

Predictive risk Project

Literature Review

June 2005

Natasha Curry, King's Fund
John Billings, New York University
Bob Darin, Health Dialog Data Services
Jennifer Dixon, King's Fund
Michael Williams, King's Fund
David Wennberg, Health Dialog Data Services

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.


Modernisation Agency



NEW YORK UNIVERSITY

King's **Fund**



Health Dialog
Understand. Collaborate. Advance.

1. Study background

1.1 Project scope

As part of the long term conditions agenda, 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 readmission to hospital. The principal output from this work will be an algorithm that PCTs can apply to HES and community data in order to identify high risk patients effectively. PCTs will then be in a better position to target interventions that aim to reduce the risk of admission for these patients. This literature review has informed the development of the 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 care for people with long term medical conditions has been rapidly climbing up the policy agenda. The World Health Organisation (WHO) describes care for long term conditions as "the health care challenge of this century", with such conditions currently responsible for 60 per cent of the global burden of disease and likely to be the leading causes of disability by 2020 (World Health Organisation 2002). Research has shown that a large number of people who have long term conditions (an estimated 17.5 million in the UK) are frequently hospitalised as an emergency (Hutt R 2004). Figures published by the Department of Health indicate that two-thirds of patients admitted as medical emergencies have an exacerbation of a long term condition or have a long term condition. It is estimated that around 10% of inpatients account for 55% of inpatient days and 5% of inpatients account for 40% of inpatient days (Department of Health 2004).

In financial terms, this minority of patients accounts for a disproportionately large proportion of NHS expenditure. Studies in the US have revealed that 10% of all Medicaid beneficiaries account for around 70% of all expenditure (Reuben DB 2003). Similarly, it is estimated that approximately 4% of the population served by managed care organisations have diabetes but that they account for around 12% of total expenditure (Selby JV 2001) and that per capita expenditure for people with COPD is two and a half times greater than for those without (Fan VS 2002). This situation may become more pronounced as people are living longer with increasingly complex conditions and as the actual number of people with a long term condition increases. Around three per cent of patients over 65 account for 35 per cent of admissions (National Primary Care Research and Development Centre 2005) and projections show that the population aged over 75 will grow substantially with the number of people over 85 in 2011 set to double what it was in 1981 (Bowns I 1991).

In response to these figures, attention has been turned to the possibility of putting in place so-called 'upstream' care to prevent the deterioration of individuals' conditions to the point where an expensive acute emergency admission is required. It is hoped that such preventative care will bring both financial and health benefits. The content and format of interventions to manage individuals with long term conditions (such as case management) may vary widely but are based on the underlying principles of seamless, pro-active and preventative care from a multi-disciplinary team in the least intensive setting appropriate (Department of Health 2002). The aspiration is that the individual's health is not allowed to deteriorate to such a point that they need to be hospitalised and, because this admission is avoided, the cost of care may be reduced. The intention is that the quality of care is enhanced and the patient's experience is improved. In order to drive through such changes, Public Service Agreement (PSA) targets have been introduced to reduce emergency bed days by 5% (from the level in 2004) by 2008 and to offer personalised care plans for vulnerable people

most at risk. Evidence suggests that case management can have a positive effect on health outcomes, and is often popular with patients, but it must also be recognised that some frequent users may not be amenable to intervention to reduce the risk of admission. There is an implicit assumption that runs through all of this current policy, and that is that upstream, community-based care will be cheaper in the long run than traditional, reactive episodic models of care. Evidence to date on this is weak.

A key challenge now facing health and social services now is how to identify patients who are at high risk of emergency admission, and for whom an intervention might reduce that risk. 'Case-finding' is the term given to the practice of identifying at-risk patients. There are various approaches for case-finding, but little consensus as to what is most effective. Evercare, a US model of care management, was recently piloted in the UK. Whilst the evaluation of Evercare (National Primary Care Research and Development Centre 2005) reported high levels of patient satisfaction, the approach used to identify high risk cases was suboptimal. This evaluation report concluded that the approach to case-finding used by Evercare may not predict sufficiently accurately which older people are at risk of unplanned admission in the future. It also estimates that with the (then) current scale of Evercare, emergency admissions are unlikely to be reduced by more than 1% (National Primary Care Research and Development Centre 2005). Further analysis of HES data, as published in the British Medical Journal, highlights that the criteria used in case-finding by the Evercare pilots (i.e. being 65 or older and having two or more admissions in the index year) has "low sensitivity in detecting older patients who will have high admissions in the following year" (Roland M 2005). In order for case management programmes to have an impact on the health of people with long term conditions and on the rate of (re)admission, it is essential that an effective and accurate system of case-finding is developed.

This literature review aims to summarise key points of learning about the identification of 'high risk' patients from around the world. The review highlights the different approaches taken to identify patients in order to inform the next stage of this project - the development of an algorithm to identify high risk patients for use by PCTs. The findings from this review will inform the development of the algorithm and identify any areas for future research.

The review is organised under the following headings:

1. Study background: the project scope and policy context;
2. Conceptual framework: what is meant by risk and why it is useful to measure it;
3. Identifying patients: a summary of the key approaches used around the world to identify high risk patients;
4. Regression models: the most common approach used and the proposed approach for this project;
5. Data availability: issues around data availability and accessibility;
6. Our approach: the next stages of the project and the approach that is to be taken.

2. Conceptual framework

2.1 What risk are we measuring?

In order to develop an effective tool for identifying 'high risk' patients, it is necessary to define what is meant by risk in this context. The literature on the subject indicates that there are a huge number of risks that could potentially be measured. Risk can include death, decline in health status and functioning, readmission to hospital, or cost of health care. The central issue is that a small number of patients could be classified as 'high risk' and using/likely to use a large amount of resources. In the US, the focus is on those who are at high risk of requiring any expensive treatment, measured in data from insurance claims. However, the cost of care could be measured in anything from hospital bed days to medications to admissions and it is these units that are more relevant to the UK context. Since each admission is a costly event, a patient who is frequently admitted to hospital could be classed as a high risk patient. However, the use of readmissions as

a measure for high risk (or high cost) may exclude individuals who are admitted fewer times but have a longer length of stay each time they are admitted. It may also mask any variation in the complexity or severity of the condition and the fact that a range of differently priced treatments and medications may be available for the same condition.

The majority of studies in the papers reviewed have defined risk in terms of the risk of hospitalisation. Some are specific and look at the risk of unplanned hospital readmission (Parker JP 2003) or early and frequent return to A&E (McCusker J 2000b) whilst others assess the risk of 'utilisation' in a much broader sense; some include use of primary care and outpatient visits as well as emergency care (Wahls TL 2004). On an individual level, other risks which are less easily quantified within the framework of cost and admission may be considered to be important. Such risks include death, functional decline (Dendukuri N 2004), clinical complications (Selby JV 2001) and ill health (Barber M 2001). Although death is not necessarily a high cost event in itself, analysis has shown the last year of life to be characterised by high healthcare costs and therefore of great significance to health providers and insurers (Petersen LA 2005). In a broader sense, readmission may be considered to be a symptom of other factors, such as unsuccessful discharge processes or inappropriate/inadequate social care (Allgar V 2002). In this case, some commentators consider it to include the quality of treatment or patient satisfaction when assessing risk (Liu CF 2004). However, such nebulous elements are not easily incorporated in a quantitative model.

This project has the specific aim of identifying those at risk of future emergency admissions for whom an intervention may reduce the risk of admission and improve health outcomes.

