evidence for
better health care

nuffieldtrust

Choosing a predictive risk model: a guide for commissioners in England

Key points

- Predictive risk models are used for predicting events such as unplanned hospital admissions, which are undesirable, costly and potentially preventable.
- Such models have been shown to be superior to other 'case finding' approaches, including threshold models and clinical opinion.
- Although the Department of Health has previously funded two predictive models for the NHS in England, the current policy is to promote an open market in terms of suppliers of risk tools.
- Commissioners should consider a range of factors when choosing whether to 'make or buy' a predictive model, including the outcome to be predicted, the accuracy of the predictions made, the cost of the model and its software, and the availability of the data on which the model is run.
- Predictive models should be seen as one component of a wider strategy for managing patients with chronic illness.
- Although there are opportunities here for improving the health status of patients with complex needs while making net savings for the NHS, the evidence for hospital-avoidance interventions is patchy and therefore robust evaluations should be built into any proposed local strategies.
- In the future, it is unclear whether predictive risk models in England should best be procured or built at a local, regional or national level.

Geraint Lewis, Natasha Curry and Martin Bardsley
November 2011

In August 2011, the Department of Health announced that it had no plans to commission national updates of the latest Patients at Risk of Re-hospitalisation tool (PARR++) or the Combined Predictive Model, which are used by the NHS in England. What does this mean for the future of chronic disease management? Predictive modelling is a complex area and there is often confusion about what it is for, what it does and how it works. We have written this short guide to explain some of the key principles involved and to provide an easy reference for people who might be new to this field or who might be required to choose a predictive model for their organisation. We hope it will be particularly useful to clinical commissioners, public health specialists and others involved in the redesign of services for patients with long-term conditions. This is by no means a comprehensive guide, but we will endeavour to keep it updated as new developments emerge.

Why is there so much talk about predictive risk models?

Health care systems in many developed countries are facing similar challenges, including:

- ageing populations
- increasing numbers of people living with long-term conditions
- rising rates of emergency hospital admissions
- financial pressures.

Older people – and younger people living with multiple long-term conditions – often experience high rates of unplanned admission to hospital. Such admissions are distressing for patients and costly to the NHS. Indeed, unplanned hospital admissions account for a considerable proportion of NHS budgets: an estimated £11 billion each year in England. So, if it were possible to predict these admissions and offer preventive care to stop them from occurring, not only would we be improving the health status of these high-risk patients, but we might also make overall savings for the NHS from reductions in unplanned admissions.

Key to this approach, however, is the ability to predict which patients are at risk of having a future unplanned hospital admission – and that is where predictive models come in.

Why not simply offer preventive care to patients who are having frequent admissions?

It might seem tempting to offer preventive care to patients who are currently having frequent hospital admissions. In other words, simply identify those patients who have had many admissions in the last few months and offer them the support of a community matron or another preventive intervention. However, although this might appear to be a sensible, pragmatic and straightforward approach, the trouble is that it is flawed (Roland and others, 2005). Why? Because of a statistical phenomenon called 'regression to the mean' (see Box 1, below).

Box 1: Regression to the mean

The phenomenon of 'regression to the mean' occurs whenever something is measured once and then measured again later. Observations made at the extreme the first time round will tend to come back to the population average the second time round. For example, the warmest place in the UK today is more likely to be relatively cooler tomorrow than warmer.

So, when we look at which people are having frequent hospital admissions at the moment, on average these individuals will have lower rates of unplanned hospital admission in the future *even without intervention*. This point is very important. If you ask a community matron to work with patients who are currently having frequent hospital admissions, the community matron may notice how the patient has fewer admissions over time. However, this reduction might well have occurred anyway due to regression to the mean, and it cannot necessarily be attributed to the input of the community matron.

Why does regression to the mean occur? Simply because after one extreme event, the next event is statistically likely to be less extreme.

Clearly, if a preventive intervention is to be successful and cost-effective, it needs to be offered to people who are at risk of the thing it is trying to prevent – namely a *future* unplanned hospital admission. The purpose of a predictive risk model is to identify which individuals in a population are in fact at risk of unplanned admissions in the future (Duncan, 2011).

Why don't we just ask our clinicians to make predictions?

It would seem intuitive that clinicians such as doctors and nurses, who often know their patients extremely well, would be best placed to make predictions about which individuals are at highest risk of unplanned hospital admission. There are three important theoretical reasons why predictive models may be preferable to predictions made by clinicians. First, predictive models are able to screen whole populations on a regular and repeated basis. This is simply not feasible for a single professional to do. Second, clinicians are unable to make predictions about patients who are not known to them. In contrast, predictive models can take account of patients' contacts with any part of the health care system, as well as other predictive factors such as deprivation and the propensity of different hospitals to admit patients. Finally, clinicians – like all human beings – are susceptible to a whole range of different cognitive biases that make it difficult to translate observation at an individual level into reliable estimations across a population.

The bottom line is that predictive models will be more accurate than clinical opinion (Curry and others, 2005). Indeed, in a recent study, the predictions made by doctors, nurses and case managers were found to be statistically no different from chance (Allaudeen and others, 2011).

Why not use simple referral criteria instead?

A third way of attempting to predict admissions is to use a rules-based approach. Known as 'threshold modelling', this is how patients were recruited for the UK Evercare pilots (Boaden and others, 2006). In these pilots, any patients aged 65 or over who had experienced two or more unplanned admissions in the previous year were eligible for the Evercare service.

A subsequent analysis showed that patients identified in this way are particularly susceptible to regression to the mean (Roland and others, 2005). It is therefore unsurprising that an independent evaluation of Evercare failed to show any reduction in unplanned admissions above and beyond what would have happened anyway due to regression to the mean (Gravelle and others, 2006). As such, although Evercare scored highly against measures of patient satisfaction, it is unlikely it could ever be cost-effective when offered to patients according to this 'threshold model' strategy.

How do I choose a predictive model?

There are various different predictive risk models available to the NHS in England for forecasting a range of health care and social care outcomes. Most of these models aim to predict unplanned hospital admissions – however, they differ in terms of the time period over which they predict (for example, 12 months) and whether they predict single or multiple admissions or readmissions. Moreover, the different models make their predictions based on different sources of routine data, so it is important to choose your model very carefully.

