

Predictive risk Project

Literature Review (summary)

This is a joint project commissioned by Essex Strategic Health Authority on behalf of the 28 strategic health authorities, the Department of Health and the NHS Modernisation Agency and undertaken by the King's Fund, in partnership with Health Dialog and New York University.



NEW YORK UNIVERSITY



Health Dialog

Understand. Collaborate. Advance.

1. Study background

1.1. Project scope

As part of the Government's agenda to improve the care for people with long term conditions, the King's Fund, along with New York University and Health Dialog Data Service (an American company specialising in health data analysis), has been commissioned by Essex Strategic Health Authority on behalf of the 28 strategic health authorities, the Department of Health and the NHS Modernisation Agency to produce a risk prediction system for use by PCTs to identify patients who are at high risk of admission to hospital. The principal output from this work will be an algorithm that PCTs can apply to hospital episode statistics (HES) and community data in order to identify high risk patients effectively. This literature review has informed the development of this algorithm.

1.2. Policy context

Whilst the focus of much recent government policy in the NHS has been on reducing waiting lists and times for people requiring elective surgery, the issue of how to improve the health of people with long term conditions has been rapidly climbing up the policy agenda. It is increasingly recognised across the world that significant morbidity and cost (to patients, carers, employers and the NHS) result from long term medical conditions. Furthermore previous research³ has shown that a large number of emergency admissions are patients with conditions for which effective primary care can reduce the risk of admission. Many of these patients experience multiple emergency admissions in a year. In the UK, around 5% of patients account for 40% of inpatient bed days⁵, and in the US, 10% of Medicaid beneficiaries account for around 70% of all expenditure⁶.

The number of people living with a long term condition is set to rise with the World Health Organisation predicting that such conditions will be the leading cause of disability by 2020⁷. As the population continues to age and more people are living with increasingly complex conditions, it is essential that appropriate and sustainable models of care are put in place.

In response to these figures, attention has been turned to the potential of so-called 'upstream' care to **prevent** the deterioration of individuals' health to the point where an 'downstream' and expensive acute emergency admission is required. The assumption is that better 'upstream' care (or case management) will improve health such that the risk of admission (and cost) will be reduced. Guidance from the DH now suggests that PCTs should offer case management to individuals who are at 'high risk' of an emergency admission. Evidence suggests that case management programmes deliver a high quality of care and are popular with patients⁸. Evidence of long term financial implications is insubstantial.

A critical challenge now facing health and social services is how to identify patients who are at future high risk of admission, and for whom an intervention might reduce that risk. There are various approaches but little consensus as to what is most effective. A key criticism of the recent Evercare pilot, whilst reporting high levels of patient satisfaction, was that it did not accurately identify individuals who were most at risk of future admission⁹. Thus the impact on admissions was estimated to be around 1%⁸. For case management programmes to have the most impact on health and admissions, it is essential that high risk individuals are identified accurately.

The following sections examine the approaches that have been developed across the world to identify high risk patients. In this context, 'high risk' refers to a high risk of admission to hospital.

2. Identifying patients - what techniques have been used?

There are a number of ways to identify patients who are likely to become high risk in the future. To an extent, the approach used depends upon the risk that is being measured, the time scale over which it is to be measured and the purpose of predicting the risk. Three main techniques have been tried and tested, within which there are numerous sub-categories:

- 1) threshold approach;
- 2) clinical knowledge; and
- 3) predictive modelling.

Threshold approach

The threshold approach (also known as rules-based and criterion-based approach) uses a set of a priori (previously designed) criteria which define or describe the 'high risk' patients. No statistical modelling is used. This technique was used in some Evercare pilots and in the Castlefields work.

The technique identifies any patients who meet a specified criterion or threshold for a parameter of interest, such as readmission¹. For example, the threshold may be anyone who is over 65 who has had 5 or more admissions in the previous 12 months. Therefore, everyone within the defined population who meets these criteria would be identified as being 'high risk' and, thus, targeted with an intervention.

Although widely used, this approach has been shown to yield low levels of accuracy in predicting future risk. This is largely because individuals who are at risk one year, may not be at risk the next and vice versa⁴. The result is that a large number of people need to be targeted to ensure those who really are high risk are included. This has been found to be inefficient and expensive.

Using clinical knowledge

One approach widely used in the UK and the US is one based on clinical knowledge, whereby the clinician uses their instinct and experience to identify individuals who are likely to become high risk of an emergency admission. The clinician then refers the patient to a case manager if it is thought an intervention would help reduce the risk.