2.2 Why measure risk?

As discussed above, there is a small number of people who account for a disproportionate amount of resource utilisation and who could be classified as 'high risk'. There are two main reasons in the literature for trying to identify which individuals are 'high risk' and therefore likely to be costly. The first is to set an accurate insurance premium that is fair and likely to cover the real cost of future healthcare. This is known as 'risk adjustment' and used within insurance-based health systems. The second is to identify patients for whom an appropriate intervention would improve care and prevent future admissions. Although these reasons are very different, they are both underpinned by a recognition that, in any system, there are scarce resources that need to be allocated in the most appropriate way. The former is purely financial and seeks to compensate insurers for the costs of care. The second is focused more on identifying the appropriate people on which to direct intensive resources to bring about better health outcomes with the aim of saving money in the long run. Outcomes for models used for setting premiums mostly focus on total healthcare cost or predicting health status whilst the most frequent outcome for models which identify individuals for interventions is hospitalisation.

Due to the nature of the insurance-based US healthcare system, the incentive to develop accurate risk prediction techniques has been strong. Risk adjustment has developed out of a need for insurers to be able to cover their costs from premiums, despite the make-up of their membership. Sick patients incur higher costs so there is an incentive for insurers to avoid taking on the sickest in society. Risk adjustment seeks to correct or, at least, reduce this incentive, by ensuring that insurers cover their costs through a fairer system of payments. In a system with no risk adjustment, insurers will seek out the healthiest patients as they will be the cheapest to care for. As Hughes explains, in 1998 Medicare's healthiest 76.3% of members consumed 14% of total expenditures whilst the sickest 15.3% accounted for 75.7% of expenditures (Hughes AS 2004). Similarly, it is estimated that 10% of all Medicaid beneficiaries account for around 70% of all expenditure (Reuben DB 2003).

Although this particular application of the technique is not currently relevant to the UK, the principles of stratifying patients according to risk are relevant and useful. Such techniques can be used to identify patients for an appropriate intervention in order to improve health outcomes, efficiently allocate resources, reduce future costs and to facilitate better planning. Emergency admission to an acute hospital is generally regarded as a poor outcome both for the patient, as it indicates that their health has deteriorated to such a

poor state that emergency admission is necessary, and for the provider, as an emergency admission to an acute hospital is generally costly. Thus, the logic goes, preventing this admission by early intervention is of benefit to all and, it is assumed, of long term economic benefit with the potential to generate long term economic savings provided the intervention costs are lower than the hospitalisation costs. If the system is able to identify those patients at the highest risk of such an admission, more intensive resources can be focused on them leading to more efficient allocation of resources and facilitating better planning of services and workload as crises are avoided or at least minimised.

The underlying assumption is that intensive medical intervention (such as hospitalisation or emergency department visit) could be better substituted with less expensive modes of care which result in better health outcomes. One gap that currently exists in the evidence is that of financial analysis and this is one area that the US experience cannot fully translate to the UK situation. It is possible to quantify the cost of an acute emergency admission and, thus, possible to cite a year on year figure of expenditure on this type of care. What is not yet available is a comparator which indicates the cost of providing an early intervention to prevent emergency admissions. Although case management programmes have been developed in the UK, there has been very little financial analysis of their impact. Likewise, in the US, the majority of studies have been focused on setting fair premiums within the context of risk adjustment and have not focused on the intervention aspect. Some commentators, such as Carlson 2003, have asserted that case management, and other interventions, are an expensive solution in the short term, but that return on investment will be seen over a longer time period (Carlson B 2003).

The stratification of patients can be illustrated using Kaiser Permanente's risk triangle, below. It is the individuals at the top of this triangle who are most at risk of emergency admission. Case management programmes attempt to target these individuals to prevent them being admitted. However, there is some debate as to whether this is the most appropriate area of the triangle on which to concentrate resources. It has been suggested that once an individual has reached this level of risk, an intervention is likely to be too late to prevent admissions. It may be of more value (both in financial and health outcomes terms) to identify those individuals in the lower two strata who are likely to move into the high risk/high cost level.

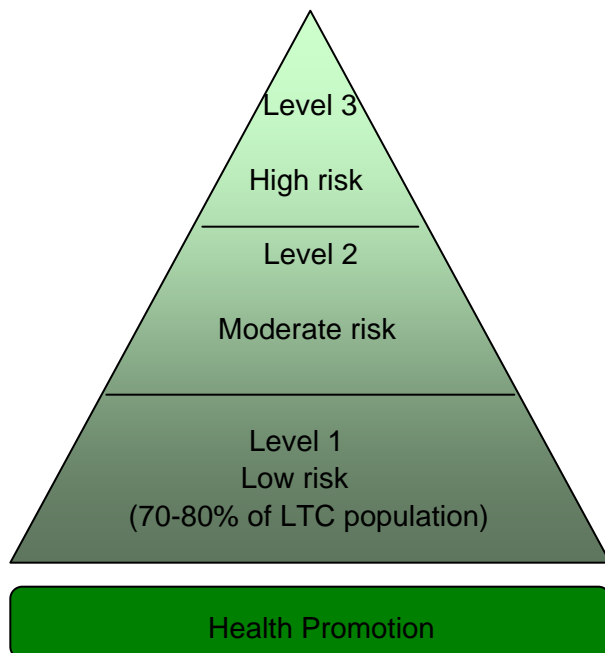


Figure 1: Risk stratification triangle as developed by Kaiser Permanente

Another issue that has been highlighted in the literature is that of changing risk profiles. Many case management programmes and US health insurance companies make an implicit assumption that an individual's level of risk and their health-related behaviour is sustained. This is unlikely to be the case and figures from the US confirm this; in one insurance company, the highest cost members represented 1% of members and accounted for 21% of total cost in 1998 but this same 1% of members accounted for just 7% of total cost in the following year without intervention (Dove HG 2003). This is because resource consumption of the highest cost patients generally decreases even in the absence of any intervention. This adds an extra layer of complexity to the process of identifying appropriate patients for an intervention. Most models seek to identify the highest cost patients (or the patients who are the highest utilisers) but this year's low cost/use patients may be next year's high cost/use patients. The challenge is how to pre-empt who will be moving up through the risk triangle and to put in preventative programmes targeted at them. Related to this is the need to be able to identify those patients who were high cost/use but who now no longer require a high intensity intervention and who could be 'stepped down' from a case management programme.

3. 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. There are three principal techniques that have been tried and tested, within which there are numerous sub-categories:

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

3.1 Threshold modelling

Threshold-based techniques are also known as rules-based or criterion-based modelling. This approach uses a set of a priori criteria which define or describe the 'high risk' patients. This is based on algorithms and no statistical modelling is used. The technique identifies any patients who meet a specified criterion or threshold for a parameter of interest, such as readmission (Cousins MS 2002). For example, in identifying those at high risk of emergency readmission, 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'.

Such models have been widely used in the UK within case-finding projects. The Castlefields project, some Evercare pilot sites and a number of case finding projects in London have adopted this approach (Castlefields Health Centre 2004). Evidence suggests that these models have not yielded a high degree of accuracy within a general population, although they have proven to be more accurate when used within a specific clinical context, such as identifying those at risk of coronary heart disease (Department of Health 2002).

Criteria used in the Castlefields project for finding patients:

People over 65, who met at least three of the following criteria were targeted:

- ***four or more active chronic diagnoses;***
- ***four or more medications, prescribed for six months or more;***
- ***two or more hospitalisations, not necessarily as an emergency, in the past twelve months;***
- ***two or more accident and emergency attendances in the past twelve months;***
- ***significant impairment in one or more major activity of daily living;***
- ***significant impairment in one or more of the instrumental activities of living, particularly where there are no support systems in place;***
- ***older people in the top 3 per cent of frequent visitors to the practice;***
- ***older people who have had two or more outpatient appointments;***
- ***older people whose total stay in hospital exceeded four weeks in a year;***
- ***older people whose social work contact exceeded four assessment visits in each three-month period;***
or
- ***older people whose pharmacy bill exceeded £100 per month.***

Figure 2: An example of case-finding criteria used in Castlefields (Castlefields Health Centre 2004)

Identifying those at risk (of CHD:) A Nottingham practice ***ran a data search to identify all registered patients over 50 years of age with a BMI over 30.*** This resulted in the identification of over 300 patients (out of a registered population of around 9,000) who had not previously been categorised as at risk of CHD and diabetes. This was then used as the basis of a call and recall programme. Once identified as high risk, NSFs provide guidelines.