Below we have considered a number of questions that commissioners should ask before deciding to implement any particular predictive model.

Note that the models described in this document are used for 'case finding' purposes. In other words, these are models that seek to identify patients who might be offered a preventive intervention. There also exists a whole range of models for predicting costs. These 'risk adjustment' cost models tend to have lower predictive accuracy than the case finding predictive models because they exclude certain types of data in order to avoid perverse incentives. For details see Nuffield Trust 2011a.

What event should we be aiming to predict?

In principle, predictive risk models can be useful for predicting any event that meets the following four criteria; the event needs to be:

- **Undesirable to the patient**. By predicting such events, it may be possible to offer a preventive service that improves the health status or quality of life of the patient.
- **Significant to the health service** (which usually means costly). For a preventive service to break even, it needs to generate net savings after taking into account the success rate of the intervention and its cost.
- **Preventable**. There is little point investing in attempts to predict an event that cannot be prevented. However, there is strong evidence, for example, that nursing home admissions can be avoided or delayed (Lewis, 2007), and inconsistent evidence that unplanned admissions can be prevented under certain circumstances (Purdy, 2010; Hansen and others, 2011).
- **Recorded in routine administrative data**. Predictive risk models are built by analysing historic data for correlations between the outcome of interest (an unplanned hospital admission, for example) and a range of potential explanatory variables from a prior period. These predictor variables may include age, deprivation, patterns of health service use, and a range of different diagnoses.

It is important to be clear about what it is you want to predict and to ensure that risk prediction is embedded within a coherent strategy for the management of long-term conditions locally (see the section about 'implementation' below).

Examples of outcomes that can be predicted and that conform to the four criteria discussed above include:

Admissions and readmissions

At the moment, predictive models in the NHS are mostly being used to predict unplanned hospital admissions. There are differences between the various models in use regarding the time-window over which they make predictions. For example, many models predict which patients are at risk of admission in the next 12 months – but models could be developed that predict admissions over longer or shorter periods. Another option might be to develop a model that purposefully includes a time delay for its prediction window. This would be to allow time for data lags and for the period necessary for preventive services to recruit and engage with patients or clients. For example, such a model might predict hospital admissions in the next 3 to 15 months, rather than in the next 0 to 12 months. Equally, certain models are designed to predict *multiple* admissions (for example, a model that predicts which patients will have two or more unplanned admissions in the next 12 months) rather than predicting which patients will have a single admission in the next 12 months.

Some predictive models only make predictions for people who meet certain criteria. For example, the PARR model, and its Scottish equivalent, Scottish Patients at Risk of Readmission and Admission (SPARRA), only make predictions of admission in the next 12 months for people who have had some contact with an NHS hospital in the previous three years (Billings and others, 2006; National Services Scotland, 2006). Another example is the PARR-30 model, currently under development, which will only make predictions of admission to hospital within the next 30 days for patients who are currently in hospital (Nuffield Trust, 2011b). The same is true for the LACE model, developed in Toronto (van Walraven and others, 2010).*

Other models are not reliant on such preconditions and instead make predictions for the entire population. For example, the Combined Predictive Model makes predictions of unplanned hospital admission for every person registered with a GP (Wennberg and others, 2006), as does the Predicting Emergency Admissions Over the Next Year (PEONY) model (Donnan and others, 2008) and the Predictive Risk Stratification Model (PRISM) (NHS Wales Informatics Service, 2009).

Specialty-specific admissions

In Scotland, the Information Services Division (ISD) has developed a special predictive model for mental health. This model, called SPARRA-MH†, predicts which people in the population are at risk of admission to a mental health hospital or to a psychiatric unit within a general hospital in the next 12 months (Information Services Division, 2009). Similarly, the $CHADS_2$ model can be used to predict the risk of stroke in patients with non-rheumatic atrial fibrillation (Gage and others, 2001).‡

In theory, analogous models could be developed for other medical specialties and conditions. However, it is important to remember that many of the patients who experience multiple hospital admissions have a combination of different health and social care problems that interact with each other. Preventive interventions should therefore recognise this phenomenon and attempt to take a broad approach to addressing these issues. So, for example, whereas specialist mental health units often look after a distinct population of patients, we would generally caution against choosing predictive models that are too specialty-specific.

---

* LACE stands for: length of stay (L); acuity of the admission (A); co-morbidity of the patient (C); and emergency department use (E).

† SPARRA-MH stands for: Scottish Patients at Risk of Readmission and Admission (Mental Health).

‡ $CHADS_2$ stands for: congestive heart failure (C); hypertension (H); age ≥ 75 years (A); diabetes mellitus (D); and prior stroke or transient ischaemic attack/TIA (S).

Nursing home admissions/intensive social care

The loss of independence that comes from being admitted to a nursing or residential home is often very undesirable and unsettling for the person concerned, and can be extremely expensive. Care home admissions funded by a local authority are often recorded in routine data and there is strong evidence that they can be averted – for example, with an intervention called 'multi-dimensional geriatric assessment and follow-up' (Stuck and others, 2002).

In February 2011, the Nuffield Trust published what we believe is the first predictive model of its type for social care (Bardsley and others, 2011a). Unlike in health care, the 'adverse' outcome to be predicted is not as clear-cut as an unplanned hospital admission. Therefore, for this social care model we developed a hybrid outcome variable. We classified patients as being at high risk if *any* of the following occurred in the target period: admission to a care home (nursing home or residential home); the start of intensive home care (defined as ten or more hours per week or a night-sitting service); or an increase in social care costs above a particular value (£1,000, £3,000 or £5,000, depending on the exact model chosen).

What data do I need?

The data you will require for risk stratification is determined in part by what it is you want to predict. Predictive models that use multiple datasets may be more complex to set up, but they tend to be more powerful and have greater scope to predict different outcomes – although there may be diminishing returns to adding more datasets.

