WestminsterResearch

http://www.wmin.ac.uk/westminsterresearch

Determining readmission time window using mixture of generalised Erlang distribution.

Eren Demir
Thierry Chaussalet
Haifeng Xie

School of Informatics

Determining readmission time window using mixture of generalised Erlang distribution

Eren Demir, Thierry Chaussalet, Haifeng Xie

Health and Social Care Modelling Group
Dept. of Information Systems and Computing, School of Informatics
University of Westminster, London, UK
{demirer, chausst, xieh}@wmin.ac.uk

Abstract

The absence of a unified definition of readmissions has motivated the development of a modelling approach, to systematically tackle the issue surrounding the appropriate choice of a time window which defines readmission. The population of discharged patients can be broadly divided in two groups - a group at high risk of readmission and a group at low risk. This approach extends previous work by the authors, without restricting the number of stages, that patients may experience in the community. Using the national data (UK), we demonstrate its usefulness in the case of chronic obstructive pulmonary disease (COPD) which is known to be one of the leading causes of readmission. We further investigate variability in the definition of readmission among 10 strategic health authorities (SHAs) in England and observe that there are differences in the estimated time window across SHAs. The novelty of this modelling approach is the ability of capturing time to readmission that exhibit a non-zero mode and to estimate an appropriate time window based on evidence objectively derived from operational data.

1. Introduction

Generally, high level of readmission is associated with poor patient care, hence, its relation to the quality of care is plausible; readmission data are easily obtained from hospital databases; and readmissions appear more frequently than other adverse outcomes, such as mortality. One of the difficulties for using readmission as a measure of quality is the absence of a unified definition of the indicator. Currently the NHS performance rating framework defines readmission as an emergency or unplanned admission to the same hospital within 28 days following discharge. Sibbritt [1] attempted validation of the use of 28 days by constructing a graphical output for the total number of readmissions in the four categories of medical, surgical, paediatrics, obstetrics and gynaecology. Each graph shows log-normal and exponential shaped distributions, with approximately 32% of readmissions occurring 28 days after discharge from the index readmission. A similar pattern was found by Chambers *et al* [2], where readmissions show an early peak (0-6 days) which then level off by 28 days. However in both cases, the justification for the choice of 28 days relied solely on visual inspection of the graphical output, therefore, it could result in an inaccurate estimation. The lack of insight within the literature into the time window of readmissions, have motivated the development of a modelling approach which systematically tackles the issue surrounding the appropriate choice of a time window to define readmission. In this paper, we extend a modelling

Table 1. Number of admissions for the 10 SHA in England for COPD patients for the calendar years 1998 to 2003. The numbers in the parentheses are percentage of readmission (as defined using 28 days).

SHA name	1998	1999	2000	2001	2002	2003
East Midlands	3938	5833	6972	7307	7567	7572
	(9.98)	(10.95)	(11.77)	(11.99)	(14.11)	(12.01)
East of England	5755	6563	7480	8067	8252	9364
	(8.84)	(8.62)	(10.05)	(9.88)	(11.32)	(11.31)
London	8025	11206	11099	11282	11665	12597
	(9.59)	(10.20)	(11.51)	(11.84)	(11.87)	(11.80)
North East	3715	5766	6027	6666	7673	9063
	(10.36)	(11.84)	(12.08)	(14.15)	(13.40)	(14.71)
North West	9498	10954	11821	13632	16110	19278
	(9.66)	(11.17)	(11.77)	(12.19)	(12.17)	(12.32)
South Central	3714	4271	4187	4233	5568	6984
	(8.94)	(8.50)	(10.10)	(11.62)	(12.21)	(11.55)
South East Coast	2223	3503	4318	4635	5534	6454
	(8.59)	(8.65)	(10.24)	(11.28)	(10.79)	(10.97)
South West	4608	6004	6176	6260	7558	8666
	(7.70)	(10.46)	(10.49)	(10.65)	(10.97)	(11.99)
West Midlands	7418	8434	8187	8572	9413	10589
	(8.45)	(9.56)	(10.98)	(10.30)	(11.48)	(11.17)
Yorkshire and The Humber	4934	9821	11173	12489	12712	13393
	(12.59)	(12.47)	(13.17)	(14.80)	(14.21)	(13.84)

approach reported in [3], where authors expressed the random variable time to readmission as a mixture of two exponential density functions. This can be restrictive when time to readmission exhibits a non-zero mode as can be observed for several chronic diseases such as chronic obstructive pulmonary disease (COPD).

