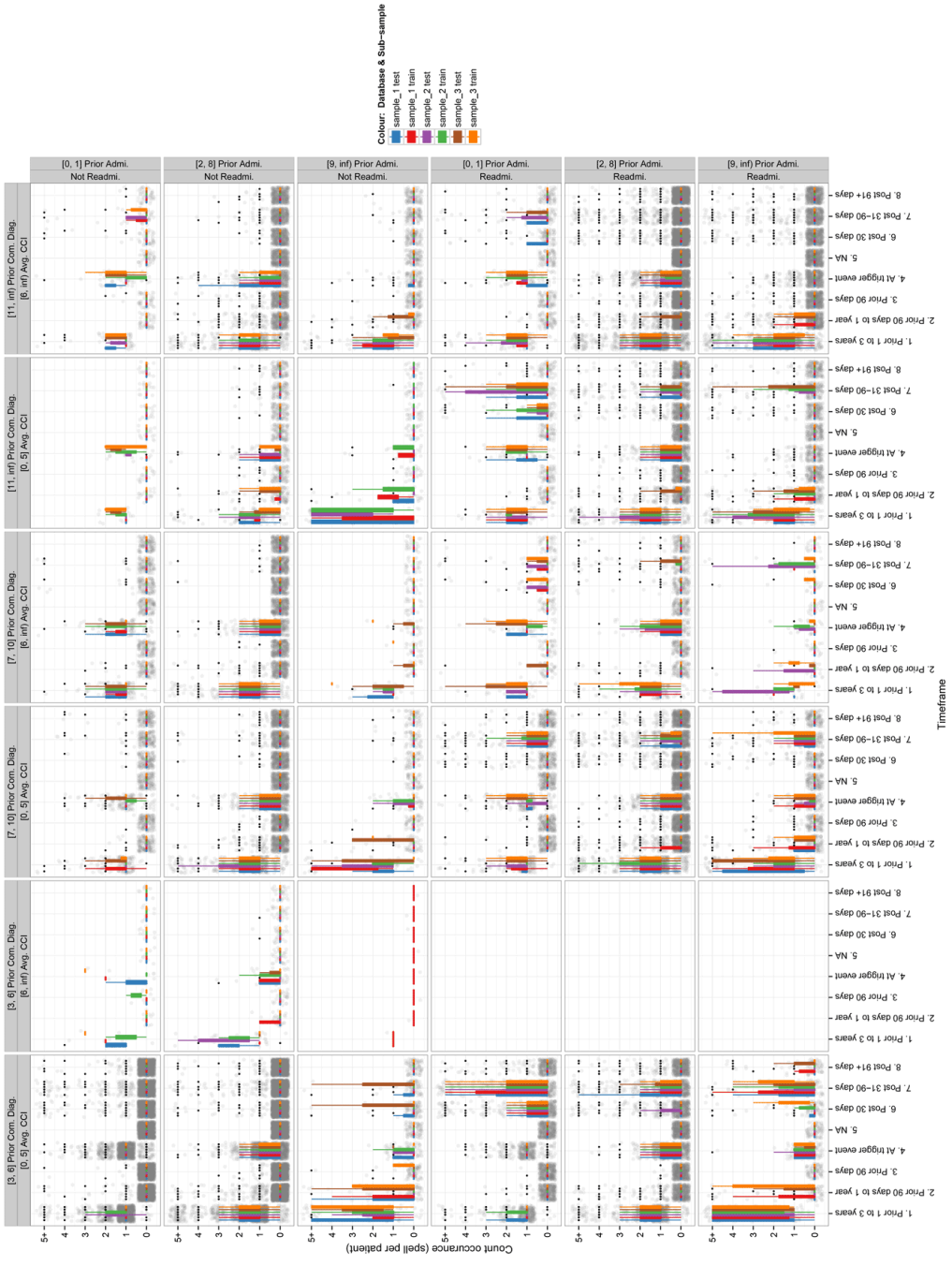

FIGURE A.42: The box-plot statistics of coagulopathy (all samples)

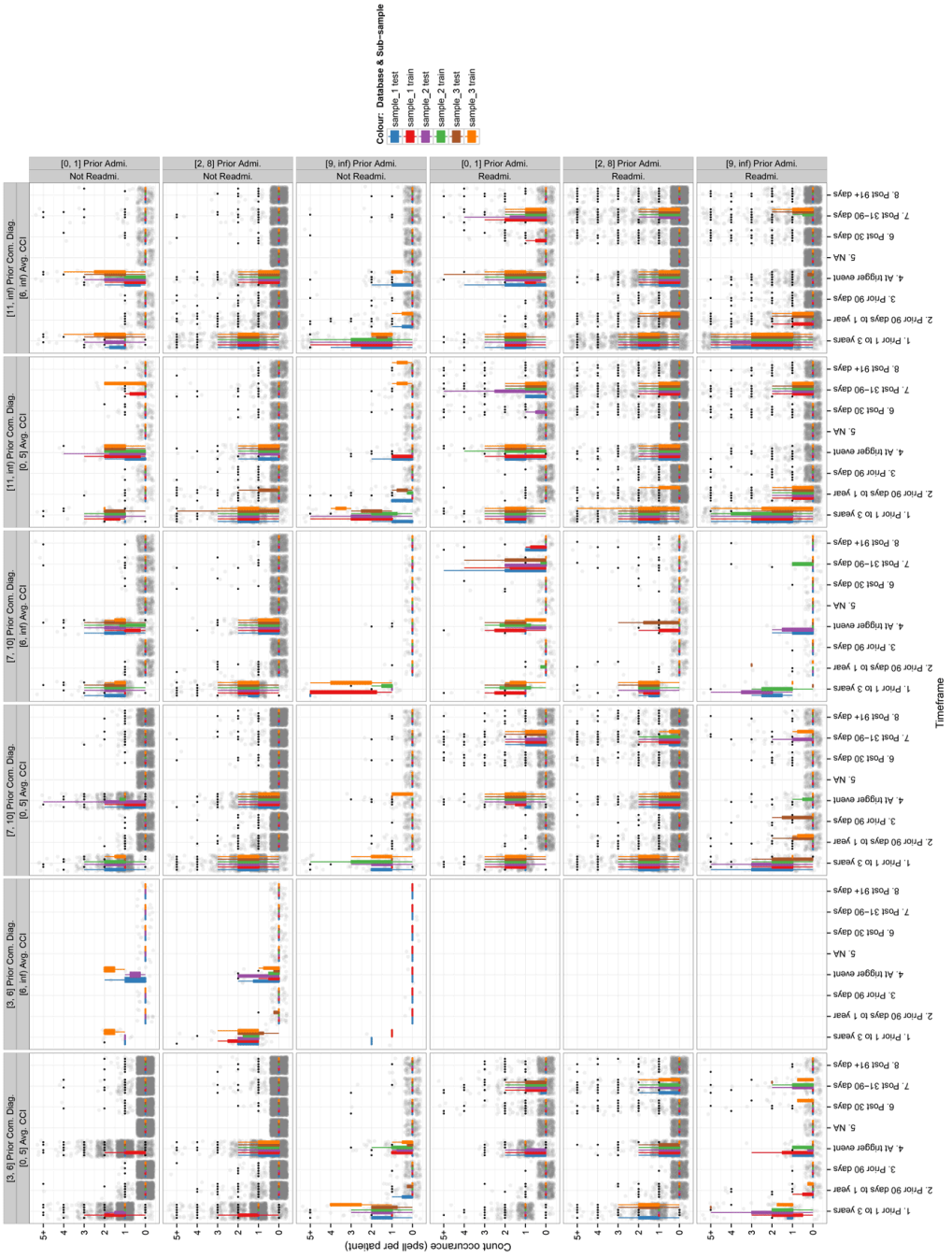

FIGURE A.43: The box-plot statistics of obesity (all samples)

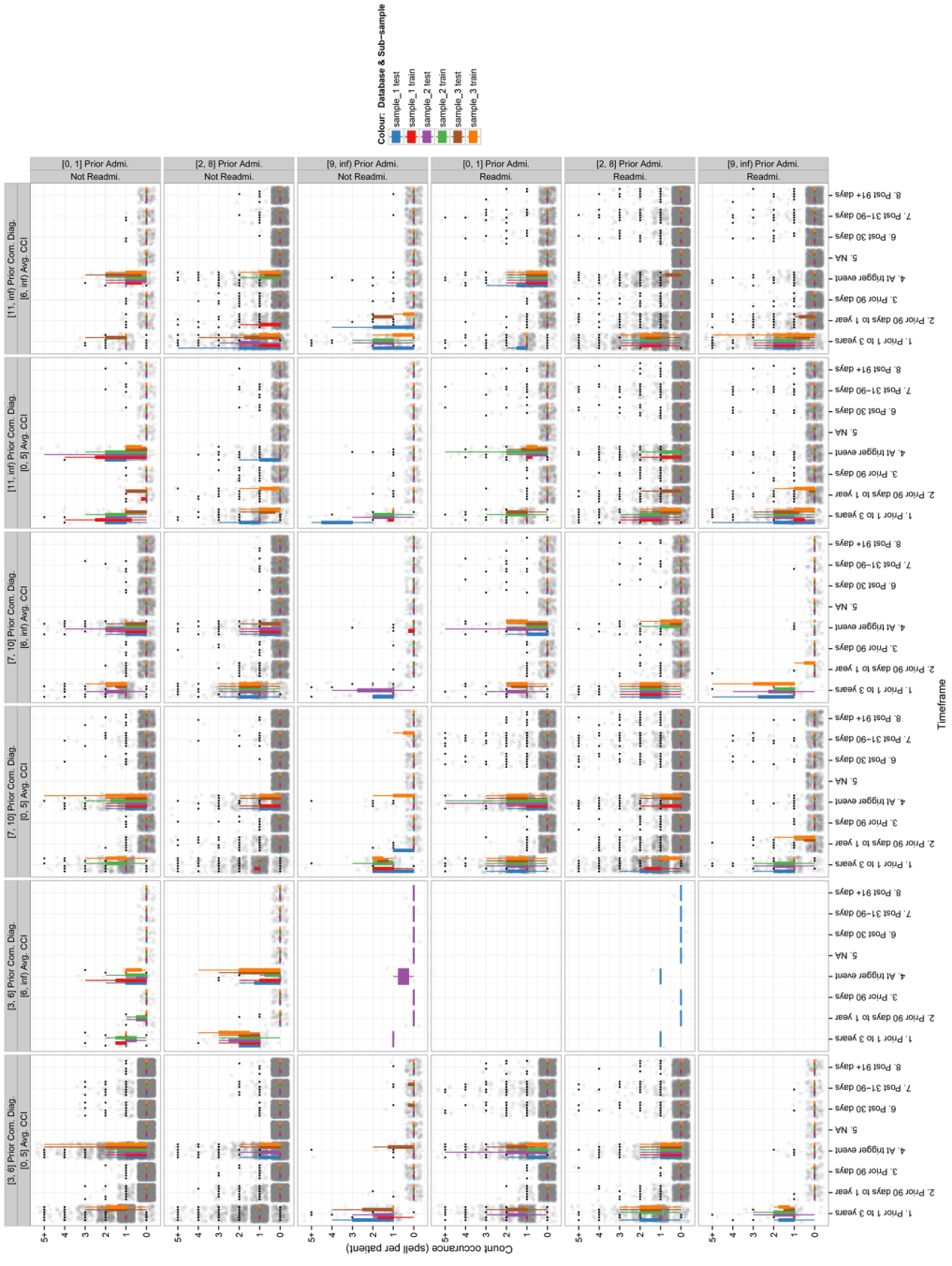

FIGURE A.44: The box-plot statistics of weight loss (all samples)

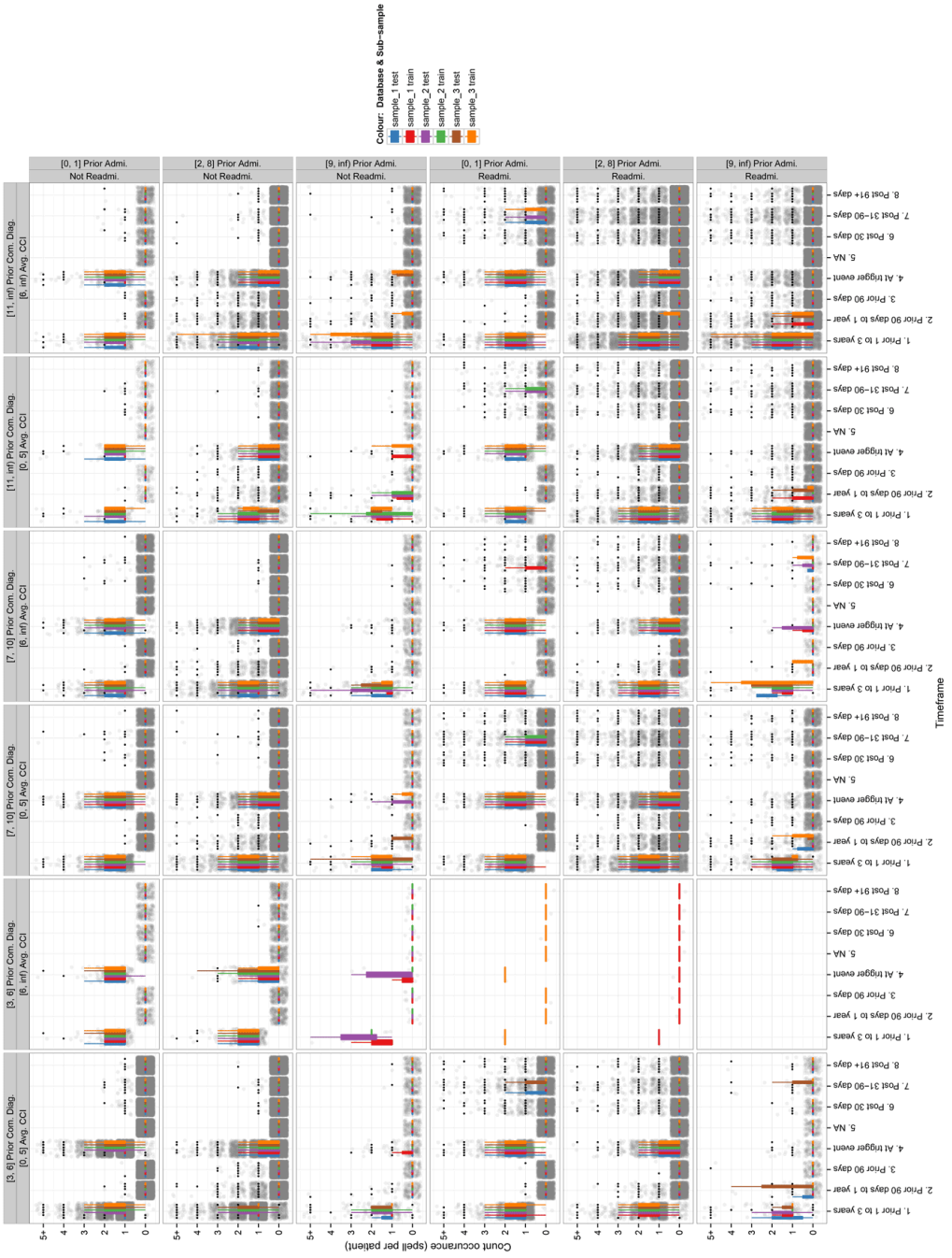

FIGURE A.45: The box-plot statistics of fluid and electrolyte disorders (all samples)

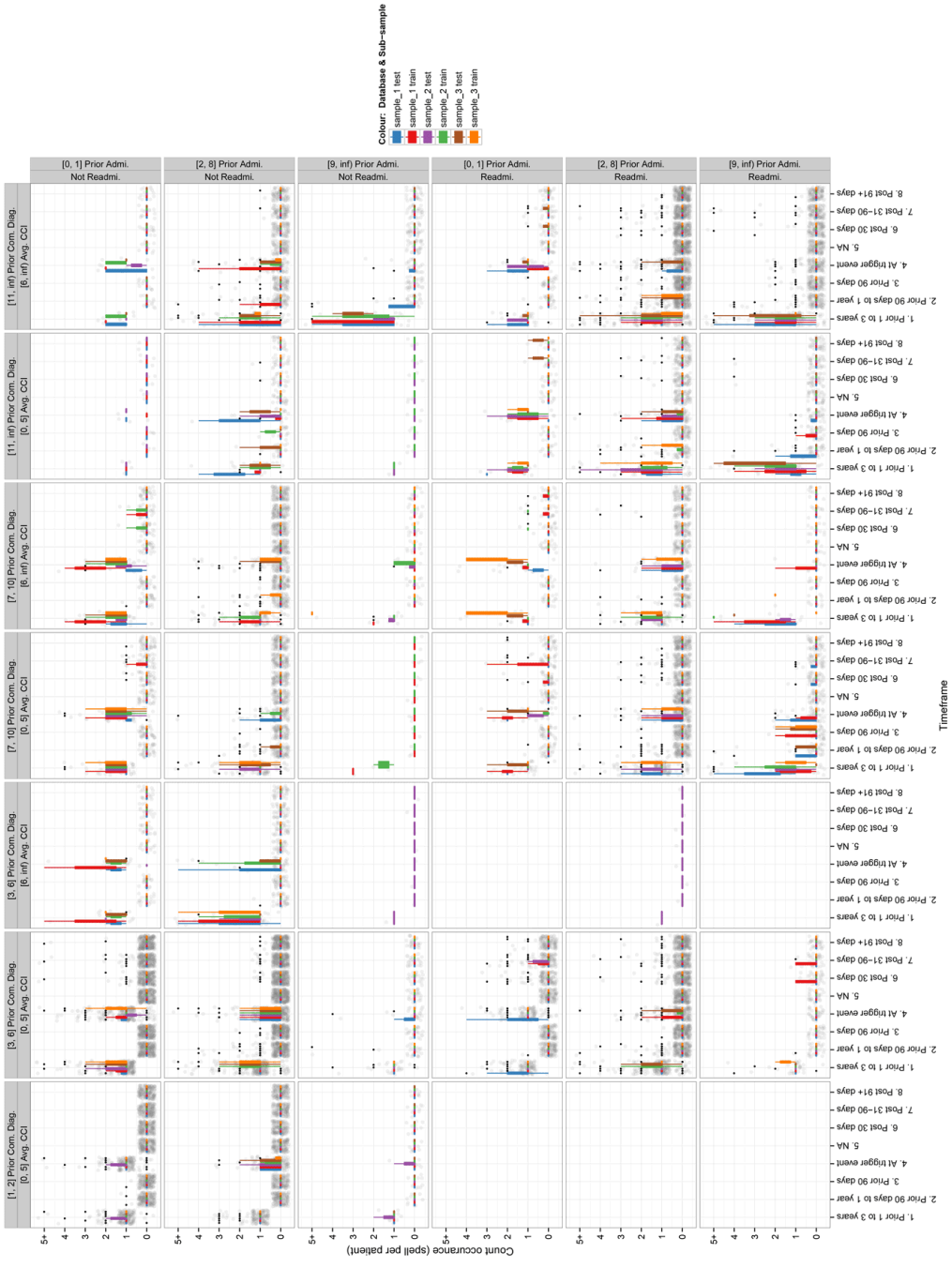

FIGURE A.46: The box-plot statistics of blood loss anemia (all samples)

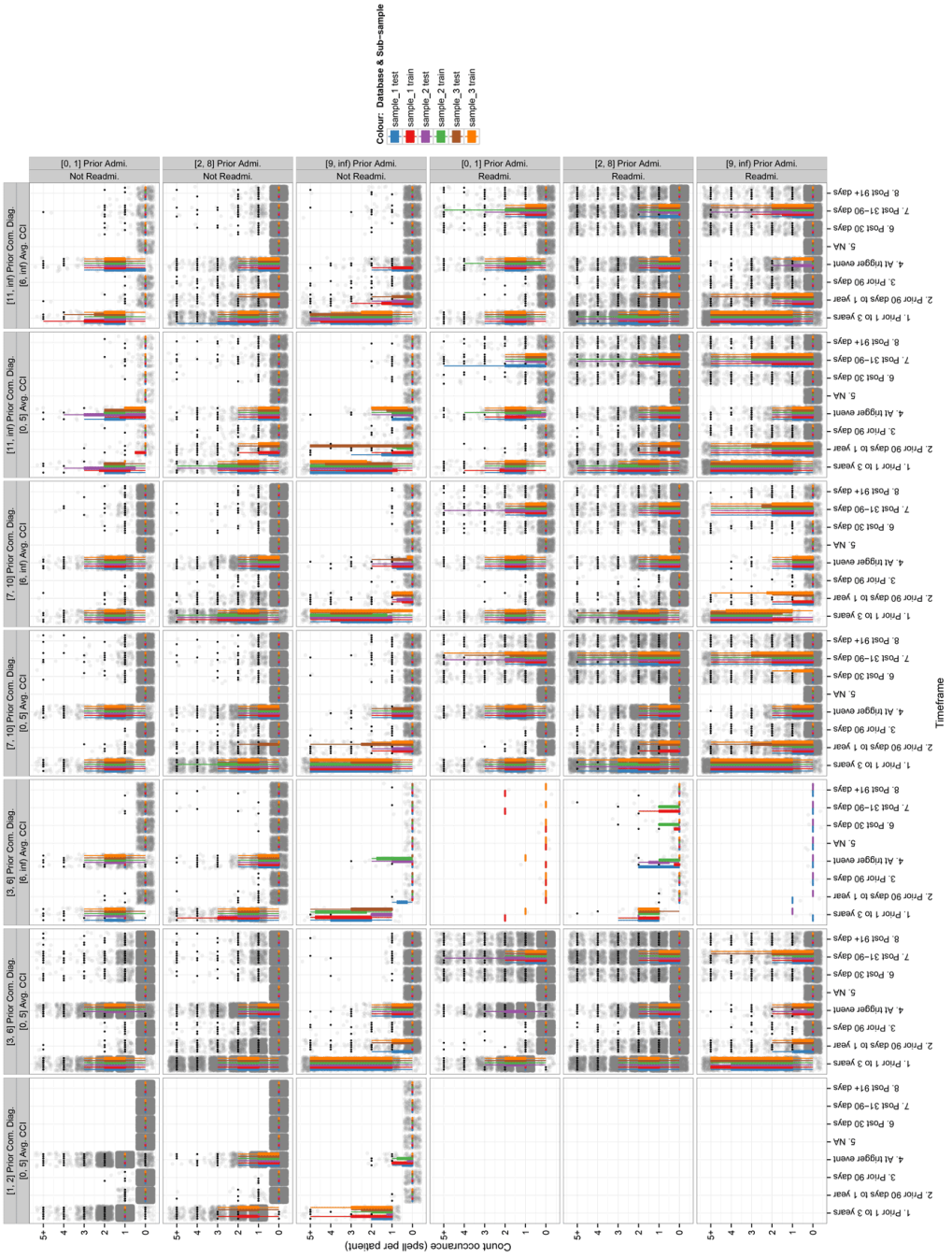

FIGURE A.47: The box-plot statistics of deficiency anemia (all samples)

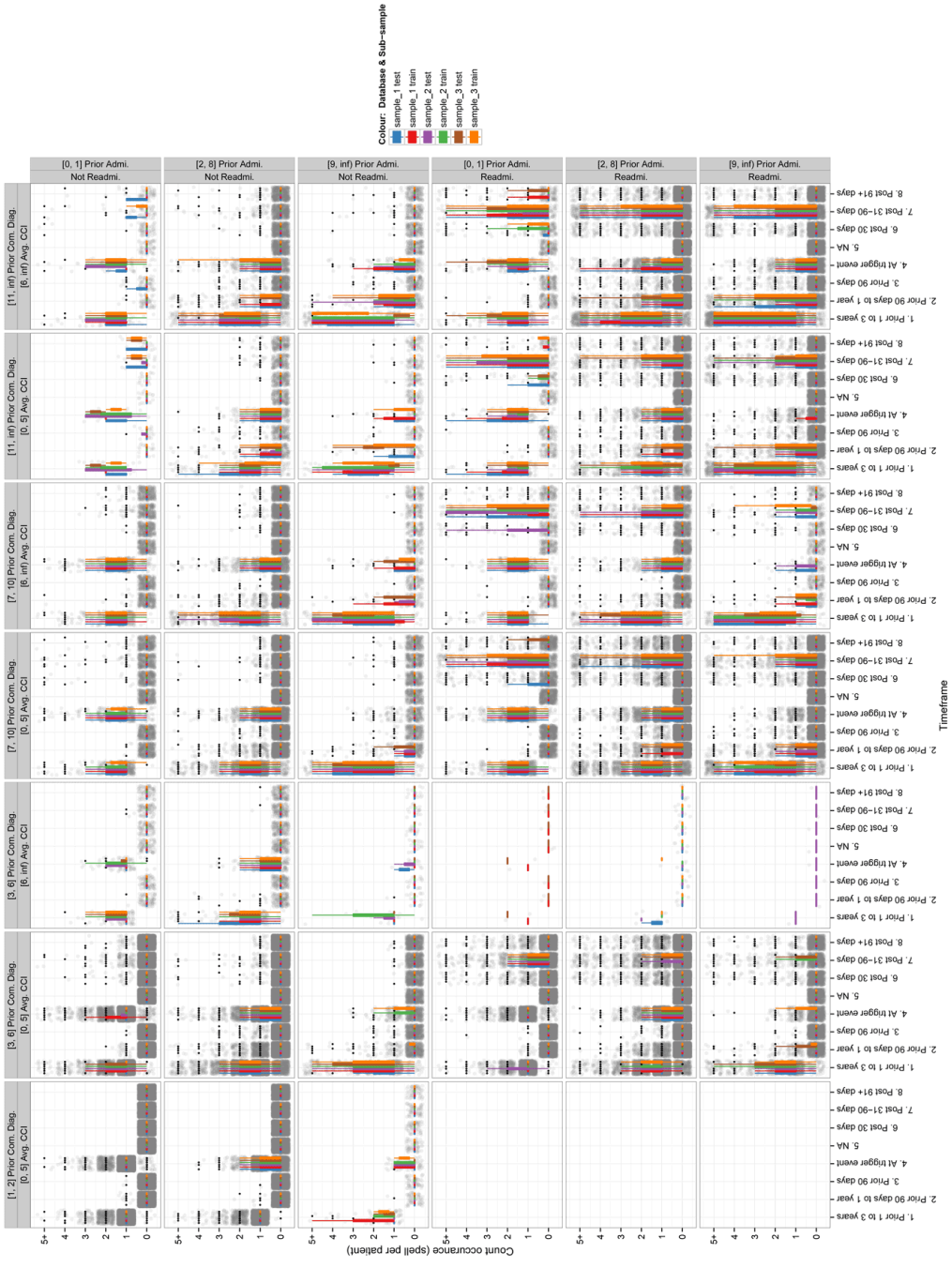

FIGURE A.48: The box-plot statistics of alcohol abuse (all samples)

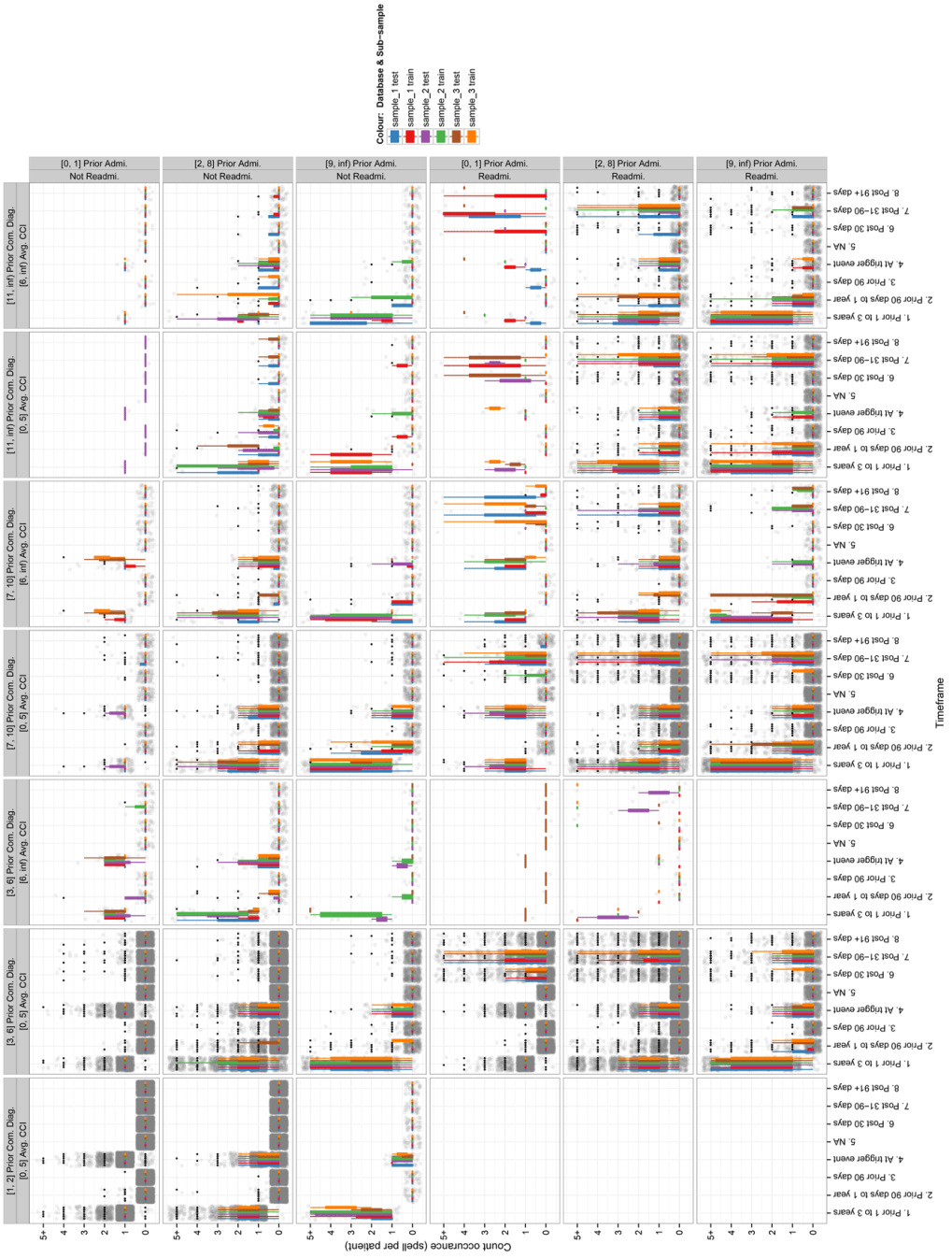

FIGURE A.49: The box-plot statistics of drug abuse (all samples)

A.6 ERMER

A.6.1 Features

A.6.1.1 Features Definitions

Table A.20: ERMER: Definition of the included features (all samples)

Feature Name	Definition	Feature's Source
readmiGap_avg	Average of days gap between spells in the past.	11. readmiGap
readmiGap_avg__trigger	Count of days gap from the previous spell at the trigger event.	11. readmiGap
s_spellAdmiMeth_acute_freq__delta	Count of the acute admission method between 12 to 36 months.	12. admimethRecoded_freq
s_spellAdmiMeth_acute_freq__365days__delta	Count of the acute admission method between 90 days to 12 months.	12. admimethRecoded_freq
s_spellAdmiMeth_acute_freq__90days__delta	Count of the acute admission method within 90 days.	12. admimethRecoded_freq
s_spellAdmiMeth_elective_freq__delta	Count of the elective admission method between 12 to 36 months.	12. admimethRecoded_freq
s_spellAdmiMeth_elective_freq__90days__delta	Count of the elective admission method within 90 days.	12. admimethRecoded_freq
admisorcRecoded_other_freq	Count of the admission source of state "NA", "Transferred from others" or "Maternity".	13. admisorcRecoded_freq
intmanigRecoded_other_freq__90days	Count of recoded intended management admission of state "NA", "Transferred from others" or "Maternity" within 90 days.	14. intmanigRecoded
epidur_maxAvg__365days	Average of spells durations within 12 months.	15. epidur
epidur_maxAvg__trigger	Average of spells durations at the trigger event.	15. epidur
epidurRecoded_avg	Average of spells durations in the past.	16. epidurRecoded
posopdur_avg	Average of post-operative durations in the past.	17. posopdur
posopdur_avg__365days	Average of post-operative durations within 12 months.	17. posopdur
posopdur_avg__trigger	Average of post-operative durations at the trigger event.	17. posopdur
preopdur_avg	Average of pre-operative durations in the past.	18. preopdur
preopdur_avg__365days	Average of pre-operative durations within 12 months.	18. preopdur
preopdur_avg__trigger	Average of pre-operative durations at the trigger event.	18. preopdur
rotreatRecoded_freq1__trigger_1	Count of recoded region of treatment of state "NA" at the trigger event.	20. rotreatRecoded
rotreatRecoded_freq1__trigger_2	Count of recoded region of treatment of state "Northern and Yorkshire" at the trigger event.	20. rotreatRecoded
rotreatRecoded_freq1__trigger_3	Count of recoded region of treatment of state "Trent" at the trigger event.	20. rotreatRecoded
rotreatRecoded_freq1__trigger_4	Count of recoded region of treatment of state "West Midlands" at the trigger event.	20. rotreatRecoded
rotreatRecoded_freq1__trigger_5	Count of recoded region of treatment of state "North West" at the trigger event.	20. rotreatRecoded
rotreatRecoded_freq1__trigger_6	Count of recoded region of treatment of state "Eastern" at the trigger event.	20. rotreatRecoded
rotreatRecoded_freq1__trigger_7	Count of recoded region of treatment of state "London" at the trigger event.	20. rotreatRecoded
rotreatRecoded_freq1__trigger_8	Count of recoded region of treatment of state "South East" at the trigger event.	20. rotreatRecoded
orgCluster_freq1__trigger_1	Count of organisation cluster of state "Acute teaching trust" at the trigger event.	21. orgCluster
orgCluster_freq1__trigger_3	Count of organisation cluster of state "Large acute trust" at the trigger event.	21. orgCluster
orgCluster_freq1__trigger_4	Count of organisation cluster of state "Medium acute trust" at the trigger event.	21. orgCluster
orgCluster_freq1__trigger_5	Count of organisation cluster of state "Small acute trust" at the trigger event.	21. orgCluster
protype_freq1_9	Count of provider type of state "TRUST" in the past.	22. protype
protype_freq1__trigger_9	Count of provider type of state "TRUST" at the trigger event.	22. protype
mainspef_uniques_freq__trigger	Count of unique main speciality seen at the trigger event.	25. mainspef
mainspefRecoded_01	Count of recoded main speciality of state "A&E" in the past.	26. mainspefRecoded
mainspefRecoded_03	Count of recoded main speciality of state "Cardiothoracic" in the past.	26. mainspefRecoded
mainspefRecoded_06	Count of recoded main speciality of state "Ear, nose & throat (ENT)" in the past.	26. mainspefRecoded
mainspefRecoded_08	Count of recoded main speciality of state "Gastroenterology " in the past.	26. mainspefRecoded
mainspefRecoded_09	Count of recoded main speciality of state "General" in the past.	26. mainspefRecoded
mainspefRecoded_10	Count of recoded main speciality of state "General Surgery" in the past.	26. mainspefRecoded
mainspefRecoded_11	Count of recoded main speciality of state "Geriatric" in the past.	26. mainspefRecoded

Continued on next page

Feature Name	Definition	Feature's Source
mainspefRecoded_12	Count of recoded main speciality of state "Gynaecology" in the past.	26. mainspefRecoded
mainspefRecoded_15	Count of recoded main speciality of state "Maternity" in the past.	26. mainspefRecoded
mainspefRecoded_19	Count of recoded main speciality of state "Ophthalmology" in the past.	26. mainspefRecoded
mainspefRecoded_27	Count of recoded main speciality of state "Plastic" in the past.	26. mainspefRecoded
mainspefRecoded_28	Count of recoded main speciality of state "Psychiatry" in the past.	26. mainspefRecoded
mainspefRecoded_31	Count of recoded main speciality of state "Respiratory" in the past.	26. mainspefRecoded
mainspefRecoded_33	Count of recoded main speciality of state "Urology" in the past.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_01	Count of recoded main speciality of state "A&E" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_03	Count of recoded main speciality of state "Cardiothoracic" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_06	Count of recoded main speciality of state "Ear, nose & throat (ENT)" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_08	Count of recoded main speciality of state "Gastroenterology " at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_09	Count of recoded main speciality of state "General" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_10	Count of recoded main speciality of state "General Surgery" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_11	Count of recoded main speciality of state "Geriatric" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_12	Count of recoded main speciality of state "Gynaecology" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_15	Count of recoded main speciality of state "Maternity" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_19	Count of recoded main speciality of state "Ophthalmology" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_27	Count of recoded main speciality of state "Plastic" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_28	Count of recoded main speciality of state "Psychiatry" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_31	Count of recoded main speciality of state "Respiratory" at the trigger event.	26. mainspefRecoded
mainspefRecoded_freq1__trigger_33	Count of recoded main speciality of state "Urology" at the trigger event.	26. mainspefRecoded
elecdur_elective_nulls_freq	Count of zero waiting time for elective admissions in the past.	29. elecdur
charlsonIndex_avg	Average value of the Charlson Index in the past.	30. charlsonIndex
charlsonIndex_max__365days	Maximum value of the Charlson Index within 12 months.	30. charlsonIndex
diagCci_44_anemia_freq__90days	Count of blood loss anemia conditions within 90 days.	36. diagCci_44_anemia_freq
diagCat_3_freq__90days__sum	Count of ischemic (coronary) heart conditions within 90 days, including the trigger event.	44. diagCat_3_freq
diagCat_5_freq__90days__sum	Count of other forms of heart conditions within 90 days, including the trigger event.	46. diagCat_5_freq
diagCat_9_freq__90days	Count of other veins, lymphatics and lymph nodes conditions within 90 days.	50. diagCat_9_freq
diagCci_02_chf_freq__90days	Count of congestive heart failure conditions within 90 days.	52. diagCci_02_chf_freq
diagCci_19_cardiac_freq__90days__sum	Count of cardiac arrhythmias conditions within 90 days, including the trigger event.	56. diagCci_19_cardiac_freq
diagRisk_3_blood_freq__90days__sum	Count of thrombocytopenia & thrombocytosis & elevated white blood cell count conditions within 90 days, including the trigger event.	66. diagRisk_3_blood_freq
diagRisk_7_kidney_freq__90days	Count of liver conditions within 90 days.	74. diagRisk_7_kidney_freq
diagOther_4_chronic_d_freq__90days__sum	Count of diabetes conditions (ACSC category 'd') within 90 days, including the trigger event.	85. diagOther_4_chronic_d_freq
diagRisk_2_Cholesterol_freq__90days	Count of disorders of lipidemias conditions within 90 days.	86. diagRisk_2_Cholesterol_freq
diagRisk_10_externalMorbidity_freq__90days	Count of external causes of morbidity conditions within 90 days.	88. diagRisk_10_externalMorbidity_freq
diagRisk_9_external_freq__90days__sum	Count of injury, poisoning & certain other consequences of external causes or complications, which not elsewhere classified within 90 days, including the trigger event.	94. diagRisk_9_external_freq
diagCci_18_depression_freq__90days	Count of depression conditions within 90 days.	96. diagCci_18_depression_freq
diagOther_8_mental_freq__90days__sum	Count of mental conditions within 90 days, including the trigger event.	100. diagOther_8_mental_freq
diagCci_14_malignancy_freq__90days	Count of malignancy conditions, including lymphoma & leukemia, except malignant neoplasm of skin within 90 days.	103. diagCci_14_malignancy_freq

Continued on next page

Feature Name	Definition	Feature's Source
diagOther_7_cancer_freq__90days	Count of neoplasm conditions within 90 days.	108. diagOther_7_cancer_freq
diagCci_04_cerebrovascular_freq__90days	Count of cerebrovascular (stroke) conditions within 90 days.	109. diagCci_04_cerebrovascular_freq
diagCci_26_neuroOther_freq__90days	Count of other neurological disorders within 90 days.	112. diagCci_26_neuroOther_freq
diagOther_4_chronic_i_freq__90days__sum	Count of mental & behavioural disorders or neurological disorders (ACSC category 'i') within 90 days, including the trigger event.	114. diagOther_4_chronic_i_freq
diagRisk_8_smoke_freq__90days	Count of exposure to tobacco smoke conditions at the trigger event.	121. diagRisk_8_smoke_freq
reference_freq__90days__sum	Count of PARR's "reference" conditions in the HRG record within 90 days, including the trigger event.	119. reference_freq
diagCci_06_cpd_freq__trigger	Count of chronic pulmonary conditions at the trigger event.	120. diagCci_06_cpd_freq
diagMorbid_1_Influenza_freq__90days	Count of influenza a pneumonia conditions within 90 days.	121. diagMorbid_1_Influenza_freq
diagOther_4_chronic_b_freq__90days__sum	Count of respiratory conditions (ACSC category 'b') within 90 days, including the trigger event.	122. diagOther_4_chronic_b_freq
diagOther_4_chronic_e_freq__90days__sum	Count of respiratory conditions (ACSC category 'e') within 90 days, including the trigger event.	123. diagOther_4_chronic_e_freq
spellOpertn_freq__365days__delta	Sum of number of operations between 90 days to 12 months.	128. spellOpertn_freq
spellOpertn_freq__90days__delta	Sum of number of operations within 90 days.	128. spellOpertn_freq
spellOpertn_freq__trigger	Sum of number of operations at the trigger event.	128. spellOpertn_freq
oper_2_heart_freq__90days	Count of heart operations within 90 days.	129. oper_2_heart_freq
oper_3_urinary_freq__90days	Count of urinary operations within 90 days.	131. oper_3_urinary_freq
triggerStartAge	State age of patient at the trigger event.	133. triggerStartAge
imd04rkRecoded	State of the recoded Index of Multiple Deprivation Overall Rank of a patient.	135. imd04rk
ethnosRecoded_0	Having recoded ethnicity of state "NA".	138. ethnosRecoded
ethnosRecoded_1	Having recoded ethnicity of state "White".	138. ethnosRecoded
gender_2	Having recoded gender of state "Female".	139. genderRecoded

A.6.1.2 Features Importance Ranks

TABLE A.21: ERMER: Average importance and average weights of features in sub-models (all samples)

#	Feature Sub-models:	Main	Age_0	Age_1	$Oper_0$	$Oper_1$
1	Sum of number of operations (trigger)	-0.744	<-1	-0.049	<-1	-0.666
2	Count of recoded main speciality of state 'Maternity' (trigger)	0.885	>1	-0.608	>1	0.289
3	Count of recoded main speciality of state 'Maternity' (3 years)	-0.021	-0.024	0.138	0.033	0.019
4	Count of recoded main speciality of state 'Gynaecology' (trigger)	0.742	0.910	-0.514	>1	0.311
5	Having recoded gender of state 'Female'	0.005	0.020	-0.032	0.059	-0.056
6	Count of recoded main speciality of state 'Gynaecology' (3 years)	-0.005	-0.004	-0.017	0.093	0.026
7	Age of patient (trigger)	0.023	0.003	0.004	0.015	0.024
8	Average of post-operative durations (trigger)	0.008	0.007	0.001	<-1	0.007
9	Count of the acute admission method (90 days)	0.054	0.053	0.043	0.044	0.049
10	Average of spells durations (3 years)	0.041	0.032	0.056	0.047	0.063
11	Sum of number of operations (90 days)	-0.011	-0.008	-0.002	>1	-0.004
12	Count of the acute admission method between (1-2 years)	-0.043	-0.031	0.075	-0.138	0.061
13	Count of recoded main speciality of state 'General' (trigger)	-0.077	-0.037	0.025	0.006	-0.183
14	Average of gaps between admissions (3 years)	0.236	0.189	0.180	0.375	0.154
15	Average of spells durations (trigger)	-0.002	-0.002	-0.002	-0.002	-0.001
16	Having recoded ethnicity of state 'others'	-0.363	-0.391	-0.368	-0.429	-0.320
17	Average of the Charlson Index (3 years)	0.018	0.040	0.009	0.011	0.018
18	Count of recoded main speciality of state 'General' (3 years)	0.015	0.021	0.003	0.013	0.012
19	Average of post-operative durations (3 years)	0.000	0.003	-0.001	>1	-0.002
20	Count of recoded main speciality of state 'General Surgery' (trigger)	-0.046	0.049	-0.064	-0.068	-0.145
21	Count of the acute admission method between 90 days to 12 months	-0.153	-0.152	-0.047	-0.314	-0.030
22	Average of pre-operative durations (trigger)	0.017	0.021	0.001	<-1	0.013
23	Count of recoded main speciality of state 'Plastic' (trigger)	0.025	0.161	-0.110	-0.164	-0.109
24	Having recoded ethnicity of state 'White'	0.010	-0.013	0.021	-0.003	0.015
25	Count of PARR's 'reference' conditions (90 days, trigger)	0.014	0.036	0.025	0.058	0.008
26	Count of recoded main speciality of state 'Geriatric' (3 years)	0.007	0.045	0.003	-0.011	0.006
27	Recoded Index of Multiple Deprivation Overall Rank (10 equal ranges)	-0.002	-0.006	0.002	0.013	-0.008
28	Maximum value of the Charlson Index (1 year)	-0.001	-0.009	0.002	-0.007	-0.003
29	Average of pre-operative durations (3 years)	-0.002	0.001	0.002	0.126	-0.004
30	Count of recoded main speciality of state 'General Surgery' (3 years)	0.018	0.024	0.002	-0.012	0.006
31	Count of recoded main speciality of state 'Plastic' (3 years)	0.002	0.013	-0.010	-0.001	-0.016
32	Count of external causes or complications (3 years)	0.007	0.003	0.003	0.020	-0.005
33	Count of recoded main speciality of state 'Geriatric' (trigger)	-0.017	-0.053	0.056	0.145	-0.136
34	Count of recoded main speciality of state 'A&E' (trigger)	-0.204	-0.182	0.013	-0.065	-0.335
35	Count of ischemic heart conditions (90 days, trigger)	-0.008	-0.012	0.002	0.003	-0.005
36	Count of unique main speciality seen (trigger)	0.119	0.150	0.029	-0.007	0.090
37	Average of post-operative durations (1 year)	-0.001	-0.002	-0.001	<-1	0.000
38	Count of other heart conditions (90 days, trigger)	0.016	0.011	0.009	0.013	0.007
39	Count of the elective admission method (90 days)	0.009	0.004	0.008	0.000	0.002
40	Count of thrombocytopenia, thrombocytosis & high WBC (90 days, trigger)	0.004	0.004	-0.002	-0.004	0.000
41	Count of recoded intended admission of states 'others' or 'Maternity' (90 days)	-0.016	-0.014	-0.006	-0.016	-0.008
42	Count of recoded main speciality of state 'A&E' (3 years)	0.032	0.037	0.034	0.074	0.037
43	Count of ACS respiratory conditions (90 days, trigger)	0.014	0.019	0.014	0.026	0.011
44	Count of ACS neurological disorders (90 days, trigger)	-0.014	0.001	-0.005	-0.019	-0.012
45	Count of mental conditions (90 days, trigger)	-0.001	0.052	-0.023	-0.003	0.003
46	Count of recoded main speciality of state 'Psychiatry' (3 years)	0.004	0.007	-0.013	0.019	-0.002
47	Count of recoded main speciality of state 'Psychiatry' (trigger)	-0.078	-0.029	0.058	0.035	-0.181
48	Count of the admission sources from 'others' or 'Maternity'	0.001	0.021	-0.012	-0.033	0.000
49	Count of chronic pulmonary conditions (trigger)	0.013	-0.039	0.027	0.034	0.018
50	Count of recoded main speciality of state 'Cardiothoracic' (3 years)	0.003	0.014	-0.013	0.047	-0.008
51	Count of ACS diabetes conditions (90 days, trigger)	0.013	0.022	0.010	0.012	0.013
52	Count of blood loss anemia conditions (90 days)	0.022	-0.019	0.008	0.032	0.019
53	Average of pre-operative durations (1 year)	0.003	0.001	-0.001	0.516	0.003
54	Count of recoded main speciality of state 'ENT' (trigger)	-0.080	-0.053	-0.127	-0.199	-0.144
55	Count of recoded region of state 'Eastern' (trigger)	0.455	0.534	0.583	0.593	0.220
56	Sum of number of operations between 90 days to 12 months	-0.020	-0.026	0.040	>1	-0.019
57	Count of organisation cluster of state 'Acute teaching trust' (trigger)	-0.029	-0.022	-0.030	-0.055	-0.020
58	Count of cardiac arrhythmias conditions (90 days, trigger)	0.002	-0.018	0.005	0.009	0.007
59	Count of congestive heart failure conditions (90 days)	-0.036	-0.046	-0.025	-0.030	-0.027
60	Count of recoded main speciality of state 'Ophthalmology' (3 years)	0.057	0.025	0.048	-0.084	0.033
61	Count of recoded main speciality of state 'Gastroenterology ' (3 years)	0.032	0.045	0.009	0.049	0.017
62	Count of ACS respiratory conditions (90 days, trigger)	-0.009	-0.015	-0.008	-0.005	-0.008
63	Count of recoded main speciality of state 'Cardiothoracic' (trigger)	0.084	0.174	0.081	0.081	-0.007
64	Count of organisation cluster of state 'Large acute trust' (trigger)	-0.058	-0.048	-0.035	-0.034	-0.052
65	Count of recoded main speciality of state 'ENT' (3 years)	0.033	0.038	0.043	-0.060	0.016
66	Count of recoded region of state 'Trent' (trigger)	0.506	0.592	0.602	0.654	0.284

Continued on next page

#	Feature Sub-models:	Main	Age_0	Age_1	Oper_0	Oper_1
67	Count of other neurological disorders (90 days)	0.007	-0.005	-0.002	0.018	0.008
68	Count of recoded region of state 'West Midlands' (trigger)	0.510	0.619	0.549	0.660	0.270
69	Count of recoded region of state 'London' (trigger)	0.511	0.604	0.616	0.627	0.291
70	Count of neoplasm conditions (90 days)	0.021	0.096	0.013	-0.039	0.026
71	Average of spells durations (1 year)	0.000	0.000	0.000	0.000	0.000
72	Count of recoded region of state 'North West' (trigger)	0.471	0.577	0.560	0.586	0.245
73	Count of the elective admission method between (1-2 years)	0.011	0.012	0.051	0.064	0.007
74	Count of heart operations (90 days)	-0.005	-0.008	-0.015	>1	-0.015
75	Count of recoded main speciality of state 'Urology' (3 years)	0.023	0.018	0.025	0.062	0.014
76	Count of organisation cluster of state 'Medium acute trust' (trigger)	-0.040	-0.053	-0.019	-0.027	-0.049
77	Count of recoded region of state 'Northern and Yorkshire' (trigger)	0.503	0.604	0.572	0.634	0.279
78	Count of provider type of state 'Trust' (trigger)	-0.107	-0.148	-0.124	0.014	-0.073
79	Count of malignancy conditions, except malignant neoplasm of skin (90 days)	-0.019	-0.096	-0.003	0.016	-0.023
80	Count of recoded main speciality of state 'Respiratory' (3 years)	0.026	0.049	-0.001	0.042	0.013
81	Count of depression conditions (90 days)	-0.007	-0.010	0.004	-0.016	0.004
82	Count of recoded main speciality of state 'Urology' (trigger)	0.069	0.080	0.012	-0.134	-0.013
83	Count of provider type of state 'Trust' (3 years)	0.085	0.137	0.126	-0.004	0.076
84	Count of recoded main speciality of state 'Gastroenterology' (trigger)	-0.012	0.064	0.029	0.011	-0.079
85	Count of recoded region of state 'others' (trigger)	0.519	0.620	0.599	0.679	0.284
86	Count of zero waiting time for elective admissions (3 years)	0.002	0.005	-0.010	-0.008	0.001
87	Count of organisation cluster of state 'Small acute trust' (trigger)	-0.054	-0.066	-0.017	-0.017	-0.050
88	Count of recoded region of state 'South East' (trigger)	0.540	0.643	0.614	0.617	0.326
89	Count of recoded main speciality of state 'Respiratory' (trigger)	-0.052	0.001	0.051	-0.028	-0.110
90	Count of liver conditions (90 days)	-0.003	-0.004	0.010	-0.009	0.001
91	Count of urinary operations (90 days)	-0.005	0.018	-0.022	>1	-0.008
92	Count of exposure to tobacco smoke conditions (trigger)	0.008	0.009	0.024	0.020	0.010
93	Count of external causes of morbidity conditions (90 days)	-0.014	-0.007	-0.008	0.037	-0.009
94	Count of cerebrovascular (stroke) conditions (90 days)	-0.022	-0.027	-0.012	0.008	-0.020
95	Count of disorders of lipidemias conditions (90 days)	-0.035	-0.043	-0.036	0.005	-0.033
96	Count of influenza a pneumonia conditions (90 days)	0.024	-0.003	0.012	-0.004	0.019
97	Count of recoded main speciality of state 'Ophthalmology' (trigger)	-0.076	0.008	-0.363	0.013	-0.250
98	Count of other veins, lymphatics and lymph nodes conditions (90 days)	-0.010	-0.005	-0.019	-0.002	-0.007
99	Count of days gap from the previous spell (trigger)	-0.727	-0.695	-0.662	-0.840	-0.521

A.6.2 Weights

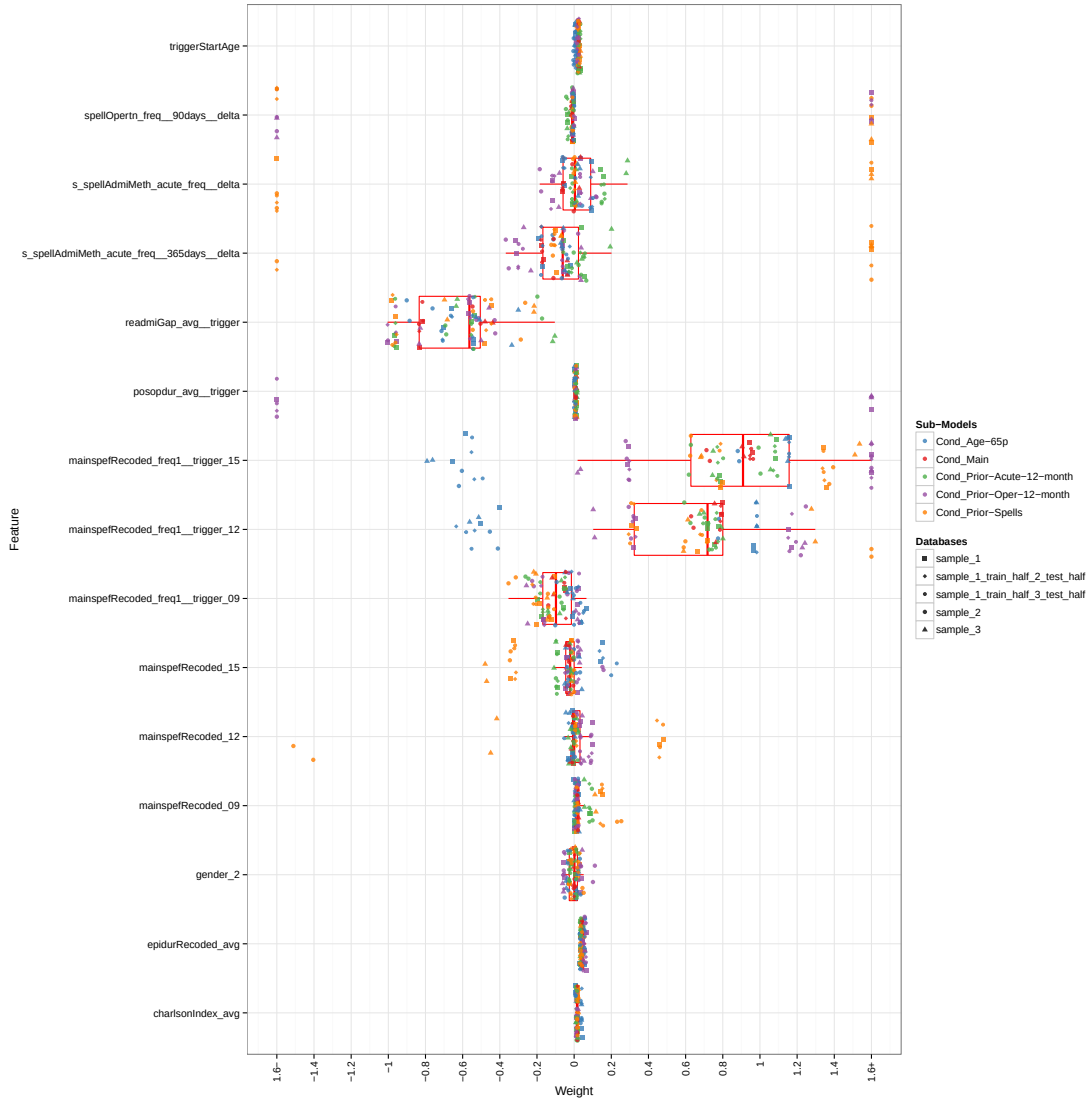

FIGURE A.50: ERMER: Sub-models' features weights (all samples - part 1)

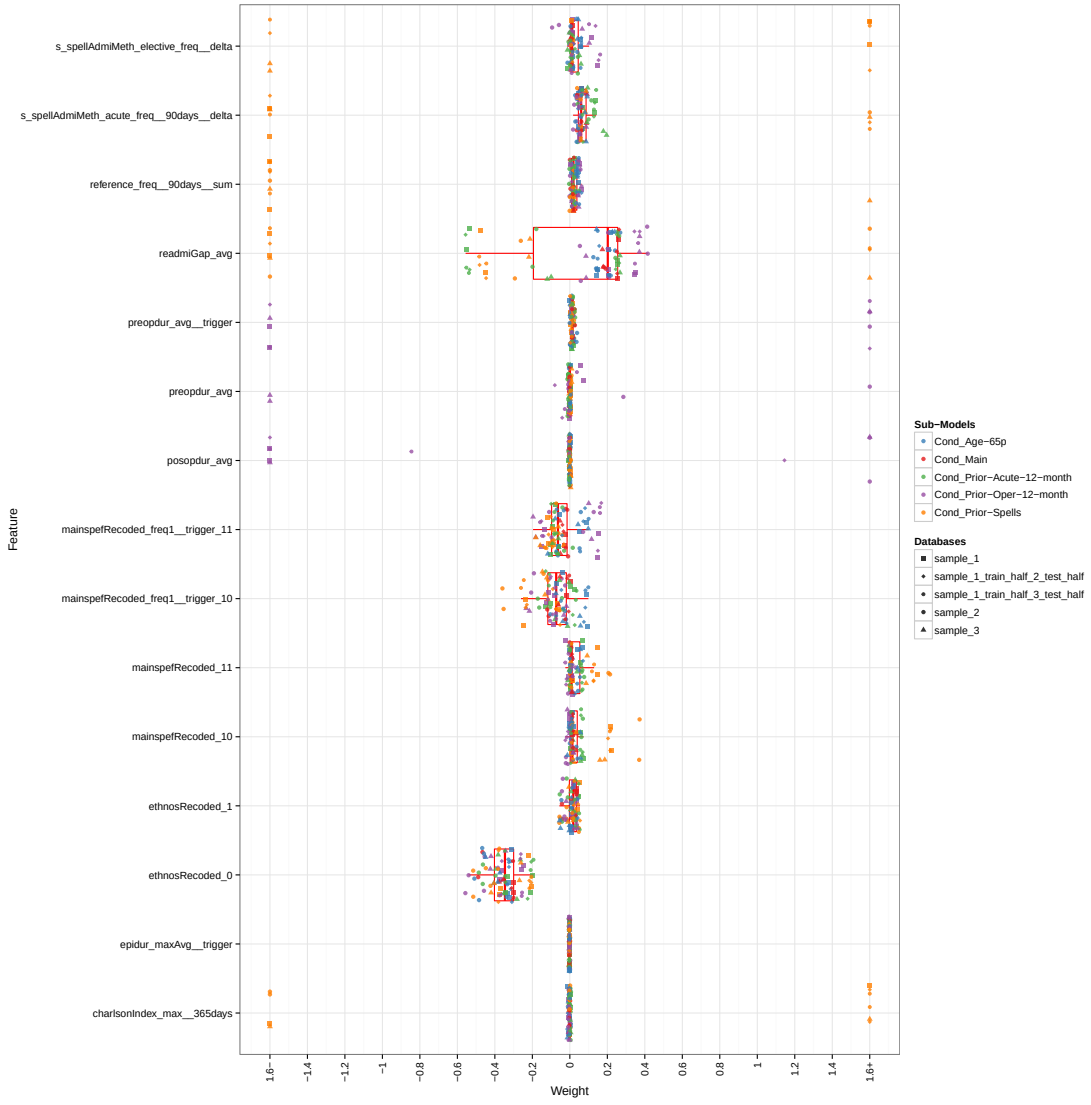

FIGURE A.51: ERMER: Sub-models' features weights (all samples - part 2)

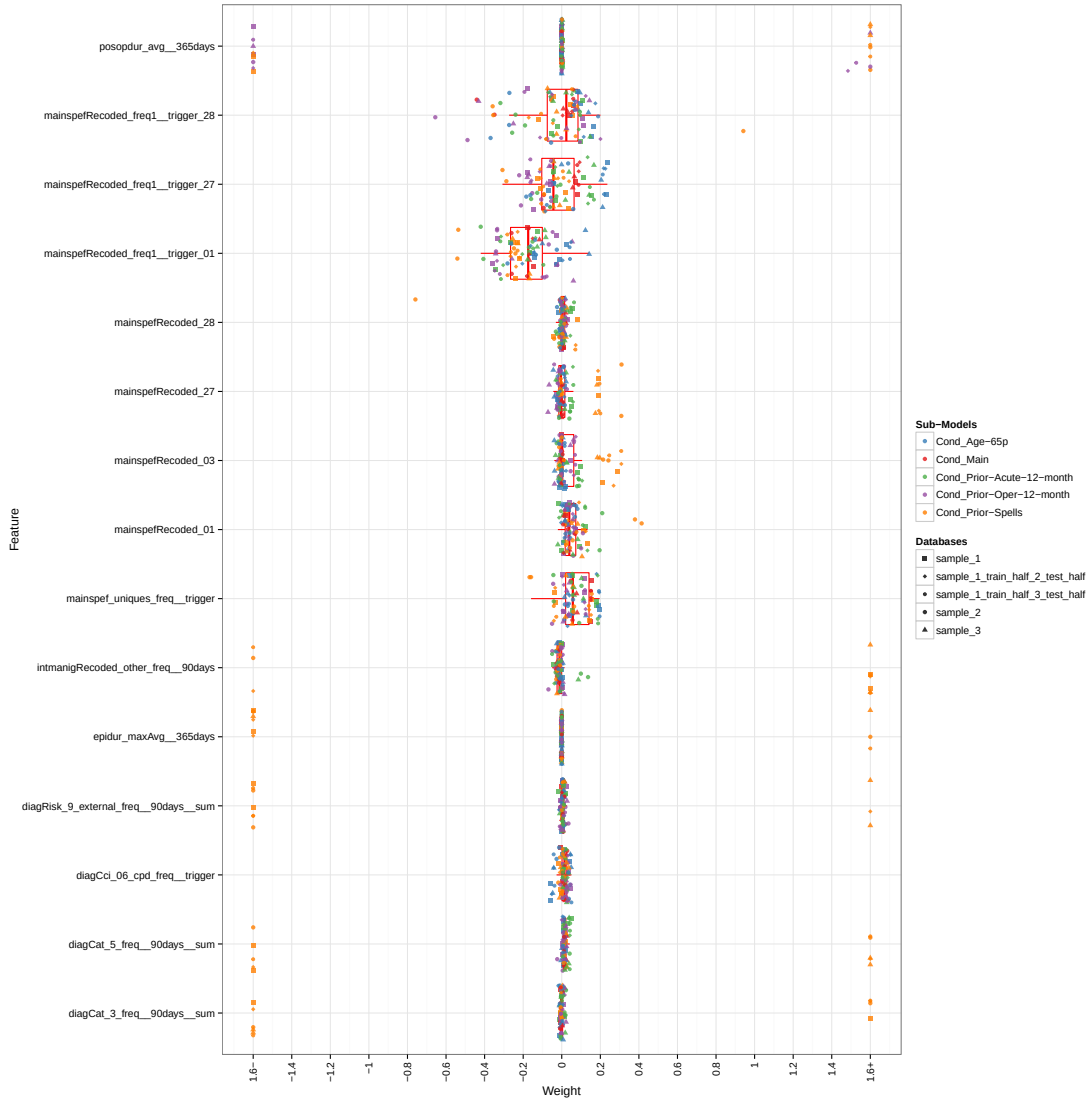

FIGURE A.52: ERMER: Sub-models' features weights (all samples - part 3)

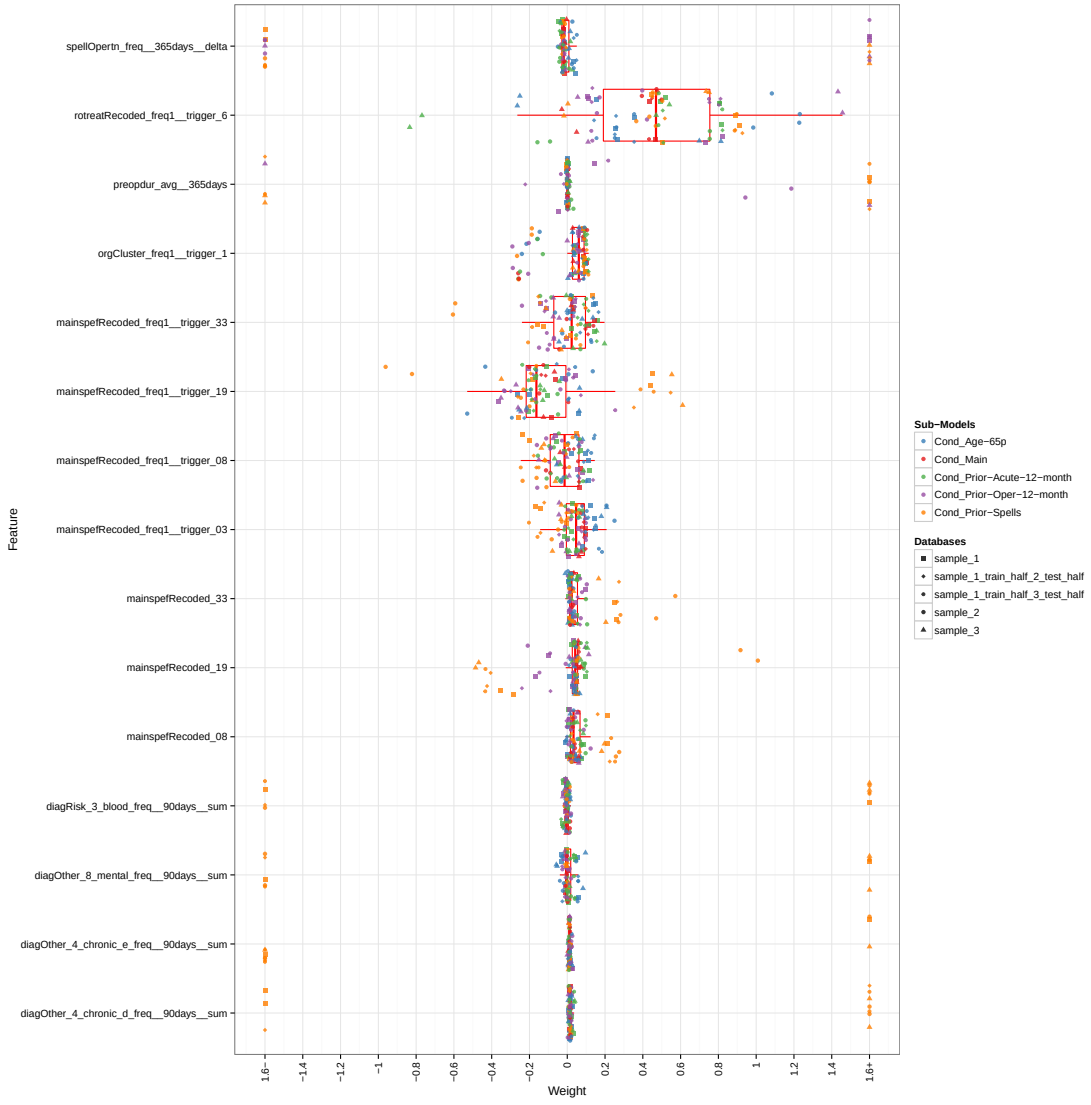

FIGURE A.53: ERMER: Sub-models' features weights (all samples - part 4)

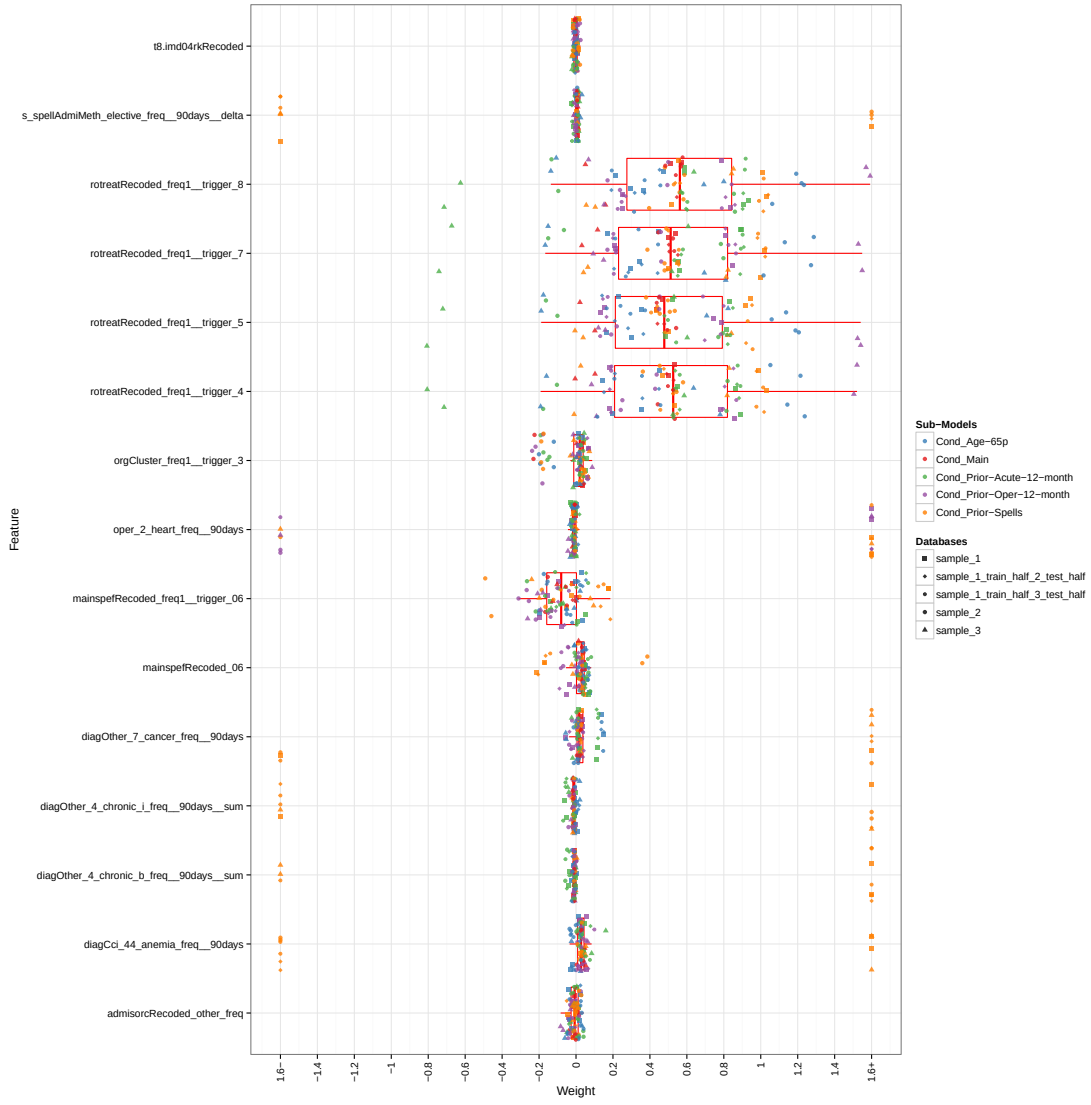

FIGURE A.54: ERMER: Sub-models' features weights (all samples - part 5)

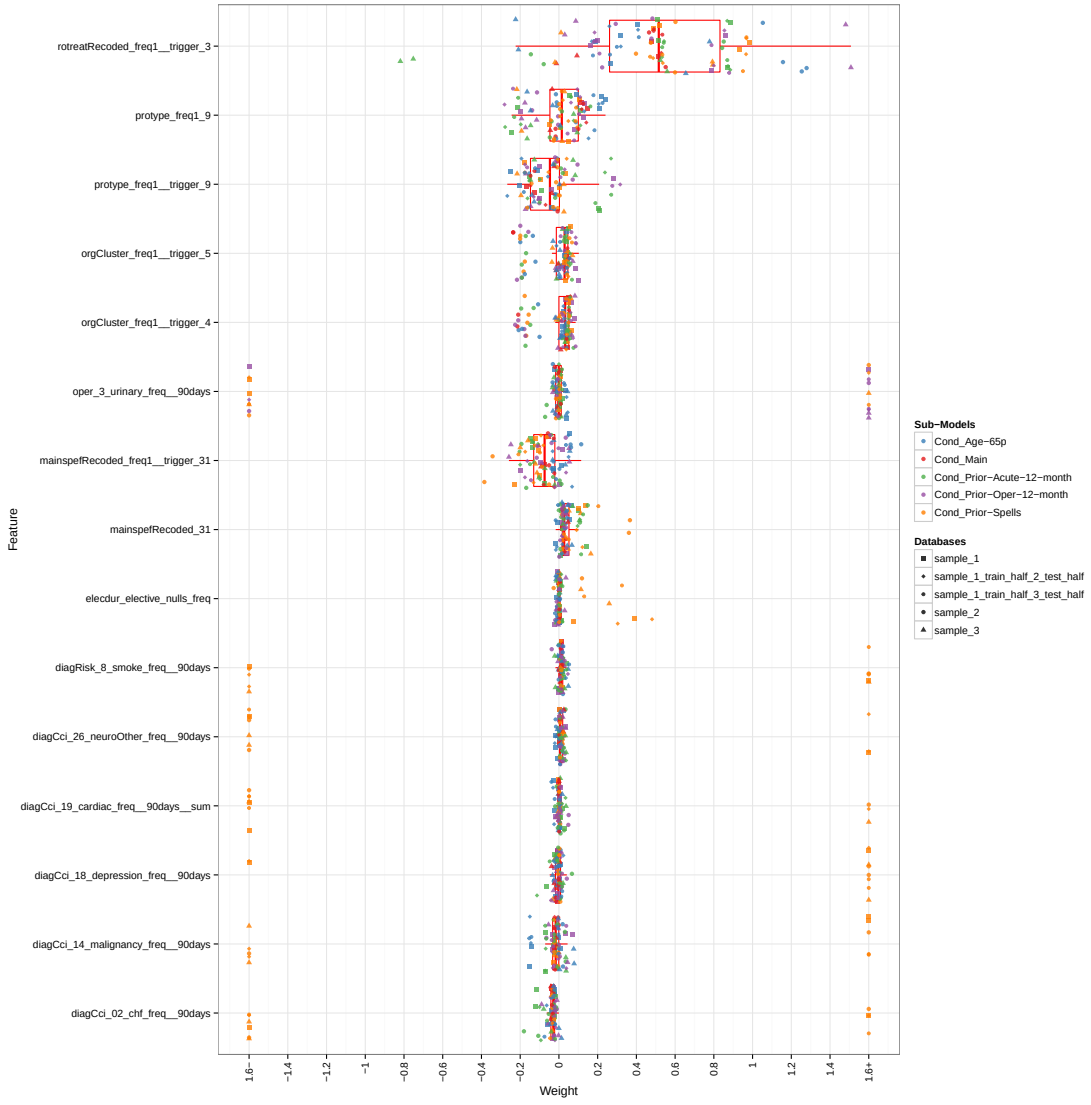

FIGURE A.55: ERMER: Sub-models' features weights (all samples - part 6)

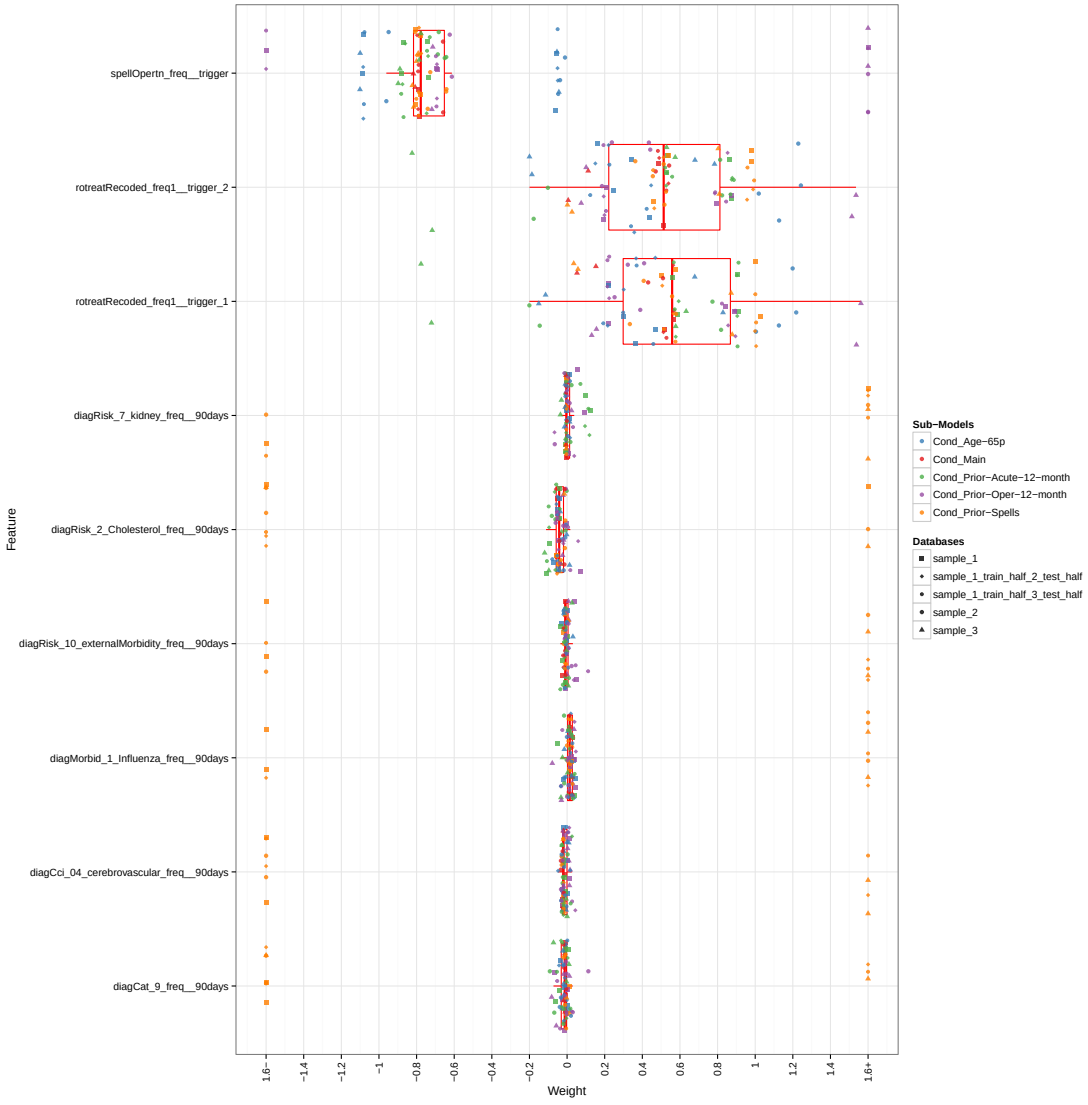

FIGURE A.56: ERMER: Sub-models' features weights (all samples - part 7)

A.6.3 Performance of Sub-Models

A.6.3.1 Summary Performance Statistics

TABLE A.22: ERMER: Performance statistics of submodels (*Sample-1*)

