A multicentre integration of a computer□led follow□up of prostate cancer is valid and safe

Salem, H., Caddeo, G., McFarlane, J., Patel, K., Cochrane, L., Soria, D., Henley, M. and Lund, J.

Title:

A multicentre integration of a computer led follow up in surgical oncology is valid and safe

Authors:

Hesham Salem, Giacomo Caddeo, Jon McFarlane, Kunjan Patel, Lynda Cochrane, Daniele Soria, Mike Henley, Jonathan Lund

Abstract

Background

Prostate cancer (CaP) has a rising number of patients requiring routine follow up. In this study, we aimed to test a computer led follow up service for prostate cancer in two UK hospitals. The testing aimed to validate the computer Expert system in making clinical decisions according to the individual patient's clinical need. The valid model should accurately identify patients with disease recurrence or treatment failure based on their blood test and clinical picture.

Methods

A clinical decision support system (CDSS) was developed from European (EAU) and national (NICE) guidelines along with knowledge acquired from Urologists. This model was then applied in two UK hospitals to review patients post CaP treatment. These patients' data (n= 200) were then reviewed by two independent Urology consultants (blinded from the CDSS and other consultant's rating) and the agreement was calculated by kappa statistics for validation. The second objective aimed to verify the system by estimating the system reliability.

Results

The two individual urology consultants identified 12 % & 15% of the patients to have potential disease progression and recommended their referral to the Urology care. The kappa coefficient for the agreement between the CDSS and the 2 consultants was 0.81 ($p < 0.001$) and 0.84 ($p < 0.001$). The agreement among both specialist was also high with $k = 0.83$ ($p < 0.001$). The system reliability was estimated on all cases and this demonstrated 100% repeatability of the decisions.

Conclusion

The computer led follow up is a valid model for providing safe follow up for prostate cancer.

1.1 Introduction

Prostate cancer is the most common malignant disease in men and the third leading cause of cancer related mortality in the United Kingdom in 2014 (1). It is a disease of the elderly, with men above the age of 75 at higher risk of disease related mortality (2). It has a heterogeneous course of progression from slow growing and potentially insignificant, to aggressive with serious impact on the patient's health and quality of life. Despite the disease rising prevalence, there has been a significant improvement in survival rates (3, 4) which may be accounted for by the availability of more accurate diagnostic and treatment modalities. However, where best to safely follow up patients remains unresolved. The current National Institute of for Health and Care Excellence (NICE) recommendation is to provide this follow up in primary care where appropriate (5). However, this recommendation was met with concerns from both GPs and Urologists because of the lack of expertise in the community (6). The National Prostate Cancer Audit 2014 annual report identified five different models for prostate cancer follow up: Consultant-led clinic, Cancer Nurse Specialist (CNS)-led clinic, Telephone clinic, Community-based specialist follow up and Radiographer-led clinic (only for radiotherapy pathway) (7). The community based model played only a minor role in follow up in this audit.

In various industries and expert simulating systems have provided an alternative cheap, reliable and available solution where expertise is lacking (8). The bottleneck to their development is the knowledge acquisition phase, which usually is exhaustive and time consuming (9). Furthermore, the system has to go through rigorous testing; system validation (are we building

the right system?) and system verification (are we building the system correctly. Previous attempts to apply this to an industry setting have suffered from a lack of consistency and formality in the model (10). The same problems were also evident in medical expert system development and has challenged their uptake in health care (11).

In this study we aimed to verify and validate an expert system simulating a Urologist in the follow up of stable prostate cancer and its application in supporting GPs in providing a follow up service in the community. Therefore, we tested the null hypothesis that the agreement between the expert system and the human domain experts is slight to fair ($k0 = 0.4$) against the alternative that it is better than moderate ($k1 > 0.61$). For verification, the expert system has to be 100 percent reliable i.e. the same clinical input triggered the system to produce the exact same outputs on each occasion.

1.2 Methods

1.2.1 Description of the prostate cancer system

The Clinical Decision Support System (CDSS) is a web based solution that can be used with current versions of browsers including Microsoft IE, Google Chrome, Mozilla and Safari. At the core of the system is a Rule Engine which processes all of the data captured and held for a patient and presents it to a complex disease domain specific algorithm compiled and tested in accordance with NICE guidelines (5). The system uses principally Microsoft technologies including MS SQL Server database, the Microsoft.NET development framework and the C# programming language along with other web technologies like JavaScript and AJAX.

1.2.2 System security and functional analysis

The expert system is based on a secure remote server held within the NHS IT network (N3). This location makes it accessible to all NHS primary and secondary care sites and NHS

partners without the need for individual copies at each site, providing one single point of access. Only authorised personnel can gain access to this system and all the stored data are encrypted in line with the standards of the Information Governance Statement of Compliance (IG SoC).

