SPARRA: Scottish Patients At Risk of Readmission and Admission

Version 1.1a  07 August 2006

Delivering for Health Information Programme

Information Services Division
Delivering for Health Information Programme Team:

Eddie Adie (Project & Information Manager)
David Baird (Senior Information Analyst)
Penny Bridger (Consultant in Public Health Medicine)
Scott Buchanan (Information Analyst)
Charles Guthrie (Senior Information Analyst)
Joanne Hattie (Information Analyst)
Steve Kendrick (Programme Principal)
Steven Snedker (Information Analyst)
Jack Vize (Senior Information Analyst)

Contacts:

To contact any of the Delivering for Health Programme team, please email:

dhip@isd.csa.scot.nhs.uk

Version History and Revisions:

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Revision</th>
<th>Actioned</th>
<th>Approved</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>30 June 2006</td>
<td>Final Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>17 July 2006</td>
<td>Executive Summary added.</td>
<td>Eddie Adie</td>
<td>Penny Bridger</td>
</tr>
<tr>
<td>1.1a</td>
<td>07 August 2006</td>
<td>Contact details amended.</td>
<td>Eddie Adie</td>
<td>Penny Bridger</td>
</tr>
</tbody>
</table>
Contents.

1. Executive Summary. ........................................... 1
2. Introduction. .................................................. 2
3. Policy context. ................................................ 3
4. Methodological context. ...................................... 9
5. The SPARRA approach. ....................................... 11
6. Overview of developmental phases. ......................... 13
7. Historic admission based model: overview. ............... 15
8. The risk prediction model. ................................... 17
9. Results. ....................................................... 23
10. Performance of the model. ................................. 32
11. Feedback of results. ......................................... 37
12. Future development. ......................................... 39
13. Conclusions. ................................................ 43

Acknowledgments. ............................................... 44
References. ....................................................... 45
1. Executive Summary

- Scottish Patients at Risk of Readmission and Admission (SPARRA) is a risk prediction algorithm, developed by the Information Services Division (ISD) to identify patients aged 65 years and over at greatest risk of emergency inpatient admission.

- The algorithm uses selected information from patient histories of hospital admission to predict their risk of future emergency inpatient admission. It will identify the majority of patients at highest risk of emergency admission, i.e. where the predicted probability of emergency inpatient admission in the next year is 60% and above.

- Feedback of information generated by the algorithm will take two forms: first, aggregate risk distributions and secondly, details of individual patients with a high predicted probability of emergency inpatient admission.

- Aggregate risk distributions show at NHS board or Community Health Partnership (CHP) how many patients are at given levels of risk of future emergency inpatient admission. This will appear on the ISD website (www.isdscotland.org/dhip) to facilitate strategic and local planning in the provision of healthcare and social services.

- Information on individual patients identified by SPARRA can be made available for NHS boards/CHPs to review and assess, allowing them to decide which patients may benefit from preventive coordinated care in the community to prevent avoidable emergency inpatient admissions. The SPARRA team at ISD will provide information about individual patients after discussion with NHS boards/CHPs to ensure that details supplied are in the most useful and appropriate format.

- Future developments of the SPARRA model include incorporating real time hospital admission and primary care data. These developments are dependent on close collaboration between the SPARRA team and key individuals in NHS boards/CHPs who are keen to take this work forward.
2. **Introduction.**

This report describes the first version of SPARRA (Scottish Patients at Risk of Readmission and Admission), a risk prediction algorithm being developed by the Information Services Division (ISD Scotland).

The algorithm aims to identify those patients aged 65 and over who are at greatest risk of emergency inpatient admission. Once identified the patients can be assessed for more co-ordinated and preventive patterns of care aimed at preventing future avoidable emergency admissions.

This first version of the algorithm is based on patient histories of hospital admission held at ISD Scotland. The approach has been pragmatic, aimed at producing relatively quickly a prediction model, which is adequate rather than perfect.

Future developments will involve the incorporation of more up-to-date admissions data and, in collaboration with local teams, primary care data. This will allow more accurate prediction of future risk and more timely feedback.

There is however a momentum in Scotland, which is rapidly gathering behind the piloting and implementation of intensive case management and care co-ordination. It was therefore considered important to provide details of high-risk patients as quickly as possible.

Preliminary analysis suggests that SPARRA, even though it is based on centrally held hospital admission data which are usually complete after around six months, will identify all but a relatively small proportion of those patients at highest risk of emergency admission and who are potentially identifiable on the basis of current data sources.

This report presents the policy and methodological context of the model. Most important however is the description of how the model works. The intention of the SPARRA team is to take this agenda forward in collaboration with local teams. This can only be achieved if there is a shared understanding of what risk prediction involves.

In addition, by providing as much information as possible on which aspects of patient histories have most impact on the likelihood of future admission, it is hoped that the model will help to inform the process of developing the best ways of caring for the most vulnerable groups in the older population.
3. Policy context.

*Increasing emergency admissions among the old.*

In 2005, the National Framework for Service Change in the NHS in Scotland (the Kerr Report) and the Ministerial response, Delivering for Health, called for a fundamental reorientation in the way healthcare is delivered in Scotland (NHS Scotland, 2005; Scottish Executive 2005a).

Figure 1 outlines the dimensions of the paradigm shift that is needed. In summary, there needs to be a shift from a system which is geared towards the hospital-based treatment of (isolated) acute conditions, admissions most often made in response to sudden crises, to a system which is founded on a preventive, anticipatory approach to the management of long-term conditions on a whole-person rather than a disease basis.

**Figure 1**

<table>
<thead>
<tr>
<th>Current Model</th>
<th>Evolving Model of Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geared towards acute conditions</td>
<td>Geared towards long-term conditions</td>
</tr>
<tr>
<td>Hospital centred</td>
<td>Locally responsive</td>
</tr>
<tr>
<td>Doctor dependent</td>
<td>Team based</td>
</tr>
<tr>
<td>Episodic care</td>
<td>Continuous care</td>
</tr>
<tr>
<td>Disjointed care</td>
<td>Integrated care</td>
</tr>
<tr>
<td>Reactive care</td>
<td>Preventative care</td>
</tr>
<tr>
<td>Patient as passive recipient</td>
<td>Patient as partner</td>
</tr>
<tr>
<td>Self care tolerated</td>
<td>Self care encouraged and facilitated</td>
</tr>
<tr>
<td>Carers undervalued</td>
<td>Carers supported as partners</td>
</tr>
<tr>
<td>Low tech</td>
<td>High tech</td>
</tr>
</tbody>
</table>

Steadily rising numbers of emergency (unplanned) inpatient admissions, particularly among older patients, have been the major source of pressure on the NHS in Scotland over the past twenty years. All of the increase in bed days for emergency inpatients over the past twenty to twenty-five years has been accounted for by patients aged 65 and over, with the vast majority being accounted for by patients aged 80 and over.

Conventional wisdom has long held that these rising numbers of emergency admissions among older people were an inevitable consequence of an ageing population.
Increasing emergency admissions among the old (continued).

In fact, it has been shown that an ageing population explains only a relatively small proportion of the increase in emergency admissions among older people. The precise proportion depends upon the period examined and the age group in question. Changes in the health status of the older population have not contributed at all to the increase. The balance of the evidence would suggest that, age for age, older people are healthier than they were twenty years ago. (ISD Scotland, 2003; NHS Scotland, 2005b).