Figure 3: an example of thresholds being used in case-finding in a specific population (Department of Health 2002)

In general, threshold models are predisposed to the negative effects of selection bias and regression to the mean. Selection bias occurs when individuals are selected because they are outliers who represent an extreme. This means the model suffers from the problem of regression to the mean. 'Regression to the mean' describes a situation whereby those who are extreme one year (e.g. in terms of number of admissions) are rarely extreme the next. For instance, patients who are 'expensive' in one 12 month period are likely to improve even without intervention. Therefore, there is a risk that the intervention would not focus on individuals who are likely to be high risk in the next 12 months which would lead to a misallocation of resources. In the US, where healthcare is insurance-based, using such a model for risk adjustment would lead to some patients being overcharged and others being undercharged. The effectiveness of a threshold model may vary according to the strata of the risk triangle (figure 1) on which the intervention is focused. Evidence is weak as to which strata an intervention is applied provides the best return on investment, although most programmes tend to target the top third. Formal evaluations of threshold models are rare, Cousins suggests that such a model is around half as accurate as other, predictive, models (Cousins MS 2002). It is for this reason that threshold models are no longer frequently used in the US for risk adjustment purposes.

3.2 Clinical knowledge

An alternative approach that has been used in the UK and the US is one based on clinical knowledge, occasionally known as 'clinical hunch'. In this approach, the clinician uses their instinct, knowledge and training to identify individuals who are likely to become high risk and who would benefit from an intervention. This approach has not been used within the realm of risk adjustment although it appears to be widely used in identifying patients for specific interventions. To date, this is perhaps the most widely used approach in the UK. GPs, social services and other health and social care professionals have referred patients to schemes and interventions on a 'hunch' that these individuals would benefit from interventionsⁱ.

Very little formal evaluation has been carried out to assess the relative accuracy of clinical knowledge in predicting future risk, but that which has been undertaken 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 future (Dudley RA 1996). One study which examined the accuracy of staff predicting readmissions of schizophrenia patients indicated that just under 20% of readmissions were predicted, but this was amongst a very small and specific population (Olfson M 1999). The use of clinician knowledge to identify individuals *currently* in need of an intervention can be effective but is limited to those patients in contact with a service. The preventative nature of case management is undermined as an 'event' would have to occur to bring about this contact. This method has been widely used in a number of health economies and, although some were effective in bringing about better health outcomes, they have not been proven to be efficient at identifying those at future risk. What the evaluation of the Evercare pilots suggested was that there is a group of high risk people who are not in regular contact with health and social care services (National Primary Care Research and Development Centre 2005). Some PCTsⁱⁱ have used validated postal questionnaires to screen whole populations but such techniques tend to have low response rates and there is little evidence to suggest the future high risk patients respond. Using clinicians to predict *future* risk in a large, general, population would be inefficient and likely to have low levels of accuracy.

3.3 Predictive modelling

The third possible way of identifying patients is through using predictive modelling. Predictive modelling seeks to establish relationships between sets of variables in order to predict future outcomes. It usually incorporates formulae to allow users to interpret historical data and make predictions about the future, map associations and statistical relationships to a specific target. It then forecasts future events based on the identified relationships (Cousins MS 2002). Evidence points to predictive models being superior to both threshold models and clinician knowledge in identifying patients at risk of future admission. 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. Most predictive models have focused on regression techniques, although there is emerging interest in artificial intelligence (discussed below).

The next section focuses on the use of regression models.

4. Regression models

The literature reveals that there are numerous predictive regression models that have varying degrees of accuracy. Models can vary in four ways: they vary in what risk they are predicting, in the type of data that they use, in the time over which they predict this risk and in the type of regression they use. The type of risk that is being predicted depends upon the purpose of risk prediction – as discussed above, it can be for the purposes of risk adjustment or for identifying high risk individuals. The majority of examples in the literature are concerned with predicting risk of high cost for the purposes of premium setting, although proxy

measures such as hospitalisation and medications have been used for cost (Bierman AS 1999;Dove HG 2003;e.g.Reuben DB 2003). In order to evaluate the relative effectiveness and usefulness of a model, two measures can be used: one is how well it explains the data and the other is its accuracy (in that it can identify individuals who are high risk and individuals who are not high risk).

What is regression analysis?

Regression analysis is a statistical technique. Regression is a generic term for all methods attempting to fit a model to observed data in order to *quantify the relationship* between two groups of variables. The fitted model may then be used either to merely *describe* the relationship between the two groups of variables, or to *predict* new values¹. In the case of this project, the regression model will analyse the relationship between variables (such as age, number of previous admissions and number of medications being taken to predict the probability of a future admission).

4.1 Use of regression

Regression is a technique used to assess the linear relationship between independent variables (these are the inputs, such as patient information) and a dependent variable (this is the outcome measure; in this case, emergency admission). The literature indicates that predictive models have used both linear and logistic regression techniques. Both aim to assess a linear relationship but differ in the type of outcome variable used. The main difference is that the linear regression model outcome variable is continuous, whereas the logistic regression outcome is binary (i.e. it has two categories that represent an event or characteristic of interest for example, whether an individual is 'high risk' or not). The outcome variable of a linear model is an actual value (such as cost) whereas a logistic model produces a predicted probability between 0 and 1 of an event, such as hospitalisation. Both types of models can be used to rank individuals by decreasing predicted value in order to target individuals with the highest risk (those that fall above a pre-determined cut-off point).

The literature indicates that a number of different regression models have been used in this field. Few papers in the literature directly compare the relative effectiveness of the different types of models, Zhao (Zhao Y 2001; 2003) and Ash (2000; 2001) favour multiple linear regression models, Schatz (2003) and Roblin (1999) favour multiple logistic regression models, while Meenan (1999) and Dove (2003) used both multiple linear and logistic regression models. Meenan and colleagues found that a linear regression model was superior to a logistic version of the same model, however, they note that both models perform similarly at policy-relevant thresholds. Dove and colleagues developed several linear and logistic regression models, then reported results for only one model type, presumably their best model. Use of either method is statistically valid for identifying high risk individuals, and logistic regression can be used as long as the variables can be appropriately transformed in order to build such a model.

4.2 Evaluating the model

There are two main ways of judging the relative quality of a model: the first is its variance/fit (or what proportion of the variation in the outcome it can explain) and the second is its accuracy (the ability of the model to identify the correct people as high risk). Although there are these principal parameters of success, there are many different graphical and numerical ways used to compare different models and there does not appear to be consistency or a shared framework for doing so. The two sections below explain how a model may be evaluated.

4.2.1 Measuring accuracy

The relative accuracy of the different models is measured in terms of ‘sensitivity’ and ‘specificity’. The sensitivity of a model is its ability to identify those at risk (of emergency readmission or death or functional decline and so on). The specificity of a model is its ability to identify those not at risk of the outcome. Sensitivity and specificity can be measured for either linear regression models or logistic regression models.

This is an important distinction to make as a model may be able to identify a high number of people who are high risk by having a large sample size, but within this population it may also include a high number of people who are not high risk. For instance, in a population of 10,000, the model may indicate that 5,000 people are going to be high risk next year and have a 100% sensitivity rate – this means that 100% of people who turn out to be high risk next year were identified by the model. However, in reality, it may be that this high risk population is only 10% of the total population identified (i.e. 1,000 people). Therefore, if the model is being used to identify people for an intervention, the PCT would have to target the intervention at 5,000 people in order to ensure the 1,000 people at real high risk are included. This is clearly inefficient and costly and a poor allocation of resources. The 1,000 real high risk people are referred to as ‘true positives’ and the 4,000 people who were identified as high risk but are not in reality high risk are known as ‘false positives’. The higher the false positive rate, the less useful the model is. However, if the model is also able to identify individuals who are not going to be high risk, it is likely to lead to a better allocation of resources. In this scenario, a perfect model would accurately indicate that 1,000 individuals are going to be high risk next year and also be able to indicate which 9,000 remaining individuals are not going to be high risk next year.