When it comes to predicting unplanned hospital admissions, the most important data source in the NHS in England is the Secondary Uses Service (SUS) dataset (Connecting for Health, 2011). SUS records information about all inpatient admissions, outpatient visits and A&E attendances. SUS has a number of advantages as a data source. First, it is the key dataset in which many events of interest such as hospital admissions are recorded (in other words, SUS is the source of the 'outcome' or 'dependent' variable of interest). Second, it is readily available to NHS organisations. Third, it uses standard coding and recording schemes. Finally, some of the most predictive factors of unplanned hospital admission are to be found within the inpatient records in SUS (in other words, SUS is also a prime source of 'predictor' or 'independent' variables).

Certain models, such as PARR, make their predictions based almost solely on SUS data (PARR also includes a few area-based variables). PARR looks back through SUS for diagnostic and health service use in the preceding three years in order to make predictions about unplanned hospital admissions in the next 12 months. PARR thereby distinguishes between people that have been admitted in the last three years who have a high risk of readmission, from people that have been admitted in the last three years who have a low risk of readmission. As such, it tackles the issue of regression to the mean (as discussed earlier) by making predictions about a future period.

However, models such as PARR and SPARRA are limited in that they can only predict the risk of readmission for people who have had contact with a hospital in the last three years. Because these models only have access to hospital data, they are unable to look at the wider population and identify other people at risk. To be able to make predictions for the whole registered population, it is necessary to incorporate other datasets, such as GP data. GP datasets, which record information in the form of Read codes, constitute a rich clinical record for a person in a population who is registered with a GP. Models such as the Combined Predictive Model (Wennberg and others, 2006) combine GP data with inpatient, outpatient and A&E data from SUS in order to offer a risk assessment of an entire primary care trust (PCT) or practice population.

The downside, however, is that GP data are generally less accessible than SUS. GP data are typically held in individual GP practices rather than being centralised. They also tend to be less standardised, with different GP practices using different Read codes more or less frequently than others. For these reasons, a certain amount of 'data-warehousing' is required before such models can be used. This involves extracting GP data, collating it centrally, and standardising and cleaning the data. As a result, it is important to weigh up the extra effort and expense involved in these processes as part of the overall aim of your long-term condition management strategy against the marginal benefit in predictive accuracy (if any) of incorporating GP data.

Isn't it a case of 'rubbish in, rubbish out'?

Clearly, the quality of data on which a predictive model is built and run will have an impact on the quality of the predictions it makes. Routine administrative databases such as SUS are often criticised for deficiencies in their data quality. However, we should always compare the costs of any inaccurate predictions against the costs of not using the predictions at all. Nor should we forget that one of the best ways to improve the quality of routine data is to use those data in practice.

Equally, it is important to remember that predictive models are built by capturing relationships that exist *within* routine databases. Therefore, if there is a systematic error inside a particular database, as long as this error continues to be made, running the model in the real world should cancel out these errors. For example, if patients with chronic obstructive pulmonary disease (COPD) are systematically miscoded as having asthma rather than COPD, and this miscoded 'asthma' turns out to be predictive of unplanned hospital admissions, then as long as patients continue to be miscoded in this way, those people who are truly at risk of future admission may still be correctly identified by the model.

One of the key considerations when choosing a predictive model is to ensure that it has been validated on the data with which it will be implemented in practice. This way, even if a model is imperfect, these imperfections can at least be quantified and taken into account (see below).

Why do we need to use pseudonymous data?

Predictive models such as PARR and the Combined Predictive Model were constructed using pseudonymous data (Rumbold and others, 2011). These are data where:

- certain variables were truncated (for example, dates of birth were replaced by years of birth, and postcodes were replaced with lower super output areas)
- other variables were removed (names and street addresses)
- the unique key (the NHS number) was replaced by a meaningless but unique pseudonym.

Where data are being obtained from more than one source (for example, from GP practices and from a local community health services provider), data linkage can be facilitated by ensuring that the different organisations agree a common passcode they will use to pseudonymise the data. Using a common passcode will ensure that the same unique key (in this case the NHS number) is converted to the same pseudonym.

Sometimes, however, there may not be a unique key present that is present in all datasets. For example, with models that combine NHS data and social care data, the NHS number may not be present in the social care data. In these circumstances, there are a number of options available. One is to generate a secure alternative identifier in all datasets. When encrypted, this then acts as a common pseudonym across all datasets (Bardsley and others, 2011b).

A few years ago, the Patient Information Advisory Group (PIAG) at the Department of Health published clear guidelines setting out the circumstances in which it was acceptable to use the PARR tool and the Combined Predictive Model (PIAG, 2008). PIAG has since been superseded by the Ethics and Confidentiality Committee of the National Information Governance Board, which is in the process of updating its guidance and is expected to publish new advice on the information governance issues concerning the use of predictive models for social care. Local Caldicott Guardians are another source of advice on information governance.

Choosing a particular model

Once you know what outcomes you want to predict, and what data sources you want to use to predict those outcomes, you should consider the different models available to you that meet these criteria.

What is the difference between a 'model', a 'tool' and a 'platform'?

In general, the terms 'predictive model', 'predictive risk model' and 'risk stratification model' are used interchangeably to mean the mathematical algorithm or the 'inner-workings' that calculate risk scores for individual patients. Technically, these models are usually 'multiple regression models', although sometimes neural networks or decision trees are used instead (see below).

However, in order to operate any algorithm, you need a software 'platform' on which to run it. In other words, the predictive 'tool' refers to the model as well as the platform or software on which the model is run in practice. The software platform may be a web portal, a bespoke software package or a database.

> Predictive tool = Predictive model + Software platform

A tool may sometimes be referred to as being 'plug and play', which essentially means that you can download the entire package, start loading your data immediately and run the model. An example of a 'plug and play' tool is the PARR++ tool, which consists of a bespoke software platform that incorporates the PARR predictive model. In contrast, a model by itself is simply an algorithm or string of code which can only be used if coupled with a software platform. An example of a standalone predictive model is the Combined Predictive Model.

When choosing a predictive model it is important to understand how you plan to run it: do you want to only consider models that are already embedded within their own software (such as PARR)? Or do you want to procure software separately to the model?