The paper is organised into the following sections: a brief description of data in Section 2; the modelling method is presented in Section 3; Section 4 demonstrates the usefulness of our proposed methodology in estimating the appropriate width of a time window, in the case of COPD and illustrate regional variation in the definition of readmission; discussion and comments on future works are in Section 5.

2. Data

The Department of Health in the UK releases annually the national database – Hospital Episode Statistics (HES). The HES dataset captures all the consultant episodes of a patient during their stay in a hospital in the UK. During a hospital stay (or called spell), a patient might encounter several successive episodes. The data period of investigation is 04/1997 - 03/2004 where each financial year starts on the 1st of April. We focus our study on chronic obstructive pulmonary disease (COPD) which is known to be one of the leading causes of readmission [4]. Spells ending with death are excluded as no further admissions is possible. To build a longitudinal dataset, all patient records were extracted using the primary diagnosis codes corresponding to COPD (ICD-10 codes J40-J44),
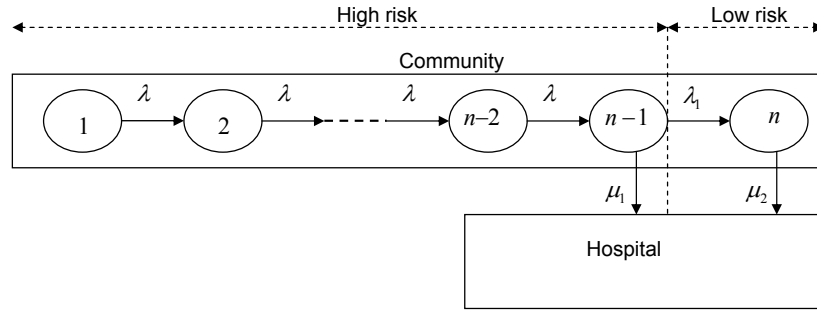
2

Figure 1. Stages that patients may experience in the community before readmission to hospital.

and joined for across financial years. We extracted 962,656 episodes and 696,911 spells. Using the time window of 28 days as currently defined by the Department of Health, the percentage of readmission within 28 day interval in England for calender years 1998 to 2003 are 28.0%, 24.2%, 23.1%, 22.3%, 22.2%, 22.0% respectively. A relatively stable level of readmission is observed for the period 2001 to 2003. The NHS in England is managed by 10 SHAs, therefore, we partition COPD data into 10 SHAs to investigate variability in the definition of readmission among regions. Table 1 shows for each SHA in England, the number of admissions and percentage of readmissions at 28 days. Regional variations in readmission rates is noticeable. Across the calendar years, Yorkshire and The Humber and North East has higher percentage of readmissions in comparison to other SHAs.

3. Modelling approach

Following discharge, patients first go through a phase with "high risk" of readmission, where they are likely to be readmitted quickly, possibly because of premature discharge; if not readmitted at the end of this phase, they enter another phase with "low risk" of readmission and stay longer in the community. In a previous paper [3], the authors assumed that the time spent in the "high risk" phase was exponentially distributed. This is no longer appropriate when the time to readmission exhibits a non-zero mode as can be observed for several chronic diseases such as COPD. To tackle this problem, we model the "high risk" time to readmission using a generalised Erlang distribution [5]. The proposed conceptual model for the movement of patients in the community is depicted in Figure 1. In the "high risk" phase, patients go through $n-2$ sequential stages at a rate λ (hence the total time spent through these stages follows an Erlang distribution), followed by stage $n-1$ (spending an exponentially time). From stage $n-1$ they either go back to hospital at rate μ_1 or, enter at rate λ_1 the "low risk" phase (stage n), which they leave at rate μ_2. This model can also be viewed as a mixture distribution model, with probability density function (pdf) of time to readmission written as