The system is triggered by an authorised user query about a patient visit. The CDSS access their records on the Hospital Information System, including their laboratory test results, before asking the user to fill a problem specific online form. The form includes routine questions enquiring about relevant symptoms as outlined in the NICE 2014 guidelines to support the system accurate decision making (5). The system output is in the form of an electronic clinic letter suggesting a tailored treatment plan for the individual patient consultation.

1.2.3 **Study configuration**

The CDSS is already part of an existing Clinical Nurse Specialist (CNS) led CaP follow up service in Derby Hospital NHS Foundation trust and University Hospitals Coventry and Warwickshire NHS Trust. The main system function in this service is acting as an intelligent database and clinic letter generator (figure 1 and figure 2). Patients included were those who had their disease diagnosed and treated by the urology cancer multidisciplinary team before being discharged for follow up to the CNS clinic. There were no specific criteria for follow up in this clinic and the clinic predated the software.

When seeing a patient for follow up the users (CNS in this study) supply the CDSS with recent symptoms and blood test results, and the system produces an electronic letter stating significant clinical details and a suggested treatment plan. The treatment plan was reviewed and overridden by the CNS if appropriate to do so.

In this study, we only aimed to evaluated the system validity, thus the letters were anonymised and the nurses' plans were eliminated before examining the cases against the inclusion criteria. Each letter generated by the CDSS had five main components (figure 1 & figure 2):

1. History of disease and treatment.
2. Current symptoms and blood test results.
3. System analysis of clinical data.
4. System suggested outcome.
5. Nurse note's text box for supplying any extra information and overriding the system decision where necessary.

The eligible cases were anonymised by eliminating all patients' identifiers and the CNS free text entry of their own clinical assessment and plan (component 5 in figure 2)

Cases were examined by the study clinical monitor against the inclusion criteria and cases were excluded from the study with justification (as not enough clinical data, newly diagnosed or treated cases i.e. not CaP etc.).

1.2.3.1 Eligibility criteria

All adult patients seen in this clinic with known prostate cancer whom had their primary treatment assigned and delivered.

1.2.3.2 Inclusion criteria

Adult patients with known CaP and presenting for routine follow up in clinic with (essential requirement):

- Known prostate cancer treatment pathway.

- Details on his Prostate specific antigen serum levels (PSA) on presentation and most recent results.

With or without (optional requirement):

- Disease or treatment related symptoms (lower urinary tract, bone pain, weight loss and erectile dysfunction)
- Abnormal blood test (haemoglobin, calcium, urea, electrolytes, creatinine, and liver function tests)

1.2.3.3 Exclusion criteria

- Cases with newly diagnosed prostate cancer that are waiting for decision to treat.
- Cases with known prostate cancer that are being evaluated for known recent disease progression.
- Cases with no clinical details of their initial cancer presentation such as Gleason score, PSA, and tumour grade which would influence the follow up decision making.

In order to reach both end points, the study divided into two branches after case selection (figure 3). In the primary end point pathway, the included cases were reviewed by two independent clinical investigators after eliminating the system analysis and suggestions (Components 3 and 4 samples in figure 1). Each investigator independently assigned an outcome for each case (figure 4) according to the disease history, current results and symptoms blinded from the system recommendation.

In the second part of the study, the system reliability and precision was estimated by codifying all cases clinical variables and output. The cases specific codes were all tabulated on a spread

sheet (Excel, Microsoft Corp, Seattle) and all cases with similar input(s) were expected to have the same output code. In case of discrepancy, errors were identified and the system was then corrected. Any system changes were followed by a retest by a new sample of cases to re-estimate reliability until 100% precision was obtained (secondary end point).

1.2.4 **Statistical evaluation**

The system validation was estimated by testing the null hypothesis that the agreement between the CDSS and the human domain experts was slight to fair (k0 = 0.4) against the alternative that it was better than moderate (k1 > 0.61) (12). Kappa was estimated between each investigator (INV1 & INV2) and then the CDSS between both investigators.

1.3 Results

The study included data of 200 patients seen in either hospitals (100 each). All patients had CaP and mean and median age was 75 (range 51 to 94) with a mean and mode Gleason score of 7. The risk stratifications according to D'Amico's classification (13) of their disease on diagnosis demonstrated 96 (48%) patients in the high risk group and 78 (39%) & 26 (13%) in the intermediate and low risk group respectively (figure 5). Radiotherapy had been used to treat 128 cases out of the total (figure 6).