Patterns of social change – including changes in living circumstances and other factors affecting the availability of informal care – may have played some part.

As a whole however, factors external to the formal system of health and social care – demographic change, levels of morbidity and patterns of social change – have explained only a proportion of the increase in emergency admissions. To fully understand the increase we have to look within the system of health and social care. It is the way in which care has been delivered in recent decades which underlies most of the increase in emergency admissions.

In particular, needs for health and social care – which manifest themselves initially as demands on primary and social care services, have been translated by the way the system operates into rising numbers of emergency hospital admissions. A lack of integration and co-ordination in the system and a lack of proactive, holistic care focused on long-term conditions has meant that the system tends to wait until medical crises occur. At this point emergency hospital admission may often be a short-term rational solution for each individual element in the system – but may not be the best outcome for the patient or for the system as a whole. A fragmented system of care has meant that problems are passed from one part of the system to another, often ending up at the one part of the system, which rarely says no: emergency inpatient admission.

The positive implication of this ‘internal’ explanation is that future demographic change involving unprecedented increases in the number of older people does not have automatic implications for future numbers of emergency admissions and pressures on the acute system. Redesign of health and social care systems can produce better care and mitigate the pressures on acute hospital services.

There has been a particularly rapid rise in recent decades in multiple emergency admissions among older people. In 1981, for example, 241 patients aged 85 and over were admitted to hospital as an emergency three or more times. By 2004, the numbers were up to 2,327. (ISD Scotland 2003; Scottish Health Statistics, 2006). We have already seen that patients aged over 80 accounted for almost all the increase in emergency bed days in recent decades. In fact most of the increase was accounted for by fewer than 5,000 patients aged over 80 with three or more emergency admissions in a year.
Increasing emergency admissions among the old (continued).

Thus a relatively small group of older patients have accounted for almost the entire increase in pressure on the acute sector with major knock on effects into other areas such as waiting times and delayed discharges.

This has happened to a large extent because these patients have not been receiving the co-ordinated, integrated community-based care, which they need. As well as having negative consequences for the system as a whole, avoidable emergency admissions, often associated with extended stays in hospital and delays in discharge, are often detrimental for patients.

However, just as poorly co-ordinated care for this relatively small group of patients has had major consequences for the health and social care system as a whole, improvements in care for these frail older patients has the potential for beneficial knock on effects throughout the system. They form a key point of leverage in improving the whole system of health and social care.

A focus on long-term conditions.

The Scottish experience is one particular manifestation of a more general international pattern. Healthcare systems are increasingly recognised as not being adequately geared towards the management of long-term conditions. In response, a worldwide shift in the focus of healthcare towards the population management of long-term conditions is beginning to gather momentum.

Several frameworks aimed at orientating and facilitating these shifts have been developed. Perhaps the most influential has been the Chronic Care Model (Improving Chronic Illness Care Program, 2006), which has particularly influenced thinking in this area in England and, more recently, Scotland. A recent survey has outlined other approaches (University of Birmingham Health Services Management Centre, 2006).

It is generally accepted that a first step in moving towards a system of population management of long-term of conditions is what is usually referred to as the risk-stratification of the population. Every person with a long-term condition is assigned to an appropriate level of risk – with the degree of risk being determined by the pattern, seriousness and complexity of their long-term conditions. Each level of risk is associated with an appropriate level of care provision and co-ordination.

Perhaps the most commonly used diagram of the risk stratification of the population with long-term conditions is the Kaiser-Permanente pyramid (see Figure 2 next page).
A focus on long-term conditions (continued).

Figure 2: (Expanded) Kaiser-Permanente Pyramid.
Population management of long-term conditions

The most fundamental benefit of the Kaiser-Permanente pyramid is that it associates an appropriate model of care with an appropriate level of coordination to each layer of the pyramid. Now of course the real world is more complicated than this. Allocation of a model of care and a level of care coordination to individual patients is a much more subtle and nuanced process. However the Kaiser-Permanente pyramid provides a valuable conceptual overview.

The appropriate form of care for people at the ‘tip’ of the pyramid – the 3-5% of the population who have complex needs, usually have multiple long-term conditions – will be some form of intensive case management or care coordination.
A focus on long-term conditions (continued).

**Intensive care/case management/co-ordination: an unofficial guide to terminology.**

There is a distinct lack of consensus in what to call the kind of intensive, preventive, co-ordinated care increasingly recognised as necessary for older people with complex problems and at high risk of emergency admission.

In England, the model has tended to be known as intensive case management, following American usage and probably as a result of the involvement of Evercare in the first Department of Health pilots.

In Scotland, Delivering for Health, along with the Kerr Report that preceded it, does not define a single model but uses phrases such as ‘intensive, co-ordinated management’. This reflects a strong emphasis on the centrality of the co-ordination of care. Thus:

“A member of the primary care team needs to be identified as a fixed point of reference for the patient. This person will take responsibility for the patient, co-ordinating the contribution of various professionals with an interest (including those in social care) and anticipating and dealing with problems before they lead to worsening health or hospitalisation”.

In Scotland there has been some uncertainty around the term ‘care management’ – particularly in social work settings. There is a sense that care management, and to some extent case management, has tended to be implemented in terms of the allocation and administration of packages of care.

This is different to the more dynamic emphasis implied by terms such as intensive case management or intensive care co-ordination which, where necessary, involve constant monitoring of patients at critical junctures and rapidly changing levels of intervention.

SPARRA – the risk prediction algorithm for Scottish patients described later in this report – represents the first stage in identifying those at the tip of the Kaiser-Permanante pyramid who might benefit most from intensive case management.

Thus the development of a risk-prediction algorithm in Scotland is situated at the confluence of two closely related drivers for change. On the one hand is Scotland’s experience of rapidly rising emergency admissions especially among older people. On the other hand is the increasing recognition across the world that the key to improving health service provision and heading off a looming crisis in the resourcing and quality of health care is better management of long-term conditions.
A focus on long-term conditions (continued).

The more general response to these drivers in Scotland has taken the form of the two reports previously mentioned: the National Framework for Service Change in the NHS in Scotland (the Kerr Report) and the Ministerial response to that report (Delivering for Health).

Delivering for Health specifies the following actions which are particularly relevant to this report.

1) A national risk prediction algorithm will be developed by June 2006.

2) NHS Boards will have introduced a system of intensive care co-ordination by the end of 2007.

It is hoped that the development of SPARRA in response to the first of these actions will, as well as fulfilling the narrow goal of identifying high-risk patients, contribute to the second commitment by informing the development of appropriate local models of care.
4. Methodological context.

Statistical modelling in order to calculate the risk that patients will experience an emergency hospital admission in a future time period is an established methodology with a growing literature. Most of this work has been carried out in the United States.

Predictive risk modelling forms an increasingly important component of the process of patient care management in the United States. Given its commercial importance, much of the detail of the algorithms employed is proprietary, however there is a growing published literature.

This literature has been reviewed and summarised by the King’s Fund (King’s Fund, 2005) as part of the background to the development of the King’s Fund risk prediction tool, Patients at Risk of Re-hospitalisation (PARR) (King’s Fund, 2006). The King’s Fund literature review also contains an excellent methodological exposition of the various possible approaches.

The development of SPARRA has benefited in particular from the experience of the King’s Fund model and discussions with its developers.