In the case of the Combined Predictive Model, some PCTs have built their own software platforms for running the model while others have commissioned outside vendors to create platforms or web portals with which to implement the model. One benefit of procuring the software platform separately is that you can tailor it to the needs and wants of those people who will use it locally, including the design and layout of the reports it will generate. When exploring the market, you are likely to come across the terms 'regression' and 'neural networks/artificial intelligence'. These terms relate to the two main 'families' to which most predictive tools belong (a third, less common family is the decision trees). They refer to the three principal methods used to generate a predictive risk model. Neural network models appear to perform slightly better than multiple regression models

(Winkelman and Mehmud, 2007). However, the perceived complexity of such models can sometimes be off-putting to clinicians.

Where should the model be run?

Alongside the question about software platform, it is important to consider where the model will run in practice. One option is to run the tool at a GP practice level, although to date most tools have commonly been run at the PCT level. Alternatively, a predictive model may be run centrally, with the results sent as secure messages to the end user. This how the SPARRA model is run in Scotland: analysts at the ISD compile the most recent data, run the SPARRA regression model, and send out messages once a quarter to each health board listing which patients are at highest risk of admission or readmission in the next 12 months. A similar approach is taken in Wales where the PRISM model is run centrally and predictions are made available to GPs through a secure website called the Welsh Predictive Risk Service.

When deciding where to run the model, it is important to consider the trade-off between flexibility and the costs of implementation and development. For example, the initial costs of establishing a central or regional tool might be relatively high, but in the long run this option might be more cost-effective than leaving every Clinical Commissioning Group (CCG) to implement its own model. However, implementing a single model across a region or nationally may mean that there is less local flexibility regarding the choice and design of the tool.

How often do predictive models need to be run, recalibrated and rebuilt?

These terms are sometimes used interchangeably, but we suggest the following definitions:

- **Running** the model – this is when an NHS organisation generates the most current list of high-risk patients by applying the model to the most up-to-date data available.
- **Recalibrating** the model – this is when researchers or analysts re-examine the relative weights of the different variables contained in the model to take account of changes in demographics, epidemiology, clinical practice and data coding. A recalibrated model uses the same set of predictor variables as the original model, but each of these variables may now be weighted differently from the previous iteration of the model.
- **Rebuilding** the model – this is when researchers rebuild the model from scratch, using a wide range of candidate variables.

Generally speaking, predictive models should be run on a regular and frequent basis – typically once a month. Depending on the systems in place, running the model more often than this can place an unsustainable administrative burden on NHS analysts and could cause confusion if patients' risk scores fluctuated wildly. In contrast, however, running the model less often than once a month may be problematic: patients may have died before being contacted or they may have begun regressing to the mean, and therefore would be at lower risk of hospitalisation than anticipated.

We have previously suggested that predictive models should be recalibrated or rebuilt every two years or so in order to ensure that their predictive accuracy does not deteriorate (Nuffield Trust, 2011a). However, recent research by David Osborne, Senior Public Health Information Analyst at NHS South West London, and subsequently by Todd Chenore, Senior Information Analyst at NHS Devon, has shown that this degradation in predictive accuracy appeared not to have occurred in practice (Osborne, 2011; Chenore, 2011).

Assessing a model's performance

Clearly, one of the most important factors with a predictive model is its predictive accuracy, as discussed below. However, it is worth bearing in mind that there may be a balance between the predictive accuracy of a model and the 'importance' of the outcome the model is designed to predict. In other words, certain outcomes may be so important to commissioners that they may be willing to tolerate lower levels of predictive accuracy from a model. Examples of 'important' outcomes might include a model that predicted the onset of diabetes in 15 years' time, or a model that predicted admission to a nursing home. In these cases, even relatively inaccurate predictions could still be very useful to commissioners and clinicians.

How do researchers measure the accuracy of their model?

A variety of different metrics can be used to assess a model's performance. Whatever measure is used to gauge predictive accuracy, however, it is important that this assessment be conducted rigorously. For example, the PARR model was built on a ten per cent sample of hospital episode statistics data for England. The newly-built PARR was then tested against a separate ten per cent sample of data, and it is the model's performance on this separate dataset that was reported. This approach – known as a split-sample method – is designed to ensure that the model performs well in practice on real-life data, that is, to ensure that the model is 'generalisable' and does not 'over-fit' the data on which it was built. An alternative to the split-sample method is to use 'bootstrapping', where repeated samples are drawn from the data and a correction is applied to take account of the phenomenon of 'optimism' (Gail and others, 2009).

What measures should we use to gauge predictive accuracy?

A variety of metrics may be used to determine the predictive accuracy of a model. These include:

- **R-squared** – a commonly used statistical term that measures the explanatory power of a model. Values range from 0 to 1, generally the higher the better. As such, it provides an overall measure of how well the model predicts future outcomes.
- **Positive predictive value (PPV)** – for any given predictive risk score threshold, this is the proportion of patients who are identified by the model as being 'high risk' that will truly experience the outcome being predicted. As such, it is a particularly useful metric when determining a business case for a preventive intervention. A high PPV means that a high proportion of the patients being offered the intervention would, without intervention, have experienced the costly adverse outcome that the intervention seeks to prevent. In contrast, with a lower PPV, many of the patients identified by the model would not have experienced the outcome in any case, and so in this sense the intervention is 'wasted' on these individuals. Of course, there may still be good reasons for offering a preventive intervention to people where the PPV is low, but it is important that the cost of the intervention should also be relatively low otherwise it will be impossible for the intervention to break even.
- **Sensitivity** – for any given risk score threshold, this is the proportion of the population who will experience the outcome of interest that the model successfully identifies. For example, a model might have a sensitivity of 40 per cent for a risk score threshold of 35. This means that if an intervention is offered to every person with a risk score of 35 or above, then 40 per cent of the people in the population who would be having an unplanned hospital admission next year will now be offered the intervention.
- **Specificity and negative predictive value (NPV)** – these two metrics relate to the ability of the model to predict which patients will *not* have a future unplanned admission. Specificity and NPV are analogous to the sensitivity and PPV, respectively. However, because the vast majority

of the population will *not* have an unplanned hospital admission in the next 12 months, so the specificity and NPV are not particularly useful metrics in this context.