Very little formal evaluation has been carried out to assess the relative accuracy of using clinical knowledge in predicting future risk, but evidence points to this approach having a low level of predictive accuracy. Clinicians may be able to identify patients who are *currently* high risk, but are less able to identify those who are going to become high risk in the future². This approach is also limited to individuals who come into contact with a health professional so is less likely to identify individuals for interventions before they become high risk.

Predictive modelling

The third possible way of identifying high risk patients is through using predictive modelling. Predictive modelling seeks to establish relationships between sets of variables, using statistical modelling, in order to predict future outcomes. It usually incorporates formulae to allow users to interpret historical data. It then forecasts future events based on the identified relationships¹.

Evidence points to predictive models having impressive predictive ability. However, within the category of predictive modelling is a large variety of techniques, some of which are more developed than others. Literature on the subject is extensive, yet it is clear that there is no single consensus as to which technique is best. The most developed approach uses regression models but there is emerging interest in using artificial intelligence. Within regression modelling, the predictive power varies according to the data variables used.

This project will be using predictive modelling as this has been shown to be the most accurate.

3. Regression models

Regression modelling is to be used in this project as it has been shown to be more accurate than threshold approaches and using clinical knowledge to identify individuals at high risk of emergency admission. Models can vary in what they are predicting, the time over which they are predicting, the type of regression used and the type of data used. In the case of this project, the model is being developed to predict admissions over the next 12 months. Logistic regression will be used as this produces predicted probabilities for admission for each individual.

What is regression? Regression is a statistical technique used to find relationships between variables for the purpose of predicting future values.

3.1. Data variables

Predictive ability depends largely upon what variables are used in the model and this, in turn, depends upon the data available on individuals. There is an extensive list of potential variables but most can be grouped under the following headings:

- Socio-demographic;
- Diagnostic;
- Prior utilisation/cost;
- Pharmacy data;
- Health status/functionality; and
- Clinical data.

Models vary in so many different ways that drawing comparisons between them is complex. Such variation also means that there is little consensus over which variables are the most predictive. However, the majority agree that *demographic* variables alone do not yield high predictive power and that the addition of *diagnostic* and *prior utilisation* data to *demographic* variables increases power significantly. The use of data on use of *pharmaceuticals* has been shown to add significant power in a small number of studies and information on health status or functionality has been shown to add modest predictive power. Very few models have included clinical information (such as blood test results, blood pressure measurements) and the relative accuracy of using these data has not been widely evaluated. However, there is a

small amount of evidence to suggest that clinical variables have the potential to be highly predictive.

Although most papers found information on clinical diagnosis to be highly predictive, there are many different ways of utilising this information. The approach that appears to yield the highest predictive power is called the DCG/HCC model (diagnostic cost group/hierarchical coexisting conditions). Unlike other diagnostic models, this takes account of combinations of conditions. Information on prior utilisation of care is also highly predictive, with some papers finding it to be more predictive than diagnostic information¹⁰. There is also evidence to suggest that two year's prior utilisation data is more predictive than just one year's worth¹¹.

Compared with data on prior utilisation of care, diagnosis and use of pharmaceuticals, information on health status¹ has not been shown to have high predictive power. However, findings have been mixed with some information on health status providing relatively high predictive power¹⁴. A complication with using this information is that it is not routinely collected and the cost of collecting it can be greater than the extra predictive power added to the model¹². Data on clinical tests (such as actual blood pressure readings) have been included in some models but the independent predictive power of these data has not been widely evaluated. Some papers suggest that data on clinical tests would add predictive power to a model as this provides detailed information about the patient's health status and is collected routinely.

3.1.1. Availability of data

A major issue to consider when developing a model is what data are readily available and of sufficient quality to be used. It is important to have historical data as well as current information about the patient as the regression model requires past information to identify associations and relationships¹¹. The quality of the data will inevitably impact upon the predictive power of the model. Within an NHS context, it is recognised that PCTs do not all have identical datasets and there is variation in the capacity to handle data.

4. Our approach

The findings from this literature review are informing phases two and three of the project, which are concerned with developing an algorithm for PCTs to use to identify individuals at risk of high admission in whom an intervention may yield results.