One of the main differences between SPARRA Version 1 and the King’s Fund tool concerns the trigger for patient’s being included in the model. The King’s Fund model has the acronym PARR – Patients at Risk of Re-hospitalisation. Entry into the risk prediction algorithm is triggered when a patient is admitted to hospital as an emergency.

The PARR algorithm was originally aimed at calculating the probability of Readmission within a year. The statistical modelling is primarily designed with real-time data in mind so that probabilities of readmission can be fed back to the hospital staff while the patient is still in hospital – and preventative measures can be initiated immediately.

The approach of SPARRA is different and is geared more directly to the use of historic data. Entry into the algorithm is not triggered by a patient being admitted as an emergency. Instead all patients who have experienced an emergency admission in the previous three years are entered into the algorithm. Their predicted probability of admission is calculated for a period beginning with an ‘index date’. The vast majority of patients for whom SPARRA predicts admission probabilities are at home – not having experienced an emergency admission for months and years. Hence – Scottish Patients at Risk of Readmission and Admission.

It should be noted that the King’s Fund algorithm also incorporates an ‘archive’ version which, like SPARRA Version 1, is based on national data holdings.
4. Methodological context (continued).

As described in more detail later in this report, the initial version of SPARRA based on historic data is only a first step. One strand of further development will incorporate ‘real-time’ hospital admission data. Another strand will combine the predictive power of hospital admission data and local primary care data. The most exciting potential development is real-time feedback of predicted risks based on a combination of historic and real-time admission data as well as primary care data.
5. **The SPARRA approach: transparent, collaborative, and evolutionary.**

Scotland is a relatively small country with a manageable number of local systems – whether NHS Boards or, increasingly Community Health Partnerships – who are engaged in developing services to manage long term conditions and in particular, programmes of intensive case management for high-risk older patients.

Scotland is fortunate in having ready access at a national level to a linked set of high quality hospital admissions data. These admission records are also linked to the General Register Office for Scotland (GROS) death records (Kendrick and Clarke, 1993).

In England, the size of the country and the number of potential users of the model meant that a more standardised and ‘remote access’ tool had to be developed. Thus the results of the King’s Fund risk prediction model are incorporated in an Access tool, which can be downloaded from the web. Some flexibility is built in to the King’s Fund tool. For example, users can decide whether to incorporate real-time data. In addition users can set their own risk-thresholds, however it is the responsibility of local systems to download national data and supplement it with local data for input to the Access tool to calculate predicted admission probabilities. This imposes a considerable data management burden on local systems.

The aim in Scotland is for ISD Scotland to do as much of the analytical and data management work as possible and to provide information based on the model in a form that is of most immediate use to local systems. This means that the SPARRA team in ISD Scotland needs to work as closely as possible with the local systems in developing methods of risk prediction and forms of output which are appropriate to local developments of care co-ordination.

The approach being adopted in developing SPARRA is to avoid as far as possible the development of a “black box” model whereby it must be taken on faith that the statistical methods adopted do in fact identify the highest risk patients.

We are aiming to make the analytical procedures as transparent and comprehensible as possible.

The narrow purpose of SPARRA is to identify those patients at highest risk of future emergency inpatient admission.
5. The SPARRA approach: (continued).

It is becoming increasingly clear in early discussions with colleagues from local systems that SPARRA can fulfil a wider purpose. The more we can make explicit the characteristics of the people most at risk of emergency admission and the more we can relate risk-levels to ‘whole-person’ portraits of the high-risk patients, the more SPARRA can contribute to the process of developing and refining appropriate models of care.

In the first instance, feedback will take the form of relatively simple descriptions of those patients identified as being at high-risk of emergency admission, together with the actual level of predictive risk. It should prove possible before long however, to provide summary admission histories giving details of diagnoses, specialties, lengths of stay etc.

Thus as well as being simply a statistical technique, we hope the model can contribute to a process of collective learning about how best to provide integrated and co-ordinated services for high-risk patients.

In this context it is extremely important to bear in mind that statistical prediction of risk is not the sum total of the case finding process. It is not a matter of ‘patient x is at 67% risk of admission in the next year, therefore they will be assigned intensive case management’. The risk prediction algorithm only identifies a pool of patients most at risk of emergency admission. The algorithm does not define the extent to which the predicted emergency admissions are preventable – or the extent to which patients identified as being at high risk of admission would benefit from a more preventative approach such as intensive case management.

Identification of the patients who will benefit most from preventive intervention and the specification of the appropriate form and level of intervention will depend upon further screening and assessment based instruments appropriate to local intervention programmes.

Thus assignment to intensive case management or other appropriate forms of preventive care is a two-stage process. First, statistical risk prediction identifies a pool of patients at high risk. Second, local screening and assessment identifies which patients would benefit from care co-ordination.

Work on risk prediction may, however be able to shed further light on these issues of preventability and appropriateness of care. There has been considerable work undertaken on trying to understand and ascertain which diagnoses offer the greatest potential for admission avoidance (King’s Fund, 2006; NHS Quality Improvement Scotland, 2004; NHS Institute for Innovation and Improvement, 2006). These are often called ambulatory care sensitive diagnoses. There is considerable scope for developing mechanisms for identifying and providing further useful information on patients at risk of emergency admission with these ambulatory care sensitive diagnoses.
6. Overview of developmental phases.

a) Historic admission-based model.

The risk-prediction model presented in this paper is based on the set of linked hospital admission and death records held at ISD Scotland. These data are currently around six months out-of-date. For example at the end of June 2006 ISD Scotland holds reasonably complete admissions data up to the end of 2005 (although this varies by NHS Board). (See Section 10 below for a discussion of the implications of this time lag in the data). We will call this the 'historic admissions-based model'.

Improvement in the model by incorporating additional data sets is likely to advance on two fronts.

b) Incorporation of real-time admissions data.

The first is by incorporating more up-to-date hospital admissions data. This will be particularly important if there is a desire to take action to prevent avoidable readmissions while the patient is still in hospital.

For further details of possibilities relating to the incorporation of real-time data see section 11 below on further development.

c) Incorporation of primary care data.

The second route forward is to link the work on admissions data to patient information held at practice level. This will allow incorporation of a wider range of patient variables and in particular those which could be described as ‘leading indicators’ predictive of emergency admission as opposed to the ‘lagging indicators’ derived from previous admissions.

ISD Scotland does not have access to a national set of comprehensive primary care data, which could be used for operational purposes. The SPARRA team is however keen to work with local systems that do have well-developed primary care patient data to further develop predictive modelling in this area. As operational primary care data are locally owned, it is especially important that these developments are driven and led by the local systems.

d) Bringing it all together: an inclusive predictive system.

As is described more fully in Section 9 where validity of the SPARRA model is discussed, if we are interested in identifying the patients at highest risk of emergency admission, then the historic version of SPARRA will probably exemplify the 80/20 rule. We will be able to identify the bulk of the highest-risk patients on the basis of historic admissions data.
d) *Bringing it all together: an inclusive predictive system (continued)*.

For purposes of developing local ownership of the risk prediction agenda and for identifying patients at lower levels of risk, it will be important to incorporate risk-prediction into primary care information systems. Where current primary care information systems are well developed and as the Community Health Index (CHI) number is incorporated into hospital admissions data, we should be able to make relatively rapid progress towards a risk prediction system that uses historic and real-time admissions data as well as local primary care patient data.