- **C-statistic** – also known as the 'area under the curve' or AUC, this is the area under the receiver operating characteristics (ROC) curve, which displays the trade-off between sensitivity and specificity for a predictive model. The area under the ROC curve is an aggregate number that reflects the distribution of sensitivities and specificities across all risk scores. Like the r-squared, the c-statistic is useful because it allows comparisons of different models based on a single number. However, it is not very intuitive and, in reality, commissioners may only be interested in a certain portion of the ROC curve, rather than the average which the c-statistic reflects.

It is important to note that a model's PPV and its sensitivity can be traded off against each other. Selecting a high risk score threshold (only offering the intervention to people with a very high risk score) will lead to a high PPV but a low sensitivity. Conversely, selecting a lower risk score threshold will diminish the PPV of the model, while increasing its sensitivity.

Model developers should be encouraged to publish the predictive accuracy of their models using standard metrics, and it is important to understand what level of accuracy you can expect if you are to procure a model, as well as which data were used to validate the models. A recent systematic review compared the accuracy of a range of published predictive models (Kansagara and others, 2011) and, in the United States (US), the Society of Actuaries assesses the predictive accuracy of a range of models on a set of test data (Winkelman and Mehmud, 2007).

Table 1: Measuring model sensitivity, specificity and PPV

		True outcome	
		Admitted to hospital	***Not admitted to hospital***
Predicted outcome	***High risk***	True positive (TP)	False positive (FP)
	Low risk	False negative (FN)	True negative (TN)

Sensitivity = TP / (TP + FN)
Specificity = TN / (FP + TN)
Positive predictive value = TP / (TP + FP)
Negative predictive value = TN / (FN + TN)

Should we ask our clinicians what they think of the model?

Clearly, it is important to engage local clinicians when implementing a predictive model: they need to understand the value and shortcomings of the model, and ideally should be engaged in the design of the software platform, as well as the design of the intervention or interventions to be offered to patients at high predicted risk. However, while the 'face validity' of a model may be interesting (in other words, whether it makes intuitive sense), we should remember that it has been shown that clinicians do not make accurate predictions of future hospital admissions (Allaudeen and others, 2011). For this reason, clinicians may be surprised at some of the predictions made by a predictive model. Remember that predictive models are designed to identify individuals whose risk is rising and who might not yet have come to the attention of their clinician. Also, note that at this stage, we are

simply interested in gauging the accuracy of the model, not whether the clinicians think that a particular patient is amenable to preventive care – the so-called 'impactibility' of a patient, where clinical characteristics may be critical (Lewis, 2010).

In some parts of the country, NHS organisations have viewed the opinions of clinicians as the 'gold standard' against which to benchmark the predictions of a model. In these organisations, when the predictions of a predictive model differed from those made by the clinicians, it was assumed that the clinicians were correct and the model was wrong. Our advice would be instead, where possible, to use the reality of who was actually admitted to hospital as the 'gold standard' and to compare the predictions of the model (and clinicians) against this benchmark.

Cost

What is the cost of running a predictive tool?

The cost of running a predictive tool consists of the following elements:

- **Cost of the predictive model** – for many open-source models there is no 'licence' cost for running the model, and the intellectual property behind these models is freely available. Examples are the Combined Model, PARR-30, the Nuffield Trust social care models and the HCC (Hierarchical Condition Categories) model (Pope and others, 2004). In contrast, for proprietary models, a licence fee is generally payable.
- **Cost of the software** – proprietary models typically come bundled with their own software. For open-source predictive models, the costs vary depending on whether or not there is free accompanying software. For example, the software on which the PARR model was run (called PARR++) could previously be downloaded by NHS organisations free of charge. In Wales, the PRISM model is run centrally and results are made freely available through a secure website. Likewise, in Scotland, the SPARRA model is run centrally free of charge but here the results are sent as secure messages. For other open-source models, such as the Combined Model or the Nuffield Trust social care models, there is no accompanying software. NHS organisations must therefore either pay to develop a software platform in-house, or they must purchase software from an outside vendor.
- **Cost of obtaining the data** – certain data sources are freely available to NHS organisations. These include SUS data (inpatient, outpatient and A&E); census data (index of multiple deprivation); Exeter data (a list of residents registered with local GPs); and, where available, community services data (district nursing, community physiotherapy, etc.) and social care data (needs assessments and service provision). In contrast, there is typically a cost involved in obtaining GP 'Read code' data. Various options exist for extracting such data – either through the developer of the GP clinical IT system, or by an external third party. In England, the NHS Information Centre is currently establishing a GP extraction service ('GPES'). So, in the future it may be possible to obtain GP data more easily.
- **Labour costs** – there are several types of labour costs to consider. These include the time taken to set up the system and to engage with local GPs and explain how the model works. The other major cost is that associated with running the model – refreshing the data and producing scores. In general, once the system is in place, the more user-friendly the model, the lower the labour costs of running it.

Implementation

Although all of the technical considerations regarding model accuracy, data availability and so on are very important, there is a danger that the procurement of a predictive model is seen as merely a technical problem with a purely technical solution. It is not. A predictive model should be regarded as just one (albeit very important) part of a wider strategy in managing the health of a population.

Usually, the aim is to reduce emergency admissions by better managing people with long-term conditions. However, a local area might choose a more specific aim than this. For example, a local strategy might focus on a particular population subgroup or a particular condition (sickle cell disease, for example). The ultimate aim of the strategy should be clearly articulated because it will dictate what constitutes 'better management' and therefore, in turn, will influence the choice of which predictive model is most appropriate.

People often ask how effective a particular predictive model is at reducing hospital admissions. The answer, of course, is always 'not at all'. A predictive model can only tell you which patients are at risk of a particular event (for example, readmission in the next 30 days or admission to a nursing home in the next 12 months). What it can never do is manage a patient and prevent their deterioration or admission. This might seem obvious, but the point is that the choice of model needs to be embedded within a wider strategy.