Modelling Approach: BPM;
Sample: Sample-1

Sub-model[a]	If[b]	N[c]	TP+FP[d]	TP[e]	Spec.[f]	Prec.[g]	Sens.[h]	F1[i]	AUC[j]
Modelling Group: Pop_Any-Acute-No-Mental									
Cond_Age-65p	0	158,180	126,801	50,470	0.93	0.79	0.52	0.62	0.78
Cond_Age-65p	1	69,281	43,235	30,270	0.81	0.61	0.39	0.47	0.66
Cond_Main		227,461	167,147	80,740	0.89	0.70	0.45	0.55	0.74
Cond_Prior-Acute-12-month	0	142,858	108,438	41,610	0.94	0.69	0.31	0.43	0.72
Cond_Prior-Acute-12-month	1	84,603	58,801	39,130	0.75	0.69	0.63	0.66	0.74
Cond_Prior-Oper-12-month	0	82,103	61,535	29,615	0.91	0.74	0.47	0.57	0.78
Cond_Prior-Oper-12-month	1	145,358	104,771	51,125	0.88	0.65	0.44	0.52	0.75
Cond_Prior-Spells	0	101,790	79,841	26,579	0.96	0.73	0.28	0.40	0.72
Cond_Prior-Spells	1	125,671	87,620	54,161	0.81	0.69	0.55	0.61	0.73
Modelling Group: Pop_Any-Acute									
Cond_Age-65p	0	161,608	129,158	51,763	0.93	0.79	0.51	0.62	0.78
Cond_Age-65p	1	70,147	43,736	30,671	0.81	0.61	0.39	0.47	0.66
Cond_Main		231,755	170,034	82,434	0.89	0.70	0.45	0.54	0.74
Cond_Prior-Acute-12-month	0	145,127	110,022	42,277	0.94	0.69	0.30	0.42	0.72
Cond_Prior-Acute-12-month	1	86,628	60,032	40,157	0.75	0.68	0.63	0.65	0.74
Cond_Prior-Oper-12-month	0	85,022	63,510	30,632	0.91	0.74	0.46	0.57	0.78
Cond_Prior-Oper-12-month	1	146,733	105,632	51,802	0.87	0.65	0.44	0.53	0.75
Cond_Prior-Spells	0	103,415	81,042	26,997	0.96	0.73	0.27	0.40	0.72
Cond_Prior-Spells	1	128,340	89,244	55,437	0.81	0.69	0.55	0.61	0.73
Modelling Group: Pop_PARR-1-Settings									
Cond_Main		20,697	12,456	10,557	0.65	0.62	0.56	0.59	0.65
Cond_Prior-Acute-12-month	0	9,781	5,769	4,162	0.83	0.54	0.27	0.36	0.60
Cond_Prior-Acute-12-month	1	10,91	6,614	6,395	0.37	0.63	0.77	0.70	0.63
Cond_Prior-Oper-12-month	0	7,928	4,757	3,698	0.74	0.60	0.44	0.51	0.64
Cond_Prior-Oper-12-month	1	12,769	7,674	6,859	0.58	0.63	0.62	0.62	0.64
Modelling Group: Pop_PARR-2-Settings									
Cond_Main		70,147	43,736	30,671	0.81	0.61	0.39	0.47	0.66
Cond_Prior-Acute-12-month	0	39,491	25,506	14,253	0.93	0.53	0.14	0.23	0.64
Cond_Prior-Acute-12-month	1	30,656	18,227	16,418	0.57	0.62	0.62	0.62	0.63
Cond_Prior-Oper-12-month	0	23,647	15,024	9,351	0.87	0.58	0.28	0.38	0.66
Cond_Prior-Oper-12-month	1	46,500	28,637	21,320	0.76	0.61	0.44	0.51	0.66

TABLE A.23: ERMER: Performance statistics of submodels (*Sample-2*)

Modelling Approach: BPM;
Sample: Sample-2

Sub-model	If	N	TP+FP	TP	Spec.	Prec.	Sens.	F1	AUC
Modelling Group: Pop_Any-Acute-No-Mental									
Cond_Age-65p	0	168,014	132,125	56,697	0.93	0.79	0.50	0.61	0.76
Cond_Age-65p	1	72,729	45,040	33,579	0.74	0.61	0.48	0.54	0.66
Cond_Main		240,743	174,571	90,276	0.88	0.70	0.47	0.56	0.73
Cond_Prior-Acute-12-month	0	115,727	88,744	33,665	0.95	0.72	0.32	0.45	0.71
Cond_Prior-Acute-12-month	1	125,016	86,433	56,611	0.78	0.69	0.58	0.63	0.72
Cond_Prior-Oper-12-month	0	56,612	44,106	19,862	0.93	0.79	0.51	0.62	0.80
Cond_Prior-Oper-12-month	1	184,131	129,862	70,414	0.86	0.67	0.46	0.54	0.73
Cond_Prior-Spells	0	72,903	58,333	18,789	0.96	0.76	0.33	0.46	0.71
Cond_Prior-Spells	1	167,840	117,090	71,487	0.82	0.69	0.53	0.60	0.72
Modelling Group: Pop_Any-Acute									
Cond_Age-65p	0	170,397	133,693	57,668	0.93	0.79	0.50	0.61	0.76
Cond_Age-65p	1	73,315	45,394	33,849	0.74	0.61	0.48	0.54	0.66
Cond_Main		243,712	176,355	91,517	0.88	0.70	0.46	0.56	0.73
Cond_Prior-Acute-12-month	0	116,721	89,399	33,978	0.95	0.72	0.32	0.44	0.71
Cond_Prior-Acute-12-month	1	126,991	87,665	57,539	0.78	0.69	0.58	0.63	0.72
Cond_Prior-Oper-12-month	0	58,133	45,103	20,429	0.93	0.79	0.50	0.61	0.80
Cond_Prior-Oper-12-month	1	185,579	130,727	71,088	0.86	0.67	0.46	0.54	0.73
Cond_Prior-Spells	0	73,536	58,784	18,969	0.96	0.76	0.33	0.46	0.71
Cond_Prior-Spells	1	170,176	118,504	72,548	0.82	0.69	0.52	0.60	0.71
Modelling Group: Pop_PARR-1-Settings									
Cond_Main		16,213	9,775	8,544	0.55	0.62	0.65	0.63	0.64
Cond_Prior-Acute-12-month	0	5,328	3,175	2,203	0.81	0.52	0.29	0.37	0.59
Cond_Prior-Acute-12-month	1	10,885	6,587	6,341	0.34	0.63	0.79	0.70	0.62
Cond_Prior-Oper-12-month	0	2,667	1,623	1,173	0.76	0.58	0.41	0.48	0.63
Cond_Prior-Oper-12-month	1	13,546	8,128	7,371	0.50	0.62	0.68	0.65	0.64
Modelling Group: Pop_PARR-2-Settings									
Cond_Main		73,315	45,394	33,849	0.74	0.61	0.48	0.54	0.66
Cond_Prior-Acute-12-month	0	29,506	19,209	10,513	0.92	0.53	0.16	0.24	0.63
Cond_Prior-Acute-12-month	1	43,809	26,453	23,336	0.53	0.62	0.67	0.64	0.64
Cond_Prior-Oper-12-month	0	11,891	7,916	4,336	0.88	0.58	0.30	0.40	0.67
Cond_Prior-Oper-12-month	1	61,424	37,568	29,513	0.70	0.61	0.52	0.56	0.66

TABLE A.24: ERMER: Performance statistics of submodels (*Sample-3*)

Modelling Approach: BPM;
Sample: Sample-3

Sub-model	If	N	TP+FP	TP	Spec.	Prec.	Sens.	F1	AUC
Modelling Group: Pop_Any-Acute-No-Mental									
Cond_Age-65p	0	209,467	167,218	66,332	0.94	0.78	0.50	0.61	0.77
Cond_Age-65p	1	90,364	55,977	40,684	0.77	0.61	0.44	0.51	0.66
Cond_Main		299,831	219,997	107,016	0.89	0.69	0.45	0.55	0.74
Cond_Prior-Acute-12-month	0	213,411	160,319	65,262	0.93	0.69	0.34	0.45	0.73
Cond_Prior-Acute-12-month	1	86,420	59,667	41,754	0.70	0.68	0.68	0.68	0.74
Cond_Prior-Oper-12-month	0	128,665	95,443	48,190	0.90	0.74	0.48	0.58	0.78
Cond_Prior-Oper-12-month	1	171,166	123,011	58,826	0.87	0.64	0.42	0.51	0.75
Cond_Prior-Spells	0	159,140	123,224	43,720	0.96	0.72	0.29	0.42	0.72
Cond_Prior-Spells	1	140,691	96,691	63,296	0.77	0.68	0.59	0.63	0.72
Modelling Group: Pop_Any-Acute									
Cond_Age-65p	0	213,519	169,926	67,890	0.94	0.78	0.50	0.61	0.77
Cond_Age-65p	1	91,369	56,502	41,155	0.76	0.60	0.44	0.51	0.66
Cond_Main		304,888	223,285	109,045	0.89	0.69	0.45	0.55	0.74
Cond_Prior-Acute-12-month	0	216,448	162,390	66,260	0.93	0.69	0.33	0.45	0.72
Cond_Prior-Acute-12-month	1	88,440	60,856	42,785	0.70	0.68	0.67	0.68	0.74
Cond_Prior-Oper-12-month	0	132,530	98,025	49,630	0.90	0.74	0.48	0.58	0.78
Cond_Prior-Oper-12-month	1	172,358	123,718	59,415	0.87	0.64	0.43	0.51	0.75
Cond_Prior-Spells	0	161,375	124,841	44,357	0.96	0.72	0.29	0.41	0.72
Cond_Prior-Spells	1	143,513	98,386	64,688	0.77	0.67	0.58	0.63	0.72
Modelling Group: Pop_PARR-1-Settings									
Cond_Main		27,854	16,817	14,434	0.59	0.62	0.61	0.62	0.65
Cond_Prior-Acute-12-month	0	16,110	9,489	7,334	0.77	0.57	0.38	0.45	0.62
Cond_Prior-Acute-12-month	1	11,744	7,344	7,100	0.27	0.64	0.86	0.73	0.64
Cond_Prior-Oper-12-month	0	13,192	7,907	6,276	0.68	0.59	0.51	0.55	0.64
Cond_Prior-Oper-12-month	1	14,662	8,881	8,158	0.49	0.63	0.69	0.66	0.64
Modelling Group: Pop_PARR-2-Settings									
Cond_Main		91,369	56,502	41,155	0.76	0.60	0.44	0.51	0.66
Cond_Prior-Acute-12-month	0	59,751	37,420	23,432	0.88	0.55	0.24	0.33	0.64
Cond_Prior-Acute-12-month	1	31,618	19,181	17,723	0.44	0.63	0.74	0.68	0.64
Cond_Prior-Oper-12-month	0	38,342	23,869	16,159	0.81	0.58	0.37	0.45	0.66
Cond_Prior-Oper-12-month	1	53,027	32,685	24,996	0.72	0.61	0.50	0.55	0.66

TABLE A.25: ERMER: Performance statistics of submodels (*Sample-1-train-half-2-test-half*)

Modelling Approach: BPM;
Sample: Sample-1-train-half-2-test-half

Sub-model	If	N	TP+FP	TP	Spec.	Prec.	Sens.	F1	AUC
Modelling Group: Pop_Any-Acute-No-Mental									
Cond_Age-65p	0	168,014	132,201	56,697	0.91	0.76	0.54	0.63	0.76
Cond_Age-65p	1	72,729	45,085	33,579	0.68	0.60	0.55	0.57	0.66
Cond_Main		240,743	174,874	90,276	0.86	0.68	0.51	0.58	0.73
Cond_Prior-Acute-12-month	0	115,727	88,998	33,665	0.95	0.72	0.33	0.46	0.70
Cond_Prior-Acute-12-month	1	125,016	85,847	56,611	0.74	0.67	0.62	0.64	0.72
Cond_Prior-Oper-12-month	0	56,612	44,079	19,862	0.91	0.76	0.53	0.63	0.80
Cond_Prior-Oper-12-month	1	184,131	128,535	70,414	0.84	0.64	0.47	0.54	0.73
Cond_Prior-Spells	0	72,903	58,238	18,789	0.96	0.76	0.32	0.45	0.71
Cond_Prior-Spells	1	167,840	116,804	71,487	0.80	0.68	0.55	0.61	0.71
Modelling Group: Pop_Any-Acute									
Cond_Age-65p	0	170,397	133,374	57,668	0.91	0.75	0.54	0.63	0.76
Cond_Age-65p	1	73,315	45,332	33,849	0.68	0.59	0.54	0.57	0.66
Cond_Main		243,712	176,429	91,517	0.86	0.68	0.51	0.58	0.73
Cond_Prior-Acute-12-month	0	116,721	89,653	33,978	0.95	0.73	0.33	0.45	0.70
Cond_Prior-Acute-12-month	1	126,991	86,884	57,539	0.74	0.66	0.61	0.64	0.72
Cond_Prior-Oper-12-month	0	58,133	45,093	20,429	0.91	0.76	0.53	0.63	0.79
Cond_Prior-Oper-12-month	1	185,579	129,296	71,088	0.84	0.64	0.47	0.54	0.73
Cond_Prior-Spells	0	73,536	58,631	18,969	0.97	0.76	0.31	0.44	0.71
Cond_Prior-Spells	1	170,176	118,069	72,548	0.80	0.67	0.55	0.60	0.71
Modelling Group: Pop_PARR-1-Settings									
Cond_Main		16,213	9,707	8,544	0.46	0.60	0.72	0.66	0.64
Cond_Prior-Acute-12-month	0	5,328	3,042	2,203	0.69	0.48	0.41	0.44	0.57
Cond_Prior-Acute-12-month	1	10,885	6,586	6,341	0.34	0.63	0.79	0.70	0.62
Cond_Prior-Oper-12-month	0	2,667	1,586	1,173	0.61	0.54	0.57	0.55	0.62
Cond_Prior-Oper-12-month	1	13,546	8,127	7,371	0.50	0.62	0.69	0.65	0.63
Modelling Group: Pop_PARR-2-Settings									
Cond_Main		73,315	45,332	33,849	0.68	0.59	0.54	0.57	0.66
Cond_Prior-Acute-12-month	0	29,506	19,118	10,513	0.91	0.52	0.18	0.26	0.62
Cond_Prior-Acute-12-month	1	43,809	26,179	23,336	0.58	0.62	0.62	0.62	0.63
Cond_Prior-Oper-12-month	0	11,891	7,805	4,336	0.85	0.55	0.31	0.40	0.66
Cond_Prior-Oper-12-month	1	61,424	37,638	29,513	0.70	0.62	0.52	0.56	0.66

TABLE A.26: ERMER: Performance statistics of submodels (*Sample-1-train-half-3-test-half*)

Modelling Approach: BPM;
Sample: Sample-1-train-half-3-test-half

Sub-model	If	N	TP+FP	TP	Spec.	Prec.	Sens.	F1	AUC
Modelling Group: Pop_Any-Acute-No-Mental									
Cond_Age-65p	0	209,467	167,249	66,332	0.94	0.80	0.49	0.60	0.77
Cond_Age-65p	1	90,364	55,544	40,684	0.83	0.63	0.36	0.45	0.66
Cond_Main		299,831	219,574	107,016	0.91	0.71	0.42	0.53	0.74
Cond_Prior-Acute-12-month	0	213,411	160,689	65,262	0.94	0.70	0.33	0.45	0.72
Cond_Prior-Acute-12-month	1	86,420	59,513	41,754	0.76	0.71	0.61	0.65	0.74
Cond_Prior-Oper-12-month	0	128,665	95,407	48,190	0.90	0.74	0.48	0.58	0.78
Cond_Prior-Oper-12-month	1	171,166	123,593	58,826	0.90	0.67	0.37	0.48	0.75
Cond_Prior-Spells	0	159,140	123,442	43,720	0.96	0.73	0.29	0.41	0.72
Cond_Prior-Spells	1	140,691	96,291	63,296	0.83	0.71	0.51	0.59	0.72
Modelling Group: Pop_Any-Acute									
Cond_Age-65p	0	213,519	169,585	67,890	0.95	0.80	0.47	0.59	0.77
Cond_Age-65p	1	91,369	55,980	41,155	0.84	0.63	0.34	0.44	0.66
Cond_Main		304,888	222,458	109,045	0.91	0.72	0.40	0.52	0.74
Cond_Prior-Acute-12-month	0	216,448	162,504	66,260	0.94	0.71	0.32	0.44	0.72
Cond_Prior-Acute-12-month	1	88,440	60,679	42,785	0.77	0.71	0.59	0.65	0.74
Cond_Prior-Oper-12-month	0	132,530	97,891	49,630	0.91	0.75	0.46	0.57	0.78
Cond_Prior-Oper-12-month	1	172,358	124,056	59,415	0.91	0.67	0.36	0.47	0.75
Cond_Prior-Spells	0	161,375	124,691	44,357	0.96	0.74	0.27	0.39	0.72
Cond_Prior-Spells	1	143,513	97,817	64,688	0.84	0.71	0.49	0.58	0.72
Modelling Group: Pop_PARR-1-Settings									
Cond_Main		27,854	16,766	14,434	0.61	0.62	0.59	0.61	0.65
Cond_Prior-Acute-12-month	0	16,110	9,360	7,334	0.65	0.54	0.50	0.52	0.62
Cond_Prior-Acute-12-month	1	11,744	7,235	7,100	0.36	0.65	0.78	0.71	0.64
Cond_Prior-Oper-12-month	0	13,192	7,869	6,276	0.61	0.57	0.58	0.58	0.64
Cond_Prior-Oper-12-month	1	14,662	8,766	8,158	0.58	0.65	0.61	0.63	0.64
Modelling Group: Pop_PARR-2-Settings									
Cond_Main		91,369	55,980	41,155	0.84	0.63	0.34	0.44	0.66
Cond_Prior-Acute-12-month	0	59,751	37,379	23,432	0.90	0.56	0.20	0.29	0.64
Cond_Prior-Acute-12-month	1	31,618	18,698	17,723	0.60	0.65	0.59	0.62	0.63
Cond_Prior-Oper-12-month	0	38,342	23,835	16,159	0.83	0.59	0.34	0.43	0.66
Cond_Prior-Oper-12-month	1	53,027	32,124	24,996	0.81	0.64	0.38	0.48	0.66

A.6.3.2 Risk Bands Statistics

TABLE A.27: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Main* model (*Sample-1*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-1;
Submodel: Cond_Main

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	1	4,805	326	241	2	26	34	134	326	416	20.61	23.68	115 (2.39%)	1 (0.02%)	45 (0.94%)	24 (0.50%)	54 (1.12%)	54 (1.12%)	13 (0.27%)	2 (0.04%)
NA	2	12,889	1,700	2,362	14	174	194	673	1,700	2,295	4.04	32.35	430 (3.34%)	38 (0.29%)	441 (3.42%)	191 (1.48%)	517 (4.01%)	296 (2.30%)	297 (2.30%)	16 (0.12%)
NA	3	16,897	2,907	5,183	48	476	566	1,162	2,907	4,411	4.79	35.61	791 (4.68%)	113 (0.67%)	926 (5.48%)	411 (2.43%)	1,112 (6.58%)	726 (4.30%)	726 (4.30%)	65 (0.38%)
NA	4	18,777	3,668	5,873	64	704	900	1,415	3,668	5,839	5.29	41.85	1,039 (5.53%)	222 (1.18%)	1,539 (8.20%)	720 (3.83%)	1,871 (9.96%)	1,041 (5.54%)	1,222 (6.51%)	177 (0.94%)
NA	5	23,019	5,026	7,083	98	990	1,257	1,868	5,026	7,970	5.67	39.63	1,266 (5.50%)	392 (1.70%)	1,943 (8.44%)	923 (4.01%)	2,476 (10.76%)	1,226 (5.33%)	1,538 (6.68%)	324 (1.41%)
NA	6	23,590	5,443	10,209	134	1,421	1,918	1,887	5,443	8,763	6.19	43.61	1,390 (5.89%)	483 (2.05%)	2,082 (8.83%)	1,051 (4.46%)	2,838 (12.03%)	1,219 (5.17%)	1,764 (7.48%)	505 (2.14%)
NA	7	24,544	6,101	14,561	182	1,878	2,498	2,091	6,101	10,158	6.94	46.49	1,561 (6.36%)	675 (2.75%)	2,496 (10.17%)	1,166 (4.75%)	3,526 (14.37%)	1,400 (5.70%)	2,145 (8.74%)	583 (2.38%)
NA	8	22,767	6,827	15,854	175	2,502	3,274	2,403	6,827	11,583	7.80	50.43	1,553 (6.82%)	780 (3.43%)	2,837 (12.46%)	1,347 (5.92%)	4,192 (18.41%)	1,638 (7.19%)	2,726 (11.97%)	735 (3.23%)
NA	9	17,986	7,024	12,787	160	2,555	3,440	2,460	7,024	12,438	9.52	57.83	1,369 (7.61%)	957 (5.32%)	3,102 (17.25%)	1,478 (8.22%)	4,349 (24.18%)	1,626 (9.04%)	2,969 (16.51%)	945 (5.25%)
NA	10	13,794	6,712	10,636	175	2,545	3,579	2,350	6,712	12,107	10.01	60.99	1,177 (8.53%)	1,101 (7.98%)	3,013 (21.84%)	1,407 (10.20%)	4,059 (29.43%)	1,572 (11.40%)	2,838 (20.57%)	1,122 (8.13%)
NA	11	12,381	7,116	8,708	199	2,525	3,734	2,603	7,116	13,885	8.57	58.26	978 (7.90%)	1,103 (8.91%)	2,610 (21.08%)	1,308 (10.56%)	3,474 (28.06%)	1,297 (10.48%)	2,555 (20.64%)	1,278 (10.32%)
NA	12	11,950	8,201	8,076	142	2,548	3,902	3,399	8,201	16,426	7.18	52.27	803 (6.72%)	1,062 (8.89%)	2,013 (16.85%)	1,071 (8.96%)	3,015 (25.23%)	1,145 (9.58%)	2,134 (17.86%)	1,286 (10.76%)
NA	13	10,061	7,163	8,882	193	2,622	4,128	3,027	7,163	14,777	6.21	50.33	778 (7.73%)	958 (9.52%)	1,631 (16.21%)	923 (9.17%)	2,543 (25.28%)	997 (9.91%)	1,775 (17.64%)	1,238 (12.30%)
NA	14	8,244	6,351	7,994	146	2,714	4,375	2,896	6,351	13,582	5.10	47.49	719 (8.72%)	853 (10.35%)	1,185 (14.37%)	666 (8.08%)	2,058 (24.96%)	771 (9.35%)	1,268 (15.38%)	933 (11.32%)
NA	15	4,262	3,308	4,162	84	2,131	3,772	1,298	3,308	7,863	6.70	52.62	574 (13.47%)	646 (15.16%)	863 (20.25%)	489 (11.47%)	1,384 (32.47%)	487 (11.43%)	929 (21.80%)	738 (17.32%)
NA	16	2,251	1,712	2,207	65	1,480	3,175	590	1,712	4,468	8.52	58.10	382 (16.97%)	454 (20.17%)	582 (25.86%)	363 (16.13%)	816 (36.25%)	341 (15.15%)	669 (29.72%)	497 (22.08%)
NA	17	1,324	1,036	1,315	38	991	2,425	362	1,036	2,904	9.71	60.89	248 (18.73%)	323 (24.40%)	397 (29.98%)	251 (18.96%)	532 (40.18%)	215 (16.24%)	474 (35.80%)	323 (24.40%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	18	903	714	900	31	748	2,391	228	714	2,281	8.54	61.14	214 (23.70%)	266 (29.46%)	297 (32.89%)	204 (22.59%)	403 (44.63%)	156 (17.28%)	344 (38.10%)	247 (27.35%)
NA	19	611	486	609	20	518	1,938	140	486	1,684	10.11	61.03	172 (28.15%)	226 (36.99%)	210 (34.37%)	117 (19.15%)	275 (45.01%)	106 (17.35%)	248 (40.59%)	186 (30.44%)
NA	20	700	613	698	72	655	3,934	222	613	3,140	10.97	57.33	244 (34.86%)	269 (38.43%)	254 (36.29%)	181 (25.86%)	320 (45.71%)	113 (16.14%)	306 (43.71%)	214 (30.57%)
NA	Total	231,755	82,434	128,340	2,042	30,203	51,434	31,208	82,434	156,990	7.09	46.77	15,803 (6.82%)	10,922 (4.71%)	28,466 (12.28%)	14,291 (6.17%)	39,814 (17.18%)	16,426 (7.09%)	26,940 (11.62%)	11,414 (4.93%)

[a] If: The condition of the sub-modelling group. [b] #.: The risk band number: 1 = [0, 0.05); 2 = [0.05, 0.10); 3 = [0.10, 0.15); 4 = [0.15, 0.20); 5 = [0.20, 0.25); 6 = [0.25, 0.30); 7 = [0.30, 0.35); 8 = [0.35, 0.40); 9 = [0.40, 0.45); 10 = [0.45, 0.50); 11 = [0.50, 0.55); 12 = [0.55, 0.60); 13 = [0.60, 0.65); 14 = [0.65, 0.70); 15 = [0.70, 0.75); 16 = [0.75, 0.80); 17 = [0.80, 0.85); 18 = [0.85, 0.90); 19 = [0.90, 0.95); 20 = [0.95, 1]. [c] TP+FP: The number of true positives (TPs) and false positives (FPs). [d] TP: The number of true positives (TPs). [e] Prior Spells: Total number of patients with any prior spell. [f] Prior 30-d Emer.: Total number of patients with prior 30-day emergency admission. [g] Prior 12-m Emer.: Total number of patients with prior 12-month emergency admission. [h] Prior 12-m Emer. Spells: Total number of spell with prior 12-month emergency admission. [i] Post 30-d Emer.: Total number of patients with post 30-day emergency admission. [j] Post 12-m Emer.: Total number of patients with post 12-month emergency admission. [k] Post 12-m Emer. Spells: Total number of spell with post 12-month emergency admission. [l] Avg. Stay: The average stay of all patients during the trigger admission. [m] Avg. Age: The average ageRecoded of all patients during the trigger admission. [n] Asth.: Total number of patients with an asthma diagnosis during the trigger admission. [o] COPD: Total number of patients with a COPD diagnosis during the trigger admission. [p] Depr.: Total number of patients with a depression diagnosis during the trigger admission. [q] Diab.: Total number of patients with a diabetes diagnosis during the trigger admission. [r] Hype.: Total number of patients with a hypertension diagnosis during the trigger admission. [s] Canc.: Total number of patients with a cancer diagnosis during the trigger admission. [t] CHD: Total number of patients with a CHD diagnosis during the trigger admission. [u] CHF: Total number of patients with a CHF diagnosis during the trigger admission.

TABLE A.28: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Age-65p* model (*Sample-1*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-1;
Submodel: Cond_Age-65p

If[a]	#[b]	TP+FP[o]TP[d]		Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	10,511	1,039	1,396	9	106	126	440	1,039	1,395	9.15	29.58	277 (2.64%)	24 (0.23%)	256 (2.44%)	99 (0.94%)	313 (2.98%)	178 (1.69%)	164 (1.56%)	3 (0.03%)
0	2	14,537	2,636	5,795	47	525	614	1,008	2,636	4,078	5.11	33.97	652 (4.49%)	64 (0.44%)	572 (3.93%)	282 (1.94%)	737 (5.07%)	594 (4.09%)	571 (3.93%)	25 (0.17%)
0	3	10,295	1,843	3,116	47	590	744	693	1,843	2,944	5.25	30.05	702 (6.82%)	90 (0.87%)	497 (4.83%)	348 (3.38%)	653 (6.34%)	482 (4.68%)	455 (4.42%)	59 (0.57%)
0	4	13,995	1,885	3,500	57	518	707	724	1,885	2,909	3.99	32.25	714 (5.10%)	115 (0.82%)	500 (3.57%)	386 (2.76%)	825 (5.89%)	308 (2.20%)	418 (2.99%)	64 (0.46%)
0	5	18,422	3,097	4,977	81	779	1,055	1,155	3,097	4,753	3.67	28.33	986 (5.35%)	133 (0.72%)	556 (3.02%)	361 (1.96%)	978 (5.31%)	348 (1.89%)	482 (2.62%)	80 (0.43%)
0	6	21,717	3,918	9,723	111	1,066	1,455	1,527	3,918	6,328	3.80	32.52	1,107 (5.10%)	158 (0.73%)	762 (3.51%)	496 (2.28%)	1,418 (6.53%)	438 (2.02%)	694 (3.20%)	71 (0.33%)
0	7	18,156	3,836	12,620	126	1,619	2,148	1,340	3,836	6,435	5.10	35.20	1,326 (7.30%)	183 (1.01%)	899 (4.95%)	526 (2.90%)	1,721 (9.48%)	572 (3.15%)	859 (4.73%)	102 (0.56%)
0	8	9,794	2,667	8,178	87	1,754	2,323	881	2,667	4,548	6.18	38.02	1,149 (11.73%)	211 (2.15%)	732 (7.47%)	574 (5.86%)	1,252 (12.78%)	681 (6.95%)	734 (7.49%)	99 (1.01%)
0	9	5,881	2,192	4,852	64	1,388	1,953	788	2,192	4,149	6.84	40.41	763 (12.97%)	218 (3.71%)	610 (10.37%)	512 (8.71%)	925 (15.73%)	572 (9.73%)	671 (11.41%)	122 (2.07%)
0	10	4,593	2,140	3,189	58	1,090	1,678	803	2,140	4,107	5.68	39.46	532 (11.58%)	193 (4.20%)	495 (10.78%)	404 (8.80%)	728 (15.85%)	399 (8.69%)	536 (11.67%)	121 (2.63%)
0	11	4,085	2,654	2,493	68	952	1,563	1,236	2,654	5,272	4.86	36.96	362 (8.86%)	195 (4.77%)	376 (9.20%)	289 (7.07%)	664 (16.25%)	276 (6.76%)	393 (9.62%)	126 (3.08%)
0	12	4,166	2,825	2,315	53	865	1,460	1,202	2,825	5,621	3.79	35.37	299 (7.18%)	142 (3.41%)	320 (7.68%)	266 (6.39%)	565 (13.56%)	269 (6.46%)	326 (7.83%)	150 (3.60%)
0	13	6,259	4,831	2,829	89	859	1,581	2,316	4,831	10,286	2.08	31.52	267 (4.27%)	116 (1.85%)	224 (3.58%)	192 (3.07%)	584 (9.33%)	244 (3.90%)	229 (3.66%)	115 (1.84%)
0	14	7,723	6,383	4,989	114	1,081	1,808	3,215	6,383	13,541	1.61	30.70	280 (3.63%)	120 (1.55%)	204 (2.64%)	148 (1.92%)	840 (10.88%)	243 (3.15%)	205 (2.65%)	97 (1.26%)
0	15	5,863	5,068	5,681	126	1,440	2,434	2,583	5,068	11,119	1.84	30.85	392 (6.69%)	99 (1.69%)	172 (2.93%)	133 (2.27%)	786 (13.41%)	308 (5.25%)	141 (2.40%)	67 (1.14%)
0	16	2,727	2,347	2,696	61	1,167	2,131	1,108	2,347	5,985	2.70	32.05	338 (12.39%)	83 (3.04%)	142 (5.21%)	106 (3.89%)	455 (16.69%)	288 (10.56%)	121 (4.44%)	62 (2.27%)
0	17	1,161	965	1,154	42	728	1,629	431	965	2,834	3.73	34.70	222 (19.12%)	60 (5.17%)	105 (9.04%)	65 (5.60%)	227 (19.55%)	143 (12.32%)	101 (8.70%)	48 (4.13%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	662	542	658	24	500	1,405	214	542	1,782	5.84	36.49	142 (21.45%)	60 (9.06%)	94 (14.20%)	82 (12.39%)	143 (21.60%)	86 (12.99%)	89 (13.44%)	33 (4.98%)
0	19	458	370	455	13	367	1,312	128	370	1,336	6.13	39.28	115 (25.11%)	49 (10.70%)	77 (16.81%)	76 (16.59%)	130 (28.38%)	66 (14.41%)	67 (14.63%)	38 (8.30%)
0	20	603	525	601	51	539	3,081	202	525	2,640	8.27	42.17	190 (31.51%)	114 (18.91%)	127 (21.06%)	130 (21.56%)	178 (29.52%)	96 (15.92%)	141 (23.38%)	72 (11.94%)
0	Total	161,608	51,763	81,217	1,328	17,933	31,207	21,994	51,763	102,062	4.61	33.00	10,815 (6.69%)	2,427 (1.50%)	7,720 (4.78%)	5,475 (3.39%)	14,122 (8.74%)	6,591 (4.08%)	7,397 (4.58%)	1,554 (0.96%)
1	1	16	0	10	0	2	2	0	0	0	987.19	79.64	0 (0.00%)	0 (0.00%)	1 (6.25%)	1 (6.25%)	1 (6.25%)	1 (6.25%)	0 (0.00%)	2 (12.50%)
1	2	261	9	8	0	1	1	2	9	12	14.31	90.78	3 (1.15%)	0 (0.00%)	5 (1.92%)	1 (0.38%)	8 (3.07%)	7 (2.68%)	4 (1.53%)	0 (0.00%)
1	3	709	45	26	0	4	8	16	45	64	10.12	79.25	7 (0.99%)	1 (0.14%)	42 (5.92%)	15 (2.12%)	50 (7.05%)	19 (2.68%)	9 (1.27%)	2 (0.28%)
1	4	1,872	296	139	0	11	11	130	296	412	2.94	75.57	38 (2.03%)	30 (1.60%)	322 (17.20%)	83 (4.43%)	342 (18.27%)	64 (3.42%)	154 (8.23%)	11 (0.59%)
1	5	2,821	606	613	1	58	71	264	606	866	8.42	76.69	86 (3.05%)	76 (2.69%)	545 (19.32%)	181 (6.42%)	611 (21.66%)	191 (6.77%)	342 (12.12%)	54 (1.91%)
1	6	5,756	1,585	1,511	16	143	163	611	1,585	2,425	9.48	76.96	169 (2.94%)	193 (3.35%)	1,236 (21.47%)	363 (6.31%)	1,445 (25.10%)	392 (6.81%)	819 (14.23%)	155 (2.69%)
1	7	7,284	2,434	3,092	31	329	388	841	2,434	3,754	11.33	77.77	318 (4.37%)	371 (5.09%)	1,656 (22.73%)	583 (8.00%)	2,006 (27.54%)	703 (9.65%)	1,364 (18.73%)	400 (5.49%)
1	8	9,247	3,462	5,944	47	633	777	1,180	3,462	5,594	12.79	77.79	424 (4.59%)	485 (5.24%)	2,470 (26.71%)	828 (8.95%)	2,940 (31.79%)	1,247 (13.49%)	1,967 (21.27%)	520 (5.62%)
1	9	12,494	5,422	8,190	82	1,038	1,347	1,655	5,422	9,044	15.10	78.93	641 (5.13%)	908 (7.27%)	3,588 (28.72%)	1,355 (10.85%)	4,400 (35.22%)	1,857 (14.86%)	3,107 (24.87%)	1,017 (8.14%)
1	10	10,065	4,871	8,304	118	1,475	1,876	1,394	4,871	8,359	14.01	79.98	746 (7.41%)	1,149 (11.42%)	3,319 (32.98%)	1,412 (14.03%)	4,167 (41.40%)	1,711 (17.00%)	3,103 (30.83%)	1,495 (14.85%)
1	11	7,193	3,864	6,890	97	1,768	2,270	1,058	3,864	6,958	14.08	80.69	604 (8.40%)	1,147 (15.95%)	2,462 (34.23%)	1,197 (16.64%)	3,220 (44.77%)	1,264 (17.57%)	2,612 (36.31%)	1,547 (21.51%)
1	12	4,791	2,746	4,763	75	1,701	2,287	705	2,746	5,028	13.84	81.26	517 (10.79%)	1,086 (22.67%)	1,794 (37.45%)	917 (19.14%)	2,333 (48.70%)	820 (17.12%)	1,940 (40.49%)	1,387 (28.95%)
1	13	3,036	1,936	3,033	56	1,576	2,316	476	1,936	3,754	12.79	80.96	450 (14.82%)	941 (30.99%)	1,206 (39.72%)	671 (22.10%)	1,562 (51.45%)	571 (18.81%)	1,446 (47.63%)	1,137 (37.45%)
1	14	1,724	1,189	1,722	54	1,161	1,968	299	1,189	2,498	12.14	80.73	289 (16.76%)	672 (38.98%)	727 (42.17%)	417 (24.19%)	926 (53.71%)	345 (20.01%)	904 (52.44%)	739 (42.87%)
1	15	1,096	773	1,096	23	820	1,637	199	773	1,751	11.42	80.73	220 (20.07%)	482 (43.98%)	485 (44.25%)	296 (27.01%)	603 (55.02%)	231 (21.08%)	636 (58.03%)	494 (45.07%)
1	16	662	516	662	37	548	1,315	126	516	1,286	10.89	79.93	135 (20.39%)	294 (44.41%)	321 (48.49%)	193 (29.15%)	394 (59.52%)	149 (22.51%)	393 (59.37%)	306 (46.22%)
1	17	434	344	434	21	377	1,058	87	344	971	9.64	79.79	105 (24.19%)	235 (54.15%)	208 (47.93%)	118 (27.19%)	254 (58.53%)	96 (22.12%)	270 (62.21%)	217 (50.00%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	305	240	305	15	272	912	63	240	728	8.92	79.06	91 (29.84%)	159 (52.13%)	155 (50.82%)	78 (25.57%)	186 (60.98%)	69 (22.62%)	205 (67.21%)	168 (55.08%)
1	19	194	161	194	9	174	733	44	161	595	11.27	78.45	74 (38.14%)	136 (70.10%)	101 (52.06%)	58 (29.90%)	126 (64.95%)	46 (23.71%)	126 (64.95%)	106 (54.64%)
1	20	187	172	187	32	179	1,087	64	172	829	7.73	77.17	71 (37.97%)	130 (69.52%)	103 (55.08%)	49 (26.20%)	118 (63.10%)	52 (27.81%)	142 (75.94%)	103 (55.08%)
1	Total	70,147	30,671	47,123	714	12,270	20,227	9,214	30,671	54,928	12.81	79.01	4,988 (7.11%)	8,495 (12.11%)	20,746 (29.58%)	8,816 (12.57%)	25,692 (36.63%)	9,835 (14.02%)	19,543 (27.86%)	9,860 (14.06%)

TABLE A.29: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Acute-12-month* model (*Sample-1*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-1;
Submodel: Cond_Prior-Acute-12-month

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	5,095	366	207	3	12	12	148	366	470	7.07	25.07	135 (2.65%)	5 (0.10%)	55 (1.08%)	25 (0.49%)	69 (1.35%)	52 (1.02%)	28 (0.55%)	1 (0.02%)
0	2	11,488	1,477	1,293	8	121	141	606	1,477	1,955	3.89	32.28	349 (3.04%)	38 (0.33%)	416 (3.62%)	181 (1.58%)	478 (4.16%)	243 (2.12%)	225 (1.96%)	16 (0.14%)
0	3	15,752	2,434	2,529	19	296	335	1,045	2,434	3,643	4.09	35.00	697 (4.42%)	90 (0.57%)	789 (5.01%)	392 (2.49%)	917 (5.82%)	492 (3.12%)	546 (3.47%)	47 (0.30%)
0	4	14,350	2,648	2,886	38	416	492	1,054	2,648	4,097	5.55	44.76	686 (4.78%)	176 (1.23%)	1,245 (8.68%)	545 (3.80%)	1,472 (10.26%)	750 (5.23%)	943 (6.57%)	113 (0.79%)
0	5	19,599	4,032	3,472	47	550	709	1,595	4,032	6,307	5.16	37.45	934 (4.77%)	285 (1.45%)	1,358 (6.93%)	703 (3.59%)	1,671 (8.53%)	834 (4.26%)	1,075 (5.48%)	172 (0.88%)
0	6	16,776	3,896	4,508	57	770	982	1,429	3,896	6,196	6.50	46.25	918 (5.47%)	318 (1.90%)	1,490 (8.88%)	695 (4.14%)	1,980 (11.80%)	797 (4.75%)	1,168 (6.96%)	284 (1.69%)
0	7	14,886	3,824	5,197	86	1,037	1,303	1,426	3,824	6,136	7.30	51.64	729 (4.90%)	443 (2.98%)	1,746 (11.73%)	769 (5.17%)	2,249 (15.11%)	861 (5.78%)	1,462 (9.82%)	332 (2.23%)
0	8	13,066	3,990	5,185	105	1,183	1,520	1,556	3,990	6,689	8.10	55.05	616 (4.71%)	485 (3.71%)	1,700 (13.01%)	794 (6.08%)	2,365 (18.10%)	925 (7.08%)	1,489 (11.40%)	390 (2.98%)
0	9	9,120	3,404	4,465	81	1,280	1,662	1,230	3,404	5,793	10.38	60.94	479 (5.25%)	456 (5.00%)	1,503 (16.48%)	786 (8.62%)	2,095 (22.97%)	950 (10.42%)	1,287 (14.11%)	442 (4.85%)
0	10	6,437	3,341	2,975	91	1,134	1,510	1,427	3,341	5,975	9.68	57.65	290 (4.51%)	368 (5.72%)	1,084 (16.84%)	505 (7.85%)	1,589 (24.69%)	814 (12.65%)	887 (13.78%)	416 (6.46%)
0	11	5,530	3,274	2,260	85	1,057	1,535	1,315	3,274	6,416	7.34	50.57	179 (3.24%)	272 (4.92%)	687 (12.42%)	404 (7.31%)	1,062 (19.20%)	614 (11.10%)	641 (11.59%)	437 (7.90%)
0	12	4,487	3,093	1,992	96	945	1,439	1,411	3,093	6,354	5.62	44.43	163 (3.63%)	194 (4.32%)	410 (9.14%)	253 (5.64%)	696 (15.51%)	480 (10.70%)	411 (9.16%)	325 (7.24%)
0	13	4,682	3,769	1,742	83	920	1,464	1,959	3,769	7,760	3.50	38.17	114 (2.43%)	135 (2.88%)	266 (5.68%)	140 (2.99%)	595 (12.71%)	307 (6.56%)	266 (5.68%)	227 (4.85%)
0	14	2,170	1,608	1,502	75	904	1,441	862	1,608	3,439	4.80	43.71	87 (4.01%)	92 (4.24%)	162 (7.47%)	100 (4.61%)	349 (16.08%)	246 (11.34%)	165 (7.60%)	136 (6.27%)
0	15	907	619	784	32	610	1,092	295	619	1,348	7.91	52.84	34 (3.75%)	70 (7.72%)	125 (13.78%)	74 (8.16%)	199 (21.94%)	190 (20.95%)	132 (14.55%)	102 (11.25%)
0	16	393	250	352	22	301	740	97	250	544	15.94	63.75	20 (5.09%)	41 (10.43%)	85 (21.63%)	44 (11.20%)	118 (30.03%)	105 (26.72%)	83 (21.12%)	66 (16.79%)
0	17	196	126	183	13	160	458	49	126	262	17.94	67.07	19 (9.69%)	28 (14.29%)	54 (27.55%)	29 (14.80%)	68 (34.69%)	70 (35.71%)	71 (36.22%)	54 (27.55%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	102	66	94	5	80	266	19	66	178	18.86	71.55	9 (8.82%)	19 (18.63%)	31 (30.39%)	19 (18.63%)	41 (40.20%)	44 (43.14%)	40 (39.22%)	29 (28.43%)
0	19	51	32	48	2	43	165	14	32	74	23.37	72.98	5 (9.80%)	14 (27.45%)	11 (21.57%)	13 (25.49%)	19 (37.25%)	24 (47.06%)	16 (31.37%)	13 (25.49%)
0	20	40	28	38	6	30	115	8	28	62	28.80	61.28	2 (5.00%)	4 (10.00%)	8 (20.00%)	6 (15.00%)	15 (37.50%)	20 (50.00%)	11 (27.50%)	5 (12.50%)
0	Total	145,127	42,277	41,712	954	11,849	17,381	17,545	42,277	73,698	6.34	44.53	6,465 (4.45%)	3,533 (2.43%)	13,225 (9.11%)	6,477 (4.46%)	18,047 (12.44%)	8,818 (6.08%)	10,946 (7.54%)	3,607 (2.49%)
1	1	133	13	133	0	11	17	5	13	16	458.22	30.57	5 (3.76%)	0 (0.00%)	2 (1.50%)	4 (3.01%)	2 (1.50%)	2 (1.50%)	0 (0.00%)	3 (2.26%)
1	2	1,270	215	1,270	1	22	28	55	215	320	10.27	23.76	100 (7.87%)	6 (0.47%)	17 (1.34%)	14 (1.10%)	37 (2.91%)	29 (2.28%)	23 (1.81%)	2 (0.16%)
1	3	2,483	563	2,483	7	126	135	194	563	903	6.58	35.61	220 (8.86%)	19 (0.77%)	162 (6.52%)	65 (2.62%)	221 (8.90%)	120 (4.83%)	149 (6.00%)	15 (0.60%)
1	4	3,108	881	3,108	9	218	242	289	881	1,491	6.52	46.02	268 (8.62%)	77 (2.48%)	442 (14.22%)	184 (5.92%)	566 (18.21%)	259 (8.33%)	385 (12.39%)	75 (2.41%)
1	5	4,668	1,228	4,668	24	351	425	369	1,228	2,098	7.10	44.85	382 (8.18%)	131 (2.81%)	575 (12.32%)	266 (5.70%)	807 (17.29%)	385 (8.25%)	564 (12.08%)	157 (3.36%)
1	6	8,538	1,888	8,538	24	467	548	565	1,888	3,116	5.55	37.56	606 (7.10%)	194 (2.27%)	692 (8.10%)	358 (4.19%)	1,151 (13.48%)	476 (5.58%)	613 (7.18%)	254 (2.97%)
1	7	8,823	2,176	8,823	44	865	1,019	698	2,176	3,793	6.35	38.90	792 (8.98%)	233 (2.64%)	734 (8.32%)	432 (4.90%)	1,336 (15.14%)	532 (6.03%)	656 (7.44%)	278 (3.15%)
1	8	7,235	2,358	7,235	63	1,126	1,357	666	2,358	4,113	8.44	45.52	788 (10.89%)	254 (3.51%)	933 (12.90%)	458 (6.33%)	1,438 (19.88%)	542 (7.49%)	912 (12.61%)	330 (4.56%)
1	9	6,530	2,478	6,530	59	1,127	1,503	704	2,478	4,490	9.01	53.60	754 (11.55%)	365 (5.59%)	1,263 (19.34%)	540 (8.27%)	1,672 (25.60%)	558 (8.55%)	1,395 (21.36%)	384 (5.88%)
1	10	6,975	3,144	6,975	60	1,210	1,716	905	3,144	5,701	10.01	61.17	786 (11.27%)	571 (8.19%)	1,682 (24.11%)	762 (10.92%)	2,177 (31.21%)	691 (9.91%)	1,795 (25.73%)	600 (8.60%)
1	11	7,085	3,706	7,085	68	1,309	1,841	1,120	3,706	6,961	10.05	64.97	779 (11.00%)	824 (11.63%)	1,890 (26.68%)	882 (12.45%)	2,456 (34.66%)	777 (10.97%)	2,012 (28.40%)	820 (11.57%)
1	12	7,504	4,693	7,504	83	1,567	2,264	1,707	4,693	9,140	8.76	59.84	712 (9.49%)	904 (12.05%)	1,756 (23.40%)	885 (11.79%)	2,397 (31.94%)	800 (10.66%)	1,866 (24.87%)	959 (12.78%)
1	13	8,079	5,837	8,079	97	1,810	2,662	2,371	5,837	11,759	5.90	51.67	749 (9.27%)	877 (10.86%)	1,475 (18.26%)	754 (9.33%)	2,281 (28.23%)	700 (8.66%)	1,545 (19.12%)	1,013 (12.54%)
1	14	5,502	4,133	5,502	102	2,106	3,212	1,565	4,133	8,999	6.40	54.08	636 (11.56%)	785 (14.27%)	1,144 (20.79%)	725 (13.18%)	1,767 (32.12%)	574 (10.43%)	1,224 (22.25%)	871 (15.83%)
1	15	3,251	2,463	3,251	84	1,778	3,021	927	2,463	5,912	7.42	57.57	502 (15.44%)	640 (19.69%)	808 (24.85%)	441 (13.57%)	1,182 (36.36%)	392 (12.06%)	929 (28.58%)	666 (20.49%)
1	16	1,974	1,530	1,974	74	1,356	2,921	509	1,530	4,150	7.57	58.98	362 (18.34%)	463 (23.45%)	559 (28.32%)	318 (16.11%)	805 (40.78%)	281 (14.24%)	646 (32.73%)	443 (22.44%)
1	17	1,321	1,045	1,321	74	1,012	2,617	364	1,045	2,983	7.26	57.89	264 (19.98%)	332 (25.13%)	383 (28.99%)	237 (17.94%)	529 (40.05%)	176 (13.32%)	429 (32.48%)	324 (24.53%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	885	721	885	68	734	2,332	263	721	2,316	7.87	59.41	217 (24.52%)	259 (29.27%)	287 (32.43%)	208 (23.50%)	381 (43.05%)	125 (14.12%)	334 (37.74%)	242 (27.34%)
1	19	619	510	619	53	543	2,120	153	510	1,920	7.44	58.39	182 (29.40%)	215 (34.73%)	214 (34.57%)	129 (20.84%)	284 (45.88%)	94 (15.19%)	255 (41.20%)	191 (30.86%)
1	20	645	575	645	94	616	4,073	234	575	3,111	9.95	55.05	234 (36.28%)	240 (37.21%)	223 (34.57%)	152 (23.57%)	278 (43.10%)	95 (14.73%)	262 (40.62%)	180 (27.91%)
1	Total	86,628	40,157	86,628	1,088	18,354	34,053	13,663	40,157	83,292	8.35	50.49	9,338 (10.78%)	7,389 (8.53%)	15,241 (17.59%)	7,814 (9.02%)	21,767 (25.13%)	7,608 (8.78%)	15,994 (18.46%)	7,807 (9.01%)

TABLE A.30: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Oper-12-month* model (*Sample-1*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-1;
Submodel: Cond_Prior-Oper-12-month

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	302	34	44	1	9	15	12	34	43	211.71	29.43	12 (3.97%)	1 (0.33%)	5 (1.66%)	8 (2.65%)	10 (3.31%)	0 (0.00%)	2 (0.66%)	1 (0.33%)
0	2	3,937	381	44	0	7	9	189	381	498	5.44	17.60	131 (3.33%)	11 (0.28%)	27 (0.69%)	27 (0.69%)	34 (0.86%)	13 (0.33%)	11 (0.28%)	1 (0.03%)
0	3	9,092	1,022	290	2	23	23	455	1,022	1,353	2.74	31.99	498 (5.48%)	25 (0.27%)	337 (3.71%)	220 (2.42%)	385 (4.23%)	41 (0.45%)	164 (1.80%)	15 (0.16%)
0	4	12,392	2,070	697	6	110	123	863	2,070	2,985	3.45	29.79	633 (5.11%)	106 (0.86%)	566 (4.57%)	329 (2.65%)	652 (5.26%)	79 (0.64%)	427 (3.45%)	45 (0.36%)
0	5	11,267	2,432	1,131	7	133	156	982	2,432	3,835	5.60	43.02	726 (6.44%)	169 (1.50%)	859 (7.62%)	498 (4.42%)	1,084 (9.62%)	128 (1.14%)	650 (5.77%)	134 (1.19%)
0	6	9,004	2,426	2,135	24	333	383	898	2,426	3,947	7.66	52.24	754 (8.37%)	292 (3.24%)	1,051 (11.67%)	564 (6.26%)	1,316 (14.62%)	166 (1.84%)	937 (10.41%)	202 (2.24%)
0	7	6,963	2,418	2,264	21	449	528	853	2,418	4,062	11.17	57.28	584 (8.39%)	313 (4.50%)	1,032 (14.82%)	574 (8.24%)	1,355 (19.46%)	166 (2.38%)	1,036 (14.88%)	298 (4.28%)
0	8	5,711	2,231	2,370	28	555	684	694	2,231	3,922	14.33	64.89	426 (7.46%)	381 (6.67%)	1,084 (18.98%)	563 (9.86%)	1,432 (25.07%)	169 (2.96%)	1,047 (18.33%)	347 (6.08%)
0	9	4,094	1,885	2,309	30	573	726	569	1,885	3,378	14.59	69.05	340 (8.30%)	361 (8.82%)	912 (22.28%)	513 (12.53%)	1,234 (30.14%)	147 (3.59%)	866 (21.15%)	435 (10.63%)
0	10	3,196	1,641	2,182	43	631	784	535	1,641	2,902	12.31	65.57	243 (7.60%)	348 (10.89%)	687 (21.50%)	375 (11.73%)	971 (30.38%)	102 (3.19%)	672 (21.03%)	401 (12.55%)
0	11	2,801	1,576	1,734	40	584	777	554	1,576	2,877	10.02	58.97	179 (6.39%)	316 (11.28%)	497 (17.74%)	292 (10.42%)	731 (26.10%)	79 (2.82%)	506 (18.06%)	365 (13.03%)
0	12	1,603	894	1,257	18	543	740	268	894	1,679	10.21	66.09	153 (9.54%)	298 (18.59%)	332 (20.71%)	185 (11.54%)	477 (29.76%)	49 (3.06%)	382 (23.83%)	299 (18.65%)
0	13	1,914	1,245	884	13	464	664	446	1,245	2,570	5.85	46.50	100 (5.22%)	259 (13.53%)	253 (13.22%)	146 (7.63%)	357 (18.65%)	42 (2.19%)	295 (15.41%)	220 (11.49%)
0	14	2,682	1,894	587	16	402	650	803	1,894	4,110	2.76	37.41	96 (3.58%)	174 (6.49%)	144 (5.37%)	85 (3.17%)	278 (10.37%)	50 (1.86%)	184 (6.86%)	142 (5.29%)
0	15	978	608	476	8	247	456	249	608	1,337	5.30	44.43	79 (8.08%)	111 (11.35%)	83 (8.49%)	58 (5.93%)	149 (15.24%)	28 (2.86%)	90 (9.20%)	80 (8.18%)
0	16	758	519	555	20	253	474	211	519	1,197	4.36	40.43	67 (8.84%)	81 (10.69%)	61 (8.05%)	35 (4.62%)	100 (13.19%)	17 (2.24%)	58 (7.65%)	50 (6.60%)
0	17	1,884	1,610	427	29	300	563	887	1,610	3,185	1.43	31.43	63 (3.34%)	57 (3.03%)	44 (2.34%)	37 (1.96%)	175 (9.29%)	10 (0.53%)	58 (3.08%)	44 (2.34%)

Continued on next page

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	1,700	1,488	310	21	206	467	796	1,488	3,145	1.27	28.88	32 (1.88%)	47 (2.76%)	28 (1.65%)	14 (0.82%)	107 (6.29%)	9 (0.53%)	39 (2.29%)	28 (1.65%)
0	19	3,925	3,519	698	54	329	630	1,973	3,519	7,397	0.94	29.81	71 (1.81%)	41 (1.04%)	31 (0.79%)	23 (0.59%)	357 (9.10%)	11 (0.28%)	36 (0.92%)	24 (0.61%)
0	20	819	739	778	43	527	1,122	442	739	2,031	2.62	33.60	67 (8.18%)	48 (5.86%)	28 (3.42%)	28 (3.42%)	87 (10.62%)	6 (0.73%)	39 (4.76%)	28 (3.42%)
0	Total	85,022	30,632	21,172	424	6,678	9,974	12,679	30,632	56,453	7.36	43.97	5,254 (6.18%)	3,439 (4.04%)	8,061 (9.48%)	4,574 (5.38%)	11,291 (13.28%)	1,312 (1.54%)	7,499 (8.82%)	3,159 (3.72%)
1	1	1,754	96	33	0	4	5	39	96	111	15.75	17.59	25 (1.43%)	0 (0.00%)	9 (0.51%)	4 (0.23%)	12 (0.68%)	14 (0.80%)	1 (0.06%)	1 (0.06%)
1	2	10,319	977	1,003	6	94	101	387	977	1,276	2.20	27.91	278 (2.69%)	16 (0.16%)	185 (1.79%)	87 (0.84%)	219 (2.12%)	121 (1.17%)	81 (0.78%)	4 (0.04%)
1	3	13,151	1,915	3,690	24	361	405	757	1,915	2,710	3.29	36.17	464 (3.53%)	37 (0.28%)	502 (3.82%)	193 (1.47%)	682 (5.19%)	360 (2.74%)	321 (2.44%)	18 (0.14%)
1	4	14,654	2,533	5,993	52	657	795	996	2,533	3,959	4.44	40.81	479 (3.27%)	104 (0.71%)	852 (5.81%)	324 (2.21%)	1,216 (8.30%)	687 (4.69%)	679 (4.63%)	53 (0.36%)
1	5	13,609	2,882	8,743	110	1,149	1,415	1,035	2,882	4,720	6.18	44.51	617 (4.53%)	189 (1.39%)	1,203 (8.84%)	471 (3.46%)	1,727 (12.69%)	1,068 (7.85%)	1,021 (7.50%)	168 (1.23%)
1	6	13,953	3,356	11,037	133	1,357	1,780	1,180	3,356	5,562	7.01	46.84	720 (5.16%)	267 (1.91%)	1,400 (10.03%)	632 (4.53%)	2,130 (15.27%)	1,324 (9.49%)	1,182 (8.47%)	272 (1.95%)
1	7	12,494	3,484	10,977	101	1,504	1,959	1,175	3,484	5,873	7.69	48.18	876 (7.01%)	362 (2.90%)	1,579 (12.64%)	686 (5.49%)	2,371 (18.98%)	1,370 (10.97%)	1,350 (10.81%)	415 (3.32%)
1	8	10,896	3,777	10,312	126	1,552	2,184	1,284	3,777	6,513	8.59	52.85	918 (8.43%)	442 (4.06%)	1,765 (16.20%)	745 (6.84%)	2,397 (22.00%)	1,399 (12.84%)	1,578 (14.48%)	487 (4.47%)
1	9	10,475	4,493	10,177	100	1,634	2,271	1,563	4,493	8,121	8.84	56.99	922 (8.80%)	557 (5.32%)	2,101 (20.06%)	919 (8.77%)	2,755 (26.30%)	1,549 (14.79%)	1,964 (18.75%)	606 (5.79%)
1	10	10,699	5,574	10,586	121	1,902	2,655	2,034	5,574	10,483	8.43	56.85	972 (9.08%)	678 (6.34%)	2,198 (20.54%)	962 (8.99%)	2,915 (27.25%)	1,562 (14.60%)	2,090 (19.53%)	741 (6.93%)
1	11	10,070	5,985	10,015	178	2,080	3,090	2,306	5,985	11,451	7.98	56.22	905 (8.99%)	748 (7.43%)	2,094 (20.79%)	975 (9.68%)	2,903 (28.83%)	1,448 (14.38%)	1,992 (19.78%)	878 (8.72%)
1	12	8,009	5,020	7,988	142	2,257	3,415	1,885	5,020	10,034	8.56	58.06	751 (9.38%)	771 (9.63%)	1,788 (22.32%)	898 (11.21%)	2,612 (32.61%)	1,230 (15.36%)	1,877 (23.44%)	913 (11.40%)
1	13	5,733	3,755	5,718	121	2,129	3,375	1,301	3,755	7,922	9.52	61.30	682 (11.90%)	791 (13.80%)	1,457 (25.41%)	801 (13.97%)	2,071 (36.12%)	1,005 (17.53%)	1,581 (27.58%)	973 (16.97%)
1	14	3,895	2,649	3,887	87	1,837	3,201	863	2,649	5,867	9.59	63.83	520 (13.35%)	713 (18.31%)	1,099 (28.22%)	667 (17.12%)	1,545 (39.67%)	679 (17.43%)	1,200 (30.81%)	835 (21.44%)
1	15	2,534	1,776	2,528	70	1,534	2,982	554	1,776	4,316	9.55	64.26	411 (16.22%)	598 (23.60%)	762 (30.07%)	454 (17.92%)	1,056 (41.67%)	461 (18.19%)	863 (34.06%)	655 (25.85%)
1	16	1,705	1,268	1,701	55	1,178	2,675	410	1,268	3,276	9.73	62.93	314 (18.42%)	395 (23.17%)	532 (31.20%)	304 (17.83%)	724 (42.46%)	329 (19.30%)	620 (36.36%)	470 (27.57%)
1	17	1,048	823	1,046	55	791	2,162	273	823	2,406	8.06	60.17	192 (18.32%)	264 (25.19%)	328 (31.30%)	226 (21.56%)	441 (42.08%)	197 (18.80%)	387 (36.93%)	281 (26.81%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	738	584	738	37	612	2,082	193	584	1,988	8.19	58.99	173 (23.44%)	209 (28.32%)	232 (31.44%)	153 (20.73%)	323 (43.77%)	135 (18.29%)	263 (35.64%)	208 (28.18%)
1	19	471	384	470	30	405	1,609	113	384	1,410	7.30	57.03	143 (30.36%)	166 (35.24%)	149 (31.63%)	89 (18.90%)	201 (42.68%)	88 (18.68%)	185 (39.28%)	143 (30.36%)
1	20	526	471	526	70	488	3,299	181	471	2,539	9.95	53.41	187 (35.55%)	176 (33.46%)	170 (32.32%)	127 (24.14%)	223 (42.40%)	88 (16.73%)	206 (39.16%)	134 (25.48%)
1	Total	146,733	51,802	107,168	1,618	23,525	41,460	18,529	51,802	100,537	6.94	48.38	10,549 (7.19%)	7,483 (5.10%)	20,405 (13.91%)	9,717 (6.62%)	28,523 (19.44%)	15,114 (10.30%)	19,441 (13.25%)	8,255 (5.63%)

TABLE A.31: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Spells* model (*Sample-1*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-1;
Submodel: Cond_Prior-Spells

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	5,599	415	0	0	0	0	178	415	526	5.19	26.38	151 (2.70%)	10 (0.18%)	92 (1.64%)	42 (0.75%)	111 (1.98%)	54 (0.96%)	44 (0.79%)	1 (0.02%)
0	2	10,388	1,402	0	0	0	0	581	1,402	1,929	3.85	33.16	302 (2.91%)	40 (0.39%)	361 (3.48%)	175 (1.68%)	428 (4.12%)	206 (1.98%)	248 (2.39%)	20 (0.19%)
0	3	13,225	2,005	0	0	0	0	848	2,005	2,962	3.98	33.73	549 (4.15%)	86 (0.65%)	588 (4.45%)	335 (2.53%)	695 (5.26%)	292 (2.21%)	447 (3.38%)	45 (0.34%)
0	4	11,005	1,821	0	0	0	0	742	1,821	2,733	6.00	43.17	488 (4.43%)	146 (1.33%)	793 (7.21%)	371 (3.37%)	980 (8.91%)	322 (2.93%)	621 (5.64%)	92 (0.84%)
0	5	15,570	2,927	0	0	0	0	1,173	2,927	4,459	4.78	33.86	626 (4.02%)	199 (1.28%)	778 (5.00%)	445 (2.86%)	1,045 (6.71%)	295 (1.89%)	638 (4.10%)	137 (0.88%)
0	6	12,354	2,635	0	0	0	0	1,047	2,635	4,202	6.26	43.93	575 (4.65%)	203 (1.64%)	872 (7.06%)	464 (3.76%)	1,227 (9.93%)	280 (2.27%)	670 (5.42%)	171 (1.38%)
0	7	9,556	2,399	0	0	0	0	933	2,399	3,861	7.18	51.55	387 (4.05%)	257 (2.69%)	940 (9.84%)	437 (4.57%)	1,271 (13.30%)	219 (2.29%)	856 (8.96%)	210 (2.20%)
0	8	7,514	2,353	0	0	0	0	979	2,353	3,897	7.89	53.58	260 (3.46%)	276 (3.67%)	781 (10.39%)	422 (5.62%)	1,162 (15.46%)	162 (2.16%)	738 (9.82%)	221 (2.94%)
0	9	4,474	1,616	0	0	0	0	621	1,616	2,696	11.01	60.55	170 (3.80%)	219 (4.89%)	630 (14.08%)	345 (7.71%)	953 (21.30%)	169 (3.78%)	556 (12.43%)	252 (5.63%)
0	10	3,562	2,028	0	0	0	0	943	2,028	3,750	8.90	53.32	88 (2.47%)	137 (3.85%)	442 (12.41%)	205 (5.76%)	721 (20.24%)	144 (4.04%)	392 (11.01%)	223 (6.26%)
0	11	3,427	2,214	0	0	0	0	991	2,214	4,373	5.61	42.48	54 (1.58%)	81 (2.36%)	245 (7.15%)	143 (4.17%)	491 (14.33%)	96 (2.80%)	217 (6.33%)	164 (4.79%)
0	12	2,149	1,471	0	0	0	0	684	1,471	2,954	5.33	41.51	36 (1.68%)	45 (2.09%)	133 (6.19%)	80 (3.72%)	251 (11.68%)	72 (3.35%)	105 (4.89%)	151 (7.03%)
0	13	2,650	2,237	0	0	0	0	1,168	2,237	4,608	2.56	31.33	24 (0.91%)	17 (0.64%)	48 (1.81%)	21 (0.79%)	209 (7.89%)	41 (1.55%)	42 (1.58%)	56 (2.11%)
0	14	1,615	1,317	0	0	0	0	744	1,317	2,627	1.94	34.37	24 (1.49%)	10 (0.62%)	23 (1.42%)	10 (0.62%)	182 (11.27%)	15 (0.93%)	29 (1.80%)	35 (2.17%)
0	15	221	118	0	0	0	0	61	118	232	9.99	49.97	7 (3.17%)	3 (1.36%)	12 (5.43%)	10 (4.52%)	38 (17.19%)	13 (5.88%)	9 (4.07%)	8 (3.62%)
0	16	58	21	0	0	0	0	7	21	35	26.09	61.82	2 (3.45%)	1 (1.72%)	5 (8.62%)	8 (13.79%)	14 (24.14%)	4 (6.90%)	4 (6.90%)	3 (5.17%)
0	17	24	12	0	0	0	0	4	12	23	23.21	60.33	1 (4.17%)	0 (0.00%)	2 (8.33%)	1 (4.17%)	3 (12.50%)	1 (4.17%)	1 (4.17%)	2 (8.33%)

Continued on next page

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	15	3	0	0	0	0	1	3	4	38.93	64.77	0 (0.00%)	1 (6.67%)	4 (26.67%)	2 (13.33%)	6 (40.00%)	4 (26.67%)	3 (20.00%)	3 (20.00%)
0	19	5	2	0	0	0	0	0	2	2	26.80	52.80	0 (0.00%)	0 (0.00%)	1 (20.00%)	0 (0.00%)	1 (20.00%)	2 (40.00%)	1 (20.00%)	1 (20.00%)
0	20	4	1	0	0	0	0	0	1	2	110.00	68.00	0 (0.00%)	0 (0.00%)	1 (25.00%)	0 (0.00%)	2 (50.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
0	Total	103,415	26,997	0	0	0	0	11,705	26,997	45,875	5.75	40.84	3,744 (3.62%)	1,731 (1.67%)	6,751 (6.53%)	3,516 (3.40%)	9,790 (9.47%)	2,391 (2.31%)	5,621 (5.44%)	1,795 (1.74%)
1	1	159	13	159	2	24	31	2	13	19	473.89	33.46	7 (4.40%)	0 (0.00%)	5 (3.14%)	5 (3.14%)	7 (4.40%)	9 (5.66%)	0 (0.00%)	3 (1.89%)
1	2	2,299	316	2,299	15	200	223	109	316	438	7.98	26.95	129 (5.61%)	1 (0.04%)	59 (2.57%)	25 (1.09%)	83 (3.61%)	109 (4.74%)	33 (1.44%)	2 (0.09%)
1	3	4,871	991	4,871	29	445	513	344	991	1,524	5.67	39.52	334 (6.86%)	30 (0.62%)	333 (6.84%)	125 (2.57%)	419 (8.60%)	378 (7.76%)	244 (5.01%)	14 (0.29%)
1	4	5,174	1,417	5,174	58	635	778	502	1,417	2,359	6.63	50.19	387 (7.48%)	100 (1.93%)	741 (14.32%)	270 (5.22%)	868 (16.78%)	645 (12.47%)	576 (11.13%)	73 (1.41%)
1	5	7,427	2,058	7,427	84	998	1,231	673	2,058	3,466	6.90	47.66	575 (7.74%)	186 (2.50%)	1,012 (13.63%)	451 (6.07%)	1,260 (16.97%)	864 (11.63%)	859 (11.57%)	162 (2.18%)
1	6	11,510	2,929	11,510	137	1,631	2,112	941	2,929	4,941	5.96	41.19	901 (7.83%)	262 (2.28%)	1,185 (10.30%)	554 (4.81%)	1,607 (13.96%)	981 (8.52%)	1,017 (8.84%)	339 (2.95%)
1	7	16,155	3,828	16,155	209	2,173	2,940	1,259	3,828	6,488	5.77	41.50	1,227 (7.60%)	346 (2.14%)	1,489 (9.22%)	710 (4.39%)	2,259 (13.98%)	1,296 (8.02%)	1,257 (7.78%)	353 (2.19%)
1	8	14,134	4,180	14,134	137	2,354	3,065	1,323	4,180	7,169	7.62	49.55	1,250 (8.84%)	489 (3.46%)	1,977 (13.99%)	880 (6.23%)	2,872 (20.32%)	1,352 (9.57%)	1,822 (12.89%)	479 (3.39%)
1	9	11,950	4,554	11,950	132	2,374	3,132	1,395	4,554	8,084	9.37	58.60	1,198 (10.03%)	713 (5.97%)	2,483 (20.78%)	1,136 (9.51%)	3,256 (27.25%)	1,492 (12.49%)	2,457 (20.56%)	681 (5.70%)
1	10	10,524	4,912	10,524	195	2,474	3,511	1,478	4,912	8,899	10.29	64.12	1,062 (10.09%)	941 (8.94%)	2,660 (25.28%)	1,161 (11.03%)	3,404 (32.35%)	1,380 (13.11%)	2,544 (24.17%)	906 (8.61%)
1	11	9,355	5,048	9,355	166	2,485	3,598	1,642	5,048	9,510	10.35	64.89	921 (9.85%)	1,066 (11.39%)	2,403 (25.69%)	1,243 (13.29%)	3,132 (33.48%)	1,297 (13.86%)	2,435 (26.03%)	1,081 (11.56%)
1	12	9,819	6,352	9,819	218	2,894	4,316	2,486	6,352	12,703	8.19	56.81	842 (8.58%)	1,066 (10.86%)	2,014 (20.51%)	1,015 (10.34%)	2,864 (29.17%)	1,164 (11.85%)	2,117 (21.56%)	1,245 (12.68%)
1	13	9,802	7,199	9,802	186	2,824	4,457	3,038	7,199	14,787	6.22	50.38	811 (8.27%)	946 (9.65%)	1,611 (16.44%)	920 (9.39%)	2,564 (26.16%)	988 (10.08%)	1,712 (17.47%)	1,160 (11.83%)
1	14	6,238	4,705	6,238	126	2,535	4,212	1,898	4,705	10,293	6.25	53.26	666 (10.68%)	870 (13.95%)	1,229 (19.70%)	727 (11.65%)	1,923 (30.83%)	732 (11.73%)	1,330 (21.32%)	982 (15.74%)
1	15	3,462	2,602	3,462	88	1,956	3,609	956	2,602	6,391	7.84	57.38	539 (15.57%)	671 (19.38%)	854 (24.67%)	477 (13.78%)	1,250 (36.11%)	456 (13.17%)	934 (26.98%)	714 (20.62%)
1	16	2,006	1,538	2,006	64	1,361	2,974	516	1,538	4,066	8.34	59.58	341 (17.00%)	448 (22.33%)	543 (27.07%)	335 (16.70%)	772 (38.48%)	323 (16.10%)	652 (32.50%)	480 (23.93%)
1	17	1,348	1,063	1,348	60	1,011	2,678	370	1,063	3,033	8.65	59.59	266 (19.73%)	333 (24.70%)	397 (29.45%)	264 (19.58%)	535 (39.69%)	214 (15.88%)	475 (35.24%)	326 (24.18%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	873	684	873	38	724	2,319	219	684	2,195	8.94	61.29	202 (23.14%)	261 (29.90%)	288 (32.99%)	195 (22.34%)	395 (45.25%)	147 (16.84%)	331 (37.92%)	246 (28.18%)
1	19	584	472	584	20	495	1,882	135	472	1,744	9.26	59.83	166 (28.42%)	211 (36.13%)	205 (35.10%)	116 (19.86%)	269 (46.06%)	103 (17.64%)	245 (41.95%)	180 (30.82%)
1	20	650	576	650	78	610	3,853	217	576	3,006	10.53	56.84	235 (36.15%)	251 (38.62%)	227 (34.92%)	166 (25.54%)	285 (43.85%)	105 (16.15%)	279 (42.92%)	193 (29.69%)
1	Total	128,340	55,437	128,340	2,042	30,203	51,434	19,503	55,437	111,115	8.18	51.49	12,059 (9.40%)	9,191 (7.16%)	21,715 (16.92%)	10,775 (8.40%)	30,024 (23.39%)	14,035 (10.94%)	21,319 (16.61%)	9,619 (7.49%)

TABLE A.32: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Main* model (*Sample-2*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-2;
Submodel: Cond_Main

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	1	2,115	113	170	0	7	8	60	113	143	1.19	30.12	62 (2.93%)	3 (0.14%)	81 (3.83%)	11 (0.52%)	87 (4.11%)	28 (1.32%)	9 (0.43%)	0 (0.00%)
NA	2	8,978	1,107	1,230	5	58	76	508	1,107	1,550	1.57	30.59	232 (2.58%)	20 (0.22%)	399 (4.44%)	105 (1.17%)	470 (5.24%)	134 (1.49%)	132 (1.47%)	11 (0.12%)
NA	3	17,284	2,731	7,483	21	344	391	1,239	2,731	3,954	2.11	37.83	926 (5.36%)	96 (0.56%)	1,327 (7.68%)	411 (2.38%)	1,579 (9.14%)	675 (3.91%)	476 (2.75%)	42 (0.24%)
NA	4	16,487	3,556	9,345	56	743	886	1,537	3,556	5,661	3.85	45.09	1,257 (7.62%)	234 (1.42%)	2,392 (14.51%)	762 (4.62%)	2,913 (17.67%)	1,331 (8.07%)	1,206 (7.31%)	92 (0.56%)
NA	5	21,690	5,315	8,776	60	945	1,169	2,177	5,315	8,761	4.07	43.46	1,343 (6.19%)	446 (2.06%)	3,592 (16.56%)	1,149 (5.30%)	4,218 (19.45%)	1,608 (7.41%)	1,855 (8.55%)	239 (1.10%)
NA	6	25,407	6,741	13,498	98	1,281	1,657	2,583	6,741	11,473	4.38	45.20	1,805 (7.10%)	675 (2.66%)	4,314 (16.98%)	1,506 (5.93%)	5,006 (19.70%)	1,990 (7.83%)	2,232 (8.78%)	462 (1.82%)
NA	7	29,712	7,932	22,004	150	2,131	2,825	3,097	7,932	13,455	4.52	44.84	2,470 (8.31%)	851 (2.86%)	4,749 (15.98%)	1,755 (5.91%)	5,760 (19.39%)	2,589 (8.71%)	2,531 (8.52%)	642 (2.16%)
NA	8	27,463	7,871	23,804	208	3,005	3,930	2,820	7,871	13,906	5.14	48.00	2,737 (9.97%)	1,003 (3.65%)	5,310 (19.34%)	1,972 (7.18%)	6,699 (24.39%)	2,736 (9.96%)	3,042 (11.08%)	816 (2.97%)
NA	9	18,915	6,696	17,225	189	3,348	4,418	2,298	6,696	12,349	6.45	55.41	2,575 (13.61%)	1,275 (6.74%)	5,426 (28.69%)	2,066 (10.92%)	6,612 (34.96%)	2,411 (12.75%)	3,268 (17.28%)	956 (5.05%)
NA	10	15,059	7,074	12,371	147	3,065	4,304	2,472	7,074	13,579	6.88	58.37	1,981 (13.15%)	1,441 (9.57%)	5,078 (33.72%)	1,912 (12.70%)	6,014 (39.94%)	2,146 (14.25%)	3,270 (21.71%)	1,100 (7.30%)
NA	11	14,634	8,878	10,041	105	2,716	4,123	3,752	8,878	17,550	5.81	55.44	1,650 (11.28%)	1,439 (9.83%)	4,688 (32.03%)	1,880 (12.85%)	5,549 (37.92%)	1,916 (13.09%)	3,075 (21.01%)	1,314 (8.98%)
NA	12	13,229	8,875	12,154	233	3,207	5,029	3,687	8,875	18,195	5.57	53.56	1,456 (11.01%)	1,316 (9.95%)	3,981 (30.09%)	1,531 (11.57%)	4,839 (36.58%)	1,954 (14.77%)	2,665 (20.15%)	1,382 (10.45%)
NA	13	12,632	9,345	12,160	218	3,210	5,302	4,227	9,345	19,482	4.40	50.32	1,504 (11.91%)	1,274 (10.09%)	3,246 (25.70%)	1,289 (10.20%)	4,349 (34.43%)	1,725 (13.66%)	2,285 (18.09%)	1,314 (10.40%)
NA	14	7,872	5,864	7,756	141	3,004	5,161	2,468	5,864	13,169	5.52	53.91	1,263 (16.04%)	1,032 (13.11%)	2,491 (31.64%)	1,048 (13.31%)	3,359 (42.67%)	1,242 (15.78%)	1,732 (22.00%)	1,139 (14.47%)
NA	15	4,643	3,478	4,610	125	2,504	4,716	1,334	3,478	8,493	6.47	57.73	854 (18.39%)	824 (17.75%)	1,836 (39.54%)	790 (17.01%)	2,329 (50.16%)	855 (18.41%)	1,255 (27.03%)	858 (18.48%)
NA	16	2,817	2,104	2,798	105	1,875	4,316	779	2,104	5,533	7.69	62.21	624 (22.15%)	664 (23.57%)	1,373 (48.74%)	612 (21.73%)	1,655 (58.75%)	606 (21.51%)	1,015 (36.03%)	689 (24.46%)
NA	17	1,732	1,322	1,724	64	1,284	3,427	456	1,322	3,828	7.74	62.46	445 (25.69%)	468 (27.02%)	877 (50.64%)	410 (23.67%)	1,068 (61.66%)	361 (20.84%)	687 (39.67%)	446 (25.75%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	18	1,193	946	1,187	57	948	3,038	308	946	2,924	10.04	63.30	345 (28.92%)	377 (31.60%)	631 (52.89%)	296 (24.81%)	751 (62.95%)	297 (24.90%)	461 (38.64%)	320 (26.82%)
NA	19	854	689	850	59	696	2,747	230	689	2,432	9.67	62.55	293 (34.31%)	323 (37.82%)	463 (54.22%)	233 (27.28%)	537 (62.88%)	208 (24.36%)	395 (46.25%)	255 (29.86%)
NA	20	996	880	990	119	881	5,620	340	880	4,777	8.22	55.68	439 (44.08%)	363 (36.45%)	547 (54.92%)	325 (32.63%)	634 (63.65%)	210 (21.08%)	435 (43.67%)	244 (24.50%)
NA	Total	243,712	91,517	170,176	2,160	35,252	63,143	36,372	91,517	181,214	4.79	48.08	24,261 (9.95%)	14,124 (5.80%)	52,801 (21.67%)	20,063 (8.23%)	64,428 (26.44%)	25,022 (10.27%)	32,026 (13.14%)	12,321 (5.06%)

TABLE A.33: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Age-65p* model (*Sample-2*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-2;
Submodel: Cond_Age-65p