Figure 3 shows the stages in developing and applying the historic admissions-based version of SPARRA

1. The predictive model is developed by applying logistic regression to a group of patients for whom we do know the outcome. In the present case this was the cohort of patients aged 65 and over and admitted as an emergency admission at any time between 2001 and 2003. The outcome was whether these patients were admitted as an emergency during the calendar year 2004.

2. The final logistic regression model gives us a set of coefficients and odds ratios, which quantify the precise effects of each independent variable. For example it might tell us that someone in the most deprived decile of data zones is 1.43 times more likely to be admitted than someone in the most affluent.

3. These coefficients or odds ratios are embodied in a look-up table.
7. **Historic admissions-based model: overview (continued).**

4. We calculate the values of the predictive variables for the set of patients for whom we do not know the outcome and for whom we want to calculate the predicted probabilities. This precise data set will depend upon how up to date is the local data concerned. For example, if admissions for residents in NHS Lanarkshire were complete up to the end of January 2006, our cohort for Lanarkshire would be those patients who were admitted as an emergency in the period 1st February 2003 to 31st January 2006. (We would be predicting the probability that these patients would be admitted in the period 1st February 2006 to 31st January 2007).

5. The values of the predictor variables for the ‘prediction cohort’ are fed into the look-up table. In other words the odds ratios in the look-up table are applied to the values for each of the individuals in the ‘prediction cohort’.

6. This allows us to calculate the predicted probability of admission for each of the individuals in the ‘prediction cohort’.

7. Details of high-risk individuals as well as risk-distributions are fed back to the front-line teams.
8. The risk-prediction model.

The data.

The data used to develop the risk prediction algorithm were the linked set of SMR01 hospital admission and death records held at ISD Scotland.

The records in this data set are linked using probability matching based on a wide range of personal identifiers such as name, date of birth and hospital case reference number. The accuracy of the linkage is generally estimated to be above 99%.

Description of the population.

The population used to develop SPARRA Version 1 consisted of all patients aged 65 and over admitted as an emergency in-patient in the three year period consisting of the calendar years 2001 to 2003. The outcome variable was whether or not the patient was admitted as an emergency inpatient in the calendar year 2004 (see Figure 4).

Figure 4: Time frame. SPARRA development data.

Index date
1st January 2004

<table>
<thead>
<tr>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>** Outcome year **</td>
</tr>
<tr>
<td>** Previous admissions **</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Any patient who died before 1st January 2004 was excluded from the analysis.
8. The risk-prediction model (continued).

*Brief description of variables included in analysis.*

**Independent Variables.**

**Age.**

The age of the patient was defined as age at 1st January 2004 i.e. age at the index date or the beginning of the outcome period for the development analysis.

**Gender.**

No description needed.

**Number of emergency inpatient admissions.**

This is the total number of emergency inpatient admissions in the three-year period preceding the index date: in this case 2001 to 2003.

**Time since most recent emergency admission.**

This is the number of days between the most recent emergency inpatient admission and the index date.

**Number of elective inpatient admissions.**

This is the total number of elective inpatient admissions in the three-year period preceding the index date: in this case 2001 to 2003.

**Number of day case admissions.**

Similarly, the total number of day case admissions in the three-year period preceding the index date: in this case 2001 to 2003.

**Total number of inpatient bed days.**

Total number of bed days accumulated by the patient in the three years prior to the index date.

**Broad diagnosis group.**

This shows the principal diagnosis recorded at the most recent emergency admission experienced by the patient. The diagnoses are grouped into ten broad categories as shown in Table 1. It was decided to adopt a relatively simple grouping of diagnoses rather than a more complex coding system such as Healthcare Resource Groups (HRG’s).
Broad diagnosis group (continued).

This was primarily in the interests of transparency, ease of interpretation and the desire to avoid a variable with too many categories. Inspection of a crosstabulation of ICD10 three digit codes against the proportion of positive outcomes showed relatively little variation in the outcome proportion. There was one exception however with Chronic Obstructive Pulmonary Disease (COPD) being associated with a high proportion admitted as an emergency in the outcome year. This reinforced external evidence suggesting that COPD should be given a separate category.

Number of diagnostic groupings.

This is simply a count of the number of different diagnostic groupings found in any record in any diagnostic position in the three-year admission. Each diagnostic group is only counted once. It is a simple measure of the complexity of the patient's diagnostic history.

Scottish Index of Multiple Deprivation.

Each patient was assigned the Scottish Index of Multiple Deprivation for the data zone in which they were living at the most recent emergency admission. The deprivation measure is based on deciles ranging from 1 (most affluent) to 10 (most deprived). Patients are assigned to data zones based on their postcode. The data zone is the smallest area to which the deprivation index is assigned with each data zone containing on average around 700 people.

NHS Board of Residence.

The NHS Board in which the patient was resident at their most recent emergency admission.

Outcome or dependent variable.

Whether or not the patient is admitted as an emergency inpatient in the outcome year.

Logistic regression.

The object of the exercise is to calculate the probability that a given patient is admitted as an emergency in a given (future) year. This can be regarded as a binary outcome: the patient is admitted in the outcome year or they are not.

Logistic regression is a statistical technique which:

a) from a range of candidate independent variables identifies those which have a significant independent effect on the probability of a binary outcome; and
b) specifies the precise size of the effects of each variable.
<table>
<thead>
<tr>
<th>Table 1</th>
<th>Broad diagnostic groupings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis</strong></td>
<td><strong>ICD10 codes</strong></td>
</tr>
</tbody>
</table>
| 1 Cancer | C00 to C97  
D00 to D48 |
| 2 Heart disease | I00 to I52 |
| 3 Other disorders of the circulatory system  
e.g. cerebrovascular disease (inc stroke), arterial disease, phlebitis, diseases of veins etc. | I60 to I99 |
| 4 Other diseases of the respiratory system  
e.g. acute upper respiratory disease, flu, pneumonia, bronchitis, asthma | J00 to J43  
J45 to J46  
J48 to J99 |
| 5. COPD | J40 to J44, J47 |
| 6 Diseases of the digestive and urinary system | K00 to K93  
N00 to N39 |
| 7 Mental disorders and diseases of the nervous system  
Includes dementia, other psychiatric disorders; Alzheimer's disease, other diseases of the nervous system. | F00 to F99  
G00 to G99 |
| 8 Symptoms, signs and ill-defined conditions  
Includes cardiac murmur, cough, chest pain, abdominal pain, difficulty walking, dizziness, confusion, malaise and fatigue, syncope and collapse etc. | R00 to R99 |
| 9 Injuries, etc | S00 to S99  
T00 to T35  
T51 to T98 |
| 10 Other  
Infectious and parasitic diseases | A00 to A99  
B00 to B99 |
| Endocrine, nutritional and metabolic disorders | E00 to E99 |
| Diseases of blood and blood forming organs | D50 to D89 |
| Diseases of the eye and ear | H00 to H95 |
| Diseases of reproductive system | N49 to N99  
O00 to O99 |
| Diseases of skin and subcutaneous tissue | L00 to L99 |
| Musculoskeletal system and connective tissue | M00 to M99 |
| Congenital anomalies and perinatal conditions | P00 to P96  
Q00 to Q99 |
| Poisoning | T36 to T50 |
| External causes of morbidity | V01 to Y98 |
| Factors influencing health status | Z00 to Z99 |

Logistic regression is thus perfectly suited to the task in hand and is by far the most commonly used statistical technique in this area.
**Interaction effects.**

As well as identifying which of the variables tested for inclusion in the model have significant independent effects acting on their own, logistic regression can also check whether there are any significant interaction effects. An interaction effect occurs when the effect of one variable depends upon the value of another variable. For example, if age had an effect on an outcome for men but not for women, that would be an interaction effect. Logistic regression would accommodate this by calculating a main effect for gender, a main effect for age and an interaction effect for age and gender.