	Patients not identified by model	Patients identified by model
Patients who are not admitted	a	b
Patients who are admitted	c	d

<p>sensitivity = $d / (d+b)$</p> <p>specificity = $a / (c+a)$</p>

Figure 4: calculating sensitivity and specificity

There is clearly a trade-off between the sensitivity and specificity of a model and the relative rate of each depends partly upon the proportion of the population you are attempting to identify; i.e., the definition of ‘high risk’, often known as the ‘cut point’ or ‘cut-off point’. In the context of death, this is clear cut but in the context of readmissions or cost, it is less so. For instance, are high risk users those individuals who are going to cost over a certain amount (say £25,000) or are they those individuals who are going to have 5 or more admissions? Or are they the top 5%, or perhaps the top 20%, of those at the highest risk of admission? It is relatively more difficult to identify a smaller proportion of individuals. As this percentage increases, the true positive rate increases and the false negative rate decreases (as there is a higher probability of identifying high risk individuals). However, this is likely to be at the expense of an increase in the false positive rate (Cousins MS 2002).

4.2.2 Measuring goodness of fit

A further way of evaluating the performance of a multiple linear regression model or an artificial intelligence model is by assessing how well it 'fits' the data. For linear regression models, the 'goodness of fit' is generally referred to as *R-squared*. For logistic regression models, the more commonly used measure is known as a receiver operating characteristic (ROC) curve. However, ROC curves have also been used for linear regression models.

R-squared (R^2) indicates the percentage of total variation among individual observations that can be explained by the model, either explained as a percentage or a number between 0 and 1; 0 explains none of the variance, 1 explains all the variance (Hu G 2004). Thus, the closer a model's *R-squared* value is to 1, the better the model explains the data. R^2 values vary widely between models.

For logistic regression models, the relationship between sensitivity and specificity is usually explored using a ROC curve. The ROC curve is measure of goodness of fit, or how effectively the model describes the outcome variable (Hosmer DW 2000). To construct the ROC curve, the x-axis is 1 minus the specificity (false positive) and the y-axis is the sensitivity (true positive rate) (Crichton N 2002). The area under the curve (AUC or c-statistic) can summarise the capacity of a model for discriminating those who experience the event of interest (for example, risk of admission) versus those who do not, and can therefore be used to compare models (Liu H 2003).

However, these two measures are not used in a standard way, which makes comparing models very complex. R-squared values are routinely published for assessing model fit for linear regression models but not often for logistic regression models. Hosmer and Lemeshow (2000) do not recommend routinely publishing R-squared values for logistic models although some papers have published R-squared values for logistic models (e.g. Bhattacharyya, 1998). Conversely, Meenan presents a ROC curve for a linear regression model GRAM but does not present AUC values (Meenan RT 1999). Roblin uses logistic regression but does not report a ROC curve, instead reporting probabilities and observed/expected values (1999). Further examples of this inconsistency include a study by Schatz (2003) where c-statistics are reported in addition to sensitivity and specificity and a study by (Bhattacharyya SK 1998) which reports just sensitivity and specificity.

4.2.3 What is a 'good' model?

Deciding what is an acceptable proportion of prediction depends upon what the model is being used for. In terms of R-squared, the papers report values from 0.02 (Ash AS 2000) to 0.46 (Bhattacharyya SK 1998). In terms of ROC curves, AUC values of 0.7 up to 0.8 indicate acceptable model discrimination; values of 0.8 to 0.9 indicate excellent discrimination and values greater than 0.9 indicate outstanding discrimination (Hosmer DW 2000). Where AUC values have been reported in the literature, most tend to be around 0.6-0.7. Several commentators assert that only a certain proportion of healthcare needs can be predicted, the remainder being left to chance (Fowles JB 1996). Newhouse, for example, found only around 14.5% of total variance to be accounted for by differences between individuals and is therefore predictable (Newhouse JP 1989). Similarly, Van Vliet cited 13.9% as the upper boundary of predictability (Breyer F 2003). Mukamel reinforces this, stating that 15% of variance can be predicted, although other studies reckon this figure should be between 10% and 25% (Mukamel DB 1997). For instance, Lamers and Hughes suggest that the maximum proportion of healthcare costs that can be predicted is 20-25% and that the maximum reported for Medicare was 9% (Hughes AS 2004; Lamers LM 1999). Thus, any model is going to be limited in its predictive ability. A predictive value that is acceptable will depend upon the local context and the relative cost of identifying patients and establishing interventions (Mukamel DB 1997).

A model must be adaptable to the context. If logistic regression is used, the model may be set to identify anyone with a certain probability of admission. This makes the output from a logistic regression model more appropriate for predicting events (e.g. admissions) than for predicting a continuous range (e.g. cost). For instance, the model may be set to identify anyone with a 0.5 (50%) chance or above of admission. This 'cutpoint' may be altered by the user. Alteration of the cutpoint has implications for the false positive rate;

the higher the probability of admission, the lower the false positive rate but the higher the false negative rate.

4.2.4 Testing

In order to test the accuracy and predictive power of a model, it is good practice to use what is known as a split sample. This means that the model developer should divide the dataset randomly into two parts (not necessarily equal halves). The model is developed using one part of the data and then the finished product is tested and evaluated on the other half of the dataset. This is important as outliers or extreme value may have influenced variable selection and weighting. Testing the model on the other part of the dataset, which will contain different individuals with extreme values, will reveal how well it performs on new data and whether it will perform as well as new individuals are added. Models that do not perform well on datasets other than the one they have been developed on are said to 'overfit' the original dataset (Cousins MS 2002).

4.3 Concurrent versus prospective models

The time period over which the model is predicting has been shown to have a major impact upon the predictive accuracy of the model. Models are designed to predict risk either concurrently or prospectively. Concurrent modelling is that which seeks to predict relative risk for the current year. Prospective modelling seeks to predict risk into the future – usually the following 12 months or longer. There is consensus in the literature that concurrent prediction is more accurate than prospective prediction (e.g. (Dudley RA 1996; Hughes AS 2004). One paper illustrates how different these two can be; in predicting costs using diagnosis data, the prospective model produced an R^2 value of 10.66 whereas the concurrent model produced an R^2 value of 42.75 (Hughes, 2004). Although a better predictor, this technique is less useful for both risk adjustment and identifying patients for an intervention. In order to set premiums, insurers need to know the potential risk of an individual before the year begins. Likewise, it has advantages for planning if the patients have been identified ahead of time so that the intervention may prevent deterioration. Mukamel has found that predictive ability declines as predictive period increases. Thus, the predictive ability would be higher for next year than it would be for the subsequent year (Mukamel, 1997).

One paper has explored the potential of using a hybrid model which uses prospective techniques for the majority of patients, but also uses concurrent data where certain high cost diagnoses were made. In effect, this technique updates the model as and when diagnoses are made (Dudley RA 1996). The paper concludes that this hybrid approach is more accurate than both concurrent and predictive models. However, no other papers have looked at this approach, so the evidence as to its accuracy is very limited.

4.4 Variables used in models

Evidence indicates that the accuracy of regression models depends largely upon the variables used. The variables used, in turn, depend upon what data are available for patients. There is an extensive list of potential variables but they can be grouped under the following headings:

- 1) socio-demographic;
- 2) diagnostic;
- 3) prior utilisation/cost;
- 4) pharmacy data;
- 5) health status/functionality;
- 6) clinical data.