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	4,406	387	771	3	31	39	205	387	511	1.61	33.73	146 (3.31%)	10 (0.23%)	205 (4.65%)	38 (0.86%)	238 (5.40%)	81 (1.84%)	55 (1.25%)	2 (0.05%)
0	2	16,899	2,745	7,030	17	318	360	1,215	2,745	4,037	2.28	34.00	722 (4.27%)	86 (0.51%)	946 (5.60%)	295 (1.75%)	1,210 (7.16%)	704 (4.17%)	391 (2.31%)	19 (0.11%)
0	3	13,510	2,714	7,577	47	683	816	1,156	2,714	4,286	4.44	37.54	1,257 (9.30%)	189 (1.40%)	1,305 (9.66%)	568 (4.20%)	1,712 (12.67%)	888 (6.57%)	743 (5.50%)	53 (0.39%)
0	4	8,936	1,843	4,636	39	715	917	739	1,843	3,064	4.48	36.92	811 (9.08%)	161 (1.80%)	976 (10.92%)	458 (5.13%)	1,260 (14.10%)	559 (6.26%)	701 (7.84%)	81 (0.91%)
0	5	12,263	2,539	3,586	36	579	822	1,014	2,539	4,376	2.96	36.28	668 (5.45%)	178 (1.45%)	993 (8.10%)	404 (3.29%)	1,243 (10.14%)	531 (4.33%)	491 (4.00%)	98 (0.80%)
0	6	25,013	4,672	12,155	88	954	1,331	1,931	4,672	7,577	2.09	32.45	1,153 (4.61%)	209 (0.84%)	1,291 (5.16%)	579 (2.31%)	1,854 (7.41%)	1,054 (4.21%)	640 (2.56%)	124 (0.50%)
0	7	24,891	4,991	20,253	126	1,770	2,330	1,943	4,991	8,296	3.03	33.38	1,999 (8.03%)	227 (0.91%)	1,538 (6.18%)	759 (3.05%)	2,523 (10.14%)	1,349 (5.42%)	734 (2.95%)	108 (0.43%)
0	8	14,438	3,759	12,764	121	2,085	2,682	1,362	3,759	6,732	4.07	35.57	2,251 (15.59%)	294 (2.04%)	1,363 (9.44%)	803 (5.56%)	2,195 (15.20%)	945 (6.55%)	805 (5.58%)	136 (0.94%)
0	9	8,511	3,094	7,291	85	1,846	2,540	1,245	3,094	5,888	4.61	37.86	1,511 (17.75%)	350 (4.11%)	1,203 (14.13%)	773 (9.08%)	1,711 (20.10%)	684 (8.04%)	749 (8.80%)	136 (1.60%)
0	10	4,992	2,173	4,626	62	1,509	2,215	841	2,173	4,295	4.92	40.04	1,039 (20.81%)	303 (6.07%)	938 (18.79%)	578 (11.58%)	1,259 (25.22%)	487 (9.76%)	624 (12.50%)	153 (3.06%)
0	11	3,236	1,705	2,908	43	1,086	1,785	636	1,705	3,570	5.48	41.16	666 (20.58%)	229 (7.08%)	678 (20.95%)	436 (13.47%)	918 (28.37%)	385 (11.90%)	454 (14.03%)	128 (3.96%)
0	12	5,498	3,840	2,162	37	959	1,743	1,785	3,840	8,077	2.31	33.77	457 (8.31%)	203 (3.69%)	527 (9.59%)	338 (6.15%)	842 (15.31%)	314 (5.71%)	381 (6.93%)	128 (2.33%)
0	13	8,155	6,492	5,123	108	1,149	2,102	3,306	6,492	13,380	1.39	31.44	433 (5.31%)	171 (2.10%)	418 (5.13%)	260 (3.19%)	868 (10.64%)	517 (6.34%)	281 (3.45%)	119 (1.46%)
0	14	8,528	7,233	8,078	160	1,514	2,728	3,784	7,233	15,251	1.16	30.92	597 (7.00%)	103 (1.21%)	347 (4.07%)	223 (2.61%)	1,142 (13.39%)	552 (6.47%)	223 (2.61%)	72 (0.84%)
0	15	5,006	4,305	4,909	107	1,487	2,567	2,156	4,305	10,103	1.68	31.72	729 (14.56%)	134 (2.68%)	305 (6.09%)	210 (4.19%)	967 (19.32%)	336 (6.71%)	185 (3.70%)	78 (1.56%)
0	16	2,457	2,117	2,448	65	1,233	2,409	1,029	2,117	5,569	2.13	33.21	528 (21.49%)	111 (4.52%)	252 (10.26%)	163 (6.63%)	611 (24.87%)	212 (8.63%)	154 (6.27%)	63 (2.56%)
0	17	1,361	1,122	1,358	65	924	2,348	488	1,122	3,375	3.47	35.78	352 (25.86%)	93 (6.83%)	216 (15.87%)	153 (11.24%)	410 (30.12%)	157 (11.54%)	141 (10.36%)	67 (4.92%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	809	678	807	34	622	1,844	289	678	2,264	3.39	37.58	229 (28.31%)	76 (9.39%)	162 (20.02%)	114 (14.09%)	269 (33.25%)	100 (12.36%)	99 (12.24%)	42 (5.19%)
0	19	569	453	568	35	452	1,776	189	453	1,788	5.29	40.24	180 (31.63%)	66 (11.60%)	131 (23.02%)	97 (17.05%)	208 (36.56%)	78 (13.71%)	89 (15.64%)	35 (6.15%)
0	20	919	806	908	101	796	4,930	318	806	4,384	8.19	43.62	385 (41.89%)	182 (19.80%)	329 (35.80%)	237 (25.79%)	427 (46.46%)	147 (16.00%)	240 (26.12%)	93 (10.12%)
0	Total	170,397	57,668	109,958	1,379	20,712	38,284	25,631	57,668	116,823	3.03	34.65	16,113 (9.46%)	3,375 (1.98%)	14,123 (8.29%)	7,486 (4.39%)	21,867 (12.83%)	10,080 (5.92%)	8,180 (4.80%)	1,735 (1.02%)
1	1	39	3	1	0	1	1	1	3	4	0.36	94.80	0 (0.00%)	0 (0.00%)	2 (5.13%)	0 (0.00%)	2 (5.13%)	0 (0.00%)	0 (0.00%)	1 (2.56%)
1	2	143	7	5	0	0	0	4	7	10	0.67	80.49	2 (1.40%)	0 (0.00%)	10 (6.99%)	4 (2.80%)	11 (7.69%)	5 (3.50%)	3 (2.10%)	0 (0.00%)
1	3	1,094	142	37	0	0	0	65	142	196	1.03	77.61	28 (2.56%)	23 (2.10%)	258 (23.58%)	72 (6.58%)	264 (24.13%)	24 (2.19%)	77 (7.04%)	7 (0.64%)
1	4	939	169	295	2	18	18	84	169	218	4.31	76.21	39 (4.15%)	31 (3.30%)	321 (34.19%)	89 (9.48%)	336 (35.78%)	53 (5.64%)	111 (11.82%)	14 (1.49%)
1	5	3,028	715	608	3	39	44	295	715	1,144	3.66	76.95	67 (2.21%)	75 (2.48%)	994 (32.83%)	232 (7.66%)	1,050 (34.68%)	178 (5.88%)	306 (10.11%)	50 (1.65%)
1	6	3,774	1,070	1,109	5	81	97	441	1,070	1,728	3.82	76.89	202 (5.35%)	210 (5.56%)	1,345 (35.64%)	381 (10.10%)	1,454 (38.53%)	311 (8.24%)	589 (15.61%)	106 (2.81%)
1	7	5,495	1,704	3,666	14	188	228	703	1,704	2,791	5.85	77.24	317 (5.77%)	217 (3.95%)	2,145 (39.04%)	468 (8.52%)	2,281 (41.51%)	757 (13.78%)	908 (16.52%)	167 (3.04%)
1	8	9,458	3,447	7,806	32	425	492	1,254	3,447	5,728	6.92	77.97	652 (6.89%)	511 (5.40%)	4,298 (45.44%)	1,080 (11.42%)	4,624 (48.89%)	1,715 (18.13%)	2,278 (24.09%)	330 (3.49%)
1	9	11,666	4,902	9,994	58	1,000	1,186	1,668	4,902	8,539	9.07	79.08	1,052 (9.02%)	956 (8.19%)	5,919 (50.74%)	1,667 (14.29%)	6,461 (55.38%)	2,404 (20.61%)	3,554 (30.46%)	776 (6.65%)
1	10	10,913	5,343	10,182	105	1,604	1,999	1,629	5,343	9,588	10.37	80.36	1,203 (11.02%)	1,341 (12.29%)	6,068 (55.60%)	1,980 (18.14%)	6,718 (61.56%)	2,515 (23.05%)	3,765 (34.50%)	1,365 (12.51%)
1	11	9,140	4,935	8,967	102	2,235	2,810	1,435	4,935	9,253	11.09	81.15	1,214 (13.28%)	1,587 (17.36%)	5,430 (59.41%)	1,843 (20.16%)	6,094 (66.67%)	2,161 (23.64%)	3,469 (37.95%)	1,715 (18.76%)
1	12	6,549	3,811	6,486	100	2,268	3,124	1,061	3,811	7,334	12.05	81.60	997 (15.22%)	1,513 (23.10%)	4,118 (62.88%)	1,490 (22.75%)	4,640 (70.85%)	1,718 (26.23%)	2,776 (42.39%)	1,699 (25.94%)
1	13	4,337	2,747	4,327	87	2,013	3,136	754	2,747	5,559	11.50	81.95	704 (16.23%)	1,287 (29.67%)	2,886 (66.54%)	1,091 (25.16%)	3,235 (74.59%)	1,131 (26.08%)	2,069 (47.71%)	1,428 (32.93%)
1	14	2,701	1,782	2,698	70	1,558	2,844	493	1,782	3,902	12.03	81.63	547 (20.25%)	958 (35.47%)	1,862 (68.94%)	770 (28.51%)	2,067 (76.53%)	743 (27.51%)	1,409 (52.17%)	1,038 (38.43%)
1	15	1,651	1,181	1,649	60	1,152	2,422	296	1,181	2,795	11.46	81.28	362 (21.93%)	741 (44.88%)	1,189 (72.02%)	544 (32.95%)	1,336 (80.92%)	478 (28.95%)	977 (59.18%)	719 (43.55%)
1	16	941	693	941	35	732	1,812	187	693	1,731	10.31	80.98	228 (24.23%)	410 (43.57%)	696 (73.96%)	316 (33.58%)	763 (81.08%)	267 (28.37%)	543 (57.70%)	413 (43.89%)
1	17	609	473	609	29	493	1,465	133	473	1,213	10.87	80.21	184 (30.21%)	330 (54.19%)	466 (76.52%)	211 (34.65%)	498 (81.77%)	191 (31.36%)	394 (64.70%)	306 (50.25%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	383	329	383	21	323	1,073	104	329	983	9.83	80.36	142 (37.08%)	239 (62.40%)	302 (78.85%)	150 (39.16%)	331 (86.42%)	126 (32.90%)	273 (71.28%)	203 (53.00%)
1	19	245	203	245	22	217	955	68	203	738	9.54	79.86	101 (41.22%)	164 (66.94%)	193 (78.78%)	96 (39.18%)	212 (86.53%)	93 (37.96%)	175 (71.43%)	129 (52.65%)
1	20	210	193	210	36	193	1,153	66	193	937	7.48	76.33	107 (50.95%)	156 (74.29%)	176 (83.81%)	93 (44.29%)	184 (87.62%)	72 (34.29%)	170 (80.95%)	120 (57.14%)
1	Total	73,315	33,849	60,218	781	14,540	24,859	10,741	33,849	64,391	8.91	79.58	8,148 (11.11%)	10,749 (14.66%)	38,678 (52.76%)	12,577 (17.15%)	42,561 (58.05%)	14,942 (20.38%)	23,846 (32.53%)	10,586 (14.44%)

TABLE A.34: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Acute-12-month* model (*Sample-2*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-2;
Submodel: Cond_Prior-Acute-12-month

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	2,352	133	153	0	6	9	68	133	164	1.48	31.08	71 (3.02%)	4 (0.17%)	98 (4.17%)	20 (0.85%)	105 (4.46%)	41 (1.74%)	14 (0.60%)	1 (0.04%)
0	2	10,191	1,237	1,024	4	43	67	576	1,237	1,751	1.88	31.70	281 (2.76%)	29 (0.28%)	444 (4.36%)	135 (1.32%)	524 (5.14%)	193 (1.89%)	130 (1.28%)	15 (0.15%)
0	3	12,512	1,990	3,518	14	224	274	1,008	1,990	2,924	3.00	41.92	707 (5.65%)	106 (0.85%)	1,295 (10.35%)	414 (3.31%)	1,484 (11.86%)	606 (4.84%)	418 (3.34%)	34 (0.27%)
0	4	12,007	2,528	3,402	25	349	438	1,103	2,528	4,116	3.92	44.80	600 (5.00%)	167 (1.39%)	1,709 (14.23%)	515 (4.29%)	1,972 (16.42%)	821 (6.84%)	694 (5.78%)	68 (0.57%)
0	5	17,271	3,655	4,520	30	390	494	1,602	3,655	5,740	3.33	40.95	912 (5.28%)	270 (1.56%)	2,225 (12.88%)	724 (4.19%)	2,520 (14.59%)	1,000 (5.79%)	933 (5.40%)	119 (0.69%)
0	6	14,828	3,616	6,031	40	604	741	1,509	3,616	5,949	4.22	47.75	1,079 (7.28%)	374 (2.52%)	2,588 (17.45%)	905 (6.10%)	2,908 (19.61%)	1,200 (8.09%)	1,067 (7.20%)	190 (1.28%)
0	7	12,050	3,127	5,646	67	926	1,141	1,231	3,127	5,214	5.21	51.19	821 (6.81%)	391 (3.24%)	2,382 (19.77%)	801 (6.65%)	2,684 (22.27%)	1,195 (9.92%)	1,124 (9.33%)	242 (2.01%)
0	8	9,842	2,752	4,866	71	1,029	1,327	1,178	2,752	4,727	5.55	51.37	612 (6.22%)	412 (4.19%)	2,003 (20.35%)	728 (7.40%)	2,338 (23.76%)	967 (9.83%)	974 (9.90%)	265 (2.69%)
0	9	6,251	2,147	3,685	77	1,024	1,379	937	2,147	3,650	6.73	55.36	429 (6.86%)	329 (5.26%)	1,593 (25.48%)	559 (8.94%)	1,886 (30.17%)	827 (13.23%)	774 (12.38%)	251 (4.02%)
0	10	4,187	1,850	2,554	57	907	1,233	751	1,850	3,469	7.51	56.17	273 (6.52%)	276 (6.59%)	1,161 (27.73%)	443 (10.58%)	1,386 (33.10%)	624 (14.90%)	630 (15.05%)	238 (5.68%)
0	11	4,344	2,646	1,978	51	715	1,078	1,159	2,646	5,292	5.55	46.53	214 (4.93%)	214 (4.93%)	829 (19.08%)	319 (7.34%)	1,043 (24.01%)	497 (11.44%)	470 (10.82%)	247 (5.69%)
0	12	4,691	3,569	1,795	62	675	1,073	1,924	3,569	7,077	3.59	39.23	152 (3.24%)	155 (3.30%)	548 (11.68%)	203 (4.33%)	824 (17.57%)	417 (8.89%)	298 (6.35%)	197 (4.20%)
0	13	3,573	2,855	1,834	77	772	1,256	1,537	2,855	5,943	3.42	38.98	141 (3.95%)	113 (3.16%)	379 (10.61%)	152 (4.25%)	608 (17.02%)	338 (9.46%)	226 (6.33%)	157 (4.39%)
0	14	1,436	1,073	1,111	39	637	1,015	566	1,073	2,285	5.30	45.37	93 (6.48%)	74 (5.15%)	256 (17.83%)	101 (7.03%)	366 (25.49%)	215 (14.97%)	140 (9.75%)	96 (6.69%)
0	15	608	430	535	20	399	795	191	430	966	8.78	53.98	41 (6.74%)	53 (8.72%)	160 (26.32%)	60 (9.87%)	205 (33.72%)	136 (22.37%)	90 (14.80%)	52 (8.55%)
0	16	258	165	238	16	187	454	63	165	333	18.35	64.76	22 (8.53%)	18 (6.98%)	105 (40.70%)	42 (16.28%)	124 (48.06%)	75 (29.07%)	52 (20.16%)	41 (15.89%)
0	17	142	92	134	11	119	337	29	92	210	21.62	71.37	12 (8.45%)	18 (12.68%)	77 (54.23%)	34 (23.94%)	85 (59.86%)	58 (40.85%)	41 (28.87%)	26 (18.31%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	81	54	77	6	61	203	15	54	133	20.15	72.84	8 (9.88%)	14 (17.28%)	41 (50.62%)	21 (25.93%)	52 (64.20%)	29 (35.80%)	19 (23.46%)	20 (24.69%)
0	19	57	30	49	6	40	152	7	30	67	28.89	69.47	5 (8.77%)	7 (12.28%)	32 (56.14%)	11 (19.30%)	35 (61.40%)	21 (36.84%)	16 (28.07%)	9 (15.79%)
0	20	40	29	35	3	19	118	10	29	66	33.23	66.56	5 (12.50%)	8 (20.00%)	19 (47.50%)	4 (10.00%)	22 (55.00%)	13 (32.50%)	10 (25.00%)	2 (5.00%)
0	Total	116,721	33,978	43,185	676	9,126	13,584	15,464	33,978	60,076	4.26	44.88	6,478 (5.55%)	3,032 (2.60%)	17,944 (15.37%)	6,191 (5.30%)	21,171 (18.14%)	9,273 (7.94%)	8,120 (6.96%)	2,270 (1.94%)
1	1	52	7	52	0	2	2	4	7	12	0.92	23.10	1 (1.92%)	0 (0.00%)	3 (5.77%)	0 (0.00%)	3 (5.77%)	2 (3.85%)	0 (0.00%)	0 (0.00%)
1	2	502	62	502	0	4	4	24	62	91	1.81	31.64	41 (8.17%)	2 (0.40%)	41 (8.17%)	12 (2.39%)	47 (9.36%)	11 (2.19%)	14 (2.79%)	1 (0.20%)
1	3	3,403	614	3,403	1	38	38	220	614	852	1.69	28.65	285 (8.37%)	16 (0.47%)	130 (3.82%)	51 (1.50%)	186 (5.47%)	132 (3.88%)	77 (2.26%)	10 (0.29%)
1	4	5,721	1,267	5,721	6	250	263	474	1,267	2,006	3.19	39.77	606 (10.59%)	56 (0.98%)	623 (10.89%)	206 (3.60%)	854 (14.93%)	406 (7.10%)	353 (6.17%)	27 (0.47%)
1	5	5,075	1,556	5,075	12	456	526	579	1,556	2,615	5.22	52.23	634 (12.49%)	187 (3.68%)	1,300 (25.62%)	413 (8.14%)	1,600 (31.53%)	645 (12.71%)	834 (16.43%)	88 (1.73%)
1	6	8,586	2,630	8,586	21	513	626	911	2,630	4,619	4.65	44.61	797 (9.28%)	298 (3.47%)	1,817 (21.16%)	621 (7.23%)	2,133 (24.84%)	909 (10.59%)	1,158 (13.49%)	234 (2.73%)
1	7	17,494	4,421	17,494	47	841	1,034	1,523	4,421	7,715	3.79	39.74	1,575 (9.00%)	454 (2.60%)	2,224 (12.71%)	896 (5.12%)	3,068 (17.54%)	1,300 (7.43%)	1,406 (8.04%)	423 (2.42%)
1	8	15,571	4,666	15,571	58	1,745	2,008	1,574	4,666	8,380	4.95	45.71	1,890 (12.14%)	586 (3.76%)	2,906 (18.66%)	1,088 (6.99%)	3,938 (25.29%)	1,459 (9.37%)	1,754 (11.26%)	519 (3.33%)
1	9	12,077	4,531	12,077	89	2,186	2,704	1,456	4,531	8,346	6.44	53.81	1,860 (15.40%)	759 (6.28%)	3,493 (28.92%)	1,347 (11.15%)	4,304 (35.64%)	1,471 (12.18%)	2,322 (19.23%)	667 (5.52%)
1	10	9,883	4,365	9,883	84	2,254	3,089	1,312	4,365	8,353	7.27	61.51	1,644 (16.63%)	1,040 (10.52%)	3,809 (38.54%)	1,461 (14.78%)	4,433 (44.85%)	1,475 (14.92%)	2,689 (27.21%)	842 (8.52%)
1	11	9,532	5,219	9,532	84	2,064	3,036	1,796	5,219	10,325	6.35	61.92	1,446 (15.17%)	1,178 (12.36%)	3,814 (40.01%)	1,513 (15.87%)	4,375 (45.90%)	1,510 (15.84%)	2,667 (27.98%)	992 (10.41%)
1	12	11,706	7,964	11,706	117	2,278	3,442	3,269	7,964	15,653	5.10	54.10	1,456 (12.44%)	1,194 (10.20%)	3,605 (30.80%)	1,375 (11.75%)	4,465 (38.14%)	1,620 (13.84%)	2,503 (21.38%)	1,194 (10.20%)
1	13	9,693	6,968	9,693	108	2,709	4,050	2,844	6,968	14,931	4.93	54.96	1,439 (14.85%)	1,220 (12.59%)	3,141 (32.40%)	1,261 (13.01%)	4,080 (42.09%)	1,441 (14.87%)	2,242 (23.13%)	1,226 (12.65%)
1	14	6,452	4,616	6,452	149	2,789	4,578	1,765	4,616	10,618	5.92	58.95	1,228 (19.03%)	1,091 (16.91%)	2,522 (39.09%)	1,044 (16.18%)	3,206 (49.69%)	1,048 (16.24%)	1,779 (27.57%)	1,088 (16.86%)
1	15	4,017	2,905	4,017	144	2,353	4,411	1,076	2,905	7,048	6.56	61.01	796 (19.82%)	847 (21.09%)	1,769 (44.04%)	775 (19.29%)	2,162 (53.82%)	775 (19.29%)	1,226 (30.52%)	844 (21.01%)
1	16	2,580	1,971	2,580	114	1,791	4,047	679	1,971	5,387	7.33	63.35	571 (22.13%)	670 (25.97%)	1,274 (49.38%)	569 (22.05%)	1,564 (60.62%)	533 (20.66%)	1,010 (39.15%)	663 (25.70%)
1	17	1,737	1,325	1,737	115	1,347	3,700	475	1,325	3,833	7.57	62.14	454 (26.14%)	480 (27.63%)	876 (50.43%)	418 (24.06%)	1,036 (59.64%)	385 (22.16%)	656 (37.77%)	459 (26.42%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	1,120	893	1,120	88	915	3,068	318	893	2,897	7.98	62.19	336 (30.00%)	358 (31.96%)	571 (50.98%)	289 (25.80%)	687 (61.34%)	237 (21.16%)	452 (40.36%)	298 (26.61%)
1	19	845	711	845	83	729	2,989	257	711	2,582	8.25	59.58	318 (37.63%)	330 (39.05%)	442 (52.31%)	236 (27.93%)	534 (63.20%)	192 (22.72%)	357 (42.25%)	245 (28.99%)
1	20	945	848	945	164	862	5,944	352	848	4,875	7.27	53.63	406 (42.96%)	326 (34.50%)	497 (52.59%)	297 (31.43%)	582 (61.59%)	198 (20.95%)	407 (43.07%)	231 (24.44%)
1	Total	126,991	57,539	126,991	1,484	26,126	49,559	20,908	57,539	121,138	5.29	51.02	17,783 (14.00%)	11,092 (8.73%)	34,857 (27.45%)	13,872 (10.92%)	43,257 (34.06%)	15,749 (12.40%)	23,906 (18.82%)	10,051 (7.91%)

TABLE A.35: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Oper-12-month* model (*Sample-2*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-2;
Submodel: Cond_Prior-Oper-12-month

If[a]	**#**[b]	**TP+FP**[o]	**TP**[d]	**Prior Spells**[e]	**Prior 30-d**[f]	**Prior 12-m**[g]	**Prior 12-m Spells**[h]	**Post 30-d**[i]	**Post 12-m**[j]	**Post 12-m Spells**[k]	**Avg. Stay.**[l]	**Avg. Age**[m]	**Asth.**[n]	**COPD**[o]	**Depr.**[p]	**Diab.**[q]	**Hype.**[r]	**Canc.**[s]	**CHD**[t]	**CHF**[u]
0	1	188	20	6	0	0	0	11	20	20	38.32	28.16	1 (0.53%)	0 (0.00%)	1 (0.53%)	2 (1.06%)	2 (1.06%)	0 (0.00%)	1 (0.53%)	0 (0.00%)
0	2	4,044	293	52	1	2	2	160	293	373	1.46	32.86	164 (4.06%)	23 (0.57%)	229 (5.66%)	111 (2.74%)	241 (5.96%)	12 (0.30%)	61 (1.51%)	9 (0.22%)
0	3	4,921	603	317	0	11	18	313	603	805	2.82	35.36	217 (4.41%)	53 (1.08%)	346 (7.03%)	143 (2.91%)	362 (7.36%)	35 (0.71%)	115 (2.34%)	16 (0.33%)
0	4	11,638	2,044	543	3	46	55	886	2,044	3,186	1.64	34.07	595 (5.11%)	128 (1.10%)	844 (7.25%)	362 (3.11%)	888 (7.63%)	59 (0.51%)	273 (2.35%)	27 (0.23%)
0	5	9,185	2,036	1,484	6	81	88	872	2,036	3,347	2.62	42.87	721 (7.85%)	213 (2.32%)	1,225 (13.34%)	566 (6.16%)	1,283 (13.97%)	93 (1.01%)	477 (5.19%)	86 (0.94%)
0	6	5,465	1,489	2,572	6	195	215	623	1,489	2,539	5.00	46.22	547 (10.01%)	182 (3.33%)	910 (16.65%)	407 (7.45%)	973 (17.80%)	89 (1.63%)	450 (8.23%)	106 (1.94%)
0	7	3,799	1,273	2,384	20	417	488	470	1,273	2,209	7.52	51.76	447 (11.77%)	182 (4.79%)	801 (21.08%)	383 (10.08%)	865 (22.77%)	86 (2.26%)	426 (11.21%)	141 (3.71%)
0	8	2,609	968	1,925	16	437	527	336	968	1,757	9.09	57.60	338 (12.96%)	180 (6.90%)	675 (25.87%)	320 (12.27%)	745 (28.56%)	64 (2.45%)	363 (13.91%)	129 (4.94%)
0	9	1,869	812	1,518	13	375	466	251	812	1,454	10.95	62.06	229 (12.25%)	175 (9.36%)	543 (29.05%)	243 (13.00%)	608 (32.53%)	54 (2.89%)	299 (16.00%)	141 (7.54%)
0	10	1,480	724	1,219	10	353	471	237	724	1,372	10.34	63.12	163 (11.01%)	180 (12.16%)	467 (31.55%)	208 (14.05%)	530 (35.81%)	41 (2.77%)	299 (20.20%)	159 (10.74%)
0	11	1,019	529	819	3	268	342	181	529	962	9.24	63.88	129 (12.66%)	143 (14.03%)	338 (33.17%)	160 (15.70%)	383 (37.59%)	30 (2.94%)	180 (17.66%)	138 (13.54%)
0	12	653	398	574	5	235	347	132	398	853	9.00	66.40	81 (12.40%)	133 (20.37%)	245 (37.52%)	114 (17.46%)	286 (43.80%)	25 (3.83%)	147 (22.51%)	103 (15.77%)
0	13	908	579	382	6	182	291	228	579	1,222	4.79	44.78	51 (5.62%)	85 (9.36%)	150 (16.52%)	67 (7.38%)	193 (21.26%)	18 (1.98%)	86 (9.47%)	70 (7.71%)
0	14	1,748	1,249	256	7	148	272	526	1,249	2,810	1.62	34.32	75 (4.29%)	66 (3.78%)	117 (6.69%)	48 (2.75%)	177 (10.13%)	22 (1.26%)	66 (3.78%)	54 (3.09%)
0	15	1,145	796	239	8	112	201	378	796	1,681	1.89	34.78	70 (6.11%)	56 (4.89%)	74 (6.46%)	42 (3.67%)	132 (11.53%)	25 (2.18%)	47 (4.10%)	32 (2.79%)
0	16	934	744	387	7	113	222	379	744	1,555	1.39	32.19	64 (6.85%)	48 (5.14%)	48 (5.14%)	27 (2.89%)	84 (8.99%)	21 (2.25%)	43 (4.60%)	25 (2.68%)
0	17	631	523	387	15	207	353	305	523	1,129	1.40	33.38	57 (9.03%)	26 (4.12%)	40 (6.34%)	22 (3.49%)	68 (10.78%)	7 (1.11%)	33 (5.23%)	17 (2.69%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	2,422	2,172	258	23	169	319	1,270	2,172	4,313	0.46	27.95	47 (1.94%)	36 (1.49%)	35 (1.45%)	22 (0.91%)	166 (6.85%)	6 (0.25%)	30 (1.24%)	20 (0.83%)
0	19	2,580	2,339	525	17	137	275	1,377	2,339	4,765	0.68	28.61	95 (3.68%)	32 (1.24%)	28 (1.09%)	17 (0.66%)	179 (6.94%)	9 (0.35%)	24 (0.93%)	19 (0.74%)
0	20	895	838	865	64	559	1,194	506	838	1,995	1.02	29.50	89 (9.94%)	29 (3.24%)	32 (3.58%)	22 (2.46%)	94 (10.50%)	1 (0.11%)	19 (2.12%)	16 (1.79%)
0	Total	58,133	20,429	16,712	230	4,047	6,146	9,441	20,429	38,347	3.72	40.94	4,180 (7.19%)	1,970 (3.39%)	7,148 (12.30%)	3,286 (5.65%)	8,259 (14.21%)	697 (1.20%)	3,439 (5.92%)	1,308 (2.25%)
1	1	1,238	59	66	0	10	13	29	59	77	0.85	26.67	20 (1.62%)	1 (0.08%)	31 (2.50%)	2 (0.16%)	33 (2.67%)	12 (0.97%)	2 (0.16%)	0 (0.00%)
1	2	5,547	597	900	6	74	85	268	597	834	1.32	30.14	168 (3.03%)	12 (0.22%)	248 (4.47%)	67 (1.21%)	278 (5.01%)	84 (1.51%)	68 (1.23%)	5 (0.09%)
1	3	12,767	1,842	4,960	21	336	401	846	1,842	2,651	1.96	35.40	546 (4.28%)	37 (0.29%)	773 (6.05%)	192 (1.50%)	937 (7.34%)	417 (3.27%)	220 (1.72%)	24 (0.19%)
1	4	12,854	2,741	8,457	27	546	639	1,174	2,741	4,313	3.00	43.43	900 (7.00%)	118 (0.92%)	1,562 (12.15%)	471 (3.66%)	1,928 (15.00%)	992 (7.72%)	686 (5.34%)	48 (0.37%)
1	5	15,858	3,698	9,229	66	1,166	1,433	1,521	3,698	6,053	4.08	45.17	1,023 (6.45%)	273 (1.72%)	2,571 (16.21%)	727 (4.58%)	3,198 (20.17%)	1,431 (9.02%)	1,377 (8.68%)	121 (0.76%)
1	6	20,014	4,991	16,333	88	1,368	1,790	1,927	4,991	8,389	4.36	44.57	1,276 (6.38%)	447 (2.23%)	3,095 (15.46%)	981 (4.90%)	4,017 (20.07%)	2,121 (10.60%)	1,834 (9.16%)	307 (1.53%)
1	7	21,927	5,913	20,075	105	1,782	2,269	2,237	5,913	10,217	4.81	46.40	1,841 (8.40%)	615 (2.80%)	3,863 (17.62%)	1,234 (5.63%)	5,094 (23.23%)	2,448 (11.16%)	2,170 (9.90%)	496 (2.26%)
1	8	18,202	5,793	17,188	102	2,212	2,785	2,037	5,793	10,104	5.77	52.07	2,112 (11.60%)	763 (4.19%)	4,460 (24.50%)	1,530 (8.41%)	5,606 (30.80%)	2,543 (13.97%)	2,531 (13.91%)	700 (3.85%)
1	9	14,565	5,716	14,092	126	2,506	3,347	1,935	5,716	10,537	6.56	57.98	2,001 (13.74%)	1,021 (7.01%)	4,767 (32.73%)	1,732 (11.89%)	5,632 (38.67%)	2,477 (17.01%)	2,885 (19.81%)	851 (5.84%)
1	10	14,033	7,333	13,830	146	2,618	3,746	2,711	7,333	14,174	6.26	57.10	1,878 (13.38%)	1,231 (8.77%)	4,752 (33.86%)	1,725 (12.29%)	5,603 (39.93%)	2,436 (17.36%)	3,038 (21.65%)	1,010 (7.20%)
1	11	13,553	8,283	13,428	138	2,655	3,952	3,443	8,283	16,276	5.85	56.00	1,711 (12.62%)	1,273 (9.39%)	4,449 (32.83%)	1,716 (12.66%)	5,386 (39.74%)	2,197 (16.21%)	2,978 (21.97%)	1,138 (8.40%)
1	12	11,060	7,238	11,005	143	2,986	4,460	2,822	7,238	14,746	6.21	57.63	1,599 (14.46%)	1,286 (11.63%)	3,901 (35.27%)	1,517 (13.72%)	4,924 (44.52%)	2,005 (18.13%)	2,671 (24.15%)	1,284 (11.61%)
1	13	8,012	5,257	7,989	161	2,931	4,642	1,916	5,257	11,523	6.79	61.85	1,408 (17.57%)	1,235 (15.41%)	3,314 (41.36%)	1,310 (16.35%)	4,101 (51.19%)	1,577 (19.68%)	2,315 (28.89%)	1,241 (15.49%)
1	14	5,448	3,687	5,439	152	2,600	4,651	1,308	3,687	8,611	7.65	63.82	1,013 (18.59%)	1,013 (18.59%)	2,485 (45.61%)	1,055 (19.36%)	2,993 (54.94%)	1,182 (21.70%)	1,762 (32.34%)	1,093 (20.06%)
1	15	3,686	2,640	3,675	124	2,177	4,323	886	2,640	6,391	8.63	65.54	698 (18.94%)	791 (21.46%)	1,852 (50.24%)	804 (21.81%)	2,198 (59.63%)	840 (22.79%)	1,308 (35.49%)	885 (24.01%)
1	16	2,480	1,800	2,473	121	1,716	4,018	617	1,800	4,710	8.81	65.67	517 (20.85%)	658 (26.53%)	1,278 (51.53%)	566 (22.82%)	1,549 (62.46%)	574 (23.15%)	1,003 (40.44%)	654 (26.37%)
1	17	1,649	1,256	1,648	110	1,250	3,511	422	1,256	3,788	8.57	63.59	419 (25.41%)	464 (28.14%)	861 (52.21%)	388 (23.53%)	1,023 (62.04%)	389 (23.59%)	635 (38.51%)	440 (26.68%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	1,072	850	1,069	67	846	2,882	282	850	2,689	8.99	62.75	315 (29.38%)	347 (32.37%)	558 (52.05%)	287 (26.77%)	671 (62.59%)	235 (21.92%)	428 (39.93%)	304 (28.36%)
1	19	763	626	760	79	655	2,702	233	626	2,309	8.22	59.18	271 (35.52%)	277 (36.30%)	393 (51.51%)	202 (26.47%)	476 (62.39%)	177 (23.20%)	311 (40.76%)	219 (28.70%)
1	20	851	768	848	148	771	5,348	317	768	4,475	7.91	53.05	365 (42.89%)	292 (34.31%)	440 (51.70%)	271 (31.84%)	522 (61.34%)	188 (22.09%)	365 (42.89%)	193 (22.68%)
1	Total	185,579	71,088	153,464	1,930	31,205	56,997	26,931	71,088	142,867	5.13	50.31	20,081 (10.82%)	12,154 (6.55%)	45,653 (24.60%)	16,777 (9.04%)	56,169 (30.27%)	24,325 (13.11%)	28,587 (15.40%)	11,013 (5.93%)

TABLE A.36: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Spells* model (*Sample-2*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-2;
Submodel: Cond_Prior-Spells

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	3,194	243	0	0	0	0	121	243	315	3.12	31.75	93 (2.91%)	11 (0.34%)	137 (4.29%)	29 (0.91%)	169 (5.29%)	38 (1.19%)	26 (0.81%)	1 (0.03%)
0	2	10,416	1,403	0	0	0	0	678	1,403	2,023	2.27	33.07	282 (2.71%)	43 (0.41%)	551 (5.29%)	173 (1.66%)	661 (6.35%)	140 (1.34%)	164 (1.57%)	16 (0.15%)
0	3	8,224	1,281	0	0	0	0	684	1,281	1,896	3.64	42.25	429 (5.22%)	87 (1.06%)	914 (11.11%)	335 (4.07%)	1,073 (13.05%)	182 (2.21%)	350 (4.26%)	35 (0.43%)
0	4	8,742	1,656	0	0	0	0	700	1,656	2,684	3.66	39.73	358 (4.10%)	117 (1.34%)	860 (9.84%)	313 (3.58%)	1,014 (11.60%)	216 (2.47%)	422 (4.83%)	54 (0.62%)
0	5	11,618	2,279	0	0	0	0	1,052	2,279	3,493	3.06	37.29	513 (4.42%)	175 (1.51%)	1,185 (10.20%)	439 (3.78%)	1,341 (11.54%)	206 (1.77%)	470 (4.05%)	89 (0.77%)
0	6	8,366	1,836	0	0	0	0	820	1,836	3,058	3.90	45.24	508 (6.07%)	205 (2.45%)	1,167 (13.95%)	480 (5.74%)	1,327 (15.86%)	195 (2.33%)	500 (5.98%)	107 (1.28%)
0	7	5,727	1,348	0	0	0	0	558	1,348	2,219	5.19	47.29	273 (4.77%)	170 (2.97%)	795 (13.88%)	293 (5.12%)	972 (16.97%)	187 (3.27%)	400 (6.98%)	99 (1.73%)
0	8	4,646	1,138	0	0	0	0	520	1,138	1,931	5.03	44.86	212 (4.56%)	123 (2.65%)	588 (12.66%)	221 (4.76%)	753 (16.21%)	126 (2.71%)	296 (6.37%)	125 (2.69%)
0	9	2,833	899	0	0	0	0	449	899	1,525	6.57	46.21	137 (4.84%)	102 (3.60%)	366 (12.92%)	146 (5.15%)	510 (18.00%)	88 (3.11%)	222 (7.84%)	96 (3.39%)
0	10	1,547	666	0	0	0	0	337	666	1,172	7.73	48.97	58 (3.75%)	62 (4.01%)	240 (15.51%)	99 (6.40%)	340 (21.98%)	73 (4.72%)	141 (9.11%)	72 (4.65%)
0	11	1,669	1,014	0	0	0	0	475	1,014	2,074	4.81	39.26	59 (3.54%)	36 (2.16%)	175 (10.49%)	68 (4.07%)	244 (14.62%)	45 (2.70%)	87 (5.21%)	53 (3.18%)
0	12	2,215	1,628	0	0	0	0	809	1,628	3,275	2.75	33.81	40 (1.81%)	27 (1.22%)	102 (4.60%)	48 (2.17%)	211 (9.53%)	57 (2.57%)	51 (2.30%)	52 (2.35%)
0	13	2,677	2,234	0	0	0	0	1,263	2,234	4,472	1.42	30.24	43 (1.61%)	19 (0.71%)	50 (1.87%)	26 (0.97%)	191 (7.13%)	35 (1.31%)	29 (1.08%)	31 (1.16%)
0	14	1,318	1,108	0	0	0	0	646	1,108	2,279	2.00	33.04	35 (2.66%)	10 (0.76%)	42 (3.19%)	21 (1.59%)	127 (9.64%)	25 (1.90%)	19 (1.44%)	21 (1.59%)
0	15	235	184	0	0	0	0	95	184	334	4.85	39.54	14 (5.96%)	7 (2.98%)	19 (8.09%)	7 (2.98%)	45 (19.15%)	14 (5.96%)	9 (3.83%)	10 (4.26%)
0	16	58	31	0	0	0	0	16	31	52	23.90	56.98	7 (12.07%)	2 (3.45%)	9 (15.52%)	2 (3.45%)	15 (25.86%)	10 (17.24%)	4 (6.90%)	5 (8.62%)
0	17	26	12	0	0	0	0	0	12	17	25.69	73.46	1 (3.85%)	3 (11.54%)	13 (50.00%)	7 (26.92%)	16 (61.54%)	4 (15.38%)	8 (30.77%)	8 (30.77%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	8	2	0	0	0	0	0	2	3	42.25	69.50	0 (0.00%)	0 (0.00%)	2 (25.00%)	0 (0.00%)	3 (37.50%)	2 (25.00%)	1 (12.50%)	1 (12.50%)
0	19	8	6	0	0	0	0	3	6	9	69.00	70.75	0 (0.00%)	0 (0.00%)	4 (50.00%)	1 (12.50%)	4 (50.00%)	5 (62.50%)	2 (25.00%)	1 (12.50%)
0	20	9	1	0	0	0	0	0	1	2	51.11	66.50	0 (0.00%)	0 (0.00%)	4 (44.44%)	1 (11.11%)	6 (66.67%)	2 (22.22%)	1 (11.11%)	1 (11.11%)
0	Total	73,536	18,969	0	0	0	0	9,226	18,969	32,833	3.71	39.68	3,062 (4.16%)	1,199 (1.63%)	7,223 (9.82%)	2,709 (3.68%)	9,022 (12.27%)	1,650 (2.24%)	3,202 (4.35%)	877 (1.19%)
1	1	89	11	89	0	3	3	6	11	17	0.79	25.47	2 (2.25%)	0 (0.00%)	4 (4.49%)	0 (0.00%)	5 (5.62%)	5 (5.62%)	1 (1.12%)	0 (0.00%)
1	2	1,112	124	1,112	4	49	58	53	124	169	1.65	34.52	69 (6.21%)	3 (0.27%)	89 (8.00%)	21 (1.89%)	100 (8.99%)	64 (5.76%)	25 (2.25%)	0 (0.00%)
1	3	6,396	1,094	6,396	20	302	351	420	1,094	1,574	1.74	34.13	422 (6.60%)	26 (0.41%)	377 (5.89%)	103 (1.61%)	471 (7.36%)	449 (7.02%)	158 (2.47%)	19 (0.30%)
1	4	9,046	2,030	9,046	46	657	776	796	2,030	3,190	2.89	45.27	864 (9.55%)	110 (1.22%)	1,355 (14.98%)	390 (4.31%)	1,655 (18.30%)	986 (10.90%)	623 (6.89%)	34 (0.38%)
1	5	8,864	2,607	8,864	58	904	1,111	1,014	2,607	4,339	5.60	53.32	883 (9.96%)	268 (3.02%)	2,272 (25.63%)	664 (7.49%)	2,684 (30.28%)	1,308 (14.76%)	1,302 (14.69%)	132 (1.49%)
1	6	13,972	4,264	13,972	110	1,292	1,652	1,537	4,264	7,419	4.45	46.27	1,234 (8.83%)	456 (3.26%)	2,840 (20.33%)	913 (6.53%)	3,292 (23.56%)	1,776 (12.71%)	1,650 (11.81%)	301 (2.15%)
1	7	23,970	6,244	23,970	167	2,241	2,929	2,298	6,244	10,588	3.93	43.81	2,113 (8.82%)	603 (2.52%)	3,805 (15.87%)	1,326 (5.53%)	4,703 (19.62%)	2,455 (10.24%)	2,061 (8.60%)	496 (2.07%)
1	8	22,586	6,612	22,586	204	3,076	4,001	2,287	6,612	11,792	4.79	48.34	2,579 (11.42%)	861 (3.81%)	4,704 (20.83%)	1,723 (7.63%)	5,914 (26.18%)	2,598 (11.50%)	2,623 (11.61%)	677 (3.00%)
1	9	16,768	6,123	16,768	161	3,172	4,238	2,055	6,123	11,373	6.22	55.87	2,366 (14.11%)	1,115 (6.65%)	5,073 (30.25%)	1,934 (11.53%)	6,103 (36.40%)	2,324 (13.86%)	3,148 (18.77%)	870 (5.19%)
1	10	12,311	5,470	12,311	142	2,937	4,080	1,684	5,470	10,373	7.67	62.88	1,907 (15.49%)	1,342 (10.90%)	4,895 (39.76%)	1,796 (14.59%)	5,666 (46.02%)	2,090 (16.98%)	3,173 (25.77%)	1,060 (8.61%)
1	11	12,209	7,017	12,209	170	3,082	4,632	2,628	7,017	13,953	6.71	60.54	1,606 (13.15%)	1,433 (11.74%)	4,634 (37.96%)	1,852 (15.17%)	5,406 (44.28%)	2,055 (16.83%)	3,086 (25.28%)	1,225 (10.03%)
1	12	13,671	9,382	13,671	224	3,421	5,333	4,021	9,382	18,752	5.37	53.15	1,559 (11.40%)	1,327 (9.71%)	4,068 (29.76%)	1,563 (11.43%)	5,020 (36.72%)	1,999 (14.62%)	2,697 (19.73%)	1,400 (10.24%)
1	13	10,813	7,803	10,813	199	3,244	5,285	3,296	7,803	16,814	5.15	54.17	1,539 (14.23%)	1,314 (12.15%)	3,343 (30.92%)	1,376 (12.73%)	4,374 (40.45%)	1,672 (15.46%)	2,346 (21.70%)	1,314 (12.15%)
1	14	6,949	5,019	6,949	141	2,990	5,198	1,928	5,019	11,614	6.18	57.62	1,218 (17.53%)	1,068 (15.37%)	2,573 (37.03%)	1,087 (15.64%)	3,311 (47.65%)	1,171 (16.85%)	1,781 (25.63%)	1,130 (16.26%)
1	15	4,272	3,135	4,272	129	2,432	4,775	1,152	3,135	7,646	7.42	61.24	791 (18.52%)	860 (20.13%)	1,877 (43.94%)	805 (18.84%)	2,316 (54.21%)	847 (19.83%)	1,315 (30.78%)	919 (21.51%)
1	16	2,611	1,950	2,611	83	1,784	4,117	683	1,950	5,214	7.42	63.22	567 (21.72%)	656 (25.12%)	1,295 (49.60%)	574 (21.98%)	1,559 (59.71%)	573 (21.95%)	964 (36.92%)	666 (25.51%)
1	17	1,678	1,280	1,678	73	1,262	3,478	446	1,280	3,721	8.26	62.57	447 (26.64%)	470 (28.01%)	849 (50.60%)	413 (24.61%)	1,032 (61.50%)	338 (20.14%)	648 (38.62%)	434 (25.86%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	1,136	908	1,136	59	911	2,982	295	908	2,877	9.28	62.83	336 (29.58%)	378 (33.27%)	595 (52.38%)	288 (25.35%)	712 (62.68%)	276 (24.30%)	462 (40.67%)	310 (27.29%)
1	19	811	668	811	53	672	2,739	232	668	2,406	9.03	61.66	295 (36.37%)	311 (38.35%)	439 (54.13%)	230 (28.36%)	510 (62.89%)	193 (23.80%)	368 (45.38%)	240 (29.59%)
1	20	912	807	912	117	821	5,405	315	807	4,550	7.55	54.56	402 (44.08%)	324 (35.53%)	491 (53.84%)	296 (32.46%)	573 (62.83%)	193 (21.16%)	393 (43.09%)	217 (23.79%)
1	Total	170,176	72,548	170,176	2,160	35,252	63,143	27,146	72,548	148,381	5.27	51.68	21,199 (12.46%)	12,925 (7.60%)	45,578 (26.78%)	17,354 (10.20%)	55,406 (32.56%)	23,372 (13.73%)	28,824 (16.94%)	11,444 (6.72%)

TABLE A.37: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Main* model (*Sample-3*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-3;
Submodel: Cond_Main

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	1	5,880	420	283	2	49	56	196	420	489	20.48	23.58	157 (2.67%)	5 (0.09%)	72 (1.22%)	35 (0.60%)	78 (1.33%)	45 (0.77%)	23 (0.39%)	2 (0.03%)
NA	2	16,091	2,106	2,066	20	256	297	870	2,106	3,035	3.66	31.09	445 (2.77%)	37 (0.23%)	539 (3.35%)	204 (1.27%)	639 (3.97%)	291 (1.81%)	334 (2.08%)	17 (0.11%)
NA	3	21,396	3,638	4,837	31	642	751	1,412	3,638	5,607	4.22	35.13	908 (4.24%)	120 (0.56%)	1,199 (5.60%)	515 (2.41%)	1,373 (6.42%)	668 (3.12%)	864 (4.04%)	76 (0.36%)
NA	4	22,121	4,174	5,854	75	910	1,149	1,561	4,174	6,471	4.91	43.44	1,173 (5.30%)	270 (1.22%)	1,965 (8.88%)	821 (3.71%)	2,358 (10.66%)	1,006 (4.55%)	1,393 (6.30%)	182 (0.82%)
NA	5	30,950	6,816	7,058	121	1,303	1,711	2,494	6,816	10,931	4.83	36.97	1,457 (4.71%)	440 (1.42%)	2,437 (7.87%)	1,100 (3.55%)	3,007 (9.72%)	1,233 (3.98%)	1,786 (5.77%)	280 (0.90%)
NA	6	33,453	7,595	9,652	124	1,804	2,346	2,703	7,595	12,384	5.13	41.29	1,902 (5.69%)	616 (1.84%)	2,766 (8.27%)	1,291 (3.86%)	3,587 (10.72%)	1,336 (3.99%)	1,973 (5.90%)	434 (1.30%)
NA	7	33,729	8,198	15,842	234	2,778	3,661	2,825	8,198	13,625	5.67	45.46	2,017 (5.98%)	778 (2.31%)	3,299 (9.78%)	1,565 (4.64%)	4,404 (13.06%)	1,454 (4.31%)	2,556 (7.58%)	662 (1.96%)
NA	8	28,978	8,666	17,220	175	3,192	4,210	2,944	8,666	14,949	7.08	50.08	2,018 (6.96%)	983 (3.39%)	3,689 (12.73%)	1,689 (5.83%)	5,091 (17.57%)	1,581 (5.46%)	3,084 (10.64%)	786 (2.71%)
NA	9	23,701	9,757	14,217	191	3,120	4,209	3,600	9,757	17,412	8.97	57.66	1,844 (7.78%)	1,284 (5.42%)	4,184 (17.65%)	1,849 (7.80%)	5,544 (23.39%)	1,693 (7.14%)	3,528 (14.89%)	936 (3.95%)
NA	10	17,591	8,455	12,186	189	3,242	4,568	2,931	8,455	15,174	10.00	64.27	1,520 (8.64%)	1,477 (8.40%)	4,088 (23.24%)	1,901 (10.81%)	5,211 (29.62%)	1,660 (9.44%)	3,649 (20.74%)	1,154 (6.56%)
NA	11	16,114	9,305	10,515	195	3,162	4,624	3,206	9,305	17,975	8.57	59.40	1,224 (7.60%)	1,442 (8.95%)	3,547 (22.01%)	1,773 (11.00%)	4,584 (28.45%)	1,473 (9.14%)	3,133 (19.44%)	1,409 (8.74%)
NA	12	17,333	11,932	9,495	191	3,212	4,867	5,083	11,932	24,280	6.18	51.12	1,136 (6.55%)	1,364 (7.87%)	2,847 (16.43%)	1,498 (8.64%)	3,985 (22.99%)	1,369 (7.90%)	2,750 (15.87%)	1,506 (8.69%)
NA	13	13,454	9,674	10,980	284	3,908	6,061	4,095	9,674	19,961	6.24	51.20	929 (6.91%)	1,250 (9.29%)	2,215 (16.46%)	1,187 (8.82%)	3,280 (24.38%)	1,116 (8.29%)	2,307 (17.15%)	1,477 (10.98%)
NA	14	10,101	7,633	9,702	198	3,449	5,620	3,397	7,633	16,097	5.76	50.13	868 (8.59%)	1,061 (10.50%)	1,676 (16.59%)	914 (9.05%)	2,612 (25.86%)	958 (9.48%)	1,763 (17.45%)	1,243 (12.31%)
NA	15	5,792	4,225	5,550	150	2,880	5,206	1,638	4,225	9,742	6.87	56.29	673 (11.62%)	885 (15.28%)	1,306 (22.55%)	746 (12.88%)	1,879 (32.44%)	645 (11.14%)	1,346 (23.24%)	1,007 (17.39%)
NA	16	3,193	2,372	3,121	89	2,058	4,325	820	2,372	6,246	8.69	60.77	510 (15.97%)	609 (19.07%)	934 (29.25%)	557 (17.44%)	1,269 (39.74%)	457 (14.31%)	960 (30.07%)	704 (22.05%)
NA	17	1,870	1,460	1,834	67	1,340	3,399	473	1,460	4,169	9.44	62.37	337 (18.02%)	455 (24.33%)	575 (30.75%)	361 (19.30%)	781 (41.76%)	289 (15.45%)	620 (33.16%)	462 (24.71%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	18	1,285	1,013	1,266	47	1,039	3,095	297	1,013	3,191	10.35	63.87	291 (22.65%)	420 (32.68%)	455 (35.41%)	307 (23.89%)	582 (45.29%)	216 (16.81%)	537 (41.79%)	380 (29.57%)
NA	19	844	715	841	56	736	2,784	236	715	2,608	10.84	61.80	204 (24.17%)	281 (33.29%)	331 (39.22%)	227 (26.90%)	420 (49.76%)	135 (16.00%)	341 (40.40%)	242 (28.67%)
NA	20	1,012	891	994	128	927	5,689	329	891	4,632	12.65	57.36	310 (30.63%)	394 (38.93%)	384 (37.94%)	272 (26.88%)	466 (46.05%)	140 (13.83%)	447 (44.17%)	247 (24.41%)
NA	Total	304,888	109,045	143,513	2,567	40,007	68,628	41,110	109,045	208,978	6.53	46.65	19,923 (6.53%)	14,171 (4.65%)	38,508 (12.63%)	18,812 (6.17%)	51,150 (16.78%)	17,765 (5.83%)	33,394 (10.95%)	13,206 (4.33%)

TABLE A.38: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Age-65p* model (*Sample-3*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-3;
Submodel: Cond_Age-65p

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	12,721	1,348	1,307	13	178	199	607	1,348	1,891	7.93	28.94	355 (2.79%)	23 (0.18%)	291 (2.29%)	115 (0.90%)	361 (2.84%)	237 (1.86%)	185 (1.45%)	9 (0.07%)
0	2	18,719	3,292	5,242	32	674	800	1,299	3,292	5,165	4.50	33.32	751 (4.01%)	88 (0.47%)	728 (3.89%)	371 (1.98%)	892 (4.77%)	619 (3.31%)	638 (3.41%)	39 (0.21%)
0	3	14,813	2,466	3,538	53	655	871	905	2,466	3,948	4.40	31.00	843 (5.69%)	116 (0.78%)	694 (4.69%)	377 (2.55%)	894 (6.04%)	481 (3.25%)	584 (3.94%)	52 (0.35%)
0	4	16,702	2,347	4,091	87	900	1,202	879	2,347	3,677	3.86	32.84	920 (5.51%)	152 (0.91%)	717 (4.29%)	464 (2.78%)	1,054 (6.31%)	415 (2.48%)	554 (3.32%)	64 (0.38%)
0	5	24,214	4,103	4,519	65	892	1,268	1,565	4,103	6,401	3.07	28.20	1,221 (5.04%)	158 (0.65%)	695 (2.87%)	450 (1.86%)	1,128 (4.66%)	440 (1.82%)	484 (2.00%)	81 (0.33%)
0	6	31,589	5,686	10,252	154	1,706	2,285	2,151	5,686	9,296	3.05	31.57	1,578 (5.00%)	265 (0.84%)	1,078 (3.41%)	674 (2.13%)	1,766 (5.59%)	732 (2.32%)	831 (2.63%)	90 (0.28%)
0	7	24,512	5,308	13,628	160	2,356	3,038	1,934	5,308	8,763	4.89	34.76	1,673 (6.83%)	297 (1.21%)	1,148 (4.68%)	762 (3.11%)	1,998 (8.15%)	730 (2.98%)	1,020 (4.16%)	107 (0.44%)
0	8	13,459	3,784	9,200	104	2,036	2,711	1,226	3,784	6,756	6.50	38.40	1,439 (10.69%)	278 (2.07%)	1,029 (7.65%)	813 (6.04%)	1,617 (12.01%)	510 (3.79%)	935 (6.95%)	94 (0.70%)
0	9	7,620	2,875	6,007	75	1,669	2,307	973	2,875	5,369	7.35	40.82	1,026 (13.46%)	283 (3.71%)	867 (11.38%)	656 (8.61%)	1,180 (15.49%)	418 (5.49%)	802 (10.52%)	118 (1.55%)
0	10	6,137	3,016	4,110	70	1,390	2,147	1,188	3,016	5,783	5.68	39.24	740 (12.06%)	240 (3.91%)	651 (10.61%)	489 (7.97%)	913 (14.88%)	333 (5.43%)	663 (10.80%)	119 (1.94%)
0	11	5,210	3,308	3,151	63	1,269	2,111	1,526	3,308	6,371	4.93	37.51	504 (9.67%)	200 (3.84%)	520 (9.98%)	400 (7.68%)	833 (15.99%)	310 (5.95%)	505 (9.69%)	164 (3.15%)
0	12	4,533	3,068	2,725	80	1,187	2,020	1,311	3,068	6,261	4.35	36.18	403 (8.89%)	193 (4.26%)	416 (9.18%)	322 (7.10%)	674 (14.87%)	261 (5.76%)	381 (8.41%)	127 (2.80%)
0	13	8,462	6,365	2,815	79	1,057	1,904	2,908	6,365	13,543	2.00	31.82	311 (3.68%)	136 (1.61%)	341 (4.03%)	213 (2.52%)	711 (8.40%)	348 (4.11%)	305 (3.60%)	153 (1.81%)
0	14	10,820	9,055	5,574	203	1,742	3,002	4,604	9,055	19,421	1.62	30.49	348 (3.22%)	145 (1.34%)	262 (2.42%)	194 (1.79%)	873 (8.07%)	377 (3.48%)	238 (2.20%)	130 (1.20%)
0	15	6,883	5,845	6,426	148	1,889	3,198	3,036	5,845	12,861	1.73	31.17	418 (6.07%)	124 (1.80%)	192 (2.79%)	153 (2.22%)	751 (10.91%)	218 (3.17%)	198 (2.88%)	92 (1.34%)
0	16	3,344	2,836	3,293	84	1,414	2,617	1,381	2,836	6,986	2.79	32.43	371 (11.09%)	94 (2.81%)	172 (5.14%)	141 (4.22%)	525 (15.70%)	181 (5.41%)	159 (4.75%)	77 (2.30%)
0	17	1,518	1,278	1,507	42	939	2,145	561	1,278	3,915	3.41	34.02	272 (17.92%)	77 (5.07%)	143 (9.42%)	118 (7.77%)	290 (19.10%)	108 (7.11%)	112 (7.38%)	61 (4.02%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	825	683	817	37	599	1,744	259	683	2,283	4.72	38.22	173 (20.97%)	86 (10.42%)	129 (15.64%)	94 (11.39%)	214 (25.94%)	76 (9.21%)	117 (14.18%)	58 (7.03%)
0	19	619	515	616	28	514	1,824	184	515	1,970	5.54	39.85	140 (22.62%)	83 (13.41%)	122 (19.71%)	101 (16.32%)	185 (29.89%)	65 (10.50%)	101 (16.32%)	57 (9.21%)
0	20	819	712	802	93	723	4,321	272	712	3,723	12.59	41.49	238 (29.06%)	143 (17.46%)	189 (23.08%)	173 (21.12%)	253 (30.89%)	85 (10.38%)	195 (23.81%)	100 (12.21%)
0	Total	213,519	67,890	89,620	1,670	23,789	41,714	28,769	67,890	134,383	4.25	32.90	13,724 (6.43%)	3,181 (1.49%)	10,384 (4.86%)	7,080 (3.32%)	17,112 (8.01%)	6,944 (3.25%)	9,007 (4.22%)	1,792 (0.84%)
1	1	39	2	9	0	3	5	0	2	2	523.23	88.74	1 (2.56%)	0 (0.00%)	3 (7.69%)	1 (2.56%)	4 (10.26%)	0 (0.00%)	2 (5.13%)	1 (2.56%)
1	2	269	7	8	0	2	2	5	7	14	16.17	92.62	7 (2.60%)	0 (0.00%)	3 (1.12%)	3 (1.12%)	7 (2.60%)	2 (0.74%)	0 (0.00%)	0 (0.00%)
1	3	608	32	18	0	4	4	14	32	45	8.44	84.35	9 (1.48%)	1 (0.16%)	21 (3.45%)	9 (1.48%)	25 (4.11%)	15 (2.47%)	9 (1.48%)	1 (0.16%)
1	4	2,327	342	79	0	12	14	131	342	479	4.42	75.34	37 (1.59%)	29 (1.25%)	431 (18.52%)	99 (4.25%)	462 (19.85%)	54 (2.32%)	207 (8.90%)	8 (0.34%)
1	5	3,219	690	486	5	26	32	253	690	987	7.93	76.97	118 (3.67%)	124 (3.85%)	680 (21.12%)	197 (6.12%)	763 (23.70%)	171 (5.31%)	455 (14.13%)	66 (2.05%)
1	6	6,393	1,770	1,164	14	122	147	587	1,770	2,657	9.68	77.23	154 (2.41%)	195 (3.05%)	1,463 (22.88%)	396 (6.19%)	1,662 (26.00%)	420 (6.57%)	803 (12.56%)	156 (2.44%)
1	7	9,056	3,044	2,254	19	267	335	1,049	3,044	4,767	7.96	77.67	316 (3.49%)	468 (5.17%)	2,257 (24.92%)	770 (8.50%)	2,646 (29.22%)	574 (6.34%)	1,474 (16.28%)	362 (4.00%)
1	8	11,095	4,160	5,084	47	562	706	1,430	4,160	6,713	11.30	78.19	465 (4.19%)	587 (5.29%)	2,992 (26.97%)	875 (7.89%)	3,424 (30.86%)	1,120 (10.09%)	2,074 (18.69%)	534 (4.81%)
1	9	14,497	6,288	7,975	71	1,052	1,284	1,988	6,288	10,567	13.68	78.77	663 (4.57%)	846 (5.84%)	4,299 (29.65%)	1,338 (9.23%)	5,139 (35.45%)	1,803 (12.44%)	3,321 (22.91%)	818 (5.64%)
1	10	13,666	6,576	8,918	130	1,733	2,130	1,971	6,576	11,293	13.14	79.73	911 (6.67%)	1,404 (10.27%)	4,385 (32.09%)	1,820 (13.32%)	5,354 (39.18%)	1,821 (13.33%)	3,763 (27.54%)	1,410 (10.32%)
1	11	10,225	5,442	8,564	153	2,404	3,165	1,537	5,442	9,744	13.29	80.51	819 (8.01%)	1,467 (14.35%)	3,497 (34.20%)	1,621 (15.85%)	4,360 (42.64%)	1,558 (15.24%)	3,324 (32.51%)	1,685 (16.48%)
1	12	7,248	4,077	6,717	125	2,375	3,258	1,081	4,077	7,387	13.30	80.88	686 (9.46%)	1,410 (19.45%)	2,671 (36.85%)	1,393 (19.22%)	3,421 (47.20%)	1,165 (16.07%)	2,746 (37.89%)	1,654 (22.82%)
1	13	4,937	3,087	4,847	91	2,216	3,167	787	3,087	6,043	12.58	81.10	588 (11.91%)	1,283 (25.99%)	1,879 (38.06%)	1,046 (21.19%)	2,403 (48.67%)	782 (15.84%)	2,050 (41.52%)	1,548 (31.36%)
1	14	3,057	2,035	3,039	64	1,766	2,890	512	2,035	4,072	12.06	81.14	413 (13.51%)	951 (31.11%)	1,299 (42.49%)	696 (22.77%)	1,640 (53.65%)	512 (16.75%)	1,425 (46.61%)	1,100 (35.98%)
1	15	1,819	1,261	1,817	43	1,226	2,282	329	1,261	2,895	11.64	80.70	317 (17.43%)	698 (38.37%)	808 (44.42%)	503 (27.65%)	987 (54.26%)	302 (16.60%)	934 (51.35%)	735 (40.41%)
1	16	1,099	829	1,099	29	836	1,820	230	829	2,094	10.96	80.27	213 (19.38%)	496 (45.13%)	511 (46.50%)	345 (31.39%)	621 (56.51%)	176 (16.01%)	629 (57.23%)	494 (44.95%)
1	17	709	558	709	20	592	1,505	146	558	1,470	10.14	79.34	178 (25.11%)	359 (50.63%)	340 (47.95%)	228 (32.16%)	416 (58.67%)	135 (19.04%)	434 (61.21%)	312 (44.01%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	482	403	482	25	434	1,328	110	403	1,180	9.65	78.54	112 (23.24%)	257 (53.32%)	241 (50.00%)	166 (34.44%)	301 (62.45%)	88 (18.26%)	311 (64.52%)	238 (49.38%)
1	19	316	268	316	19	292	1,114	74	268	885	9.47	78.69	87 (27.53%)	197 (62.34%)	176 (55.70%)	123 (38.92%)	208 (65.82%)	51 (16.14%)	212 (67.09%)	153 (48.42%)
1	20	308	284	308	42	294	1,726	107	284	1,301	7.63	76.74	105 (34.09%)	218 (70.78%)	168 (54.55%)	103 (33.44%)	195 (63.31%)	72 (23.38%)	214 (69.48%)	139 (45.13%)
1	Total	91,369	41,155	53,893	897	16,218	26,914	12,341	41,155	74,595	11.87	79.15	6,199 (6.78%)	10,990 (12.03%)	28,124 (30.78%)	11,732 (12.84%)	34,038 (37.25%)	10,821 (11.84%)	24,387 (26.69%)	11,414 (12.49%)

TABLE A.39: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Acute-12-month* model (*Sample-3*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-3;
Submodel: Cond_Prior-Acute-12-month

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	5,901	404	181	3	17	40	194	404	477	12.95	23.67	162 (2.75%)	4 (0.07%)	70 (1.19%)	32 (0.54%)	76 (1.29%)	45 (0.76%)	24 (0.41%)	1 (0.02%)
0	2	14,857	1,908	1,192	13	152	170	804	1,908	2,717	3.09	30.94	390 (2.63%)	35 (0.24%)	489 (3.29%)	185 (1.25%)	570 (3.84%)	285 (1.92%)	291 (1.96%)	16 (0.11%)
0	3	20,927	3,237	2,632	21	398	461	1,298	3,237	4,780	3.53	34.52	887 (4.24%)	105 (0.50%)	1,062 (5.07%)	470 (2.25%)	1,197 (5.72%)	594 (2.84%)	726 (3.47%)	61 (0.29%)
0	4	18,521	3,371	3,115	46	581	729	1,323	3,371	5,240	4.99	45.61	864 (4.66%)	226 (1.22%)	1,737 (9.38%)	690 (3.73%)	2,012 (10.86%)	871 (4.70%)	1,148 (6.20%)	134 (0.72%)
0	5	28,328	5,989	3,669	75	763	983	2,280	5,989	9,430	4.42	35.36	1,309 (4.62%)	371 (1.31%)	1,953 (6.89%)	905 (3.19%)	2,380 (8.40%)	985 (3.48%)	1,411 (4.98%)	204 (0.72%)
0	6	26,390	6,131	5,323	99	1,084	1,394	2,221	6,131	10,006	5.35	43.97	1,478 (5.60%)	459 (1.74%)	2,196 (8.32%)	1,035 (3.92%)	2,752 (10.43%)	1,077 (4.08%)	1,505 (5.70%)	308 (1.17%)
0	7	25,124	6,128	6,392	138	1,646	2,068	2,241	6,128	9,914	5.82	48.82	1,231 (4.90%)	621 (2.47%)	2,682 (10.68%)	1,170 (4.66%)	3,384 (13.47%)	1,120 (4.46%)	1,999 (7.96%)	407 (1.62%)
0	8	19,525	5,947	6,955	145	1,937	2,479	2,191	5,947	10,119	7.59	54.81	1,091 (5.59%)	723 (3.70%)	2,669 (13.67%)	1,193 (6.11%)	3,503 (17.94%)	1,205 (6.17%)	2,173 (11.13%)	517 (2.65%)
0	9	14,760	6,001	6,152	130	2,053	2,699	2,341	6,001	10,438	9.65	61.44	888 (6.02%)	852 (5.77%)	2,683 (18.18%)	1,183 (8.01%)	3,530 (23.92%)	1,125 (7.62%)	2,100 (14.23%)	572 (3.88%)
0	10	10,319	5,145	4,595	136	1,814	2,498	2,096	5,145	9,126	9.98	63.93	623 (6.04%)	752 (7.29%)	2,102 (20.37%)	998 (9.67%)	2,781 (26.95%)	967 (9.37%)	1,740 (16.86%)	667 (6.46%)
0	11	8,957	5,294	3,670	139	1,799	2,423	2,004	5,294	10,289	8.20	55.81	408 (4.56%)	642 (7.17%)	1,552 (17.33%)	810 (9.04%)	2,086 (23.29%)	807 (9.01%)	1,351 (15.08%)	718 (8.02%)
0	12	8,847	6,180	3,209	148	1,644	2,499	2,807	6,180	12,658	5.14	45.56	299 (3.38%)	438 (4.95%)	923 (10.43%)	539 (6.09%)	1,418 (16.03%)	686 (7.75%)	945 (10.68%)	629 (7.11%)
0	13	8,019	6,398	3,102	153	1,741	2,716	3,232	6,398	12,921	4.06	41.58	265 (3.30%)	332 (4.14%)	654 (8.16%)	372 (4.64%)	1,198 (14.94%)	510 (6.36%)	636 (7.93%)	507 (6.32%)
0	14	3,088	2,219	2,471	140	1,842	2,812	1,086	2,219	4,718	6.87	51.24	187 (6.06%)	248 (8.03%)	450 (14.57%)	271 (8.78%)	730 (23.64%)	344 (11.14%)	448 (14.51%)	330 (10.69%)
0	15	1,497	990	1,219	52	1,052	1,979	462	990	2,019	7.92	56.31	109 (7.28%)	158 (10.55%)	264 (17.64%)	139 (9.29%)	378 (25.25%)	194 (12.96%)	279 (18.64%)	217 (14.50%)
0	16	685	455	590	25	538	1,332	165	455	999	10.73	63.96	64 (9.34%)	109 (15.91%)	157 (22.92%)	104 (15.18%)	217 (31.68%)	128 (18.69%)	168 (24.53%)	154 (22.48%)
0	17	325	217	288	19	262	786	75	217	522	14.39	70.50	33 (10.15%)	65 (20.00%)	100 (30.77%)	62 (19.08%)	134 (41.23%)	84 (25.85%)	93 (28.62%)	80 (24.62%)

Continued on next page

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	212	146	180	10	159	549	35	146	333	19.24	73.71	21 (9.91%)	45 (21.23%)	67 (31.60%)	43 (20.28%)	92 (43.40%)	58 (27.36%)	62 (29.25%)	51 (24.06%)
0	19	109	72	96	7	87	377	19	72	227	38.12	73.08	9 (8.26%)	23 (21.10%)	36 (33.03%)	26 (23.85%)	43 (39.45%)	28 (25.69%)	37 (33.94%)	30 (27.52%)
0	20	57	28	42	2	32	216	10	28	116	81.74	59.46	7 (12.28%)	17 (29.82%)	20 (35.09%)	10 (17.54%)	25 (43.86%)	12 (21.05%)	17 (29.82%)	7 (12.28%)
0	Total	216,448	66,260	55,073	1,501	19,601	29,210	26,884	66,260	117,049	6.03	45.08	10,325 (4.77%)	6,225 (2.88%)	21,866 (10.10%)	10,237 (4.73%)	28,506 (13.17%)	11,125 (5.14%)	17,153 (7.92%)	5,610 (2.59%)
1	1	72	9	72	0	10	13	5	9	11	611.83	33.51	2 (2.78%)	0 (0.00%)	5 (6.94%)	1 (1.39%)	5 (6.94%)	0 (0.00%)	4 (5.56%)	0 (0.00%)
1	2	758	115	758	0	15	20	51	115	187	12.38	23.63	44 (5.80%)	2 (0.26%)	16 (2.11%)	4 (0.53%)	27 (3.56%)	13 (1.72%)	12 (1.58%)	0 (0.00%)
1	3	2,070	475	2,070	2	85	94	136	475	800	8.01	31.75	155 (7.49%)	13 (0.63%)	116 (5.60%)	67 (3.24%)	156 (7.54%)	66 (3.19%)	78 (3.77%)	6 (0.29%)
1	4	2,838	743	2,838	2	169	196	233	743	1,276	6.48	43.07	199 (7.01%)	55 (1.94%)	321 (11.31%)	123 (4.33%)	428 (15.08%)	148 (5.21%)	285 (10.04%)	50 (1.76%)
1	5	3,859	1,020	3,859	6	300	368	300	1,020	1,789	6.46	44.08	305 (7.90%)	104 (2.69%)	516 (13.37%)	208 (5.39%)	709 (18.37%)	277 (7.18%)	418 (10.83%)	90 (2.33%)
1	6	7,192	1,528	7,192	15	425	499	446	1,528	2,546	5.09	37.99	404 (5.62%)	148 (2.06%)	585 (8.13%)	268 (3.73%)	932 (12.96%)	349 (4.85%)	455 (6.33%)	171 (2.38%)
1	7	8,388	2,005	8,388	7	682	802	583	2,005	3,411	5.36	36.14	705 (8.40%)	172 (2.05%)	667 (7.95%)	341 (4.07%)	1,125 (13.41%)	379 (4.52%)	535 (6.38%)	203 (2.42%)
1	8	7,188	2,270	7,188	28	964	1,135	630	2,270	4,016	7.01	41.56	820 (11.41%)	250 (3.48%)	875 (12.17%)	465 (6.47%)	1,285 (17.88%)	422 (5.87%)	756 (10.52%)	273 (3.80%)
1	9	6,548	2,552	6,548	43	1,060	1,339	714	2,552	4,640	8.12	49.59	790 (12.06%)	337 (5.15%)	1,186 (18.11%)	533 (8.14%)	1,533 (23.41%)	434 (6.63%)	1,168 (17.84%)	304 (4.64%)
1	10	7,068	3,238	7,068	39	1,186	1,612	924	3,238	5,908	8.71	57.75	800 (11.32%)	494 (6.99%)	1,617 (22.88%)	722 (10.22%)	2,026 (28.66%)	509 (7.20%)	1,568 (22.18%)	431 (6.10%)
1	11	7,446	3,847	7,446	45	1,340	1,869	1,115	3,847	7,347	9.41	63.12	878 (11.79%)	768 (10.31%)	2,019 (27.12%)	935 (12.56%)	2,516 (33.79%)	665 (8.93%)	1,863 (25.02%)	613 (8.23%)
1	12	7,526	4,475	7,526	44	1,585	2,223	1,460	4,475	8,846	8.64	62.49	810 (10.76%)	940 (12.49%)	1,967 (26.14%)	978 (12.99%)	2,567 (34.11%)	650 (8.64%)	1,944 (25.83%)	845 (11.23%)
1	13	8,888	6,283	8,888	65	1,945	2,760	2,452	6,283	12,529	6.59	53.87	833 (9.37%)	955 (10.74%)	1,783 (20.06%)	919 (10.34%)	2,507 (28.21%)	708 (7.97%)	1,860 (20.93%)	1,007 (11.33%)
1	14	6,801	4,997	6,801	65	2,334	3,378	1,956	4,997	10,796	6.38	53.80	738 (10.85%)	918 (13.50%)	1,433 (21.07%)	815 (11.98%)	2,104 (30.94%)	612 (9.00%)	1,462 (21.50%)	957 (14.07%)
1	15	4,219	3,080	4,219	97	2,249	3,616	1,095	3,080	7,425	7.13	58.16	530 (12.56%)	762 (18.06%)	1,073 (25.43%)	626 (14.84%)	1,509 (35.77%)	473 (11.21%)	1,144 (27.12%)	808 (19.15%)
1	16	2,664	2,033	2,664	108	1,842	3,631	682	2,033	5,367	7.68	59.09	424 (15.92%)	539 (20.23%)	768 (28.83%)	466 (17.49%)	1,034 (38.81%)	287 (10.77%)	836 (31.38%)	584 (21.92%)
1	17	1,718	1,369	1,718	103	1,321	3,165	463	1,369	3,973	7.85	60.12	349 (20.31%)	435 (25.32%)	560 (32.60%)	312 (18.16%)	736 (42.84%)	212 (12.34%)	581 (33.82%)	418 (24.33%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	1,261	1,038	1,261	92	1,087	3,259	348	1,038	3,272	8.11	59.98	280 (22.20%)	369 (29.26%)	431 (34.18%)	305 (24.19%)	557 (44.17%)	169 (13.40%)	488 (38.70%)	322 (25.54%)
1	19	867	745	867	93	792	2,918	244	745	2,639	8.83	59.45	209 (24.11%)	292 (33.68%)	308 (35.52%)	204 (23.53%)	410 (47.29%)	135 (15.57%)	346 (39.91%)	250 (28.84%)
1	20	1,069	963	1,069	212	1,015	6,521	389	963	5,151	8.40	54.97	323 (30.22%)	393 (36.76%)	396 (37.04%)	283 (26.47%)	478 (44.71%)	132 (12.35%)	438 (40.97%)	264 (24.70%)
1	Total	88,440	42,785	88,440	1,066	20,406	39,418	14,226	42,785	91,929	7.78	50.47	9,598 (10.85%)	7,946 (8.98%)	16,642 (18.82%)	8,575 (9.70%)	22,644 (25.60%)	6,640 (7.51%)	16,241 (18.36%)	7,596 (8.59%)

TABLE A.40: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Oper-12-month* model (*Sample-3*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-3;
Submodel: Cond_Prior-Oper-12-month

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	71	8	32	1	12	13	1	8	8	1,013.75	45.54	2 (2.82%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (1.41%)	0 (0.00%)
0	2	5,582	470	19	0	7	12	234	470	579	4.11	14.45	149 (2.67%)	5 (0.09%)	18 (0.32%)	35 (0.63%)	23 (0.41%)	7 (0.13%)	6 (0.11%)	0 (0.00%)
0	3	11,746	1,295	173	1	9	10	525	1,295	1,739	2.49	35.41	627 (5.34%)	47 (0.40%)	527 (4.49%)	272 (2.32%)	586 (4.99%)	41 (0.35%)	272 (2.32%)	11 (0.09%)
0	4	16,703	2,727	739	5	153	162	1,164	2,727	3,923	2.68	27.89	839 (5.02%)	194 (1.16%)	767 (4.59%)	413 (2.47%)	882 (5.28%)	85 (0.51%)	566 (3.39%)	60 (0.36%)
0	5	18,123	3,911	892	8	183	214	1,457	3,911	6,304	3.86	39.25	1,080 (5.96%)	216 (1.19%)	1,147 (6.33%)	660 (3.64%)	1,373 (7.58%)	136 (0.75%)	720 (3.97%)	129 (0.71%)
0	6	14,935	3,974	2,094	18	362	408	1,472	3,974	6,524	5.71	53.19	1,108 (7.42%)	455 (3.05%)	1,951 (13.06%)	959 (6.42%)	2,319 (15.53%)	186 (1.25%)	1,316 (8.81%)	251 (1.68%)
0	7	10,752	3,607	2,838	23	649	758	1,228	3,607	6,183	9.10	55.61	906 (8.43%)	503 (4.68%)	1,628 (15.14%)	831 (7.73%)	1,971 (18.33%)	210 (1.95%)	1,399 (13.01%)	318 (2.96%)
0	8	9,235	3,686	3,023	34	819	1,017	1,219	3,686	6,425	12.56	61.54	767 (8.31%)	518 (5.61%)	1,688 (18.28%)	801 (8.67%)	2,143 (23.21%)	270 (2.92%)	1,439 (15.58%)	415 (4.49%)
0	9	7,617	3,499	3,033	39	858	1,027	1,103	3,499	6,097	14.06	68.63	632 (8.30%)	680 (8.93%)	1,672 (21.95%)	821 (10.78%)	2,175 (28.55%)	247 (3.24%)	1,447 (19.00%)	544 (7.14%)
0	10	5,659	2,837	3,104	56	969	1,205	850	2,837	5,171	13.08	68.07	453 (8.00%)	630 (11.13%)	1,294 (22.87%)	714 (12.62%)	1,733 (30.62%)	225 (3.98%)	1,159 (20.48%)	624 (11.03%)
0	11	4,908	2,723	3,084	64	1,126	1,471	903	2,723	5,039	10.35	62.80	398 (8.11%)	537 (10.94%)	1,008 (20.54%)	537 (10.94%)	1,386 (28.24%)	178 (3.63%)	1,049 (21.37%)	655 (13.35%)
0	12	2,955	1,628	2,474	50	1,035	1,433	458	1,628	3,055	12.22	70.74	296 (10.02%)	520 (17.60%)	810 (27.41%)	443 (14.99%)	1,090 (36.89%)	116 (3.93%)	795 (26.90%)	517 (17.50%)
0	13	2,587	1,616	1,944	43	972	1,393	515	1,616	3,269	9.85	63.42	225 (8.70%)	487 (18.82%)	560 (21.65%)	340 (13.14%)	770 (29.76%)	82 (3.17%)	666 (25.74%)	484 (18.71%)
0	14	5,052	3,657	1,239	30	776	1,187	1,442	3,657	7,749	3.27	39.13	162 (3.21%)	343 (6.79%)	355 (7.03%)	218 (4.32%)	557 (11.03%)	84 (1.66%)	414 (8.19%)	312 (6.18%)
0	15	1,914	1,204	757	15	493	830	456	1,204	2,651	5.55	48.12	143 (7.47%)	237 (12.38%)	219 (11.44%)	161 (8.41%)	339 (17.71%)	59 (3.08%)	259 (13.53%)	200 (10.45%)
0	16	1,265	863	892	22	478	875	332	863	2,177	4.61	44.71	98 (7.75%)	151 (11.94%)	141 (11.15%)	87 (6.88%)	200 (15.81%)	32 (2.53%)	195 (15.42%)	110 (8.70%)
0	17	2,914	2,424	830	41	543	1,026	1,372	2,424	4,892	1.56	32.98	103 (3.53%)	108 (3.71%)	88 (3.02%)	59 (2.02%)	257 (8.82%)	33 (1.13%)	106 (3.64%)	71 (2.44%)