**Specifying the model.**

Specifying the model is the phase of the analysis that identifies which of the variables (and interactions) tested for inclusion do in fact have a significant effect on the outcome.

Forwards stepwise selection was used to identify the independent variables to be included in the model. Using forwards stepwise selection; the first step is to assess which variable on its own has the greatest predictive power. This variable is included in the model. Then the remaining variables and interactions are tested to see which one adds the most predictive power. Thus effects are successively added until none of the remaining variables adds any significant predictive power to the model.

Specification of the model was carried out on two different 10% samples of the data. This was to assess the stability of the model: whether the same set of independent variables identified as significant on one set of data is identified as significant on another set of data.

In fact the same main effects were identified as significant in both halves of the analysis. All but one of the independent variables tested for inclusion were found to have significant effects. The exception was NHS Board of Residence.

There was however a good deal of instability in the interactions identified as significant.

Partly for this reason, it was decided to exclude all interactions but one from the model. The only interaction retained was that between age and number of previous emergency admissions. Previous work had shown that age has an independent effect on the probability of emergency admission for those people with a relatively small number of previous emergency admissions.
Specifying the model (continued).

For people with a relatively high number of previous emergency admissions – three or more – age has only a slight or no effect. Indeed, results of the present analysis suggest a reverse effect for very high numbers of previous admissions. The coefficients of the interaction term reflect this pattern (See Section 8 for a fuller discussion). This was unlike the pattern of coefficients for the other interactions, which made no interpretable sense.

Another reason for excluding the interactions which were on the margins of significance is that the inclusion of too many interactions can have distorting effects on the apparent main effects of the variables involved.

As well as a powerful model we want a model whose results 'make sense'. Inclusion of too many interactions generates a complex model which it is almost impossible to interpret.

An important point to emphasise is that, at the margin, tweaking the model in terms of exclusion or inclusion of borderline variables or interactions makes relatively little difference to the predictive power of the model. Most of the predictive work is done by those variables added first in the course of forwards stepwise selection with rapidly decreasing returns for the variables added towards the end of the process.

**NHS Board of Residence.**

Somewhat surprisingly, the only variable entered into the specification process that was not identified as having a significant independent effect was NHS Board of Residence.

This may be a surprise because of the recent perception of variation in rates of emergency admission between NHS Boards. It must be remembered however, that in this analysis, much of any variation in emergency admission rates between NHS Boards will have been absorbed by the fact that an individual’s number of previous emergency admissions is in the model as a very powerful predictor. Variation in emergency admission rates between NHS Boards is likely to have a greater effect on 'number of previous emergency admissions' – an independent variable – than it is on likelihood of admission over the next year. In addition, the current model controls for the effect of deprivation.

The present analysis does not imply therefore that there is no significant variation in admission rates between NHS Boards. It is saying however that once numbers of previous admissions, deprivation and all the other independent variables have been taken account of, NHS Board has no significant effect on the likelihood that a given person will be admitted in the coming year.
9. Results.

Effects of independent variables.

Table 2 shows the odds ratios for the main effects for the variables in the model. The odds ratios show how much the probability of emergency admission for a patient in a given category of the independent variable compares with the probability for someone in the reference category of that variable.