Studies of models using the various categories are often specific to a context or clinical condition so it is difficult to draw generalised comparisons. In assessing relative accuracy, it must be remembered that the purpose for which the model is being employed is an important consideration; it could be for setting premiums or identifying individuals. Unfortunately, the latter are very much in the minority. A further complexity stems from the fact that the majority of models use a combination of variables from the above list. In addition, the measures of accuracy are not consistent between papers so detailed comparison of R-squared/AUC values is not possible. Bearing these limitations in mind, the next section aims to summarise the key findings from a large number of different studies in order to inform the development of the tool for use in the NHS.

There is no consensus amongst the papers as to which variables produce the highest predictive power. 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 relative additional power added varies hugely between studies, possibly because of different contextual factors. The use of pharmacy data has been shown to add power in a small number of studies as has the use of health status/functionality information. Very few have included clinical and procedural information and the relative accuracy of using these data has not been evaluated.

A small number of papers reported findings from disease-specific studies. Within these, R-squared and AUC values vary and, due to differences in sample size and other similar factors, it is difficult to compare these with utilisation or cost models. For example, one study on asthma reported good sensitivity/specificity and AUC values of 0.781 and 0.712 which reflects good predictive value (Schatz M 2003). However, a similar study in 2004 reports lower AUC values of 0.615 and 0.614 (Schatz M 2004). A diabetes-focused study also reported sensitivity/specificity values comparable with Schatz 2003 (Bhattacharyya SK 1998). This could be compared with Axelrod's AUC of 0.89 for predicting total resource utilisation but would not be a like with like comparison (Axelrod RC 2003).

4.4.1 Socio-demographic data

Models that are based solely on socio-demographic data have been shown repeatedly to have low predictive power (Hendryx MS; Petersen LA 2005; Reuben DB 2003). Medicare, pre-1998, used a demographic model (average adjusted per capita cost or AAPCC) based on age and sex alone which could only account for less than 1% of Medicare's expenditure variance (Hendryx MS). Similar models tested on different populations concur, with the highest value being 4%, as predicted in a Dutch study which used age, sex, employment and disability as its variables (van Barneveld EM 1997). Numerous other studies that have compared demographic models with others have consistently found such variables, alone, to be the weakest predictor of high risk, whether that be rehospitalisation (Reuben DB 2003), cost (e.g. Lamers LM 1999), death (e.g. Petersen LA 2005), outpatient visits or hospital days (e.g. Wahls TL 2004). However, within the category of demographic models, it is evident that certain variables add more power than others. For instance, Carr-Hill asserts that including indicators of poverty makes a model more predictive of inpatient service utilisation (Carr-Hill RA 2002) and van Barneveld found that information on employment and disability combined with age and sex produced an R^2 value of 0.04 compared with 0.03 for a simple age/sex model (van Barneveld EM 1997). Breyer also found that the addition of income variables and marital status further increased the predictive value (Breyer F 2003).

4.4.2 Diagnostic data

Despite demographic variables being weak predictors when used alone, it is evident that most models include demographics in addition to other data. Evidence appears to be divided about the relative accuracy of diagnostic data, but it is clear from the literature that information on clinical diagnosis adds predictive power to a model with demographic variables. There are many different ways of incorporating diagnostic variables into a predictive model, each with a varying level of accuracy. Petersen concludes that although diagnosis is more predictive of death and hospitalisation than demographic data, the best predictive value is

produced when diagnostic data is combined with age (Petersen LA 2005). Lamers' study agrees; this study found diagnoses (DCG) to add predictive power to a demographic model when predicting expenditure (Lamers LM 1999). Two studies on mental health also reported that diagnostic data added power when predicting rehospitalisation and use of outpatient services (Cuffel BJ 2002;Hendryx MS). Fowles reported that, in terms of risk assessment, a diagnostic model (using ACGs) was more predictive than both demographic and health status models (Fowles JB 1996).

Types of diagnostic model:

There are a number of different models within the realm of diagnostic models and they have been shown to have varying degrees of accuracy. The most widely used are the ambulatory cost groups (ACGs) and the Diagnostic Cost Groups (DCGs) and latterly the hierarchical DCG/HCC. Other models include diagnoses, such as clinical risk groups (CRGs). A further, more recent development has been the global risk adjustment model (GRAM). Little information was provided in the literature as to how diagnoses had been grouped and coded in the various models.

ACGs and DCGs are both based on age, sex and diagnoses but differ in the way they characterise disease burden (Hughes AS 2004) ACGs are based on the assumption that a patient's illness burden better characterises the patient's need for health services than the presence of specific diseases. Thus, patients are assigned to an ACG based on evidence that certain groups of medical conditions have similar utilisation patterns (Rosen AK 2001). These assignments are based on ICD9-CM codes, which are then assigned to one of 32 diagnosis groups (ADGs). A person can theoretically be assigned to any number of ADGs. ADGs are further grouped until, finally, ACG weights are produced which represent the expected resource use level associated with a particular ACG (Meenan RT 2003).

DCGs were developed to predict future costs for the Medicare population based on the 'worst' inpatient diagnosis recorded in a time period. DCGs have since been developed to include ambulatory and inpatient diagnoses and the cumulative effect of multiple conditions in predicting total medical expenditure; this is known as the DCG/HCC (hierarchical coexisting conditions) model (Rosen AK 2001). Instead of just assigning people to diagnosis categories, this approach assigns severity weights to diagnoses and takes into account combinations of diagnoses where the individual has more than one condition. Thus, there are two types of DCG model: the DCG/HCC and the PIP-DCG (principle inpatient diagnosis). Like ACGs, DCGs use ICD9-CM diagnoses to classify patients based on clinical similarity with special attention paid to individuals with expensive chronic conditions. Each ICD9-CM code is mapped to one of 545 DxGroups which are then grouped into 118 condition categories (CCs) (Meenan RT 2003).

A more recent development in diagnosis-based prediction is the global risk adjustment model (GRAM) which includes demographic and diagnostic variables to classify patients using ICD9-CM costs and clinical resource intensity (CRI). Diseases within diagnosis categories are split into terminal classes based on their impact on next year's overall costliness (Meenan RT 2003).

4.4.3 Prior utilisation data

When predicting the risk of hospitalisation in the subsequent 12 months, many models have used prior utilisation data. In many US studies, this is expressed in terms of cost, although some do use the actual number of different medical encounters. Although prior cost/utilisation data may be a powerful predictor, there is caution around using this in the US as there is a concern that this can leave insurers open to manipulation by care providers. However, this is not an issue in the UK currently and, in this project, the NHS is indifferent to the underlying cause of the high utilisation. The literature reviewed does not distinguish between prior costs and the composition of utilisation. For instance, in the models tested it does not appear

to be possible to distinguish between an individual who had four admissions evenly spaced over 12 months and an individual who had four admissions within a period of two months. Some papers have looked at the ability of models to predict readmission within a certain time frame (e.g. Marcantonio ER 1999;Meldon SW 2003) but none have evaluated the relative predictive value of recent utilisation versus less recent utilisation.

There are mixed findings over the relative predictive power of prior utilisation data with some papers maintaining that diagnoses are more predictive. It is clear, however, that prior utilisation/cost data significantly increases the predictive power of a demographic model. Examples include Barneveld's Dutch model where the addition of one year's prior utilisation data increased the R² from 0.03 in the age/sex model to 0.26. The addition of two year's prior utilisation data further increased this figures to 0.44 (van Barneveld EM 1997). Similar findings are presented in a paper by Breyer; the basic age/sex model explained around 5% of variance but the addition of one year's prior utilisation data meant the model then explained around 26% - more than was explained by the addition of diagnoses, which was 11% (Breyer F 2003). Similarly, a Dutch study found prior utilisation data to have more predictive power than diagnosis data in predicting future costs (Lamers LM 2001). Several papers advocate using outpatient utilisation data as well as inpatient data to add to predictive value (e.g.Hughes AS 2004;Lamers LM 2001). It is possible that use of this data allows individuals who have not yet had an acute emergency admission to be identified and this may prevent a future acute emergency admission. The independent predictive value of using outpatient data has not been evaluated, although it is likely to assist in the early identification of individuals who may be in the lower two strata of the Kaiser Permanente risk triangle (figure one). Phase three of the predictive risk project will utilise such data, linked to HES data, in order to test its predictive power.