$$f(x) = pf_{n-1}(x) + (1-p)f_n(x), \tag{1}$$

where $p = \mu_1/(\mu_1 + \lambda_1)$ is the probability of being in the high risk group, i.e. the group of those readmitted to hospital at the end of the phase with "high risk" of readmission, and $f_{n-1}(x)$ and $f_n(x)$ are the pdf of time to readmission for each group ("high risk" and "low risk"). This approach divides the readmitted patient population in two groups (ϕ_1 and ϕ_2) in a probabilistic way, that
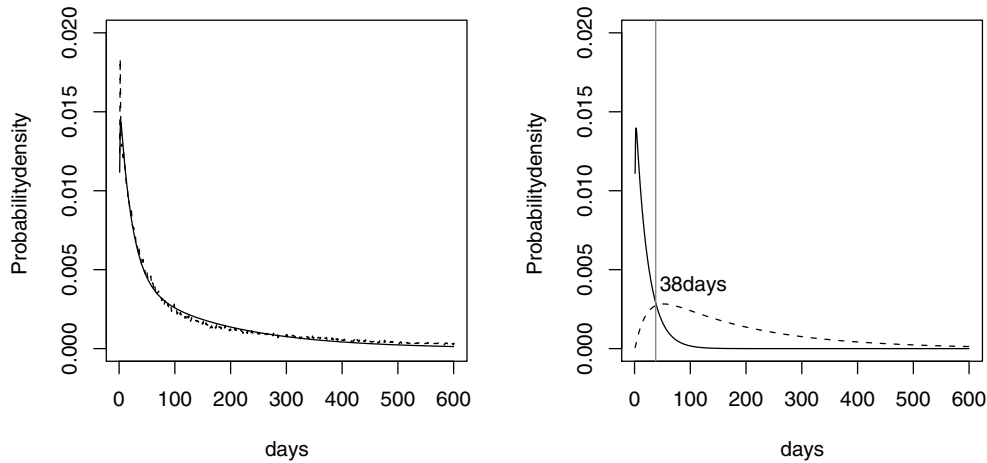
Figure 2. Empirical (solid line) and fitted (dotted line) density function of time to readmission for the COPD data (left). The two curves represent (right) $f_{n-1}(x)$ (high risk pdf) and $f_n(x)$ (low risk pdf). The vertical line indicates the point of intersection.

is with membership determined by posterior probabilities derived via Bayes'a theorem as $p(\phi_1|x)$ $= pf_{n-1}(x)/f(x)$ and $p(\phi_2|x) = (1-p)f_n(x)/f(x)$. Therefore using a Bayesian classification argument, one can show that the optimal time window of readmission is determined by solving $p(\phi_1|x) = p(\phi_2|x)$ or equivalently given by the time value where $pf_{n-1}(x) = (1-p)f_n(x)$ that is, where the two corresponding curves intersect.

4. Application of the Model

The modelling approach outlined in the previous section is applied to the COPD dataset. A sequential procedure was adopted whereby increasing the number of stages n were tried, starting with $n = 1$ (corresponding to exponential distribution), until a fit was obtained that satisfactorily described patients experience in the community before readmission. We tested models with up to $n = 4$ stages. A program is developed, to fit the model to the overall COPD data, by using a general method of maximising the log-likelihood function using the open source program R [6]. Models with three stages were consistently shown to provide the "best" fit, judging by Akaike information criterion (AIC) [7] and Bayesian information criterion (BIC) [8], which are measures of goodness-of-fit taking into account the complexity of the model. Using the whole dataset, we estimated the rates of the model depicted in Figure 1, where $n = 3$, as $\lambda = 1.271739$, $\lambda_1 = 0.029829$, $\mu_1 = 0.015853$ and $\mu_2 = 0.005859$. Thus the pdf of time to readmission for COPD patients are estimated to be $f(x) = 0.011892e^{-0.045682x} - 0.016300e^{-1.271739x} + 0.004409e^{-0.005858x}$. Figure 2 (left) shows the empirical and the fitted pdf of time to readmission for the COPD dataset. The close agreement between these two curves suggests that the model with three stages, is able to capture non-zero mode and the overall pattern of time to readmission for COPD patients. The estimated optimal time window is 38 days. This is graphically illustrated in Figure 2 (right) where the two curves represent high and low risk pdf, and the vertical line indicating the point of intersection at about 38 days. The probability of belonging to the high risk (of readmission) group is estimated to be at about 35% for a COPD patient who is just discharged from hospital. Probability of readmission within x days can be estimated by $\int_0^x f(u)\,du$, where $f(u)$ is the pdf as

Table 2. Optimal time window for the 10 SHAs in England and probability of being in high risk group.