Continued on next page

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	1,801	1,576	527	45	376	878	833	1,576	3,405	1.70	31.17	72 (4.00%)	98 (5.44%)	78 (4.33%)	47 (2.61%)	148 (8.22%)	20 (1.11%)	92 (5.11%)	61 (3.39%)
0	19	6,958	6,330	866	107	533	1,202	3,596	6,330	13,234	0.79	29.70	134 (1.93%)	63 (0.91%)	48 (0.69%)	41 (0.59%)	466 (6.70%)	14 (0.20%)	63 (0.91%)	36 (0.52%)
0	20	1,753	1,595	1,638	106	1,132	2,650	948	1,595	4,212	2.73	34.47	130 (7.42%)	101 (5.76%)	78 (4.45%)	70 (3.99%)	200 (11.41%)	21 (1.20%)	93 (5.31%)	46 (2.62%)
0	Total	132,530	49,630	30,198	708	11,485	17,771	20,108	49,630	92,636	6.66	45.21	8,324 (6.28%)	5,893 (4.45%)	14,077 (10.62%)	7,509 (5.67%)	18,618 (14.05%)	2,046 (1.54%)	12,057 (9.10%)	4,844 (3.66%)
1	1	857	48	43	0	12	18	24	48	55	35.56	14.03	6 (0.70%)	0 (0.00%)	7 (0.82%)	3 (0.35%)	11 (1.28%)	5 (0.58%)	3 (0.35%)	0 (0.00%)
1	2	10,701	910	803	12	138	148	404	910	1,141	1.99	26.59	222 (2.07%)	10 (0.09%)	175 (1.64%)	62 (0.58%)	203 (1.90%)	92 (0.86%)	44 (0.41%)	2 (0.02%)
1	3	18,875	2,358	3,676	41	692	803	965	2,358	3,431	2.39	32.58	518 (2.74%)	30 (0.16%)	517 (2.74%)	161 (0.85%)	771 (4.08%)	305 (1.62%)	276 (1.46%)	14 (0.07%)
1	4	21,853	3,517	6,775	93	1,166	1,409	1,346	3,517	5,353	3.49	38.16	645 (2.95%)	87 (0.40%)	1,057 (4.84%)	407 (1.86%)	1,586 (7.26%)	645 (2.95%)	666 (3.05%)	60 (0.27%)
1	5	17,922	3,662	10,259	152	1,740	2,266	1,392	3,662	5,976	5.44	42.72	740 (4.13%)	191 (1.07%)	1,414 (7.89%)	508 (2.83%)	2,082 (11.62%)	1,063 (5.93%)	1,116 (6.23%)	126 (0.70%)
1	6	14,619	3,880	9,834	113	1,486	1,982	1,366	3,880	6,482	7.65	49.36	756 (5.17%)	262 (1.79%)	1,725 (11.80%)	685 (4.69%)	2,431 (16.63%)	1,298 (8.88%)	1,317 (9.01%)	204 (1.40%)
1	7	12,134	3,976	9,220	100	1,598	2,128	1,345	3,976	6,809	8.78	52.33	793 (6.54%)	387 (3.19%)	1,861 (15.34%)	752 (6.20%)	2,471 (20.36%)	1,295 (10.67%)	1,433 (11.81%)	329 (2.71%)
1	8	11,752	4,351	10,402	110	1,634	2,297	1,454	4,351	7,600	8.33	52.48	896 (7.62%)	405 (3.45%)	1,972 (16.78%)	799 (6.80%)	2,491 (21.20%)	1,454 (12.37%)	1,662 (14.14%)	498 (4.24%)
1	9	11,999	5,307	11,325	132	1,897	2,689	1,862	5,307	9,739	7.39	54.40	1,059 (8.83%)	577 (4.81%)	2,344 (19.53%)	991 (8.26%)	2,965 (24.71%)	1,565 (13.04%)	1,940 (16.17%)	577 (4.81%)
1	10	11,807	6,099	11,481	113	2,177	3,125	2,259	6,099	11,549	7.65	56.19	1,060 (8.98%)	701 (5.94%)	2,418 (20.48%)	1,090 (9.23%)	3,111 (26.35%)	1,556 (13.18%)	2,155 (18.25%)	655 (5.55%)
1	11	10,851	6,212	10,682	148	2,510	3,654	2,345	6,212	11,893	7.73	57.28	990 (9.12%)	835 (7.70%)	2,480 (22.86%)	1,143 (10.53%)	3,211 (29.59%)	1,554 (14.32%)	2,180 (20.09%)	760 (7.00%)
1	12	8,458	5,033	8,390	131	2,410	3,713	1,728	5,033	10,090	8.28	60.05	906 (10.71%)	866 (10.24%)	2,093 (24.75%)	1,064 (12.58%)	2,749 (32.50%)	1,262 (14.92%)	2,001 (23.66%)	881 (10.42%)
1	13	6,379	3,953	6,342	120	2,361	3,804	1,288	3,953	8,307	8.91	62.43	741 (11.62%)	841 (13.18%)	1,815 (28.45%)	926 (14.52%)	2,394 (37.53%)	1,066 (16.71%)	1,781 (27.92%)	898 (14.08%)
1	14	4,652	3,060	4,626	106	2,194	3,826	991	3,060	6,675	9.31	64.84	598 (12.85%)	774 (16.64%)	1,364 (29.32%)	768 (16.51%)	1,805 (38.80%)	824 (17.71%)	1,393 (29.94%)	895 (19.24%)
1	15	3,383	2,280	3,368	116	1,889	3,647	683	2,280	5,494	10.04	65.23	436 (12.89%)	660 (19.51%)	1,084 (32.04%)	577 (17.06%)	1,456 (43.04%)	644 (19.04%)	1,095 (32.37%)	770 (22.76%)
1	16	2,233	1,603	2,225	88	1,483	3,293	481	1,603	4,080	8.88	65.38	367 (16.44%)	510 (22.84%)	755 (33.81%)	445 (19.93%)	996 (44.60%)	420 (18.81%)	818 (36.63%)	595 (26.65%)
1	17	1,512	1,148	1,503	69	1,089	2,854	355	1,148	3,321	9.74	64.43	277 (18.32%)	396 (26.19%)	529 (34.99%)	342 (22.62%)	696 (46.03%)	286 (18.92%)	541 (35.78%)	452 (29.89%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	981	802	979	47	808	2,553	261	802	2,569	9.28	60.78	217 (22.12%)	296 (30.17%)	341 (34.76%)	236 (24.06%)	477 (48.62%)	173 (17.64%)	374 (38.12%)	288 (29.36%)
1	19	674	578	673	63	595	2,400	197	578	2,197	8.83	57.95	158 (23.44%)	213 (31.60%)	246 (36.50%)	173 (25.67%)	326 (48.37%)	122 (18.10%)	271 (40.21%)	210 (31.16%)
1	20	716	638	709	105	643	4,248	256	638	3,581	8.83	51.87	214 (29.89%)	237 (33.10%)	234 (32.68%)	171 (23.88%)	300 (41.90%)	90 (12.57%)	271 (37.85%)	148 (20.67%)
1	Total	172,358	59,415	113,315	1,859	28,522	50,857	21,002	59,415	116,342	6.44	47.75	11,599 (6.73%)	8,278 (4.80%)	24,431 (14.17%)	11,303 (6.56%)	32,532 (18.87%)	15,719 (9.12%)	21,337 (12.38%)	8,362 (4.85%)

TABLE A.41: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Spells* model (*Sample-3*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample-3;
Submodel: Cond_Prior-Spells

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	6,504	471	0	0	0	0	219	471	550	9.07	25.85	175 (2.69%)	8 (0.12%)	114 (1.75%)	46 (0.71%)	123 (1.89%)	56 (0.86%)	43 (0.66%)	2 (0.03%)
0	2	14,548	1,936	0	0	0	0	825	1,936	2,804	3.18	31.53	361 (2.48%)	41 (0.28%)	491 (3.38%)	200 (1.37%)	571 (3.92%)	214 (1.47%)	333 (2.29%)	21 (0.14%)
0	3	18,205	2,742	0	0	0	0	1,161	2,742	4,076	3.61	34.19	784 (4.31%)	109 (0.60%)	892 (4.90%)	431 (2.37%)	1,053 (5.78%)	334 (1.83%)	641 (3.52%)	52 (0.29%)
0	4	16,078	2,670	0	0	0	0	1,048	2,670	3,979	4.80	43.08	707 (4.40%)	231 (1.44%)	1,253 (7.79%)	552 (3.43%)	1,468 (9.13%)	395 (2.46%)	865 (5.38%)	110 (0.68%)
0	5	22,856	4,567	0	0	0	0	1,793	4,567	7,176	4.36	33.43	1,037 (4.54%)	279 (1.22%)	1,274 (5.57%)	651 (2.85%)	1,607 (7.03%)	449 (1.96%)	914 (4.00%)	162 (0.71%)
0	6	20,731	4,621	0	0	0	0	1,688	4,621	7,402	5.11	43.01	999 (4.82%)	362 (1.75%)	1,485 (7.16%)	736 (3.55%)	1,890 (9.12%)	374 (1.80%)	997 (4.81%)	218 (1.05%)
0	7	18,313	4,273	0	0	0	0	1,595	4,273	6,918	5.69	49.80	748 (4.08%)	465 (2.54%)	1,807 (9.87%)	808 (4.41%)	2,317 (12.65%)	330 (1.80%)	1,324 (7.23%)	298 (1.63%)
0	8	12,110	3,579	0	0	0	0	1,408	3,579	6,084	7.87	55.16	569 (4.70%)	468 (3.86%)	1,513 (12.49%)	748 (6.18%)	2,031 (16.77%)	294 (2.43%)	1,282 (10.59%)	309 (2.55%)
0	9	8,322	3,468	0	0	0	0	1,388	3,468	5,767	10.54	61.42	414 (4.97%)	474 (5.70%)	1,304 (15.67%)	655 (7.87%)	1,830 (21.99%)	276 (3.32%)	1,097 (13.18%)	313 (3.76%)
0	10	5,973	3,251	0	0	0	0	1,485	3,251	5,656	10.19	59.47	249 (4.17%)	354 (5.93%)	925 (15.49%)	450 (7.53%)	1,378 (23.07%)	245 (4.10%)	756 (12.66%)	357 (5.98%)
0	11	4,275	2,525	0	0	0	0	1,018	2,525	4,783	9.41	54.34	130 (3.04%)	223 (5.22%)	531 (12.42%)	286 (6.69%)	839 (19.63%)	171 (4.00%)	484 (11.32%)	349 (8.16%)
0	12	5,111	3,531	0	0	0	0	1,516	3,531	7,085	4.75	40.54	92 (1.80%)	114 (2.23%)	296 (5.79%)	187 (3.66%)	535 (10.47%)	147 (2.88%)	269 (5.26%)	291 (5.69%)
0	13	5,825	4,955	0	0	0	0	2,593	4,955	10,223	2.16	33.01	64 (1.10%)	59 (1.01%)	136 (2.33%)	99 (1.70%)	471 (8.09%)	79 (1.36%)	155 (2.66%)	179 (3.07%)
0	14	1,919	1,514	0	0	0	0	847	1,514	2,931	4.44	40.59	57 (2.97%)	32 (1.67%)	87 (4.53%)	37 (1.93%)	264 (13.76%)	44 (2.29%)	78 (4.06%)	102 (5.32%)
0	15	319	120	0	0	0	0	66	120	220	11.16	63.74	10 (3.13%)	11 (3.45%)	31 (9.72%)	22 (6.90%)	66 (20.69%)	31 (9.72%)	37 (11.60%)	37 (11.60%)
0	16	165	85	0	0	0	0	44	85	143	12.33	54.65	3 (1.82%)	3 (1.82%)	14 (8.48%)	7 (4.24%)	28 (16.97%)	18 (10.91%)	12 (7.27%)	12 (7.27%)
0	17	41	18	0	0	0	0	9	18	30	23.27	72.00	0 (0.00%)	2 (4.88%)	5 (12.20%)	2 (4.88%)	12 (29.27%)	6 (14.63%)	7 (17.07%)	5 (12.20%)

Continued on next page

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	41	21	0	0	0	0	3	21	34	30.39	80.58	0 (0.00%)	0 (0.00%)	7 (17.07%)	2 (4.88%)	11 (26.83%)	9 (21.95%)	2 (4.88%)	3 (7.32%)
0	19	22	7	0	0	0	0	2	7	13	88.64	76.68	3 (13.64%)	0 (0.00%)	8 (36.36%)	3 (13.64%)	9 (40.91%)	5 (22.73%)	4 (18.18%)	6 (27.27%)
0	20	17	3	0	0	0	0	0	3	3	228.06	46.24	1 (5.88%)	0 (0.00%)	3 (17.65%)	1 (5.88%)	3 (17.65%)	1 (5.88%)	1 (5.88%)	0 (0.00%)
0	Total	161,375	44,357	0	0	0	0	18,708	44,357	75,877	5.58	42.03	6,403 (3.97%)	3,235 (2.00%)	12,176 (7.55%)	5,923 (3.67%)	16,506 (10.23%)	3,478 (2.16%)	9,301 (5.76%)	2,826 (1.75%)
1	1	144	20	144	2	32	39	7	20	26	451.62	31.45	5 (3.47%)	0 (0.00%)	8 (5.56%)	4 (2.78%)	10 (6.94%)	0 (0.00%)	3 (2.08%)	1 (0.69%)
1	2	1,806	244	1,806	16	210	232	105	244	364	8.32	26.18	89 (4.93%)	3 (0.17%)	60 (3.32%)	10 (0.55%)	82 (4.54%)	90 (4.98%)	26 (1.44%)	0 (0.00%)
1	3	4,420	912	4,420	23	560	664	298	912	1,481	6.07	36.43	287 (6.49%)	27 (0.61%)	286 (6.47%)	93 (2.10%)	351 (7.94%)	289 (6.54%)	176 (3.98%)	17 (0.38%)
1	4	5,030	1,385	5,030	62	701	846	440	1,385	2,314	5.92	48.08	319 (6.34%)	68 (1.35%)	699 (13.90%)	238 (4.73%)	805 (16.00%)	512 (10.18%)	491 (9.76%)	54 (1.07%)
1	5	7,034	1,895	7,034	126	1,248	1,608	638	1,895	3,210	6.31	47.23	482 (6.85%)	159 (2.26%)	989 (14.06%)	380 (5.40%)	1,223 (17.39%)	733 (10.42%)	786 (11.17%)	120 (1.71%)
1	6	10,591	2,823	10,591	132	1,787	2,285	881	2,823	4,850	6.13	42.11	705 (6.66%)	245 (2.31%)	1,229 (11.60%)	515 (4.86%)	1,593 (15.04%)	964 (9.10%)	896 (8.46%)	198 (1.87%)
1	7	17,231	4,143	17,231	242	2,985	3,887	1,344	4,143	6,996	4.81	38.77	1,251 (7.26%)	304 (1.76%)	1,497 (8.69%)	707 (4.10%)	2,223 (12.90%)	1,206 (7.00%)	1,207 (7.00%)	347 (2.01%)
1	8	15,518	4,570	15,518	175	3,049	4,024	1,398	4,570	7,954	6.05	45.68	1,487 (9.58%)	442 (2.85%)	2,044 (13.17%)	930 (5.99%)	2,751 (17.73%)	1,200 (7.73%)	1,657 (10.68%)	419 (2.70%)
1	9	13,174	5,062	13,174	171	2,903	4,006	1,568	5,062	9,222	8.20	55.98	1,386 (10.52%)	736 (5.59%)	2,656 (20.16%)	1,131 (8.59%)	3,344 (25.38%)	1,415 (10.74%)	2,325 (17.65%)	563 (4.27%)
1	10	12,448	5,795	12,448	200	3,184	4,438	1,798	5,795	10,811	9.15	62.37	1,244 (9.99%)	1,013 (8.14%)	3,154 (25.34%)	1,357 (10.90%)	3,864 (31.04%)	1,396 (11.21%)	2,815 (22.61%)	787 (6.32%)
1	11	11,345	6,028	11,345	218	3,292	4,733	1,867	6,028	11,459	9.76	65.27	1,145 (10.09%)	1,216 (10.72%)	3,105 (27.37%)	1,538 (13.56%)	3,857 (34.00%)	1,398 (12.32%)	2,791 (24.60%)	1,067 (9.41%)
1	12	11,831	7,409	11,831	256	3,764	5,614	2,709	7,409	15,066	8.11	59.58	1,078 (9.11%)	1,314 (11.11%)	2,754 (23.28%)	1,392 (11.77%)	3,612 (30.53%)	1,312 (11.09%)	2,646 (22.36%)	1,264 (10.68%)
1	13	12,121	8,675	12,121	278	4,196	6,470	3,650	8,675	17,697	6.32	53.13	983 (8.11%)	1,246 (10.28%)	2,255 (18.60%)	1,192 (9.83%)	3,279 (27.05%)	1,127 (9.30%)	2,302 (18.99%)	1,364 (11.25%)
1	14	8,140	5,955	8,140	186	3,364	5,572	2,391	5,955	12,883	6.76	55.38	843 (10.36%)	1,114 (13.69%)	1,730 (21.25%)	965 (11.86%)	2,501 (30.72%)	908 (11.15%)	1,766 (21.70%)	1,201 (14.75%)
1	15	4,854	3,541	4,854	109	2,726	4,965	1,237	3,541	8,378	7.88	59.73	638 (13.14%)	882 (18.17%)	1,263 (26.02%)	744 (15.33%)	1,741 (35.87%)	604 (12.44%)	1,332 (27.44%)	973 (20.05%)
1	16	2,961	2,218	2,961	86	1,978	4,236	741	2,218	5,805	8.56	61.97	460 (15.54%)	635 (21.45%)	893 (30.16%)	549 (18.54%)	1,215 (41.03%)	415 (14.02%)	976 (32.96%)	702 (23.71%)
1	17	1,861	1,459	1,861	59	1,403	3,558	475	1,459	4,225	8.33	61.77	344 (18.48%)	473 (25.42%)	598 (32.13%)	354 (19.02%)	781 (41.97%)	273 (14.67%)	628 (33.75%)	467 (25.09%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	1,219	994	1,219	51	993	3,019	303	994	3,226	9.20	62.64	274 (22.48%)	402 (32.98%)	425 (34.86%)	310 (25.43%)	558 (45.78%)	194 (15.91%)	507 (41.59%)	356 (29.20%)
1	19	818	699	818	51	737	2,819	230	699	2,562	10.50	61.50	201 (24.57%)	284 (34.72%)	332 (40.59%)	220 (26.89%)	412 (50.37%)	123 (15.04%)	349 (42.67%)	249 (30.44%)
1	20	967	861	967	124	895	5,613	322	861	4,572	9.40	56.51	299 (30.92%)	373 (38.57%)	355 (36.71%)	260 (26.89%)	442 (45.71%)	128 (13.24%)	414 (42.81%)	231 (23.89%)
1	Total	143,513	64,688	143,513	2,567	40,007	68,628	22,402	64,688	133,101	7.60	51.80	13,520 (9.42%)	10,936 (7.62%)	26,332 (18.35%)	12,889 (8.98%)	34,644 (24.14%)	14,287 (9.96%)	24,093 (16.79%)	10,380 (7.23%)

TABLE A.42: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Main* model (*Sample_1_train_half_2_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_2_test_half;
Submodel: Cond_Main

If[a]	#[b]	TP+FP[o]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	1	2,330	170	140	2	10	42	85	170	245	9.62	26.96	67 (2.88%)	2 (0.09%)	73 (3.13%)	14 (0.60%)	81 (3.48%)	25 (1.07%)	11 (0.47%)	2 (0.09%)
NA	2	12,257	1,718	2,710	9	106	136	806	1,718	2,498	2.50	31.64	486 (3.97%)	36 (0.29%)	619 (5.05%)	172 (1.40%)	762 (6.22%)	256 (2.09%)	208 (1.70%)	14 (0.11%)
NA	3	15,765	2,867	8,055	40	452	546	1,264	2,867	4,397	3.23	40.59	1,027 (6.51%)	132 (0.84%)	1,775 (11.26%)	541 (3.43%)	2,134 (13.54%)	860 (5.46%)	831 (5.27%)	62 (0.39%)
NA	4	17,739	4,090	8,431	69	727	907	1,732	4,090	6,672	4.47	44.51	1,222 (6.89%)	324 (1.83%)	2,978 (16.79%)	902 (5.08%)	3,537 (19.94%)	1,463 (8.25%)	1,531 (8.63%)	177 (1.00%)
NA	5	21,459	5,352	8,732	85	938	1,244	2,219	5,352	8,831	4.34	42.64	1,435 (6.69%)	499 (2.33%)	3,519 (16.40%)	1,208 (5.63%)	4,120 (19.20%)	1,535 (7.15%)	1,797 (8.37%)	331 (1.54%)
NA	6	23,938	5,888	13,026	131	1,368	1,815	2,200	5,888	9,993	4.21	44.54	1,671 (6.98%)	620 (2.59%)	3,933 (16.43%)	1,402 (5.86%)	4,695 (19.61%)	1,822 (7.61%)	1,967 (8.22%)	516 (2.16%)
NA	7	25,492	6,526	18,841	164	1,970	2,702	2,360	6,526	11,094	4.47	45.75	2,192 (8.60%)	773 (3.03%)	4,362 (17.11%)	1,577 (6.19%)	5,356 (21.01%)	2,157 (8.46%)	2,145 (8.41%)	606 (2.38%)
NA	8	23,332	6,404	20,032	155	2,471	3,232	2,323	6,404	11,174	4.81	48.47	2,293 (9.83%)	876 (3.75%)	4,569 (19.58%)	1,651 (7.08%)	5,729 (24.55%)	2,213 (9.48%)	2,441 (10.46%)	709 (3.04%)
NA	9	18,788	6,268	16,657	163	2,737	3,680	2,268	6,268	11,336	5.74	53.23	2,211 (11.77%)	1,020 (5.43%)	4,792 (25.51%)	1,712 (9.11%)	5,828 (31.02%)	2,245 (11.95%)	2,785 (14.82%)	791 (4.21%)
NA	10	14,286	5,954	12,577	132	2,626	3,659	2,093	5,954	11,213	6.44	58.22	1,814 (12.70%)	1,139 (7.97%)	4,656 (32.59%)	1,747 (12.23%)	5,422 (37.95%)	2,025 (14.17%)	2,907 (20.35%)	921 (6.45%)
NA	11	13,127	6,962	10,093	126	2,498	3,685	2,662	6,962	13,776	6.07	57.31	1,595 (12.15%)	1,237 (9.42%)	4,267 (32.51%)	1,715 (13.06%)	4,978 (37.92%)	1,775 (13.52%)	2,823 (21.51%)	1,044 (7.95%)
NA	12	12,937	8,337	9,557	160	2,585	3,979	3,596	8,337	16,584	5.21	53.91	1,334 (10.31%)	1,223 (9.45%)	3,805 (29.41%)	1,419 (10.97%)	4,568 (35.31%)	1,795 (13.87%)	2,482 (19.19%)	1,162 (8.98%)
NA	13	11,282	7,722	10,596	189	2,822	4,407	3,297	7,722	15,772	5.14	53.04	1,256 (11.13%)	1,049 (9.30%)	3,219 (28.53%)	1,268 (11.24%)	4,061 (36.00%)	1,550 (13.74%)	2,143 (18.99%)	1,179 (10.45%)
NA	14	10,265	7,531	10,106	164	2,809	4,600	3,391	7,531	15,691	4.23	49.91	1,223 (11.91%)	991 (9.65%)	2,561 (24.95%)	1,039 (10.12%)	3,502 (34.12%)	1,421 (13.84%)	1,866 (18.18%)	1,062 (10.35%)
NA	15	7,332	5,521	7,275	120	2,728	4,643	2,388	5,521	12,107	4.62	51.97	1,047 (14.28%)	877 (11.96%)	2,078 (28.34%)	872 (11.89%)	2,778 (37.89%)	1,138 (15.52%)	1,546 (21.09%)	934 (12.74%)
NA	16	4,648	3,412	4,626	79	2,240	4,201	1,335	3,412	8,058	5.66	56.41	856 (18.42%)	786 (16.91%)	1,634 (35.15%)	744 (16.01%)	2,094 (45.05%)	832 (17.90%)	1,239 (26.66%)	748 (16.09%)
NA	17	3,004	2,212	2,998	71	1,842	4,051	788	2,212	5,773	5.87	59.53	669 (22.27%)	678 (22.57%)	1,188 (39.55%)	571 (19.01%)	1,476 (49.13%)	613 (20.41%)	939 (31.26%)	606 (20.17%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	18	2,076	1,578	2,072	64	1,432	3,645	530	1,578	4,419	6.60	62.25	528 (25.43%)	566 (27.26%)	980 (47.21%)	490 (23.60%)	1,158 (55.78%)	480 (23.12%)	813 (39.16%)	511 (24.61%)
NA	19	1,547	1,209	1,546	74	1,149	3,525	407	1,209	3,707	6.39	61.57	457 (29.54%)	499 (32.26%)	740 (47.83%)	399 (25.79%)	901 (58.24%)	339 (21.91%)	621 (40.14%)	395 (25.53%)
NA	20	2,108	1,796	2,106	163	1,742	8,444	628	1,796	7,874	7.34	57.98	878 (41.65%)	797 (37.81%)	1,053 (49.95%)	620 (29.41%)	1,248 (59.20%)	478 (22.68%)	931 (44.17%)	551 (26.14%)
NA	Total	243,712	91,517	170,176	2,160	35,252	63,143	36,372	91,517	181,214	4.79	48.08	24,261 (9.95%)	14,124 (5.80%)	52,801 (21.67%)	20,063 (8.23%)	64,428 (26.44%)	25,022 (10.27%)	32,026 (13.14%)	12,321 (5.06%)

TABLE A.43: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Age-65p* model (*Sample_1_train_half_2_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_2_test_half;
Submodel: Cond_Age-65p

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	8,608	1,096	1,567	5	62	81	535	1,096	1,627	3.19	32.06	246 (2.86%)	22 (0.26%)	388 (4.51%)	76 (0.88%)	503 (5.84%)	147 (1.71%)	121 (1.41%)	5 (0.06%)
0	2	16,963	3,162	9,042	43	529	622	1,383	3,162	4,795	3.59	35.98	1,169 (6.89%)	127 (0.75%)	1,353 (7.98%)	448 (2.64%)	1,736 (10.23%)	758 (4.47%)	754 (4.44%)	40 (0.24%)
0	3	9,907	2,109	5,549	47	681	887	869	2,109	3,472	4.78	37.71	951 (9.60%)	179 (1.81%)	1,107 (11.17%)	530 (5.35%)	1,405 (14.18%)	747 (7.54%)	717 (7.24%)	73 (0.74%)
0	4	10,921	2,112	3,700	48	581	852	875	2,112	3,513	3.29	34.50	735 (6.73%)	151 (1.38%)	879 (8.05%)	403 (3.69%)	1,145 (10.48%)	516 (4.72%)	492 (4.51%)	97 (0.89%)
0	5	17,606	3,257	5,986	71	661	970	1,305	3,257	5,398	2.23	30.92	853 (4.84%)	164 (0.93%)	965 (5.48%)	413 (2.35%)	1,372 (7.79%)	474 (2.69%)	418 (2.37%)	92 (0.52%)
0	6	21,869	4,046	12,830	103	1,110	1,538	1,646	4,046	6,754	2.41	33.32	1,369 (6.26%)	185 (0.85%)	1,235 (5.65%)	563 (2.57%)	1,968 (9.00%)	589 (2.69%)	564 (2.58%)	97 (0.44%)
0	7	17,875	3,760	14,466	107	1,626	2,192	1,462	3,760	6,346	3.05	35.01	1,697 (9.49%)	216 (1.21%)	1,261 (7.05%)	603 (3.37%)	2,030 (11.36%)	857 (4.79%)	613 (3.43%)	97 (0.54%)
0	8	11,848	2,847	10,890	94	1,602	2,093	1,095	2,847	4,934	3.50	36.43	1,618 (13.66%)	226 (1.91%)	1,149 (9.70%)	611 (5.16%)	1,756 (14.82%)	905 (7.64%)	622 (5.25%)	96 (0.81%)
0	9	8,047	2,316	7,537	60	1,529	2,118	812	2,316	4,208	3.91	38.35	1,270 (15.78%)	212 (2.63%)	1,017 (12.64%)	624 (7.75%)	1,494 (18.57%)	803 (9.98%)	604 (7.51%)	93 (1.16%)
0	10	5,522	2,025	4,981	52	1,255	1,821	735	2,025	3,829	4.39	39.60	940 (17.02%)	252 (4.56%)	859 (15.56%)	519 (9.40%)	1,132 (20.50%)	675 (12.22%)	561 (10.16%)	116 (2.10%)
0	11	4,212	2,007	3,501	48	1,054	1,698	883	2,007	4,044	4.32	39.86	709 (16.83%)	226 (5.37%)	697 (16.55%)	448 (10.64%)	922 (21.89%)	508 (12.06%)	495 (11.75%)	126 (2.99%)
0	12	4,562	2,744	2,717	50	925	1,602	1,150	2,744	5,673	3.09	36.15	503 (11.03%)	187 (4.10%)	550 (12.06%)	372 (8.15%)	767 (16.81%)	407 (8.92%)	395 (8.66%)	113 (2.48%)
0	13	6,716	4,967	3,508	84	956	1,759	2,538	4,967	10,419	1.95	32.71	494 (7.36%)	180 (2.68%)	489 (7.28%)	301 (4.48%)	890 (13.25%)	336 (5.00%)	333 (4.96%)	112 (1.67%)
0	14	7,513	6,098	5,591	133	1,217	2,089	3,176	6,098	12,532	1.46	31.64	492 (6.55%)	148 (1.97%)	383 (5.10%)	219 (2.91%)	955 (12.71%)	396 (5.27%)	265 (3.53%)	99 (1.32%)
0	15	6,215	5,177	6,111	102	1,402	2,406	2,618	5,177	11,184	1.58	31.82	574 (9.24%)	134 (2.16%)	311 (5.00%)	201 (3.23%)	967 (15.56%)	453 (7.29%)	208 (3.35%)	80 (1.29%)
0	16	4,668	3,881	4,654	65	1,392	2,463	1,978	3,881	8,891	1.69	32.54	604 (12.94%)	123 (2.63%)	304 (6.51%)	217 (4.65%)	832 (17.82%)	414 (8.87%)	192 (4.11%)	68 (1.46%)
0	17	2,889	2,424	2,880	62	1,134	2,257	1,097	2,424	6,015	2.58	33.49	560 (19.38%)	123 (4.26%)	236 (8.17%)	196 (6.78%)	580 (20.08%)	328 (11.35%)	177 (6.13%)	73 (2.53%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	1,628	1,339	1,624	36	914	2,141	600	1,339	3,889	2.78	35.66	410 (25.18%)	107 (6.57%)	224 (13.76%)	183 (11.24%)	392 (24.08%)	244 (14.99%)	141 (8.66%)	56 (3.44%)
0	19	1,129	893	1,128	42	743	2,162	358	893	2,871	4.38	38.46	308 (27.28%)	119 (10.54%)	231 (20.46%)	166 (14.70%)	343 (30.38%)	209 (18.51%)	149 (13.20%)	63 (5.58%)
0	20	1,699	1,408	1,696	127	1,339	6,533	516	1,408	6,429	5.59	42.51	611 (35.96%)	294 (17.30%)	485 (28.55%)	393 (23.13%)	678 (39.91%)	314 (18.48%)	359 (21.13%)	139 (8.18%)
0	Total	170,397	57,668	109,958	1,379	20,712	38,284	25,631	57,668	116,823	3.03	34.65	16,113 (9.46%)	3,375 (1.98%)	14,123 (8.29%)	7,486 (4.39%)	21,867 (12.83%)	10,080 (5.92%)	8,180 (4.80%)	1,735 (1.02%)
1	1	6	1	1	0	0	0	1	1	2	145.33	80.00	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (16.67%)	0 (0.00%)	0 (0.00%)
1	2	70	5	5	0	2	7	0	5	7	55.01	86.69	0 (0.00%)	0 (0.00%)	3 (4.29%)	1 (1.43%)	3 (4.29%)	7 (10.00%)	0 (0.00%)	2 (2.86%)
1	3	318	13	17	1	3	4	9	13	19	4.53	85.76	8 (2.52%)	1 (0.31%)	22 (6.92%)	10 (3.14%)	24 (7.55%)	12 (3.77%)	3 (0.94%)	0 (0.00%)
1	4	867	133	76	1	3	3	64	133	187	3.80	75.56	22 (2.54%)	20 (2.31%)	241 (27.80%)	66 (7.61%)	247 (28.49%)	36 (4.15%)	70 (8.07%)	7 (0.81%)
1	5	1,330	263	304	0	24	29	106	263	366	7.03	75.36	50 (3.76%)	41 (3.08%)	446 (33.53%)	115 (8.65%)	466 (35.04%)	81 (6.09%)	131 (9.85%)	22 (1.65%)
1	6	4,055	991	868	5	65	78	407	991	1,595	3.91	76.05	137 (3.38%)	118 (2.91%)	1,373 (33.86%)	338 (8.34%)	1,457 (35.93%)	239 (5.89%)	447 (11.02%)	50 (1.23%)
1	7	5,222	1,563	2,837	7	164	196	636	1,563	2,571	6.35	76.92	290 (5.55%)	257 (4.92%)	2,120 (40.60%)	519 (9.94%)	2,252 (43.13%)	575 (11.01%)	903 (17.29%)	152 (2.91%)
1	8	8,429	2,916	6,545	32	366	430	1,060	2,916	4,795	7.78	77.30	545 (6.47%)	432 (5.13%)	3,893 (46.19%)	942 (11.18%)	4,175 (49.53%)	1,349 (16.00%)	1,782 (21.14%)	296 (3.51%)
1	9	11,336	4,662	9,276	60	809	995	1,636	4,662	8,010	10.00	78.83	897 (7.91%)	815 (7.19%)	5,641 (49.76%)	1,549 (13.66%)	6,162 (54.36%)	2,164 (19.09%)	3,060 (26.99%)	692 (6.10%)
1	10	10,788	4,922	9,712	80	1,264	1,547	1,541	4,922	8,587	10.02	80.17	1,037 (9.61%)	1,124 (10.42%)	5,850 (54.23%)	1,794 (16.63%)	6,460 (59.88%)	2,361 (21.89%)	3,428 (31.78%)	1,196 (11.09%)
1	11	9,202	4,677	8,931	81	1,804	2,212	1,369	4,677	8,627	10.03	81.00	1,089 (11.83%)	1,381 (15.01%)	5,265 (57.22%)	1,762 (19.15%)	5,869 (63.78%)	2,101 (22.83%)	3,294 (35.80%)	1,513 (16.44%)
1	12	6,965	3,947	6,930	75	1,970	2,574	1,151	3,947	7,487	9.93	81.56	954 (13.70%)	1,337 (19.20%)	4,185 (60.09%)	1,482 (21.28%)	4,674 (67.11%)	1,735 (24.91%)	2,842 (40.80%)	1,525 (21.90%)
1	13	4,884	2,928	4,877	91	1,923	2,773	824	2,928	5,728	9.98	82.12	789 (16.15%)	1,231 (25.20%)	2,999 (61.40%)	1,148 (23.51%)	3,392 (69.45%)	1,275 (26.11%)	2,231 (45.68%)	1,377 (28.19%)
1	14	3,362	2,121	3,359	76	1,673	2,749	582	2,121	4,328	9.45	82.02	582 (17.31%)	1,013 (30.13%)	2,165 (64.40%)	817 (24.30%)	2,409 (71.65%)	979 (29.12%)	1,693 (50.36%)	1,050 (31.23%)
1	15	2,228	1,495	2,227	55	1,343	2,478	437	1,495	3,306	8.96	81.66	462 (20.74%)	794 (35.64%)	1,448 (64.99%)	604 (27.11%)	1,630 (73.16%)	668 (29.98%)	1,209 (54.26%)	787 (35.32%)
1	16	1,517	1,067	1,517	46	990	2,146	285	1,067	2,551	8.75	81.23	363 (23.93%)	660 (43.51%)	1,023 (67.44%)	444 (29.27%)	1,152 (75.94%)	464 (30.59%)	883 (58.21%)	618 (40.74%)
1	17	963	704	963	43	699	1,651	188	704	1,658	9.19	80.90	254 (26.38%)	458 (47.56%)	685 (71.13%)	340 (35.31%)	752 (78.09%)	301 (31.26%)	594 (61.68%)	418 (43.41%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	750	581	750	37	583	1,595	158	581	1,529	8.14	80.87	224 (29.87%)	390 (52.00%)	544 (72.53%)	265 (35.33%)	594 (79.20%)	234 (31.20%)	528 (70.40%)	349 (46.53%)
1	19	464	372	464	29	377	1,270	113	372	1,110	8.51	79.81	178 (38.36%)	279 (60.13%)	337 (72.63%)	161 (34.70%)	366 (78.88%)	161 (34.70%)	305 (65.73%)	227 (48.92%)
1	20	559	488	559	62	478	2,122	174	488	1,928	7.08	78.62	267 (47.76%)	398 (71.20%)	438 (78.35%)	220 (39.36%)	477 (85.33%)	199 (35.60%)	443 (79.25%)	305 (54.56%)
1	Total	73,315	33,849	60,218	781	14,540	24,859	10,741	33,849	64,391	8.91	79.58	8,148 (11.11%)	10,749 (14.66%)	38,678 (52.76%)	12,577 (17.15%)	42,561 (58.05%)	14,942 (20.38%)	23,846 (32.53%)	10,586 (14.44%)

TABLE A.44: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Acute-12-month* model (*Sample_1_train_half_2_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_2_test_half;
Submodel: Cond_Prior-Acute-12-month

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	2,221	149	171	1	7	24	72	149	207	5.38	27.76	68 (3.06%)	3 (0.14%)	83 (3.74%)	19 (0.86%)	91 (4.10%)	29 (1.31%)	19 (0.86%)	3 (0.14%)
0	2	10,697	1,448	1,712	7	74	92	689	1,448	2,101	2.42	31.20	405 (3.79%)	37 (0.35%)	533 (4.98%)	144 (1.35%)	637 (5.95%)	190 (1.78%)	157 (1.47%)	10 (0.09%)
0	3	11,783	1,956	3,448	15	264	314	949	1,956	2,981	3.29	42.85	634 (5.38%)	112 (0.95%)	1,453 (12.33%)	432 (3.67%)	1,669 (14.16%)	585 (4.96%)	518 (4.40%)	43 (0.36%)
0	4	12,290	2,587	3,764	31	365	441	1,134	2,587	4,138	4.07	44.78	638 (5.19%)	194 (1.58%)	1,939 (15.78%)	564 (4.59%)	2,208 (17.97%)	848 (6.90%)	841 (6.84%)	87 (0.71%)
0	5	16,979	3,766	4,487	43	454	583	1,633	3,766	6,035	3.65	39.16	899 (5.29%)	277 (1.63%)	2,158 (12.71%)	765 (4.51%)	2,475 (14.58%)	940 (5.54%)	972 (5.72%)	154 (0.91%)
0	6	14,978	3,455	5,614	46	695	877	1,411	3,455	5,560	4.27	47.16	1,041 (6.95%)	340 (2.27%)	2,412 (16.10%)	856 (5.72%)	2,733 (18.25%)	980 (6.54%)	984 (6.57%)	193 (1.29%)
0	7	12,530	3,033	5,521	71	910	1,130	1,230	3,033	5,014	4.54	51.33	792 (6.32%)	382 (3.05%)	2,488 (19.86%)	796 (6.35%)	2,846 (22.71%)	1,033 (8.24%)	1,020 (8.14%)	236 (1.88%)
0	8	9,651	2,524	4,825	71	1,012	1,299	1,034	2,524	4,349	5.35	53.72	619 (6.41%)	396 (4.10%)	2,044 (21.18%)	697 (7.22%)	2,363 (24.48%)	1,082 (11.21%)	968 (10.03%)	258 (2.67%)
0	9	6,116	1,981	3,791	68	962	1,282	793	1,981	3,444	7.07	58.62	467 (7.64%)	374 (6.12%)	1,634 (26.72%)	601 (9.83%)	1,875 (30.66%)	928 (15.17%)	837 (13.69%)	265 (4.33%)
0	10	4,206	1,989	2,399	52	822	1,157	899	1,989	3,664	7.01	57.64	251 (5.97%)	281 (6.68%)	1,105 (26.27%)	434 (10.32%)	1,311 (31.17%)	706 (16.79%)	564 (13.41%)	245 (5.83%)
0	11	4,014	2,490	1,702	41	673	1,010	1,112	2,490	4,995	5.03	47.39	197 (4.91%)	221 (5.51%)	753 (18.76%)	319 (7.95%)	942 (23.47%)	493 (12.28%)	421 (10.49%)	207 (5.16%)
0	12	3,910	2,888	1,671	56	626	986	1,476	2,888	6,006	3.44	40.83	146 (3.73%)	122 (3.12%)	480 (12.28%)	178 (4.55%)	701 (17.93%)	428 (10.95%)	286 (7.31%)	177 (4.53%)
0	13	4,227	3,481	1,558	73	716	1,128	1,967	3,481	6,905	2.21	36.25	121 (2.86%)	100 (2.37%)	282 (6.67%)	114 (2.70%)	545 (12.89%)	321 (7.59%)	184 (4.35%)	147 (3.48%)
0	14	1,782	1,342	1,309	39	602	1,008	684	1,342	2,726	4.33	43.22	99 (5.56%)	76 (4.26%)	208 (11.67%)	92 (5.16%)	321 (18.01%)	249 (13.97%)	130 (7.30%)	94 (5.27%)
0	15	697	481	609	23	434	791	228	481	1,040	6.86	51.38	46 (6.60%)	33 (4.73%)	127 (18.22%)	61 (8.75%)	167 (23.96%)	183 (26.26%)	87 (12.48%)	58 (8.32%)
0	16	311	201	292	20	242	544	91	201	456	10.14	59.35	24 (7.72%)	29 (9.32%)	97 (31.19%)	51 (16.40%)	117 (37.62%)	111 (35.69%)	39 (12.54%)	34 (10.93%)
0	17	136	78	127	8	115	320	21	78	168	15.98	70.18	12 (8.82%)	18 (13.24%)	56 (41.18%)	29 (21.32%)	67 (49.26%)	69 (50.74%)	35 (25.74%)	21 (15.44%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	93	63	89	4	75	253	22	63	148	11.67	70.86	11 (11.83%)	16 (17.20%)	49 (52.69%)	20 (21.51%)	54 (58.06%)	51 (54.84%)	28 (30.11%)	22 (23.66%)
0	19	54	35	52	4	43	169	9	35	69	10.30	68.02	3 (5.56%)	8 (14.81%)	24 (44.44%)	8 (14.81%)	29 (53.70%)	22 (40.74%)	19 (35.19%)	8 (14.81%)
0	20	46	31	44	3	35	176	10	31	70	12.37	58.09	5 (10.87%)	13 (28.26%)	19 (41.30%)	11 (23.91%)	20 (43.48%)	25 (54.35%)	11 (23.91%)	8 (17.39%)
0	Total	116,721	33,978	43,185	676	9,126	13,584	15,464	33,978	60,076	4.26	44.88	6,478 (5.55%)	3,032 (2.60%)	17,944 (15.37%)	6,191 (5.30%)	21,171 (18.14%)	9,273 (7.94%)	8,120 (6.96%)	2,270 (1.94%)
1	1	68	10	68	0	4	5	2	10	12	137.06	23.69	4 (5.88%)	0 (0.00%)	3 (4.41%)	0 (0.00%)	3 (4.41%)	3 (4.41%)	0 (0.00%)	1 (1.47%)
1	2	2,041	352	2,041	1	20	20	127	352	509	3.73	24.78	187 (9.16%)	3 (0.15%)	55 (2.69%)	28 (1.37%)	92 (4.51%)	57 (2.79%)	35 (1.71%)	4 (0.20%)
1	3	4,636	1,042	4,636	5	177	191	382	1,042	1,677	3.89	37.27	514 (11.09%)	39 (0.84%)	494 (10.66%)	155 (3.34%)	657 (14.17%)	300 (6.47%)	297 (6.41%)	25 (0.54%)
1	4	5,094	1,549	5,094	11	362	412	621	1,549	2,635	5.51	49.87	584 (11.46%)	133 (2.61%)	1,263 (24.79%)	385 (7.56%)	1,529 (30.02%)	581 (11.41%)	798 (15.67%)	98 (1.92%)
1	5	6,410	1,948	6,410	29	467	562	681	1,948	3,343	5.35	49.38	648 (10.11%)	281 (4.38%)	1,652 (25.77%)	553 (8.63%)	1,996 (31.14%)	738 (11.51%)	1,033 (16.12%)	255 (3.98%)
1	6	11,844	3,073	11,844	32	668	825	1,010	3,073	5,330	3.96	41.05	1,006 (8.49%)	397 (3.35%)	1,965 (16.59%)	724 (6.11%)	2,603 (21.98%)	1,042 (8.80%)	1,176 (9.93%)	365 (3.08%)
1	7	13,194	3,602	13,194	53	1,144	1,350	1,218	3,602	6,412	4.46	41.35	1,438 (10.90%)	459 (3.48%)	1,958 (14.84%)	797 (6.04%)	2,774 (21.02%)	1,106 (8.38%)	1,177 (8.92%)	441 (3.34%)
1	8	11,461	3,495	11,461	66	1,507	1,893	1,135	3,495	6,315	5.10	45.74	1,454 (12.69%)	484 (4.22%)	2,267 (19.78%)	875 (7.63%)	2,990 (26.09%)	1,074 (9.37%)	1,366 (11.92%)	461 (4.02%)
1	9	9,907	3,436	9,907	61	1,616	2,124	1,073	3,436	6,242	6.03	52.56	1,434 (14.47%)	564 (5.69%)	2,709 (27.34%)	1,028 (10.38%)	3,274 (33.05%)	1,220 (12.31%)	1,753 (17.69%)	522 (5.27%)
1	10	9,024	3,660	9,024	76	1,698	2,359	1,106	3,660	6,892	6.60	58.65	1,399 (15.50%)	744 (8.24%)	3,158 (35.00%)	1,170 (12.97%)	3,662 (40.58%)	1,226 (13.59%)	2,145 (23.77%)	658 (7.29%)
1	11	8,457	4,043	8,457	77	1,773	2,571	1,326	4,043	7,842	6.63	63.18	1,259 (14.89%)	950 (11.23%)	3,415 (40.38%)	1,285 (15.19%)	3,880 (45.88%)	1,255 (14.84%)	2,399 (28.37%)	818 (9.67%)
1	12	9,189	5,450	9,189	93	1,832	2,745	1,987	5,450	10,446	5.67	58.92	1,266 (13.78%)	1,022 (11.12%)	3,313 (36.05%)	1,236 (13.45%)	3,894 (42.38%)	1,325 (14.42%)	2,152 (23.42%)	934 (10.16%)
1	13	9,626	6,516	9,626	96	2,068	3,062	2,675	6,516	13,299	4.83	54.64	1,243 (12.91%)	1,036 (10.76%)	3,010 (31.27%)	1,174 (12.20%)	3,738 (38.83%)	1,330 (13.82%)	2,103 (21.85%)	1,048 (10.89%)
1	14	7,979	5,696	7,979	120	2,367	3,666	2,416	5,696	12,097	4.58	54.36	1,147 (14.38%)	981 (12.29%)	2,437 (30.54%)	975 (12.22%)	3,195 (40.04%)	1,178 (14.76%)	1,804 (22.61%)	948 (11.88%)
1	15	5,925	4,296	5,925	113	2,418	4,020	1,692	4,296	9,599	4.95	55.87	984 (16.61%)	879 (14.84%)	1,976 (33.35%)	871 (14.70%)	2,578 (43.51%)	964 (16.27%)	1,498 (25.28%)	874 (14.75%)
1	16	3,927	2,913	3,927	116	2,083	3,906	1,105	2,913	7,122	5.41	58.17	785 (19.99%)	769 (19.58%)	1,518 (38.66%)	676 (17.21%)	1,909 (48.61%)	707 (18.00%)	1,144 (29.13%)	701 (17.85%)
1	17	2,820	2,076	2,820	102	1,794	4,051	753	2,076	5,395	5.48	59.92	640 (22.70%)	660 (23.40%)	1,191 (42.23%)	558 (19.79%)	1,461 (51.81%)	533 (18.90%)	930 (32.98%)	595 (21.10%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	1,926	1,491	1,926	107	1,347	3,584	525	1,491	4,269	5.86	59.90	499 (25.91%)	506 (26.27%)	856 (44.44%)	453 (23.52%)	1,058 (54.93%)	391 (20.30%)	719 (37.33%)	450 (23.36%)
1	19	1,511	1,199	1,511	100	1,125	3,591	431	1,199	3,842	6.07	59.46	467 (30.91%)	478 (31.63%)	685 (45.33%)	361 (23.89%)	832 (55.06%)	308 (20.38%)	574 (37.99%)	363 (24.02%)
1	20	1,952	1,692	1,952	226	1,656	8,622	643	1,692	7,860	6.77	55.30	825 (42.26%)	707 (36.22%)	932 (47.75%)	568 (29.10%)	1,132 (57.99%)	411 (21.06%)	803 (41.14%)	490 (25.10%)
1	Total	126,991	57,539	126,991	1,484	26,126	49,559	20,908	57,539	121,138	5.29	51.02	17,783 (14.00%)	11,092 (8.73%)	34,857 (27.45%)	13,872 (10.92%)	43,257 (34.06%)	15,749 (12.40%)	23,906 (18.82%)	10,051 (7.91%)

TABLE A.45: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Oper-12-month* model (*Sample_1_train_half_2_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_2_test_half;
Submodel: Cond_Prior-Oper-12-month

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	16	4	10	0	2	3	0	4	4	793.94	38.56	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (6.25%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
0	2	1,006	84	3	0	2	2	50	84	93	6.48	21.65	17 (1.69%)	0 (0.00%)	8 (0.80%)	4 (0.40%)	9 (0.89%)	0 (0.00%)	1 (0.10%)	0 (0.00%)
0	3	3,999	314	81	2	4	4	172	314	399	2.78	32.33	175 (4.38%)	12 (0.30%)	199 (4.98%)	103 (2.58%)	212 (5.30%)	15 (0.38%)	45 (1.13%)	4 (0.10%)
0	4	8,477	1,297	300	0	8	10	613	1,297	1,935	1.64	26.73	324 (3.82%)	46 (0.54%)	329 (3.88%)	152 (1.79%)	340 (4.01%)	34 (0.40%)	88 (1.04%)	8 (0.09%)
0	5	10,509	1,934	686	2	51	60	827	1,934	3,058	1.95	37.12	673 (6.40%)	72 (0.69%)	821 (7.81%)	388 (3.69%)	865 (8.23%)	60 (0.57%)	221 (2.10%)	25 (0.24%)
0	6	7,203	1,632	1,992	7	156	181	699	1,632	2,669	3.12	48.03	689 (9.57%)	191 (2.65%)	1,178 (16.35%)	501 (6.96%)	1,237 (17.17%)	88 (1.22%)	425 (5.90%)	57 (0.79%)
0	7	4,842	1,325	2,436	13	309	356	536	1,325	2,251	5.01	49.60	525 (10.84%)	206 (4.25%)	934 (19.29%)	436 (9.00%)	989 (20.43%)	80 (1.65%)	456 (9.42%)	93 (1.92%)
0	8	3,382	1,203	2,087	17	367	420	441	1,203	2,083	7.50	56.53	411 (12.15%)	178 (5.26%)	817 (24.16%)	373 (11.03%)	874 (25.84%)	83 (2.45%)	438 (12.95%)	124 (3.67%)
0	9	2,460	921	1,732	13	361	431	317	921	1,659	8.79	62.59	288 (11.71%)	208 (8.46%)	751 (30.53%)	322 (13.09%)	825 (33.54%)	63 (2.56%)	405 (16.46%)	146 (5.93%)
0	10	1,826	814	1,466	10	322	392	256	814	1,495	9.21	66.25	208 (11.39%)	186 (10.19%)	604 (33.08%)	259 (14.18%)	687 (37.62%)	57 (3.12%)	357 (19.55%)	163 (8.93%)
0	11	1,467	718	1,097	7	296	390	256	718	1,334	7.77	61.70	157 (10.70%)	165 (11.25%)	453 (30.88%)	207 (14.11%)	511 (34.83%)	48 (3.27%)	268 (18.27%)	162 (11.04%)
0	12	960	510	843	9	237	322	172	510	954	8.79	67.89	114 (11.88%)	163 (16.98%)	346 (36.04%)	155 (16.15%)	395 (41.15%)	32 (3.33%)	227 (23.65%)	140 (14.58%)
0	13	1,216	824	610	6	258	368	323	824	1,793	4.32	47.60	73 (6.00%)	147 (12.09%)	231 (19.00%)	116 (9.54%)	274 (22.53%)	27 (2.22%)	146 (12.01%)	124 (10.20%)
0	14	2,255	1,587	411	5	210	297	687	1,587	3,388	2.32	35.03	81 (3.59%)	96 (4.26%)	152 (6.74%)	77 (3.41%)	248 (11.00%)	38 (1.69%)	101 (4.48%)	78 (3.46%)
0	15	829	497	331	10	156	292	195	497	1,033	3.46	44.07	69 (8.32%)	73 (8.81%)	88 (10.62%)	42 (5.07%)	134 (16.16%)	28 (3.38%)	62 (7.48%)	55 (6.63%)
0	16	626	463	540	12	171	280	202	463	1,138	2.84	37.80	70 (11.18%)	49 (7.83%)	63 (10.06%)	45 (7.19%)	93 (14.86%)	16 (2.56%)	48 (7.67%)	35 (5.59%)
0	17	1,030	869	404	26	257	428	501	869	1,793	1.56	32.45	55 (5.34%)	49 (4.76%)	54 (5.24%)	33 (3.20%)	104 (10.10%)	7 (0.68%)	44 (4.27%)	24 (2.33%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	1,189	1,061	237	10	142	322	637	1,061	2,226	0.77	29.22	46 (3.87%)	50 (4.21%)	42 (3.53%)	26 (2.19%)	98 (8.24%)	7 (0.59%)	46 (3.87%)	26 (2.19%)
0	19	3,984	3,587	610	50	265	546	2,131	3,587	7,152	0.54	28.51	113 (2.84%)	35 (0.88%)	36 (0.90%)	16 (0.40%)	272 (6.83%)	9 (0.23%)	26 (0.65%)	23 (0.58%)
0	20	857	785	836	31	473	1,042	426	785	1,890	1.41	30.78	92 (10.74%)	44 (5.13%)	42 (4.90%)	30 (3.50%)	92 (10.74%)	5 (0.58%)	35 (4.08%)	21 (2.45%)
0	Total	58,133	20,429	16,712	230	4,047	6,146	9,441	20,429	38,347	3.72	40.94	4,180 (7.19%)	1,970 (3.39%)	7,148 (12.30%)	3,286 (5.65%)	8,259 (14.21%)	697 (1.20%)	3,439 (5.92%)	1,308 (2.25%)
1	1	1,085	60	37	1	6	7	34	60	88	6.70	25.01	18 (1.66%)	1 (0.09%)	29 (2.67%)	4 (0.37%)	31 (2.86%)	13 (1.20%)	2 (0.18%)	1 (0.09%)
1	2	8,691	1,066	1,092	0	57	73	485	1,066	1,524	1.63	28.76	289 (3.33%)	13 (0.15%)	288 (3.31%)	89 (1.02%)	347 (3.99%)	101 (1.16%)	78 (0.90%)	7 (0.08%)
1	3	13,208	2,166	5,634	28	323	398	1,003	2,166	3,371	2.24	37.28	696 (5.27%)	67 (0.51%)	1,057 (8.00%)	284 (2.15%)	1,295 (9.80%)	462 (3.50%)	411 (3.11%)	28 (0.21%)
1	4	15,952	3,213	8,561	40	664	785	1,352	3,213	5,150	3.36	41.98	940 (5.89%)	166 (1.04%)	2,042 (12.80%)	560 (3.51%)	2,553 (16.00%)	1,057 (6.63%)	1,040 (6.52%)	82 (0.51%)
1	5	16,596	3,970	12,805	99	1,256	1,530	1,602	3,970	6,572	4.79	45.59	1,176 (7.09%)	339 (2.04%)	2,865 (17.26%)	887 (5.34%)	3,654 (22.02%)	1,675 (10.09%)	1,593 (9.60%)	219 (1.32%)
1	6	18,134	4,783	15,810	113	1,585	2,048	1,849	4,783	8,045	5.25	47.45	1,396 (7.70%)	507 (2.80%)	3,518 (19.40%)	1,055 (5.82%)	4,529 (24.98%)	2,004 (11.05%)	1,859 (10.25%)	411 (2.27%)
1	7	17,243	4,910	16,059	131	1,775	2,354	1,750	4,910	8,456	5.33	48.85	1,632 (9.46%)	600 (3.48%)	3,715 (21.54%)	1,221 (7.08%)	4,765 (27.63%)	2,096 (12.16%)	2,043 (11.85%)	547 (3.17%)
1	8	15,658	5,201	15,044	113	2,012	2,689	1,811	5,201	9,256	5.92	53.72	1,732 (11.06%)	785 (5.01%)	4,232 (27.03%)	1,414 (9.03%)	5,088 (32.49%)	2,355 (15.04%)	2,361 (15.08%)	704 (4.50%)
1	9	14,094	5,883	13,783	127	2,094	2,862	2,111	5,883	10,905	6.31	56.50	1,721 (12.21%)	913 (6.48%)	4,440 (31.50%)	1,528 (10.84%)	5,213 (36.99%)	2,290 (16.25%)	2,637 (18.71%)	850 (6.03%)
1	10	13,567	6,758	13,427	130	2,404	3,389	2,594	6,758	12,827	6.01	56.81	1,758 (12.96%)	1,054 (7.77%)	4,457 (32.85%)	1,576 (11.62%)	5,217 (38.45%)	2,276 (16.78%)	2,829 (20.85%)	920 (6.78%)
1	11	12,315	6,887	12,244	143	2,558	3,799	2,719	6,887	13,499	5.88	56.89	1,606 (13.04%)	1,140 (9.26%)	4,061 (32.98%)	1,581 (12.84%)	4,920 (39.95%)	2,134 (17.33%)	2,702 (21.94%)	1,015 (8.24%)
1	12	10,595	6,518	10,572	164	2,751	4,288	2,567	6,518	13,025	5.99	58.16	1,426 (13.46%)	1,185 (11.18%)	3,600 (33.98%)	1,402 (13.23%)	4,449 (41.99%)	1,916 (18.08%)	2,423 (22.87%)	1,116 (10.53%)
1	13	8,148	5,215	8,127	129	2,639	4,216	1,945	5,215	11,015	6.31	60.02	1,248 (15.32%)	1,039 (12.75%)	3,021 (37.08%)	1,198 (14.70%)	3,739 (45.89%)	1,604 (19.69%)	2,126 (26.09%)	1,132 (13.89%)
1	14	6,091	4,025	6,078	144	2,437	4,216	1,449	4,025	8,910	6.69	61.21	1,040 (17.07%)	954 (15.66%)	2,425 (39.81%)	989 (16.24%)	2,994 (49.15%)	1,248 (20.49%)	1,762 (28.93%)	999 (16.40%)
1	15	4,334	2,981	4,329	102	2,104	3,998	1,045	2,981	6,961	6.59	62.27	750 (17.31%)	810 (18.69%)	1,797 (41.46%)	807 (18.62%)	2,206 (50.90%)	921 (21.25%)	1,336 (30.83%)	787 (18.16%)
1	16	3,220	2,258	3,218	95	1,865	3,897	801	2,258	5,563	6.29	61.35	639 (19.84%)	711 (22.08%)	1,312 (40.75%)	610 (18.94%)	1,645 (51.09%)	745 (23.14%)	1,041 (32.33%)	706 (21.93%)
1	17	2,200	1,625	2,198	67	1,376	3,354	536	1,625	4,365	6.39	61.52	528 (24.00%)	544 (24.73%)	938 (42.64%)	488 (22.18%)	1,166 (53.00%)	475 (21.59%)	765 (34.77%)	502 (22.82%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	1,636	1,239	1,635	70	1,102	3,135	405	1,239	3,579	6.32	58.90	439 (26.83%)	446 (27.26%)	687 (41.99%)	376 (22.98%)	881 (53.85%)	376 (22.98%)	577 (35.27%)	369 (22.56%)
1	19	1,223	964	1,222	66	897	3,078	359	964	3,353	6.33	55.75	386 (31.56%)	373 (30.50%)	486 (39.74%)	259 (21.18%)	611 (49.96%)	258 (21.10%)	413 (33.77%)	275 (22.49%)
1	20	1,589	1,366	1,589	168	1,300	6,881	514	1,366	6,403	5.73	52.35	661 (41.60%)	507 (31.91%)	683 (42.98%)	449 (28.26%)	866 (54.50%)	319 (20.08%)	589 (37.07%)	343 (21.59%)
1	Total	185,579	71,088	153,464	1,930	31,205	56,997	26,931	71,088	142,867	5.13	50.31	20,081 (10.82%)	12,154 (6.55%)	45,653 (24.60%)	16,777 (9.04%)	56,169 (30.27%)	24,325 (13.11%)	28,587 (15.40%)	11,013 (5.93%)

TABLE A.46: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Spells* model (*Sample_1_train_half_2_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_2_test_half;
Submodel: Cond_Prior-Spells

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	2,664	187	0	0	0	0	96	187	256	2.75	29.01	88 (3.30%)	7 (0.26%)	93 (3.49%)	21 (0.79%)	106 (3.98%)	30 (1.13%)	20 (0.75%)	2 (0.08%)
0	2	9,386	1,309	0	0	0	0	646	1,309	1,898	2.48	32.12	320 (3.41%)	35 (0.37%)	487 (5.19%)	158 (1.68%)	593 (6.32%)	126 (1.34%)	166 (1.77%)	10 (0.11%)
0	3	8,375	1,333	0	0	0	0	673	1,333	2,021	3.70	41.19	363 (4.33%)	84 (1.00%)	887 (10.59%)	305 (3.64%)	1,058 (12.63%)	209 (2.50%)	397 (4.74%)	39 (0.47%)
0	4	8,002	1,469	0	0	0	0	658	1,469	2,396	4.26	40.60	327 (4.09%)	114 (1.42%)	913 (11.41%)	331 (4.14%)	1,074 (13.42%)	229 (2.86%)	427 (5.34%)	64 (0.80%)
0	5	12,559	2,462	0	0	0	0	1,072	2,462	3,863	3.12	33.56	542 (4.32%)	147 (1.17%)	906 (7.21%)	378 (3.01%)	1,100 (8.76%)	232 (1.85%)	379 (3.02%)	93 (0.74%)
0	6	9,442	1,898	0	0	0	0	832	1,898	3,017	3.89	44.33	515 (5.45%)	187 (1.98%)	1,092 (11.57%)	455 (4.82%)	1,293 (13.69%)	214 (2.27%)	410 (4.34%)	101 (1.07%)
0	7	6,779	1,485	0	0	0	0	646	1,485	2,454	4.29	49.44	333 (4.91%)	175 (2.58%)	1,062 (15.67%)	381 (5.62%)	1,262 (18.62%)	172 (2.54%)	434 (6.40%)	90 (1.33%)
0	8	4,363	1,035	0	0	0	0	470	1,035	1,725	5.55	49.61	217 (4.97%)	146 (3.35%)	649 (14.88%)	234 (5.36%)	794 (18.20%)	129 (2.96%)	351 (8.04%)	106 (2.43%)
0	9	2,208	740	0	0	0	0	337	740	1,275	7.92	55.45	130 (5.89%)	133 (6.02%)	453 (20.52%)	165 (7.47%)	563 (25.50%)	76 (3.44%)	254 (11.50%)	107 (4.85%)
0	10	1,960	1,120	0	0	0	0	568	1,120	2,160	5.76	46.70	64 (3.27%)	80 (4.08%)	307 (15.66%)	114 (5.82%)	389 (19.85%)	67 (3.42%)	147 (7.50%)	71 (3.62%)
0	11	2,233	1,476	0	0	0	0	718	1,476	2,912	3.21	36.98	56 (2.51%)	41 (1.84%)	173 (7.75%)	79 (3.54%)	276 (12.36%)	59 (2.64%)	101 (4.52%)	82 (3.67%)
0	12	1,736	1,292	0	0	0	0	675	1,292	2,727	2.60	34.48	32 (1.84%)	22 (1.27%)	92 (5.30%)	41 (2.36%)	179 (10.31%)	45 (2.59%)	47 (2.71%)	40 (2.30%)
0	13	2,628	2,250	0	0	0	0	1,295	2,250	4,475	1.16	30.38	37 (1.41%)	13 (0.49%)	51 (1.94%)	22 (0.84%)	193 (7.34%)	28 (1.07%)	38 (1.45%)	31 (1.18%)
0	14	1,023	840	0	0	0	0	506	840	1,526	2.13	34.47	30 (2.93%)	8 (0.78%)	31 (3.03%)	15 (1.47%)	101 (9.87%)	17 (1.66%)	12 (1.17%)	20 (1.96%)
0	15	124	49	0	0	0	0	23	49	91	6.78	51.93	5 (4.03%)	5 (4.03%)	11 (8.87%)	6 (4.84%)	21 (16.94%)	5 (4.03%)	10 (8.06%)	11 (8.87%)
0	16	30	13	0	0	0	0	7	13	19	11.70	53.30	2 (6.67%)	2 (6.67%)	5 (16.67%)	2 (6.67%)	8 (26.67%)	3 (10.00%)	2 (6.67%)	5 (16.67%)
0	17	14	7	0	0	0	0	3	7	10	21.43	57.00	1 (7.14%)	0 (0.00%)	5 (35.71%)	2 (14.29%)	6 (42.86%)	5 (35.71%)	4 (28.57%)	3 (21.43%)

Continued on next page

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	6	2	0	0	0	0	0	2	5	32.83	74.33	0 (0.00%)	0 (0.00%)	4 (66.67%)	0 (0.00%)	4 (66.67%)	4 (66.67%)	1 (16.67%)	1 (16.67%)
0	19	1	1	0	0	0	0	1	1	1	13.00	74.00	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (100.00%)	0 (0.00%)
0	20	3	1	0	0	0	0	0	1	2	45.00	56.67	0 (0.00%)	0 (0.00%)	2 (66.67%)	0 (0.00%)	2 (66.67%)	0 (0.00%)	1 (33.33%)	1 (33.33%)
0	Total	73,536	18,969	0	0	0	0	9,226	18,969	32,833	3.71	39.68	3,062 (4.16%)	1,199 (1.63%)	7,223 (9.82%)	2,709 (3.68%)	9,022 (12.27%)	1,650 (2.24%)	3,202 (4.35%)	877 (1.19%)
1	1	98	15	98	1	8	10	6	15	20	175.55	27.77	2 (2.04%)	0 (0.00%)	5 (5.10%)	1 (1.02%)	5 (5.10%)	5 (5.10%)	1 (1.02%)	1 (1.02%)
1	2	3,283	496	3,283	12	157	221	187	496	759	2.99	28.08	219 (6.67%)	7 (0.21%)	145 (4.42%)	41 (1.25%)	194 (5.91%)	163 (4.96%)	51 (1.55%)	4 (0.12%)
1	3	7,872	1,613	7,872	32	485	566	651	1,613	2,527	3.16	40.78	728 (9.25%)	52 (0.66%)	959 (12.18%)	245 (3.11%)	1,174 (14.91%)	648 (8.23%)	432 (5.49%)	24 (0.30%)
1	4	8,353	2,397	8,353	54	735	899	943	2,397	3,934	5.50	52.42	821 (9.83%)	205 (2.45%)	2,083 (24.94%)	599 (7.17%)	2,432 (29.12%)	1,202 (14.39%)	1,123 (13.44%)	109 (1.30%)
1	5	9,762	3,004	9,762	78	1,015	1,263	1,163	3,004	5,180	5.38	51.24	933 (9.56%)	375 (3.84%)	2,593 (26.56%)	787 (8.06%)	3,013 (30.86%)	1,410 (14.44%)	1,433 (14.68%)	262 (2.68%)
1	6	16,071	4,380	16,071	132	1,699	2,216	1,532	4,380	7,536	4.06	43.22	1,395 (8.68%)	479 (2.98%)	2,913 (18.13%)	1,001 (6.23%)	3,552 (22.10%)	1,695 (10.55%)	1,571 (9.78%)	422 (2.63%)
1	7	20,819	5,449	20,819	173	2,235	2,935	1,919	5,449	9,323	4.02	44.00	2,016 (9.68%)	593 (2.85%)	3,506 (16.84%)	1,263 (6.07%)	4,483 (21.53%)	2,145 (10.30%)	1,831 (8.79%)	521 (2.50%)
1	8	18,405	5,351	18,405	132	2,488	3,299	1,857	5,351	9,370	4.62	49.31	2,072 (11.26%)	754 (4.10%)	4,020 (21.84%)	1,406 (7.64%)	4,983 (27.07%)	2,085 (11.33%)	2,208 (12.00%)	635 (3.45%)
1	9	14,561	5,058	14,561	149	2,516	3,366	1,645	5,058	9,324	6.10	56.59	1,958 (13.45%)	933 (6.41%)	4,471 (30.71%)	1,636 (11.24%)	5,263 (36.14%)	2,081 (14.29%)	2,641 (18.14%)	744 (5.11%)
1	10	11,877	5,027	11,877	120	2,591	3,616	1,603	5,027	9,480	6.96	61.78	1,690 (14.23%)	1,070 (9.01%)	4,451 (37.48%)	1,640 (13.81%)	5,086 (42.82%)	1,943 (16.36%)	2,798 (23.56%)	859 (7.23%)
1	11	10,548	5,312	10,548	151	2,576	3,825	1,856	5,312	10,412	7.01	62.82	1,475 (13.98%)	1,207 (11.44%)	4,143 (39.28%)	1,591 (15.08%)	4,752 (45.05%)	1,829 (17.34%)	2,770 (26.26%)	1,050 (9.95%)
1	12	11,418	7,276	11,418	195	2,856	4,349	2,913	7,276	14,406	5.84	56.49	1,333 (11.67%)	1,242 (10.88%)	3,698 (32.39%)	1,436 (12.58%)	4,469 (39.14%)	1,724 (15.10%)	2,415 (21.15%)	1,165 (10.20%)
1	13	11,010	7,730	11,010	167	2,863	4,598	3,296	7,730	15,859	4.83	52.72	1,305 (11.85%)	1,036 (9.41%)	3,118 (28.32%)	1,227 (11.14%)	3,965 (36.01%)	1,638 (14.88%)	2,120 (19.26%)	1,177 (10.69%)
1	14	8,387	6,073	8,387	145	2,712	4,368	2,640	6,073	13,044	4.83	53.26	1,178 (14.05%)	1,005 (11.98%)	2,440 (29.09%)	1,017 (12.13%)	3,276 (39.06%)	1,305 (15.56%)	1,790 (21.34%)	992 (11.83%)
1	15	5,908	4,317	5,908	130	2,599	4,664	1,702	4,317	9,798	5.25	56.12	968 (16.38%)	878 (14.86%)	1,982 (33.55%)	863 (14.61%)	2,583 (43.72%)	1,031 (17.45%)	1,522 (25.76%)	882 (14.93%)
1	16	3,888	2,844	3,888	92	2,045	3,996	1,082	2,844	6,982	6.04	58.32	739 (19.01%)	741 (19.06%)	1,479 (38.04%)	678 (17.44%)	1,809 (46.53%)	769 (19.78%)	1,090 (28.03%)	710 (18.26%)
1	17	2,657	1,969	2,657	67	1,698	3,958	664	1,969	5,134	5.87	60.52	625 (23.52%)	652 (24.54%)	1,099 (41.36%)	540 (20.32%)	1,388 (52.24%)	538 (20.25%)	903 (33.99%)	563 (21.19%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	1,849	1,420	1,849	78	1,276	3,447	486	1,420	4,051	6.56	61.20	481 (26.01%)	519 (28.07%)	853 (46.13%)	443 (23.96%)	1,020 (55.16%)	433 (23.42%)	720 (38.94%)	468 (25.31%)
1	19	1,467	1,146	1,467	73	1,069	3,346	400	1,146	3,631	6.44	59.79	425 (28.97%)	442 (30.13%)	670 (45.67%)	370 (25.22%)	822 (56.03%)	293 (19.97%)	569 (38.79%)	357 (24.34%)
1	20	1,943	1,671	1,943	179	1,629	8,201	601	1,671	7,611	7.21	56.95	836 (43.03%)	735 (37.83%)	950 (48.89%)	570 (29.34%)	1,137 (58.52%)	435 (22.39%)	836 (43.03%)	499 (25.68%)
1	Total	170,176	72,548	170,176	2,160	35,252	63,143	27,146	72,548	148,381	5.27	51.68	21,199 (12.46%)	12,925 (7.60%)	45,578 (26.78%)	17,354 (10.20%)	55,406 (32.56%)	23,372 (13.73%)	28,824 (16.94%)	11,444 (6.72%)

TABLE A.47: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Main* model (*Sample_1_train_half_3_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_3_test_half;
Submodel: Cond_Main

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	1	6,168	438	253	3	28	46	207	438	519	17.97	23.76	163 (2.64%)	6 (0.10%)	74 (1.20%)	39 (0.63%)	81 (1.31%)	47 (0.76%)	26 (0.42%)	3 (0.05%)
NA	2	18,365	2,534	2,602	18	237	274	1,053	2,534	3,646	3.64	32.03	577 (3.14%)	43 (0.23%)	656 (3.57%)	249 (1.36%)	765 (4.17%)	352 (1.92%)	363 (1.98%)	19 (0.10%)
NA	3	22,595	4,003	5,501	43	706	854	1,525	4,003	6,225	4.16	36.29	1,109 (4.91%)	168 (0.74%)	1,408 (6.23%)	626 (2.77%)	1,616 (7.15%)	812 (3.59%)	1,029 (4.55%)	98 (0.43%)
NA	4	26,379	5,372	6,221	86	920	1,174	1,999	5,372	8,467	4.88	40.41	1,357 (5.14%)	348 (1.32%)	2,277 (8.63%)	972 (3.68%)	2,675 (10.14%)	1,251 (4.74%)	1,644 (6.23%)	223 (0.85%)
NA	5	33,273	7,574	7,852	134	1,378	1,833	2,702	7,574	12,333	5.04	39.10	1,895 (5.70%)	543 (1.63%)	2,700 (8.11%)	1,296 (3.90%)	3,333 (10.02%)	1,368 (4.11%)	1,962 (5.90%)	367 (1.10%)
NA	6	35,252	8,249	12,291	156	2,062	2,749	2,912	8,249	13,704	5.43	43.45	2,075 (5.89%)	755 (2.14%)	3,244 (9.20%)	1,484 (4.21%)	4,193 (11.89%)	1,432 (4.06%)	2,375 (6.74%)	576 (1.63%)
NA	7	33,259	8,893	17,310	217	2,844	3,804	2,892	8,893	15,225	6.45	47.78	2,174 (6.54%)	958 (2.88%)	3,685 (11.08%)	1,808 (5.44%)	4,957 (14.90%)	1,574 (4.73%)	3,042 (9.15%)	781 (2.35%)
NA	8	28,764	9,321	18,246	251	3,702	5,028	3,189	9,321	16,194	7.89	52.64	2,044 (7.11%)	1,199 (4.17%)	4,256 (14.80%)	1,870 (6.50%)	5,727 (19.91%)	1,842 (6.40%)	3,647 (12.68%)	928 (3.23%)
NA	9	22,336	9,625	14,412	181	3,591	4,871	3,407	9,625	17,242	9.34	60.19	1,832 (8.20%)	1,477 (6.61%)	4,392 (19.66%)	1,992 (8.92%)	5,790 (25.92%)	1,822 (8.16%)	3,848 (17.23%)	1,125 (5.04%)
NA	10	17,034	8,997	11,905	214	3,497	5,084	3,117	8,997	16,713	9.84	62.62	1,437 (8.44%)	1,553 (9.12%)	4,079 (23.95%)	1,986 (11.66%)	5,249 (30.81%)	1,628 (9.56%)	3,631 (21.32%)	1,457 (8.55%)
NA	11	16,318	10,149	9,984	221	3,536	5,284	3,805	10,149	20,231	7.77	56.86	1,131 (6.93%)	1,504 (9.22%)	3,349 (20.52%)	1,699 (10.41%)	4,393 (26.92%)	1,470 (9.01%)	3,160 (19.37%)	1,585 (9.71%)
NA	12	15,429	11,087	9,183	206	3,394	5,297	4,756	11,087	22,743	6.03	50.64	963 (6.24%)	1,333 (8.64%)	2,467 (15.99%)	1,336 (8.66%)	3,622 (23.48%)	1,200 (7.78%)	2,566 (16.63%)	1,521 (9.86%)
NA	13	11,726	8,644	10,347	285	3,685	6,002	3,692	8,644	18,172	5.96	50.26	852 (7.27%)	1,188 (10.13%)	1,982 (16.90%)	1,032 (8.80%)	2,958 (25.23%)	1,006 (8.58%)	1,978 (16.87%)	1,424 (12.14%)
NA	14	8,765	6,899	8,429	208	3,641	6,155	3,173	6,899	15,416	5.16	48.20	744 (8.49%)	924 (10.54%)	1,371 (15.64%)	808 (9.22%)	2,254 (25.72%)	763 (8.71%)	1,385 (15.80%)	1,047 (11.95%)
NA	15	4,100	3,080	3,950	89	2,485	4,680	1,225	3,080	7,617	7.23	54.89	509 (12.41%)	639 (15.59%)	878 (21.41%)	528 (12.88%)	1,324 (32.29%)	459 (11.20%)	918 (22.39%)	744 (18.15%)
NA	16	2,031	1,583	1,979	54	1,554	3,670	563	1,583	4,378	8.58	60.30	319 (15.71%)	452 (22.26%)	569 (28.02%)	321 (15.81%)	792 (39.00%)	262 (12.90%)	585 (28.80%)	458 (22.55%)
NA	17	1,265	1,009	1,240	52	1,064	3,166	316	1,009	3,101	9.75	62.76	263 (20.79%)	381 (30.12%)	436 (34.47%)	282 (22.29%)	567 (44.82%)	213 (16.84%)	481 (38.02%)	361 (28.54%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
NA	18	768	648	761	34	681	2,547	238	648	2,395	10.95	62.73	172 (22.40%)	269 (35.03%)	277 (36.07%)	205 (26.69%)	347 (45.18%)	101 (13.15%)	298 (38.80%)	224 (29.17%)
NA	19	516	456	512	33	488	2,231	146	456	1,899	9.08	59.14	144 (27.91%)	204 (39.53%)	187 (36.24%)	127 (24.61%)	243 (47.09%)	93 (18.02%)	209 (40.50%)	132 (25.58%)
NA	20	545	484	535	82	514	3,879	193	484	2,758	14.65	57.37	163 (29.91%)	227 (41.65%)	221 (40.55%)	152 (27.89%)	264 (48.44%)	70 (12.84%)	247 (45.32%)	133 (24.40%)
NA	Total	304,888	109,045	143,513	2,567	40,007	68,628	41,110	109,045	208,978	6.53	46.65	19,923 (6.53%)	14,171 (4.65%)	38,508 (12.63%)	18,812 (6.17%)	51,150 (16.78%)	17,765 (5.83%)	33,394 (10.95%)	13,206 (4.33%)

TABLE A.48: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Age-65p* model (*Sample_1_train_half_3_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_3_test_half;
Submodel: Cond_Age-65p