Kappa statistics were estimated to test the hypothesis. The unweighted kappa for INV1 INV2 was 0.8, and with quadratic weighting the kappa remained, demonstrating substantial agreement between CDSS and investigator at 0.86 and 0.96 respectively for INV1 and INV2 respectively. This refuted the null hypothesis as k0 > 0.4 and alternative hypothesis k1 > 0.6 was accepted with the conclusion that the CDSS is validation and its assigned outcomes are acceptable by the domain experts (tables 1, 2 & 3).

Both investigators identified 31 cases with possible disease progression or recurrence (INV1 = 25, INV2 = 29). Most of those cases were biochemical failure (BCF) post radiotherapy treatment (21 out of the total of 31) based on either ASTRO (14-16) or PHOENIX (17) criteria or both. Six cases were on watchful waiting had either high or rising PSA and were identified as disease progression (those patients did not receive any treatment). Two patients had radical surgery with detectable PSA and one was treated with hormone ablation, with significant PSA doubling time. CDSS identified all of the above cases as disease progression or recurrence except one case who was only classified by INV1 as potential recurrence post radiotherapy. They had only two consecutive rises in their PSA, thus not identified as BCF by CDSS or INV2.

The reliability testing was estimated by examining the outcome of cases with similar input codes. This demonstrated 100% reliability and so the study secondary objective was met.

1.4 Discussion

This multicentre study validated and verified a novel CDSS led follow up service that can be used in a prostate cancer follow up setting, which could be based in either primary or secondary care. This is the first study to combine quantitative methodologies to perform full system validation and verification on real patients. This critical test has been lacking in the development of medical Expert Systems, and this has led to poor uptake health care because of lack of confidence in these systems (11). Standalone software as CDSS are recognised by European regulatory bodies as medical devices and they require registration via strict criteria (18). These criteria are outlined in the declarations of the International Organization for Standardization (ISO) 13485 (quality management systems), 14971 (risk management) and 62304 (software development cycle). The latter addresses the standards for testing that should met in the validation. In this study, we quantified the validity by estimating the kappa statistics

in a 2 centre live clinical evaluation. Furthermore, the objective reliability testing verified the system and completed the development cycle. These quantified tests support the utility of the Expert System as it has met state of the art verification and validation methodologies (18).

It has been estimated that CaP costs more than €8.43 billion across Europe (£0.8 billion in the UK) with most spent on treatment in the first year after diagnosis (19). Prostate cancer follow up is also expensive overall and cost effective service improvement is much needed. This cost is expected to increase with improvement in disease specific survival and increase in the population life expectancy. The CDSS model may be able to safely move follow up in to primary care with potentially significant cost saving (6). The system developers believe that the valid system can be used by any healthcare worker regardless their Urological background with a potential of adopting an interface for direct patient interaction. This could have significant benefits and cost savings but given the age group of the patients and lack of flexibility of NHS health care, the later would be challenging to implement.

This type of study is limited by the quality of the clinical data and experience of the human experts. The data in our study were all real patients' data collected in real time clinics from two large tertiary and secondary care centres. Furthermore, uncommon and rare presentations have been validated in the knowledge validation study. Cases studied were allocated to the CNS follow up clinic by the MDT and may explain the low recurrence rate in this study even though 48% of the cases were in the high risk group.

The clinical investigators were both qualified consultants and they received independent specialist training. One had special interest in CaP community follow up and had previously developed a follow up model for a remote area. The other investigator had an interest in core Urology and stone disease. Both were not familiar with the software development and had no conflict of interest. The rating was performed blinded from the other investigator and the

CDSS, thus bias has been eliminated and the agreement calculated by the kappa coefficient should be the true agreement.

In medicine, the validation of an intervention usually requires a comparison against the gold standard to estimate the sensitivity and specificity. This is different to expert system validation where it is based on the Turing test concept in comparing the machine's cognitive performance to human domain experts (10, 20). Kappa statistics is more appropriate for this type of validation as they estimate the agreement between human and machine cognitive function and has been applied by other studies to validate expert systems (21, 22). The sensitivity and specificity is only accurate if compared the gold standard as histopathology results and this is more widely applied by artificial neural networks as they learn directly from data and the role of human opinion is limited (9).

The long term outcome for those cases was not examined as part of this study. The validation aimed to estimate the validity of the system reasoning against consultant urologists. The long term outcome can be useful to estimate the accuracy of both investigators and the CDSS, however this would require long term follow up and more invasive tests.

1.5 Conclusion

The CDSS demonstrated high validity and accuracy in its decision making. This tool has potential for safe use in supporting follow up of surgical oncology in the primary and secondary care. Mixed methodology approach is required to perform the mandatory system validation and verification.