Key to the success of a long-term conditions strategy is the efficacy, equity and cost-effectiveness of the intervention or interventions being offered to patients at high predicted risk. If a hospital avoidance scheme is to make net savings for the NHS, the cost of that intervention per patient must be lower than the average expected cost of unplanned hospital admissions for those patients to whom it is being offered. There exists a trade-off between the cost of the intervention, the risk scores of the patients, and the effectiveness of the intervention in preventing unplanned hospital admissions. For example, in an article describing the PARR model, Billings and others (2006) present a table that outlines a range of potential business cases based on a combination of:

- **Risk score thresholds** – for example, choosing a risk score threshold of 70 would mean that all patients with a PARR score greater than 70 would be offered the preventive service.
- **Assumed reduction in admissions** – for example, an assumed reduction of ten per cent means that patients receiving the intervention would be expected to have ten per cent fewer admissions than would otherwise have occurred.
- **Cost of the intervention**.

Purdy (2010) has compiled a useful summary of the evidence for a range of hospital-avoidance schemes, and Hansen and colleagues (2011) have published a systematic review of interventions to reduce 30-day readmissions. Several major evaluations of other hospital-avoidance interventions are currently underway, however, including evaluations of telehealth and telecare, virtual wards, community matrons and integrated care organisations (Nuffield Trust, 2011c; Clinicaltrials.gov, 2010; Lewis and others, 2011; New York State, 2008).

Who to target?

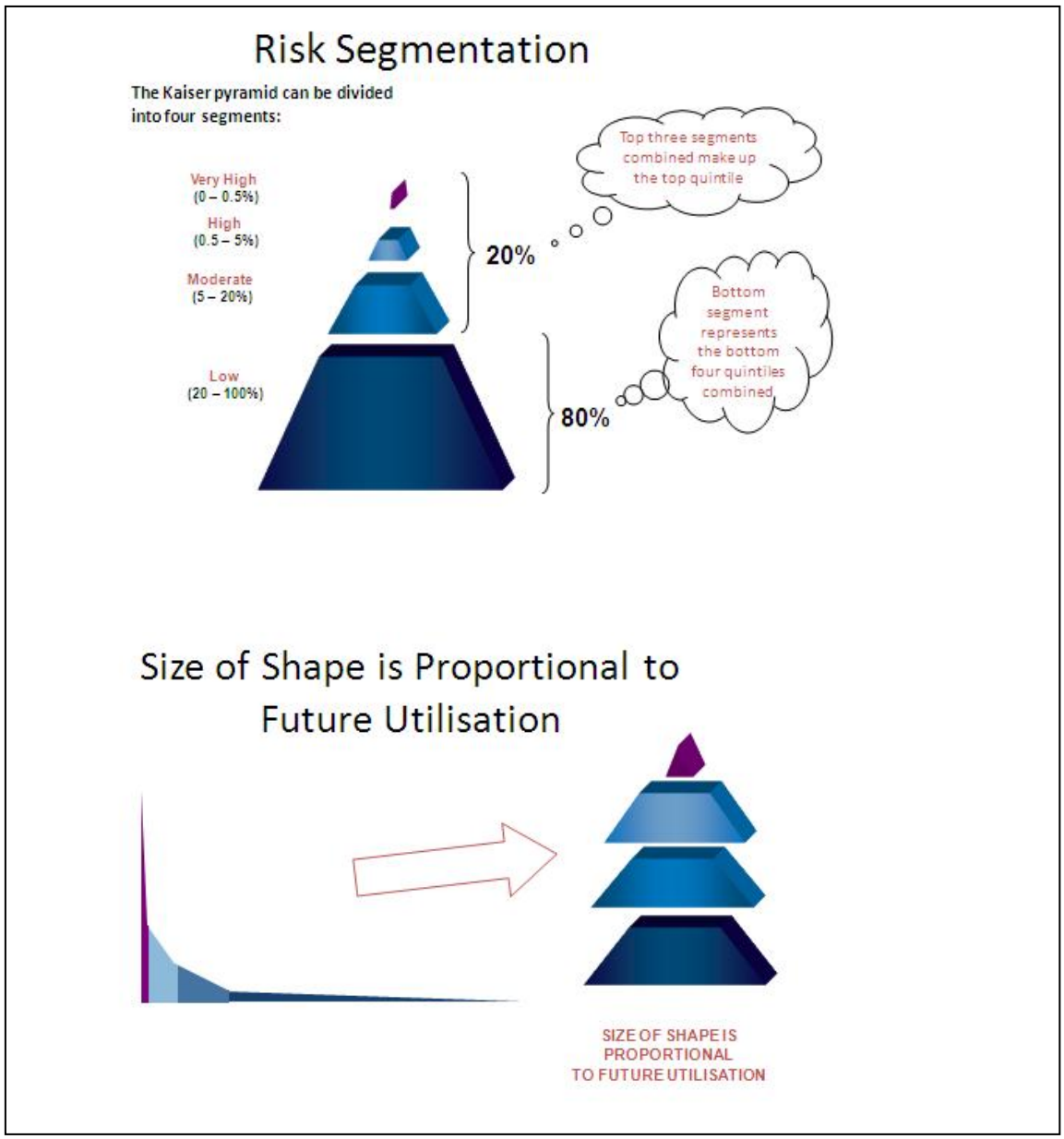
The population on which the intervention should focus will depend upon the local strategy for managing patients with long-term conditions (as discussed above), the availability of data and the assumed cost-effectiveness of any intervention.

Individually, people with high predicted risk scores present the greatest opportunity for making savings from averted hospital admissions. However, there are relatively few of these people so it is

essential that they are accurately identified. Enrolling people who are not in reality very high risk in intensive interventions such as community matrons or virtual wards is not cost-effective. Targeting interventions at larger populations at lower risk might offer more opportunity to intervene early but, the lower the risk threshold chosen, the cheaper and/or more effective the intervention needs to be.