Thus to take a relatively simple example we can look at the effect of the Scottish Index of Multiple Deprivation. The odds ratio for the reference category, the most affluent decile or tenth, is set at 1.00. Compared with this group, other things being equal, individuals in the fifth most affluent decile are 1.13 times more likely to be admitted and individuals in the most deprived decile are 1.41 times more likely to be admitted.
### Table 2  SPARRA Version 1. Main effects. Full sample.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Cases</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>44,686</td>
<td>1.00</td>
</tr>
<tr>
<td>70-74</td>
<td>46,139</td>
<td>1.16</td>
</tr>
<tr>
<td>75-79</td>
<td>45,086</td>
<td>1.49</td>
</tr>
<tr>
<td>80-84</td>
<td>39,775</td>
<td>1.83</td>
</tr>
<tr>
<td>85-89</td>
<td>23,375</td>
<td>2.11</td>
</tr>
<tr>
<td>90+</td>
<td>14,986</td>
<td>2.26</td>
</tr>
<tr>
<td><strong>Previous emergency admissions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One</td>
<td>132,113</td>
<td>1.00</td>
</tr>
<tr>
<td>Two</td>
<td>44,733</td>
<td>1.43</td>
</tr>
<tr>
<td>Three</td>
<td>18,341</td>
<td>1.98</td>
</tr>
<tr>
<td>Four</td>
<td>8,614</td>
<td>2.46</td>
</tr>
<tr>
<td>Five</td>
<td>4,236</td>
<td>3.38</td>
</tr>
<tr>
<td>Six or more</td>
<td>6,010</td>
<td>4.86</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>90,515</td>
<td>1.00</td>
</tr>
<tr>
<td>Female</td>
<td>123,532</td>
<td>0.94</td>
</tr>
<tr>
<td><strong>Time since most recent emergency admission</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than one month</td>
<td>14,518</td>
<td>1.00</td>
</tr>
<tr>
<td>1 to 2 months</td>
<td>11,517</td>
<td>0.75</td>
</tr>
<tr>
<td>2 to 3 months</td>
<td>10,451</td>
<td>0.69</td>
</tr>
<tr>
<td>3 to 4 months</td>
<td>9,009</td>
<td>0.65</td>
</tr>
<tr>
<td>4 to 5 months</td>
<td>8,458</td>
<td>0.61</td>
</tr>
<tr>
<td>5 to 6 months</td>
<td>7,817</td>
<td>0.60</td>
</tr>
<tr>
<td>6 to 7 months</td>
<td>7,740</td>
<td>0.56</td>
</tr>
<tr>
<td>7 to 8 months</td>
<td>7,445</td>
<td>0.57</td>
</tr>
<tr>
<td>8 to 9 months</td>
<td>6,895</td>
<td>0.54</td>
</tr>
<tr>
<td>9 to 10 months</td>
<td>7,022</td>
<td>0.51</td>
</tr>
<tr>
<td>10 to 11 months</td>
<td>6,509</td>
<td>0.52</td>
</tr>
<tr>
<td>11 to 12 months</td>
<td>6,822</td>
<td>0.51</td>
</tr>
<tr>
<td>12 to 24 months (1 to 2 years)</td>
<td>62,612</td>
<td>0.45</td>
</tr>
<tr>
<td>24 to 36 months (2 to 3 years)</td>
<td>47,232</td>
<td>0.40</td>
</tr>
<tr>
<td><strong>Total bed days</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 2 days</td>
<td>26,608</td>
<td>1.00</td>
</tr>
<tr>
<td>2 to 6 days</td>
<td>51,894</td>
<td>1.10</td>
</tr>
<tr>
<td>7 to 13 days</td>
<td>42,252</td>
<td>1.18</td>
</tr>
<tr>
<td>14 to 27 days</td>
<td>37,888</td>
<td>1.30</td>
</tr>
<tr>
<td>28 to 55 days</td>
<td>27,688</td>
<td>1.38</td>
</tr>
<tr>
<td>56 days or more</td>
<td>27,717</td>
<td>1.24</td>
</tr>
</tbody>
</table>
Table 2 (cont.) SPARRA Version 1. Main effects. Full sample.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Cases</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principal diagnosis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>5,680</td>
<td>1.45</td>
</tr>
<tr>
<td>Heart disease</td>
<td>31,054</td>
<td>1.38</td>
</tr>
<tr>
<td>Other circulatory</td>
<td>14,128</td>
<td>1.15</td>
</tr>
<tr>
<td>Other respiratory</td>
<td>10,907</td>
<td>1.27</td>
</tr>
<tr>
<td>COPD</td>
<td>7,122</td>
<td>2.13</td>
</tr>
<tr>
<td>Digestive/urinary</td>
<td>30,105</td>
<td>1.18</td>
</tr>
<tr>
<td>Mental/nervous system</td>
<td>7,714</td>
<td>1.20</td>
</tr>
<tr>
<td>Symptoms, signs, ill-defined</td>
<td>46,242</td>
<td>1.00</td>
</tr>
<tr>
<td>Injuries etc.</td>
<td>32,195</td>
<td>1.28</td>
</tr>
<tr>
<td>Other</td>
<td>28,900</td>
<td>1.24</td>
</tr>
<tr>
<td><strong>Number of different diagnosis groups.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One</td>
<td>29,992</td>
<td>1.00</td>
</tr>
<tr>
<td>Two</td>
<td>52,089</td>
<td>1.06</td>
</tr>
<tr>
<td>Three</td>
<td>52,194</td>
<td>1.25</td>
</tr>
<tr>
<td>Four</td>
<td>39,808</td>
<td>1.37</td>
</tr>
<tr>
<td>Five</td>
<td>24,445</td>
<td>1.49</td>
</tr>
<tr>
<td>Six or more</td>
<td>18,519</td>
<td>1.58</td>
</tr>
<tr>
<td><strong>Scottish Index of Multiple Deprivation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decile 1: most affluent</td>
<td>16,216</td>
<td>1.00</td>
</tr>
<tr>
<td>Decile 2</td>
<td>17,037</td>
<td>1.08</td>
</tr>
<tr>
<td>Decile 3</td>
<td>18,831</td>
<td>1.11</td>
</tr>
<tr>
<td>Decile 4</td>
<td>20,110</td>
<td>1.10</td>
</tr>
<tr>
<td>Decile 5</td>
<td>21,829</td>
<td>1.13</td>
</tr>
<tr>
<td>Decile 6</td>
<td>23,153</td>
<td>1.20</td>
</tr>
<tr>
<td>Decile 7</td>
<td>23,689</td>
<td>1.25</td>
</tr>
<tr>
<td>Decile 8</td>
<td>24,818</td>
<td>1.29</td>
</tr>
<tr>
<td>Decile 9</td>
<td>24,206</td>
<td>1.34</td>
</tr>
<tr>
<td>Decile 10: most deprived</td>
<td>24,158</td>
<td>1.41</td>
</tr>
<tr>
<td><strong>Previous elective inpatient admissions.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>157,817</td>
<td>1.00</td>
</tr>
<tr>
<td>One</td>
<td>36,271</td>
<td>1.00</td>
</tr>
<tr>
<td>Two</td>
<td>11,938</td>
<td>1.00</td>
</tr>
<tr>
<td>Three</td>
<td>4,209</td>
<td>1.12</td>
</tr>
<tr>
<td>Four or more</td>
<td>3,812</td>
<td>1.26</td>
</tr>
<tr>
<td><strong>Previous day case admissions.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>166,494</td>
<td>1.00</td>
</tr>
<tr>
<td>One</td>
<td>49,040</td>
<td>1.02</td>
</tr>
<tr>
<td>Two</td>
<td>18,947</td>
<td>1.07</td>
</tr>
<tr>
<td>Three</td>
<td>6,748</td>
<td>1.12</td>
</tr>
<tr>
<td>Four or more</td>
<td>8,279</td>
<td>1.25</td>
</tr>
</tbody>
</table>
Age and previous emergency admissions: main effects and interaction.

Age.

The main effects odds ratios for age run in the expected direction, increasing with the older age groups. However because the model includes an interaction term between age and number of previous emergency admissions, the overall effect of age can only by assessed by looking at the main effect and this interaction effect in combination.

Number of emergency admissions in previous three years.

The main effect for previous emergency admissions is the most powerful of all. The odds ratio for people with six or more previous emergency admissions is 4.86 compared with patients with only one previous emergency admission. Again however, because of the interaction effect between age and number of previous emergency admissions, the overall effect of number of previous emergency admissions can only be assessed in combination with the effect of age.

Table 3 Odds Ratios: age and previous admissions interaction.

<table>
<thead>
<tr>
<th>Age group</th>
<th>One</th>
<th>Two</th>
<th>Three</th>
<th>Four</th>
<th>Five</th>
<th>Six or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>65 to 69</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 to 74</td>
<td>0.97</td>
<td>0.92</td>
<td>0.85</td>
<td>0.86</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>75 to 79</td>
<td>0.86</td>
<td>0.79</td>
<td>0.77</td>
<td>0.69</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>80 to 84</td>
<td>0.77</td>
<td>0.67</td>
<td>0.62</td>
<td>0.55</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>85 to 89</td>
<td>0.74</td>
<td>0.61</td>
<td>0.62</td>
<td>0.41</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>90+</td>
<td>0.68</td>
<td>0.56</td>
<td>0.46</td>
<td>0.43</td>
<td>0.31</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 shows the interaction effects for age and number of previous emergency admissions.
Age and previous emergency admissions (continued).

What is going on here can be better understood by looking at Figure 5. This does not show the results of the logistic regression but simply shows the proportion admitted in the forthcoming twelve months within each category defined by age and number of previous admissions.

If we look at the first set of columns, showing percentage admitted by age for those with only one previous emergency admission, there is a strong and straightforward positive association with age. As the number of previous emergency admissions increases however, the relationship with age attenuates. In fact by the time we reach people with five previous emergency admissions, the relationship has reversed so that the older the patient, the less is the chance that they will experience an emergency admission.

This is a classic interaction effect. The relationship between age and the probability of future emergency admissions changes depending upon the value of another variable: in this case, number of previous emergency admissions.

When two variables are represented in a model only by main effects, their combined effect can be calculated by multiplying the odds ratios for given categories. Where there is an interaction effect it is necessary to multiply the main effects and the interaction effect for the relevant categories.

As an example, how do we assess the effect of someone being aged 85 to 89 and having experienced five previous emergency admissions on the probability of emergency admission? The effect is calculated as an odds ratio to the reference category: someone aged 65 to 69 who has only experienced one previous emergency admission.
Age and previous emergency admissions (continued).

The main age effect for someone aged 85 to 89 is 2.11; the main effect for having five previous admissions is 3.38; the interaction effect for aged 85 to 89 and five previous admissions is 0.41. The combined effect is this 2.11 x 3.38 x 0.41 or 2.92. Thus the interaction odds ratios for older patients with higher numbers of previous admissions act to dampen or even reverse the impact of the main effects.

Gender.

Gender has a small effect on the probability of emergency admission, with women being slightly less likely to be admitted – the odds ratio compared with men being 0.94.

Time since most recent emergency admission.