If[a]	#[b]	TP+FP[o]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	15,250	1,706	1,617	14	168	206	745	1,706	2,428	7.59	29.54	400 (2.62%)	24 (0.16%)	378 (2.48%)	136 (0.89%)	461 (3.02%)	227 (1.49%)	220 (1.44%)	11 (0.07%)
0	2	18,632	3,557	6,367	51	797	966	1,369	3,557	5,705	4.92	34.35	918 (4.93%)	119 (0.64%)	842 (4.52%)	439 (2.36%)	1,034 (5.55%)	656 (3.52%)	725 (3.89%)	49 (0.26%)
0	3	15,446	2,551	3,130	63	723	959	923	2,551	4,069	4.35	30.04	908 (5.88%)	141 (0.91%)	715 (4.63%)	392 (2.54%)	910 (5.89%)	507 (3.28%)	611 (3.96%)	63 (0.41%)
0	4	19,018	2,801	3,965	64	771	1,081	1,065	2,801	4,460	3.68	32.31	1,029 (5.41%)	158 (0.83%)	726 (3.82%)	483 (2.54%)	1,077 (5.66%)	343 (1.80%)	557 (2.93%)	84 (0.44%)
0	5	30,838	5,350	6,190	102	1,095	1,556	2,005	5,350	8,517	2.82	28.64	1,494 (4.84%)	228 (0.74%)	955 (3.10%)	583 (1.89%)	1,547 (5.02%)	438 (1.42%)	596 (1.93%)	82 (0.27%)
0	6	31,583	6,230	13,619	153	1,781	2,458	2,320	6,230	10,189	3.84	33.28	1,911 (6.05%)	288 (0.91%)	1,248 (3.95%)	760 (2.41%)	2,100 (6.65%)	544 (1.72%)	1,052 (3.33%)	115 (0.36%)
0	7	21,345	5,071	13,814	174	2,681	3,532	1,751	5,071	8,833	5.51	36.11	1,755 (8.22%)	295 (1.38%)	1,211 (5.67%)	790 (3.70%)	2,036 (9.54%)	720 (3.37%)	1,119 (5.24%)	121 (0.57%)
0	8	10,450	3,358	8,293	120	2,398	3,168	1,077	3,358	6,203	6.92	39.49	1,371 (13.12%)	337 (3.22%)	1,051 (10.06%)	839 (8.03%)	1,523 (14.57%)	659 (6.31%)	954 (9.13%)	131 (1.25%)
0	9	6,328	2,698	4,913	75	1,832	2,642	890	2,698	5,196	6.76	41.06	889 (14.05%)	279 (4.41%)	756 (11.95%)	577 (9.12%)	1,037 (16.39%)	504 (7.96%)	760 (12.01%)	128 (2.02%)
0	10	5,157	2,854	3,203	54	1,419	2,304	1,129	2,854	5,581	5.41	39.49	565 (10.96%)	249 (4.83%)	633 (12.27%)	502 (9.73%)	868 (16.83%)	393 (7.62%)	618 (11.98%)	133 (2.58%)
0	11	4,601	3,175	2,457	65	1,171	2,050	1,524	3,175	6,400	4.15	36.65	371 (8.06%)	194 (4.22%)	423 (9.19%)	369 (8.02%)	709 (15.41%)	295 (6.41%)	440 (9.56%)	181 (3.93%)
0	12	6,400	4,602	2,427	106	1,129	2,101	2,009	4,602	9,707	2.57	33.08	312 (4.88%)	162 (2.53%)	358 (5.59%)	264 (4.13%)	653 (10.20%)	287 (4.48%)	354 (5.53%)	155 (2.42%)
0	13	9,118	7,352	3,421	96	1,034	1,949	3,758	7,352	15,600	1.56	30.93	270 (2.96%)	153 (1.68%)	253 (2.77%)	184 (2.02%)	754 (8.27%)	263 (2.88%)	237 (2.60%)	120 (1.32%)
0	14	8,911	7,552	6,038	186	1,577	2,843	3,738	7,552	16,360	1.42	30.32	331 (3.71%)	117 (1.31%)	187 (2.10%)	156 (1.75%)	760 (8.53%)	264 (2.96%)	174 (1.95%)	104 (1.17%)
0	15	5,903	5,142	5,704	160	2,026	3,431	2,754	5,142	11,815	1.67	30.70	369 (6.25%)	93 (1.58%)	155 (2.63%)	143 (2.42%)	704 (11.93%)	308 (5.22%)	140 (2.37%)	89 (1.51%)
0	16	2,149	1,852	2,111	48	1,226	2,405	868	1,852	5,102	2.66	32.20	302 (14.05%)	66 (3.07%)	130 (6.05%)	118 (5.49%)	342 (15.91%)	201 (9.35%)	95 (4.42%)	45 (2.09%)
0	17	1,001	849	986	33	752	1,959	374	849	2,679	3.65	35.36	209 (20.88%)	69 (6.89%)	97 (9.69%)	81 (8.09%)	207 (20.68%)	127 (12.69%)	101 (10.09%)	56 (5.59%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	558	469	554	26	479	1,690	187	469	1,811	4.29	38.22	111 (19.89%)	53 (9.50%)	103 (18.46%)	84 (15.05%)	157 (28.14%)	82 (14.70%)	82 (14.70%)	44 (7.89%)
0	19	379	330	373	27	328	1,482	126	330	1,456	9.64	39.70	94 (24.80%)	67 (17.68%)	59 (15.57%)	66 (17.41%)	90 (23.75%)	60 (15.83%)	61 (16.09%)	30 (7.92%)
0	20	452	391	438	53	402	2,932	157	391	2,272	13.78	42.04	115 (25.44%)	89 (19.69%)	104 (23.01%)	114 (25.22%)	143 (31.64%)	66 (14.60%)	111 (24.56%)	51 (11.28%)
0	Total	213,519	67,890	89,620	1,670	23,789	41,714	28,769	67,890	134,383	4.25	32.90	13,724 (6.43%)	3,181 (1.49%)	10,384 (4.86%)	7,080 (3.32%)	17,112 (8.01%)	6,944 (3.25%)	9,007 (4.22%)	1,792 (0.84%)
1	1	14	1	6	0	2	3	0	1	1	1,210.21	78.33	0 (0.00%)	0 (0.00%)	3 (21.43%)	0 (0.00%)	3 (21.43%)	0 (0.00%)	3 (21.43%)	0 (0.00%)
1	2	276	8	4	0	0	0	4	8	14	8.45	93.40	9 (3.26%)	0 (0.00%)	4 (1.45%)	3 (1.09%)	8 (2.90%)	4 (1.45%)	0 (0.00%)	0 (0.00%)
1	3	825	54	25	0	5	6	22	54	78	5.85	80.56	6 (0.73%)	1 (0.12%)	44 (5.33%)	12 (1.45%)	47 (5.70%)	24 (2.91%)	12 (1.45%)	0 (0.00%)
1	4	2,808	457	155	3	17	21	175	457	646	3.68	76.01	81 (2.88%)	59 (2.10%)	525 (18.70%)	133 (4.74%)	563 (20.05%)	79 (2.81%)	263 (9.37%)	19 (0.68%)
1	5	3,882	905	657	8	75	86	336	905	1,320	7.37	76.75	125 (3.22%)	152 (3.92%)	837 (21.56%)	254 (6.54%)	921 (23.72%)	228 (5.87%)	503 (12.96%)	96 (2.47%)
1	6	8,526	2,523	1,586	11	217	272	899	2,523	3,984	7.10	76.93	265 (3.11%)	315 (3.69%)	1,976 (23.18%)	614 (7.20%)	2,215 (25.98%)	498 (5.84%)	1,150 (13.49%)	208 (2.44%)
1	7	10,490	3,705	3,690	36	482	593	1,275	3,705	5,822	9.74	78.13	433 (4.13%)	609 (5.81%)	2,686 (25.61%)	919 (8.76%)	3,150 (30.03%)	810 (7.72%)	1,825 (17.40%)	490 (4.67%)
1	8	12,556	5,029	7,080	51	964	1,207	1,630	5,029	8,276	12.01	78.25	651 (5.18%)	760 (6.05%)	3,641 (29.00%)	1,154 (9.19%)	4,184 (33.32%)	1,463 (11.65%)	2,820 (22.46%)	702 (5.59%)
1	9	16,746	7,828	9,580	113	1,620	2,059	2,399	7,828	13,391	14.54	79.22	910 (5.43%)	1,284 (7.67%)	5,047 (30.14%)	1,940 (11.58%)	6,149 (36.72%)	2,116 (12.64%)	4,250 (25.38%)	1,258 (7.51%)
1	10	12,998	6,638	9,564	130	2,252	2,923	1,856	6,638	11,747	13.66	80.33	924 (7.11%)	1,668 (12.83%)	4,349 (33.46%)	1,978 (15.22%)	5,435 (41.81%)	1,876 (14.43%)	3,980 (30.62%)	1,988 (15.29%)
1	11	8,697	4,853	8,065	130	2,533	3,368	1,323	4,853	8,848	13.78	80.91	763 (8.77%)	1,527 (17.56%)	3,206 (36.86%)	1,618 (18.60%)	4,036 (46.41%)	1,365 (15.70%)	3,177 (36.53%)	1,836 (21.11%)
1	12	5,589	3,448	5,536	101	2,251	3,097	902	3,448	6,488	13.16	81.37	623 (11.15%)	1,330 (23.80%)	2,190 (39.18%)	1,103 (19.74%)	2,832 (50.67%)	921 (16.48%)	2,309 (41.31%)	1,605 (28.72%)
1	13	3,348	2,194	3,332	79	1,926	2,947	528	2,194	4,391	12.44	81.43	453 (13.53%)	1,069 (31.93%)	1,409 (42.08%)	727 (21.71%)	1,806 (53.94%)	585 (17.47%)	1,513 (45.19%)	1,211 (36.17%)
1	14	1,828	1,266	1,827	61	1,374	2,496	351	1,266	2,906	12.31	81.00	289 (15.81%)	707 (38.68%)	818 (44.75%)	451 (24.67%)	1,005 (54.98%)	304 (16.63%)	927 (50.71%)	738 (40.37%)
1	15	1,089	822	1,089	44	914	1,974	224	822	2,022	11.80	80.48	209 (19.19%)	503 (46.19%)	521 (47.84%)	318 (29.20%)	633 (58.13%)	201 (18.46%)	600 (55.10%)	475 (43.62%)
1	16	648	513	648	35	590	1,577	127	513	1,367	10.51	79.72	150 (23.15%)	343 (52.93%)	307 (47.38%)	174 (26.85%)	384 (59.26%)	125 (19.29%)	378 (58.33%)	285 (43.98%)
1	17	433	359	433	22	402	1,282	106	359	1,041	9.74	78.80	118 (27.25%)	240 (55.43%)	224 (51.73%)	144 (33.26%)	273 (63.05%)	85 (19.63%)	268 (61.89%)	213 (49.19%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	255	219	255	17	245	927	59	219	750	10.29	79.39	67 (26.27%)	156 (61.18%)	137 (53.73%)	89 (34.90%)	161 (63.14%)	53 (20.78%)	162 (63.53%)	129 (50.59%)
1	19	187	169	187	21	181	828	54	169	604	7.65	77.53	62 (33.16%)	141 (75.40%)	101 (54.01%)	53 (28.34%)	120 (64.17%)	42 (22.46%)	119 (63.64%)	80 (42.78%)
1	20	174	164	174	35	168	1,248	71	164	899	8.00	76.33	61 (35.06%)	126 (72.41%)	99 (56.90%)	48 (27.59%)	113 (64.94%)	42 (24.14%)	128 (73.56%)	81 (46.55%)
1	Total	91,369	41,155	53,893	897	16,218	26,914	12,341	41,155	74,595	11.87	79.15	6,199 (6.78%)	10,990 (12.03%)	28,124 (30.78%)	11,732 (12.84%)	34,038 (37.25%)	10,821 (11.84%)	24,387 (26.69%)	11,414 (12.49%)

TABLE A.49: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Acute-12-month* model (*Sample_1_train_half_3_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_3_test_half;
Submodel: Cond_Prior-Acute-12-month

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	5,822	403	163	3	10	32	188	403	475	11.18	23.74	155 (2.66%)	4 (0.07%)	55 (0.94%)	32 (0.55%)	61 (1.05%)	33 (0.57%)	18 (0.31%)	2 (0.03%)
0	2	16,300	2,113	1,298	14	162	188	894	2,113	3,003	3.29	31.22	490 (3.01%)	41 (0.25%)	560 (3.44%)	217 (1.33%)	638 (3.91%)	272 (1.67%)	291 (1.79%)	18 (0.11%)
0	3	21,447	3,366	2,759	16	394	450	1,366	3,366	4,987	3.58	35.64	970 (4.52%)	125 (0.58%)	1,206 (5.62%)	535 (2.49%)	1,351 (6.30%)	574 (2.68%)	759 (3.54%)	62 (0.29%)
0	4	20,170	3,867	3,258	42	599	757	1,475	3,867	5,994	4.91	43.81	1,012 (5.02%)	301 (1.49%)	1,874 (9.29%)	759 (3.76%)	2,154 (10.68%)	875 (4.34%)	1,230 (6.10%)	139 (0.69%)
0	5	30,009	6,585	4,060	64	746	945	2,489	6,585	10,648	4.63	36.70	1,530 (5.10%)	403 (1.34%)	2,138 (7.12%)	1,013 (3.38%)	2,579 (8.59%)	981 (3.27%)	1,511 (5.04%)	239 (0.80%)
0	6	26,267	6,311	5,876	100	1,286	1,600	2,277	6,311	10,197	5.76	47.03	1,527 (5.81%)	577 (2.20%)	2,566 (9.77%)	1,188 (4.52%)	3,152 (12.00%)	1,031 (3.93%)	1,805 (6.87%)	338 (1.29%)
0	7	23,157	6,044	6,528	120	1,551	1,981	2,117	6,044	9,926	6.72	51.56	1,180 (5.10%)	727 (3.14%)	2,874 (12.41%)	1,287 (5.56%)	3,591 (15.51%)	1,065 (4.60%)	2,270 (9.80%)	498 (2.15%)
0	8	19,812	6,198	6,911	166	1,909	2,449	2,322	6,198	10,586	7.72	54.81	1,043 (5.26%)	835 (4.21%)	2,769 (13.98%)	1,196 (6.04%)	3,617 (18.26%)	1,193 (6.02%)	2,221 (11.21%)	542 (2.74%)
0	9	13,811	5,232	6,071	148	2,099	2,720	1,918	5,232	9,061	10.03	61.66	827 (5.99%)	863 (6.25%)	2,539 (18.38%)	1,190 (8.62%)	3,404 (24.65%)	1,211 (8.77%)	2,022 (14.64%)	653 (4.73%)
0	10	9,947	5,130	4,381	163	2,025	2,720	2,160	5,130	9,274	9.48	59.62	482 (4.85%)	693 (6.97%)	1,861 (18.71%)	887 (8.92%)	2,577 (25.91%)	1,006 (10.11%)	1,613 (16.22%)	739 (7.43%)
0	11	8,857	5,482	3,416	124	1,843	2,597	2,171	5,482	10,691	7.15	51.82	330 (3.73%)	506 (5.71%)	1,207 (13.63%)	679 (7.67%)	1,745 (19.70%)	870 (9.82%)	1,148 (12.96%)	700 (7.90%)
0	12	7,051	4,888	2,782	129	1,507	2,309	2,214	4,888	9,881	5.25	46.01	239 (3.39%)	356 (5.05%)	757 (10.74%)	410 (5.81%)	1,172 (16.62%)	597 (8.47%)	776 (11.01%)	505 (7.16%)
0	13	7,609	6,133	2,697	151	1,684	2,708	3,156	6,133	12,559	3.53	39.71	203 (2.67%)	266 (3.50%)	543 (7.14%)	308 (4.05%)	1,037 (13.63%)	473 (6.22%)	544 (7.15%)	426 (5.60%)
0	14	3,345	2,545	2,356	140	1,583	2,542	1,300	2,545	5,358	5.35	46.17	144 (4.30%)	204 (6.10%)	368 (11.00%)	203 (6.07%)	638 (19.07%)	329 (9.84%)	366 (10.94%)	282 (8.43%)
0	15	1,522	1,066	1,314	50	1,113	2,055	523	1,066	2,298	6.77	52.09	93 (6.11%)	122 (8.02%)	202 (13.27%)	126 (8.28%)	331 (21.75%)	238 (15.64%)	235 (15.44%)	172 (11.30%)
0	16	674	465	609	35	553	1,282	169	465	1,022	11.70	61.83	45 (6.68%)	74 (10.98%)	152 (22.55%)	82 (12.17%)	207 (30.71%)	163 (24.18%)	158 (23.44%)	133 (19.73%)
0	17	336	230	301	21	270	808	80	230	523	17.29	69.73	29 (8.63%)	56 (16.67%)	96 (28.57%)	63 (18.75%)	124 (36.90%)	103 (30.65%)	84 (25.00%)	79 (23.51%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	156	104	151	5	143	473	33	104	232	19.81	71.56	12 (7.69%)	36 (23.08%)	46 (29.49%)	29 (18.59%)	66 (42.31%)	44 (28.21%)	53 (33.97%)	49 (31.41%)
0	19	94	61	90	5	79	340	16	61	214	27.41	70.41	10 (10.64%)	20 (21.28%)	31 (32.98%)	21 (22.34%)	37 (39.36%)	40 (42.55%)	32 (34.04%)	27 (28.72%)
0	20	62	37	52	5	45	254	16	37	120	58.97	62.24	4 (6.45%)	16 (25.81%)	22 (35.48%)	12 (19.35%)	25 (40.32%)	27 (43.55%)	17 (27.42%)	7 (11.29%)
0	Total	216,448	66,260	55,073	1,501	19,601	29,210	26,884	66,260	117,049	6.03	45.08	10,325 (4.77%)	6,225 (2.88%)	21,866 (10.10%)	10,237 (4.73%)	28,506 (13.17%)	11,125 (5.14%)	17,153 (7.92%)	5,610 (2.59%)
1	1	119	16	119	0	13	19	8	16	21	384.01	25.27	4 (3.36%)	0 (0.00%)	4 (3.36%)	2 (1.68%)	4 (3.36%)	0 (0.00%)	3 (2.52%)	0 (0.00%)
1	2	1,529	277	1,529	2	37	39	98	277	473	6.28	22.69	99 (6.47%)	2 (0.13%)	29 (1.90%)	17 (1.11%)	50 (3.27%)	23 (1.50%)	15 (0.98%)	1 (0.07%)
1	3	2,356	653	2,356	2	152	175	193	653	1,122	7.41	37.76	208 (8.83%)	27 (1.15%)	184 (7.81%)	88 (3.74%)	238 (10.10%)	109 (4.63%)	158 (6.71%)	15 (0.64%)
1	4	3,212	961	3,212	12	242	289	278	961	1,680	8.44	44.95	282 (8.78%)	86 (2.68%)	443 (13.79%)	187 (5.82%)	556 (17.31%)	243 (7.57%)	393 (12.24%)	71 (2.21%)
1	5	5,385	1,383	5,385	11	417	504	390	1,383	2,360	5.80	41.04	429 (7.97%)	153 (2.84%)	663 (12.31%)	307 (5.70%)	903 (16.77%)	372 (6.91%)	527 (9.79%)	176 (3.27%)
1	6	10,332	2,384	10,332	24	607	735	695	2,384	4,213	4.90	35.22	809 (7.83%)	192 (1.86%)	700 (6.78%)	387 (3.75%)	1,207 (11.68%)	423 (4.09%)	594 (5.75%)	235 (2.27%)
1	7	8,654	2,423	8,654	49	1,081	1,266	669	2,423	4,295	6.25	39.45	890 (10.28%)	233 (2.69%)	843 (9.74%)	507 (5.86%)	1,342 (15.51%)	469 (5.42%)	715 (8.26%)	258 (2.98%)
1	8	7,070	2,638	7,070	58	1,308	1,662	759	2,638	4,753	8.00	47.26	820 (11.60%)	299 (4.23%)	1,137 (16.08%)	520 (7.36%)	1,554 (21.98%)	477 (6.75%)	1,068 (15.11%)	331 (4.68%)
1	9	6,800	2,951	6,800	51	1,273	1,791	811	2,951	5,585	8.97	56.80	847 (12.46%)	471 (6.93%)	1,579 (23.22%)	721 (10.60%)	1,946 (28.62%)	491 (7.22%)	1,577 (23.19%)	420 (6.18%)
1	10	7,397	3,794	7,397	34	1,420	2,054	1,043	3,794	7,081	9.69	63.31	856 (11.57%)	741 (10.02%)	2,059 (27.84%)	925 (12.51%)	2,546 (34.42%)	650 (8.79%)	1,944 (26.28%)	590 (7.98%)
1	11	7,447	4,272	7,447	50	1,636	2,395	1,283	4,272	8,699	9.14	66.04	836 (11.23%)	979 (13.15%)	2,137 (28.70%)	1,038 (13.94%)	2,691 (36.14%)	737 (9.90%)	2,123 (28.51%)	904 (12.14%)
1	12	8,515	5,823	8,515	72	1,896	2,729	2,155	5,823	11,739	7.15	56.77	805 (9.45%)	1,034 (12.14%)	1,929 (22.65%)	994 (11.67%)	2,632 (30.91%)	725 (8.51%)	1,987 (23.34%)	1,045 (12.27%)
1	13	7,687	5,815	7,687	97	2,201	3,330	2,380	5,815	12,204	6.19	52.75	745 (9.69%)	943 (12.27%)	1,545 (20.10%)	804 (10.46%)	2,283 (29.70%)	613 (7.97%)	1,534 (19.96%)	990 (12.88%)
1	14	4,800	3,608	4,800	104	2,361	3,740	1,392	3,608	8,635	6.55	56.85	570 (11.88%)	827 (17.23%)	1,110 (23.13%)	655 (13.65%)	1,636 (34.08%)	446 (9.29%)	1,169 (24.35%)	848 (17.67%)
1	15	2,729	2,049	2,729	109	1,903	3,694	746	2,049	5,420	7.66	59.33	389 (14.25%)	543 (19.90%)	772 (28.29%)	438 (16.05%)	1,100 (40.31%)	298 (10.92%)	792 (29.02%)	566 (20.74%)
1	16	1,663	1,344	1,663	94	1,342	3,332	439	1,344	3,671	7.45	60.50	322 (19.36%)	441 (26.52%)	519 (31.21%)	316 (19.00%)	688 (41.37%)	202 (12.15%)	559 (33.61%)	420 (25.26%)
1	17	1,023	854	1,023	61	889	2,701	289	854	2,806	7.61	60.66	227 (22.19%)	336 (32.84%)	351 (34.31%)	238 (23.26%)	472 (46.14%)	143 (13.98%)	392 (38.32%)	277 (27.08%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	709	621	709	63	654	2,534	233	621	2,248	8.25	59.96	166 (23.41%)	242 (34.13%)	258 (36.39%)	170 (23.98%)	324 (45.70%)	89 (12.55%)	269 (37.94%)	199 (28.07%)
1	19	492	439	492	58	471	2,269	146	439	1,868	6.81	57.87	131 (26.63%)	189 (38.41%)	193 (39.23%)	127 (25.81%)	243 (49.39%)	74 (15.04%)	200 (40.65%)	121 (24.59%)
1	20	521	480	521	115	503	4,160	219	480	3,056	8.76	54.52	163 (31.29%)	208 (39.92%)	187 (35.89%)	134 (25.72%)	229 (43.95%)	56 (10.75%)	222 (42.61%)	129 (24.76%)
1	Total	88,440	42,785	88,440	1,066	20,406	39,418	14,226	42,785	91,929	7.78	50.47	9,598 (10.85%)	7,946 (8.98%)	16,642 (18.82%)	8,575 (9.70%)	22,644 (25.60%)	6,640 (7.51%)	16,241 (18.36%)	7,596 (8.59%)

TABLE A.50: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Oper-12-month* model (*Sample_1_train_half_3_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_3_test_half;
Submodel: Cond_Prior-Oper-12-month

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	92	11	43	1	17	21	2	11	16	936.23	47.22	1 (1.09%)	0 (0.00%)	0 (0.00%)	1 (1.09%)	0 (0.00%)	0 (0.00%)	1 (1.09%)	0 (0.00%)
0	2	3,089	231	9	0	3	5	126	231	285	3.97	16.73	54 (1.75%)	3 (0.10%)	13 (0.42%)	10 (0.32%)	13 (0.42%)	5 (0.16%)	2 (0.06%)	0 (0.00%)
0	3	12,806	1,345	166	0	12	13	575	1,345	1,769	2.61	29.37	619 (4.83%)	36 (0.28%)	393 (3.07%)	222 (1.73%)	436 (3.40%)	39 (0.30%)	174 (1.36%)	7 (0.05%)
0	4	17,017	2,817	737	5	119	133	1,149	2,817	4,126	2.81	30.41	867 (5.09%)	178 (1.05%)	815 (4.79%)	407 (2.39%)	927 (5.45%)	84 (0.49%)	532 (3.13%)	43 (0.25%)
0	5	19,608	4,171	1,120	10	213	252	1,595	4,171	6,611	3.99	38.26	1,211 (6.18%)	240 (1.22%)	1,190 (6.07%)	729 (3.72%)	1,421 (7.25%)	141 (0.72%)	748 (3.81%)	131 (0.67%)
0	6	15,882	4,254	2,553	20	397	457	1,542	4,254	6,996	6.21	52.39	1,267 (7.98%)	479 (3.02%)	2,035 (12.81%)	1,032 (6.50%)	2,414 (15.20%)	199 (1.25%)	1,414 (8.90%)	264 (1.66%)
0	7	11,648	4,061	3,197	40	770	931	1,335	4,061	7,031	9.60	56.53	1,017 (8.73%)	570 (4.89%)	1,839 (15.79%)	948 (8.14%)	2,259 (19.39%)	238 (2.04%)	1,534 (13.17%)	367 (3.15%)
0	8	9,491	3,862	3,248	35	885	1,056	1,242	3,862	6,830	12.63	63.72	779 (8.21%)	602 (6.34%)	1,813 (19.10%)	884 (9.31%)	2,310 (24.34%)	300 (3.16%)	1,655 (17.44%)	463 (4.88%)
0	9	7,385	3,532	3,247	56	1,013	1,273	1,123	3,532	6,216	13.36	69.96	564 (7.64%)	706 (9.56%)	1,729 (23.41%)	877 (11.88%)	2,272 (30.77%)	264 (3.57%)	1,548 (20.96%)	602 (8.15%)
0	10	5,165	2,677	3,288	63	1,039	1,333	749	2,677	4,803	12.61	71.34	462 (8.94%)	612 (11.85%)	1,261 (24.41%)	698 (13.51%)	1,706 (33.03%)	214 (4.14%)	1,279 (24.76%)	725 (14.04%)
0	11	4,355	2,461	2,833	60	1,121	1,468	838	2,461	4,499	9.87	63.18	332 (7.62%)	536 (12.31%)	954 (21.91%)	476 (10.93%)	1,314 (30.17%)	151 (3.47%)	951 (21.84%)	583 (13.39%)
0	12	2,846	1,636	2,190	42	1,046	1,486	509	1,636	3,246	9.66	66.98	257 (9.03%)	497 (17.46%)	700 (24.60%)	398 (13.98%)	958 (33.66%)	109 (3.83%)	749 (26.32%)	517 (18.17%)
0	13	2,117	1,355	1,498	20	863	1,272	463	1,355	2,822	8.85	60.58	169 (7.98%)	433 (20.45%)	487 (23.00%)	278 (13.13%)	654 (30.89%)	64 (3.02%)	501 (23.67%)	394 (18.61%)
0	14	4,630	3,348	971	20	667	1,041	1,313	3,348	7,202	2.68	37.41	140 (3.02%)	325 (7.02%)	293 (6.33%)	183 (3.95%)	457 (9.87%)	68 (1.47%)	326 (7.04%)	274 (5.92%)
0	15	2,067	1,351	690	20	478	843	537	1,351	2,996	3.95	43.90	119 (5.76%)	213 (10.30%)	191 (9.24%)	119 (5.76%)	301 (14.56%)	60 (2.90%)	208 (10.06%)	166 (8.03%)
0	16	1,025	673	755	17	360	699	231	673	1,699	4.37	44.21	89 (8.68%)	128 (12.49%)	121 (11.80%)	58 (5.66%)	170 (16.59%)	27 (2.63%)	144 (14.05%)	103 (10.05%)
0	17	2,442	2,024	701	30	494	887	1,163	2,024	4,176	1.28	31.82	73 (2.99%)	104 (4.26%)	66 (2.70%)	49 (2.01%)	174 (7.13%)	26 (1.06%)	90 (3.69%)	67 (2.74%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	1,870	1,620	471	56	347	905	917	1,620	3,298	1.63	32.37	68 (3.64%)	87 (4.65%)	70 (3.74%)	49 (2.62%)	183 (9.79%)	23 (1.23%)	84 (4.49%)	62 (3.32%)
0	19	7,394	6,728	962	107	551	1,220	3,805	6,728	14,165	0.80	29.25	131 (1.77%)	61 (0.82%)	43 (0.58%)	32 (0.43%)	467 (6.32%)	16 (0.22%)	44 (0.60%)	31 (0.42%)
0	20	1,601	1,473	1,519	106	1,090	2,476	894	1,473	3,850	2.43	33.80	105 (6.56%)	83 (5.18%)	64 (4.00%)	59 (3.69%)	182 (11.37%)	18 (1.12%)	73 (4.56%)	45 (2.81%)
0	Total	132,530	49,630	30,198	708	11,485	17,771	20,108	49,630	92,636	6.66	45.21	8,324 (6.28%)	5,893 (4.45%)	14,077 (10.62%)	7,509 (5.67%)	18,618 (14.05%)	2,046 (1.54%)	12,057 (9.10%)	4,844 (3.66%)
1	1	2,850	189	40	0	5	6	96	189	216	8.96	18.82	52 (1.82%)	2 (0.07%)	20 (0.70%)	12 (0.42%)	23 (0.81%)	14 (0.49%)	6 (0.21%)	0 (0.00%)
1	2	15,376	1,650	1,266	14	158	182	694	1,650	2,264	1.82	28.32	403 (2.62%)	18 (0.12%)	319 (2.07%)	121 (0.79%)	369 (2.40%)	165 (1.07%)	124 (0.81%)	6 (0.04%)
1	3	18,521	2,770	4,667	31	664	757	1,067	2,770	4,267	2.98	37.26	638 (3.44%)	65 (0.35%)	822 (4.44%)	308 (1.66%)	1,069 (5.77%)	463 (2.50%)	491 (2.65%)	28 (0.15%)
1	4	20,111	3,647	6,674	86	994	1,247	1,371	3,647	5,668	4.28	40.41	687 (3.42%)	163 (0.81%)	1,275 (6.34%)	506 (2.52%)	1,813 (9.01%)	800 (3.98%)	938 (4.66%)	99 (0.49%)
1	5	18,295	4,145	11,254	166	1,901	2,438	1,545	4,145	6,882	5.89	44.49	801 (4.38%)	239 (1.31%)	1,681 (9.19%)	633 (3.46%)	2,385 (13.04%)	1,330 (7.27%)	1,333 (7.29%)	203 (1.11%)
1	6	16,591	4,364	12,365	176	2,001	2,698	1,485	4,364	7,286	7.26	47.96	927 (5.59%)	391 (2.36%)	1,946 (11.73%)	798 (4.81%)	2,761 (16.64%)	1,467 (8.84%)	1,438 (8.67%)	312 (1.88%)
1	7	13,994	4,499	11,996	108	1,952	2,664	1,431	4,499	7,727	7.83	49.89	1,132 (8.09%)	453 (3.24%)	2,067 (14.77%)	864 (6.17%)	2,794 (19.97%)	1,541 (11.01%)	1,712 (12.23%)	492 (3.52%)
1	8	11,794	4,605	10,943	140	2,010	2,839	1,434	4,605	8,218	8.54	55.92	1,096 (9.29%)	540 (4.58%)	2,405 (20.39%)	1,038 (8.80%)	3,042 (25.79%)	1,544 (13.09%)	2,043 (17.32%)	609 (5.16%)
1	9	11,445	5,552	11,053	133	2,148	3,097	1,932	5,552	10,599	8.68	58.57	1,046 (9.14%)	732 (6.40%)	2,629 (22.97%)	1,138 (9.94%)	3,317 (28.98%)	1,599 (13.97%)	2,352 (20.55%)	708 (6.19%)
1	10	11,528	6,511	11,358	153	2,493	3,727	2,456	6,511	12,420	8.05	57.68	1,033 (8.96%)	892 (7.74%)	2,640 (22.90%)	1,175 (10.19%)	3,336 (28.94%)	1,660 (14.40%)	2,281 (19.79%)	820 (7.11%)
1	11	10,169	6,404	10,106	151	2,668	4,066	2,443	6,404	12,629	7.96	57.06	906 (8.91%)	906 (8.91%)	2,294 (22.56%)	1,149 (11.30%)	3,066 (30.15%)	1,447 (14.23%)	2,154 (21.18%)	911 (8.96%)
1	12	7,552	4,911	7,514	158	2,648	4,183	1,734	4,911	10,361	8.51	59.61	763 (10.10%)	854 (11.31%)	1,838 (24.34%)	961 (12.73%)	2,565 (33.96%)	1,150 (15.23%)	1,852 (24.52%)	961 (12.73%)
1	13	5,138	3,441	5,113	126	2,419	4,128	1,124	3,441	7,759	9.71	63.58	608 (11.83%)	820 (15.96%)	1,500 (29.19%)	786 (15.30%)	1,969 (38.32%)	874 (17.01%)	1,483 (28.86%)	891 (17.34%)
1	14	3,334	2,315	3,324	83	1,926	3,600	739	2,315	5,566	10.15	65.75	425 (12.75%)	610 (18.30%)	1,057 (31.70%)	568 (17.04%)	1,437 (43.10%)	597 (17.91%)	1,056 (31.67%)	788 (23.64%)
1	15	2,215	1,558	2,207	100	1,562	3,532	461	1,558	3,996	10.19	66.49	337 (15.21%)	512 (23.12%)	730 (32.96%)	441 (19.91%)	988 (44.60%)	448 (20.23%)	776 (35.03%)	583 (26.32%)
1	16	1,373	1,073	1,371	54	1,101	2,905	372	1,073	2,978	9.67	64.67	243 (17.70%)	374 (27.24%)	463 (33.72%)	276 (20.10%)	620 (45.16%)	262 (19.08%)	469 (34.16%)	386 (28.11%)
1	17	827	685	827	51	727	2,363	203	685	2,296	9.00	62.59	166 (20.07%)	252 (30.47%)	312 (37.73%)	210 (25.39%)	416 (50.30%)	147 (17.78%)	335 (40.51%)	245 (29.63%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	518	448	516	38	471	1,935	161	448	1,680	12.10	61.08	130 (25.10%)	196 (37.84%)	189 (36.49%)	140 (27.03%)	243 (46.91%)	100 (19.31%)	208 (40.15%)	154 (29.73%)
1	19	375	327	371	22	346	1,735	105	327	1,382	7.19	55.02	104 (27.73%)	131 (34.93%)	118 (31.47%)	84 (22.40%)	163 (43.47%)	60 (16.00%)	142 (37.87%)	85 (22.67%)
1	20	352	321	350	69	328	2,755	149	321	2,148	9.33	52.97	102 (28.98%)	128 (36.36%)	126 (35.80%)	95 (26.99%)	156 (44.32%)	51 (14.49%)	144 (40.91%)	81 (23.01%)
1	Total	172,358	59,415	113,315	1,859	28,522	50,857	21,002	59,415	116,342	6.44	47.75	11,599 (6.73%)	8,278 (4.80%)	24,431 (14.17%)	11,303 (6.56%)	32,532 (18.87%)	15,719 (9.12%)	21,337 (12.38%)	8,362 (4.85%)

TABLE A.51: ERMER: Risk bands statistics of the *Pop_Any-Acute_Cond_Prior-Spells* model (*Sample_1_train_half_3_test_half*)

Modelling Approach: BPM;
Modelling Group: Pop_Any-Acute;
Sample: Sample_1_train_half_3_test_half;
Submodel: Cond_Prior-Spells

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	1	6,895	535	0	0	0	0	248	535	667	7.64	26.19	201 (2.92%)	10 (0.15%)	125 (1.81%)	58 (0.84%)	141 (2.04%)	60 (0.87%)	50 (0.73%)	4 (0.06%)
0	2	15,814	2,169	0	0	0	0	903	2,169	3,124	3.54	32.46	443 (2.80%)	47 (0.30%)	581 (3.67%)	251 (1.59%)	666 (4.21%)	253 (1.60%)	361 (2.28%)	24 (0.15%)
0	3	19,207	2,916	0	0	0	0	1,223	2,916	4,341	3.93	35.22	832 (4.33%)	153 (0.80%)	996 (5.19%)	475 (2.47%)	1,164 (6.06%)	399 (2.08%)	692 (3.60%)	68 (0.35%)
0	4	17,107	3,100	0	0	0	0	1,197	3,100	4,746	5.01	42.10	790 (4.62%)	270 (1.58%)	1,383 (8.08%)	634 (3.71%)	1,620 (9.47%)	452 (2.64%)	966 (5.65%)	125 (0.73%)
0	5	25,551	5,188	0	0	0	0	1,979	5,188	8,243	4.43	34.37	1,149 (4.50%)	310 (1.21%)	1,334 (5.22%)	708 (2.77%)	1,697 (6.64%)	447 (1.75%)	947 (3.71%)	199 (0.78%)
0	6	20,030	4,612	0	0	0	0	1,712	4,612	7,440	5.85	46.55	984 (4.91%)	440 (2.20%)	1,704 (8.51%)	835 (4.17%)	2,144 (10.70%)	399 (1.99%)	1,191 (5.95%)	262 (1.31%)
0	7	16,221	3,998	0	0	0	0	1,465	3,998	6,514	6.48	51.82	666 (4.11%)	489 (3.01%)	1,809 (11.15%)	847 (5.22%)	2,341 (14.43%)	274 (1.69%)	1,405 (8.66%)	349 (2.15%)
0	8	11,605	3,720	0	0	0	0	1,477	3,720	6,281	8.03	54.97	509 (4.39%)	487 (4.20%)	1,400 (12.06%)	668 (5.76%)	1,938 (16.70%)	318 (2.74%)	1,156 (9.96%)	301 (2.59%)
0	9	7,071	2,817	0	0	0	0	1,092	2,817	4,742	10.85	63.59	364 (5.15%)	441 (6.24%)	1,220 (17.25%)	574 (8.12%)	1,718 (24.30%)	260 (3.68%)	1,042 (14.74%)	367 (5.19%)
0	10	5,807	3,432	0	0	0	0	1,543	3,432	6,155	8.91	54.88	175 (3.01%)	297 (5.11%)	754 (12.98%)	383 (6.60%)	1,192 (20.53%)	209 (3.60%)	641 (11.04%)	373 (6.42%)
0	11	5,451	3,569	0	0	0	0	1,558	3,569	7,001	5.92	44.51	127 (2.33%)	151 (2.77%)	444 (8.15%)	241 (4.42%)	754 (13.83%)	158 (2.90%)	412 (7.56%)	329 (6.04%)
0	12	3,632	2,558	0	0	0	0	1,205	2,558	5,091	4.68	41.33	61 (1.68%)	81 (2.23%)	206 (5.67%)	144 (3.96%)	436 (12.00%)	110 (3.03%)	219 (6.03%)	217 (5.97%)
0	13	4,542	3,870	0	0	0	0	2,062	3,870	7,925	2.03	32.65	50 (1.10%)	33 (0.73%)	120 (2.64%)	49 (1.08%)	385 (8.48%)	58 (1.28%)	114 (2.51%)	114 (2.51%)
0	14	1,984	1,673	0	0	0	0	949	1,673	3,248	2.99	36.68	38 (1.92%)	18 (0.91%)	55 (2.77%)	32 (1.61%)	235 (11.84%)	36 (1.81%)	57 (2.87%)	54 (2.72%)
0	15	288	122	0	0	0	0	67	122	223	7.88	55.48	7 (2.43%)	6 (2.08%)	19 (6.60%)	11 (3.82%)	34 (11.81%)	16 (5.56%)	22 (7.64%)	17 (5.90%)
0	16	91	47	0	0	0	0	16	47	86	20.91	65.41	4 (4.40%)	0 (0.00%)	11 (12.09%)	6 (6.59%)	20 (21.98%)	11 (12.09%)	11 (12.09%)	11 (12.09%)
0	17	50	20	0	0	0	0	8	20	32	29.98	70.74	2 (4.00%)	2 (4.00%)	9 (18.00%)	3 (6.00%)	14 (28.00%)	15 (30.00%)	9 (18.00%)	7 (14.00%)

Continued on next page

If[a]	#[b]	TP+FP	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
0	18	12	7	0	0	0	0	3	7	12	53.33	75.92	0 (0.00%)	0 (0.00%)	3 (25.00%)	3 (25.00%)	4 (33.33%)	2 (16.67%)	4 (33.33%)	5 (41.67%)
0	19	8	2	0	0	0	0	0	2	3	202.75	54.00	1 (12.50%)	0 (0.00%)	3 (37.50%)	1 (12.50%)	3 (37.50%)	1 (12.50%)	1 (12.50%)	0 (0.00%)
0	20	9	2	0	0	0	0	1	2	3	215.33	39.22	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (11.11%)	0 (0.00%)
0	Total	161,375	44,357	0	0	0	0	18,708	44,357	75,877	5.58	42.03	6,403 (3.97%)	3,235 (2.00%)	12,176 (7.55%)	5,923 (3.67%)	16,506 (10.23%)	3,478 (2.16%)	9,301 (5.76%)	2,826 (1.75%)
1	1	211	22	211	3	39	47	8	22	27	300.45	23.17	6 (2.84%)	0 (0.00%)	5 (2.37%)	2 (0.95%)	6 (2.84%)	1 (0.47%)	3 (1.42%)	0 (0.00%)
1	2	3,115	499	3,115	21	347	395	177	499	754	4.83	26.98	176 (5.65%)	6 (0.19%)	96 (3.08%)	23 (0.74%)	134 (4.30%)	137 (4.40%)	47 (1.51%)	2 (0.06%)
1	3	5,214	1,259	5,214	41	702	843	410	1,259	2,106	5.36	42.66	379 (7.27%)	42 (0.81%)	488 (9.36%)	182 (3.49%)	573 (10.99%)	429 (8.23%)	317 (6.08%)	27 (0.52%)
1	4	6,197	1,886	6,197	85	987	1,252	627	1,886	3,234	7.62	49.79	473 (7.63%)	149 (2.40%)	1,003 (16.19%)	353 (5.70%)	1,153 (18.61%)	739 (11.93%)	756 (12.20%)	103 (1.66%)
1	5	9,112	2,596	9,112	144	1,673	2,178	834	2,596	4,490	6.12	44.80	705 (7.74%)	234 (2.57%)	1,238 (13.59%)	529 (5.81%)	1,538 (16.88%)	947 (10.39%)	937 (10.28%)	193 (2.12%)
1	6	16,287	4,213	16,287	196	2,711	3,499	1,287	4,213	7,209	4.89	38.32	1,335 (8.20%)	325 (2.00%)	1,503 (9.23%)	713 (4.38%)	2,025 (12.43%)	1,141 (7.01%)	1,182 (7.26%)	334 (2.05%)
1	7	18,860	5,077	18,860	244	3,461	4,673	1,592	5,077	8,963	5.47	43.33	1,632 (8.65%)	438 (2.32%)	2,066 (10.95%)	989 (5.24%)	2,965 (15.72%)	1,403 (7.44%)	1,747 (9.26%)	439 (2.33%)
1	8	15,017	5,294	15,017	187	3,318	4,466	1,620	5,294	9,544	7.65	53.95	1,480 (9.86%)	724 (4.82%)	2,806 (18.69%)	1,226 (8.16%)	3,592 (23.92%)	1,479 (9.85%)	2,470 (16.45%)	599 (3.99%)
1	9	12,924	5,725	12,924	186	3,266	4,548	1,668	5,725	10,616	9.37	62.64	1,380 (10.68%)	1,075 (8.32%)	3,330 (25.77%)	1,460 (11.30%)	4,061 (31.42%)	1,503 (11.63%)	3,016 (23.34%)	802 (6.21%)
1	10	11,714	6,190	11,714	199	3,424	4,979	1,906	6,190	11,712	10.18	66.02	1,157 (9.88%)	1,318 (11.25%)	3,333 (28.45%)	1,559 (13.31%)	4,091 (34.92%)	1,480 (12.63%)	3,011 (25.70%)	1,120 (9.56%)
1	11	10,647	6,261	10,647	211	3,537	5,384	2,079	6,261	12,714	9.59	64.42	1,032 (9.69%)	1,349 (12.67%)	2,865 (26.91%)	1,494 (14.03%)	3,706 (34.81%)	1,287 (12.09%)	2,759 (25.91%)	1,347 (12.65%)
1	12	12,026	8,569	12,026	327	4,231	6,455	3,515	8,569	17,631	6.65	53.59	924 (7.68%)	1,316 (10.94%)	2,299 (19.12%)	1,207 (10.04%)	3,260 (27.11%)	1,171 (9.74%)	2,343 (19.48%)	1,380 (11.48%)
1	13	9,678	7,363	9,678	225	3,632	6,069	3,136	7,363	15,785	6.05	51.65	847 (8.75%)	1,087 (11.23%)	1,771 (18.30%)	960 (9.92%)	2,659 (27.47%)	929 (9.60%)	1,759 (18.18%)	1,239 (12.80%)
1	14	5,235	3,919	5,235	135	2,849	5,030	1,523	3,919	9,389	7.24	56.74	612 (11.69%)	855 (16.33%)	1,216 (23.23%)	713 (13.62%)	1,790 (34.19%)	606 (11.58%)	1,249 (23.86%)	961 (18.36%)
1	15	2,897	2,159	2,897	79	2,041	4,187	756	2,159	5,710	8.73	61.03	422 (14.57%)	606 (20.92%)	810 (27.96%)	478 (16.50%)	1,140 (39.35%)	379 (13.08%)	862 (29.75%)	669 (23.09%)
1	16	1,683	1,343	1,683	60	1,335	3,445	435	1,343	3,764	8.92	61.34	296 (17.59%)	429 (25.49%)	510 (30.30%)	308 (18.30%)	677 (40.23%)	244 (14.50%)	540 (32.09%)	412 (24.48%)
1	17	1,033	846	1,033	61	903	2,797	289	846	2,676	9.61	62.85	213 (20.62%)	332 (32.14%)	368 (35.62%)	265 (25.65%)	493 (47.73%)	168 (16.26%)	408 (39.50%)	310 (30.01%)

Continued on next page

If[a]	#[b]	TP+FP[c]	TP[d]	Prior Spells[e]	Prior 30-d[f]	Prior 12-m[g]	Prior 12-m Spells[h]	Post 30-d[i]	Post 12-m[j]	Post 12-m Spells[k]	Avg. Stay.[l]	Avg. Age[m]	Asth.[n]	COPD[o]	Depr.[p]	Diab.[q]	Hype.[r]	Canc.[s]	CHD[t]	CHF[u]
1	18	709	611	709	41	642	2,510	220	611	2,395	9.09	61.46	173 (24.40%)	259 (36.53%)	261 (36.81%)	178 (25.11%)	326 (45.98%)	101 (14.25%)	278 (39.21%)	203 (28.63%)
1	19	476	420	476	34	449	2,198	137	420	1,717	7.42	58.60	126 (26.47%)	184 (38.66%)	178 (37.39%)	114 (23.95%)	231 (48.53%)	80 (16.81%)	191 (40.13%)	116 (24.37%)
1	20	478	436	478	88	460	3,673	183	436	2,665	10.37	57.19	152 (31.80%)	208 (43.51%)	186 (38.91%)	136 (28.45%)	224 (46.86%)	63 (13.18%)	218 (45.61%)	124 (25.94%)
1	Total	143,513	64,688	143,513	2,567	40,007	68,628	22,402	64,688	133,101	7.60	51.80	13,520 (9.42%)	10,936 (7.62%)	26,332 (18.35%)	12,889 (8.98%)	34,644 (24.14%)	14,287 (9.96%)	24,093 (16.79%)	10,380 (7.23%)

A.6.3.3 Performance Ensemble Model

A.6.3.4 Generated Models

TABLE A.52: ERMER: The generated ensemble models (*Sample-1*)

Sample[a]	**Ensemble Comb.**[b]	**AUC**[c]	**TP**[d]	**Spec.**[e]	**Prec.**[f]	**Sens.**[g]	**F1**[h]
Sample-1	1.1.8.0.1.3.1.0.0	76.97	38,018	90.15	72.10	46.12	56.26
Sample-1	1.1.6.0.0.4.1.0.0	76.91	38,091	90.11	72.06	46.21	56.31
Sample-1	1.0.8.0.1.3.1.0.1	76.96	37,840	90.12	71.94	45.90	56.05
Sample-1	1.1.8.0.0.4.1.0.0	76.96	38,111	90.13	72.10	46.23	56.34
Sample-1	1.1.9.0.0.4.2.0.0	77.05	37,979	90.05	71.87	46.07	56.15
Sample-1	1.1.7.0.1.3.1.0.0	76.95	38,037	90.13	72.08	46.14	56.26
Sample-1	1.1.8.0.0.4.2.0.0	77.03	37,983	90.02	71.82	46.08	56.14
Sample-2	1.1.8.0.1.3.1.0.0	75.83	44,932	88.32	71.65	49.10	58.27
Sample-2	1.1.6.0.0.4.1.0.0	75.81	44,937	88.36	71.72	49.10	58.30
Sample-2	1.0.8.0.1.3.1.0.1	75.73	44,720	88.37	71.65	48.87	58.10
Sample-2	1.1.8.0.0.4.1.0.0	75.85	45,015	88.31	71.67	49.19	58.34
Sample-2	1.1.9.0.0.4.2.0.0	75.88	45,032	88.25	71.58	49.21	58.32
Sample-2	1.1.7.0.1.3.1.0.0	75.81	44,887	88.35	71.68	49.05	58.24
Sample-2	1.1.8.0.0.4.2.0.0	75.86	45,009	88.26	71.58	49.18	58.30
Sample-3	1.1.8.0.1.3.1.0.0	77.10	51,440	89.33	71.11	47.17	56.72
Sample-3	1.1.6.0.0.4.1.0.0	77.05	51,396	89.32	71.08	47.13	56.68
Sample-3	1.0.8.0.1.3.1.0.1	77.16	51,269	89.29	70.97	47.02	56.56
Sample-3	1.1.8.0.0.4.1.0.0	77.09	51,453	89.32	71.09	47.19	56.72
Sample-3	1.1.9.0.0.4.2.0.0	77.22	51,328	89.22	70.86	47.07	56.57
Sample-3	1.1.7.0.1.3.1.0.0	77.08	51,420	89.33	71.10	47.15	56.70
Sample-3	1.1.8.0.0.4.2.0.0	77.20	51,322	89.22	70.85	47.06	56.56
Sample-1-*train-half*-3-*test-half*	1.1.8.0.1.3.1.0.0	76.87	45,896	91.83	74.14	42.09	53.70
Sample-1-*train-half*-3-*test-half*	1.1.6.0.0.4.1.0.0	76.82	46,178	91.70	73.97	42.35	53.86
Sample-1-*train-half*-3-*test-half*	1.0.8.0.1.3.1.0.1	76.93	45,575	91.87	74.11	41.79	53.45
Sample-1-*train-half*-3-*test-half*	1.1.8.0.0.4.1.0.0	76.86	46,175	91.73	74.03	42.34	53.88
Sample-1-*train-half*-3-*test-half*	1.1.9.0.0.4.2.0.0	77.02	45,950	91.70	73.86	42.14	53.66
Sample-1-*train-half*-3-*test-half*	1.1.7.0.1.3.1.0.0	76.85	45,898	91.80	74.09	42.09	53.68
Sample-1-*train-half*-3-*test-half*	1.1.8.0.0.4.2.0.0	77.00	45,980	91.68	73.84	42.17	53.68

[a] Sample: The selected sample data. [b] Ensemble Comb.: The ensemble model combination, which is defined according to the $Cond_Ensemble$. [c] AUC: The AUC of ROC of the classification. [d] TP: The number of true positives (TPs). [e] Spec.: The specificity measure, also known as the true negative rate (TNR). [f] Prec.: The precision measure, also known as the positive predictive value (PPV). [g] Sens.: The sensitivity measure, also known as the true positive rate (TPR), or the recall. [h] F1: The F1 score of the classification.

A.6.3.5 Summary Performance Statistics

TABLE A.53: ERMER: The benchmark of the *Cond_Ensemble* sub-model (*Sample-1*)

Statistic[a]	**Orig. PARR**[b]			**Orig. PARR-30**[c]	**Orig. Billings-13 (IP)**[d]	**Orig. Billings-13 (full)**[e]	**Sub_PARR-2-Settings**[f]			**Sub_IPAEOPGP**[g]			**Sub_Any-Acute**[h]		
Threshold	**0.50**	**0.60**	**0.70**	**0.50**	**0.50**	**0.50**	**0.50**	**0.60**	**0.70**	**0.50**	**0.60**	**0.70**	**0.50**	**0.60**	**0.70**
TP+FP[i]	17,455	4,810	2,011	6,395	8,743	10,545	19,646	7,946	2,991	51,422	30,361	14,719	52,842	31,260	15,231
TP[j]	NA	NA	NA	3,786	4,627	5,669	11,962	5,512	2,291	36,966	24,051	12,432	37,979	24,759	12,878
Sensitivity[k]	0.543	0.178	0.081	0.054	0.049	0.060	0.390	0.180	0.075	0.478	0.311	0.161	0.461	0.300	0.156
Specificity[l]	0.722	0.950	0.986	0.995	NA	NA	0.805	0.938	0.982	0.887	0.950	0.982	0.900	0.956	0.984
Precision[m]	0.653	0.774	0.843	0.592	0.529	0.538	0.609	0.694	0.766	0.719	0.792	0.845	0.719	0.792	0.846
Emer. admi. post 12 m.[n]	1.47	2.23	3.0	NA	NA	NA	24,397	12,717	6,295	81,296	56,381	31,588	83,786	58,246	32,861
							(1.242)	(1.600)	(2.105)	(1.581)	(1.857)	(2.146)	(1.586)	(1.863)	(2.158)
Emer. admi. prior 12 m.[o]	2.22	3.43	4.59	NA	NA	NA	9,069	4,820	2,213	18,057	11,068	5,410	18,563	11,364	5,594
							(0.462)	(0.607)	(0.740)	(0.351)	(0.365)	(0.368)	(0.351)	(0.364)	(0.367)
Emer. admi. prior 13-24 m.[p]	0.93	1.84	2.80	NA	NA	NA	7,870	4,228	1,932	16,378	10,216	4,820	16,796	10,468	4,961
							(0.401)	(0.532)	(0.646)	(0.319)	(0.336)	(0.327)	(0.318)	(0.335)	(0.326)
Emer. admi. prior 25-36 m.[q]	0.73	1.48	2.25	NA	NA	NA	116	55	27	217	129	71	222	131	73
							(0.006)	(0.007)	(0.009)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)
HL Test[r]	NA			NA	NA	NA	7,042.149***			13,900.775***			16,487.336***		
AUC ROC[s]	0.69			0.70	0.73	0.78	0.661			0.767			0.771		
N[t]	42,778			576,868	1,836,099	1,836,099	70,147			204,672			231,755		

[a] Statistic: Name of the statistical measure. [b] Orig. PARR: The performance of the original PARR model. Some of the values are estimated. [c] Orig. PARR-30: The performance of the original PARR-30 model. Some of the values are estimated. [d] Orig. Billings-13 (IP): The performance of the original Billings et. al. (2013) model with inpatient data. Some of the values are estimated. [e] Orig. Billings-13 (full): The performance of the original Billings et. al. (2013) model with inpatient, A&E, outpatient and GP data. Some of the values are estimated. [f] Orig. Sub_PARR-2-Settings: The performance of the model for the sub-population $Sub\_PARR-2-Settings$. [g] Orig. Sub_IPAEOPGP: The performance of the model for the sub-population $Sub\_IPAEOPGP$. [h] Orig. Sub_Any-Acute: The performance of the model for the sub-population $Sub\_Any-Acute$. [i] TP+FP: The number of true positives (TPs) and false positives (FPs). [j] TP: The number of true positives (TPs). [k] Sensitivity: The sensitivity measure, also known as the true positive rate (TPR), or the recall. [l] Specificity: The specificity measure, also known as the true negative rate (TNR). [m] Precision: The precision measure, also known as the positive predictive value (PPV). [n] Emer. admi. post 12 m.: The number of emergency admissions in the next 12 months. [o] Emer. admi. prior 12 m.: The number of emergency admissions in the past 12 months. [q] Emer. admi. prior 13-24 m.: The number of emergency admissions in the past 13 to 24 months. [r] Emer. admi. prior 25-36 m.: The number of emergency admissions in the past 25 to 36 months. [s] HL Test: the Hosmer-Lemeshow test. [t] N: The total number of patients.

TABLE A.54: ERMER: The benchmark of the *Cond_Ensemble* sub-model (*Sample-2*)

Statistic	**Orig. PARR**			**Orig. PARR-30**	**Orig. Billings-13 (IP)**	**Orig. Billings-13 (full)**	**Sub_PARR-2-Settings**			**Sub_IPAEOPGP**			**Sub_Any-Acute**		
Threshold	**0.50**	**0.60**	**0.70**	**0.50**	**0.50**	**0.50**	**0.50**	**0.60**	**0.70**	**0.50**	**0.60**	**0.70**	**0.50**	**0.60**	**0.70**
TP+FP	17,455	4,810	2,011	6,395	8,743	10,545	25,972	11,121	4,212	61,229	34,292	15,745	62,910	35,230	16,177
TP	NA	NA	NA	3,786	4,627	5,669	15,916	7,577	3,169	43,858	26,920	13,180	45,032	27,611	13,539
Sensitivity	0.543	0.178	0.081	0.054	0.049	0.060	0.470	0.224	0.094	0.503	0.309	0.151	0.492	0.302	0.148
Specificity	0.722	0.950	0.986	0.995	NA	NA	0.745	0.910	0.974	0.873	0.946	0.981	0.883	0.950	0.983
Precision	0.653	0.774	0.843	0.592	0.529	0.538	0.613	0.681	0.752	0.716	0.785	0.837	0.716	0.784	0.837
Emer. admi. post 12 m.	1.47	2.23	3.0	NA	NA	NA	33,656	17,835	8,637	99,397	66,013	35,776	102,144	67,710	36,720
							(1.296)	(1.604)	(2.051)	(1.623)	(1.925)	(2.272)	(1.624)	(1.922)	(2.270)
Emer. admi. prior 12 m.	2.22	3.43	4.59	NA	NA	NA	11,739	6,572	3,046	22,340	13,821	6,951	22,953	14,167	7,111
							(0.452)	(0.591)	(0.723)	(0.365)	(0.403)	(0.441)	(0.365)	(0.402)	(0.440)
Emer. admi. prior 13-24 m.	0.93	1.84	2.80	NA	NA	NA	10,089	5,639	2,674	20,041	12,367	6,226	20,560	12,671	6,363
							(0.388)	(0.507)	(0.635)	(0.327)	(0.361)	(0.395)	(0.327)	(0.360)	(0.393)
Emer. admi. prior 25-36 m.	0.73	1.48	2.25	NA	NA	NA	183	105	44	316	201	103	324	205	106
							(0.007)	(0.009)	(0.010)	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)
HL Test	NA			NA	NA	NA	3,405.223***			12,659.019***			14,319.221***		
AUC ROC	0.69			0.70	0.73	0.78	0.663			0.756			0.759		
N	42,778			576,868	1,836,099	1,836,099	73,315			224,001			243,712		

TABLE A.55: ERMER: The benchmark of the *Cond_Ensemble* sub-model (*Sample-1-train-half-3-test-half*)

Statistic	Orig. PARR			Orig. PARR-30	Orig. Billings-13 (IP)	Orig. Billings-13 (full)	Sub_PARR-2-Settings			Sub_IPAEOPGP			Sub_Any-Acute		
Threshold	**0.50**	**0.60**	**0.70**	**0.50**	**0.50**	**0.50**	**0.50**	**0.60**	**0.70**	**0.50**	**0.60**	**0.70**	**0.50**	**0.60**	**0.70**
TP+FP	17,455	4,810	2,011	6,395	8,743	10,545	22,351	8,351	2,896	60,515	35,642	18,487	62,213	36,753	19,117
TP	NA	NA	NA	3,786	4,627	5,669	14,003	5,942	2,337	44,730	28,783	16,114	45,950	29,654	16,678
Sensitivity	0.543	0.178	0.081	0.054	0.049	0.060	0.340	0.144	0.057	0.438	0.282	0.158	0.421	0.272	0.153
Specificity	0.722	0.950	0.986	0.995	NA	NA	0.834	0.952	0.989	0.905	0.959	0.986	0.917	0.964	0.988
Precision	0.653	0.774	0.843	0.592	0.529	0.538	0.627	0.712	0.807	0.739	0.808	0.872	0.739	0.807	0.872
Emer. admi. post 12 m.	1.47	2.23	3.0	NA	NA	NA	29,302	14,447	6,838	99,630	68,182	39,992	102,944	70,771	41,788
							(1.311)	(1.730)	(2.361)	(1.646)	(1.913)	(2.163)	(1.655)	(1.926)	(2.186)
Emer. admi. prior 12 m.	2.22	3.43	4.59	NA	NA	NA	11,657	5,711	2,331	22,046	12,401	5,614	22,651	12,767	5,813
							(0.522)	(0.684)	(0.805)	(0.364)	(0.348)	(0.304)	(0.364)	(0.347)	(0.304)
Emer. admi. prior 13-24 m.	0.93	1.84	2.80	NA	NA	NA	9,730	4,719	1,933	19,498	10,923	4,812	19,988	11,211	4,960
							(0.435)	(0.565)	(0.667)	(0.322)	(0.306)	(0.260)	(0.321)	(0.305)	(0.259)
Emer. admi. prior 25-36 m.	0.73	1.48	2.25	NA	NA	NA	111	51	28	239	137	75	248	143	80
							(0.005)	(0.006)	(0.010)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
HL Test	NA			NA	NA	NA	7,042.149***			27,251.451***			31,312.081***		
AUC ROC	0.69			0.70	0.73	0.78	0.658			0.767			0.771		
N	42,778			576,868	1,836,099	1,836,099	91,369			268,575			304,888		

A.6.3.6 Top Risk Segments Statistics

TABLE A.56: ERMER: The top risk segments profile of the *Cond_Ensemble* sub-model (*Sample-1*)

Risk Seg.[a]	Model[b]	Sub-population[c]	Min Risk[d]	Asthma (%)[e]	COPD (%)[f]	Depres. (%)[g]	Diab. (%)[h]	Hyper. (%)[i]	Cancer (%)[j]	CHD (%)[k]	CHF (%)[l]	Avg. Age[m]	Avg. LoS[n]	5-9 Meds[o]	10+ Meds[p]
Train Sample: 50% of *Sample-1*; Test Sample: 50% of *Sample-1*															
	Orig. CPM[q]	All	NA	20.10	8.20	17.90	13.70	45.20	14.90	15.50	6.50	67.30	11.40	26.00	6.00
	Orig. PARR[r]	All	NA	14.30	5.10	11.80	9.30	29.00	9.40	12.00	4.60	55.40	10.40	14.10	3.00
10,000		PARR-2-Settings	0.576	16.69	35.68	41.94	23.49	53.20	19.65	50.93	39.82	80.80	11.06	NA	NA
	Ensemble[s]	t													
		IPAEOPGP	0.759	11.54	12.19	12.24	8.25	20.71	6.51	14.88	10.34	39.68	4.49	NA	NA
		u													
		Any-Acute	0.766	11.25	11.52	11.57	7.93	19.74	6.24	14.01	9.72	38.61	4.39	NA	NA
		v													
	Orig. CPM	All	NA	23.30	11.00	20.60	16.20	51.40	15.30	18.50	9.40	69.70	11.50	29.50	8.60
	Orig. PARR	All	NA	16.60	8.80	13.70	12.90	38.70	15.10	18.80	8.30	66.20	11.00	17.60	4.30
5,000		PARR-2-Settings	0.647	20.84	44.14	45.10	26.00	56.28	21.24	57.00	45.10	80.33	11.32	NA	NA
	Ensemble	IPAEOPGP	0.817	15.80	15.78	15.28	10.14	25.20	7.52	18.58	12.96	42.36	4.91	NA	NA
		Any-Acute	0.818	15.84	15.72	15.32	10.38	25.12	7.60	18.48	12.92	41.99	4.93	NA	NA
	Orig. CPM	All	NA	31.40	19.60	22.70	18.80	61.20	16.40	24.50	15.90	71.60	10.30	31.50	13.40
	Orig. PARR	All	NA	24.30	20.60	16.90	19.90	47.70	20.00	24.80	19.20	69.50	9.90	14.60	7.40
1,000		PARR-2-Settings	0.815	31.40	59.10	50.70	26.90	61.70	22.90	66.40	53.30	78.95	10.04	NA	NA
	Ensemble	IPAEOPGP	0.910	33.40	35.70	30.50	21.90	39.70	14.30	38.60	26.30	53.38	6.98	NA	NA
		Any-Acute	0.912	33.20	34.80	29.90	21.80	39.00	14.40	37.50	25.40	52.21	6.85	NA	NA
	Orig. CPM	All	NA	34.00	22.60	25.20	19.20	63.20	16.40	28.80	19.40	70.70	9.80	30.60	14.60
	Orig. PARR	All	NA	28.40	24.60	16.40	23.00	49.00	21.00	28.80	24.40	69.10	9.40	14.80	8.60
500		PARR-2-Settings	0.881	37.40	67.60	52.00	26.40	63.20	25.20	69.60	55.20	77.98	9.35	NA	NA
	Ensemble	IPAEOPGP	0.957	38.20	38.60	34.80	25.20	42.80	14.20	43.20	27.40	54.49	7.43	NA	NA
		Any-Acute	0.958	37.80	37.60	33.80	25.00	41.20	14.00	41.80	26.80	52.95	7.37	NA	NA
	Orig. CPM	All	NA	40.80	28.40	24.40	21.20	62.80	17.60	29.60	23.20	69.50	8.90	32.00	18.00
	Orig. PARR	All	NA	30.80	26.80	20.00	22.80	52.40	20.80	30.80	29.20	68.80	9.80	14.40	10.00
250		PARR-2-Settings	0.933	36.40	70.00	53.60	27.60	63.20	25.60	69.60	53.60	77.34	9.57	NA	NA
	Ensemble	IPAEOPGP	0.985	40.80	39.20	36.40	27.20	42.80	11.60	42.00	29.60	53.88	7.99	NA	NA
		Any-Acute	0.986	40.40	38.80	36.40	27.60	42.40	11.60	41.20	28.80	52.76	7.89	NA	NA
Train Sample: 50% of *Sample-2*; Test Sample: 50% of *Sample-2*															
	Orig. CPM	All	NA	20.10	8.20	17.90	13.70	45.20	14.90	15.50	6.50	67.30	11.40	26.00	6.00
	Orig. PARR	All	NA	14.30	5.10	11.80	9.30	29.00	9.40	12.00	4.60	55.40	10.40	14.10	3.00
10,000		PARR-2-Settings	0.612	22.03	38.78	70.50	29.45	78.14	28.56	54.89	39.58	81.52	8.15	NA	NA
	Ensemble	IPAEOPGP	0.771	17.96	15.53	23.70	13.10	31.93	10.40	18.76	11.97	42.08	3.85	NA	NA
		Any-Acute	0.774	18.03	15.38	23.51	12.94	31.66	10.41	18.59	11.98	41.80	3.82	NA	NA
	Orig. CPM	All	NA	23.30	11.00	20.60	16.20	51.40	15.30	18.50	9.40	69.70	11.50	29.50	8.60
	Orig. PARR	All	NA	16.60	8.80	13.70	12.90	38.70	15.10	18.80	8.30	66.20	11.00	17.60	4.30
5,000		PARR-2-Settings	0.683	26.30	46.88	74.00	33.40	81.30	29.56	60.98	44.56	80.96	8.32	NA	NA
	Ensemble	IPAEOPGP	0.815	26.62	22.48	31.78	18.50	41.02	14.20	26.06	16.32	46.35	4.55	NA	NA
		Any-Acute	0.818	26.70	22.34	31.72	18.44	40.78	14.10	26.02	16.38	46.09	4.50	NA	NA

Continued on next page

Risk Seg.	Model	Sub-population	Min Risk	Asthma (%)	COPD (%)	Depres. (%)	Diab. (%)	Hyper. (%)	Cancer (%)	CHD (%)	CHF (%)	Avg. Age	Avg. LoS	5-9 Meds	10+ Meds
	Orig. CPM	All	NA	31.40	19.60	22.70	18.80	61.20	16.40	24.50	15.90	71.60	10.30	31.50	13.40
	Orig. PARR	All	NA	24.30	20.60	16.90	19.90	47.70	20.00	24.80	19.20	69.50	9.90	14.60	7.40
1,000		PARR-2-Settings	0.839	40.20	64.40	78.80	38.70	85.70	33.90	71.70	52.70	79.16	8.23	NA	NA
	Ensemble	IPAEOPGP	0.939	43.80	33.70	49.40	31.90	58.30	20.10	41.40	21.90	52.04	5.98	NA	NA
		Any-Acute	0.940	44.40	33.10	48.90	31.60	57.50	20.00	40.60	21.40	51.55	5.92	NA	NA
	Orig. CPM	All	NA	34.00	22.60	25.20	19.20	63.20	16.40	28.80	19.40	70.70	9.80	30.60	14.60
	Orig. PARR	All	NA	28.40	24.60	16.40	23.00	49.00	21.00	28.80	24.40	69.10	9.40	14.80	8.60
500		PARR-2-Settings	0.895	44.60	70.20	80.80	42.20	87.00	34.20	75.60	55.20	78.46	9.09	NA	NA
	Ensemble	IPAEOPGP	0.980	45.60	32.00	51.00	31.80	58.80	18.60	41.20	20.20	50.48	4.66	NA	NA
		Any-Acute	0.981	45.80	31.60	50.60	31.40	58.40	18.80	40.20	19.80	50.08	4.60	NA	NA
	Orig. CPM	All	NA	40.80	28.40	24.40	21.20	62.80	17.60	29.60	23.20	69.50	8.90	32.00	18.00
	Orig. PARR	All	NA	30.80	26.80	20.00	22.80	52.40	20.80	30.80	29.20	68.80	9.80	14.40	10.00
250		PARR-2-Settings	0.941	49.60	72.00	84.80	44.00	88.40	33.20	82.00	56.00	76.69	9.51	NA	NA
	Ensemble	IPAEOPGP	0.996	48.40	28.00	50.80	30.80	57.60	14.00	37.60	16.80	47.88	3.97	NA	NA
		Any-Acute	0.996	48.00	28.00	50.80	30.00	57.60	13.60	37.60	16.80	47.56	3.99	NA	NA
Train Sample: 50% of *Sample-1*; Test Sample: 50% of *Sample-3*															
	Orig. CPM	All	NA	20.10	8.20	17.90	13.70	45.20	14.90	15.50	6.50	67.30	11.40	26.00	6.00
	Orig. PARR	All	NA	14.30	5.10	11.80	9.30	29.00	9.40	12.00	4.60	55.40	10.40	14.10	3.00
10,000		PARR-2-Settings	0.583	16.11	37.70	43.99	23.96	54.83	17.25	49.15	38.29	80.99	10.50	NA	NA
	Ensemble	IPAEOPGP	0.801	7.81	9.10	8.22	6.13	15.47	3.50	9.63	6.25	35.81	3.57	NA	NA
		Any-Acute	0.803	8.05	9.08	8.27	6.23	15.51	3.61	9.58	6.22	35.41	3.57	NA	NA
	Orig. CPM	All	NA	23.30	11.00	20.60	16.20	51.40	15.30	18.50	9.40	69.70	11.50	29.50	8.60
	Orig. PARR	All	NA	16.60	8.80	13.70	12.90	38.70	15.10	18.80	8.30	66.20	11.00	17.60	4.30
5,000		PARR-2-Settings	0.648	19.52	45.78	46.76	27.08	57.54	18.18	54.60	42.68	80.39	11.10	NA	NA
	Ensemble	IPAEOPGP	0.824	13.24	15.10	13.70	10.10	22.38	5.54	15.70	10.06	40.38	4.53	NA	NA
		Any-Acute	0.825	13.54	15.16	13.74	10.22	22.24	5.72	15.60	10.06	39.77	4.54	NA	NA
	Orig. CPM	All	NA	31.40	19.60	22.70	18.80	61.20	16.40	24.50	15.90	71.60	10.30	31.50	13.40
	Orig. PARR	All	NA	24.30	20.60	16.90	19.90	47.70	20.00	24.80	19.20	69.50	9.90	14.60	7.40
1,000		PARR-2-Settings	0.808	28.90	64.10	53.10	31.70	63.40	19.50	63.60	47.40	78.37	9.66	NA	NA
	Ensemble	IPAEOPGP	0.901	28.50	38.00	32.00	23.80	40.80	12.40	38.50	22.20	53.53	7.00	NA	NA
		Any-Acute	0.903	28.30	37.40	31.40	23.70	39.90	12.60	37.50	21.60	52.10	6.96	NA	NA
	Orig. CPM	All	NA	34.00	22.60	25.20	19.20	63.20	16.40	28.80	19.40	70.70	9.80	30.60	14.60
	Orig. PARR	All	NA	28.40	24.60	16.40	23.00	49.00	21.00	28.80	24.40	69.10	9.40	14.80	8.60
500		PARR-2-Settings	0.871	32.20	70.80	54.00	29.60	63.20	20.60	65.20	44.80	77.47	9.50	NA	NA
	Ensemble	IPAEOPGP	0.943	31.20	42.60	37.60	26.80	46.20	12.60	44.80	25.00	55.54	7.65	NA	NA
		Any-Acute	0.947	30.20	41.40	36.80	27.60	45.40	12.60	44.00	24.60	53.95	7.57	NA	NA
	Orig. CPM	All	NA	40.80	28.40	24.40	21.20	62.80	17.60	29.60	23.20	69.50	8.90	32.00	18.00
	Orig. PARR	All	NA	30.80	26.80	20.00	22.80	52.40	20.80	30.80	29.20	68.80	9.80	14.40	10.00
250		PARR-2-Settings	0.923	34.80	75.60	56.00	25.60	64.00	22.00	70.80	46.80	76.87	9.08	NA	NA
	Ensemble	IPAEOPGP	0.977	30.80	42.00	36.80	27.60	44.80	11.20	48.40	24.40	54.74	7.36	NA	NA
		Any-Acute	0.979	29.60	40.80	36.40	26.00	44.40	10.80	46.80	23.20	53.46	7.10	NA	NA

Continued on next page

Risk Seg.	Model	Sub-population	Min Risk	Asthma (%)	COPD (%)	Depres. (%)	Diab. (%)	Hyper. (%)	Cancer (%)	CHD (%)	CHF (%)	Avg. Age	Avg. LoS	5-9 Meds	10+ Meds

[a] Risk Seg.: The top predicted risk segment. [b] Model: The selected model. [c] Sub-population: The selected sub-population. [d] Min Risk: The minimum risk in the segment. [e] Asthma (%): The number of patients with a history of Asthma diagnosis (ICD-10: J45-J46). [f] COPD (%): The number of patients with a history of Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-10: J20, J41-J44, J47). [g] Depres. (%): The number of patients with a history of Depression diagnosis (ICD-10: I10-I15). [h] Diab. (%): The number of patients with a history of Diabetes diagnosis (ICD-10: E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9, E10.2-E10.5, E10.7, E11.2-E11.5, E11.7, E12.2-E12.5, E12.7, E13.2-E13.5, E13.7, E14.2-E14.5, E14.7). [i] Hyper. (%): The number of patients with a history of Hypertension diagnosis (ICD-10: I10-I15, I27, I6, I87.0, I87, I97, K76.6, H35.0, R03, O13, O14, O16, O10, G93.2, H40.0, P292, P293). [j] Cancer (%): The number of patients with a history of Cancer diagnosis (ICD-10: C00-D49). [k] CHD (%): The number of patients with a history of Coronary Heart Disease (CHD) diagnosis (ICD-10: I20-I25). [l] CHF (%): The number of patients with a history of Congestive Heart Failure (CHF) diagnosis (ICD-10: I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0). [m] Avg. Age: The average age of patients at the trigger event. [n] Avg. LoS: The average length of stay (LoS) of patient at the trigger event. [o] 5-9 Meds: The number of patients with 5-9 medication prescription. [p] 10+ Meds: The number of patients with 10+ medication prescription. [q] Orig. CPM: The performance of the original CPM model. Some of the values are estimated. [r] Orig. PARR: The performance of the original PARR model. Some of the values are estimated. [s] Ensemble: The performance of the ensemble models. [t] PARR-2-Settings: The performance of the model for the sub-population $Sub\_PARR-2-Settings$. [u] IPAEOPGP: The performance of the model for the sub-population $Sub\_IPAEOPGP$. [v] Any-Acute: The performance of the model for the sub-population $Sub\_Any-Acute$.

Continued on next page

A.6.3.7 Risk Bands Statistics

TABLE A.57: ERMER: The risk bands statistics of of the benchmarking models.

	Orig. PARR-30[a]						Orig. Billings-13 (IP)[b]						Orig. Billings-13 (full)[c]					
#[d]	TP+FP (%)[e]	TP[f]	Prec.[g]	Sens.[h]	Avg.[i]	C.I.[j]	N (%)	TP	Prec.	Sens.	Avg.	C.I.	N (%)	TP	Prec.	Sens.	Avg.	C.I.
1	32,653 (5.66)	NA	NA	NA	3.90	3.60;4.00	1,388,857 (76.27)	40,242	0.052	1.000	NA	NA	1,358,609 (74.61)	33,194	0.052	1.000	NA	NA
2	283,165 (49.09)	NA	NA	NA	7.10	7.00;7.20	281,216 (15.44)	21,397	0.126	0.575	NA	NA	280,636 (15.41)	21,512	0.133	0.649	NA	NA
3	146,626 (25.42)	NA	NA	NA	12.70	12.6;12.9	70,583 (3.88)	10,155	0.219	0.349	NA	NA	80,352 (4.41)	11,289	0.220	0.422	NA	NA
4	48,596 (8.42)	NA	NA	NA	18.90	18.6;19.3	33,578 (1.84)	6,717	0.285	0.242	NA	NA	36,535 (2.01)	7,096	0.283	0.303	NA	NA
5	25,193 (4.37)	NA	NA	NA	23.70	23.2;24.3	13,857 (0.76)	3,511	0.346	0.171	NA	NA	20,762 (1.14)	4,929	0.333	0.228	NA	NA
6	14,282 (2.48)	NA	NA	NA	28.00	27.5;28.9	9,011 (0.49)	2,609	0.385	0.134	NA	NA	12,461 (0.68)	3,476	0.378	0.176	NA	NA
7	8,559 (1.48)	NA	NA	NA	32.00	31.3;33.0	5,791 (0.32)	1,931	0.421	0.106	NA	NA	8,276 (0.45)	2,680	0.417	0.139	NA	NA
8	5,514 (0.96)	NA	NA	NA	36.30	35.1;37.9	4,061 (0.22)	1,430	0.449	0.086	NA	NA	5,636 (0.31)	2,022	0.450	0.111	NA	NA
9	3,472 (0.60)	NA	NA	NA	39.00	37.4;41.0	2,998 (0.16)	1,165	0.477	0.071	NA	NA	4,162 (0.23)	1,573	0.479	0.090	NA	NA
10	2,413 (0.42)	NA	NA	NA	44.90	43.0;46.9	2,301 (0.13)	908	0.501	0.058	NA	NA	3,034 (0.17)	1,252	0.510	0.073	NA	NA
11	1,543 (0.27)	NA	NA	NA	47.70	45.2;50.7	1,738 (0.10)	765	0.529	0.049	NA	NA	2,386 (0.13)	1,088	0.538	0.060	NA	NA
12	1,174 (0.20)	NA	NA	NA	50.60	48.0;53.3	1,366 (0.08)	623	0.551	0.041	NA	NA	1,788 (0.10)	846	0.562	0.048	NA	NA
13	840 (0.15)	NA	NA	NA	54.30	51.1;57.8	1,071 (0.06)	528	0.574	0.034	NA	NA	1,454 (0.08)	701	0.587	0.039	NA	NA
14	617 (0.11)	NA	NA	NA	60.60	56.5;65.1	933 (0.05)	466	0.593	0.029	NA	NA	1,106 (0.06)	581	0.618	0.032	NA	NA
15	518 (0.09)	NA	NA	NA	63.20	59.8;67.2	775 (0.04)	429	0.617	0.024	NA	NA	919 (0.05)	532	0.645	0.026	NA	NA
16	425 (0.07)	NA	NA	NA	65.00	60.1;69.3	735 (0.04)	398	0.634	0.019	NA	NA	760 (0.04)	443	0.666	0.020	NA	NA
17	276 (0.05)	NA	NA	NA	66.30	60.4;72.4	562 (0.03)	354	0.666	0.015	NA	NA	557 (0.03)	364	0.696	0.016	NA	NA
18	289 (0.05)	NA	NA	NA	75.40	70.2;80.6	484 (0.03)	295	0.679	0.011	NA	NA	545 (0.03)	360	0.711	0.012	NA	NA
19	263 (0.05)	NA	NA	NA	83.00	77.6;87.6	444 (0.02)	291	0.710	0.008	NA	NA	455 (0.02)	317	0.738	0.008	NA	NA
20	450 (0.08)	NA	NA	NA	88.70	85.3;91.4	639 (0.04)	478	0.748	0.005	NA	NA	567 (0.03)	437	0.771	0.005	NA	NA
N	576,868.00	NA	NA	NA	12.20	12.1;12.3	1,821,000	94,692	0.53	0.05	NA	NA	1,821,000	94,692	0.54	0.06	NA	NA

k

[a] Orig. PARR-30: The performance of the original PARR-30 model. Some of the values are estimated. [b] Orig. Billings-13 (IP): The performance of the original Billings et. al. (2013) model with inpatient data. Some of the values are estimated. [c] Orig. Billings-13 (full): The performance of the original Billings et. al. (2013) model with inpatient, A&E, outpatient and GP data. Some of the values are estimated. [d] #: The risk band number: 1 = [0, 0.05); 2 = [0.05, 0.10); 3 = [0.10, 0.15); 4 = [0.15, 0.20); 5 = [0.20, 0.25); 6 = [0.25, 0.30); 7 = [0.30, 0.35); 8 = [0.35, 0.40); 9 = [0.40, 0.45); 10 = [0.45, 0.50); 11 = [0.50, 0.55); 12 = [0.55, 0.60); 13 = [0.60, 0.65); 14 = [0.65, 0.70); 15 = [0.70, 0.75); 16 = [0.75, 0.80); 17 = [0.80, 0.85); 18 = [0.85, 0.90); 19 = [0.90, 0.95); 20 = [0.95, 1]. [e] TP+FP: The number of true positives (TPs) and false positives (FPs). [f] Avg.: The average of number of readmissions after the trigger event. [g] C.I.: The confidence interval for the average of readmissions using the bootstrapped central estimate with 95% CI. [h] TP: The number of true positives (TPs). [i] Prec.: The precision measure, also known as the positive predictive value (PPV). [j] Sens.: The sensitivity measure, also known as the true positive rate (TPR), or the recall. [k] N: The total number of patients. [l] PARR-2-Settings: The performance of the model for the sub-population $Sub\_PARR-2-Settings$. [m] IPAEOPGP: The performance of the model for the sub-population $Sub\_IPAEOPGP$. [n] Any-Acute: The performance of the model for the sub-population $Sub\_Any-Acute$.