Based on this evidence, the phase two algorithm is utilising logistic regression to predict the future risk of admission. The design of the model is empirically driven and the team has been able to utilise whatever information is available that helps predict future admissions. Consistent with evidence from the literature, variables in the algorithm include cost, utilisation, data on diagnosis and use of pharmaceuticals and demography. Within the utilisation category, the algorithm is able to both flag individuals with prior admission and take account of the number of admissions that an individual has had within a defined period. What is unique to this algorithm is that, in addition to the above factors, it also includes the number of specialists a patient had seen, area rates for supply-sensitive conditions¹⁵ and the relative rate of subsequent admissions for individual hospitals. Thus, the model takes into account the subsequent admission rates for the hospital to which the patient has been admitted in order to pick up geographical differences in physician style. Clinical data are also included in the model. There is a paucity of evidence as to the predictive value of using clinical data but some papers point to their potential. As health status has been shown to be

¹ This may include an assessment of functionality in terms of activities of daily living. Examples of tools used include the short form 36¹² and the Identification of Seniors at Risk (ISAR) questionnaire¹³.

relatively predictive, it is likely that clinical data will add to the power of the model. The phase two algorithm is using HES data and inpatient data from Clearnet, but is also exploring the added predictive power of data on use of accident and emergency and outpatients. Where records are complete, the algorithm is using three years' worth of prior data.

Phase three will build on the findings from phase two and the literature review. The phase three algorithm will link HES data from community services, such as GP records, district nursing records and social services data. In doing so, the work will fill in some of gaps identified in the current literature. This includes exploring the predictive value of outpatient data and its application for identifying individuals who are not yet in the 'high risk' strata of the risk triangle and identifying additional diagnostic data that may not be present in the phase two data sets. The phase three model will also have a broader set of pharmacy data that can be used as predictors of future admissions. All of this data will be linked to the phase two dataset in order to explore whether such variables add significant predictive power in the identification of future high risk individuals. A principal aim of the phase three work is to enable PCTs to identify those individuals who are not yet considered 'high risk' in order to prevent further deterioration and consequent high cost emergency admission.

1. Cousins MS. **Introduction to predictive modelling.** Shickle LM and Bander JA. *Disease Management* 5(3). 2002.
2. Dudley RA. **The best of both worlds? Potential of hybrid prospective/concurrent risk adjustment.** Medlin CA, Hammann LB, Cisternas MG, Brand R, Rennie DJ, and Luft HS. *Medical Care* 41(1), 56-69. 1996.
3. Hutt R. Case-managing long term conditions. Rosen R and McCauley J. 2004.
4. Dove HG. **A prediction model for targeting low-cost, high-risk members of managed care organizations.** Duncan I and Robb A. *American Journal of Managed Care* 9(5), 381-389. 2003.
5. Department of Health. *Improving Chronic Disease Management* . 2004.
6. Reuben DB. **Identification of risk for high hospital use: cost comparisons of four strategies and performance across subgroups.** Keeler E, Seeman TE, Sewall A, Hirsch SH, and Guralnik JM. *Journal of the American Geriatric Society* 51(5), 615-620. 2003.
7. World Health Organisation. **Innovative Care for Chronic Conditions: Building Blocks for Action.** 2002.
8. National Primary Care Research and Development Centre. *Evercare Evaluation Interim Report: Implications for supporting people with long-term conditions.* 2005.
9. Roland M. **Follow up of people aged 65 and over with a history of emergency admission: analysis of routine admission data.** Dusheiko M, Gravelle H, and Parker S. *British Medical Journal* 330, 289-292. 2005.
10. Lamers LM. **Health-based risk adjustment: is inpatient and outpatient diagnostic information sufficient?** *Inquiry* 38(4), 423-431. 2001.

11. van Barneveld EM. **Risk-adjusted capitation payments for catastrophic risks based on multi-year prior costs.** van Vliet RC and van de Ven WP. Health Policy 32(2), 123-135. 1997.
12. Pietz K. **Predicting healthcare costs in a population of veterans affairs beneficiaries using diagnosis-based risk adjustment and self-reported health status.** Ashton CM, McDonnell M, and Wray NP. Medical Care 42(10), 1027-1035. 2004.
13. McCusker J. **Prediction of hospital utilization among elderly patients during the 6 months after an emergency department visit.** Bellavance F, Belzile E Verdon J and Cardin S. Annals of Emergency Medicine 5(5), 438-445. 2000.
14. Meldon SW. **A brief risk-stratification tool to predict repeat emergency department visits and hospitalizations in older patients discharged from the emergency department.** Mion LC, Palmer RM, Drew BL, Connor JT, Lewicki LJ, Bass DM, and Emerman CL. Academic Emergency Medicine 10(3), 224. 2003.
15. Wennberg JE. **Finding High Quality, Efficient Providers for Value Purchasing: Cohort Methods Better Than Methods Based on Events.** Fischer ES. Medical Care 40(10), 853-855. 2002.