The reference category consists of those patients who on 1st January 2004 were within one month of their most recent admission. Initially there is a rapid fall off in the outcome probability after the first month. Patients in their second month from most recent emergency admission have only 75% of the probability of being admitted of those in their first month. Thereafter there is a gradual fall off as the interval extends to a year so that patients who are twelve months away from their most recent admission are down to an odds ratio of 0.51 – around half the probability of those within a month.

Thereafter there is a gradual falling off as the interval extends to two and then three years.

Total bed days.

The pattern of odds ratios for total bed days occupied in the previous three years is in the expected direction, with the exception of the last category consisting of those who have been in hospital more than 56 days in the last three years.

Principal diagnosis at most recent admission.

Here the reference category is ‘signs and symptoms’: the group of individuals with a principal diagnosis of ‘signs and symptoms’ at the most recent emergency admission. Most other diagnoses have odds ratios between 1.1 and 1.3 compared to this group.

The main exception of course is COPD with an odds ratio of 2.13. The other diagnosis groups with relatively high odds ratios are cancer and heart disease.
Number of different diagnosis groups.

Here there is a clear relationship with outcome. Those who have had six or more diagnoses coded in the previous three years have an odds ratio of 1.58 i.e. they are 1.58 times more likely to be admitted than those who have had only one diagnosis group coded in the previous three years.

Scottish Index of Multiple Deprivation.

For the more affluent half of the population there is quite a gentle gradient so that those in the fifth most affluent decile are only 1.13 times more likely to be admitted than those in the most affluent decile. The gradient steepens however in the more deprived half of the population so that the odds ratio for the most deprived decile is 1.41.

Number of previous elective inpatient admissions.

One or two previous elective admissions have no effect on the probability of future emergency admission. Only those with three or more elective admissions show higher probability of emergency admission.

Number of previous day case admissions.

These have a similar effect to elective inpatient admissions.

Summary of effects of independent variables.

An important aspect of the pattern of odds ratios for the independent variables is that they all make sense. By far the most powerful effect is that of number of previous admissions. As we have seen, the presence of an interaction effect means that this must be viewed in combination with the effect of age. Beyond this powerful complex effect, probably the next most influential variable is time since most recent emergency admission. Other reasonably sized effects are diagnosis, with COPD standing out as the strongest predictor of future emergency admission; number of different diagnoses; Scottish Index of Multiple Deprivation and total bed days. Previous elective inpatient admissions, previous day case admissions and gender have only moderate effects.
Distribution of predicted probabilities.

Figure 6 shows the distribution of predicted probabilities for the full SPARRA cohort of 214,047 individuals. 2,751 individuals or 1.3% had a predicted probability of emergency admission of 70% or higher with another 5,456 (2.5%) in the range from 60 to 70%.

This group of 8,207 people, or 3.8% of the cohort, with a probability of 60% or more of experiencing an emergency in the next year might, might be regarded as the first step in identifying the top level of the Kaiser-Permanente pyramid among older people.
Examples.

Table 4 shows three examples selected at random from different ‘probability bands’ of the predicted admission probabilities. The predicted probability of admission is shown alongside values of the independent variables. NHS Board of Residence is shown but it should be remembered that this was not used as a predictive variable.

<table>
<thead>
<tr>
<th>Table 4.</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case A</td>
</tr>
<tr>
<td></td>
<td>Very high probability</td>
</tr>
<tr>
<td>Predicted admission probability</td>
<td>86%</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
</tr>
<tr>
<td>Age</td>
<td>65 to 69</td>
</tr>
<tr>
<td>Previous emergency admissions</td>
<td>6 or more</td>
</tr>
<tr>
<td>Since most recent emergency admission</td>
<td>Less than one month</td>
</tr>
<tr>
<td>Bed days in previous 3 years</td>
<td>14 to 28 days</td>
</tr>
<tr>
<td>Most recent diagnosis</td>
<td>COPD</td>
</tr>
<tr>
<td>Number of diagnoses</td>
<td>6 or more</td>
</tr>
<tr>
<td>Deprivation decile</td>
<td>10 (most deprived)</td>
</tr>
<tr>
<td>Previous elective inpatient admissions</td>
<td>None</td>
</tr>
<tr>
<td>Previous day case admissions</td>
<td>None</td>
</tr>
<tr>
<td>NHS Board (not in predictive model)</td>
<td>Greater Glasgow</td>
</tr>
<tr>
<td>Outcome</td>
<td>Admitted</td>
</tr>
</tbody>
</table>
10. **Performance of the model.**

Perhaps the simplest way to look at the performance of the model is to compare the pattern of predicted probabilities with the actual outcomes experienced by the cohort.

Ideally this should be done on a sample which is different to that on which the model was developed. Here we are looking at the outcomes of the same people: the full SPARRA cohort. The effects of the independent variables were stable when tested on different samples and combined with the size of the data set; this would suggest we do not have an over-fitted model.

Figure 7 presents the distribution of predicted probabilities with each ‘risk category’ split according to whether its members were admitted in the outcome year. In other words we are comparing the outcome predicted by the model with the actual outcome.

**Figure 7:** Predicted risk categories split by whether patients admitted in outcome year or not. Full SPARRA cohort.

The numbers underlying this chart are presented in Table 5.
10. Performance of the model (continued).

Table 5: Predicted versus actual outcomes

<table>
<thead>
<tr>
<th>Predicted admission probability</th>
<th>Total in risk category</th>
<th>Outcome</th>
<th>Per cent admitted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Not admitted</td>
<td>Admitted</td>
</tr>
<tr>
<td>Under 10%</td>
<td>1,287</td>
<td>1,210</td>
<td>77</td>
</tr>
<tr>
<td>10-20%</td>
<td>58,088</td>
<td>49,075</td>
<td>9,013</td>
</tr>
<tr>
<td>20-30%</td>
<td>68,692</td>
<td>51,249</td>
<td>17,443</td>
</tr>
<tr>
<td>30-40%</td>
<td>44,135</td>
<td>28,925</td>
<td>15,210</td>
</tr>
<tr>
<td>40-50%</td>
<td>22,696</td>
<td>12,656</td>
<td>10,040</td>
</tr>
<tr>
<td>50-60%</td>
<td>10,942</td>
<td>5,185</td>
<td>5,757</td>
</tr>
<tr>
<td>60-70%</td>
<td>5,456</td>
<td>1,924</td>
<td>3,532</td>
</tr>
<tr>
<td>70% plus</td>
<td>2,751</td>
<td>656</td>
<td>2,095</td>
</tr>
</tbody>
</table>

A commonly used method to assess the power or adequacy of a predictive model is to calculate the area under the Receiver Operating Curve (ROC).

Figure 8 shows the ROC for the main SPARRA analysis. The area under the curve measures the accuracy of the predictive technique in question. In a perfect predictive model, the curve would run straight up the left-hand axis and straight along the top of the graph and the area under the curve would be 1. In a model with no predictive power at all, the curve is equal to the diagonal and the area under the curve is 0.5.

Figure 8: Receiver Operator Curve

ROC Curve

Area Under Curve = 0.679

1 - Specificity

Diagonal segments are produced by ties.
10. Performance of the model (continued).

The model developed here has an area under the ROC of 0.68. This is reasonable but not spectacular. Risk prediction models which have incorporated, in addition to information based on previous hospital admissions, such 'leading' variables as prescribing data or data on long term conditions derived from primary care registers seem to be achieving areas under the ROC closer to 0.8 (personal communications: Simon Steer and Paul Leak, NHS Highland; Peter Donnan, Dundee University).