TABLE A.58: ERMER: The risk bands statistics of of the *Cond_Ensemble* sub-model on the $Sample-1$ data.

	PARR-2-Settings[l]						IPAEOPGP[m]						Any-Acute[n]					
#	TP+FP (%)	TP	Prec.	Sens.	Avg.	C.I.	N (%)	TP	Prec.	Sens.	Avg.	C.I.	N (%)	TP	Prec.	Sens.	Avg.	C.I.
1	14 (0.02)	0	0.000	0.000	0.00	0.00;0.00	2,101 (1.03)	140	0.067	1.000	6.66	0.06;0.08	3,797 (1.64)	240	0.063	1.000	6.30	0.06;0.07
2	103 (0.15)	5	0.049	1.000	4.85	0.00;0.10	8,065 (3.94)	945	0.117	0.871	11.68	0.11;0.12	12,435 (5.37)	1,411	0.113	0.855	11.36	0.11;0.12
3	522 (0.74)	48	0.092	0.906	9.19	0.07;0.12	14,916 (7.29)	2,046	0.137	0.653	13.70	0.13;0.14	20,067 (8.66)	2,675	0.133	0.618	13.33	0.13;0.14
4	2,329 (3.32)	369	0.158	0.874	15.8	0.14;0.17	15,054 (7.36)	2,643	0.176	0.458	17.55	0.17;0.18	20,237 (8.73)	3,457	0.171	0.444	17.08	0.17;0.18
5	3,404 (4.85)	742	0.218	0.637	21.79	0.20;0.23	20,850 (10.19)	3,979	0.191	0.408	19.09	0.19;0.20	24,368 (10.51)	4,613	0.189	0.372	18.92	0.18;0.19
6	6,356 (9.06)	1,832	0.288	0.611	28.80	0.28;0.30	20,969 (10.25)	4,585	0.219	0.320	21.87	0.21;0.22	23,313 (10.06)	5,075	0.218	0.290	21.77	0.21;0.22
7	7,681 (10.95)	2,618	0.341	0.466	34.09	0.33;0.35	21,445 (10.48)	5,593	0.261	0.281	26.09	0.25;0.27	23,063 (9.95)	5,968	0.259	0.255	25.87	0.25;0.26
8	9,604 (13.69)	3,705	0.386	0.398	38.57	0.38;0.40	18,623 (9.10)	6,271	0.337	0.239	33.64	0.33;0.34	19,461 (8.40)	6,523	0.335	0.218	33.49	0.33;0.34
9	11,501 (16.40)	5,080	0.442	0.353	44.18	0.43;0.45	17,265 (8.44)	7,216	0.418	0.216	41.77	0.41;0.43	17,827 (7.69)	7,425	0.417	0.199	41.66	0.41;0.42
10	8,987 (12.81)	4,310	0.480	0.230	47.95	0.47;0.49	13,962 (6.82)	6,896	0.494	0.171	49.38	0.49;0.50	14,345 (6.19)	7,068	0.493	0.159	49.27	0.48;0.50
11	6,913 (9.86)	3,713	0.537	0.166	53.66	0.53;0.55	10,921 (5.34)	6,160	0.564	0.133	56.38	0.55;0.57	11,191 (4.83)	6,313	0.564	0.124	56.38	0.55;0.57
12	4,787 (6.82)	2,737	0.572	0.109	57.21	0.56;0.59	10,140 (4.95)	6,755	0.666	0.127	66.57	0.66;0.67	10,391 (4.48)	6,907	0.665	0.120	66.50	0.66;0.67
13	3,076 (4.39)	1,948	0.633	0.072	63.32	0.62;0.65	10,109 (4.94)	7,426	0.735	0.122	73.43	0.73;0.74	10,357 (4.47)	7,585	0.732	0.116	73.21	0.72;0.74
14	1,879 (2.68)	1,273	0.677	0.045	67.80	0.66;0.70	5,533 (2.70)	4,193	0.758	0.065	75.81	0.75;0.77	5,672 (2.45)	4,296	0.757	0.062	75.74	0.75;0.77
15	1,116 (1.59)	800	0.717	0.027	71.68	0.69;0.74	4,301 (2.10)	3,423	0.796	0.050	79.58	0.78;0.81	4,424 (1.91)	3,517	0.795	0.048	79.49	0.78;0.81
16	721 (1.03)	547	0.759	0.018	75.86	0.73;0.79	2,975 (1.45)	2,447	0.823	0.035	82.31	0.81;0.84	3,089 (1.33)	2,549	0.825	0.034	82.55	0.81;0.84
17	460 (0.66)	364	0.791	0.012	79.13	0.75;0.83	4,595 (2.25)	4,076	0.887	0.054	88.72	0.88;0.90	4,757 (2.05)	4,223	0.888	0.053	88.77	0.88;0.90
18	306 (0.44)	240	0.784	0.008	78.43	0.74;0.83	1,697 (0.83)	1,475	0.869	0.019	86.91	0.85;0.89	1,769 (0.76)	1,542	0.872	0.019	87.22	0.86;0.89
19	199 (0.28)	167	0.839	0.005	83.92	0.79;0.89	597 (0.29)	509	0.853	0.007	85.26	0.82;0.88	619 (0.27)	527	0.851	0.006	85.13	0.82;0.88
20	189 (0.27)	173	0.915	0.006	91.53	0.87;0.95	554 (0.27)	502	0.906	0.006	90.61	0.88;0.93	573 (0.25)	520	0.908	0.006	90.75	0.88;0.93
N	70,147	30,671	0.609	0.390	43.72	0.43;0.44	204,672	77,280	0.719	0.478	37.75	0.38;0.38	231,755	82,434	0.719	0.461	35.56	0.35;0.36

[a] Orig. PARR-30: The performance of the original PARR-30 model. Some of the values are estimated. [b] Orig. Billings-13 (IP): The performance of the original Billings et. al. (2013) model with inpatient data. Some of the values are estimated. [c] Orig. Billings-13 (full): The performance of the original Billings et. al. (2013) model with inpatient, A&E, outpatient and GP data. Some of the values are estimated. [d] #: The risk band number: 1 = [0, 0.05); 2 = [0.05, 0.10); 3 = [0.10, 0.15); 4 = [0.15, 0.20); 5 = [0.20, 0.25); 6 = [0.25, 0.30); 7 = [0.30, 0.35); 8 = [0.35, 0.40); 9 = [0.40, 0.45); 10 = [0.45, 0.50); 11 = [0.50, 0.55); 12 = [0.55, 0.60); 13 = [0.60, 0.65); 14 = [0.65, 0.70); 15 = [0.70, 0.75); 16 = [0.75, 0.80); 17 = [0.80, 0.85); 18 = [0.85, 0.90); 19 = [0.90, 0.95); 20 = [0.95, 1]. [e] TP+FP: The number of true positives (TPs) and false positives (FPs). [f] Avg.: The average of number of readmissions after the trigger event. [g] C.I.: The confidence interval for the average of readmissions using the bootstrapped central estimate with 95% CI. [h] TP: The number of true positives (TPs). [i] Prec.: The precision measure, also known as the positive predictive value (PPV). [j] Sens.: The sensitivity measure, also known as the true positive rate (TPR), or the recall. [k] N: The total number of patients. [l] PARR-2-Settings: The performance of the model for the sub-population $Sub\_PARR-2-Settings$. [m] IPAEOPGP: The performance of the model for the sub-population $Sub\_IPAEOPGP$. [n] Any-Acute: The performance of the model for the sub-population $Sub\_Any-Acute$.

TABLE A.59: ERMER: The risk bands statistics of the *Cond_Ensemble* sub-model (*Sample-2*)

	PARR-2-Settings[k]						IPAEOPGP[l]						Any-Acute[m]					
#	TP+FP (%)	TP	Prec.	Sens.	Avg.	C.I.	N (%)	TP	Prec.	Sens.	Avg.	C.I.	N (%)	TP	Prec.	Sens.	Avg.	C.I.
1	2 (0.00)	2	1.000	1.000	100.00	1.00;1.00	1,397 (0.62)	79	0.057	1.000	5.65	0.05;0.07	1,873 (0.77)	98	0.052	1.000	5.28	0.04;0.06
2	129 (0.18)	13	0.101	0.867	10.07	0.05;0.15	7,513 (3.35)	852	0.113	0.915	11.34	0.11;0.12	10,212 (4.19)	1,125	0.110	0.920	11.00	0.10;0.12
3	951 (1.30)	145	0.152	0.906	15.35	0.13;0.18	13,940 (6.22)	2,132	0.153	0.696	15.29	0.15;0.16	16,565 (6.80)	2,459	0.148	0.668	14.85	0.14;0.15
4	1,212 (1.65)	206	0.170	0.563	16.91	0.15;0.19	16,343 (7.30)	3,168	0.194	0.508	19.39	0.19;0.20	19,209 (7.88)	3,569	0.186	0.492	18.58	0.18;0.19
5	3,574 (4.87)	862	0.241	0.702	24.09	0.23;0.26	18,859 (8.42)	4,402	0.233	0.414	23.34	0.23;0.24	22,137 (9.08)	4,978	0.225	0.407	22.46	0.22;0.23
6	4,013 (5.47)	1,149	0.286	0.483	28.65	0.27;0.30	20,608 (9.20)	4,674	0.227	0.305	22.67	0.22;0.23	22,608 (9.28)	5,066	0.224	0.293	22.40	0.22;0.23
7	6,422 (8.76)	2,030	0.316	0.461	31.59	0.30;0.33	26,838 (11.98)	6,349	0.237	0.293	23.64	0.23;0.24	28,758 (11.80)	6,770	0.235	0.281	23.52	0.23;0.24
8	10,049 (13.71)	3,823	0.380	0.465	38.03	0.37;0.39	23,409 (10.45)	7,266	0.310	0.251	31.04	0.30;0.32	24,519 (10.06)	7,560	0.308	0.239	30.80	0.30;0.31
9	10,945 (14.93)	4,718	0.431	0.364	43.12	0.42;0.44	18,994 (8.48)	7,442	0.392	0.205	39.17	0.39;0.40	19,603 (8.04)	7,682	0.392	0.195	39.19	0.38;0.40
10	10,046 (13.70)	4,985	0.496	0.278	49.63	0.49;0.51	14,871 (6.64)	6,971	0.469	0.161	46.87	0.46;0.48	15,318 (6.29)	7,178	0.469	0.154	46.84	0.46;0.48
11	8,431 (11.50)	4,548	0.539	0.202	53.95	0.53;0.55	13,271 (5.92)	7,597	0.572	0.149	57.26	0.56;0.58	13,670 (5.61)	7,848	0.574	0.144	57.41	0.57;0.58
12	6,420 (8.76)	3,791	0.590	0.144	59.03	0.58;0.60	13,666 (6.10)	9,341	0.684	0.155	68.36	0.68;0.69	14,010 (5.75)	9,573	0.683	0.150	68.31	0.68;0.69
13	4,153 (5.66)	2,557	0.616	0.089	61.54	0.60;0.63	11,454 (5.11)	8,386	0.732	0.122	73.21	0.72;0.74	11,776 (4.83)	8,596	0.730	0.119	72.99	0.72;0.74
14	2,756 (3.76)	1,851	0.672	0.060	67.12	0.66;0.69	7,093 (3.17)	5,354	0.755	0.072	75.51	0.75;0.76	7,277 (2.99)	5,476	0.753	0.070	75.21	0.74;0.76
15	1,679 (2.29)	1,186	0.706	0.037	70.63	0.68;0.73	4,056 (1.81)	3,116	0.768	0.040	76.84	0.75;0.78	4,167 (1.71)	3,200	0.768	0.039	76.81	0.76;0.78
16	1,020 (1.39)	735	0.721	0.023	72.05	0.69;0.75	5,402 (2.41)	4,627	0.857	0.057	85.65	0.85;0.87	5,580 (2.29)	4,783	0.857	0.056	85.71	0.85;0.87
17	647 (0.88)	514	0.794	0.016	79.44	0.76;0.83	2,999 (1.34)	2,575	0.859	0.031	85.86	0.85;0.87	3,076 (1.26)	2,636	0.857	0.030	85.72	0.84;0.87
18	405 (0.55)	331	0.817	0.010	81.72	0.78;0.85	1,691 (0.75)	1,454	0.860	0.017	85.98	0.84;0.88	1,730 (0.71)	1,487	0.860	0.017	85.95	0.84;0.88
19	244 (0.33)	207	0.848	0.006	84.83	0.80;0.89	728 (0.32)	625	0.859	0.007	85.98	0.83;0.88	742 (0.30)	638	0.860	0.007	85.98	0.83;0.88
20	217 (0.30)	196	0.903	0.006	90.32	0.86;0.94	869 (0.39)	783	0.901	0.009	90.21	0.88;0.92	882 (0.36)	795	0.901	0.009	90.24	0.88;0.92
N	73,315	33,849	0.613	0.470	46.18	0.46;0.47	224,001	87,193	0.716	0.503	38.93	0.39;0.39	243,712	91,517	0.716	0.492	37.55	0.37;0.38

TABLE A.60: ERMER: The risk bands statistics of the *Cond_Ensemble* sub-model (*Sample-1-train-half-3-test-half*)

	PARR-2-Settings[k]						**IPAEOPGP**[l]						**Any-Acute**[m]					
#	**TP+FP (%)**	**TP**	**Prec.**	**Sens.**	**Avg.**	**C.I.**	**N (%)**	**TP**	**Prec.**	**Sens.**	**Avg.**	**C.I.**	**N (%)**	**TP**	**Prec.**	**Sens.**	**Avg.**	**C.I.**
1	13 (0.01)	1	0.077	1.000	7.69	0.00;0.23	3,051 (1.14)	203	0.067	1.000	6.65	0.06;0.08	5,281 (1.73)	352	0.067	1.000	6.66	0.06;0.07
2	104 (0.11)	3	0.029	0.750	2.88	0.00;0.07	11,714 (4.36)	1,564	0.134	0.885	13.38	0.13;0.14	17,324 (5.68)	2,165	0.125	0.860	12.51	0.12;0.13
3	677 (0.74)	78	0.115	0.951	11.52	0.09;0.14	19,759 (7.36)	2,816	0.143	0.614	14.24	0.14;0.15	26,158 (8.58)	3,550	0.136	0.585	13.56	0.13;0.14
4	3,229 (3.53)	482	0.149	0.855	14.92	0.14;0.16	20,869 (7.77)	3,677	0.176	0.445	17.62	0.17;0.18	27,760 (9.10)	4,733	0.170	0.438	17.05	0.17;0.17
5	4,596 (5.03)	1,092	0.238	0.659	23.78	0.22;0.25	31,421 (11.70)	6,083	0.194	0.424	19.37	0.19;0.20	37,608 (12.34)	7,252	0.193	0.402	19.30	0.19;0.20
6	9,196 (10.06)	2,807	0.305	0.629	30.51	0.30;0.31	31,522 (11.74)	7,189	0.228	0.334	22.78	0.22;0.23	34,732 (11.39)	7,854	0.226	0.303	22.59	0.22;0.23
7	11,108 (12.16)	4,065	0.366	0.477	36.59	0.36;0.37	28,363 (10.56)	8,370	0.295	0.280	29.50	0.29;0.30	30,379 (9.96)	8,948	0.295	0.257	29.44	0.29;0.30
8	12,992 (14.22)	5,373	0.414	0.387	41.34	0.41;0.42	23,247 (8.66)	8,738	0.376	0.226	37.58	0.37;0.38	24,173 (7.93)	9,058	0.375	0.206	37.46	0.37;0.38
9	15,240 (16.68)	7,179	0.471	0.341	47.12	0.46;0.48	21,255 (7.91)	9,853	0.464	0.203	46.36	0.46;0.47	21,930 (7.19)	10,116	0.461	0.187	46.11	0.46;0.47
10	11,863 (12.98)	6,072	0.512	0.224	51.19	0.50;0.52	16,859 (6.28)	8,816	0.523	0.154	52.31	0.52;0.53	17,330 (5.68)	9,067	0.523	0.144	52.32	0.52;0.53
11	8,348 (9.14)	4,584	0.549	0.144	54.89	0.54;0.56	13,268 (4.94)	7,798	0.588	0.120	58.76	0.58;0.60	13,602 (4.46)	7,991	0.587	0.112	58.74	0.58;0.60
12	5,652 (6.19)	3,477	0.615	0.099	61.53	0.60;0.63	11,605 (4.32)	8,149	0.702	0.111	70.19	0.69;0.71	11,858 (3.89)	8,305	0.700	0.105	70.03	0.69;0.71
13	3,452 (3.78)	2,254	0.653	0.060	65.29	0.64;0.67	11,069 (4.12)	8,278	0.748	0.102	74.78	0.74;0.76	11,394 (3.74)	8,479	0.744	0.096	74.39	0.74;0.75
14	2,003 (2.19)	1,351	0.674	0.035	67.39	0.65;0.70	6,086 (2.27)	4,391	0.721	0.051	72.16	0.71;0.73	6,242 (2.05)	4,497	0.720	0.049	72.04	0.71;0.73
15	1,123 (1.23)	844	0.752	0.021	75.15	0.73;0.77	4,851 (1.81)	3,925	0.809	0.044	80.91	0.80;0.82	4,998 (1.64)	4,049	0.810	0.042	81.03	0.80;0.82
16	694 (0.76)	559	0.805	0.014	80.54	0.77;0.84	3,526 (1.31)	2,986	0.847	0.032	84.69	0.84;0.86	3,656 (1.20)	3,101	0.848	0.031	84.76	0.84;0.86
17	446 (0.49)	367	0.823	0.009	82.28	0.78;0.85	6,851 (2.55)	6,231	0.910	0.063	90.95	0.90;0.92	7,061 (2.32)	6,424	0.910	0.061	90.96	0.90;0.92
18	263 (0.29)	228	0.867	0.006	86.69	0.82;0.91	2,252 (0.84)	2,056	0.913	0.020	91.29	0.90;0.92	2,348 (0.77)	2,145	0.914	0.020	91.35	0.90;0.92
19	202 (0.22)	183	0.906	0.004	90.59	0.87;0.94	559 (0.21)	507	0.907	0.005	90.69	0.88;0.93	584 (0.19)	529	0.906	0.005	90.75	0.88;0.93
20	168 (0.18)	156	0.929	0.004	92.85	0.89;0.96	448 (0.17)	409	0.913	0.004	91.29	0.89;0.94	470 (0.15)	430	0.915	0.004	91.48	0.89;0.94
N	91,369	41,155	0.627	0.340	45.04	0.45;0.45	268,575	102,039	0.739	0.438	37.99	0.38;0.38	304,888	109,045	0.739	0.421	35.76	0.36;0.36

A.7 T-CARER

A.7.1 Features

TABLE A.61: T-CARER: The weights of features using the Random Forest method (*Sample-1*)

#	Name	Weight	Temporal	Definition
1	admimeth_0t30d_prevalence_2_cnt	1.30E-01	0-30	Admission method: Unknown
2	admimeth_0t30d_prevalence_1_cnt	1.21E-01	0-30	Admission method: Elective
3	mainspef_0t30d_prevalence_5_cnt	5.83E-02	0-30	Main Speciality: Maternity
4	epidur_0t30d_avg	5.68E-02	0-30	Average episode duration
5	diagCCS_0t30d_prevalence_16_cnt	4.63E-02	0-30	CCS group: Other complications of pregnancy
6	ethnos_0	4.13E-02	Trigger	ethnicity: Not known
7	posopdur_0t30d_avg	4.06E-02	0-30	Average Post-operation duration
8	diagCCS_0t30d_others_cnt	2.33E-02	0-30	CCS group: Others
9	epidur_365t730d_avg	2.07E-02	365-730	Average episode duration
10	operOPCSL1_0t30d_others_cnt	1.97E-02	0-30	Operation group: Others
11	diagCCS_0t30d_prevalence_5_cnt	1.91E-02	0-30	CCS group: Normal pregnancy and delivery
12	epidur_0t30d_others_cnt	1.80E-02	0-30	Episode duration
13	ethnos_1	1.72E-02	Trigger	ethnicity: White
14	mainspef_0t30d_prevalence_4_cnt	1.65E-02	0-30	Main Speciality: Gynaecology
15	preopdur_0t30d_avg	1.59E-02	0-30	Average Post-operation duration
16	gender_1	1.56E-02	Trigger	Gender: Male
17	epidur_365t730d_others_cnt	1.45E-02	365-730	Episode duration
18	mainspef_0t30d_prevalence_3_cnt	1.37E-02	0-30	Main Speciality: Plastic
19	admimeth_180t365d_prevalence_2_cnt	1.35E-02	180-365	Admission method: Unknown
20	admimeth_365t730d_prevalence_2_cnt	1.20E-02	365-730	Admission method: Unknown
21	gapDays_365t730d_others_cnt	1.08E-02	365-730	Gap-Days
22	operOPCSL1_0t30d_prevalence_7_cnt	1.08E-02	0-30	Operation group: Female Genital Tract
23	preopdur_365t730d_others_cnt	1.06E-02	365-730	Post-operation duration
24	mainspef_0t30d_prevalence_1_cnt	1.02E-02	0-30	Main Speciality: General
25	diagCCS_365t730d_others_cnt	9.16E-03	365-730	CCS group: Others
26	admimeth_365t730d_others_cnt	8.87E-03	365-730	Admission method: Others
27	epidur_180t365d_avg	8.05E-03	180-365	Average episode duration
28	diagCCS_180t365d_others_cnt	8.04E-03	180-365	CCS group: Others
29	diagCCS_0t30d_prevalence_1_cnt	7.50E-03	0-30	CCS group: Others
30	mainspef_0t30d_prevalence_2_cnt	7.10E-03	0-30	Main Speciality: General Surgery
31	gapDays_180t365d_others_cnt	6.92E-03	180-365	Gap-Days
32	diagCCS_0t30d_prevalence_8_cnt	6.63E-03	0-30	CCS group: Abdominal pain
33	posopdur_365t730d_others_cnt	6.21E-03	365-730	Post-operation duration
34	epidur_180t365d_others_cnt	5.89E-03	180-365	Episode duration
35	mainspef_0t30d_others_cnt	5.46E-03	0-30	Main Speciality: Others
36	gapDays_180t365d_avg	5.31E-03	180-365	Average Gap-Days
37	gapDays_365t730d_avg	5.20E-03	365-730	Average Gap-Days
38	preopdur_30t90d_others_cnt	4.97E-03	30-90	Post-operation duration
39	diagCCS_0t30d_prevalence_19_cnt	4.88E-03	0-30	CCS group: Other screening for suspected conditions (not mental disorders or infectious disease)
40	admimeth_90t180d_prevalence_2_cnt	4.10E-03	90-180	Admission method: Unknown
41	admimeth_365t730d_prevalence_1_cnt	4.04E-03	365-730	Admission method: Elective
42	operOPCSL1_0t30d_prevalence_4_cnt	3.98E-03	0-30	Operation group: Miscellaneous Operations
43	posopdur_180t365d_others_cnt	3.92E-03	180-365	Post-operation duration
44	diagCCS_0t30d_prevalence_24_cnt	3.87E-03	0-30	CCS group: Chronic obstructive pulmonary disease & bronchiectasis
45	diagCCS_0t30d_prevalence_4_cnt	3.76E-03	0-30	CCS group: Coronary atherosclerosis & other heart disease
46	mainspef_0t30d_prevalence_8_cnt	3.60E-03	0-30	Main Speciality: Geriatric
47	mainspef_90t180d_others_cnt	3.56E-03	90-180	Main Speciality: Others
48	epidur_90t180d_others_cnt	3.50E-03	90-180	Episode duration
49	preopdur_90t180d_others_cnt	3.44E-03	90-180	Post-operation duration
50	operOPCSL1_0t30d_prevalence_3_cnt	3.21E-03	0-30	Operation group: Upper Female Genital Tract
51	epidur_90t180d_avg	3.12E-03	90-180	Average episode duration
52	posopdur_30t90d_others_cnt	2.60E-03	30-90	Post-operation duration
53	diagCCS_30t90d_others_cnt	2.46E-03	30-90	CCS group: Others
54	admimeth_90t180d_others_cnt	2.38E-03	90-180	Admission method: Others
55	mainspef_180t365d_others_cnt	2.29E-03	180-365	Main Speciality: Others
56	admimeth_30t90d_prevalence_2_cnt	2.21E-03	30-90	Admission method: Unknown
57	posopdur_0t30d_others_cnt	2.18E-03	0-30	Post-operation duration
58	diagCCS_90t180d_others_cnt	1.98E-03	90-180	CCS group: Others
59	admimeth_180t365d_others_cnt	1.98E-03	180-365	Admission method: Others
60	posopdur_90t180d_others_cnt	1.98E-03	90-180	Post-operation duration
61	mainspef_0t30d_prevalence_9_cnt	1.94E-03	0-30	Main Speciality: Ear, nose & throat
62	diagCCS_0t30d_prevalence_10_cnt	1.87E-03	0-30	CCS group: Cardiac dysrhythmias
63	gapDays_90t180d_avg	1.82E-03	90-180	Average Gap-Days
64	preopdur_180t365d_others_cnt	1.76E-03	180-365	Post-operation duration
65	diagCCS_180t365d_prevalence_1_cnt	1.68E-03	180-365	CCS group: Others
66	gapDays_90t180d_others_cnt	1.62E-03	90-180	Gap-Days

Continued on next page

#	Name	Weight	Temporal	Definition
67	diagCCS_0t30d_prevalence_2_cnt	1.46E-03	0-30	CCS group: Residual codes; unclassified
68	epidur_30t90d_avg	1.42E-03	30-90	Average episode duration
69	admimeth_180t365d_prevalence_1_cnt	1.40E-03	180-365	Admission method: Elective
70	diagCCS_0t30d_prevalence_6_cnt	1.39E-03	0-30	CCS group: Other upper respiratory disease
71	diagCCS_365t730d_prevalence_1_cnt	1.39E-03	365-730	CCS group: Others
72	mainspef_30t90d_others_cnt	1.38E-03	30-90	Main Speciality: Others
73	diagCCS_0t30d_prevalence_3_cnt	1.35E-03	0-30	CCS group: Essential hypertension
74	mainspef_180t365d_prevalence_1_cnt	1.32E-03	180-365	Main Speciality: General
75	mainspef_0t30d_prevalence_7_cnt	1.28E-03	0-30	Main Speciality: Ophthalmology
76	admimeth_30t90d_others_cnt	1.25E-03	30-90	Admission method: Others
77	gapDays_30t90d_others_cnt	1.25E-03	30-90	Gap-Days
78	gapDays_30t90d_avg	1.17E-03	30-90	Average Gap-Days
79	epidur_30t90d_others_cnt	1.14E-03	30-90	Episode duration
80	diagCCS_0t30d_prevalence_20_cnt	1.14E-03	0-30	CCS group: Other lower respiratory disease
81	diagCCS_0t30d_prevalence_27_cnt	1.09E-03	0-30	CCS group: Other female genital disorders
82	posopdur_365t730d_avg	1.07E-03	365-730	Average Post-operation duration
83	operOPCSL1_0t30d_prevalence_10_cnt	1.06E-03	0-30	Operation group: Male Genital Organs
84	diagCCS_365t730d_prevalence_5_cnt	1.04E-03	365-730	CCS group: Normal pregnancy and/or delivery
85	diagCCS_0t30d_prevalence_28_cnt	1.03E-03	0-30	CCS group: Deficiency & other anemia
86	admimeth_90t180d_prevalence_1_cnt	1.02E-03	90-180	Admission method: Elective
87	diagCCS_0t30d_prevalence_9_cnt	1.02E-03	0-30	CCS group: Cataract
88	imd04rk_1	1.02E-03	Trigger	imd04rk: 0 to 3248
89	diagCCS_0t30d_prevalence_7_cnt	9.86E-04	0-30	CCS group: Diabetes mellitus without complication
90	imd04rk_2	9.74E-04	Trigger	imd04rk: 3248 to 6496
91	mainspef_0t30d_prevalence_10_cnt	9.63E-04	0-30	Main Speciality: Cardiothoracic
92	ageTrigger_80	8.88E-04	Trigger	Age: 80-85
93	operOPCSL1_0t30d_prevalence_2_cnt	8.51E-04	0-30	Operation group: Lower Digestive Tract
94	gapDays_0t30d_others_cnt	8.45E-04	0-30	Gap-Days
95	diagCCS_0t30d_prevalence_30_cnt	8.39E-04	0-30	CCS group: Other aftercare
96	admimeth_30t90d_prevalence_1_cnt	8.16E-04	30-90	Admission method: Elective
97	diagCCS_0t30d_prevalence_21_cnt	8.13E-04	0-30	CCS group: Other & unspecified benign neoplasm
98	operOPCSL1_0t30d_prevalence_6_cnt	7.79E-04	0-30	Operation group: Mental Health
99	ethnos_9	7.17E-04	Trigger	ethnicity: Any other
100	diagCCS_0t30d_prevalence_23_cnt	6.87E-04	0-30	CCS group: Other skin disorders
101	imd04rk_3	6.52E-04	Trigger	imd04rk: 6496 to 9745
102	preopdur_0t30d_others_cnt	6.48E-04	0-30	Post-operation duration
103	posopdur_180t365d_avg	6.39E-04	180-365	Average Post-operation duration
104	diagCCS_0t30d_prevalence_13_cnt	6.33E-04	0-30	CCS group: Other gastrointestinal disorders
105	diagCCS_0t30d_prevalence_15_cnt	6.30E-04	0-30	CCS group: Other connective tissue disease
106	mainspef_180t365d_prevalence_5_cnt	5.80E-04	180-365	Main Speciality: Gynaecology
107	diagCCS_0t30d_prevalence_12_cnt	5.78E-04	0-30	CCS group: Abdominal hernia
108	diagCCS_365t730d_prevalence_4_cnt	5.77E-04	365-730	CCS group: Coronary atherosclerosis & other heart disease
109	imd04rk_4	5.76E-04	Trigger	imd04rk: 9745 to 12993
110	mainspef_0t30d_prevalence_6_cnt	5.75E-04	0-30	Main Speciality: Urology
111	operOPCSL1_365t730d_others_cnt	5.64E-04	365-730	Operation group: Others
112	diagCCS_0t30d_prevalence_14_cnt	5.63E-04	0-30	CCS group: External cause codes: Fall
113	diagCCS_0t30d_prevalence_17_cnt	5.55E-04	0-30	CCS group: Osteoarthritis
114	preopdur_365t730d_avg	5.40E-04	365-730	Average Post-operation duration
115	diagCCS_0t30d_prevalence_11_cnt	5.14E-04	0-30	CCS group: Asthma
116	ageTrigger_75	5.04E-04	Trigger	Age: 75-80
117	diagCCS_365t730d_prevalence_21_cnt	5.04E-04	365-730	CCS group: Chronic obstructive pulmonary disease & bronchiectasis
118	operOPCSL1_0t30d_prevalence_16_cnt	4.98E-04	0-30	Operation group: Other Bones & Joints
119	diagCCS_0t30d_prevalence_18_cnt	4.92E-04	0-30	CCS group: Genitourinary symptoms & ill-defined conditions
120	operOPCSL1_0t30d_prevalence_9_cnt	4.88E-04	0-30	Operation group: Heart
121	ageTrigger_30	4.86E-04	Trigger	Age: 30-35
122	imd04rk_6	4.82E-04	Trigger	imd04rk: 16241 to 19489
123	operOPCSL1_0t30d_prevalence_5_cnt	4.79E-04	0-30	Operation group: Upper Digestive Tract
124	ageTrigger_85	4.77E-04	Trigger	Age: 85+
125	imd04rk_5	4.72E-04	Trigger	imd04rk: 12993 to 16241
126	imd04rk_9	4.69E-04	Trigger	imd04rk: 25986 to 29234
127	imd04rk_7	4.68E-04	Trigger	imd04rk: 19489 to 22737
128	mainspef_90t180d_prevalence_1_cnt	4.54E-04	90-180	Main Speciality: General
129	imd04rk_8	4.46E-04	Trigger	imd04rk: 22737 to 25986
130	operOPCSL1_0t30d_prevalence_1_cnt	4.11E-04	0-30	Operation group: Urinary
131	diagCCS_0t30d_prevalence_26_cnt	4.11E-04	0-30	CCS group: Esophageal disorders
132	mainspef_180t365d_prevalence_3_cnt	4.09E-04	180-365	Main Speciality: Plastic
133	diagCCS_0t30d_prevalence_22_cnt	4.04E-04	0-30	CCS group: Other nervous system disorders

Continued on next page

#	Name	Weight	Temporal	Definition
134	diagCCS_180t365d_prevalence_4_cnt	3.70E-04	180-365	CCS group: Coronary atherosclerosis & other heart disease
135	diagCCS_365t730d_prevalence_8_cnt	3.31E-04	365-730	CCS group: Diabetes mellitus without complication
136	preopdur_180t365d_avg	3.10E-04	180-365	Average Post-operation duration
137	ethnos_2	3.06E-04	Trigger	ethnicity: Indian
138	operOPCSL1_365t730d_prevalence_2_cnt	3.04E-04	365-730	Operation group: Miscellaneous Operations
139	diagCCS_365t730d_prevalence_24_cnt	3.03E-04	365-730	CCS group: Other complications of pregnancy
140	operOPCSL1_180t365d_others_cnt	2.87E-04	180-365	Operation group: Others
141	admimeth_0t30d_others_cnt	2.87E-04	0-30	Admission method: Others
142	diagCCS_180t365d_prevalence_14_cnt	2.78E-04	180-365	CCS group: Chronic obstructive pulmonary disease & bronchiectasis
143	ethnos_3	2.71E-04	Trigger	ethnicity: Pakistani
144	diagCCS_365t730d_prevalence_11_cnt	2.71E-04	365-730	CCS group: Cataract
145	diagCCS_0t30d_prevalence_25_cnt	2.69E-04	0-30	CCS group: Allergic reactions
146	gapDays_0t30d_avg	2.52E-04	0-30	Average Gap-Days
147	ethnos_7	2.41E-04	Trigger	ethnicity: Black - Other
148	mainspef_180t365d_prevalence_7_cnt	2.30E-04	180-365	Main Speciality: Geriatric
149	ageTrigger_35	2.27E-04	Trigger	Age: 35-40
150	diagCCS_0t30d_prevalence_29_cnt	2.20E-04	0-30	CCS group: Spondylosis; intervertebral disc disorders; other back problems
151	ageTrigger_20	2.18E-04	Trigger	Age: 20-25
152	mainspef_30t90d_prevalence_1_cnt	2.13E-04	30-90	Main Speciality: General
153	posopdur_90t180d_avg	2.11E-04	90-180	Average Post-operation duration
154	ageTrigger_55	2.08E-04	Trigger	Age: 55-60
155	operOPCSL1_365t730d_prevalence_5_cnt	2.08E-04	365-730	Operation group: Upper Female Genital Tract
156	diagCCS_365t730d_prevalence_2_cnt	2.05E-04	365-730	CCS group: Residual codes; unclassified
157	operOPCSL1_0t30d_prevalence_11_cnt	2.04E-04	0-30	Operation group: Respiratory Tract
158	ageTrigger_25	2.02E-04	Trigger	Age: 25-30
159	diagCCS_90t180d_prevalence_1_cnt	1.99E-04	90-180	CCS group: Others
160	diagCCS_180t365d_prevalence_24_cnt	1.92E-04	180-365	CCS group: Normal pregnancy and/or delivery
161	ageTrigger_50	1.91E-04	Trigger	Age: 50-55
162	operOPCSL1_180t365d_prevalence_2_cnt	1.88E-04	180-365	Operation group: Miscellaneous Operations
163	diagCCS_180t365d_prevalence_2_cnt	1.87E-04	180-365	CCS group: Residual codes; unclassified
164	mainspef_180t365d_prevalence_2_cnt	1.67E-04	180-365	Main Speciality: General Surgery
165	ageTrigger_40	1.60E-04	Trigger	Age: 40-45
166	posopdur_30t90d_avg	1.58E-04	30-90	Average Post-operation duration
167	mainspef_90t180d_prevalence_5_cnt	1.56E-04	90-180	Main Speciality: Gynaecology
168	diagCCS_365t730d_prevalence_12_cnt	1.55E-04	365-730	CCS group: Cardiac dysrhythmias
169	diagCCS_365t730d_prevalence_6_cnt	1.52E-04	365-730	CCS group: Other gastrointestinal disorders
170	diagCCS_365t730d_prevalence_3_cnt	1.47E-04	365-730	CCS group: Essential hypertension
171	mainspef_180t365d_prevalence_4_cnt	1.47E-04	180-365	Main Speciality: Urology
172	diagCCS_30t90d_prevalence_7_cnt	1.45E-04	30-90	CCS group: Other complications of pregnancy
173	operOPCSL1_365t730d_prevalence_1_cnt	1.42E-04	365-730	Operation group: Urinary
174	ageTrigger_70	1.35E-04	Trigger	Age: 70-75
175	ageTrigger_45	1.34E-04	Trigger	Age: 45-50
176	ageTrigger_15	1.31E-04	Trigger	Age: 15-20
177	ageTrigger_65	1.31E-04	Trigger	Age: 65-70
178	diagCCS_365t730d_prevalence_18_cnt	1.28E-04	365-730	CCS group: Other lower respiratory disease
179	mainspef_30t90d_prevalence_5_cnt	1.18E-04	30-90	Main Speciality: Maternity
180	ageTrigger_60	1.12E-04	Trigger	Age: 60-65
181	diagCCS_90t180d_prevalence_21_cnt	1.11E-04	90-180	CCS group: Other complications of pregnancy
182	mainspef_180t365d_prevalence_10_cnt	1.10E-04	180-365	Main Speciality: Maternity
183	mainspef_30t90d_prevalence_6_cnt	1.10E-04	30-90	Main Speciality: Plastic
184	mainspef_90t180d_prevalence_9_cnt	1.09E-04	90-180	Main Speciality: Maternity
185	diagCCS_365t730d_prevalence_7_cnt	1.08E-04	365-730	CCS group: Other upper respiratory disease
186	diagCCS_365t730d_prevalence_10_cnt	1.07E-04	365-730	CCS group: Abdominal pain
187	operOPCSL1_30t90d_prevalence_1_cnt	1.07E-04	30-90	Operation group: Miscellaneous Operations
188	operOPCSL1_30t90d_others_cnt	1.06E-04	30-90	Operation group: Others
189	operOPCSL1_0t30d_prevalence_8_cnt	1.03E-04	0-30	Operation group: Soft Tissue
190	diagCCS_365t730d_prevalence_9_cnt	1.02E-04	365-730	CCS group: Other aftercare
191	diagCCS_30t90d_prevalence_1_cnt	9.74E-05	30-90	CCS group: Others
192	diagCCS_180t365d_prevalence_3_cnt	9.67E-05	180-365	CCS group: Essential hypertension
193	diagCCS_180t365d_prevalence_6_cnt	9.60E-05	180-365	CCS group: Other aftercare
194	diagCCS_180t365d_prevalence_10_cnt	9.35E-05	180-365	CCS group: Cancer; other & unspecified primary
195	operOPCSL1_0t30d_prevalence_14_cnt	9.26E-05	0-30	Operation group: Bones & Joints of Skull & Spine
196	operOPCSL1_90t180d_others_cnt	9.20E-05	90-180	Operation group: Others
197	operOPCSL1_365t730d_prevalence_3_cnt	9.16E-05	365-730	Operation group: Lower Digestive Tract
198	preopdur_30t90d_avg	9.06E-05	30-90	Average Post-operation duration
199	mainspef_90t180d_prevalence_4_cnt	9.05E-05	90-180	Main Speciality: Plastic
200	preopdur_90t180d_avg	8.95E-05	90-180	Average Post-operation duration

Continued on next page

#	Name	Weight	Temporal	Definition
201	ethnos_6	8.58E-05	Trigger	ethnicity: Black - African
202	diagCCS_365t730d_prevalence_28_cnt	8.47E-05	365-730	CCS group: Other & unspecified benign neoplasm
203	diagCCS_365t730d_prevalence_16_cnt	8.37E-05	365-730	CCS group: Cancer; other & unspecified primary
204	diagCCS_180t365d_prevalence_11_cnt	8.32E-05	180-365	CCS group: Other upper respiratory disease
205	mainspef_30t90d_prevalence_4_cnt	7.64E-05	30-90	Main Speciality: Gynaecology
206	ethnos_5	7.57E-05	Trigger	ethnicity: Black - Caribbean
207	operOPCSL1_0t30d_prevalence_12_cnt	7.54E-05	0-30	Operation group: Nervous System
208	mainspef_180t365d_prevalence_6_cnt	7.06E-05	180-365	Main Speciality: Ophthalmology
209	diagCCS_365t730d_prevalence_22_cnt	6.83E-05	365-730	CCS group: Spondylosis; intervertebral disc disorders; other back problems
210	diagCCS_365t730d_prevalence_19_cnt	6.46E-05	365-730	CCS group: Osteoarthritis
211	diagCCS_365t730d_prevalence_20_cnt	6.30E-05	365-730	CCS group: Other screening for suspected conditions (not mental disorders or infectious disease)
212	diagCCS_180t365d_prevalence_7_cnt	6.28E-05	180-365	CCS group: Cataract
213	ethnos_4	6.28E-05	Trigger	ethnicity: Bangladeshi
214	diagCCS_90t180d_prevalence_3_cnt	6.26E-05	90-180	CCS group: Coronary atherosclerosis & other heart disease
215	admimeth_0t30d_prevalence_3_cnt	5.98E-05	0-30	Admission method: Acute
216	diagCCS_365t730d_prevalence_27_cnt	5.91E-05	365-730	CCS group: Urinary tract infections
217	diagCCS_365t730d_prevalence_17_cnt	5.82E-05	365-730	CCS group: Abdominal hernia
218	diagCCS_365t730d_prevalence_30_cnt	5.75E-05	365-730	CCS group: Phlebitis; thrombophlebitis & thromboembolism
219	mainspef_30t90d_prevalence_10_cnt	5.71E-05	30-90	Main Speciality: Ophthalmology
220	mainspef_90t180d_prevalence_2_cnt	5.58E-05	90-180	Main Speciality: General Surgery
221	diagCCS_90t180d_prevalence_10_cnt	5.46E-05	90-180	CCS group: Other aftercare
222	diagCCS_365t730d_prevalence_23_cnt	5.38E-05	365-730	CCS group: Other nervous system disorders
223	diagCCS_365t730d_prevalence_26_cnt	5.22E-05	365-730	CCS group: Esophageal disorders
224	diagCCS_180t365d_prevalence_8_cnt	5.11E-05	180-365	CCS group: Other gastrointestinal disorders
225	diagCCS_365t730d_prevalence_14_cnt	5.08E-05	365-730	CCS group: Genitourinary symptoms & ill-defined conditions
226	diagCCS_365t730d_prevalence_13_cnt	4.95E-05	365-730	CCS group: Asthma
227	diagCCS_365t730d_prevalence_29_cnt	4.93E-05	365-730	CCS group: Complication of device; implant or graft
228	diagCCS_180t365d_prevalence_21_cnt	4.90E-05	180-365	CCS group: Abdominal hernia
229	operOPCSL1_365t730d_prevalence_10_cnt	4.72E-05	365-730	Operation group: Mental Health
230	operOPCSL1_180t365d_prevalence_1_cnt	4.71E-05	180-365	Operation group: Urinary
231	operOPCSL1_365t730d_prevalence_6_cnt	4.62E-05	365-730	Operation group: Female Genital Tract
232	diagCCS_180t365d_prevalence_5_cnt	4.58E-05	180-365	CCS group: Diabetes mellitus without complication
233	operOPCSL1_0t30d_prevalence_15_cnt	4.31E-05	0-30	Operation group: Arteries & Veins
234	operOPCSL1_90t180d_prevalence_2_cnt	4.07E-05	90-180	Operation group: Urinary
235	mainspef_90t180d_prevalence_3_cnt	4.00E-05	90-180	Main Speciality: Urology
236	diagCCS_180t365d_prevalence_12_cnt	3.97E-05	180-365	CCS group: Abdominal pain
237	diagCCS_90t180d_prevalence_2_cnt	3.93E-05	90-180	CCS group: Residual codes; unclassified
238	diagCCS_365t730d_prevalence_15_cnt	3.72E-05	365-730	CCS group: Other connective tissue disease
239	operOPCSL1_180t365d_prevalence_5_cnt	3.49E-05	180-365	Operation group: Upper Female Genital Tract
240	operOPCSL1_365t730d_prevalence_4_cnt	3.40E-05	365-730	Operation group: Upper Digestive Tract
241	diagCCS_180t365d_prevalence_22_cnt	3.35E-05	180-365	CCS group: Other nervous system disorders
242	diagCCS_30t90d_prevalence_5_cnt	3.27E-05	30-90	CCS group: Maintenance chemotherapy; radiotherapy
243	diagCCS_180t365d_prevalence_18_cnt	3.23E-05	180-365	CCS group: Spondylosis; intervertebral disc disorders; other back problems
244	diagCCS_90t180d_prevalence_8_cnt	2.95E-05	90-180	CCS group: Abdominal pain
245	operOPCSL1_0t30d_prevalence_13_cnt	2.95E-05	0-30	Operation group: Lower Female Genital Tract
246	diagCCS_30t90d_prevalence_12_cnt	2.85E-05	30-90	CCS group: Secondary malignancies
247	mainspef_90t180d_prevalence_6_cnt	2.79E-05	90-180	Main Speciality: Geriatric
248	mainspef_30t90d_prevalence_3_cnt	2.75E-05	30-90	Main Speciality: Urology
249	diagCCS_30t90d_prevalence_2_cnt	2.42E-05	30-90	CCS group: Residual codes; unclassified
250	diagCCS_180t365d_prevalence_27_cnt	2.27E-05	180-365	CCS group: Administrative/social admission
251	diagCCS_30t90d_prevalence_20_cnt	2.15E-05	30-90	CCS group: Other screening for suspected conditions (not mental disorders or infectious disease)
252	diagCCS_90t180d_prevalence_4_cnt	2.05E-05	90-180	CCS group: Essential hypertension
253	diagCCS_180t365d_prevalence_19_cnt	2.04E-05	180-365	CCS group: Other screening for suspected conditions (not mental disorders or infectious disease)
254	mainspef_30t90d_prevalence_8_cnt	2.04E-05	30-90	Main Speciality: Haemotology
255	operOPCSL1_365t730d_prevalence_12_cnt	2.03E-05	365-730	Operation group: Arteries & Veins
256	diagCCS_180t365d_prevalence_28_cnt	1.96E-05	180-365	CCS group: Phlebitis; thrombophlebitis & thromboembolism
257	diagCCS_180t365d_prevalence_23_cnt	1.96E-05	180-365	CCS group: Osteoarthritis
258	diagCCS_180t365d_prevalence_16_cnt	1.95E-05	180-365	CCS group: Other lower respiratory disease
259	diagCCS_90t180d_prevalence_20_cnt	1.85E-05	90-180	CCS group: Administrative/social admission

Continued on next page

#	Name	Weight	Temporal	Definition
260	diagCCS_180t365d_prevalence_17_cnt	1.81E-05	180-365	CCS group: Other connective tissue disease
261	diagCCS_180t365d_prevalence_26_cnt	1.80E-05	180-365	CCS group: Congestive heart failure; nonhypertensive
262	diagCCS_90t180d_prevalence_19_cnt	1.80E-05	90-180	CCS group: Maintenance chemotherapy; radiotherapy
263	diagCCS_180t365d_prevalence_9_cnt	1.77E-05	180-365	CCS group: Cardiac dysrhythmias
264	mainspef_90t180d_prevalence_8_cnt	1.65E-05	90-180	Main Speciality: Cardiothoracic
265	operOPCSL1_365t730d_prevalence_8_cnt	1.60E-05	365-730	Operation group: Nervous System
266	diagCCS_180t365d_prevalence_29_cnt	1.56E-05	180-365	CCS group: Urinary tract infections
267	diagCCS_90t180d_prevalence_18_cnt	1.52E-05	90-180	CCS group: Other screening for suspected conditions (not mental disorders or infectious disease)
268	operOPCSL1_180t365d_prevalence_3_cnt	1.48E-05	180-365	Operation group: Lower Digestive Tract
269	operOPCSL1_180t365d_prevalence_4_cnt	1.47E-05	180-365	Operation group: Upper Digestive Tract
270	mainspef_180t365d_prevalence_9_cnt	1.43E-05	180-365	Main Speciality: Gastroenterology
271	diagCCS_30t90d_prevalence_3_cnt	1.39E-05	30-90	CCS group: Coronary atherosclerosis & other heart disease
272	diagCCS_30t90d_prevalence_21_cnt	1.38E-05	30-90	CCS group: Cancer of breast
273	operOPCSL1_0t30d_prevalence_17_cnt	1.37E-05	0-30	Operation group: Other Abdominal Organs
274	operOPCSL1_365t730d_prevalence_7_cnt	1.24E-05	365-730	Operation group: Heart
275	operOPCSL1_90t180d_prevalence_1_cnt	1.24E-05	90-180	Operation group: Miscellaneous Operations
276	diagCCS_180t365d_prevalence_13_cnt	1.19E-05	180-365	CCS group: Asthma
277	operOPCSL1_365t730d_prevalence_9_cnt	1.17E-05	365-730	Operation group: Soft Tissue
278	diagCCS_180t365d_prevalence_15_cnt	1.13E-05	180-365	CCS group: Genitourinary symptoms & ill-defined conditions
279	diagCCS_90t180d_prevalence_14_cnt	1.11E-05	90-180	CCS group: Cancer; other & unspecified primary
280	diagCCS_365t730d_prevalence_25_cnt	1.10E-05	365-730	CCS group: Deficiency & other anemia
281	diagCCS_180t365d_prevalence_30_cnt	1.04E-05	180-365	CCS group: Esophageal disorders
282	operOPCSL1_180t365d_prevalence_9_cnt	9.90E-06	180-365	Operation group: Female Genital Tract
283	mainspef_30t90d_prevalence_2_cnt	9.71E-06	30-90	Main Speciality: General Surgery
284	mainspef_90t180d_prevalence_7_cnt	9.69E-06	90-180	Main Speciality: Ophthalmology
285	diagCCS_180t365d_prevalence_25_cnt	9.64E-06	180-365	CCS group: Complication of device; implant or graft
286	diagCCS_90t180d_prevalence_15_cnt	9.26E-06	90-180	CCS group: Asthma
287	diagCCS_30t90d_prevalence_10_cnt	8.89E-06	30-90	CCS group: Abdominal pain
288	diagCCS_90t180d_prevalence_9_cnt	8.26E-06	90-180	CCS group: Chronic obstructive pulmonary disease & bronchiectasis
289	diagCCS_90t180d_prevalence_22_cnt	7.94E-06	90-180	CCS group: Spondylosis; intervertebral disc disorders; other back problems
290	mainspef_180t365d_prevalence_8_cnt	7.27E-06	180-365	Main Speciality: Cardiothoracic
291	diagCCS_90t180d_prevalence_6_cnt	6.77E-06	90-180	CCS group: Cardiac dysrhythmias
292	diagCCS_30t90d_prevalence_4_cnt	6.73E-06	30-90	CCS group: Essential hypertension
293	diagCCS_180t365d_prevalence_20_cnt	6.69E-06	180-365	CCS group: Deficiency & other anemia
294	operOPCSL1_0t30d_prevalence_18_cnt	6.32E-06	0-30	Operation group: Skin
295	diagCCS_30t90d_prevalence_8_cnt	5.64E-06	30-90	CCS group: Deficiency & other anemia
296	operOPCSL1_30t90d_prevalence_2_cnt	5.58E-06	30-90	Operation group: Urinary
297	operOPCSL1_365t730d_prevalence_14_cnt	5.55E-06	365-730	Operation group: Lower Female Genital Tract
298	operOPCSL1_365t730d_prevalence_11_cnt	5.32E-06	365-730	Operation group: Respiratory Tract
299	diagCCS_90t180d_prevalence_16_cnt	4.98E-06	90-180	CCS group: Cataract
300	diagCCS_90t180d_prevalence_30_cnt	4.90E-06	90-180	CCS group: Osteoarthritis
301	diagCCS_30t90d_prevalence_14_cnt	4.74E-06	30-90	CCS group: Other lower respiratory disease
302	operOPCSL1_180t365d_prevalence_7_cnt	4.59E-06	180-365	Operation group: Nervous System
303	diagCCS_30t90d_prevalence_6_cnt	4.56E-06	30-90	CCS group: Diabetes mellitus without complication
304	diagCCS_90t180d_prevalence_26_cnt	4.19E-06	90-180	CCS group: Other nervous system disorders
305	mainspef_30t90d_prevalence_7_cnt	4.09E-06	30-90	Main Speciality: Geriatric
306	operOPCSL1_90t180d_prevalence_4_cnt	3.52E-06	90-180	Operation group: Upper Digestive Tract
307	diagCCS_30t90d_prevalence_9_cnt	2.70E-06	30-90	CCS group: Cardiac dysrhythmias
308	diagCCS_30t90d_prevalence_25_cnt	2.27E-06	30-90	CCS group: Chronic kidney disease
309	operOPCSL1_180t365d_prevalence_6_cnt	2.14E-06	180-365	Operation group: Heart
310	diagCCS_90t180d_prevalence_23_cnt	1.41E-06	90-180	CCS group: Congestive heart failure; nonhypertensive
311	mainspef_90t180d_prevalence_10_cnt	1.36E-06	90-180	Main Speciality: Haemotology
312	mainspef_30t90d_prevalence_9_cnt	1.97E-07	30-90	Main Speciality: Cardiothoracic

TABLE A.62: T-CARER: The weights of features using the Random Forest method (*Sample-2*)

#	Name	Weight	Temporal	Definition
1	mainspef_0t30d_prevalence_2_cnt	9.33E-02	0-30	Main Speciality: Maternity
2	diagCCS_0t30d_prevalence_10_cnt	6.60E-02	0-30	CCS group: Other complications of pregnancy
3	epidur_0t30d_avg	5.93E-02	0-30	Average episode duration
4	posopdur_0t30d_avg	5.79E-02	0-30	Average Post-operation duration
5	gender_1	5.09E-02	Trigger	Gender: Male
6	diagCCS_0t30d_prevalence_3_cnt	4.77E-02	0-30	CCS group: Normal pregnancy and delivery
7	diagCCS_0t30d_prevalence_13_cnt	4.23E-02	0-30	CCS group: OB-related trauma to perineum & vulva
8	diagCCS_0t30d_others_cnt	3.50E-02	0-30	CCS group: others
9	ethnos_0	3.11E-02	Trigger	ethnicity: Not known
10	mainspef_0t30d_prevalence_5_cnt	2.80E-02	0-30	Main Speciality: Gynaecology
11	operOPCSL1_0t30d_prevalence_1_cnt	2.05E-02	0-30	Operation group: Female Genital Tract
12	posopdur_365t730d_others_cnt	2.04E-02	365-730	Post-operation duration
13	diagCCS_0t30d_prevalence_22_cnt	1.93E-02	0-30	CCS group: Fetal distress & abnormal forces of labor
14	mainspef_0t30d_prevalence_6_cnt	1.90E-02	0-30	Main Speciality: Plastic
15	preopdur_180t365d_others_cnt	1.80E-02	180-365	Pre-operation duration
16	epidur_365t730d_avg	1.80E-02	365-730	Average episode duration
17	gapDays_365t730d_avg	1.74E-02	365-730	Average Gap-Days
18	preopdur_365t730d_others_cnt	1.71E-02	365-730	Pre-operation duration
19	mainspef_0t30d_prevalence_1_cnt	1.71E-02	0-30	Main Speciality: General
20	gapDays_365t730d_others_cnt	1.54E-02	365-730	Gap-Days
21	preopdur_0t30d_avg	1.47E-02	0-30	Average Pre-operation duration
22	admimeth_180t365d_prevalence_1_cnt	1.33E-02	180-365	Admission method: Unknown
23	epidur_365t730d_others_cnt	1.30E-02	365-730	Episode duration
24	epidur_0t30d_others_cnt	1.10E-02	0-30	Episode duration
25	mainspef_0t30d_prevalence_9_cnt	1.06E-02	0-30	Main Speciality: Paediatrics
26	mainspef_0t30d_prevalence_4_cnt	1.06E-02	0-30	Main Speciality: A&E
27	diagCCS_365t730d_others_cnt	9.36E-03	365-730	CCS group: others
28	admimeth_365t730d_prevalence_1_cnt	7.48E-03	365-730	Admission method: Unknown
29	mainspef_0t30d_prevalence_3_cnt	7.40E-03	0-30	Main Speciality: General Surgery
30	diagCCS_90t180d_others_cnt	7.32E-03	90-180	CCS group: others
31	diagCCS_180t365d_others_cnt	7.27E-03	180-365	CCS group: others
32	diagCCS_0t30d_prevalence_7_cnt	7.18E-03	0-30	CCS group: Abdominal pain
33	mainspef_180t365d_others_cnt	6.95E-03	180-365	Main Speciality: Others
34	diagCCS_0t30d_prevalence_24_cnt	6.77E-03	0-30	CCS group: Other birth complications; mother's puerperium
35	epidur_90t180d_others_cnt	6.22E-03	90-180	Episode duration
36	ethnos_1	5.97E-03	Trigger	ethnicity: White
37	epidur_180t365d_avg	5.79E-03	180-365	Average episode duration
38	mainspef_0t30d_prevalence_7_cnt	5.76E-03	0-30	Main Speciality: Geriatric
39	posopdur_180t365d_others_cnt	5.35E-03	180-365	Post-operation duration
40	diagCCS_0t30d_prevalence_16_cnt	5.25E-03	0-30	CCS group: Chronic obstructive pulmonary & bronchiectasis
41	admimeth_90t180d_prevalence_1_cnt	5.13E-03	90-180	Admission method: Unknown
42	epidur_180t365d_others_cnt	4.97E-03	180-365	Episode duration
43	diagCCS_0t30d_prevalence_11_cnt	4.85E-03	0-30	CCS group: Other upper respiratory disease
44	operOPCSL1_0t30d_others_cnt	4.50E-03	0-30	Operation group: Others
45	gapDays_180t365d_avg	4.34E-03	180-365	Average Gap-Days
46	diagCCS_0t30d_prevalence_2_cnt	3.80E-03	0-30	CCS group: Essential hypertension
47	posopdur_0t30d_others_cnt	3.74E-03	0-30	Post-operation duration
48	diagCCS_0t30d_prevalence_21_cnt	3.68E-03	0-30	CCS group: Delirium dementia & amnestic & other cognitives
49	mainspef_180t365d_prevalence_1_cnt	3.49E-03	180-365	Main Speciality: General
50	diagCCS_0t30d_prevalence_1_cnt	3.37E-03	0-30	CCS group: Others
51	diagCCS_0t30d_prevalence_5_cnt	3.14E-03	0-30	CCS group: Coronary atherosclerosis & other heart disease
52	mainspef_90t180d_others_cnt	3.09E-03	90-180	Main Speciality: Others
53	mainspef_0t30d_others_cnt	2.88E-03	0-30	Main Speciality: Others
54	diagCCS_30t90d_others_cnt	2.86E-03	30-90	CCS group: others
55	gapDays_90t180d_avg	2.73E-03	90-180	Average Gap-Days
56	gapDays_90t180d_others_cnt	2.69E-03	90-180	Gap-Days
57	admimeth_90t180d_others_cnt	2.47E-03	90-180	Admission method: Others
58	admimeth_30t90d_prevalence_1_cnt	2.46E-03	30-90	Admission method: Unknown
59	gapDays_30t90d_avg	2.23E-03	30-90	Average Gap-Days
60	posopdur_30t90d_others_cnt	2.09E-03	30-90	Post-operation duration
61	diagCCS_0t30d_prevalence_29_cnt	2.03E-03	0-30	CCS group: Other injuries & conditions due to external causes
62	gapDays_180t365d_others_cnt	1.99E-03	180-365	Gap-Days

Continued on next page

#	Name	Weight	Temporal	Definition
63	preopdur_0t30d_others_cnt	1.88E-03	0-30	Pre-operation duration
64	diagCCS_0t30d_prevalence_6_cnt	1.83E-03	0-30	CCS group: Cardiac dysrhythmias
65	admimeth_180t365d_others_cnt	1.82E-03	180-365	Admission method: Others
66	posopdur_90t180d_others_cnt	1.75E-03	90-180	Post-operation duration
67	epidur_30t90d_others_cnt	1.67E-03	30-90	Episode duration
68	diagCCS_365t730d_prevalence_4_cnt	1.62E-03	365-730	CCS group: Normal pregnancy and/or delivery
69	admimeth_365t730d_others_cnt	1.60E-03	365-730	Admission method: Others
70	mainspef_30t90d_others_cnt	1.58E-03	30-90	Main Speciality: Others
71	diagCCS_0t30d_prevalence_4_cnt	1.54E-03	0-30	CCS group: Residual codes; unclassified
72	diagCCS_0t30d_prevalence_9_cnt	1.45E-03	0-30	CCS group: Diabetes mellitus without complication
73	epidur_90t180d_avg	1.44E-03	90-180	Average episode duration
74	mainspef_0t30d_prevalence_8_cnt	1.40E-03	0-30	Main Speciality: Cardiothoracic
75	admimeth_365t730d_prevalence_2_cnt	1.31E-03	365-730	Admission method: Elective
76	diagCCS_0t30d_prevalence_17_cnt	1.29E-03	0-30	CCS group: Urinary tract infections
77	diagCCS_365t730d_prevalence_12_cnt	1.19E-03	365-730	CCS group: Other complications of pregnancy
78	diagCCS_365t730d_prevalence_1_cnt	1.15E-03	365-730	CCS group: Others
79	diagCCS_0t30d_prevalence_12_cnt	9.98E-04	0-30	CCS group: Asthma
80	mainspef_180t365d_prevalence_3_cnt	9.82E-04	180-365	Main Speciality: Gynaecology
81	preopdur_30t90d_others_cnt	9.59E-04	30-90	Pre-operation duration
82	posopdur_365t730d_avg	9.39E-04	365-730	Average Post-operation duration
83	mainspef_90t180d_prevalence_1_cnt	9.33E-04	90-180	Main Speciality: General
84	diagCCS_0t30d_prevalence_8_cnt	8.77E-04	0-30	CCS group: External cause codes: Fall
85	mainspef_0t30d_prevalence_10_cnt	8.66E-04	0-30	Main Speciality: Gastroenterology
86	diagCCS_0t30d_prevalence_20_cnt	8.31E-04	0-30	CCS group: Other connective tissue disease
87	diagCCS_180t365d_prevalence_1_cnt	8.23E-04	180-365	CCS group: Others
88	imd04rk_2	8.15E-04	Trigger	imd04rk: 3248 to 6496
89	admimeth_180t365d_prevalence_2_cnt	7.73E-04	180-365	Admission method: Elective
90	diagCCS_0t30d_prevalence_14_cnt	7.62E-04	0-30	CCS group: Other lower respiratory disease
91	diagCCS_0t30d_prevalence_28_cnt	7.05E-04	0-30	CCS group: Deficiency & other anemia
92	diagCCS_0t30d_prevalence_26_cnt	6.92E-04	0-30	CCS group: Alcohol-related disorders
93	imd04rk_3	6.76E-04	Trigger	imd04rk: 6496 to 9745
94	diagCCS_0t30d_prevalence_19_cnt	6.69E-04	0-30	CCS group: Phlebitis; thrombophlebitis & thromboembolism
95	imd04rk_1	6.60E-04	Trigger	imd04rk: 0 to 3248
96	epidur_30t90d_avg	6.27E-04	30-90	Average episode duration
97	ageTrigger_25	6.14E-04	Trigger	Age: 25-30
98	diagCCS_0t30d_prevalence_23_cnt	6.11E-04	0-30	CCS group: Allergic reactions
99	operOPCSL1_0t30d_prevalence_3_cnt	5.94E-04	0-30	Operation group: Lower Digestive Tract
100	ageTrigger_85	5.85E-04	Trigger	Age: 85+
101	admimeth_30t90d_others_cnt	5.73E-04	30-90	Admission method: Others
102	diagCCS_365t730d_prevalence_7_cnt	5.61E-04	365-730	CCS group: Cardiac dysrhythmias
103	diagCCS_0t30d_prevalence_25_cnt	5.57E-04	0-30	CCS group: Complication of device; implant or graft
104	diagCCS_0t30d_prevalence_18_cnt	5.49E-04	0-30	CCS group: Other gastrointestinal disorders
105	diagCCS_90t180d_prevalence_9_cnt	5.45E-04	90-180	CCS group: Other complications of pregnancy
106	preopdur_90t180d_others_cnt	5.44E-04	90-180	Pre-operation duration
107	posopdur_180t365d_avg	5.17E-04	180-365	Average Post-operation duration
108	diagCCS_365t730d_prevalence_13_cnt	4.91E-04	365-730	CCS group: Chronic obstructive pulmonary & bronchiectasis
109	imd04rk_6	4.84E-04	Trigger	imd04rk: 16241 to 19489
110	gapDays_30t90d_others_cnt	4.40E-04	30-90	Gap-Days
111	imd04rk_5	4.30E-04	Trigger	imd04rk: 12993 to 16241
112	ageTrigger_35	4.22E-04	Trigger	Age: 35-40
113	preopdur_90t180d_avg	4.11E-04	90-180	Average Pre-operation duration
114	operOPCSL1_0t30d_prevalence_8_cnt	4.04E-04	0-30	Operation group: Upper Female Genital Tract
115	ageTrigger_15	4.03E-04	Trigger	Age: 15-20
116	imd04rk_4	4.02E-04	Trigger	imd04rk: 9745 to 12993
117	imd04rk_8	3.96E-04	Trigger	imd04rk: 22737 to 25986
118	diagCCS_180t365d_prevalence_2_cnt	3.93E-04	180-365	CCS group: Essential hypertension
119	imd04rk_7	3.88E-04	Trigger	imd04rk: 19489 to 22737
120	diagCCS_0t30d_prevalence_15_cnt	3.85E-04	0-30	CCS group: Disorders of lipid metabolism
121	mainspef_180t365d_prevalence_6_cnt	3.83E-04	180-365	Main Speciality: Plastic
122	ageTrigger_80	3.75E-04	Trigger	Age: 80-85
123	diagCCS_365t730d_prevalence_5_cnt	3.68E-04	365-730	CCS group: Coronary atherosclerosis & other heart disease
124	ageTrigger_60	3.55E-04	Trigger	Age: 60-65
125	diagCCS_180t365d_prevalence_9_cnt	3.44E-04	180-365	CCS group: Chronic obstructive pulmonary & bronchiectasis
126	imd04rk_9	3.42E-04	Trigger	imd04rk: 25986 to 29234
127	preopdur_365t730d_avg	3.27E-04	365-730	Average Pre-operation duration
128	admimeth_30t90d_prevalence_2_cnt	3.23E-04	30-90	Admission method: Elective

Continued on next page

#	Name	Weight	Temporal	Definition
129	diagCCS_365t730d_prevalence_3_cnt	3.15E-04	365-730	CCS group: Essential hypertension
130	ageTrigger_30	3.12E-04	Trigger	Age: 30-35
131	operOPCSL1_0t30d_prevalence_4_cnt	3.11E-04	0-30	Operation group: Diagnostics & Tests
132	diagCCS_0t30d_prevalence_30_cnt	3.05E-04	0-30	CCS group: Thyroid disorders
133	ethnos_3	3.02E-04	Trigger	ethnicity: Pakistani
134	operOPCSL1_0t30d_prevalence_2_cnt	2.99E-04	0-30	Operation group: Miscellaneous Operations
135	diagCCS_365t730d_prevalence_2_cnt	2.96E-04	365-730	CCS group: Residual codes; unclassified
136	mainspef_180t365d_prevalence_4_cnt	2.87E-04	180-365	Main Speciality: Geriatric
137	diagCCS_0t30d_prevalence_27_cnt	2.79E-04	0-30	CCS group: Other nervous system disorders
138	admimeth_90t180d_prevalence_2_cnt	2.73E-04	90-180	Admission method: Elective
139	diagCCS_365t730d_prevalence_28_cnt	2.71E-04	365-730	CCS group: Cataract
140	ageTrigger_45	2.57E-04	Trigger	Age: 45-50
141	ethnos_9	2.39E-04	Trigger	ethnicity: Any other
142	diagCCS_365t730d_prevalence_6_cnt	2.36E-04	365-730	CCS group: Diabetes mellitus without complication
143	operOPCSL1_365t730d_others_cnt	2.36E-04	365-730	Operation group: Others
144	mainspef_90t180d_prevalence_5_cnt	1.95E-04	90-180	Main Speciality: Maternity
145	diagCCS_180t365d_prevalence_20_cnt	1.95E-04	180-365	CCS group: Other complications of pregnancy
146	diagCCS_365t730d_prevalence_29_cnt	1.94E-04	365-730	CCS group: Other birth complications; mother's puerperium
147	mainspef_90t180d_prevalence_3_cnt	1.78E-04	90-180	Main Speciality: Gynaecology
148	diagCCS_365t730d_prevalence_9_cnt	1.74E-04	365-730	CCS group: Abdominal pain
149	mainspef_30t90d_prevalence_2_cnt	1.69E-04	30-90	Main Speciality: Maternity
150	mainspef_30t90d_prevalence_1_cnt	1.68E-04	30-90	Main Speciality: General
151	ageTrigger_50	1.68E-04	Trigger	Age: 50-55
152	operOPCSL1_365t730d_prevalence_4_cnt	1.67E-04	365-730	Operation group: Upper Female Genital Tract
153	ageTrigger_20	1.65E-04	Trigger	Age: 20-25
154	mainspef_180t365d_prevalence_5_cnt	1.64E-04	180-365	Main Speciality: A&E
155	gapDays_0t30d_avg	1.64E-04	0-30	Average Gap-Days
156	posopdur_90t180d_avg	1.64E-04	90-180	Average Post-operation duration
157	diagCCS_90t180d_prevalence_2_cnt	1.56E-04	90-180	CCS group: Essential hypertension
158	diagCCS_365t730d_prevalence_20_cnt	1.54E-04	365-730	CCS group: Administrative/social admission
159	operOPCSL1_0t30d_prevalence_16_cnt	1.51E-04	0-30	Operation group: Lower Female Genital Tract
160	diagCCS_30t90d_prevalence_5_cnt	1.50E-04	30-90	CCS group: Other complications of pregnancy
161	diagCCS_90t180d_prevalence_1_cnt	1.45E-04	90-180	CCS group: Others
162	diagCCS_180t365d_prevalence_28_cnt	1.38E-04	180-365	CCS group: Congestive heart failure; nonhypertensive
163	preopdur_180t365d_avg	1.37E-04	180-365	Average Pre-operation duration
164	diagCCS_365t730d_prevalence_14_cnt	1.34E-04	365-730	CCS group: Other lower respiratory disease
165	operOPCSL1_180t365d_others_cnt	1.34E-04	180-365	Operation group: Others
166	mainspef_90t180d_prevalence_6_cnt	1.28E-04	90-180	Main Speciality: A&E
167	diagCCS_365t730d_prevalence_30_cnt	1.24E-04	365-730	CCS group: Fetal distress & abnormal forces of labor
168	diagCCS_365t730d_prevalence_16_cnt	1.20E-04	365-730	CCS group: Complication of device; implant or graft
169	diagCCS_90t180d_prevalence_17_cnt	1.19E-04	90-180	CCS group: Normal pregnancy and/or delivery
170	ethnos_7	1.18E-04	Trigger	ethnicity: Black - Other
171	diagCCS_180t365d_prevalence_8_cnt	1.14E-04	180-365	CCS group: Abdominal pain
172	diagCCS_365t730d_prevalence_11_cnt	1.13E-04	365-730	CCS group: Other gastrointestinal disorders
173	operOPCSL1_0t30d_prevalence_5_cnt	1.08E-04	0-30	Operation group: Urinary
174	diagCCS_30t90d_prevalence_12_cnt	1.08E-04	30-90	CCS group: Normal pregnancy and/or delivery
175	ageTrigger_55	1.07E-04	Trigger	Age: 55-60
176	mainspef_180t365d_prevalence_9_cnt	1.00E-04	180-365	Main Speciality: Maternity
177	diagCCS_90t180d_prevalence_7_cnt	9.95E-05	90-180	CCS group: Chronic obstructive pulmonary & bronchiectasis
178	ageTrigger_75	9.56E-05	Trigger	Age: 75-80
179	diagCCS_180t365d_prevalence_4_cnt	9.53E-05	180-365	CCS group: Coronary atherosclerosis & other heart disease
180	diagCCS_365t730d_prevalence_10_cnt	9.40E-05	365-730	CCS group: Other upper respiratory disease
181	diagCCS_365t730d_prevalence_8_cnt	9.34E-05	365-730	CCS group: Asthma
182	ageTrigger_40	9.26E-05	Trigger	Age: 40-45
183	diagCCS_90t180d_prevalence_3_cnt	9.12E-05	90-180	CCS group: Residual codes; unclassified
184	diagCCS_90t180d_prevalence_15_cnt	8.65E-05	90-180	CCS group: Other upper respiratory disease
185	diagCCS_365t730d_prevalence_22_cnt	8.55E-05	365-730	CCS group: OB-related trauma to perineum & vulva
186	diagCCS_365t730d_prevalence_18_cnt	8.52E-05	365-730	CCS group: External cause codes: Fall
187	diagCCS_180t365d_prevalence_19_cnt	8.33E-05	180-365	CCS group: Normal pregnancy and/or delivery
188	ethnos_6	8.31E-05	Trigger	ethnicity: Black - African
189	diagCCS_180t365d_prevalence_5_cnt	7.85E-05	180-365	CCS group: Diabetes mellitus without complication
190	diagCCS_365t730d_prevalence_26_cnt	7.83E-05	365-730	CCS group: Genitourinary symptoms & ill-defined conditions

Continued on next page

#	Name	Weight	Temporal	Definition
191	diagCCS_365t730d_prevalence_15_cnt	7.72E-05	365-730	CCS group: Phlebitis; thrombophlebitis & thromboembolism
192	diagCCS_30t90d_prevalence_1_cnt	7.68E-05	30-90	CCS group: Others
193	mainspef_180t365d_prevalence_2_cnt	7.60E-05	180-365	Main Speciality: General Surgery
194	ethnos_2	7.27E-05	Trigger	ethnicity: Indian
195	ethnos_5	6.89E-05	Trigger	ethnicity: Black - Caribbean
196	diagCCS_365t730d_prevalence_21_cnt	6.70E-05	365-730	CCS group: Disorders of lipid metabolism
197	mainspef_90t180d_prevalence_7_cnt	6.49E-05	90-180	Main Speciality: Plastic
198	operOPCSL1_180t365d_prevalence_1_cnt	6.42E-05	180-365	Operation group: Miscellaneous Operations
199	operOPCSL1_365t730d_prevalence_3_cnt	6.26E-05	365-730	Operation group: Female Genital Tract
200	diagCCS_180t365d_prevalence_7_cnt	6.23E-05	180-365	CCS group: Asthma
201	diagCCS_180t365d_prevalence_3_cnt	6.11E-05	180-365	CCS group: Residual codes; unclassified
202	diagCCS_365t730d_prevalence_24_cnt	6.08E-05	365-730	CCS group: Other screening (excl. mental & infectious)
203	diagCCS_90t180d_prevalence_24_cnt	6.04E-05	90-180	CCS group: Alcohol-related disorders
204	diagCCS_180t365d_prevalence_6_cnt	6.03E-05	180-365	CCS group: Cardiac dysrhythmias
205	ageTrigger_70	5.71E-05	Trigger	Age: 70-75
206	diagCCS_90t180d_prevalence_8_cnt	5.54E-05	90-180	CCS group: Abdominal pain
207	diagCCS_365t730d_prevalence_23_cnt	5.13E-05	365-730	CCS group: Allergic reactions
208	diagCCS_365t730d_prevalence_17_cnt	5.09E-05	365-730	CCS group: Other connective tissue disease
209	diagCCS_180t365d_prevalence_27_cnt	5.05E-05	180-365	CCS group: Delirium dementia & amnestic & other cognitives
210	operOPCSL1_0t30d_prevalence_7_cnt	4.98E-05	0-30	Operation group: Soft Tissue
211	diagCCS_180t365d_prevalence_16_cnt	4.98E-05	180-365	CCS group: Disorders of lipid metabolism
212	ageTrigger_65	4.85E-05	Trigger	Age: 65-70
213	ethnos_4	4.84E-05	Trigger	ethnicity: Bangladeshi
214	diagCCS_365t730d_prevalence_19_cnt	4.78E-05	365-730	CCS group: Urinary tract infections
215	mainspef_30t90d_prevalence_8_cnt	4.69E-05	30-90	Main Speciality: Plastic
216	operOPCSL1_365t730d_prevalence_2_cnt	4.59E-05	365-730	Operation group: Urinary
217	diagCCS_180t365d_prevalence_17_cnt	4.57E-05	180-365	CCS group: External cause codes: Fall
218	diagCCS_365t730d_prevalence_25_cnt	4.10E-05	365-730	CCS group: Other nervous system disorders
219	diagCCS_180t365d_prevalence_25_cnt	4.03E-05	180-365	CCS group: Administrative/social admission
220	diagCCS_365t730d_prevalence_27_cnt	3.87E-05	365-730	CCS group: Deficiency & other anemia
221	diagCCS_180t365d_prevalence_26_cnt	3.70E-05	180-365	CCS group: Genitourinary symptoms & ill-defined conditions
222	mainspef_90t180d_prevalence_2_cnt	3.64E-05	90-180	Main Speciality: General Surgery
223	diagCCS_180t365d_prevalence_12_cnt	3.63E-05	180-365	CCS group: Phlebitis; thrombophlebitis & thromboembolism
224	diagCCS_90t180d_prevalence_5_cnt	3.61E-05	90-180	CCS group: Diabetes mellitus without complication
225	gapDays_0t30d_others_cnt	3.42E-05	0-30	Gap-Days
226	diagCCS_180t365d_prevalence_13_cnt	3.30E-05	180-365	CCS group: Complication of device; implant or graft
227	diagCCS_180t365d_prevalence_30_cnt	3.30E-05	180-365	CCS group: Mood disorders
228	diagCCS_180t365d_prevalence_24_cnt	3.23E-05	180-365	CCS group: Alcohol-related disorders
229	operOPCSL1_365t730d_prevalence_1_cnt	2.98E-05	365-730	Operation group: Miscellaneous Operations
230	preopdur_30t90d_avg	2.90E-05	30-90	Average Pre-operation duration
231	diagCCS_30t90d_prevalence_3_cnt	2.76E-05	30-90	CCS group: Residual codes; unclassified
232	diagCCS_90t180d_prevalence_20_cnt	2.71E-05	90-180	CCS group: Congestive heart failure; nonhypertensive
233	diagCCS_180t365d_prevalence_11_cnt	2.68E-05	180-365	CCS group: Other upper respiratory disease
234	operOPCSL1_0t30d_prevalence_12_cnt	2.66E-05	0-30	Operation group: Mental Health
235	diagCCS_90t180d_prevalence_25_cnt	2.64E-05	90-180	CCS group: Other screening (excl. mental & infectious)
236	diagCCS_30t90d_prevalence_8_cnt	2.53E-05	30-90	CCS group: Abdominal pain
237	diagCCS_180t365d_prevalence_10_cnt	2.51E-05	180-365	CCS group: Other gastrointestinal disorders
238	operOPCSL1_365t730d_prevalence_6_cnt	2.25E-05	365-730	Operation group: Heart
239	diagCCS_180t365d_prevalence_22_cnt	2.15E-05	180-365	CCS group: Deficiency & other anemia
240	diagCCS_180t365d_prevalence_14_cnt	2.14E-05	180-365	CCS group: Other lower respiratory disease
241	admimeth_0t30d_prevalence_2_cnt	2.11E-05	0-30	Admission method: Elective
242	diagCCS_180t365d_prevalence_29_cnt	2.11E-05	180-365	CCS group: Epilepsy; convulsions
243	mainspef_180t365d_prevalence_8_cnt	2.09E-05	180-365	Main Speciality: Gastroenterology
244	diagCCS_90t180d_prevalence_18_cnt	1.99E-05	90-180	CCS group: Disorders of lipid metabolism
245	operOPCSL1_365t730d_prevalence_5_cnt	1.97E-05	365-730	Operation group: Lower Digestive Tract
246	mainspef_180t365d_prevalence_10_cnt	1.86E-05	180-365	Main Speciality: Urology
247	operOPCSL1_365t730d_prevalence_10_cnt	1.83E-05	365-730	Operation group: Diagnostics & Tests
248	operOPCSL1_365t730d_prevalence_7_cnt	1.72E-05	365-730	Operation group: Upper Digestive Tract
249	operOPCSL1_30t90d_others_cnt	1.60E-05	30-90	Operation group: Others
250	posopdur_30t90d_avg	1.38E-05	30-90	Average Post-operation duration
251	mainspef_90t180d_prevalence_10_cnt	1.36E-05	90-180	Main Speciality: Urology
252	diagCCS_180t365d_prevalence_23_cnt	1.18E-05	180-365	CCS group: Other nervous system disorders
253	operOPCSL1_180t365d_prevalence_2_cnt	1.13E-05	180-365	Operation group: Urinary