4.4.4 Pharmacy data

Data on use of pharmaceuticals have been found to be as predictive as, or only slightly less predictive than, data on diagnosis and have an important role to play in early identification of individuals who may not have had an emergency admission. However, pharmacy-based models are not commonly used. One author, Sales, tested the so called 'RxRisk' model (formerly known as the chronic disease score) and, although it was found to be less predictive than diagnosis, it was more predictive than demographics alone (Sales AE 2003). Zhao compared the relative predictive power of inpatient diagnosis versus pharmacy-based information and found the highest R-squared value to be produced when the two are combined (0.118 in the combination model compared with 0.084 for the inpatient diagnosis model and 0.083 for the pharmacy-based model) (Zhao Y 2001).The advantage of using pharmacy data is that they may identify people who are taking multiple medicines who may currently be in a moderate risk group with the potential to move to the higher risk category. If interventions are to be used as primary prevention, data on use of pharmaceuticals may be useful in the early identification of individuals with a high probability of an initial emergency admission. What is more difficult to capture in a regression model is the fact that patients may not comply with their medication regime. It is such behavioural aspects that are problematic to incorporate in a model but may have a significant impact on future risk of service utilisation (Hu G 2004). There is scope to include 'did not attend' data in a model as an indicator of behavioural risk, but this appears to have been used in only one instance and its independent predictive power was not assessed (Cuffel BJ 2002). It is possible that certain pharmaceuticals have greater predictive power than others, although this is not explicitly discussed in the literature. For instance, Sales states that individuals with chronic illnesses such as hypertension and diabetes are frequently prescribed the same medication and this allows the RxRisk model to identify them as potentially high risk (Sales AE 2003) but does not discuss the possibility that certain medications are more associated with high admission rates.

4.4.5 Health status and clinical data

The last two variables - data on health status and other clinical measures - have only been tested in a small number of papers. Data on health status (such as requiring help with activities of daily living) have not been shown to have impressive predictive power compared to data on diagnosis and prior utilisation. Data on

other clinical measures (such as test results like blood pressure and lipid levels) have been included in a number of models but their independent predictive power has not been widely evaluated. Where it has been evaluated, results are inconclusive with one paper finding that it did not enhance predictive power in a model with prior utilisation data (Rector TS 2004) and another finding that a model including clinical variables explained 20% more variance than a model using data on basic demography and prior use (Hendryx MS 2001). The findings of the latter suggest that clinical data have great potential in predicting rehospitalisation in the mental health field. The main reason that the predictive value of clinical data has not been more widely tested is that it does not appear in claims data.

The use of information on health status has produced similarly mixed conclusions. This includes information about physical and social functioning and difficulty with activities of daily living. Information on health status can be collected via questionnaires and in self-report formats. The major limitation to using these data is that they are not routinely collected. Bierman (1999) found that expenditure is inversely related to health status, with individuals reporting poor health status accounting for three times more expenditure than those reporting good health status. This particular study simply asked one single question about health status. This relationship would break down where obstetrics is concerned and so such measures could only be used for certain populations. However, Meldon reported moderately predictive AUC values for a triage risk screening tool that was essentially health status based with values ranging from 0.65 to 0.72 for readmission within 120 days and 30 days respectively (Meldon SW 2003). McCusker reported slightly weaker predictive values (AUC: 0.63 and 0.68) when using the Identification of Seniors at Risk (ISAR) tool to predict early and frequent visits to the emergency department (McCusker J 2000a). Pietz found that the use of the Short Form 36 (SF-36)¹ had a marginal effect on predictive value, having minimal impact on a prospective model but slightly more impact on a concurrent model. Pietz concludes that this marginal predictive value is outweighed by the cost and burden of the required administration and suggests that it could just be used for very high risk patients following initial identification (Pietz K 2004). Similar conclusions were drawn in a study by Mukamel which found that results from questionnaires on health status do not add to predictive power when prior claims are used, although a tool screening for health status alone did explain around 17% of variance (Mukamel DB 1997). Lamers concluded that information on health status can be of good predictive value but that it is often impractical to collect (Lamers LM 1999).

4.4.6 Comparisons

Different papers have explored the predictive power of various types of model but few have sought to compare the whole range of models. In addition, papers have sought to use models to predict different outcomes (e.g. hospitalisation, need for long term care and cost) which makes direct comparison difficult. For instance, both ACGs and DCGs were found to predict costs better than age/gender models alone, but age/gender models were actually better than ACGs at predicting death. This same paper, however, found DCGs to be superior to both ACGs and age/gender models in predicting clinical outcomes (Petersen LA 2005). Fishman found HCCs to compare favourably with ACGs, with each model predicting 15.4% and 10.2% respectively. This paper also found HCCs to be better at picking up the extreme groups (Fishman PA 2003). Compared to RxRisk, HCCs and ACGs are both superior, with RxRisk predicting just 8.7% compared to the 15.4% and 10.2% (Fishman PA 2003). However, as discussed above, Zhao's paper found the predictive value of inpatient HCCs and pharmacy data (Rx) to be similar, and that combining these two variables yielded the most predictive power (Zhao Y 2001).

Sales (2003) found that the DCG/HCC model to have impressive predictive power when used concurrently (45%) in comparison to ADGs (31%), RxRisk (20%) and an age/sex model (1%). In a prospective model, the age/sex model's predictive power remains constant, the RxRisk model's power reduced to 12% and, likewise, the ADG model reduced to 12%. However, the DCG/HCC model experiences the largest reduction in predictive power, with only 15% of variance being explained (Sales AE 2003). Meenan has undertaken the most comprehensive comparison of models; GRAM, DCGs, ACGs, RxRisk and prior-expense were evaluated as to their ability to identify high cost individuals and enrollee groups. Predictive ability varied as to

¹ SF36 - a tool that measure social and physical functioning. Can be filled in by the patient.

the cutpoint used but the overall conclusion was that a the DCG/HCC model can predict future costs at least as well as, in some cases better than, a prior-expense model (Meenan RT 2003). The issue of local variation was raised in one UK- based paper by Allgar. This highlights that the same model can have different predictive power in one location to another. No R-squared values were provided but sensitivity varied from 96% to 99% and specificity varied from 16% to 22% across three hospital sites (Allgar V 2002). Although the variation may appear small here but, when applied to a national context, the variation is likely to increase.

4.5 Modelling with artificial intelligence

In recent years, new models for predicting risk have been developed based on artificial intelligence. These models can utilise neural networks, regression (linear, polynomial, or logistic), decision trees, fuzzy logic, principal component analysis, rule induction, genetic algorithms, nearest neighbours and Kohonen Networks (Axelrod RC 2003). Such models are widely used in the financial, legal and actuarial sectors and are known to have been used by healthcare companies in the US for disease management². Neural networks are the most commonly used type of model in this area. Evidence suggests that models that use neural networks yield a higher predictive power than typical regression models; one study found predictive power to be double that of a traditional regression model (Axelrod RC 2003).