The well-known 'Kaiser Pyramid' illustrates the distribution of risk across a typical population (see Figure 1, below). Predictive models are able to assign names (or, more strictly, pseudonyms) to next year's Kaiser Pyramid. What the pyramid shows is that the very high-risk group are a tiny proportion of any given population, but that they each account for a disproportionately large share of future service utilisation. Moving down the pyramid, the population size increases so although each individual accounts for a smaller proportion of future utilisation than those in the high-risk category, in aggregate these lower-risk populations will represent a greater proportion of future utilisation because there are far more such people. It is important, therefore, that any intervention be targeted carefully at the right population after having taken account of the expected cost-effectiveness of the intervention.

Figure 1: Risk segmentation according to the 'Kaiser pyramid'



Roemer's law

Roemer's law states that a 'hospital bed built is a hospital bed filled' (Ginsburg and Koretz, 1983). In other words, should hospital avoidance interventions be successful, it is likely that any financial gains made will be undermined by local hospitals now admitting lower-risk individuals. Commissioners need to be mindful of this phenomenon and take steps to monitor and mitigate the risk.

Who should be involved in implementation?

There has been a tendency in the past, as mentioned above, for people to regard risk prediction as a technical solution to a technical problem. As a result, the procurement or development of predictive models has, on occasion, been left to IT staff or to a lead commissioner who then presents it to the clinical teams as a *fait accompli*. In these cases, it is unsurprising if GPs and other clinicians are unwilling to use the model or to engage with a wider programme of long-term condition management because they may not share the vision and do not see the value.

We believe it is important for a range of people to be involved from the outset in the development or procurement of a predictive model locally, so that it is embedded not only in the technical systems but also into the organisational processes, working practices and culture of the clinicians that will use the model in everyday practice.

Who should commission predictive tools?

Careful consideration needs to be given as to who should be involved in commissioning a new predictive tool in England and at what scale. In the past, PCT staff have tended to take a lead role in the implementation of such tools. However, following the abolition of PCTs, it is currently unclear who will take over this role. Emerging CCGs will need to give careful thought to the scale at which any predictive tools are commissioned. For small CCGs it may make financial sense to procure a model in partnership with other CCGs, or even regionally. The NHS Commissioning Board may also wish to consider the national procurement of a model or models that can be implemented locally. However, the economies of scale to be gained from regional or national procurement need to be weighed against the advantages of procurement at a local level. For example, where local professionals have more input into the design and development of a model, they may perceive greater ownership over it.

As discussed, the accuracy of a predictive model can be calculated by the model developers using techniques such as split-sample and bootstrapping. However, what commissioners and clinicians should really want to know is the efficacy, cost-effectiveness and equity of the local long-term condition management strategy. The evaluation of such a strategy should include both an assessment of the accuracy of the model's predictions plus an evaluation of the intervention offered on the basis of those predictions.

It is important to think about evaluation and monitoring at the outset: evaluation should be embedded within the long-term condition management strategy and it should set out some clearly defined desired outcomes. As with any evaluation, it is important to establish a valid comparator group. Cost saving is likely to be a key outcome of interest, but other factors – such as patient experience, health outcomes and health inequalities – might also feature in any evaluation.

What is the future of predictive modelling likely to hold?

Following the announcement that the Department of Health will leave predictive modelling development up to the free market (Johnson, 2011), the intention is that existing companies and universities involved in the field will innovate further; that new vendors will enter the market; that commissioners will be offered more choice of models; and that competition will drive prices down. Whether any or all of this will be borne out in reality remains to be seen.

In many parts of the country, NHS employees have developed their own predictive models (Chenore, 2011) or software platforms on which to run existing models (Binns, 2011). Aside from these in-house options, the wider market is already relatively busy, with a range of commercial companies and non-commercial organisations having developed their own models and tools. There have been some calls for the NHS Commissioning Board to undertake national procurement of a new PARR/Combined Predictive Model style model, but this seems unlikely to be approved given the current emphasis on the market for information tools and analysis.

The market and politics aside, we can look to the US to offer us an insight into where the science of predictive modelling might develop in the coming years. The insurance-based system in the US means that it is ahead of the NHS in developing predictive models. After all, predicting the future is core business for the insurance world. So, many of the lessons applied in the NHS in recent years originated in the US. We can, therefore, look at recent developments there, including the development of so-called 'impactibility models' that seek to identify the subgroup of high-risk patients who are most amenable to preventive care (Lewis, 2010).

Conclusions

As commissioners face growing financial pressures, risk prediction is likely to take an ever more central role in targeting investment and reducing unplanned hospital admission rates. The Department of Health's decision not to fund a national update to its centrally procured models means that commissioners in England will need to take a much more active role around risk prediction. Previously there were standard, proven, cost-free models that were backed by the Department of Health. Now, however, commissioners will begin navigating through what may become a highly competitive free market with all the advantages and pitfalls that that entails. It is essential that commissioners be well informed and prepared when they start to research this emerging market and that they ask the right questions.

This short guide has attempted to set out some of the key considerations that any commissioner should bear in mind when procuring a predictive risk model. It is intended as a starting point for localities to begin discussions with local stakeholders rather than a comprehensive 'how to' guide.

References

Allaudeen N, Schnipper JL, Orav EJ, Wachter RM and Vidyarthi AR (2011) 'Inability of providers to predict unplanned readmissions', *J Gen Intern Med* (26)7, 771–776.

Bardsley M, Billings J, Dixon J, Georghiou T, Lewis GH and Steventon A (2011a) 'Predicting who will use intensive social care: case finding tools based on linked health and social care data', *Age Ageing* 40(2), 265–270.

Bardsley M, Billings J, Chassin LJ, Dixon J, Eastmure E, Georghiou T, Lewis G, Vaithianathan R and Steventon A (2011b) *Predicting Social Care Costs: A feasibility study*. London: Nuffield Trust.

Billings J, Dixon J, Mijanovich T and Wennberg D (2006) 'Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients', *BMJ* 333, 327.

Binns I (2011) 'Reducing non elective admissions using the Combined Predictive Model', NHS Blackpool. Available at http://glasgows.co.uk/hicat/docs/Ian_Binns.pdf

Boaden R, Dusheiko M, Gravelle H, Parker S, Pickard S, Roland M, Sargent P and Sheaff R (2006) *Evercare Evaluation: Final report*. Manchester: National Primary Care Research and Development Centre.