1.6 References

1. CancerResearchUK. Prostate cancer mortality 2016. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-One.
2. Droz J-P, Balducci L, Bolla M, Emberton M, Fitzpatrick JM, Joniau S, et al. Management of prostate cancer in older men: recommendations of a working group of the International Society of Geriatric Oncology. Bju International. 2010 01/12/accepted;106(4):462-9. PubMed PMID: PMC3258484.
3. CancerResearchUK. Statistics and outlook for prostate cancer. http://www.cancerresearchuk.org/about-cancer/type/prostate-cancer/treatment/statistics-and-outlook-for-prostate-cancer: Cancer Research UK Web site; 2014.
4. SEER. SEER stat fact sheets: prostate cancer. National Cancer Institute Web site: National Cancer Institute; 2014.
5. NICE. Prostate cancer: diagnosis and treatment CG175. National Institute for Health and Care Excellence; 2014.
6. Watson EK, O'Brien R, Campbell C, Weller D, Neal RD, Wilkinson C, et al. Views of health professionals on the role of primary care in the follow-up of men with prostate cancer. Family Practice. 2011 Dec;28(6):647-54. PubMed PMID: WOS:000297241800009.
7. NCPA. NCPA 2014 Annual Report. http://www.npca.org.uk/reports/: National Prostate Cancer Audit; 2014.
8. Liao SH. Expert system methodologies and applications - a decade review from 1995 to 2004. Expert Systems with Applications. 2005 Jan;28(1):93-103. PubMed PMID: WOS:000225261500009. English.
9. Pandey B, Mishra RB. Knowledge and intelligent computing system in medicine. Computers in Biology and Medicine. 2009 Mar;39(3):215-30. PubMed PMID: WOS:000264521400003. English.
10. Okeefe RM, Oleary DE. EXPERT SYSTEM VERIFICATION AND VALIDATION - A SURVEY AND TUTORIAL. Artificial Intelligence Review. 1993 Feb;7(1):3-42. PubMed PMID: WOS:A1993KQ54000001.
11. Ammenwerth E, Nykanen P, Rigby M, de Keizer N. Clinical decision support systems: Need for evidence, need for evaluation. Artificial Intelligence in Medicine. 2013 Sep;59(1):1-3. PubMed PMID: WOS:000325194700001.
12. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960 04/01/;20(1):37-46.
13. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA. 1998 Sep 16;280(11):969-74. PubMed PMID: 9749478. Epub 1998/09/28. eng.
14. Bellera CA, Hanley JA, Joseph L, Albertsen PC. A STATISTICAL EVALUATION OF RULES FOR BIOCHEMICAL FAILURE AFTER RADIOTHERAPY IN MEN TREATED FOR PROSTATE CANCER. International Journal of Radiation Oncology Biology Physics. 2009 Dec;75(5):1357-63. PubMed PMID: WOS:000272341800012.
15. Kuban DA, Levy LB, Potters L, Beyer DC, Blasko JC, Moran BJ, et al. Comparison of biochemical failure definitions for permanent prostate brachytherapy. International Journal of Radiation Oncology Biology Physics. 2006 Aug;65(5):1487-93. PubMed PMID: WOS:000239442100028.
16. Horwitz EM, Thames HD, Kuban DA, Levy LB, Kupelian PA, Martinez AA, et al. Definitions of biochemical failure that best predict clinical failure in patients with prostate

cancer treated with external beam radiation alone: A multi-institutional pooled analysis. Journal of Urology. 2005 Mar;173(3):797-802. PubMed PMID: WOS:000227075500033.
17. ASTRO. Consensus statement: guidelines for PSA following radiation therapy. American Society for Therapeutic Radiology and Oncology Consensus Panel. International journal of radiation oncology, biology, physics. 1997 1997 Mar;37(5):1035-41. PubMed PMID: MEDLINE:9169810.
18. 93/42/EEC. DIRECTIVE 2007/47/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 September 2007. The European, Parliament

The Council of the

European Union; 2007.
19. Luengo-Fernandez R, Leal J, Gray A, Sullivan R. Economic burden of cancer across the European Union: a population-based cost analysis. The Lancet Oncology. 2013 11//;14(12):1165-74.
20. Turing AM. Computing machinery and intelligence. Mind. 1950;59(236):433-60.
21. Lopes M, Ortega NRS, Silveira PSP, Massad E, Higa R, Marin HD. Fuzzy cognitive map in differential diagnosis of alterations in urinary elimination: A nursing approach. International Journal of Medical Informatics. 2013 Mar;82(3):201-8. PubMed PMID: WOS:000315799800007.
22. Torshizi AD, Zarandi MHF, Torshizi GD, Eghbali K. A hybrid fuzzy-ontology based intelligent system to determine level of severity and treatment recommendation for Benign Prostatic Hyperplasia. Computer Methods and Programs in Biomedicine. 2014 Jan;113(1):301-13. PubMed PMID: WOS:000327180900028.