Adoption of systems, such as neural networks, has been slow within the healthcare arena, despite the evidence of impressive accuracy. This stems partly from the fact that there is little information about them in the public arena due to proprietary concerns. It also stems from a concern amongst clinicians about the models' usability. The user of the model is unable to know how exactly the neural networks predict risk and thus the relationship between inputs and outputs. Clinicians must trust that the outputs are medically sound. This is often referred to as the 'black box' phenomenon (Hartnell N 2003). In the context of this project, a neural network model might be able to identify accurately high risk patients but PCTs would not know why these patients had been singled out. It is for these reasons that regression-based predictive models are the most widely used and trusted prediction tools. However, it is clear that artificial intelligence has huge potential to transform the arena of predictive modelling in health and the NHS may benefit hugely from undertaking further research into this particular area.

5. Data availability

From the above summary of the literature, it appears that diagnoses and prior utilisation are the key predictive variables when combined with demographic data. Any model is only as good as the data it uses so it is important to consider what data are available and over what period of time. The majority of models have used one year's prior data (this includes utilisation, diagnoses and demographics). Barneveld tested the use of one, two and three years' prior cost data in three separate models which also varied by costing categories. In all three different models, predictive power was lowest using just one year's worth of data. There was variation in the use of two and three years' data with predictive power increasing significantly in one instance, marginally in another and actually dropping in the third model (van Barneveld EM 1997). The only other paper that discusses this issue suggests that the more years' data that is used, the higher the sensitivity value, but this is at the risk of a reduced specificity value (Rector TS 2004). One of the advantages associated with the use of artificial intelligence models is that their predictive value is high even with an extended 'look forward' period, based on a reduced 'look back' period (Disease Management 2001).

The type of model and number of variables used will inevitably vary with the type of data that are readily available. For example, it would not be efficient to interview every individual within a PCT using a health status questionnaire, so it is important that the tool will be able to use data that are already accessible. US

² E.g. Sentara Healthcare and Anderson Area Medical Centre, South Carolina

models generally use healthcare administrative claims data - these can include data on demography and prior cost, in addition to diagnosis (from ICD-9-CM claim codes), procedures (via Current Procedural Terminology (CPT) claim codes), encounters or utilisation of care (inpatient, outpatient, ER), or on use of pharmaceuticals (via National Drug Code (NDC) claim codes). Information on diagnosis can also be obtained from patient records, as may data on medications. What is not clear from the literature reviewed is whether diagnosis-based methodologies are more powerful when they have a broader dataset to identify diagnoses (e.g. outpatient and GP data). In the UK at present, prescribed medications can be included but there is no way of knowing whether the patient is complying with their medication regime. Information about health status and functionality can only be obtained from contact with the patient. If a tool is to identify patients prior to deterioration in health, it is unlikely that these data can be relied upon. The quality of data is also a vital consideration. A paper by Retchin (1998) highlighted this issue, which is also likely to be prominent when using NHS data. Although data on insurance claims, originally created for facilitating payment, have a number of drawbacks, such as lag time due to administrative reasons, integrity issues such as accuracy, completeness, and "gaming" (or upcoding), the data are commonly used because they are available, reliable, inexpensive, and scalable (Retchin SM 1998;Ridinger MH 2000;Villagra V 2004). Coding problems and lack of electronic data may be a major obstacle to predicting risk accurately when using NHS data. It is important, also, to recognise that PCTs have made different levels of progress towards implementing electronic records and that some PCTs will have access to more information than others. Some of the predictive factors associated with high admissions are found in small area census data. This poses a challenge as these data are only updated every ten years so are often relatively dated.

6. 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 conditions (Wennberg JE 2002) 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 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.

References

Allgar V. **Readmissions - Can they be predicted on admission?** 8[3]. 2002.

Ref Type: Generic

Ash AS. **Using diagnoses to describe populations and predict costs.** Ellis RP and Pope GC. Health Care Financing Review 21[3]. 2000.

Ref Type: Generic

Ash AS. Finding future high-cost cases: comparing prior cost versus diagnosis-base methods. Zhao Y, Ellis RP, and Kramer MS. Health Services Research 36[6]. 2001.

Ref Type: Generic

Axelrod RC. **Predictive modelling in health plans.** Vogel D. Disease Management Health Outcomes 11[12]. 2003.

Ref Type: Generic

Barber M. **Looking for the needle in the haystack: techniques for finding.** Gregorie D. Health Plan . 2001.

Ref Type: Generic

Bhattacharyya SK. **Predicting hospitalisation of patients with diabetes mellitus. An application of the Bayesian discriminant analysis.** Pharmacoconomics 13[5], 519-529. 1998.

Ref Type: Generic

Bierman AS. **How well does a single question about health predict the financial health of Medicare managed care plans?** Bubolz TA, Fischer ES, and Wasson JH. Effective Clinical Practice 2[2], 56-62. 1999.

Ref Type: Generic

Bowns I. Case finding in elderly people: validation of a postal questionnaire. Challis D and Tong MS. British Journal of General Practice 41, 100-104. 1991.

Ref Type: Generic

Breyer F. Determinants of health care utilization by German sickness fund members - with application to risk adjustment. Heineck M and Lorenz N. Health Economics 12, 367-376. 2003.

Ref Type: Generic

Carlson B. **Predictive modeling, sharp lens on near future.** Managed Care 12[7]. 2003.

Ref Type: Generic

Carr-Hill RA. **Risk adjustment for hospital use using social security data: cross sectional small area analysis.** British Medical Journal 16[7334], 390. 2002.

Ref Type: Generic

Castlefields Health Centre. Chronic Disease Management. 2004.

Ref Type: Generic

- Cousins MS. **Introduction to predictive modelling**. Shickle LM and Bander JA. Disease Management 5[3]. 2002.
Ref Type: Generic
- Crichton N. **Receiver Operating Characteristic (ROC) curves**. Journal of Clinical Nursing 11, 134-136. 2002.
Ref Type: Generic
- Cuffel BJ. Predictive models and the effectiveness of strategies for improving outpatient follow-up under managed care. Held M and Goldman W. Psychiatric Services 53[11]. 2002.
Ref Type: Generic
- Dendukuri N. **The identification of seniors at risk screening tool: further evidence of concurrent and predictive validity**. McCusker J and Belzile E. Journal of the American Geriatric Society 52[2], 290-296. 2004.
Ref Type: Generic
- Department of Health. Screening/Case finding: A practical aid to implementation in primary care. 2002.
Ref Type: Generic
- Department of Health. Improving Chronic Disease Management . 2004.
Ref Type: Generic
- Disease Management. **New generation of predictive modelling tools emerging**. Disease Management 6[22]. 2001.
Ref Type: Generic
- 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.
Ref Type: Generic
- 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.
Ref Type: Generic
- Fan VS. **Using quality of life to predict hospitalization and mortality in patients with obstructive lung diseases**. Curtis JR, Tu SP, McDonnell MB, and Fihn SD. Chest 122[2], 429-436. 2002.
Ref Type: Generic
- Fishman PA. **Predicting costs of care using a pharmacy-based measure risk adjustment in a veteran population**. Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, and O'Keefe-Rosetti C. Medical Care 41[1], 84-99. 2003.
Ref Type: Generic
- Fowles JB. **Taking health status into account when setting capitation rates: a comparison of risk-adjustment methods**. Weiner JP, Knutson D, Fowler E, Tucker AM, and Ireland M. Journal of the American Medical Association 276[16], 1316-1321. 1996.
Ref Type: Generic
- Hartnell N. Neural networks: from science fiction to pharmacy. American Journal of Health-Systems Pharmacy 60. 2003.
Ref Type: Generic

Hendryx MS. An examination of methods for risk adjustment of rehospitalization rates. *Mental Health Services Review* 3[1].

Ref Type: Generic

Hendryx MS. An examination of methods for risk adjustment of rehospitalization rates. *Mental Health Services Review* 3[1].

Ref Type: Generic

Hendryx MS. An examination of methods for risk adjustment of rehospitalization rates. *Mental Health Services Review* 3[1].