Continued on next page

#	Name	Weight	Temporal	Definition
254	operOPCSL1_90t180d_others_cnt	1.06E-05	90-180	Operation group: Others
255	mainspef_90t180d_prevalence_4_cnt	9.57E-06	90-180	Main Speciality: Geriatric
256	diagCCS_180t365d_prevalence_15_cnt	9.27E-06	180-365	CCS group: Urinary tract infections
257	diagCCS_30t90d_prevalence_20_cnt	7.40E-06	30-90	CCS group: Other screening (excl. mental & infectious)
258	diagCCS_30t90d_prevalence_9_cnt	6.74E-06	30-90	CCS group: Chronic obstructive pulmonary & bronchiectasis
259	operOPCSL1_90t180d_prevalence_1_cnt	6.66E-06	90-180	Operation group: Miscellaneous Operations
260	diagCCS_180t365d_prevalence_18_cnt	5.75E-06	180-365	CCS group: Other connective tissue disease
261	diagCCS_90t180d_prevalence_4_cnt	5.75E-06	90-180	CCS group: Coronary atherosclerosis & other heart disease
262	diagCCS_90t180d_prevalence_22_cnt	5.58E-06	90-180	CCS group: Other connective tissue disease
263	diagCCS_90t180d_prevalence_11_cnt	5.30E-06	90-180	CCS group: Other gastrointestinal disorders
264	operOPCSL1_180t365d_prevalence_5_cnt	5.27E-06	180-365	Operation group: Upper Female Genital Tract
265	mainspef_180t365d_prevalence_7_cnt	4.97E-06	180-365	Main Speciality: Cardiothoracic
266	mainspef_30t90d_prevalence_3_cnt	4.60E-06	30-90	Main Speciality: General Surgery
267	diagCCS_90t180d_prevalence_12_cnt	4.38E-06	90-180	CCS group: Other lower respiratory disease
268	mainspef_30t90d_prevalence_4_cnt	4.26E-06	30-90	Main Speciality: Gynaecology
269	diagCCS_30t90d_prevalence_19_cnt	4.23E-06	30-90	CCS group: Congestive heart failure; nonhypertensive
270	diagCCS_30t90d_prevalence_14_cnt	4.16E-06	30-90	CCS group: Complication of device; implant or graft
271	mainspef_30t90d_prevalence_6_cnt	4.15E-06	30-90	Main Speciality: A&E
272	diagCCS_90t180d_prevalence_19_cnt	3.96E-06	90-180	CCS group: External cause codes: Fall
273	operOPCSL1_180t365d_prevalence_3_cnt	3.58E-06	180-365	Operation group: Lower Digestive Tract
274	operOPCSL1_180t365d_prevalence_12_cnt	3.58E-06	180-365	Operation group: Mental Health
275	diagCCS_30t90d_prevalence_10_cnt	3.48E-06	30-90	CCS group: Other gastrointestinal disorders
276	diagCCS_90t180d_prevalence_29_cnt	3.48E-06	90-180	CCS group: Other nervous system disorders
277	diagCCS_90t180d_prevalence_14_cnt	3.27E-06	90-180	CCS group: Complication of device; implant or graft
278	diagCCS_30t90d_prevalence_7_cnt	2.94E-06	30-90	CCS group: Cardiac dysrhythmias
279	operOPCSL1_365t730d_prevalence_8_cnt	2.88E-06	365-730	Operation group: Soft Tissue
280	diagCCS_90t180d_prevalence_13_cnt	2.58E-06	90-180	CCS group: Phlebitis; thrombophlebitis & thromboembolism
281	operOPCSL1_90t180d_prevalence_2_cnt	1.66E-06	90-180	Operation group: Urinary
282	diagCCS_180t365d_prevalence_21_cnt	1.65E-06	180-365	CCS group: Allergic reactions
283	operOPCSL1_180t365d_prevalence_8_cnt	1.24E-06	180-365	Operation group: Female Genital Tract
284	diagCCS_90t180d_prevalence_6_cnt	9.76E-07	90-180	CCS group: Cardiac dysrhythmias
285	diagCCS_90t180d_prevalence_30_cnt	9.27E-07	90-180	CCS group: Mood disorders
286	diagCCS_30t90d_prevalence_4_cnt	8.25E-07	30-90	CCS group: Coronary atherosclerosis & other heart disease
287	operOPCSL1_0t30d_prevalence_9_cnt	7.08E-07	0-30	Operation group: Upper Digestive Tract
288	diagCCS_30t90d_prevalence_6_cnt	1.63E-07	30-90	CCS group: Diabetes mellitus without complication

A.7.2 WDNN Model Specifications

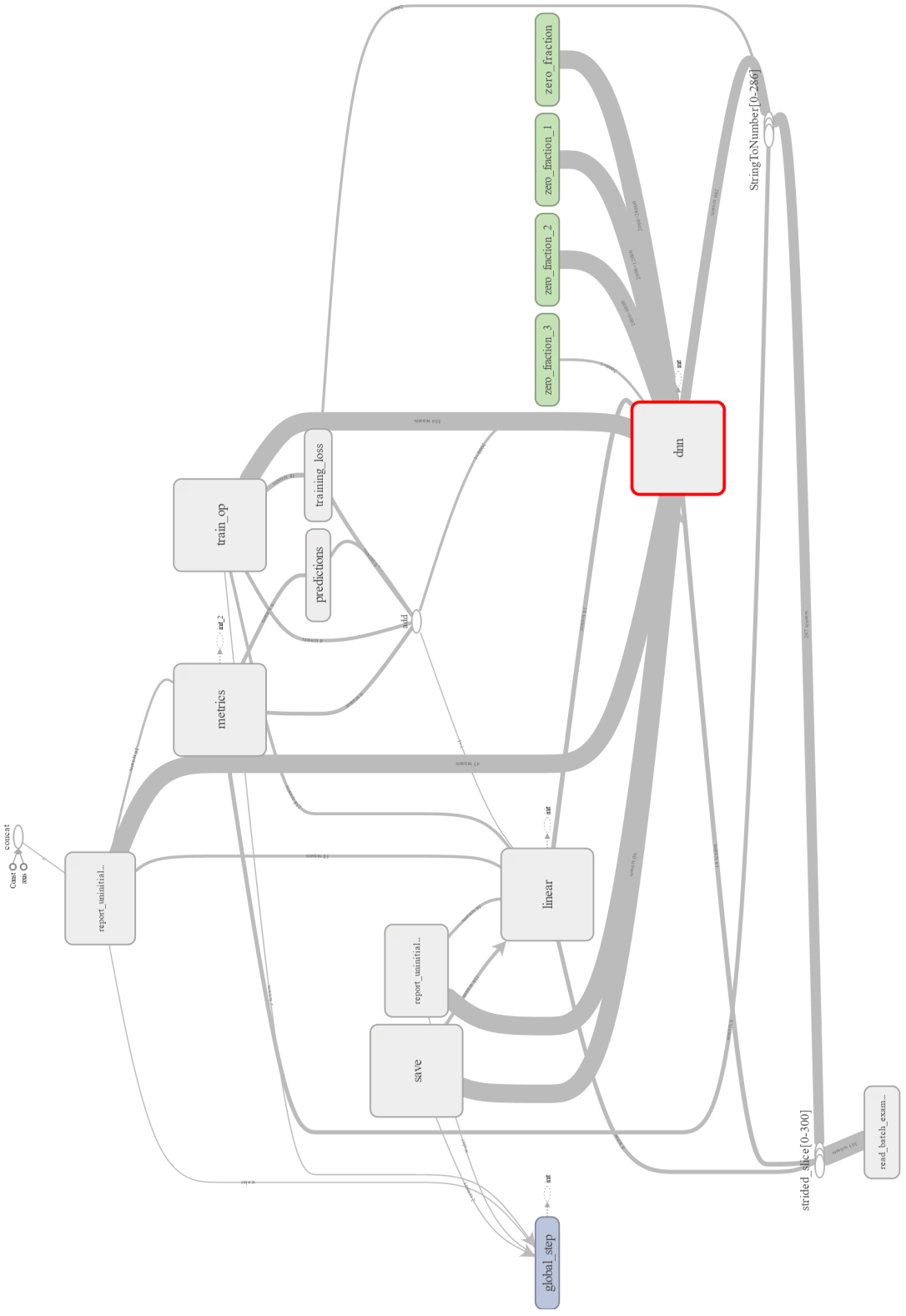

FIGURE A.57: T-CARER: The abstract graph of the designed WDNN

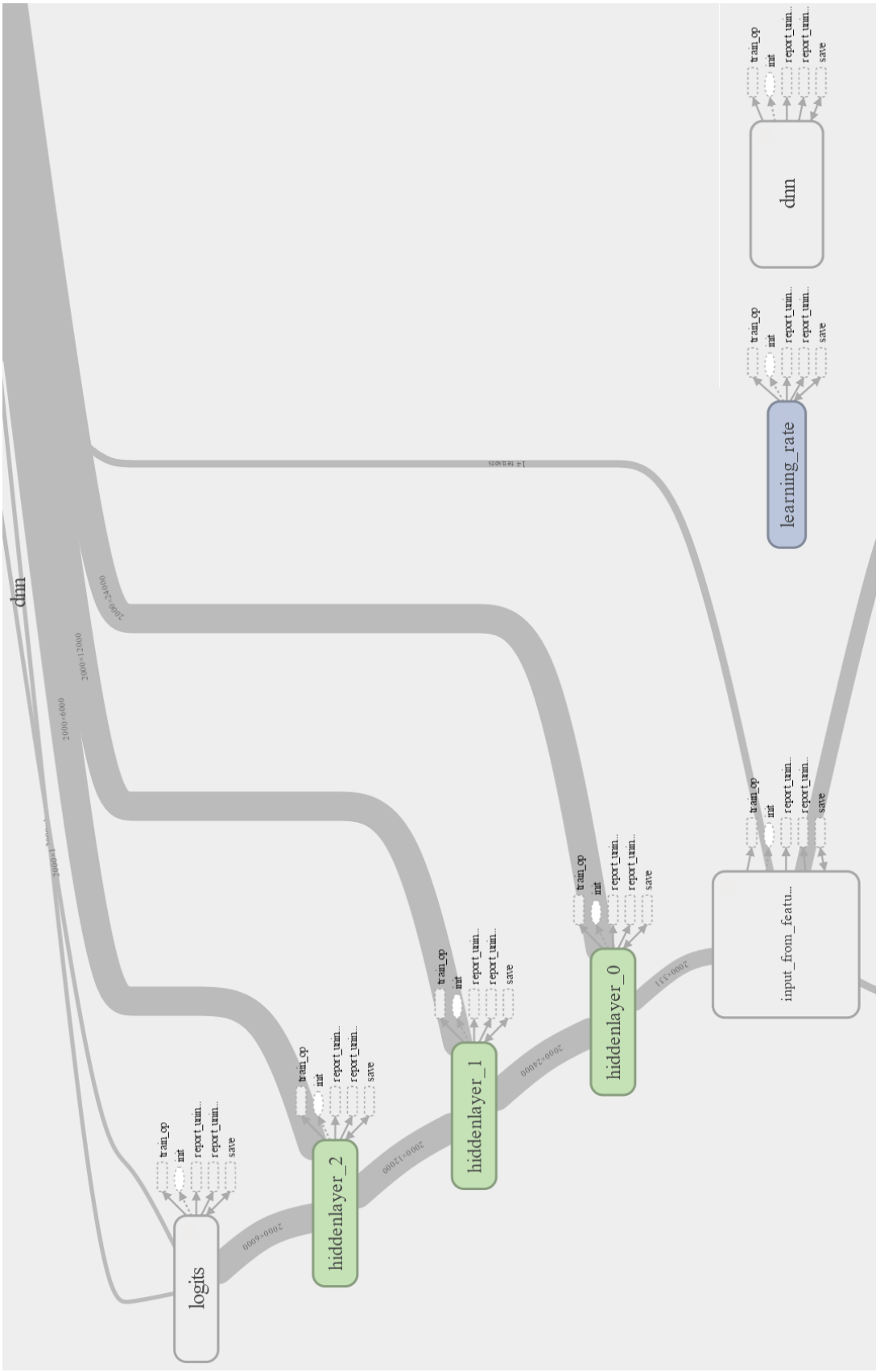

FIGURE A.58: T-CARER: The abstract graph of the deep part of the designed WDNN

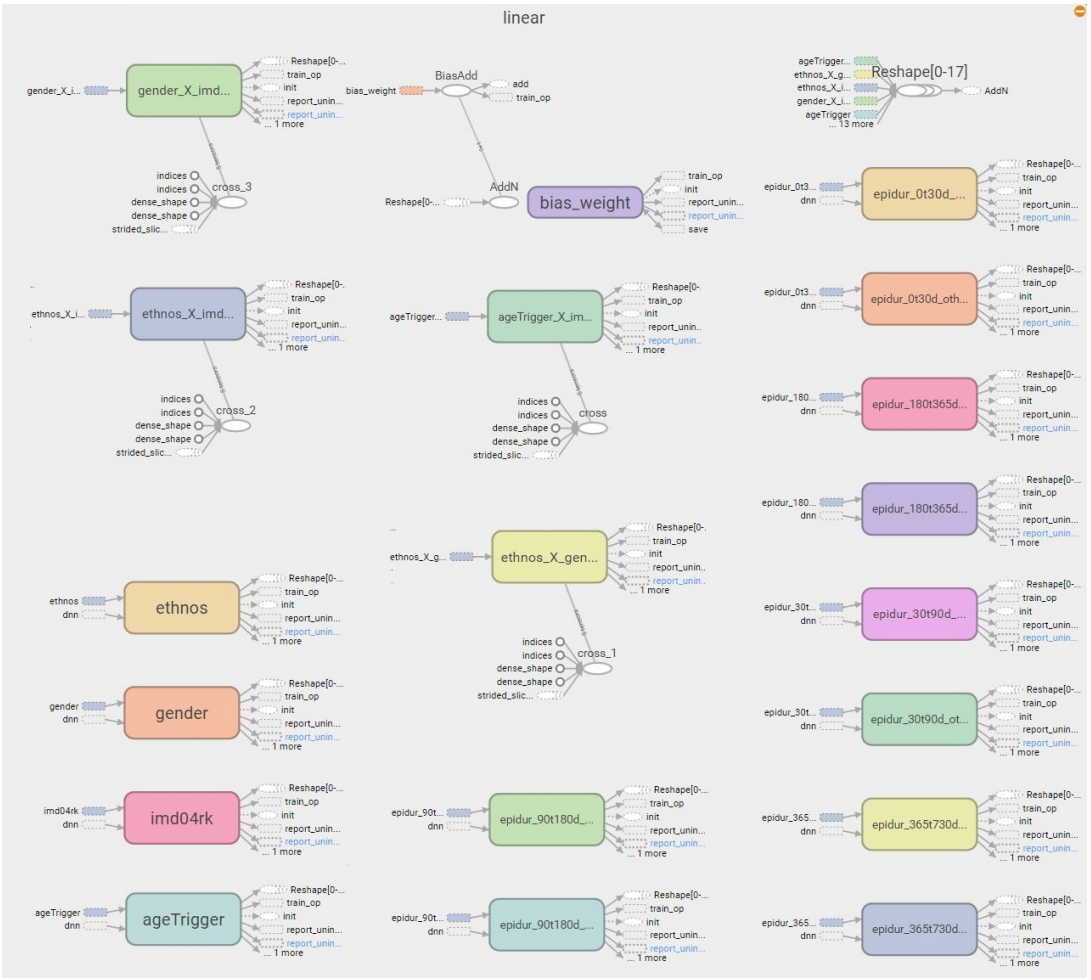

FIGURE A.59: T-CARER: The abstract graph of the linear part of the designed WDNN

A.7.3 Performance for CCI and ECI Categories

A.7.3.1 Random Forest - Sample-1

TABLE A.63: T-CARER: The profile of the model and the HSCIC-CCI for the Elixhauser diagnoses categories, using random-forest (*Sample-1*)

Elixhauser Comorbidity Index	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Congestive heart failure	4,652	2,241	14	12	11,393	0.875	0.576	4,340	962	2,007	4,919	404	199	8,619	3,468	872 (7.7%)	3,534 (31.0%)
AIDS/HIV	63	50	12	7	350	0.73	0.388	46	159	23	276	175	31	71	14	32 (9.1%)	190 (54.3%)
Depression	2,115	772	39	25	5,499	0.893	0.583	1,889	908	757	2,771	177	73	854	335	1,554 (28.3%)	10 (0.2%)
Cardiac arrhythmias	7,637	3,652	14	10	21,344	0.893	0.557	6,817	3,704	2,647	11,346	1,271	505	7,541	3,021	3,796 (17.8%)	2,925 (13.7%)
Valvular disease	1,670	772	16	11	4,536	0.832	0.574	1,390	1,087	529	2,416	225	89	1,540	643	747 (16.5%)	583 (12.9%)
Pulmonary circulation disorder	587	247	23	13	1,847	0.923	0.503	542	233	217	1,030	77	30	639	232	310 (16.8%)	224 (12.1%)
Peripheral vascular disorders	2,042	1,315	13	12	5,811	0.777	0.545	1,587	1,578	606	3,237	272	104	3,582	1,380	207 (3.6%)	1,838 (31.6%)
Hypertension, uncomplicated	11,958	5,358	16	9	42,818	0.773	0.507	9,244	15,605	3,015	28,182	4,929	1,515	12,123	4,396	4,848 (11.3%)	8,463 (19.8%)
Hypertension, complicated	571	349	20	9	1,308	0.862	0.628	492	233	253	587	61	25	875	410	82 (6.3%)	351 (26.8%)
Paralysis	991	529	24	17	2,881	0.92	0.542	912	427	379	1,525	685	206	1,482	565	347 (12.0%)	1,031 (35.8%)
Other neurological disorders	4,488	2,304	39	10	12,314	0.851	0.576	3,821	2,864	1,665	6,534	554	232	2,725	1,108	2,713 (22.0%)	647 (5.3%)
Chronic pulmonary disease	10,417	4,728	25	8	31,196	0.845	0.575	8,798	9,401	4,345	17,842	342	177	24,360	7,640	1,158 (3.7%)	13,948 (44.7%)
Diabetes, uncomplicated	7,764	3,942	19	9	22,309	0.841	0.565	6,531	5,725	2,770	12,398	13,973	4,401	6,651	2,755	3,776 (16.9%)	11,321 (50.7%)
Diabetes, complicated	986	606	21	13	2,659	0.793	0.612	782	886	345	1,470	386	163	677	319	463 (17.4%)	378 (14.2%)
Hypothyroidism	1,812	337	18	10	5,869	0.841	0.544	1,524	1,792	610	3,579	465	162	1,838	685	839 (14.3%)	978 (16.7%)
Renal failure	2,618	1,539	22	12	5,923	0.896	0.619	2,345	697	1,134	2,425	134	56	4,357	1,873	472 (8.0%)	1,682 (28.4%)
Liver disease	1,139	694	22	9	3,247	0.85	0.585	968	905	411	1,813	169	68	1,091	375	593 (18.3%)	522 (16.1%)
Peptic ulcer disease	651	341	20	9	2,704	0.696	0.482	453	1,279	125	1,938	47	16	1,369	371	82 (3.0%)	914 (33.8%)
Psychoses	1,089	619	29	42	2,864	0.906	0.571	987	394	435	1,435	72	22	191	98	889 (31.0%)	-197 (-6.9%)
Lymphoma	891	483	21	7	2,092	0.704	0.574	627	536	224	997	41	18	595	245	382 (18.3%)	169 (8.1%)
Metastatic cancer	2,454	1,123	18	10	6,069	0.747	0.541	1,834	1,129	509	2,976	49	16	5,205	2,093	-259 (-4.3%)	2,506 (41.3%)
Solid tumour	6,729	3,672	20	8	19,482	0.606	0.498	4,080	7,187	1,144	11,503	438	159	7,450	2,950	1,130 (5.8%)	3,529 (18.1%)
Rheumatoid arthritis	1,533	409	19	11	5,773	0.76	0.502	1,165	2,297	448	3,897	100	35	3,482	1,066	99 (1.7%)	2,138 (37.0%)
Coagulopathy	413	194	39	7	1,315	0.77	0.529	318	431	144	814	53	19	261	89	229 (17.4%)	118 (9.0%)
Obesity	447	168	28	9	1,687	0.743	0.503	332	698	102	1,134	212	70	400	146	186 (11.0%)	290 (17.2%)
Weight loss	729	351	20	12	2,378	0.645	0.492	470	939	114	1,538	64	30	435	173	297 (12.5%)	185 (7.8%)
Fluid & electrolyte disorders	1,319	485	19	15	3,722	0.945	0.542	1,246	372	521	1,884	188	73	1,416	501	745 (20.0%)	511 (13.7%)
Blood loss anemia	85	28	29	8	321	0.835	0.47	71	90	23	217	19	7	71	25	46 (14.3%)	39 (12.1%)
Deficiency anemia	3,187	1,850	35	10	9,134	0.853	0.575	2,718	2,405	1,131	5,195	333	127	1,707	715	2,003 (21.9%)	446 (4.9%)
Alcohol abuse	1,286	855	22	7	3,624	0.867	0.573	1,115	847	426	2,056	122	63	676	274	841 (23.2%)	179 (4.9%)
Drug abuse	673	470	11	10	2,021	0.796	0.544	536	585	218	1,217	20	11	123	53	483 (23.9%)	-52 (-2.6%)

[a] The Elixhauser Comorbidity Index (ECI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age. [d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1-score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point. [h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+.

TABLE A.64: T-CARER: The profile of the model and the HSCIC-CCI for the Charlson diagnoses categories, using random-forest (*Sample-1*)

Charlson Comorbidity Index	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Myocardial infarction	3,529	2,223	18	7	9,757	0.9	0.548	3,177	1,336	946	5,260	519	207	5,705	2,039	1,138 (11.7%)	3,010 (30.8%)
Peripheral vascular disease	2,042	1,315	13	12	5,811	0.836	0.559	1,587	1,578	606	3,237	272	104	3,582	1,380	207 (3.6%)	1,838 (31.6%)
Cerebrovascular disease	3,748	1,841	15	19	11,585	0.885	0.501	3,495	1,341	1,255	6,286	820	253	5,741	2,021	1,474 (12.7%)	2,736 (23.6%)
Dementia	2,332	841	10	24	6,340	0.893	0.538	2,110	643	768	3,037	75	35	4,496	1,610	500 (7.9%)	1,955 (30.8%)
Chronic pulmonary disease	10,417	4,728	25	8	31,196	0.975	0.603	8,798	9,401	4,345	17,842	342	177	24,360	7,640	1,158 (3.7%)	13,948 (44.7%)
Rheumatic disease	1,412	350	18	11	5,161	0.398	0.387	1,072	1,967	413	3,437	70	22	3,446	1,051	21 (0.4%)	2,131 (41.3%)
Peptic ulcer disease	910	471	20	12	3,801	0.665	0.446	674	1,558	178	2,707	86	28	1,774	484	190 (5.0%)	1,164 (30.6%)
Mild liver disease	983	600	21	9	2,733	0.782	0.581	846	768	371	1,511	148	57	892	335	511 (18.7%)	409 (15.0%)
Diabetes, uncomplicated	7,870	3,993	19	9	22,598	0.98	0.593	6,627	5,783	2,817	12,549	14,102	4,445	6,729	2,787	3,840 (17.0%)	11,420 (50.5%)
Diabetes, complicated	880	553	19	12	2,375	0.356	0.418	686	831	296	1,323	261	119	603	287	399 (16.8%)	286 (12.0%)
Hemiplegia or paraplegia	991	529	24	17	2,881	0.825	0.524	912	427	379	1,525	685	206	1,482	565	347 (12.0%)	1,031 (35.8%)
Renal disease	2,626	1,543	22	12	5,945	0.967	0.635	2,350	702	1,138	2,438	134	56	4,369	1,877	473 (8.0%)	1,689 (28.4%)
Malignancy	8,129	4,471	19	8	22,675	0.949	0.599	5,090	8,013	1,540	13,011	501	190	8,383	3,322	1,768 (7.8%)	3,837 (16.9%)
Moderate or severe liver disease	323	209	21	8	898	0.083	0.139	275	198	107	471	36	18	442	136	139 (15.5%)	220 (24.5%)
Metastatic solid tumour	2,454	1,123	18	10	6,069	0.974	0.591	1,834	1,129	509	2,976	49	16	5,205	2,093	-259 (-4.3%)	2,506 (41.3%)
Congestive heart failure	4,652	2,241	14	12	11,393	0.875	0.576	4,340	962	2,007	4,919	404	199	8,619	3,468	872 (7.7%)	3,534 (31.0%)
AIDS/HIV	63	50	12	7	350	0.73	0.388	46	159	23	276	175	31	71	14	32 (9.1%)	190 (54.3%)

[a] The Charlson Comorbidity Index (CCI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age.
[d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1-score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point.
[h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+.

A.7.3.2 Random Forest - Sample-2

TABLE A.65: T-CARER: The profile of the model and the HSCIC-CCI for the Elixhauser diagnoses categories, using random-forest (*Sample-2*)

Elixhauser Comorbidity Index	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Congestive heart failure	4,922	2,385	14	10	8,716	0.913	0.724	4,242	962	1,364	3,185	304	163	7,031	4,053	189 (2.2%)	2,510 (28.8%)
AIDS/HIV	102	67	13	6	442	0.676	0.543	69	257	20	329	219	49	89	23	46 (10.4%)	225 (50.9%)
Depression	3,029	1,114	38	8	7,196	0.647	0.585	1,961	2,454	608	3,917	327	142	1,462	733	1,228 (17.1%)	664 (9.2%)
Cardiac arrhythmias	11,194	5,377	14	8	23,006	0.812	0.66	9,089	4,546	2,565	10,603	1,550	764	10,033	5,520	3,569 (15.5%)	4,090 (17.8%)
Valvular disease	2,019	942	16	10	3,957	0.834	0.674	1,683	645	457	1,754	228	109	1,868	1,067	616 (15.6%)	736 (18.6%)
Pulmonary circulation disorder	806	377	22	10	1,745	0.797	0.649	642	408	152	868	101	41	729	397	245 (14.0%)	321 (18.4%)
Peripheral vascular disorders	1,980	1,236	14	11	3,780	0.821	0.674	1,625	583	498	1,612	108	63	3,074	1,650	-25 (-0.7%)	1,281 (33.9%)
Hypertension, uncomplicated	19,370	8,554	18	8	46,496	0.713	0.596	13,813	13,980	3,347	25,568	5,932	2,400	16,643	8,131	5,682 (12.2%)	10,486 (22.6%)
Hypertension, complicated	1,702	911	15	11	3,002	0.845	0.708	1,439	374	466	1,114	52	30	2,701	1,545	-106 (-3.5%)	992 (33.0%)
Paralysis	1,124	570	21	14	2,272	0.818	0.653	919	376	312	1,013	420	194	1,392	740	179 (7.9%)	743 (32.7%)
Other neurological disorders	5,672	2,863	35	8	11,863	0.751	0.662	4,262	3,244	1,438	5,695	604	310	3,513	1,944	2,318 (19.5%)	1,367 (11.5%)
Chronic pulmonary disease	13,510	5,865	30	6	30,516	0.754	0.656	10,184	9,672	4,018	15,589	491	269	24,090	10,669	-485 (-1.6%)	12,226 (40.1%)
Diabetes, uncomplicated	10,110	5,054	21	8	21,907	0.76	0.645	7,687	5,767	2,428	10,779	11,516	4,729	8,737	4,757	2,930 (13.4%)	9,749 (44.5%)
Diabetes, complicated	887	516	21	11	1,603	0.821	0.696	728	240	276	622	248	129	761	472	256 (16.0%)	314 (19.6%)
Hypothyroidism	3,394	685	19	9	7,859	0.767	0.632	2,602	2,233	760	4,142	695	312	2,940	1,539	1,063 (13.5%)	1,461 (18.6%)
Renal failure	3,956	2,173	16	10	6,883	0.858	0.715	3,393	780	1,222	2,492	106	61	6,161	3,556	-163 (-2.4%)	2,215 (32.2%)
Liver disease	1,527	939	22	7	2,955	0.788	0.689	1,204	663	432	1,280	189	96	1,157	668	536 (18.1%)	434 (14.7%)
Peptic ulcer disease	464	233	23	11	1,158	0.81	0.619	376	319	93	662	43	16	711	306	70 (6.0%)	400 (34.5%)
Psychoses	1,121	641	29	31	2,571	0.647	0.573	725	766	206	1,344	103	45	332	175	550 (21.4%)	109 (4.2%)
Lymphoma	462	246	18	10	743	0.764	0.709	353	100	103	222	15	9	478	289	64 (8.6%)	136 (18.3%)
Metastatic cancer	1,650	835	17	8	3,034	0.755	0.642	1,246	398	288	1,185	26	17	2,687	1,461	-215 (-7.1%)	1,036 (34.1%)
Solid tumour	3,510	2,062	18	9	6,425	0.763	0.663	2,677	1,029	684	2,542	131	74	4,750	2,621	56 (0.9%)	1,813 (28.2%)
Rheumatoid arthritis	1,603	458	19	8	3,833	0.783	0.626	1,255	1,076	342	2,045	64	28	3,176	1,369	-114 (-3.0%)	1,658 (43.3%)
Coagulopathy	416	175	40	8	1,003	0.75	0.661	312	371	120	553	42	19	265	140	172 (17.1%)	114 (11.4%)
Obesity	853	343	29	7	2,009	0.734	0.625	626	631	228	1,095	307	136	704	357	269 (13.4%)	457 (22.7%)
Weight loss	709	369	23	12	1,483	0.753	0.633	534	329	106	725	70	29	490	267	267 (18.0%)	215 (14.5%)
Fluid & electrolyte disorders	2,850	1,161	19	14	5,901	0.838	0.654	2,387	992	625	2,771	375	182	2,599	1,372	1,015 (17.2%)	1,140 (19.3%)
Blood loss anemia	69	30	28	8	204	0.638	0.494	44	70	12	127	8	2	51	26	18 (8.8%)	23 (11.3%)
Deficiency anemia	5,006	2,975	33	7	12,003	0.713	0.638	3,571	4,385	1,237	6,589	525	254	2,548	1,361	2,210 (18.4%)	1,050 (8.7%)
Alcohol abuse	2,132	1,313	23	5	5,421	0.646	0.609	1,377	2,275	461	3,148	175	81	997	547	830 (15.3%)	403 (7.4%)
Drug abuse	941	614	13	5	2,548	0.576	0.561	542	1,158	169	1,541	33	15	261	109	433 (17.0%)	104 (4.1%)

[a] The Elixhauser Comorbidity Index (ECI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age. [d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1-score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point. [h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+.

TABLE A.66: T-CARER: The profile of the model and the HSCIC-CCI for the Charlson diagnoses categories, using random-forest (*Sample-2*)

Charlson Comorbidity Index	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Myocardial infarction	4,079	2,600	21	5	7,806	0.692	0.63	2,821	1,667	602	3,509	396	221	4,493	2,373	448 (5.7%)	2,077 (26.6%)
Peripheral vascular disease	1,980	1,236	14	11	3,780	0.705	0.631	1,625	583	498	1,612	108	63	3,074	1,650	-25 (-0.7%)	1,281 (33.9%)
Cerebrovascular disease	4,651	2,206	16	14	9,911	0.911	0.671	3,640	2,044	830	4,855	598	277	5,457	2,795	845 (8.5%)	2,578 (26.0%)
Dementia	4,020	1,407	9	15	7,766	0.779	0.633	3,556	638	831	3,244	100	55	6,380	3,312	244 (3.1%)	2,611 (33.6%)
Chronic pulmonary disease	13,510	5,865	30	6	30,516	0.956	0.723	10,184	9,672	4,018	15,589	491	269	24,090	10,669	-485 (-1.6%)	12,226 (40.1%)
Rheumatic disease	1,462	394	16	9	3,438	0.261	0.349	1,174	931	318	1,801	34	15	3,110	1,342	-168 (-4.9%)	1,612 (46.9%)
Peptic ulcer disease	695	370	23	10	1,817	0.653	0.558	542	552	120	1,080	76	24	935	405	137 (7.5%)	540 (29.7%)
Mild liver disease	1,393	853	22	7	2,691	0.879	0.722	1,108	596	408	1,162	167	83	1,054	613	495 (18.4%)	389 (14.5%)
Diabetes, uncomplicated	10,162	5,087	21	8	22,027	0.964	0.709	7,730	5,796	2,449	10,842	11,560	4,747	8,780	4,780	2,950 (13.4%)	9,790 (44.4%)
Diabetes, complicated	848	496	22	11	1,501	0.223	0.327	697	213	262	564	212	115	726	456	241 (16.1%)	278 (18.5%)
Hemiplegia or paraplegia	1,124	570	21	14	2,272	0.859	0.666	919	376	312	1,013	420	194	1,392	740	179 (7.9%)	743 (32.7%)
Renal disease	3,962	2,176	16	10	6,898	0.943	0.743	3,399	784	1,222	2,500	107	62	6,168	3,558	-159 (-2.3%)	2,219 (32.2%)
Malignancy	4,216	2,435	19	9	7,589	0.85	0.702	3,199	1,217	840	2,931	155	89	5,469	3,032	167 (2.2%)	2,061 (27.2%)
Moderate or severe liver disease	359	244	20	7	607	0.227	0.334	298	75	114	211	39	23	330	196	102 (16.8%)	113 (18.6%)
Metastatic solid tumour	1,650	835	17	8	3,034	0.953	0.704	1,246	398	288	1,185	26	17	2,687	1,461	-215 (-7.1%)	1,036 (34.1%)
Congestive heart failure	4,922	2,385	14	10	8,716	0.913	0.724	4,242	962	1,364	3,185	304	163	7,031	4,053	189 (2.2%)	2,510 (28.8%)
AIDS/HIV	102	67	13	6	442	0.676	0.543	69	257	20	329	219	49	89	23	46 (10.4%)	225 (50.9%)

[a] The Charlson Comorbidity Index (CCI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age.
[d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1-score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point.
[h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+.

A.7.3.3 Wide and Deep Neural Network - Sample-1

TABLE A.67: T-CARER: The profile of the model and the HSCIC-CCI for the Elixhauser diagnoses categories, using WDNN (*Sample-1*)

Elixhauser Comorbidity Index	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Congestive heart failure	2,454	1,123	18	10	6,069	0.226	0.299	554	2,918	104	3,521	49	16	5,205	2,093	-1,539 (-25.4%)	3,051 (50.3%)
AIDS/HIV	63	50	12	7	350	0.317	0.412	20	273	6	286	175	31	71	14	6 (1.7%)	200 (57.1%)
Depression	2,115	772	39	25	5,499	0.353	0.422	746	2,710	177	3,271	177	73	854	335	411 (7.5%)	510 (9.3%)
Cardiac arrhythmias	7,637	3,652	14	10	21,344	0.39	0.444	2,976	10,901	603	13,392	1,271	505	7,541	3,021	-45 (-0.2%)	4,971 (23.3%)
Valvular disease	1,670	772	16	11	4,536	0.331	0.407	552	2,375	117	2,803	225	89	1,540	643	-91 (-2.0%)	970 (21.4%)
Pulmonary circulation disorder	587	247	23	13	1,847	0.388	0.425	228	1,001	52	1,226	77	30	639	232	-4 (-0.2%)	420 (22.7%)
Peripheral vascular disorders	2,042	1,315	13	12	5,811	0.308	0.381	629	3,135	128	3,692	272	104	3,582	1,380	-751 (-12.9%)	2,293 (39.5%)
Hypertension, uncomplicated	11,958	5,358	16	9	42,818	0.28	0.362	3,345	27,665	663	30,521	4,929	1,515	12,123	4,396	-1,051 (-2.5%)	10,802 (25.2%)
Hypertension, complicated	571	349	20	9	1,308	0.433	0.506	247	579	71	706	61	25	875	410	-163 (-12.5%)	470 (35.9%)
Paralysis	991	529	24	17	2,881	0.377	0.42	374	1,475	79	1,846	685	206	1,482	565	-191 (-6.6%)	1,352 (46.9%)
Other neurological disorders	4,488	2,304	39	10	12,314	0.361	0.435	1,621	6,479	433	7,634	554	232	2,725	1,108	513 (4.2%)	1,747 (14.2%)
Chronic pulmonary disease	10,417	4,728	25	8	31,196	0.478	0.519	4,984	16,982	1,799	20,031	342	177	24,360	7,640	-2,656 (-8.5%)	16,137 (51.7%)
Diabetes, uncomplicated	7,764	3,942	19	9	22,309	0.42	0.468	3,261	11,645	930	14,093	13,973	4,401	6,651	2,755	506 (2.3%)	13,016 (58.3%)
Diabetes, complicated	986	606	21	13	2,659	0.318	0.42	314	1,476	98	1,636	386	163	677	319	-5 (-0.2%)	544 (20.5%)
Hypothyroidism	1,812	337	18	10	5,869	0.358	0.43	648	3,500	153	3,987	465	162	1,838	685	-37 (-0.6%)	1,386 (23.6%)
Renal failure	2,618	1,539	22	12	5,923	0.423	0.48	1,108	2,415	291	3,142	134	56	4,357	1,873	-765 (-12.9%)	2,399 (40.5%)
Liver disease	1,139	694	22	9	3,247	0.309	0.394	352	1,810	89	2,069	169	68	1,091	375	-23 (-0.7%)	778 (24.0%)
Peptic ulcer disease	651	341	20	9	2,704	0.19	0.264	124	1,890	10	2,044	47	16	1,369	371	-247 (-9.1%)	1,020 (37.7%)
Psychoses	1,089	619	29	42	2,864	0.412	0.467	449	1,389	114	1,719	72	22	191	98	351 (12.3%)	87 (3.0%)
Lymphoma	891	483	21	7	2,092	0.26	0.343	232	970	60	1,167	41	18	595	245	-13 (-0.6%)	339 (16.2%)
Metastatic cancer	2,454	1,123	18	10	6,069	0.226	0.299	554	2,918	104	3,521	49	16	5,205	2,093	-1,539 (-25.4%)	3,051 (50.3%)
Solid tumour	6,729	3,672	20	8	19,482	0.192	0.27	1,289	11,236	240	12,557	438	159	7,450	2,950	-1,661 (-8.5%)	4,583 (23.5%)
Rheumatoid arthritis	1,533	409	19	11	5,773	0.3	0.386	460	3,850	98	4,209	100	35	3,482	1,066	-606 (-10.5%)	2,450 (42.4%)
Coagulopathy	413	194	39	7	1,315	0.356	0.432	147	781	43	883	53	19	261	89	58 (4.4%)	187 (14.2%)
Obesity	447	168	28	9	1,687	0.266	0.351	119	1,128	34	1,225	212	70	400	146	-27 (-1.6%)	381 (22.6%)
Weight loss	729	351	20	12	2,378	0.199	0.285	145	1,506	19	1,641	64	30	435	173	-28 (-1.2%)	288 (12.1%)
Fluid & electrolyte disorders	1,319	485	19	15	3,722	0.411	0.443	542	1,818	111	2,335	188	73	1,416	501	41 (1.1%)	962 (25.8%)
Blood loss anemia	85	28	29	8	321	0.318	0.34	27	189	4	227	19	7	71	25	2 (0.6%)	49 (15.3%)
Deficiency anemia	3,187	1,850	35	10	9,134	0.38	0.455	1,211	5,018	302	5,834	333	127	1,707	715	496 (5.4%)	1,085 (11.9%)
Alcohol abuse	1,286	855	22	7	3,624	0.292	0.382	376	2,030	90	2,305	122	63	676	274	102 (2.8%)	428 (11.8%)
Drug abuse	673	470	11	10	2,021	0.316	0.408	213	1,191	62	1,327	20	11	123	53	160 (7.9%)	58 (2.9%)

[a] The Elixhauser Comorbidity Index (ECI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age. [d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1-score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point. [h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+.

TABLE A.68: T-CARER: The profile of the model and the HSCIC-CCI for the Charlson diagnoses categories, using WDNN (*Sample-1*)

Charlson Comorbidity Index	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Myocardial infarction	3,529	2,223	18	7	9,757	0.335	0.397	1,182	4,989	216	6,067	519	207	5,705	2,039	-857 (-8.8%)	3,817 (39.1%)
Peripheral vascular disease	4,652	2,241	14	12	11,393	0.471	0.494	2,189	4,726	537	6,418	404	199	8,619	3,468	-1,279 (-11.2%)	5,033 (44.2%)
Cerebrovascular disease	2,042	1,315	13	12	5,811	0.308	0.381	629	3,135	128	3,692	272	104	3,582	1,380	-751 (-12.9%)	2,293 (39.5%)
Dementia	3,748	1,841	15	19	11,585	0.371	0.399	1,390	6,009	259	7,628	820	253	5,741	2,021	-631 (-5.4%)	4,078 (35.2%)
Chronic pulmonary disease	2,332	841	10	24	6,340	0.391	0.416	912	2,864	168	3,854	75	35	4,496	1,610	-698 (-11.0%)	2,772 (43.7%)
Rheumatic disease	10,417	4,728	25	8	31,196	0.478	0.519	4,984	16,982	1,799	20,031	342	177	24,360	7,640	-2,656 (-8.5%)	16,137 (51.7%)
Peptic ulcer disease	1,412	350	18	11	5,161	0.3	0.387	424	3,391	90	3,725	70	22	3,446	1,051	-627 (-12.1%)	2,419 (46.9%)
Mild liver disease	910	471	20	12	3,801	0.21	0.28	191	2,629	19	2,877	86	28	1,774	484	-293 (-7.7%)	1,334 (35.1%)
Diabetes, uncomplicated	983	600	21	9	2,733	0.322	0.411	317	1,506	79	1,715	148	57	892	335	-18 (-0.7%)	613 (22.4%)
Diabetes, complicated	7,870	3,993	19	9	22,598	0.421	0.469	3,310	11,796	952	14,267	14,102	4,445	6,729	2,787	523 (2.3%)	13,138 (58.1%)
Hemiplegia or paraplegia	880	553	19	12	2,375	0.298	0.4	262	1,328	75	1,465	261	119	603	287	-25 (-1.1%)	428 (18.0%)
Renal disease	991	529	24	17	2,881	0.377	0.42	374	1,475	79	1,846	685	206	1,482	565	-191 (-6.6%)	1,352 (46.9%)
Malignancy	2,626	1,543	22	12	5,945	0.423	0.48	1,112	2,426	292	3,156	134	56	4,369	1,877	-765 (-12.9%)	2,407 (40.5%)
Moderate or severe liver disease	8,129	4,471	19	8	22,675	0.211	0.293	1,712	12,694	366	14,300	501	190	8,383	3,322	-1,610 (-7.1%)	5,126 (22.6%)
Metastatic solid tumour	323	209	21	8	898	0.251	0.324	81	479	18	568	36	18	442	136	-55 (-6.1%)	317 (35.3%)
Congestive heart failure	2,454	1,123	18	10	6,069	0.226	0.299	554	2,918	104	3,521	49	16	5,205	2,093	-1,539 (-25.4%)	3,051 (50.3%)
AIDS/HIV	63	50	12	7	350	0.317	0.412	20	273	6	286	175	31	71	14	6 (1.7%)	200 (57.1%)

[a] The Charlson Comorbidity Index (CCI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age.
[d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1-score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point.
[h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+.

A.7.3.4 Wide and Deep Neural Network - Sample-2

TABLE A.69: T-CARER: The profile of the model and the HSCIC-CCI for the Elixhauser diagnoses categories, using WDNN (*Sample-2*)

Elixhauser Comorbidity Index	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Congestive heart failure	1,650	835	17	8	3,034	0.486	0.533	802	825	229	1,224	26	17	2,687	1,461	-659 (-21.7%)	1,075 (35.4%)
AIDS/HIV	102	67	13	6	442	0.422	0.446	43	292	15	331	219	49	89	23	20 (4.5%)	227 (51.4%)
Depression	3,029	1,114	38	8	7,196	0.487	0.522	1,474	3,018	498	3,940	327	142	1,462	733	741 (10.3%)	687 (9.5%)
Cardiac arrhythmias	11,194	5,377	14	8	23,006	0.573	0.581	6,409	7,345	2,416	10,650	1,550	764	10,033	5,520	889 (3.9%)	4,137 (18.0%)
Valvular disease	2,019	942	16	10	3,957	0.555	0.577	1,121	1,194	423	1,750	228	109	1,868	1,067	54 (1.4%)	732 (18.5%)
Pulmonary circulation disorder	806	377	22	10	1,745	0.476	0.525	384	665	128	885	101	41	729	397	-13 (-0.7%)	338 (19.4%)
Peripheral vascular disorders	1,980	1,236	14	11	3,780	0.579	0.597	1,146	1,085	465	1,621	108	63	3,074	1,650	-504 (-13.3%)	1,290 (34.1%)
Hypertension, uncomplicated	19,370	8,554	18	8	46,496	0.482	0.512	9,344	19,325	3,029	25,608	5,932	2,400	16,643	8,131	1,213 (2.6%)	10,526 (22.6%)
Hypertension, complicated	1,702	911	15	11	3,002	0.612	0.624	1,041	709	431	1,128	52	30	2,701	1,545	-504 (-16.8%)	1,006 (33.5%)
Paralysis	1,124	570	21	14	2,272	0.558	0.571	627	702	262	1,040	420	194	1,392	740	-113 (-5.0%)	770 (33.9%)
Other neurological disorders	5,672	2,863	35	8	11,863	0.567	0.594	3,214	4,257	1,161	5,762	604	310	3,513	1,944	1,270 (10.7%)	1,434 (12.1%)
Chronic pulmonary disease	13,510	5,865	30	6	30,516	0.589	0.598	7,958	11,845	3,453	15,782	491	269	24,090	10,669	-2,711 (-8.9%)	12,419 (40.7%)
Diabetes, uncomplicated	10,110	5,054	21	8	21,907	0.571	0.578	5,773	7,712	2,260	10,793	11,516	4,729	8,737	4,757	1,016 (4.6%)	9,763 (44.6%)
Diabetes, complicated	887	516	21	11	1,603	0.59	0.619	523	435	217	643	248	129	761	472	51 (3.2%)	335 (20.9%)
Hypothyroidism	3,394	685	19	9	7,859	0.534	0.551	1,812	3,089	723	4,150	695	312	2,940	1,539	273 (3.5%)	1,469 (18.7%)
Renal failure	3,956	2,173	16	10	6,883	0.634	0.643	2,510	1,580	1,081	2,516	106	61	6,161	3,556	-1,046 (-15.2%)	2,239 (32.5%)
Liver disease	1,527	939	22	7	2,955	0.599	0.623	914	937	353	1,319	189	96	1,157	668	246 (8.3%)	473 (16.0%)
Peptic ulcer disease	464	233	23	11	1,158	0.502	0.531	233	514	75	654	43	16	711	306	-73 (-6.3%)	392 (33.9%)
Psychoses	1,121	641	29	31	2,571	0.498	0.518	558	974	149	1,372	103	45	332	175	383 (14.9%)	137 (5.3%)
Lymphoma	462	246	18	10	743	0.543	0.598	251	155	91	238	15	9	478	289	-38 (-5.1%)	152 (20.5%)
Metastatic cancer	1,650	835	17	8	3,034	0.486	0.533	802	825	229	1,224	26	17	2,687	1,461	-659 (-21.7%)	1,075 (35.4%)
Solid tumour	3,510	2,062	18	9	6,425	0.512	0.557	1,796	1,775	587	2,618	131	74	4,750	2,621	-825 (-12.8%)	1,889 (29.4%)
Rheumatoid arthritis	1,603	458	19	8	3,833	0.558	0.563	895	1,548	315	2,068	64	28	3,176	1,369	-474 (-12.4%)	1,681 (43.9%)
Coagulopathy	416	175	40	8	1,003	0.55	0.586	229	451	90	547	42	19	265	140	89 (8.9%)	108 (10.8%)
Obesity	853	343	29	7	2,009	0.522	0.561	445	868	184	1,112	307	136	704	357	88 (4.4%)	474 (23.6%)
Weight loss	709	369	23	12	1,483	0.434	0.494	308	543	88	728	70	29	490	267	41 (2.8%)	218 (14.7%)
Fluid & electrolyte disorders	2,850	1,161	19	14	5,901	0.55	0.567	1,567	1,937	538	2,770	375	182	2,599	1,372	195 (3.3%)	1,139 (19.3%)
Blood loss anemia	69	30	28	8	204	0.435	0.451	30	101	12	128	8	2	51	26	4 (2.0%)	24 (11.8%)
Deficiency anemia	5,006	2,975	33	7	12,003	0.552	0.576	2,763	5,172	995	6,656	525	254	2,548	1,361	1,402 (11.7%)	1,117 (9.3%)
Alcohol abuse	2,132	1,313	23	5	5,421	0.514	0.554	1,095	2,560	366	3,179	175	81	997	547	548 (10.1%)	434 (8.0%)
Drug abuse	941	614	13	5	2,548	0.454	0.5	427	1,268	131	1,559	33	15	261	109	318 (12.5%)	122 (4.8%)

[a] The Elixhauser Comorbidity Index (ECI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age. [d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1-score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point. [h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+.

TABLE A.70: T-CARER: The profile of the model and the HSCIC-CCI for the Charlson diagnoses categories, using WDNN (*Sample-2*)

Charlson Comorbidity Index	**Population Profile**				**T-CARER Profile**							**HSCIC-CCI Profile**				**Comparisons**	
Diagnoses Group [a]	**Prior** [b]	**Male**	**Age** [c]	**LoS** [d]	**Total**	**Sens. (0.5)** [e]	**F1 (.5)** [f]	**TP (.5)** [g]	**TN (.5)** [h]	**TP (.7)**	**TN (.7)**	**CCI 1-3** [i]	**CCI 1-3 (TP)**	**CCI 4+**	**CCI 4+ (TP)**	**Delta Score (.5, 4+)** [j]	**Delta Score (.7, 4+)**
Myocardial infarction	4,079	2,600	21	5	7,806	0.41	0.491	1,671	2,670	536	3,504	396	221	4,493	2,373	-702 (-9.0%)	2,072 (26.5%)
Peripheral vascular disease	4,922	2,385	14	10	8,716	0.611	0.623	3,009	2,065	1,271	3,231	304	163	7,031	4,053	-1,044 (-12.0%)	2,556 (29.3%)
Cerebrovascular disease	1,980	1,236	14	11	3,780	0.579	0.597	1,146	1,085	465	1,621	108	63	3,074	1,650	-504 (-13.3%)	1,290 (34.1%)
Dementia	4,651	2,206	16	14	9,911	0.488	0.525	2,269	3,529	768	4,896	598	277	5,457	2,795	-526 (-5.3%)	2,619 (26.4%)
Chronic pulmonary disease	4,020	1,407	9	15	7,766	0.621	0.591	2,496	1,819	795	3,235	100	55	6,380	3,312	-816 (-10.5%)	2,602 (33.5%)
Rheumatic disease	13,510	5,865	30	6	30,516	0.589	0.598	7,958	11,845	3,453	15,782	491	269	24,090	10,669	-2,711 (-8.9%)	12,419 (40.7%)
Peptic ulcer disease	1,462	394	16	9	3,438	0.573	0.572	837	1,349	296	1,823	34	15	3,110	1,342	-505 (-14.7%)	1,634 (47.5%)
Mild liver disease	695	370	23	10	1,817	0.453	0.498	315	867	98	1,069	76	24	935	405	-90 (-5.0%)	529 (29.1%)
Diabetes, uncomplicated	1,393	853	22	7	2,691	0.614	0.633	855	843	337	1,195	167	83	1,054	613	242 (9.0%)	422 (15.7%)
Diabetes, complicated	10,162	5,087	21	8	22,027	0.571	0.578	5,807	7,756	2,273	10,856	11,560	4,747	8,780	4,780	1,027 (4.7%)	9,804 (44.5%)
Hemiplegia or paraplegia	848	496	22	11	1,501	0.586	0.62	497	394	208	585	212	115	726	456	41 (2.7%)	299 (19.9%)
Renal disease	1,124	570	21	14	2,272	0.558	0.571	627	702	262	1,040	420	194	1,392	740	-113 (-5.0%)	770 (33.9%)
Malignancy	3,962	2,176	16	10	6,898	0.634	0.642	2,512	1,586	1,081	2,524	107	62	6,168	3,558	-1,046 (-15.2%)	2,243 (32.5%)
Moderate or severe liver disease	4,216	2,435	19	9	7,589	0.511	0.561	2,156	2,062	722	3,020	155	89	5,469	3,032	-876 (-11.5%)	2,150 (28.3%)
Metastatic solid tumour	359	244	20	7	607	0.577	0.617	207	143	91	224	39	23	330	196	11 (1.8%)	126 (20.8%)
Congestive heart failure	1,650	835	17	8	3,034	0.486	0.533	802	825	229	1,224	26	17	2,687	1,461	-659 (-21.7%)	1,075 (35.4%)
AIDS/HIV	102	67	13	6	442	0.422	0.446	43	292	15	331	219	49	89	23	20 (4.5%)	227 (51.4%)

[a] The Charlson Comorbidity Index (CCI) diagnoses groups. [b] Total number of patients with prior spells. [c] The Inter-Quartile Range (IQR) of patients' age.
[d] The IQR of patients' length-of-stay. [e] Sensitivity, 50% cut-off point. [f] F1-score, 50% cut-off point. [g] True Positive (TP), 50% cut-off point.
[h] True Negative (TN), 50% cut-off point. [i] Total number of patients scored between 1 to 3 by the HSCIC-CCI. [j] Subtraction of TCARER's True Positive (50% cut-off point) from the HSCIC-CCI of 4+.

A.8 Toolkits

A.8.1 UML Diagrams

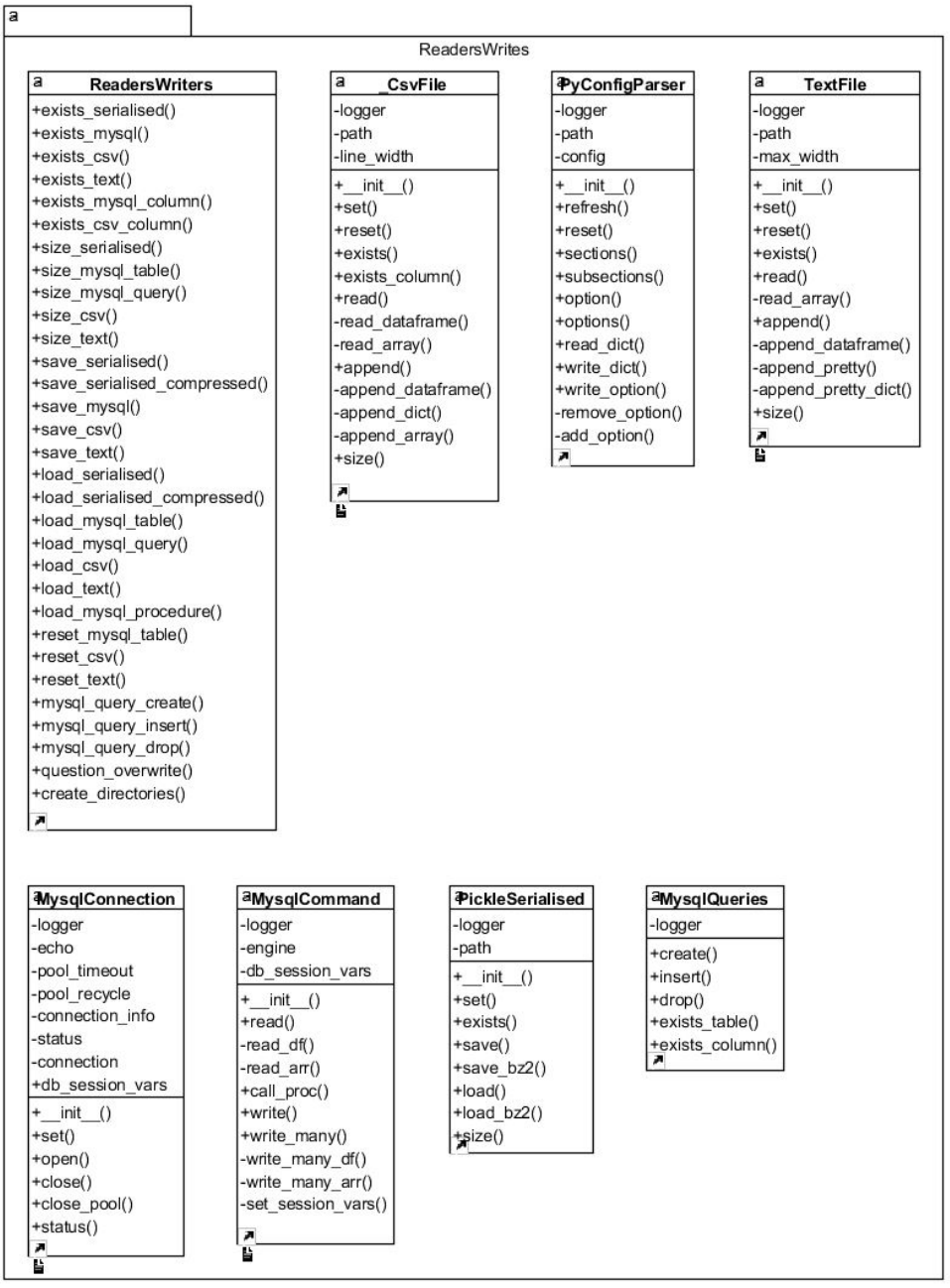

FIGURE A.60: Toolkits: UML diagram of the developed classes in the *ReadersWriters* sub-package

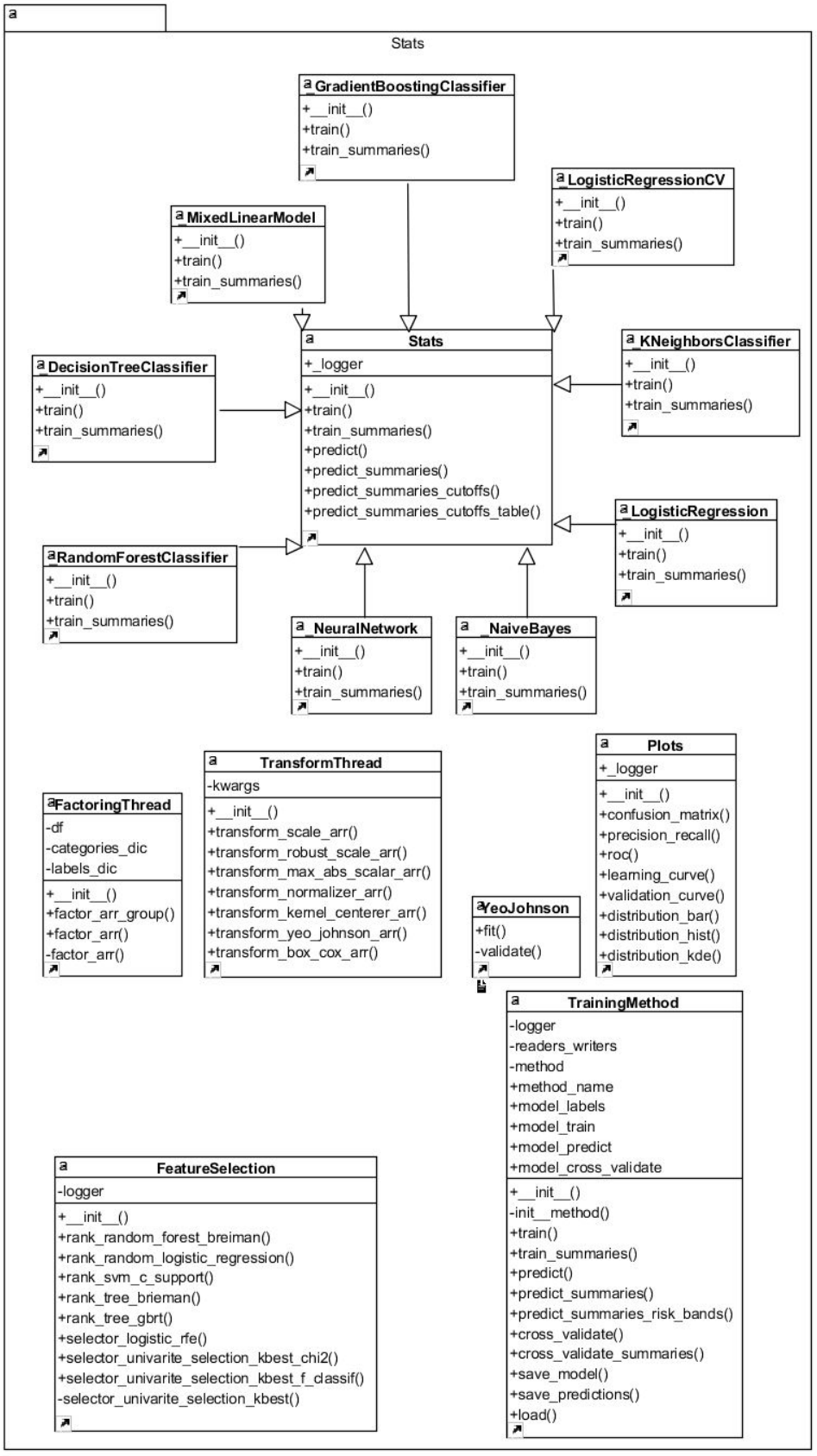

FIGURE A.61: Toolkits: UML diagram of the developed classes in the *Stats* sub-package

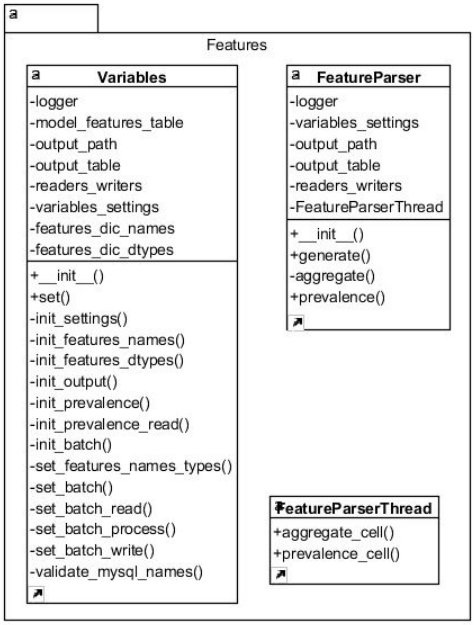

FIGURE A.62: Toolkits: UML diagram of the developed classes in the *Features* sub-package

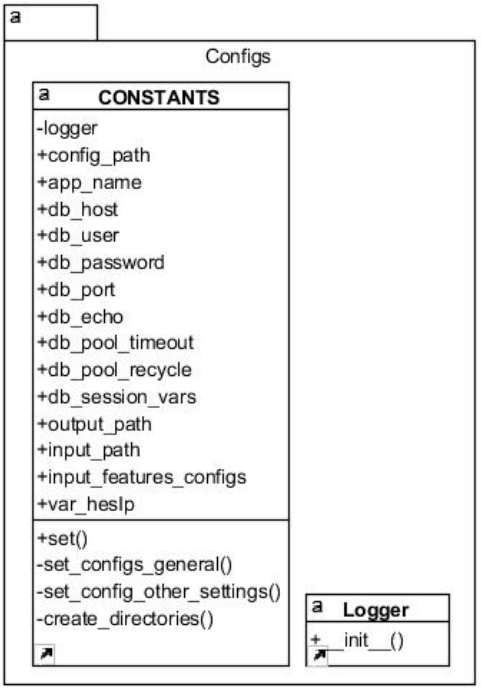

FIGURE A.63: Toolkits: UML diagram of the developed classes in the *Config* sub-packages

Bibliography

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zhang. Tensorflow: A system for large-scale machine learning. CoRR, abs/1605.08695, 2016. URL http://arxiv.org/abs/1605.08695. 51, 81, 118, 139

Rachael Addicott and Steve Dewar. Improving choice at end of life: a descriptive analysis of the impact and costs of the Marie Curie Delivering Choice Programme in Lincolnshire. The King's Fund, 2008. 159

Shola Adeyemi and Thierry J Chaussalet. Models for extracting information on patient pathways. In Intelligent Patient Management, pages 171–182. Springer, 2009. 38

Shola Adeyemi, Eren Demir, and Thierry Chaussalet. Towards an evidence-based decision making healthcare system management: Modelling patient pathways to improve clinical outcomes. Decision Support Systems, 55, 1:117–125, 2013. doi: 10.1016/j.dss.2012.12.039. 7, 15, 62, 155

Agena. Agenarisk software solution, 2016. URL http://www.agenarisk.com. [Retrieved 02.09.2016]. 165

Amr Ahmed and Eric P Xing. Recovering time-varying networks of dependencies in social and biological studies. Proceedings of the National Academy of Sciences, 106 (29):11878–11883, 2009. 161, 162, 163

Dincer Aktuerk, David McNulty, Daniel Ray, Irena Begaj, Neil Howell, Nick Freemantle, and Domenico Pagano. National administrative data produces an accurate and stable risk prediction model for short-term and 1-year mortality following cardiac surgery. International journal of cardiology, 203:196–203, 2016. 22

Nazima Allaudeen, Jeffrey L Schnipper, E John Orav, Robert M Wachter, and Arpana R Vidyarthi. Inability of providers to predict unplanned readmissions. Journal of general internal medicine, 26(7):771–776, 2011. 143

Rachel MacKay Altman. Mixed hidden markov models: an extension of the hidden markov model to the longitudinal data setting. Journal of the American Statistical Association, 102(477):201–210, 2007. 156

George A Alvarez. Representing multiple objects as an ensemble enhances visual cognition. Trends in cognitive sciences, 15(3):122–131, 2011. doi: 10.1016/j.tics.2011.01.003. 93

Alejandro Álvaro-Meca, Thomas Kneib, Ruth Gil Prieto, and Ángel Gil de Miguel. Impact of comorbidities and surgery on health related transitions in pancreatic cancer admissions: A multi state model. Cancer epidemiology, 36(2):e142–e146, 2012. doi: 10.1016/j.canep.2011.12.005. 6, 63

Ruben Amarasingham, Billy J Moore, Ying P Tabak, Mark H Drazner, Christopher A Clark, Song Zhang, W Gary Reed, Timothy S Swanson, Ying Ma, and Ethan A Halm. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Medical care, 48(11):981–988, 2010. 150

Zahid Ansari, James N Laditka, and Sarah B Laditka. Access to health care and hospitalization for ambulatory care sensitive conditions. Medical care research and review, 63(6):719–741, 2006. 16

Apache. Apache license, version 2.0, 2016. URL https://www.apache.org/. [Retrieved 02.09.2016]. 132

JN Armitage and JH Van der Meulen. Identifying co-morbidity in surgical patients using administrative data with the royal college of surgeons charlson score. British Journal of Surgery, 97(5):772–781, 2010. 22, 114

Peter C Austin. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting ami mortality. Statistics in medicine, 26(15):2937–2957, 2007. 39

Peter C Austin, Matthew B Stanbrook, Geoffrey M Anderson, Alice Newman, and Andrea S Gershon. Comparative ability of comorbidity classification methods for administrative data to predict outcomes in patients with chronic obstructive pulmonary disease. Annals of epidemiology, 22(12):881–887, 2012. 20, 68, 114

Steven R Austin, Yu-Ning Wong, Robert G Uzzo, J Robert Beck, and Brian L Egleston. Why summary comorbidity measures such as the charlson comorbidity index and elixhauser score work. Medical care, 53(9):e65, 2015. 124

P Aylin, R Alexandrescu, M H Jen, E K Mayer, and A Bottle. Day of week of procedure and 30 day mortality for elective surgery: retrospective analysis of hospital episode statistics. BMJ, 346, 2013. doi: 10.1136/bmj.f2424. URL http://www.bmj.com/content/346/bmj.f2424. 21

Paul Aylin, Alex Bottle, and Azeem Majeed. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. BMJ: British Medical Journal, 334(7602):1044, 2007. 53

Paul Aylin, Alex Bottle, Min Hua Jen, Steven Middleton, and Foster Intelligence. HSMR mortality indicators, Nov. 2010. URL https://www1.imperial.ac.uk. [Retrieved 02.09.2016]. 20, 69, 79, 91, 105, 114

Timothy L Bailey, Charles Elkan, et al. Fitting a mixture model by expectation maximization to discover motifs in bipolymers, 1994. URL http://www.cs.toronto.edu/~brudno/csc2417_15/10.1.1.121.7056.pdf. [Retrieved 02.09.2016]. 44

Adrian Baker, Paul Leak, Lewis D Ritchie, Amanda J Lee, and Shona Fielding. Anticipatory care planning and integration: a primary care pilot study aimed at reducing unplanned hospitalisation. British Journal of General Practice, 62(595):e113–e120, 2012. 148

Laura-Mae Baldwin, Carrie N Klabunde, Pam Green, William Barlow, and George Wright. In search of the perfect comorbidity measure for use with administrative claims data: does it exist? Medical care, 44(8), 2006. 20, 68

David Barber. Bayesian reasoning and machine learning. Cambridge University Press, 2012. ISBN 9780521518147. 44

Martin Bardsley. Predictive models and data linkage - sharing international experience: Linking disease registry information and predictive modelling to improve quality and efficiency, Sep. 2012. URL http://www.nuffieldtrust.org.uk. [Retrieved 02.09.2016]. 33, 63

Martin Bardsley, Theo Georghiou, Ludovic Chassin, Geraint Lewis, Adam Steventon, and Jennifer Dixon. Overlap of hospital use and social care in older people in england. Journal of health services research & policy, 17(3):133–139, 2012. doi: 10.1258/jhsrp.2011.010171. 3, 14

Martin Bardsley, Ian Blunt, Sian Davies, and Jennifer Dixon. Is secondary preventive care improving? observational study of 10-year trends in emergency admissions for conditions amenable to ambulatory care. BMJ open, 3(1), 2013. doi: 10.1136/bmjopen-2012-002007. 16, 20, 21, 24, 81

Sarah Barnett and Suneetha Ramani Moonesinghe. Clinical risk scores to guide perioperative management. Postgraduate medical journal, 87(1030):535–541, 2011. 21

Mr. Bayes and Mr Price. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, FRS communicated by mr. price, in a letter to john canton, AMFRS. Philosophical Transactions (1683-1775), pages 370–418, 1763. 43

Estela Bee Dagum and Silvia Bianconcini. Seasonal adjustment methods and real time trend-cycle estimation. Springer, 2016. 6

Riccardo Bellazzi, Fulvia Ferrazzi, and Lucia Sacchi. Predictive data mining in clinical medicine: a focus on selected methods and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(5):416–430, 2011. 26

Yoshua Bengio et al. Learning deep architectures for ai. Foundations and trends® in Machine Learning, 2(1):1–127, 2009. 49, 80

James Berger et al. The case for objective bayesian analysis. Bayesian analysis, 1(3): 385–402, 2006. 43

JM Bernardo, MJ Bayarri, JO Berger, AP Dawid, D Heckerman, AFM Smith, M West, et al. The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures. Bayesian statistics, 7:453–464, 2003. 44

John Billings and Tod Mijanovich. Improving the management of care for high-cost medicaid patients. Health Affairs, 26(6):1643–1654, 2007. 151

John Billings, Lisa Zeitel, Joanne Lukomnik, Timothy S Carey, Arthur E Blank, and Laurie Newman. Impact of socioeconomic status on hospital use in new york city. Health affairs, 12(1):162–173, 1993. 16

John Billings, Jennifer Dixon, Tod Mijanovich, David Wennberg, et al. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. Bmj, 333(7563):327, 2006a. doi: 10.1136/bmj.38870.657917.ae. 3, 10, 17, 62, 63, 64, 81, 90, 91, 103, 152

John Billings, Tod Mijanovich, J Dixon, N Curry, D Wennberg, B Darin, and K Steinort. Case finding algorithms for patients at risk of re-hospitalisation. PARR1 and PARR2, Feb. 2006b. URL http://www.kingsfund.org.uk. [Retrieved 02.09.2016]. 64

John Billings, Tod Mijanovich, J Dixon, N Curry, D Wennberg, B Darin, and K Steinort. Risk prediction in sussex with SPOKE, Feb. 2006c. URL http://www.kingsfund.org.uk/sites/files/kf/field/field_document/PARR-case-finding-algorithms-feb06.pdf. [Retrieved 02.01.2016]. 152

John Billings, Ian Blunt, Adam Steventon, Theo Georghiou, Geraint Lewis, and Martin Bardsley. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). BMJ open, 2(4):e001667, 2012. doi: 10.1136/bmjopen-2012-001667. 3, 18, 21, 81, 148

John Billings, Theo Georghiou, Ian Blunt, and Martin Bardsley. Choosing a model to predict hospital admission: an observational study of new variants of predictive models for case finding. BMJ open, 3(8):e003352, 2013. doi: 10.1136/bmjopen-2013-003352. 10, 18, 38, 53, 54, 62, 63, 65, 81, 90, 91, 102, 103, 104, 105, 147

Jeff A Bilmes. What HMMs can do. IEICE TRANSACTIONS on Information and Systems, 89(3):869–891, 2006. [Retrieved 02.09.2016]. 165

Christopher M Bishop and Nasser M Nasrabadi. Pattern recognition and machine learning, volume 1. springer New York, 2006. ISBN 9780387310732. 33, 44, 80

A. Blake, B. North, and M. Isard. Learning multi-class dynamics. In Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II, pages 389–395, Cambridge, MA, USA, 1999. MIT Press. ISBN 0-262-11245-0. URL http://dl.acm.org/citation.cfm?id=340534.340675. 162

Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K Warmuth. Occam's razor. Information processing letters, 24(6):377–380, 1987. 37, 95

Ian Blunt. Focus on preventable admissions - trends in emergency admissions for ambulatory care sensitive conditions, 2001 to 2013, Oct. 2013a. URL http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/131010_qualitywatch_focus_preventable_admissions_0.pdf. [Retrieved 02.01.2016]. 20

Ian Blunt. Focus on preventable admissions - trends in emergency admissions for ambulatory care sensitive conditions, 2001 to 2013, Oct. 2013b. URL http://www.qualitywatch.org.uk. [Retrieved 02.09.2016]. 16, 20

Leonardo Bottaci, Philip J Drew, John E Hartley, Matthew B Hadfield, Ridzuan Farouk, Peter WR Lee, Iain MC Macintyre, Graeme S Duthie, and John RT Monson. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. The Lancet, 350(9076):469–472, 1997. 6

Alex Bottle and Paul Aylin. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. Journal of clinical epidemiology, 64(12):1426–1433, 2011. doi: http://dx.doi.org/10.1016/j.jclinepi.2011.04.004. 21, 28, 69, 105, 114, 121

Alex Bottle, Paul Aylin, and Azeem Majeed. Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis. Journal of the Royal Society of Medicine, 99(8):406–414, 2006. 152

Alex Bottle, Brian Jarman, and Paul Aylin. Strengths and weaknesses of hospital standardised mortality ratios. BMJ, 342:c7116, 2011. doi: 10.1136/bmj.c7116. 69, 79, 91, 105

Alex Bottle, Steven Middleton, Cor J Kalkman, Edward H Livingston, and Paul Aylin. Global comparators project: international comparison of hospital outcomes using administrative data. Health services research, 48(6pt1):2081–2100, 2013. 21, 24, 25

Alex Bottle, Rene Gaudoin, S Jones, and P Aylin. Can valid and practical risk-prediction or casemix adjustment models, including adjustment for comorbidity, be generated from english hospital administrative data (hospital episode statistics)? a national observational study. Health Serv Deliv Res, 2(40), 2014. doi: 10.3310/hsdr02400. 6, 18, 20, 21, 25, 63, 69, 105, 114, 121, 145

Kathryn H Bowles and Jacqueline R Cater. Screening for risk of rehospitalization from home care: Use of the outcomes assessment information set and the probability of readmission instrument. Research in nursing & health, 26(2):118–127, 2003. 143

Leo Breiman. Bagging predictors. Machine learning, 24(2):123–140, 1996. 35, 39

Leo Breiman. Random forests. Machine learning, 45(1):5–32, 2001. doi: doi:10.1023/A:1010933404324. 39, 40, 86, 87, 116, 117, 131

Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. Classification and regression trees. CRC press, 1984. 36, 39

Samuel L. Brilleman, Hugh Gravelle, Sandra Hollinghurst, Sarah Purdy, Chris Salisbury, and Frank Windmeijer. Keep it simple? predicting primary health care costs with clinical morbidity measures. Journal of Health Economics, 35(0):109 – 122, 2014. ISSN 0167-6296. 28

Gavin Brown. Ensemble learning. In Encyclopedia of Machine Learning, pages 312–320. Springer, 2011. doi: 10.1007/978-0-387-30164-8_252. 93

Bupa. Bupa schedule of procedures, 2016. URL http://codes.bupa.co.uk/procedures. [Retrieved 02.01.2016]. 22, 166

Pierre-Régis Burgel, Jean-Louis Paillasseur, and Nicolas Roche. Identification of clinical phenotypes using cluster analyses in copd patients with multiple comorbidities. BioMed research international, 2014, 2014. 139

Reinhard Busse, Alexander Geissler, and Wilm Quentin. Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals. McGraw-Hill Education (UK), 2011. ISBN 9780335245574. 16

Amaia Calderón-Larrañaga, Leanne Carney, Michael Soljak, Alex Bottle, Martyn Partridge, Derek Bell, Gerrard Abi-Aad, Paul Aylin, and Azeem Majeed. Association of population and primary healthcare factors with hospital admission rates for chronic obstructive pulmonary disease in england: national cross-sectional study. Thorax, 66(3):191–196, 2011. 16

Michael Caley and Khesh Sidhu. Estimating the future healthcare costs of an aging population in the UK: expansion of morbidity and the need for preventative care. Journal of Public Health, 33(1):117–122, 2011. doi: 10.1093/pubmed/fdq044. 3, 13

Iain M Carey, Sunil M Shah, Tess Harris, Stephen DeWilde, and Derek G Cook. A new simple primary care morbidity score predicted mortality and better explains between practice variations than the charlson index. Journal of clinical epidemiology, 66(4): 436–444, 2013. 28

Rich Caruana. Multitask learning. In Learning to learn, pages 95–133. Springer, 1998. 34

Rich Caruana, Alexandru Niculescu-Mizil, Geoff Crew, and Alex Ksikes. Ensemble selection from libraries of models. In Proceedings of the twenty-first international conference on Machine learning, page 18. ACM, 2004. doi: 10.1145/1015330.1015432. 93

Gavin C Cawley and Nicola LC Talbot. Preventing over-fitting during model selection via bayesian regularisation of the hyper-parameters. The Journal of Machine Learning Research, 8:841–861, 2007. 91

Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. The Journal of Machine Learning Research, 11:2079–2107, 2010. 91

Anita Charlesworth, Loraine Hawkins, and Louise Marshall. NHS payment reform: lessons from the past and directions for the future, 2014. URL http://nuffield.dh.bytemark.co.uk. [Retrieved 02.09.2016]. 16

Mary Charlson, Ted P Szatrowski, Janey Peterson, and Jeffrey Gold. Validation of a combined comorbidity index. Journal of clinical epidemiology, 47(11):1245–1251, 1994. 20

Mary E Charlson, Peter Pompei, Kathy L Ales, and C Ronald MacKenzie. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases, 40(5):373–383, 1987. 4, 10, 20, 28, 67, 69, 113, 166

Abraham Charnes and William Wager Cooper. Preface to topics in data envelopment analysis. Annals of Operations research, 2(1):59–94, 1984. 93

Yi-Wei Chen and Chih-Jen Lin. Combining svms with various feature selection strategies. In Feature extraction, pages 315–324. Springer, 2006. 36

Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, pages 7–10. ACM, 2016. 50, 81, 118

Xavier Chitnis, T Georghiou, A Steventon, and M Bardsley. The impact of the marie curie nursing service on place of death and hospital use at the end of life, 2012. URL http://www.nuffieldtrust.org.uk. [Retrieved 02.09.2016]. 13

A Clarke, I Blunt, and M Bardsley. Ps18 analysis of emergency 30-day readmissions in england using routine hospital data 2004-2010. is there scope for reduction? Journal of Epidemiology and Community Health, 66(Suppl 1):A45–A45, 2012. doi: 10.1136/jech-2012-201753.117. 2, 3