Chenore T (2011) Personal communication to the authors.

ClinicalTrials.gov (2010) 'A virtual ward to reduce readmissions after hospital discharge'. Available at http://clinicaltrials.gov/ct2/show/NCT01108172

Connecting for Health (2011) *The SUS Programme*. Available at www.connectingforhealth.nhs.uk/systemsandservices/sus

Curry N, Billings J, Darin B, Dixon J, Williams M, Wennberg D (2005) *Predictive Risk Project Literature Review*. London: The King's Fund.

Donnan PT, Dorward DWT, Mutch B and Morris AD (2008) 'Development and validation of a model for predicting emergency admissions over the next year (PEONY)', *Arch Intern Med* 168(13), 1416–1422.

Duncan I (2011) *Healthcare Risk Adjustment & Predictive Modeling*. Winsted: ACTEX Publications.

Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW and Radford MJ (2001). 'Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation', *JAMA* 285(22), 2864–2870.

Gail M, Krickeberg K, Samet JM, Tsiatis A and Wong W (2009) 'Overfitting and optimism in prediction models', in *Clinical Prediction Models*. New York: Springer, 1–18.

Ginsburg PB and Koretz DM (1983) 'Bed availability and hospital utilization: estimates of the "Roemer effect"', *Health Care Financing Review* 5(1), 87–92.

Gravelle H, Dusheiko M, Sheaff R, Sargent P, Boaden R, Pickard S, Parker S and Roland M (2006) 'Impact of case management (Evercare) on frail elderly patients: controlled before and after analysis of quantitative outcome data', *BMJ* 334, 31.

Hansen LO, Young RS, Hinami K, Leung A, and Williams MV (2011) 'Interventions to reduce 30-day rehospitalization: a systematic review', *Ann Intern Med* 155(8), 520–528.

Information Services Division (2009) *SPARRA Mental Health: Scottish patients at risk of readmission and admission (to psychiatric hospitals or units)*. Edinburgh: NHS National Services Scotland.

Johnson S (2011) Letter to PCT Chief Executives. Available at www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_129005.pdf

Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M and Kripalani S (2011) 'Risk prediction models for hospital readmission: a systematic review', *JAMA* 306(15), 1688–1698.

Lewis G (2007) *Predicting Who Will Need Costly Care: How best to target preventive health, housing and social programmes.* London: The King's Fund.

Lewis G (2010) "'Impactibility models": identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs', *Milbank Q* 88(2), 240–255.

Lewis G, Bardsley M, Vaithianathan R, Steventon A, Georghiou T, Billings J and Dixon J (2011) 'Do "Virtual Wards" reduce rates of unplanned hospital admissions and at what cost? A research protocol using propensity matched controls', *Int J Integr Care* 11.

National Services Scotland (2006) *SPARRA: Scottish patients at risk of readmission and admission.* Edinburgh: NHS National Services Scotland. Available at http://www.isdscotland.org/Health-Topics/Health-and-Social-Community-Care/SPARRA/

New York State (2008) Chronic illness demonstration projects: request for proposals. Albany, NY: New York State Office of Health Insurance Programs. Available at www.health.ny.gov/funding/rfp/inactive/08010310003/0801031003.pdf

NHS Wales Informatics Service (2009) *GP practices trial tool to identify patients at risk.* Available at www.wales.nhs.uk/nwis/news/17929

Nuffield Trust (2011a) *Predictive Risk and Health Care: An overview.* London: Nuffield Trust.

Nuffield Trust (2011b) *Predicting Risk of Hospital Readmission with PARR-30.* Available at www.nuffieldtrust.org.uk/our-work/projects/predicting-risk-hospital-readmission-parr-30

Nuffield Trust (2011c) *Evaluation.* Available at www.nuffieldtrust.org.uk/our-work/evaluation.

Osborne D (2011) Personal communication to the authors.

Patient Information Advisory Group (2008) *The Use of Patient Information in the Long Term Conditions Programme.* Available at www.advisorybodies.doh.gov.uk/piag/piag-ltc-oct2008.pdf

Pope G, Kautter J, Ellis R, Ash A, Ayanian J, Iezzoni L, Ingber M, Levy J and Robst J (2004) 'Risk adjustment of Medicare capitation payments using the CMS-HCC Model', *Health Care Financing Review* 25(4), 119–141.

Purdy S (2010) *Avoiding Hospital Admissions: What does the research evidence say?* London: The King's Fund.

Roland M, Dusheiko M, Gravelle H and Parker S (2005) 'Follow up of people aged 65 and over with a history of emergency admissions: analysis of routine admission data', *BMJ* 330, 289.

Rumbold B, Lewis G and Bardsley M (2011) *Access to Person-level Data in Health Care: Understanding information governance.* London: Nuffield Trust.

Stuck AE, Egger M, Hammer A, Minder CE and Beck JC (2002) 'Home visits to prevent nursing home admission and functional decline in elderly people: systematic review and meta-regression analysis', *JAMA* 287, 1022–1028.

van Walraven and others (2010) 'Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community', *CMAJ* 182(6), 551–557.

Wennberg D and others (2006) *Combined Predictive Model: Final report and technical documentation*. London: The King's Fund.

Winkelman R and Mehmud S (2007) 'A comparative analysis of claims-based tools for health risk assessment', Society of Actuaries. Available at www.soa.org/files/pdf/risk-assessmentc.pdf

evidence for
better health care

nuffieldtrust

For more information about the Nuffield Trust, including details of our latest research and analysis, please visit www.nuffieldtrust.org.uk

Download further copies of this paper from www.nuffieldtrust.org.uk/publications

Subscribe to our newsletter: www.nuffieldtrust.org.uk/newsletter

Follow us on Twitter: Twitter.com/NuffieldTrust

Nuffield Trust is an authoritative and independent source of evidence-based research and policy analysis for improving health care in the UK

59 New Cavendish Street
London W1G 7LP
Telephone: 020 7631 8450
Facsimile: 020 7631 8451
Email: info@nuffieldtrust.org.uk

www.nuffieldtrust.org.uk

Published by the Nuffield Trust.