Ref Type: Generic

Hendryx MS. An examination of methods for risk adjustment of rehospitalization rates. *Mental Health Services Review* 3[1]. 2001.

Ref Type: Generic

Hosmer DW. *Applied logistic regression* 2nd edition. Lemeshow S. 160-167. 2000. New York, John Wiley & Sons, Inc.

Ref Type: Generic

Hu G. **The differences between claim-based health risk adjustment models and cost prediction models.** Lesneski E. *Disease Management* 7[2], 153-158. 2004.

Ref Type: Generic

Hughes AS. **Clinical Risk Groups (CRGs): a classification system for risk-adjusted capitation-based payment and health care management.** Averill RF, Eisenhandler J, Goldfield NI, Muldoon J, Neff JM, and Gay JC. *Medical Care* 42[1], 81-90. 2004.

Ref Type: Generic

Hutt R. Case-managing long term conditions. Rosen R and McCauley J. 2004.

Ref Type: Generic

Lamers LM. **Risk-adjusted capitation based on the Diagnostic Cost Group Model: an empirical evaluation with health survey information.** *Health Service Research* 33[6], 1727-1744. 1999.

Ref Type: Generic

Lamers LM. **Health-based risk adjustment: is inpatient and outpatient diagnostic information sufficient?** *Inquiry* 38[4], 423-431. 2001.

Ref Type: Generic

Liu CF. *Introduction to risk adjustment.* 2004.

Ref Type: Generic

Liu H. **Estimating the Area under a Receiver Operating Characteristic (ROC) Curve For Repeated Measures Design.** Tong W. *Journal of statistical software* 8[12]. 2003.

Ref Type: Generic

Marcantonio ER. **Factors associated with unplanned hospital readmission among patients 65 years of age and older in a Medicare managed care plan.** McKean S, Goldfinger M, Kleefield S, Yurkofsky M, and Brennan TA. *American Journal of Medicine* 107[1], 13-17. 1999.

Ref Type: Generic

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. 2000a.

Ref Type: Generic

McCusker J. **Return to the emergency department among elders: patterns and predictors.** Cardin S, Bellavance F, and Belzile E. Academic Emergency Medicine 7[3], 249-259. 2000b.

Ref Type: Generic

Meenan RT. **The sensitivity and specificity of forecasting high-cost users of medical care.** O'Keefe-Rosetti C, Hornbrook MC, Bachman DJ, Fishman PA, and Hurtado AV. Medical Care 37[8], 815-823. 1999.

Ref Type: Generic

Meenan RT. **Using risk-adjustment models to identify high-cost risks.** Goodman MJ, Fishman PA, Hornbrook MC, O'Keefe-Rosetti C, and Bachman DJ. Medical Care 41[11], 1301-1312. 2003.

Ref Type: Generic

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.

Ref Type: Generic

Mukamel DB. **The effect of accurate patient screening on the cost-effectiveness of case management programs.** Chou CC, Zimmer JG, and Rothenberg BM. Gerontologist 37[6], 777-784. 1997.

Ref Type: Generic

National Primary Care Research and Development Centre. Evercare Evaluation Interim Report: Implications for supporting people with long-term conditions. 2005.

Ref Type: Generic

Newhouse JP. **Adjusting capitation rates using objective health measures and prior utilization.** Manning WG, Keeler E, and Sloss EM. Health Care Financing Review 10[3], 41-54. 1989.

Ref Type: Generic

Olfson M. **Assessing Clinical Predictions of Early Rehospitalization in Schizophrenia.** Mechanic D, Boyer C, Hansell S, Walkup J, and Weiden P. Journal of Nervous and Mental Diseases 187[12]. 1999.

Ref Type: Generic

Parker JP. **Can pharmacy data improve prediction of hospital outcomes? Comparisons with a diagnosis-based comorbidity measure.** McCombs JS and Graddy EA. Medical Care 41[3], 407-419. 2003.

Ref Type: Generic

Petersen LA. **Comparison of the predictive validity of diagnosis-based risk adjusters for clinical outcomes.** Pietz K, Woodard LD, and Byrne M. Medical Care 43[1], 61-67. 2005.

Ref Type: Generic

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.

Ref Type: Generic

Rector TS. **Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions.** Wickstrom SL, Shah M, Greenlee N, Rheault P, Rogowski J, Freedman V, Adams J, and Escarce JJ. Health Service Research 39[6], 1839-1857. 2004.

Ref Type: Generic

- Retchin SM. **Commentary: establishing standards for the utility of administrative claims data.** Ballard DM. Health Services Research 32[6], 861-866. 1998.
Ref Type: Generic
- 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.
Ref Type: Generic
- Ridinger MH. **Predictive modeling points way to future risk status.** Rice JJ. Health Management Technology 21[2], 10-12. 2000.
Ref Type: Generic
- Roblin DW. **A low-cost approach to prospective identification of impending high cost outcomes.** Juhn PL, Preston BJ, Della Penna R, Feitelberg SP, Khoury A, and Scott JC. Medical Care 37[11], 1155-1163. 1999.
Ref Type: Generic
- 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.
Ref Type: Generic
- Rosen AK. **Evaluating diagnosis-based case-mix measures: how well do they apply to the VA population?** Loveland S, Anderson JJ, Rothendler JA, Hankin CS, Rakovski CC, Moskowitz MA, and Berlowitz DR. Medical Care 39[7], 692-704. 2001.
Ref Type: Generic
- Sales AE. **Predicting costs of care using a pharmacy-based measure risk adjustment in a veteran population.** Liu CF, Sloan KL, Malkin J, Fishman PA, Rosen AK, Loveland S, Paul-Nichol W, Suzuki NT, Perrin E, Sharp ND, and Todd-Sternberg J. Medical Care 41[6], 753-760. 2003.
Ref Type: Generic
- Schatz M. **Risk factors for asthma hospitalizations in a managed care organization: development of a clinical prediction rule.** Cook EF, Joshua A, and Petitti D. American Journal of Managed Care 9[8], 538-547. 2003.
Ref Type: Generic
- Schatz M. **Asthma population management: development and validation of a practical 3-level risk stratification scheme.** Nakahiro R, Jones CH, Roth RM, Joshua A, and Petitti D. American Journal of Managed Care 10[1], 25-32. 2004.
Ref Type: Generic
- Selby JV. **Developing a prediction rule from automated clinical databases to identify high-risk patients in a large population with diabetes.** Karter AJ, Ackerson LM, Ferrara A, and Liu J. 24. 2001.
Ref Type: Generic
- 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.
Ref Type: Generic
- Villagra V. **Strategies to Control Costs and Quality: A Focus on Outcomes Research for Disease Management.** Medical Care 42[4]. 2004.
Ref Type: Generic

Wahls TL. **Predicting resource utilization in a Veterans Health Administration primary care population: comparison of methods based on diagnoses and medications.** Barnett MJ and Rosenthal GE. Medical Care 42[2], 123-128. 2004.

Ref Type: Generic

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.

Ref Type: Generic

World Health Organisation. **Innovative Care for Chronic Conditions: Building Blocks for Action.** 2002.

Ref Type: Generic

Zhao Y. **Measuring population health risks using inpatient diagnoses and outpatient pharmacy data.** Ellis RP, Ash AS, Calabrese D, Ayanian JZ, Slaughter JP, Weyuker L, and Bowen B. Health Services Research 36[6]. 2001.

Ref Type: Generic

Zhao Y. **Identifying future high-cost cases through predictive modeling.** Ash AS, Haughton J, and McMillan B. Disease Management Health Outcomes 11[6]. 2003.

Ref Type: Generic

ⁱ <http://www.london.nhs.uk/olderpeople/projects/index.htm>

ⁱⁱ <http://www.london.nhs.uk/olderpeople/projects/index.htm>