Adam Coates and Andrew Y Ng. Learning feature representations with k-means. In Neural networks: Tricks of the trade, pages 561–580. Springer, 2012. 80

Eric A Coleman, Jodi D Smith, Janet C Frank, Sung-Joon Min, Carla Parry, and Andrew M Kramer. Preparing patients and caregivers to participate in care delivered across settings: the care transitions intervention. Journal of the American Geriatrics Society, 52(11):1817–1825, 2004. 153

Peter D Congdon. Applied Bayesian hierarchical methods. CRC Press, 2010. ISBN 9781584887218. doi: 10.1201/9781584887218. 7

Nancy R Cook. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation, 115(7):928–935, 2007. doi: 10.1161/circulationaha.106.672402. 27, 29, 103

Richard J Cook and Jerald F Lawless. Statistical issues in modeling chronic disease in cohort studies. Statistics in Biosciences, 6:127–161, 2013. 161

Chiara Cornalba. Clinical and operational risk: A bayesian approach. Methodology and Computing in Applied Probability, 11(1):47–63, 2009. 32

D. R. Cox. A use of complex probabilities in the theory of stochastic processes. Mathematical Proceedings of the Cambridge Philosophical Society, 51:313–319, 4 1955. ISSN 1469-8064. doi: 10.1017/S0305004100030231. URL http://journals.cambridge.org/article_S0305004100030231. 156

David R Cox. The regression analysis of binary sequences. Journal of the Royal Statistical Society. Series B (Methodological), pages 215–242, 1958. 37

Sarah J Crane, Ericka E Tung, Gregory J Hanson, Stephen Cha, Rajeev Chaudhry, and Paul Y Takahashi. Use of an electronic administrative database to identify older community dwelling adults at high-risk for hospitalization or emergency department visits: the elders risk assessment index. BMC health services research, 10(1):338, 2010. 149

Patrick R Cronin, Jeffrey L Greenwald, Gwen C Crevensten, Henry C Chueh, and Adrian H Zai. Development and implementation of a real-time 30-day readmission predictive model. In AMIA Annual Symposium Proceedings, volume 2014, page 424. American Medical Informatics Association, 2014. 145

Martin J Crowder. Multivariate survival analysis and competing risks. CRC Press, 2012. ISBN 9781439875216. 160

Yang Cui, Mahmoud Torabi, Evelyn L Forget, Colleen Metge, Xibiao Ye, Michael Moffatt, and Luis Oppenheimer. Geographical variation analysis of all-cause hospital readmission cases in winnipeg, canada. BMC health services research, 15(1):129, 2015. doi: 10.1186/s12913-015-0807-2. 6, 63

AJ Culyer and JP Newhouse. Handbook of Health Economics: Vol. 1A. Elsevier, 2000. 33

AJ Culyer, MV Pauly, JP Newhouse, TG McGuire, and PP Barros. Handbook of Health Economics: Vol. 2. Elsevier, 2012. 33

Marie Curie. The role of the palliative care network nurse: Focusing on end of life care, Mar. 2010. URL http://www.liv.ac.uk. [Retrieved 02.09.2016]. 159

George E Dahl, Dong Yu, Li Deng, and Alex Acero. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Transactions on Audio, Speech, and Language Processing, 20(1):30–42, 2012. 139

Belur V Dasarathy and Belur V Sheela. A composite classifier system design: concepts and methodology. Proceedings of the IEEE, 67(5):708–713, 1979. 34

Wennberg David, Siegel Matt, Bob Darin, Nadya Filipova, Ronald Russell, Linda Kenney, Klaus Steinort, Tae-Ryong Park, Gokhan Cakmakci, Jennifer Dixon, Natasha Curry, and John Billings. Combined predictive model: Final report, 2006. URL http://www.kingsfund.org.uk. [Retrieved 02.09.2016]. 53

Alisha Davies, Cono Ariti, Theo Georghiou, and Martin Bardsley. Evaluation of complex health and care interventions using retrospective matched control methods, Aug. 2015. URL http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/evaluation_report_final_0.pdf. [Retrieved 02.01.2016]. 61

Roger B Davis and James R Anderson. Exponential survival trees. Statistics in Medicine, 8(8):947–961, 1989. 160

Thomas Dean and Keiji Kanazawa. A model for reasoning about persistence and causation. Computational intelligence, 5(2):142–150, 1989. 45, 139

DeltaRho. Github - delta-rho: Trelliscope - detailed visualization of large complex data in r, 2016. URL www.github.com/delta-rho/trelliscope. [Retrieved 02.09.2016]. 72

Eren Demir. A decision support tool for predicting patients at risk of readmission: A comparison of classification trees, logistic regression, generalized additive models, and multivariate adaptive regression splines. Decision Sciences, 45(5):849–880, 2014. 38

Eren Demir and Thierry Chaussalet. Capturing the re-admission process: focus on time window. Journal of Applied Statistics, 38(5):951–960, 2011. doi: 10.1080/02664761003692415. 6, 63

Eren Demir, Thierry Chaussalet, Haifeng Xie, and Peter H Millard. Modelling risk of readmission with phase-type distribution and transition models. IMA Journal of Management Mathematics, 20(4):357–367, 2009. 38, 150

Howard B Demuth, Mark H Beale, Orlando De Jess, and Martin T Hagan. Neural network design. Martin Hagan, 2014. 48

Li Deng, Dong Yu, et al. Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7(3–4):197–387, 2014. 49

Susan DesHarnais. Current uses of large data sets to assess the quality of providers: construction of risk-adjusted indexes of hospital performance. International journal of technology assessment in health care, 6(02):229–238, 1990. 70

Nancy J Devlin and John Appleby. Getting the most out of PROMS, Apr. 2010. URL https://www.aqesss.qc.ca/docs/pdf/i-media/20100401/PROM_NHS.pdf. [Retrieved 02.01.2016]. 22

Richard A Deyo, Daniel C Cherkin, and Marcia A Ciol. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. Journal of clinical epidemiology, 45(6):613–619, 1992. 20

William D'Hoore, C Sicotte, and C Tilquin. Risk adjustment in outcome assessment: the charlson comorbidity index. Methods of information in medicine, 32(5):382–387, 1993. 20

William D'Hoore, Andr'e Bouckaert, and Charles Tilquin. Practical considerations on the use of the charlson comorbidity index with administrative data bases. Journal of clinical epidemiology, 49(12):1429–1433, 1996. 20, 21, 28

Health Dialog. NHS wales - informing healthcare: Wales predictive model: Final report and technical documentation, Aug. 2008. URL http://www.nliah.com. [Retrieved 02.09.2016]. 17, 53, 151

Ramón Díaz-Uriarte and Sara Alvarez De Andres. Gene selection and classification of microarray data using random forest. BMC bioinformatics, 7(1):3, 2006. 87

Ascensión Doñate-Martínez, Jorge Garcés Ferrer, and Francisco Ródenas Rigla. Application of screening tools to detect risk of hospital readmission in elderly patients in valencian healthcare system (VHS)(spain). Archives of gerontology and geriatrics, 59(2):408–414, 2014. 143, 154

Peter T Donnan, David WT Dorward, Bill Mutch, and Andrew D Morris. Development and validation of a model for predicting emergency admissions over the next year (PEONY): a uk historical cohort study. Archives of Internal Medicine, 168(13): 1416–1422, 2008. 150

Crispin Dowler. Emergency readmissions payment ban being ignored, 2011. URL http://www.hsj.co.uk. [Retrieved 02.09.2016]. 17

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12 (Jul):2121–2159, 2011. 48, 118

Thi Duong, Dinh Phung, Hung Bui, and Svetha Venkatesh. Efficient duration and hierarchical modeling for human activity recognition. Artificial Intelligence, 173(7): 830–856, 2009. 162, 163

Eamonn MP Eeles, Susan V White, Sinead M O'Mahony, Antony J Bayer, and Ruth E Hubbard. The impact of frailty and delirium on mortality in older inpatients. Age and ageing, 41(3):412–416, 2012. 15

Bradley Efron and Robert J Tibshirani. An introduction to the bootstrap. CRC press, 1994. 35

Yves Eggli, Béatrice Desquins, Erol Seker, and Patricia Halfon. Comparing potentially avoidable hospitalization rates related to ambulatory care sensitive conditions in switzerland: the need to refine the definition of health conditions and to adjust for population health status. BMC health services research, 14(1):1, 2014. 61

A Elixhauser and C Steiner. Readmissions to us hospitals by diagnosis, 2010: Statistical brief# 153, 2006. 115, 125

Anne Elixhauser and Claudia Steiner. Readmissions to us hospitals by diagnosis, 2010. Healthcare Cost and Utilization Project Statistical Brief, Apr. 2013. URL http://www.ncbi.nlm.nih.gov/books/NBK154385/. 125

Anne Elixhauser, Claudia Steiner, D Robert Harris, and Rosanna M Coffey. Comorbidity measures for use with administrative data. Medical care, 36(1):8–27, 1998. 4, 10, 20, 67, 70, 113, 166

Cristóbal Esteban, Danilo Schmidt, Denis Krompaß, and Volker Tresp. Predicting sequences of clinical events by using a personalized temporal latent embedding model. In Healthcare Informatics (ICHI), 2015 International Conference on, pages 130–139. IEEE, 2015. 81

Mark Fackrell. Modelling healthcare systems with phase-type distributions. Health care management science, 12(1):11–26, 2009. 156

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. Journal of machine learning research, 9(Aug):1871–1874, 2008. 116

Tom Fawcett. An introduction to ROC analysis. Pattern recognition letters, 27(8): 861–874, 2006. 29

Willliam Feller. An introduction to probability theory and its applications, volume 2. John Wiley & Sons, 2008. 37

Norman E Fenton and Martin D Neil. Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press, 2012. ISBN 9781439809105. 5, 32

Robert B Fetter, Youngsoo Shin, Jean L Freeman, Richard F Averill, and John D Thompson. Case mix definition by diagnosis-related groups. Medical care, 18(2): i–53, 1980. 16, 67, 70, 113

Jonathan F Finks, Nicholas H Osborne, and John D Birkmeyer. Trends in hospital volume and operative mortality for high-risk surgery. New England Journal of Medicine, 364(22):2128–2137, 2011. 21

Claudia Fischer, Helen A. Anema, and Niek S. Klazinga. The validity of indicators for assessing quality of care: a review of the european literature on hospital readmission rate. The European Journal of Public Health, 2011. ISSN 1101-1262. doi: 10.1093/eurpub/ckr165. 28, 105

Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs. Longitudinal data analysis. CRC Press, 2008. ISBN 9781420011579. 24

Jonathan Agner Forsberg, Rikard Wedin, Bjarne H Bauer, Henrik CF andHansen, Minna Laitinen, Clement S Trovik, Johnny Keller, Patrick J Boland, and John H Healey. External validation of the bayesian estimated tools for survival (BETS) models in patients with surgically treated skeletal metastases. BMC cancer, 12(1): 493, 2012. 160

Tobias Freund, Michel Wensing, Cornelia Mahler, Jochen Gensichen, Antje Erler, Martin Beyer, Ferdinand M Gerlach, Joachim Szecsenyi, and Frank Peters-Klimm. Development of a primary care-based complex care management intervention for chronically ill patients at high risk for hospitalization: a study protocol. Implement Sci, 5 (70DOI):5908–5, 2010. 142

Tobias Freund, Stephen M Campbell, Stefan Geissler, Cornelia U Kunz, Cornelia Mahler, Frank Peters-Klimm, and Joachim Szecsenyi. Strategies for reducing potentially avoidable hospitalizations for ambulatory care-sensitive conditions. The Annals of Family Medicine, 11(4):363–370, 2013. 142

Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In icml, volume 96, pages 148–156, 1996. 35

Jerome H Friedman. Multivariate adaptive regression splines. The annals of statistics, pages 1–67, 1991. 39

Jerome H Friedman. Greedy function approximation: a gradient boosting machine. Annals of statistics, pages 1189–1232, 2001. 86

Keinosuke Fukunaga. Introduction to statistical pattern recognition. Academic press, 2013. doi: 10.1016/c2009-0-27872-x. 93

The King's Fund. Improving the quality of care in general practice, 2011. URL http://www.kingsfund.org.uk. [Retrieved 02.09.2016]. 164

The King's Fund. The health and social care system in 2025 - a view of the future, 2013. URL http://www.kingsfund.org.uk. [Retrieved 02.09.2016]. 14

The King's Fund. Predicting and reducing re-admission to hospital, 2016a. URL http://www.kingsfund.org.uk. [Retrieved 02.09.2016]. 3, 17

The King's Fund. WinBUGS, 2016b. URL http://www.openbugs.net. [Retrieved 02.09.2016]. 165

The King's Fund. Avoiding hospital admissions, Dec. 2016c. URL http://www.kingsfund.org.uk/sites/files/kf/Avoiding-Hospital-Admissions-Sarah-Purdy-December2010_0.pdf. [Retrieved 02.01.2016]. 20

Joshua J Gagne, Robert J Glynn, Jerry Avorn, Raisa Levin, and Sebastian Schneeweiss. A combined comorbidity score predicted mortality in elderly patients better than existing scores. Journal of clinical epidemiology, 64(7):749–759, 2011. 20, 114

Lalit Garg, Sally McClean, Brian Meenan, and Peter Millard. Non-homogeneous markov models for sequential pattern mining of healthcare data. IMA Journal of Management Mathematics, 20(4):327–344, 2009. 160

Lalit Garg, Sally McClean, Brian Meenan, and Peter Millard. A non-homogeneous discrete time markov model for admission scheduling and resource planning in a cost or capacity constrained healthcare system. Health care management science, 13(2): 155–169, 2010. 160

Lalit Garg, Sally McCLEAN, Brian J Meenan, and Peter Millard. Phase-type survival trees and mixed distribution survival trees for clustering patients' hospital length of stay. Informatica, 22(1):57–72, 2011. 160

Lalit Garg, Sally I McClean, Maria Barton, Brian J Meenan, and Ken Fullerton. Intelligent patient management and resource planning for complex, heterogeneous, and stochastic healthcare systems. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 42(6):1332–1345, 2012. 160

G David Garson. Hierarchical linear modeling: Guide and applications. Sage, 2012. ISBN 9781412998857. 155, 158

Seymour Geisser. Predictive interference: an introduction, volume 55. CRC Press, 1993. 32

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. Bayesian data analysis, volume 2. Taylor & Francis, 2013. ISBN 9781439840955. 43, 44

D Gencaga, EE Kuruoglu, and A Ertuzun. Modeling non-gaussian time-varying vector autoregressive processes by particle filtering. Multidimensional Systems and Signal Processing, 21(1):73–85, 2010. 163

T. Georghiou, S. Davies, A. Davies, and M. Bradley. Understanding patterns of health and social care at the end of life, Oct. 2012. URL http://www.nuffieldtrust.org.uk. [Retrieved 02.09.2016]. 13

Zoubin Ghahramani and Geoffrey E Hinton. Variational learning for switching state-space models. Neural computation, 12(4):831–864, 2000. 162

Zoubin Ghahramani, Thomas L Griffiths, and Peter Sollich. Bayesian nonparametric latent feature models. BAYESIAN STATISTICS, 8(1):1–25, 2007. 34

William A Ghali, Ruth E Hall, Amy K Rosen, Arlene S Ash, and Mark A Moskowitz. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. Journal of clinical epidemiology, 49(3):273–278, 1996. 20

Asghar Ghasemi, Saleh Zahediasl, et al. Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2): 486–489, 2012. 85

Walter R Gilks. Markov chain monte carlo. Encyclopedia of Biostatistics, 2005. 45

Jeff Gill. Bayesian methods: A social and behavioral sciences approach, volume 20. CRC press, 2014. ISBN 9781439862490. 43

Alex Goldstein, Adam Kapelner, Justin Bleich, and Emil Pitkin. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics, 24(1):44–65, 2015. 36

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. MIT Press, 2016. 5

Google. Tensorflow - an open-source software library for machine intelligence, 2016. URL https://www.tensorflow.org/. [Retrieved 02.09.2016]. 118, 131

Susan Graham, Deborah Estrin, Eric Horvitz, Isaac Kohane, Elizabeth Mynatt, and Ida Sim. Information technology research challenges for healthcare: From discovery to delivery. ACM SIGHIT Record, 1(1):4–9, 2011. 34

Michael Green, Jonas Bjork, Jakob Forberg, Ulf Ekelund, Lars Edenbrandt, and Mattias Ohlsson. Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. Artificial intelligence in medicine, 38(3):305–318, 2006. doi: 10.1016/j.artmed.2006.07.006. 6, 63

Piotr S Gromski, Yun Xu, Elon Correa, David I Ellis, Michael L Turner, and Royston Goodacre. A comparative investigation of modern feature selection and classification approaches for the analysis of mass spectrometry data. Analytica chimica acta, 829: 1–8, 2014. 87

Murat M Gunal and Michael Pidd. Discrete event simulation for performance modelling in health care: a review of the literature. Journal of Simulation, 4(1):42–51, 2010. 157

Fan Guo, Steve Hanneke, Wenjie Fu, and Eric P Xing. Recovering temporally rewiring networks: A model-based approach. In Proceedings of the 24th international conference on Machine learning, pages 321–328. ACM, 2007. 162

Sunil Kumar Gupta, Santu Rana, Dinh Phung, and Svetha Venkatesh. Keeping up with innovation: A predictive framework for modeling healthcare data with evolving clinical interventions. In Proceedings of the 2014 SIAM International Conference on Data Mining, pages 235–243, 2014. doi: 10.1137/1.9781611973440.27. 6, 63

Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. Machine learning, 46(1-3): 389–422, 2002. doi: 10.1109/csb.2003.1227389. 42, 86, 87

Aurlien Gron. Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly Media, 2017. ISBN 9781491962299. 48

Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas, and H Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature, 405(6789):947–951, 2000. 47

Alon Halevy, Peter Norvig, and Fernando Pereira. The unreasonable effectiveness of data. IEEE Intelligent Systems, 24(2):8–12, 2009. 37

Patricia Halfon, Yves Eggli, Isaline Prêtre-Rohrbach, Danielle Meylan, Alfio Marazzi, and Bernard Burnand. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. Medical care, 44(11): 972–981, 2006. 152

Bradley G Hammill, Lesley H Curtis, Gregg C Fonarow, Paul A Heidenreich, Clyde W Yancy, Eric D Peterson, and Adrian F Hernandez. Incremental value of clinical data

beyond claims data in predicting 30-day outcomes after heart failure hospitalization. Circulation: Cardiovascular Quality and Outcomes, 4(1):60–67, 2011. 149

David J. Hand. Measuring classifier performance: acoherent alternative to the area under the roc curve. Machine Learning, 77(1):103–123, 2009. ISSN 1573-0565. doi: 10.1007/s10994-009-5119-5. URL http://dx.doi.org/10.1007/s10994-009-5119-5. 29

Steve Hanneke, Wenjie Fu, Eric P Xing, et al. Discrete temporal models of social networks. Electronic Journal of Statistics, 4:585–605, 2010. 162

Lars Kai Hansen and Peter Salamon. Neural network ensembles. IEEE transactions on pattern analysis and machine intelligence, 12(10):993–1001, 1990. 34

Penny Hansford and Helen Meehan. Gold standards framework: Improving community care, 2005. URL http://www.goldstandardsframework.org.uk. [Retrieved 02.09.2016]. 159

Omar Hasan, David O Meltzer, Shimon A Shaykevich, Chaim M Bell, Peter J Kaboli, Andrew D Auerbach, Tosha B Wetterneck, Vineet M Arora, James Zhang, and Jeffrey L Schnipper. Hospital readmission in general medicine patients: a prediction model. Journal of general internal medicine, 25(3):211–219, 2010. 150

Jeff Hawkins and Sandra Blakeslee. On intelligence. Macmillan, 2007. 2

Jeff Heaton. Introduction to neural networks with Java. Heaton Research, Inc., 2008. 50

Jonathan E Helm, Adel Alaeddini, Jon M Stauffer, Kurt M Bretthauer, and Ted A Skolarus. Reducing hospital readmissions by integrating empirical prediction with resource optimization. Production and Operations Management, 2015. doi: 10.1111/poms.12377. 6, 63

Claire Henry and Pam Fenner. An introduction to the NHS end of life care programme. Clinical Practice Development, 1(1):56–60, 2007. 158

Ralf Herbrich, Thore Graepel, and Colin Campbell. Bayes point machines. The Journal of Machine Learning Research, 1:245–279, 2001. 9, 45, 86, 108

Geoffrey E Hinton. Learning multiple layers of representation. Trends in cognitive sciences, 11(10):428–434, 2007. 34

Geoffrey E Hinton and Drew Van Camp. Keeping the neural networks simple by minimizing the description length of the weights. In Proceedings of the sixth annual conference on Computational learning theory, pages 5–13. ACM, 1993. 27

Julia Hippisley-Cox and Carol Coupland. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score. BMJ open, 3(8):e003482, 2013. 18, 146

Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Synthesis Lectures on Artificial Intelligence and Machine Learning, 12(1):55–67, 1970. 82

Thomas Hofmann, Bernhard Schölkopf, and Alexander J Smola. Kernel methods in machine learning. The annals of statistics, pages 1171–1220, 2008. 42, 46

CD'Arcy J Holman, David B Preen, Natalya J Baynham, Judith C Finn, and James B Semmens. A multipurpose comorbidity scoring system performed better than the charlson index. Journal of clinical epidemiology, 58(10):1006–1014, 2005. 21, 28, 69, 114, 123, 153

Dawn E Holmes and Lakhmi C Jain. Data Mining: Foundations and Intelligent Paradigms: Volume 3: Medical, Health, Social, Biological and Other Applications, volume 3. Springer, 2012. 32

Eric Horvitz. From data to predictions and decisions: Enabling evidence-based healthcare, 2010. URL http://research.microsoft.com. [Retrieved 02.09.2016]. 34

Harold Hotelling. Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6):417, 1933. 82

Torsten Hothorn, Kurt Hornik, and Achim Zeileis. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics, 15(3):651–674, 2006. 40

Torsten Hothorn, Kurt Hornik, Carolin Strobl, and Achim Zeileis. Party: A laboratory for recursive partytioning, 2010. URL http://cran.r-project.org. [Retrieved 02.09.2016]. 87

Stuart Howell, Michael Coory, Jennifer Martin, and Stephen Duckett. Using routine inpatient data to identify patients at risk of hospital readmission. BMC Health Services Research, 9(1):96, 2009. 38, 150

William H. Hsu. Bayesian network tools in java (BNJ) software solution, 2016. URL http://bnj.sourceforge.net. [Retrieved 02.09.2016]. 165

Jeremy Hurst and Sally Williams. Can NHS hospitals do more with less?, Jan. 2012. URL http://www.nuffieldtrust.org.uk/sites/files/nuffield/can-nhs-hospitals-do-more-with-less_full-report-120112.pdf. [Retrieved 02.01.2016]. 21

Hayley A Hutchings, Bridie A Evans, Deborah Fitzsimmons, Jane Harrison, Martin Heaven, Peter Huxley, Mark-Rhys Kingston, Leo Lewis, Ceri Phillips, Alison Porter, et al. Predictive risk stratification model: a progressive cluster-randomised trial in chronic conditions management (PRISMATIC) research protocol. Trials, 14(1):301, 2013. 17, 61, 151

Dyfed W Huws, Deborah Cashmore, Robert G Newcombe, Catherine Roberts, Judith Vincent, and Glyn Elwyn. Impact of case management by advanced practice nurses in primary care on unplanned hospital admissions: a controlled intervention study. BMC Health Services Research, 8(1):115, 2008. doi: 10.1186/1472-6963-8-115. 6, 63

IPython. Ipython - interactive computing, 2016. URL https://ipython.org/. [Retrieved 02.09.2016]. 127

Alan Julian Izenman. Modern multivariate statistical techniques. Regression, classification and manifold learning, 2008. 26, 72

Robert A Jacobs, Michael I Jordan, Steven J Nowlan, and Geoffrey E Hinton. Adaptive mixtures of local experts. Neural computation, 3(1):79–87, 1991. 36

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An introduction to statistical learning, volume 112. Springer, 2013. 99

Jean-Marie Januel, Jean-Christophe Luthi, Hude Quan, François Borst, Patrick Taffé, William A Ghali, and Bernard Burnand. Improved accuracy of co-morbidity coding over time after the introduction of icd-10 administrative data. BMC health services research, 11(1):194, 2011. 114

Harold Jeffreys. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 186(1007):453–461, 1946. ISSN 00804630. URL http://www.jstor.org/stable/97883. 44

Min-Hua Jen, Alex Bottle, Graham Kirkwood, Ron Johnston, and Paul Aylin. The performance of automated case-mix adjustment regression model building methods in a health outcome prediction setting. Health care management science, 14(3): 267–278, 2011. 149

Peter B Jensen, Lars J Jensen, and Søren Brunak. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics, 13 (6):395–405, 2012. doi: 10.1038/nrg3208. 23, 26

S Jhanji, B Thomas, A Ely, D Watson, CJ Hinds, and RM Pearse. Mortality and utilisation of critical care resources amongst high-risk surgical patients in a large nhs trust. Anaesthesia, 63(7):695–700, 2008. 21

Lorne D Johnson and Georgios Sakoulis. Maximizing equity market sector predictability in a bayesian time-varying parameter model. Computational Statistics & Data Analysis, 52(6):3083–3106, 2008. 163

A Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. Advances in neural information processing systems, 14:841, 2002. 43

Michael I Jordan and Robert A Jacobs. Hierarchical mixtures of experts and the em algorithm. Neural computation, 6(2):181–214, 1994. 36

Jupyter. Jupyter kernels, 2016. URL https://github.com/jupyter/. [Retrieved 02.09.2016]. 127

Ashok Kay Kanagarajah, Peter Lindsay, Anne Miller, and David Parker. An exploration into the uses of agent-based modeling to improve quality of healthcare. In Unifying Themes in Complex Systems, pages 471–478. Springer, 2008. 157

Devan Kansagara, Honora Englander, Amanda Salanitro, David Kagen, Cecelia Theobald, Michele Freeman, and Sunil Kripalani. Risk prediction models for hospital readmission: a systematic review. JAMA: the journal of the American Medical Association, 306(15):1688–1698, 2011. doi: 10.1001/jama.2011.1515. 33, 53, 63

Moreson H Kaplan and Alvan R Feinstein. The importance of classifying initial comorbidity in evaluating the outcome of diabetes mellitus. Journal of chronic diseases, 27(7-8):387–404, 1974. 69

Panagiotis Kasteridis, Andrew Street, Matthew Dolman, Lesley Gallier, Kevin Hudson, Jeremy Martin, and Ian Wyer. Who would most benefit from improved integrated care? implementing an analytical strategy in south somerset. International journal of integrated care, 15, 2015. 142

Korina Katsaliaki and Navonil Mustafee. Applications of simulation within the healthcare context. Journal of the Operational Research Society, 62(8):1431–1451, 2010. 157

John Kautter, Gregory C Pope, Melvin Ingber, Sara Freeman, Lindsey Patterson, Michael Cohen, and Patricia Keenan. The HHS-HCC risk adjustment model for individual and small group markets under the affordable care act, 2014. URL https://

www.cms.gov/mmrr/Downloads/MMRR2014_004_03_a03.pdf. [Retrieved 02.01.2016]. 21, 144

Maria Kelly, Linda Sharp, Fiona Dwane, Tracy Kelleher, and Harry Comber. Factors predicting hospital length-of-stay and readmission after colorectal resection: a population-based study of elective and emergency admissions. BMC Health Services Research, 12(1):77, 2012. ISSN 1472-6963. doi: 10.1186/1472-6963-12-77. URL http://dx.doi.org/10.1186/1472-6963-12-77. 156

Kim Keltie, Helen Cole, Mick Arber, Hannah Patrick, John Powell, Bruce Campbell, and Andrew Sims. Identifying complications of interventional procedures from UK routine healthcare databases: a systematic search for methods using clinical codes. BMC medical research methodology, 14(1):126, 2014. 22

Wayne Khuu, Vincy Chan, and Angela Colantonio. A systematic review protocol for measuring comorbidity in inpatient rehabilitation for non-traumatic brain injury. Systematic reviews, 4(1):1, 2015. 20, 68

Linda T Kohn, Janet M Corrigan, Molla S Donaldson, et al. To err is human: building a safer health system, volume 627. National Academies Press, 2000. 32

Daphne Koller and Nir Friedman. Probabilistic Graphical Models: Principles and Techniques. MIT Press, 2009. 5, 26, 27, 34, 43, 44, 45, 81, 139, 161, 162

Harlan M Krumholz, Ya-Ting Chen, Yun Wang, Viola Vaccarino, Martha J Radford, and Ralph I Horwitz. Predictors of readmission among elderly survivors of admission with heart failure. American heart journal, 139(1):72–77, 2000. 153

Max Kuhn and Kjell Johnson. Applied Predictive Modeling. Springer, 2013. 5, 25, 26, 27, 32, 83, 99, 139

Elena Kulinskaya, Diana Kornbrot, and Haiyan Gao. Length of stay as a performance indicator: robust statistical methodology. IMA Journal of Management Mathematics, 16(4):369–381, 2005. 155

Raymond NC Kuo and Mei-Shu Lai. Comparison of rx-defined morbidity groups and diagnosis-based risk adjusters for predicting healthcare costs in taiwan. BMC health services research, 10(1):126, 2010. 20, 68

Martijn Lappenschaar, Arjen Hommersom, Peter JF Lucas, Joep Lagro, and Stefan Visscher. Multilevel bayesian networks for the analysis of hierarchical health care data. Artificial intelligence in medicine, 57, 3:171–183, 2013a. 162, 163

Martijn Lappenschaar, Arjen Hommersom, Peter JF Lucas, Joep Lagro, Stefan Visscher, Joke C Korevaar, and Francois G Schellevis. Multilevel temporal bayesian networks can model longitudinal change in multimorbidity. Journal of clinical epidemiology, 66(12):1405–1416, 2013b. 162, 163

Mauro Laudicella, Paolo Li Donni, and Peter C Smith. Hospital readmission rates: signal of failure or success? Journal of health economics, 32(5):909–921, 2013. 164

Williams Lea. More care, less pathway: A review of the liverpool care pathway, 2013. URL https://www.gov.uk. [Retrieved 02.09.2016]. 159

Charles Leadbeater and Jake Garber. Dying for change, 2010. URL http://www.demos.co.uk. [Retrieved 02.09.2016]. 13

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. Nature, 521(7553): 436–444, 2015. 118

Eva K Lee, Fan Yuan, Daniel A Hirsh, Michael D Mallory, and Harold K Simon. A clinical decision tool for predicting patient care characteristics: patients returning within 72 hours in the emergency department. In AMIA Annual Symposium Proceedings, volume 2012, page 495. American Medical Informatics Association, 2012. 6, 63

Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In Proceedings of the 26th annual international conference on machine learning, pages 609–616. ACM, 2009. 81

Klaus Lemke. A predictive model to identify patients at risk of unplanned 30-day acute care hospital readmission. In Healthcare Informatics (ICHI), 2013 IEEE International Conference on, pages 551–556. IEEE, 2013. 143

Geraint Lewis. PARR++ is dead: long live predictive modelling, Aug. 2011. URL http://www.nuffieldtrust.org.uk. [Retrieved 02.09.2016]. 3, 17, 38, 62, 63, 64, 91

Geraint Lewis, Natasha Curry, and Martin Bardsley. Choosing a predictive risk model: a guide for commissioners in england, 2011. URL https://www.primis.nottingham.ac.uk. [Retrieved 02.09.2016]. 1, 2, 3, 13, 14, 17, 23, 31, 32, 33, 53, 63

Bing Li, Dewey Evans, Peter Faris, Stafford Dean, and Hude Quan. Risk adjustment performance of charlson and elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. BMC Health services research, 8(1):12, 2008. 70

Pengxiang Li, Michelle M Kim, and Jalpa A Doshi. Comparison of the performance of the cms hierarchical condition category (cms-hcc) risk adjuster with the charlson and elixhauser comorbidity measures in predicting mortality. BMC health services research, 10(1):245, 2010. 21

Jessica R Lieffers, Vickie E Baracos, Marcy Winget, and Konrad Fassbender. A comparison of charlson and elixhauser comorbidity measures to predict colorectal cancer survival using administrative health data. Cancer, 117(9):1957–1965, 2011. 20, 68

Richard Lilford, Mohammed A Mohammed, David Spiegelhalter, and Richard Thomson. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. The Lancet, 363(9415):1147–1154, 2004. 27

Nan Liu, Zhi Xiong Koh, Eric Chern-Pin Chua, Licia Mei-Ling Tan, Zhiping Lin, Bilal Mirza, and Marcus Eng Hock Ong. Risk scoring for prediction of acute cardiac complications from imbalanced clinical data. IEEE journal of biomedical and health informatics, 18(6):1894–1902, 2014. 37

David Lunn, Chris Jackson, Nicky Best, Andrew Thomas, and David Spiegelhalter. The BUGS book: A practical introduction to Bayesian analysis. CRC press, 2012. 43

David Lyon, Gillian A Lancaster, Steve Taylor, Chris Dowrick, and Hannah Chellaswamy. Predicting the likelihood of emergency admission to hospital of older people: development and validation of the emergency admission risk likelihood index (EARLI). Family practice, 24(2):158–167, 2007. 19, 151

David JC MacKay. Information theory, inference and learning algorithms. Cambridge university press, 2003. 27

Ronny S Mans, Wil van der Aalst, and Rob JB Vanwersch. Process mining in healthcare: evaluating and exploiting operational healthcare processes. Springer, 2015. 139

Adele H Marshall and Mariangela Zenga. Recent developments in fitting coxian phase-type distributions in healthcare. In Applied stochastic models and data analysis. The XIII International Conference, Vilnius, Lithuania, page 482, 2009. 156

Deborah A Marshall, Lina Burgos-Liz, Maarten J IJzerman, Nathaniel D Osgood, William V Padula, Mitchell K Higashi, Peter K Wong, Kalyan S Pasupathy, and William Crown. Applying dynamic simulation modeling methods in health care delivery researchthe simulate checklist: report of the ispor simulation modeling emerging good practices task force. Value in health, 18(1):5–16, 2015. 139

Louise Marshall, Anita Charlesworth, and Jeremy Hurst. The NHS payment system: evolving policy and emerging evidence, 2014. URL http://www.nuffieldtrust.org.uk. [Retrieved 02.09.2016]. 17

Mark E Maruish. User's manual for the SF-36v2 Health Survey. Quality Metric Incorporated, 2011. 142

Antonello Maruotti. Mixed hidden markov models for longitudinal data: An overview. International Statistical Review, 79(3):427–454, 2011. 156

Mathworks. Matlab software solution, 2016. URL http://www.mathworks.co.uk. [Retrieved 02.09.2016]. 165

Randall Matignon. Neural network modeling using SAS enterprise miner. AuthorHouse, 2005. 7

3M. 3M grouping software for potentially preventable events, 2016. URL http://solutions.3m.com/wps/portal/3M/en_US/Health-Information-Systems/HIS/Products-and-Services/Products-List-A-Z/PPR-and-PPC-Grouping-Software/. [Retrieved 02.01.2016]. 143

ACI. Risk stratification - a discussion paper for NSW health's approach to risk stratification, Dec. 2014. URL http://www.aci.health.nsw.gov.au. [Retrieved 02.09.2016]. 33, 63

ACI. Evidence check - targeting integrated care social and clinical risk factors, May 2015. URL http://www.aci.health.nsw.gov.au/__data/assets/pdf_file/0008/274742/Evidence_Check_Targeting_integrated_care_Social_and_clinical_risk_factors.pdf. [Retrieved 02.01.2016]. 20

AHRQ. Guide to prevention quality indicators: Hospital admission for ambulatory care sensitive conditions, Oct. 2001. URL http://www.qualityindicators.ahrq.gov/. [Retrieved 02.01.2016]. 166

AHRQ. AHRQ - HCUP - clinical classifications software (ccs) for ICD-10-CM, 2016a. URL https://www.hcup-us.ahrq.gov. [Retrieved 02.09.2016]. 22, 115, 125, 166

AHRQ. AHRQ - HCUP - elixhauser comorbidity software for ICD-10-CM, 2016b. URL https://www.hcup-us.ahrq.gov. [Retrieved 02.09.2016]. 4, 10, 20, 67, 70, 113, 116, 166

BGS. Fit for frailty (part 1), Jun. 2014. URL http://www.bgs.org.uk/index.php/fit-for-frailty. [Retrieved 02.01.2016]. 15

BGS. Fit for frailty (part 2), Jan. 2015. URL http://www.bgs.org.uk/index.php/fitforfrailty-2m. [Retrieved 02.01.2016]. 15

CDC. A framework for patient-centered health risk assessments, 2016. URL http://www.cdc.gov/policy/hst/hra/. [Retrieved 02.01.2016]. 143

CMS. ICD-10 MS-DRG conversion project, 2016a. URL https://www.cms.gov/Medicare/Coding/ICD10/ICD-10-MS-DRG-Conversion-Project.html. [Retrieved 02.01.2016]. 144

CMS. Medicare managed care manual, 2016b. URL https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Internet-Only-Manuals-IOMs-Items/CMS019326.html. [Retrieved 02.01.2016]. 21, 144

CMS. Acute inpatient PPS, 2016c. URL https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/index.html. [Retrieved 02.01.2016]. 144

DCLG. English indices of deprivation, Dec. 2012. URL https://www.gov.uk/government/collections/english-indices-of-deprivation. [Retrieved 02.01.2016]. 166

DFI. My hospital guide: Essential reading for smart spending, Nov. 2010. URL http://myhospitalguide.drfosterintelligence.co.uk/. [Retrieved 02.01.2016]. 20

DFI. My hospital guide, 2013. URL http://myhospitalguide.drfosterintelligence.co.uk. [Retrieved 02.09.2016]. 114

DFI. Palliative care - dr foster intelligence, Mar. 2014. URL www.drfoster.com/wp-content/uploads/2014/08/Palliative-report-final-260314.pdf. [Retrieved 02.01.2016]. 78

DH. Combined predictive model - final report and technical documentation, Dec. 2006. URL http://www.kingsfund.org.uk. [Retrieved 02.09.2016]. 3, 10, 17, 38, 62, 63, 64, 90, 103, 105

DH. Advance care planning: a guide for health and social care staff, Feb. 2007. URL http://www.ncpc.org.uk. [Retrieved 02.09.2016]. 159

DH. End of life care strategy: Promoting high quality care for all adults at the end of life, Jul. 2008. URL https://www.gov.uk. [Retrieved 02.09.2016]. 158

DH. End of life care strategy: quality markers and measures for end of life care, Jun. 2009. URL http:/www.dh.gov.uk. [Retrieved 02.09.2016]. 158

DH. The operating framework for the NHS in england 2011/12, 2010. URL https://www.gov.uk. [Retrieved 02.09.2016]. 17

DH. Risk stratification and next steps with DH risk prediction tools: Patients at risk of re-hospitalisation and the combined predictive model, Aug. 2011a. URL https://www.gov.uk. [Retrieved 02.09.2016]. 3, 18, 33, 63

DH. Preferred priorities for care (PPC), Feb. 2011b. URL http://www.endoflifecare.nhs.uk. [Retrieved 02.09.2016]. 159

DH. Long term conditions compendium of information (third edition), Apr. 2012. URL https://www.gov.uk. [Retrieved 02.09.2016]. 3, 14

DH. An evaluation of the route to success resources, related tools and frameworks covering disease specific areas: heart failure; advanced kidney disease; dementia; and long term neurological conditions, Dec. 2013a. URL http://www.endoflifecare.nhs.uk. [Retrieved 02.09.2016]. 158

DH. Business case: for the health and care modernisation transition programme, 2013b. URL https://www.gov.uk. [Retrieved 02.09.2016]. 1

DH. The financial sustainability of nhs bodies, 2014. URL http://www.nao.org.uk. [Retrieved 02.09.2016]. 16, 17

DH. Department of health - policies, 2016. URL https://www.gov.uk. [Retrieved 02.09.2016]. 25

DSL. GeNIe & SMILE software solution, 2016. URL http://genie.sis.pitt.edu. [Retrieved 02.09.2016]. 165

FCS. Fairer care funding - the report of the commission on funding of care and support, 2011. URL http://www.ilis.co.uk. [Retrieved 02.09.2016]. 13

Health Dialog. Delaying disease progression across a population, Aug. 2016a. URL http://healthdialog.com/insights/research/delaying-disease-progression-across-population. [Retrieved 02.01.2016]. 144

Health Dialog. Uncovering insights that enable action using innovative analytics and the care pathways framework, 2016b. URL http://www.healthdialog.com/solutions/analytics-and-care-pathways. [Retrieved 02.01.2016]. 144

HI. Risk stratification and case finding, Dec. 2013. URL http://www.health-intelligence.com/wp-content/uploads/2013/03/Risk-Stratification-Booklet-A4.pdf. [Retrieved 02.01.2016]. 151

HSCIC. Hes data dictionary, Sep 2010. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 96

HSCIC. National clinical coding standards ICD-10 4th edition (2013), Aug. 2011. URL https://isd.hscic.gov.uk. [Retrieved 02.09.2016]. 90

HSCIC. Hospital episode statistics, emergency readmissions to hospital within 28 days of discharge - financial year 2011/12, Dec. 2013a. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 3

HSCIC. Hospital episode statistics - admitted patient care, england - 2013-14, Jan 2013b. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 24

HSCIC. Hospital episode statistics (HES) analysis guide, Jan. 2014a. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 54, 72

HSCIC. National clinical coding standards OPCS-4 - accurate data for quality information, Apr. 2014b. URL http://systems.hscic.gov.uk. [Retrieved 02.09.2016]. 116

HSCIC. Quality and outcomes framework - prevalence, achievements and exceptions report, Oct. 2014c. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 16, 164

HSCIC. Summary hospital-level mortality indicator, Oct. 2014d. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 20, 69, 79, 90

HSCIC. methodology to create provider and CIP spells from HES APC data, 2014e. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 78, 90

HSCIC. Summary hospital-level mortality indicator (SHMI) - methodology development log, Dec. 2015. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 69, 79

HSCIC. Healthcare resource groups, 2016a. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 21, 166

HSCIC. HRG v3.5 toolkit, 2016b. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 76

HSCIC. Summary hospital-level mortality indicator, 2016c. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 69, 79, 91

HSCIC. Hospital episode statistics (HES), Jan. 2016d. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 2, 54, 55

HSCIC. The processing cycle and hes data quality, 2016e. URL http://content.digital.nhs.uk/. [Retrieved 02.09.2016]. 77

HSCIC. What HES data are available?, 2016f. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 61, 164

HSCIC. General and personal medical services, england - 2004-2014, Mar. 2016g. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 164

HSCIC. Hospital episode statistics, 2016h. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 164, 166

HSCIC. Methodological changes, 2016i. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 24, 25

HSCIC. National clinical coding standards OPCS-4 classification, 2016j. URL http://systems.hscic.gov.uk/data/clinicalcoding/codingstandards/opcs4. [Retrieved 02.01.2016]. 22, 166

HSCIC. NHS outcomes framework indicators, 2016k. URL http://www.hscic.gov.uk/nhsof. [Retrieved 02.01.2016]. 20

HSCIC. Quality and outcomes framework, 2016l. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 164

HSCIC. Aligning statistics to the new health landscape - health and social care information centre, 2016m. URL http://www.hscic.gov.uk. [Retrieved 02.09.2016]. 164

JHU. The johns hopkins ACG system - version 11.0 technical reference guide, Nov. 2014. URL http://acg.jhsph.org/index.php/component/content/article/95-about-the-acg-system/system-development/286-new-in-version-11. [Retrieved 02.01.2016]. 21, 143, 166

JHU. The johns hopkins adjusted clinical groups (ACG) system - care management, 2016a. URL http://acg.jhsph.org/index.php/solutions/care-management. [Retrieved 02.01.2016]. 143

JHU. The johns hopkins adjusted clinical groups (ACG) system - population health, 2016b. URL http://acg.jhsph.org/index.php/solutions/population-profiling. [Retrieved 02.01.2016]. 143

JHU. Pra and PraPlus screening instruments, 2016c. URL http://www.jhsph.edu/research/centers-and-institutes/roger-c-lipitz-center-for-integrated-health-care/Pra_PraPlus/. [Retrieved 02.01.2016]. 143

LexisNexis. LexisNexis healthcare, 2016a. URL http://www.lexisnexis.com/risk/health-care/. [Retrieved 02.01.2016]. 144

LexisNexis. LexisNexis population health monitor, 2016b. URL http://www.lexisnexis.com/risk/products/health-care/population-health-monitor.aspx. [Retrieved 02.01.2016]. 144

LexisNexis. LexisNexis risk navigator performance, 2016c. URL http://www.lexisnexis.com/risk/products/health-care/risk-navigator-perfomance.aspx. [Retrieved 02.01.2016]. 144

NCEPOD. Knowing the risk: a review of the peri-operative care of surgical patients, Dec. 2011. URL http://www.ncepod.org.uk/2011report2/downloads/POC_fullreport.pdf. [Retrieved 02.01.2016]. 22

NHS. SPARRA mental health: Scottish patients at risk of readmission and admission (to psychiatric hospitals or units), Jan. 2009. URL http://www.isdscotland.org. [Retrieved 02.09.2016]. 150

NHS. Anticipatory care planning - frequently asked questions, Mar. 2010. URL http://www.scottishneurological.org.uk/content/res/Anticipatory_Care_Planning.pdf. [Retrieved 02.01.2016]. 148

NHS. Scottish patients at risk of readmission and admission (SPARRA) - version 3 (developing risk prediction to support preventative and anticipatory care in scotland), Oct. 2011. URL http://www.isdscotland.org. [Retrieved 02.09.2016]. 17, 81, 148

NHS. Whole system partnership - end of life care engagement event. Presentation, Jul. 2012. 14

NHS. Reducing in-hospital mortality - observations arising from AQuA's work, May 2013a. URL https://www.aquanw.nhs.uk. [Retrieved 02.09.2016]. 24

NHS. Nhs hospital data and datasets: A consultation - nhs england, 2013b. URL http://england.nhs.uk. [Retrieved 02.09.2016]. 55

NHS. NHS england publishes CCG funding allocations for next two years following adoption of new formula, 2013c. URL http://www.england.nhs.uk. [Retrieved 02.09.2016]. 1

NHS. Improving end of life care through early recognition of need, 2013d. URL http://www.thewholesystem.co.uk. [Retrieved 02.09.2016]. 14

NHS. National tariff payment system 2014/15, Dec. 2013e. URL https://www.gov.uk. [Retrieved 02.09.2016]. 74

NHS. Five year forward view, Oct. 2014. URL https://www.england.nhs.uk. [Retrieved 02.09.2016]. 3

NHS. QOF database, 2016a. URL http://www.gpcontract.co.uk. [Retrieved 02.09.2016]. 164

NHS. Care outside hospital programme, 2016b. URL http://www.institute.nhs.uk/news/care_outside_hospitals_-_banner_image/care_outside_hospital_programme.html. [Retrieved 02.01.2016]. 166

NHS. Payment by results data, 2016c. URL http://content.digital.nhs.uk. [Retrieved 02.09.2016]. 55

NHS. Read codes, 2016d. URL https://digital.nhs.uk. [Retrieved 02.09.2016]. 54, 77

NHS. Sus essentials - secondary uses service - essential user guide, 2016e. URL http://content.digital.nhs.uk. [Retrieved 02.09.2016]. 55

NHS. Directory of organisation patient safety incident reports, 2016f. URL http://www.nrls.npsa.nhs.uk. [Retrieved 02.09.2016]. 164

NICE. NICE guidance, 2016. URL http://www.nice.org.uk. [Retrieved 02.09.2016]. 15, 16

NRLS. Quarterly data summary issue 2 - patient safety, Sep. 2010a. URL http://www.nrls.npsa.nhs.uk. [Retrieved 02.09.2016]. 164

NRLS. Organisation patient safety incident reports workbook - NRLS cluster list, Mar. 2010b. URL http://www.nrls.npsa.nhs.uk. [Retrieved 02.09.2016]. 76, 164

NRLS. NRLS data: NRLS clusters and benchmarking information, 2014. URL http://www.nrls.npsa.nhs.uk. [Retrieved 02.09.2016]. 164, 166

NTIS. Medicare severity diagnosis related grouper (MS-DRG) and medicare code editor (MCE) ICD-10 software, 2016. URL https://www.ntis.gov/products/grouper/. [Retrieved 02.01.2016]. 144

OECD. Health at a glance: Europe 2014, Dec. 2014. URL http://www.oecd.org. [Retrieved 02.09.2016]. 15, 20

ONS. Health expectancies at birth and age 65 in the united kingdom, 2008-10, Aug. 2012. URL http://www.ons.gov.uk/ons/rel/disability-and-health-measurement/health-expectancies-at-birth-and-age-65-in-the-united-kingdom/2008-10/index.html. [Retrieved 02.01.2016]. 15

ONS. Population estimates for england and wales, mid-2002 to mid-2010 revised (sub-national), Apr. 2013. URL http://www.ons.gov.uk. [Retrieved 02.09.2016]. 164

ONS. Avoidable mortality in england and wales, 2001 - 2012, May 2014a. URL http://www.ons.gov.uk. [Retrieved 02.09.2016]. 164

ONS. Quality and methodology information: Mortality statistics in england and wales, Jul. 2014b. URL http://www.ons.gov.uk. [Retrieved 02.09.2016]. 102

ONS. Population estimates, 2016. URL http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Population+Estimates. [Retrieved 02.01.2016]. 166

OPTUM. HealthNumerics-RISC - predictive models: A successful approach to risk stratification, Jan. 2014. URL https://www.optum.com/content/dam/optum/resources/whitePapers/Optum_HNRISC_White%20Paper.pdf. [Retrieved 02.01.2016]. 142, 166

OPTUM. Episode treatment groups, 2016a. URL https://etg.optum.com/etg-links/episode-treatment-groups/. [Retrieved 02.01.2016]. 166

OPTUM. HealthNumerics-RISC - easily identify high-risk patients in order to proactively improve quality of care, 2016b. URL http://www.optum.co.uk/healthcare/risk-prediction/healthnumerics-risc.html. [Retrieved 02.01.2016]. 142

OPTUM. OPTUM: Health care technology, health care management and health services, 2016c. URL https://www.optum.com/. [Retrieved 02.01.2016]. 142

OPTUM. SF health surveys, 2016d. URL https://www.optum.com/optum-outcomes/what-we-do/health-surveys.html. [Retrieved 02.01.2016]. 142

R Foundation. The r project for statistical computing, 2016. URL www.r-project.org. [Retrieved 02.09.2016]. 72

SGV. Hospital admission risk program (HARP) background paper, Mar. 2002. URL http://www.health.vic.gov.au/archive/archive2005/harp/harpbgpap.pdf. [Retrieved 02.01.2016]. 146

SGV. Hospital admission risk program monitoring and evaluation framework, Oct. 2011. URL https://www2.health.vic.gov.au/Api/downloadmedia/%7B7EBC12EB-385B-4395-A542-B7C9DB8399E5%7D. [Retrieved 02.01.2016]. 146

SGV. Policies and guidelines, 2016. URL https://www2.health.vic.gov.au/about/publications/policiesandguidelines. [Retrieved 02.01.2016]. 146

TPP. SystemOne, 2016a. URL http://www.tpp-uk.com/products/systmone. [Retrieved 02.01.2016]. 77

TPP. Systmone - tpp, 2016b. URL https://www.tpp-uk.com. [Retrieved 02.09.2016]. 54

TRUST. Predictive risk stratification: impact on care for people with or at risk of chronic conditions (PRISMATIC), 2016. URL http://www.trustresearch.org.uk/en/prismatic.htm. [Retrieved 02.01.2016]. 151

Verisk Health. DxCG risk analytics, 2016. URL https://www.veriskhealth.com/answers/population-answers/dxcg-risk-analytics. [Retrieved 02.01.2016]. 21, 142

YNHHSC/CORE. Hospital-wide all-cause unplanned readmission measure - final technical report, Mar. 2012. URL http://www.qualitynet.org. [Retrieved 02.09.2016]. 18, 147

YNHHSC/CORE. 2015 measure updates and specifications report - hospital-wide all-cause unplanned readmission measure - version 4.0, Mar. 2015. URL http://www.qualitynet.org. [Retrieved 02.09.2016]. 18, 147

Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4):115–133, 1943. 47

Hemalkumar B Mehta, Francesca Dimou, Deepak Adhikari, Nina P Tamirisa, Eric Sieloff, Taylor P Williams, Yong-Fang Kuo, and Taylor S Riall. Comparison of comorbidity scores in predicting surgical outcomes. Medical care, 54(2):180–187, 2016. 21, 22, 69, 114, 121

Mohsen Mesgarpour, Thierry Chaussalet, and Salma Chahed. Ensemble risk model of emergency readmission (ERMER) to hospital, using bayes point machine method and hospital episode statistics inpatient data. Manuscript submitted for publication, 2016. 145

Borislava Mihaylova, Andrew Briggs, Anthony O'Hagan, and Simon G Thompson. Review of statistical methods for analysing healthcare resources and costs. Health economics, 20(8):897–916, 2011. 23, 26, 27, 105

Milliman. Milliman advanced risk adjusters, 2016. URL http://us.milliman.com. [Retrieved 02.09.2016]. 142

Thomas P Minka. A family of algorithms for approximate bayesian inference, Sep. 2001a. URL http://research.microsoft.com. [Retrieved 02.09.2016]. 46, 108

Thomas P Minka. Expectation propagation for approximate bayesian inference. In Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence, pages 362–369. Morgan Kaufmann Publishers Inc., 2001b. ISBN 1-55860-800-1. 9, 27, 45, 46, 108, 109

Thomas P Minka. Expectation propagation for approximate bayesian inference. In Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence, pages 362–369. Morgan Kaufmann Publishers Inc., 2001c. 44

Tom Minka. A roadmap to research on EP, 2016. URL http://research.microsoft.com. [Retrieved 02.09.2016]. 109

Judith Mistichelli. Diagnosis related groups (drgs) and the prospective payment system: Forecasting social implications. Bioethics Research Library, 1984. 16, 67, 70, 113

Kristine Monteith, James L Carroll, Kevin Seppi, and Tony Martinez. Turning bayesian model averaging into bayesian model combination. In Neural Networks (IJCNN), The 2011 International Joint Conference on, pages 2657–2663. IEEE, 2011. 35

Suneetha Ramani Moonesinghe, Michael G Mythen, Priya Das, KM Rowan, and MP Grocott. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery. Anesthesiology, 119(4):959, 2013. 21, 22

Zoe Morrison, Bernard Fernando, Dipak Kalra, Kathrin Cresswell, Ann Robertson, Akiko Hemmi, and Aziz Sheikh. An evaluation of different levels of structuring within the clinical record, 2012. URL http://www.birmingham.ac.uk. [Retrieved 02.09.2016]. 54

EF Ruth Morrissey, James C McElnay, Michael Scott, and Brian J McConnell. Influence of drugs, demographics and medical history on hospital readmission of elderly patients. Clinical drug investigation, 23(2):119–128, 2003. 153

David Glen Mosley. The hierarchical condition category model-an improved comorbidity adjustment tool for predicting mortality in medicare populations?, 2013. URL http://arizona.openrepository.com. [Retrieved 02.09.2016]. 28

Irene M Mullins, Mir S Siadaty, Jason Lyman, Ken Scully, Carleton T Garrett, W Greg Miller, Rudy Muller, Barry Robson, Chid Apte, Sholom Weiss, et al. Data mining and clinical data repositories: Insights from a 667,000 patient data set. Computers in biology and medicine, 36(12):1351–1377, 2006. doi: 10.1016/j.compbiomed.2005.08.003. 81

Kevin Murphy. Bayes net toolbox software solution for matlab, 2016. URL https://code.google.com. [Retrieved 02.09.2016]. 165

Kevin P Murphy. Machine learning: a probabilistic perspective. MIT press, 2012. ISBN 9780262304320. 26, 27, 35, 36, 39, 80, 92, 96, 97, 124

Kevin P Murphy, Yair Weiss, and Michael I Jordan. Loopy belief propagation for approximate inference: An empirical study. In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, pages 467–475. Morgan Kaufmann Publishers Inc., 1999. 44

Samer AM Nashef, Frangois Roques, Philippe Michel, E Gauducheau, S Lemeshow, R Salamon, EuroSCORE Study Group, et al. European system for cardiac operative risk evaluation (EuroSCORE). European Journal of Cardio-Thoracic Surgery, 16(1):9–13, 1999. 22

Radford M Neal. Probabilistic inference using markov chain monte carlo methods, Sep. 1993. URL https://www.cs.princeton.edu. [Retrieved 02.09.2016]. 27

Yurii Nesterov and Arkadii Nemirovskii. Interior-point polynomial algorithms in convex programming. SIAM, 1994. 48

Marcel F Neuts. Probability distributions of phase type. Purdue University. Department of Statistics, 1974. 156

Marcel F Neuts. Probability distributions of phase type. In Marcel F Neuts, editor, Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach. Courier Dover Publications, 1981. 156

Joseph P Newhouse. Reimbursing health plans and health providers: efficiency in production versus selection. Journal of economic literature, 34(3):1236–1263, 1996. 33

Kenney Ng, Amol Ghoting, Steven R Steinhubl, Walter F Stewart, Bradley Malin, and Jimeng Sun. PARAMO: A parallel predictive modeling platform for healthcare analytic research using electronic health records. Journal of biomedical informatics, 48:160–170, 2014. 2, 5

Jon Nicholl. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. Journal of epidemiology and community health, 61(11):1010–1013, 2007. 28, 105

Johan Nilsson, Mattias Ohlsson, Lars Thulin, Peter Höglund, Samer AM Nashef, and Johan Brandt. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. The Journal of thoracic and cardiovascular surgery, 132(1):12–19, 2006. doi: 10.1016/j.jtcvs.2005.12.055. 6, 63

Ellen Nolte and Martin McKee. Caring for people with chronic conditions: a health system perspective. McGraw-Hill, 2008. 15

Klaus Nordhausen. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition by Trevor Hastie, Robert Tibshirani, Jerome Friedman. Blackwell Publishing Ltd, 2009. doi: 10.1111/j.1751-5823.2009.00095_18.x. 96, 99, 102

James R Norris. Markov chains. Cambridge university press, 1998. 156

Nancy L Novotny and Mary Ann Anderson. Prediction of early readmission in medical inpatients using the probability of repeated admission instrument. Nursing research, 57(6):406–415, 2008. 143

Nuffield-Trust. Predicting risk of hospital readmission with PARR-30, Aug. 2012. URL http://www.nuffieldtrust.org.uk. [Retrieved 02.09.2016]. 2, 3, 18, 38

Royal College of Physicians. HES for physicians - a guide to the use of information derived from hospital episode statistics, Mar. 2006. URL https://www.rcplondon.ac.uk. [Retrieved 02.09.2016]. 24

University of Washington. The graphical models toolkit (GMTK), 2016. URL http://melodi.ee.washington.edu. [Retrieved 02.09.2016]. 165

Kyoung-Su Oh and Keechul Jung. Gpu implementation of neural networks. Pattern Recognition, 37(6):1311–1314, 2004. 34

Michael O'Kelly and Bohdana Ratitch. Clinical Trials with Missing Data: A Guide for Practitioners. Wiley, 2 2014. ISBN 9781118762516. doi: 10.1002/9781118762516. 24

Health Quality Ontario. Early identification of people at-risk of hospitalization: Hospital admission risk prediction (HARP) - a new tool for supporting providers and patients, 2013a. URL http://www.hqontario.ca. [Retrieved 02.09.2016]. 146

Health Quality Ontario. Early identification of people at-risk of hospitalization: Hospital admission risk prediction (HARP) - a new tool for supporting providers and patients - technical appendix, 2013b. URL http://www.hqontario.ca. [Retrieved 02.09.2016]. 146

David Opitz and Richard Maclin. Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research, pages 169–198, 1999. 93

Oracle. Mysql, 2016. URL www.mysql.com. [Retrieved 02.09.2016]. 71

Jason W Osborne. Improving your data transformations: Applying the box-cox transformation. Practical Assessment, Research & Evaluation, 15(12):1–9, 2010. 85

Robert B Ohara and D Johan Kotze. Do not log-transform count data. Methods in Ecology and Evolution, 1(2):118–122, 2010. 85

James T Pacala, Chad Boult, and Lisa Boult. Predictive validity of a questionnaire that identifies older persons at risk for hospital admission. Journal of the American Geriatrics Society, 43(4):374–377, 1995. 143

James T Pacala, Chad Boult, Richard L Reed, and Ellen Aliberti. Predictive validity of the Pra instrument among older recipients of managed care. Journal of the american geriatrics society, 45(5):614–617, 1997. 143

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359, 2010. 139

Ulrich Paquet. Bayesian inference for latent variable models, Jul. 2008. URL https://www.cl.cam.ac.uk. [Retrieved 02.09.2016]. 27

Fiona Paton, Paul Wilson, and Kath Wright. Predictive validity of tools used to assess the risk of unplanned admissions: A rapid review of the evidence, 2014. URL https://www.york.ac.uk. [Retrieved 02.09.2016]. 10, 33, 63, 103, 151

Vladimir Pavlovic, James M Rehg, and John MacCormick. Learning switching linear models of human motion. In NIPS, pages 981–987. Citeseer, 2000. 162

Sheila Payne, Jane Seymour, Christine Ingleton, anita Sargeant, and Hugh Kidd. Independent evaluation of the marie curie cancer care: Delivering choice programme, Jan. 2008. URL http://uat-w1-www.mariecurie.org.uk. [Retrieved 02.09.2016]. 158

Rupert M Pearse, David A Harrison, Philip James, David Watson, Charles Hinds, Andrew Rhodes, R Michael Grounds, and E David Bennett. Identification and

characterisation of the high-risk surgical population in the united kingdom. Critical Care, 10(3):R81, 2006. 22

Karl Pearson. Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58:240–242, 1895. 86

Karl Pearson. Principal components analysis. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 6(2):559, 1901. 82

Linda Peelen, Nicolette F de Keizer, Evert de Jonge, Robert-Jan Bosman, Ameen Abu-Hanna, and Niels Peek. Using hierarchical dynamic bayesian networks to investigate dynamics of organ failure in patients in the intensive care unit. Journal of biomedical informatics, 43(2):273–286, 2010. doi: 10.1016/j.jbi.2009.10.002. 6, 63

Michael J Pencina, Ralph B D'Agostino, Ralph B D'Agostino, and Ramachandran S Vasan. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine, 27(2):157, 2008. doi: 10.1002/sim.2929. 27, 29, 103

Margaret S Pepe, Holly Janes, and Christopher I Li. Net risk reclassification p values: valid or misleading? Journal of the National Cancer Institute, page dju041, 2014. 29

Bruno-Edouard Perrin, Liva Ralaivola, Aurelien Mazurie, Samuele Bottani, Jacques Mallet, and Florence d'Alche Buc. Gene networks inference using dynamic bayesian networks. Bioinformatics, 19(suppl 2):ii138–ii148, 2003. 163

Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh. Deepcare: A deep dynamic memory model for predictive medicine. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pages 30–41. Springer, 2016. 50

Lorien Y Pratt, SJ Hanson, CL Giles, and JD Cowan. Discriminability-based transfer between neural networks. Advances in neural information processing systems, pages 204–204, 1993. 34

S James Press. Subjective and objective Bayesian statistics: principles, models, and applications, volume 590. John Wiley & Sons, 2009. 43

KL Protopapa. Is there a place for the surgical outcome risk tool app in routine clinical practice? British journal of hospital medicine (London, England: 2005), 77 (11):612–613, 2016. 22

KL Protopapa, JC Simpson, NCE Smith, and SR Moonesinghe. Development and validation of the surgical outcome risk tool (SORT). British Journal of Surgery, 101 (13):1774–1783, 2014. 22

DR Prytherch, MS Whiteley, B Higgins, PC Weaver, WG Prout, and SJ Powell. POSSUM and portsmouth POSSUM for predicting mortality. British Journal of Surgery, 85(9):1217–1220, 1998. 22

S Purdy, T Griffin, C Salisbury, and D Sharp. Ambulatory care sensitive conditions: terminology and disease coding need to be more specific to aid policy makers and clinicians. Public health, 123(2):169–173, 2009a. 16

Sarah Purdy, Tom Griffin, Chris Salisbury, and Debbie Sharp. Prioritizing ambulatory care sensitive hospital admissions in england for research and intervention: a delphi exercise. Primary Health Care Research & Development, 11(1):41, 2009b. 16, 17

QualityWatch. Focus on: International comparisons of healthcare quality, Jul. 2016. URL http://www.qualitywatch.org.uk. [Retrieved 02.09.2016]. 20

Hude Quan, Vijaya Sundararajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L Duncan Saunders, Cynthia A Beck, Thomas E Feasby, and William A Ghali. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Medical care, pages 1130–1139, 2005. doi: 10.1097/01.mlr.0000182534.19832.83. 20, 21, 28, 105

Wilm Quentin, Alexander Geissler, David Scheller-Kreinsen, and Reinhard Busse. Understanding DRGs and DRG-based hospital payment in europe. Diagnosis-Related groups in Europe: Moving towards transparency, efficiency and quality in hospitals. Open University Press and Mc Graw Hill: Berkshire, England, pages 23–36, 2011. 16

J. Ross Quinlan. Induction of decision trees. Machine learning, 1(1):81–106, 1986. 39

J Ross Quinlan. C4. 5: programs for machine learning. Elsevier, 2014. 39

R-Project. Package party, 2010. URL http://cran.r-project.org. [Retrieved 02.09.2016]. 87

James Raftery, Paul Roderick, and Andrew Stevens. Potential use of routine databases in health technology assessment. Health Technology Assessment, 9(20):1–106, 2005. doi: 10.3310/hta9200. 53

Arvind Rao, Alfred O Hero III, James Douglas Engel, et al. Inferring time-varying network topologies from gene expression data. EURASIP Journal on Bioinformatics and Systems Biology, 2007:7–7, 2007. 163

Andrew P Reimer, Alex Milinovich, and Elizabeth A Madigan. Data quality assessment framework to assess electronic medical record data for use in research. International

journal of medical informatics, 90:40–47, 2016. doi: 10.1016/j.ijmedinf.2016.03.006. 105

Microsoft Research. Infer.net software solution, 2016. URL http://research.microsoft.com. [Retrieved 02.09.2016]. 46, 108, 109, 131, 165

Thomas E Rix and Tom Bates. Pre-operative risk scores for the prediction of outcome in elderly people who require emergency surgery. World Journal of Emergency Surgery, 2:16, 2007. 21

Christian Robert and George Casella. Monte Carlo Statistical Methods. Springer, 2004. 27

Joshua W. Robinson and Alexander J. Hartemink. Learning non-stationary dynamic bayesian networks. J. Mach. Learn. Res., 11:3647–3680, December 2010. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=1756006.1953047. 161, 163

M Rodriguez. The failure of predictive modeling and why we follow the herd, 2011. URL http://www.cfclaw.com. [Retrieved 02.09.2016]. 5

Lior Rokach. Ensemble methods for classifiers. In Data mining and knowledge discovery handbook, pages 957–980. Springer, 2005. 93

Lior Rokach. Pattern classification using ensemble methods, volume 75. World Scientific, 2010. 35, 92

Patrick S Romano, Leslie L Roos, and James G Jollis. Presentation adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. Journal of clinical epidemiology, 46(10):1075–1079, 1993. 20, 21, 28

Sheldon M Ross. Introduction to Probability Models. Academic Press, 1993. 156

Sebastian Ruder. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747, 2016. 48

Stuart J Russell and Peter Norvig. Artificial intelligence: a modern approach (International Edition). {Pearson US Imports & PHIPEs}, 2002. 93

RM Sakia. The box-cox transformation technique: a review. The statistician, pages 169–178, 1992. 85

Neil J Salkind. Encyclopedia of measurement and statistics. Sage Publications, 2006. 85

Claude Sammut and Geoffrey I Webb. Encyclopedia of machine learning. Springer Science & Business Media, 2011. 35, 36, 92

Colin Sanderson and Jennifer Dixon. Conditions for which onset or hospital admission is potentially preventable by timely and effective ambulatory care. Journal of health services research & policy, 5(4):222–230, 2000. 16

Robert E Schapire. The strength of weak learnability. Machine learning, 5(2):197–227, 1990. 34

Robert E Schapire and Yoav Freund. Boosting: Foundations and algorithms. MIT press, 2012. 35

Jürgen Schmidhuber. Deep learning in neural networks: An overview. Neural networks, 61:85–117, 2015. 49

Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In International Conference on Artificial Neural Networks, pages 583–588. Springer, 1997. 26

scikit learn. Scikit-learn - machine learning in python, 2016. URL http://scikit-learn.org. [Retrieved 02.09.2016]. 116, 117

John F Scoggins and Donald L Patrick. The use of patient-reported outcomes instruments in registered clinical trials: evidence from clinicaltrials.gov. Contemporary clinical trials, 30(4):289–292, 2009. 142

Bayes Server. Bayes server software solution, 2016. URL http://www.bayesserver.com. [Retrieved 02.09.2016]. 165

Martin Sewell. Ensemble learning. RN, 11(02), 2008. 35, 36, 92

Cosma Rohilla Shalizi. Advanced data analysis from an elementary point of view, May. 2015. URL http://www.stat.cmu.edu. [Retrieved 02.09.2016]. 109

Mansour TA Sharabiani, Paul Aylin, and Alex Bottle. Systematic review of comorbidity indices for administrative data. Medical care, 50(12):1109–1118, 2012. 20, 68, 114

Karen Louise Shaw, Colette Clifford, Keri Thomas, and Helen Meehan. Review: improving end-of-life care: a critical review of the gold standards framework in primary care. Palliative Medicine, 24(3):317–329, 2010. 159

Paul Shelton, Mark A Sager, and Cheryl Schraeder. The community assessment risk screen (CARS): identifying elderly persons at risk for hospitalization or emergency department visit. The American journal of managed care, 6(8):925–33, 2000. 154

Mollie Shulan, Kelly Gao, and Crystal Dea Moore. Predicting 30-day all-cause hospital readmissions. Health care management science, 16(2):167–175, 2013. 147

Jaan Sidorov and Robert Shull. My patients are sicker: using the PRA risk survey for case finding and examining primary care site utilization patterns in a medicare-risk MCO. Am J Manag Care, 8(6):569–75, 2002. 143

Marc D Silverstein, Huanying Qin, S Quay Mercer, Jaclyn Fong, and Ziad Haydar. Risk factors for 30-day hospital readmission in patients ≥65 years of age. Proceedings (Baylor University. Medical Center), 21(4):363, 2008. 151

Judith D Singer and John B Willett. Applied longitudinal data analysis: Modeling change and event occurrence. Oxford university press, 2003. ISBN 9780195152968. 6, 27, 81, 105

Sidhartha Sinha, George Peach, Jan D Poloniecki, Matt M Thompson, and Peter J Holt. Studies using english administrative data (hospital episode statistics) to assess health-care outcomes - systematic review and recommendations for reporting. The European Journal of Public Health, 23(1):86–92, 2013. 33

Nicholas G Smedira, Katherine J Hoercher, Brian Lima, Maria M Mountis, Randall C Starling, Lucy Thuita, Darlene M Schmuhl, and Eugene H Blackstone. Unplanned hospital readmissions after heartmate ii implantation: frequency, risk factors, and impact on resource use and survival. JACC: Heart Failure, 1(1):31–39, 2013. 35

Sollis. Sollis: Population health analytics, 2016. URL http://www.sollis.co.uk/. [Retrieved 02.01.2016]. 143

Le Song, Mladen Kolar, and Eric P Xing. Time-varying dynamic bayesian networks. In NIPS, pages 1732–1740, 2009. 162

Xiaowei Song, Arnold Mitnitski, Jafna Cox, and Kenneth Rockwood. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. Stud Health Technol Inform, 107(Pt 1):736–40, 2004. 6, 63

Danielle A Southern, Hude Quan, and William A Ghali. Comparison of the elixhauser and charlson/deyo methods of comorbidity measurement in administrative data. Medical care, 42(4):355–360, 2004. 70

Roger Stedman. Scoring systems and outcomes. In Fang Gao Smith, editor, Core Topics in Critical Care Medicine. Cambridge University Press, 2010. 31

Ewout W Steyerberg, Andrew J Vickers, Nancy R Cook, Thomas Gerds, Mithat Gonen, Nancy Obuchowski, Michael J Pencina, and Michael W Kattan. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass.), 21(1):128, 2010. doi: 10.1097/EDE.0b013e3181c30fb2. 27, 29, 104

TE Strandberg, KH Pitkälä, and RS Tilvis. Frailty in older people. European geriatric medicine, 2(6):344–355, 2011. 1

Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC bioinformatics, 8(1):25, 2007. 40, 117

R. S. Sutton. Two problems with backpropagation and other steepest-descent learning procedures for networks. In Proc. 8th Annual Conf. Cognitive Science Society. Cognitive Science Society, 1986. 48

NRA Symons, K Moorthy, AM Almoudaris, A Bottle, P Aylin, CA Vincent, and OD Faiz. Mortality in high-risk emergency general surgical admissions. British Journal of Surgery, 100(10):1318–1325, 2013. 21

Tableau. Tableau, 2016. URL www.tableau.com. [Retrieved 02.09.2016]. 72

Vincent YF Tan, John Winn, Angela Simpson, and Adnan Custovic. Immune system modeling with Infer.NET. In eScience, 2008. eScience'08. IEEE Fourth International Conference on, pages 364–365. IEEE, 2008. 46

Kim Thomas, Mary VanOyen Force, Debbie Rasmussen, Dee Dodd, and Susan Whildin. Rapid response team challenges, solutions, benefits. Critical Care Nurse, 27(1):20–27, 2007. 159

Sebastian Thrun and Lorien Pratt. Learning to learn, chapter learning to learn: Introduction and overview. Thrun and Pratt [1997a], pages 3–13, 1997. 34

Y Tian, A Dixon, and H Gao. Emergency hospital admissions for ambulatory care-sensitive conditions: identifying the potential for reductions, Apr. 2012. URL http://www.kingsfund.org.uk. [Retrieved 02.09.2016]. 16

Nicholas Timmins. Never again? the story of the health and social care act 2012, may 2013. 3

Laura Tolo̧si and Thomas Lengauer. Classification with correlated features: unreliability of feature ranking and solutions. Bioinformatics, 27(14):1986–1994, 2011. 40, 117

Lior Turgeman and Jerrold H May. A mixed-ensemble model for hospital readmission. Artificial Intelligence in Medicine, 72:72–82, 2016. 35

UCLA. SAMIAM software solution, 2016. URL http://reasoning.cs.ucla.edu. [Retrieved 02.09.2016]. 165

MRC Biostatistics Unit. The BUGS project, 2016. URL http://www.mrc-bsu.cam.ac.uk. [Retrieved 02.09.2016]. 165

James Victor Uspensky. Introduction to mathematical probability. McGraw-Hill, 1937. 37

Carl van Walraven, Peter C Austin, Alison Jennings, Hude Quan, and Alan J Forster. A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. Medical care, pages 626–633, 2009. 20

Carl van Walraven, Irfan A Dhalla, Chaim Bell, Edward Etchells, Ian G Stiell, Kelly Zarnke, Peter C Austin, and Alan J Forster. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. Canadian Medical Association Journal, 182(6):551–557, 2010. 19, 40, 149

Carl van Walraven, Jenna Wong, and Alan J Forster. LACE+ index: extension of a validated index to predict early death or urgent readmission after hospital discharge using administrative data. Open Medicine, 6(3):e80, 2012. 19, 148

Vladimir Naumovich Vapnik and Vlamimir Vapnik. Statistical learning theory, volume 1. Wiley New York, 1998. 45

Antanas Verikas, Adas Gelzinis, and Marija Bacauskiene. Mining data with random forests: A survey and results of new tests. Pattern Recognition, 44(2):330–349, 2011. 87

Strother H Walker and David B Duncan. Estimation of the probability of an event as a function of several independent variables. Biometrika, 54(1-2):167–179, 1967. 37

RE Walpole, RH Myers, SL Myers, and K Ye. Probability and Statistics for Engineers and Scientists. Pearson, 2014. ISBN 978-0321629111. 26, 105

Jeremy Walston, Evan C Hadley, Luigi Ferrucci, Jack M Guralnik, Anne B Newman, Stephanie A Studenski, William B Ershler, Tamara Harris, and Linda P Fried. Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the american geriatrics society/national institute on aging research conference on frailty in older adults. Journal of the American Geriatrics Society, 54(6):991–1001, 2006. 15

CY Wang, Laura-Mae Baldwin, Barry G Saver, Sharon A Dobie, Pamela K Green, Yong Cai, and Carrie N Klabunde. The contribution of longitudinal comorbidity measurements to survival analysis. Medical care, 47(7):813, 2009a. 114

Peiming Wang. A bivariate zero-inflated negative binomial regression model for count data with excess zeros. Economics Letters, 78(3):373–378, 2003. 27

Zhaowen Wang, Ercan E Kuruoglu, Xiaokang Yang, Yi Xu, and Thomas S Huang. Time varying dynamic bayesian network for nonstationary events modeling and online inference. Signal Processing, IEEE Transactions on, 59(4):1553–1568, 2011. 162, 163

Zhu Wang, Wayne A Woodward, and Henry L Gray. The application of the kalman filter to nonstationary time series through time deformation. Journal of Time Series Analysis, 30(5):559–574, 2009b. 161

Andrew R Webb. Statistical pattern recognition. John Wiley & Sons, 2003. ISBN 9780470854778. doi: 10.1002/0470854774. 29

Rick Weber, Akila Gothandaraman, Robert J Hinde, and Gregory D Peterson. Comparing hardware accelerators in scientific applications: A case study. IEEE Transactions on Parallel and Distributed Systems, 22(1):58–68, 2011. 34

J.P. Weiner and C Abrams. The johns hopkins ACG system - technical reference guide - version 10, Dec. 2011. URL http://acg.jhsph.org. [Retrieved 02.09.2016]. 4, 5, 7, 67, 113

D Wennberg, M Siegel, B Darin, N Filipova, R Russell, L Kenney, K Steinort, T Park, G Cakmakci, J Dixon, N Curry, and J Billings. Combined predictive model: Final report and technical documentation, Dec. 2006. URL http://tna.europarchive.org. [Retrieved 02.09.2016]. 38, 151

Hadley Wickham. ggplot2: elegant graphics for data analysis. Springer, 2016. 72

Jennifer L Wiler, Richard T Griffey, and Tava Olsen. Review of modeling approaches for emergency department patient flow and crowding research. Academic Emergency Medicine, 18(12):1371–1379, 2011. 155

David H Wolpert. Stacked generalization. Neural networks, 5(2):241–259, 1992. 36

Robert S Woodworth and EL Thorndike. The influence of improvement in one mental function upon the efficiency of other functions.(i). Psychological review, 8(3):247, 1901. 34

Haifeng Xie, Thierry J Chaussalet, and Peter H Millard. A continuous time markov model for the length of stay of elderly people in institutional long-term care. Journal of the Royal Statistical Society: Series A (Statistics in Society), 168(1):51–61, 2005. 156

Eric P Xing, Michael I Jordan, Richard M Karp, et al. Feature selection for high-dimensional genomic microarray data. In ICML, volume 1, pages 601–608. Citeseer, 2001. 86

Stanley Xu, Richard H Jones, and Gary K Grunwald. Analysis of longitudinal count data with serial correlation. Biometrical journal, 49(3):416–428, 2007. 161, 163

Ying Yang, Geoffrey I Webb, and Xindong Wu. Discretization methods. In Data mining and knowledge discovery handbook, pages 113–130. Springer, 2005a. doi: 10.1007/0-387-25465-X_6. 105

Ying Yang, Geoffrey I Webb, and Xindong Wu. Discretization methods. In Data mining and knowledge discovery handbook, pages 113–130. Springer, 2005b. 7, 26

In-Kwon Yeo and Richard A Johnson. A new family of power transformations to improve normality or symmetry. Biometrika, pages 954–959, 2000. 85

Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In Advances in neural information processing systems, pages 3320–3328, 2014. 34

Shipeng Yu, Faisal Farooq, Alexander Van Esbroeck, Glenn Fung, Vikram Anand, and Balaji Krishnapuram. Predicting readmission risk with institution-specific prediction models. Artificial Intelligence in Medicine, 65(2):89–96, 2015. 42

Matthew D Zeiler. Adadelta: an adaptive learning rate method. arXiv preprint arXiv:1212.5701, 2012. 49

Bichen Zheng, Jinghe Zhang, Sang Won Yoon, Sarah S Lam, Mohammad Khasawneh, and Srikanth Poranki. Predictive modeling of hospital readmissions using meta-heuristics and data mining. Expert Systems with Applications, 42(20):7110–7120, 2015. 40

Shuheng Zhou, John Lafferty, and Larry Wasserman. Time varying undirected graphs. arXiv preprint arXiv:0802.2758, 2008. 162

Zhi-Hua Zhou. Ensemble methods: foundations and algorithms. CRC press, 2012. 35, 36, 39, 92