SHA name	Optimal time window (days)	Probability of belonging to the high risk group
East Midlands	41	0.36
East of England	36	0.33
London	43	0.38
North East	43	0.37
North West	37	0.33
South Central	36	0.35
South East Coast	40	0.36
South West	31	0.30
West Midlands	36	0.34
Yorkshire and The Humber	47	0.39

in (1). Therefore the estimated probability of readmission for COPD using 38 days is 35%.

We also fitted the model for each of the 10 SHA in England. Testing models up to $n = 4$ stages, models with three stages were again found to be consistent for all 10 SHAs, judging by AIC and BIC. Table 2 summarises the estimated time windows, together with their corresponding probability of being in the high risk group. There are some variability in the estimated optimal time window among SHAs. Yorkshire and The Humber, which has the highest level of readmission as shown in Table 1, has the highest probability of being in the high risk group following discharge. Our estimate suggests that the optimal time window for Yorkshire and The Humber region, in the case of COPD is significantly greater than the currently used 28 days by the Department of Health. East of England, South Central, South West and West Midlands SHAs provides some support for the use of 28 days. We may speculate that one possible cause of variation in readmission rate could be deprivation differences among the regions. Based on a report published by the office of the deputy prime minister [9], Yorkshire region is recognised as one of the most deprived and Southern England is within the least deprived region. A recent study in the Greater Manchester area (UK) [10] showed that deprivation indeed exerted a significant effect on the risk of emergency readmission.

5. Conclusion

The model that was presented by the authors in [3] assumed that time to readmission can be expressed by a mixture of exponential distributions. In most cases patterns of readmission exhibits a non-zero mode or exponential shaped distributions and mixture of exponential distributions lacks to capture non-zero mode. Having recognised this limitation, we extended the approach without restricting the number of stages that patients may experience in the community before readmission to hospital. This generalised Erlang model, has the capability of capturing, time to readmission that exhibits a non-zero mode. Using the national Hospital Episode Statistics dataset, we showed that in the case of COPD, there are potential problems in defining readmission. Our estimated time window using the complete COPD dataset is 38 days, which is third greater than the current 28 days used by Department of Health. This provides some support for the use of 28 days. Further examination of the estimated time windows for each of 10 SHA in England revealed that there are some differences between SHAs. Estimated time windows for northern England SHAs are slightly higher than the

IEEE
COMPUTER
SOCIETY

south. This variability may indicate that the use of a single number in defining readmission may not be an appropriate choice of time window. However, more research is needed to understand readmission, especially, the variability between SHAs and hospitals. Particularly differences in case-mix i.e. deprivation and mix of local populations may contribute to the variation of the estimated time windows for different diagnosis and regions. The novelty of this approach is the ability to capture various patterns of readmission (exponential and non-zero mode distributions) and the estimation of an appropriate time window based on operational data. The estimated probability of being in the high risk group could also be used to monitor quality of healthcare.

References

[1] D. W. Sibbritt, "Validation of a 28 day interval between discharge and readmission for emergency readmission rates," *Journal of Quality in Clinical Practice*, vol. 15, pp. 211–220, 1995.

[2] M. Chambers and A. Clarke, "Measuring readmission rates," *British Medical Journal*, vol. 301, pp. 1134–1136, 1990.

[3] E. Demir, T. J. Chaussalet, H. Xie and P. H. Millard, "A Method for Determining an Emergency Readmission Time Window for Better Patient Management," in *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems*, pp. 789–793, 2006.

[4] M. Roland, M. Dusheiko, H. Gravelle and S. Parker, " Follow up of people aged 65 and over with a history of emergency admissions: Analysis of routine admission data", *British Medical Journal*, vol. 330, pp. 289-292, 2005.

[5] S. M. Ross, *Introduction to Probability Models*, Academic Press, 2003.

[6] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2005.

[7] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.

[8] G. Schwarz, "Estimating the dimension of a model," *Annal of Statistics*, vol. 6, pp. 461–464, 1978.

[9] Office of the Deputy Prime Minister, "The English indices of deprivation 2004", report, 2004.

[10] G. Lyratzopoulos, D. Havely, I. Gemmell and A. G. Cook, "Factors influencing emergency medical readmission risk in a UK district general hospital: A prospective study," *BioMed Central Emergency Medicine*, vol. 5, pp. 1–9, 2005.