It is however important to recognise here that the ROC is a method of describing and quantifying how the predictive technique operates across the entire range of predicted probabilities. In other words, success in predicting patients with low probabilities of admission contributes to the shape of the ROC and the magnitude of the area under the curve as much as success in predicting which patients have a high-risk of admission.

For the purposes of helping to identify patients who will benefit most from intensive case management, we are primarily interested in how good the model is at identifying these high-risk patients – not necessarily how good it is at predicting risk across the entire risk spectrum.

In terms of assessing the adequacy of SPARRA Version 1 in identifying the high-risk patients who are our primary interest, we can ask a different question.

This question is: compared with a model incorporating a wider range of data from primary care or other sources, how adequate is SPARRA Version 1 in identifying the pool of patients who are at highest risk of future emergency admission? Will it identify only a small proportion of those who would be identified based on wider data? Will it identify around half? Or will it identify most of the potentially identifiable high-risk patients?

Preliminary analysis provides at least a tentative, order-of-magnitude answer to this question. This analysis is based on the data set, which links the 2003 Scottish Health Survey with the SMR hospital admission records. The data set provides us with information on 1,628 individuals' aged 65 and over included in the 2003 Scottish Health Survey.

Two models were developed for these individuals. The first replicated as closely as possible the SPARRA modeling and was based solely on information form previous hospital admissions. The second included additional variables from the Scottish Health Survey.

Comparison of the two models allows us to identify how many additional high-risk patients were identified by the inclusion of variables over and above those available to the SPARRA model.
10. Performance of the model (continued).

In specifying the first model all the independent admission-based variables included in SPARRA were tested for inclusion using forward stepwise regression. Variables which were found to have significant independent effects were: total bed days in the previous three years; time since most recent admission; age and diagnosis at most recent admission. (This is much fewer than the number of significant independent variables in the full SPARRA model because of the much smaller number of cases in the data set).

The model was then rerun, this time including variables from the Scottish Health Survey. The Scottish Health Survey variables assessed for significance were: number of long-standing illnesses; self-assessed health; smoking status; limiting long-standing illness and thirteen separate long-term condition flags. Again, using forward stepwise regression, two of these variables entered the model as having significant independent effects. These were the number of long-term conditions and self-assessed health.

Figure 9: Cumulative predictive risks based on models using
a) 4 SPARRA variables
b) same plus 2 Scottish Health Survey variables

Figure 9 shows the number of patients identified by the model as exceeding a given risk threshold across the entire range from 0% to 100%.

It can be seen that for patients above a risk threshold of 50%, addition of the Scottish Health Survey variables identifies only a relatively small number of additional patients.
10. **Performance of the model (continued).**

The number of additional patients identified at the higher risk thresholds is given in Table 6. We are dealing with relatively small numbers of high-risk patients but as an order of magnitude, the results suggest that we are in the area of the 80/20 rule.

**Table 6: Additional high risk patients identified by expanded model**

<table>
<thead>
<tr>
<th></th>
<th>Identified by SPARRA type model</th>
<th>Identified by SPARRA plus SHS model</th>
<th>Additional numbers identified</th>
<th>Additional percentage identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>80% and above</td>
<td>8</td>
<td>9</td>
<td>1</td>
<td>13%</td>
</tr>
<tr>
<td>70% and above</td>
<td>22</td>
<td>29</td>
<td>7</td>
<td>32%</td>
</tr>
<tr>
<td>60% and above</td>
<td>46</td>
<td>47</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>50% and above</td>
<td>63</td>
<td>72</td>
<td>9</td>
<td>14%</td>
</tr>
</tbody>
</table>

It is almost certain that based on a wider range of data from other sources and applying the model to a larger number of cases, the proportion of additional patients identified will increase.

In terms of answering the order of magnitude question posed earlier – is the SPARRA model identifying most high-risk patients or only a small proportion of them? – this relatively simple analysis would suggest an answer closer to the former. SPARRA is identifying the majority of high-risk patients in the community who are potentially identifiable on the basis of the kinds of data currently available.

This analysis is offered on purely pragmatic short-term grounds. Any local systems contemplating implementation of intensive care co-ordination and thus requiring the identification of a pool of potential candidates for further assessment and screening can be reasonably confident that the SPARRA admissions based algorithm will identify most of the patients at highest risk.

As discussed elsewhere in the report however, the sooner we develop further our understanding of the role of ‘leading indicators’ relating to primary care usage, patterns of long-term conditions and clinical risk factors, the better.
11. Feedback of results.

Feed back of results to local teams wishing to identify high-risk patients as potential candidates for intensive case management or other forms of preventive care can take two main forms: aggregate risk distributions or details of identifiable patients.

**A note on the data**

The data used to calculate predicted probabilities is the linked set of SMR01 hospital admission and death records held at ISD Scotland. *This data set was linked using probability matching and is reckoned to be around 99% accurate. Thus errors in the patient data because of incorrect links or missed links will creep into the list of high-risk patients.* This should be kept in mind when using the information.

In both cases, the date from which the risk is calculated will depend upon the latest date for which ISD Scotland has a reasonably complete set of SMR01 admissions data for the NHS Board or locality in question.

**Aggregate risk distributions.**

For purposes of scoping or planning services, it may be useful to show how many patients in a given GP Practice or Community Health Partnership (CHP) are at given levels of risk of future emergency admission.

Section 8 above, showed the overall distribution of predicted risk for the Scotland level cohort on which the model was developed.

Similar data can be generated for the 'prediction cohort' for each local area. This would show for example that a given GP Practice contained X number of patient at 80% or over risk of emergency admission, Y between 70 and 80% and Z between 60 and 70%.

It is intended that these distributions will be made available on the web at CHP and NHS Board level.

**Identifiable details of high-risk patients.**

ISD Scotland will send to local teams details of patients predicted to be at high risk of emergency admission.

Details will include identifying information such as date of birth, hospital case reference number at most recent admission, postcode and CHI (where available from the SMR01 record). In the first instance names are not being provided. Simple additional information relating to the patient's admission history will also be provided.
11. Feedback of results (continued).

Confidentiality procedures.

ISD Scotland is returning information to NHS Boards about their residents. These data essentially belong to the NHS Board. Release of data will require the completion of a confidentiality form by the NHS Board concerned.

Implications of using historic admissions data.

As the data used to calculate predicted risks are likely to be around six months out-of-date, some of the time in the 'outcome year' will have already passed. For example, if the admissions data for a given Health Board were reasonably complete up to the end January 2006, we would be predicting admission probabilities for the period 1st February 2006 to 31st January 2007. If a local team were intending to identify high-risk patients in July 2006, by then almost half the 'prediction period' will have passed.

How big an issue is this? Does it invalidate the predictions or is the algorithm still a useful guide to high-risk patients? It might be worth thinking about what might have happened to the patient in the relevant period.

One of three eventualities will have occurred between 1st February 2006 and July 2006.

1. **The patient has died.**
   A small number of patients on the list of high-risk patients will have died. This possibility should be checked through local records.

2. **The patient has been admitted to hospital.**
   Due to the strong influence of recency of latest admission, on the probability of future admission, any recent admission is likely to increase the risk-level of a given patient. Thus the predicted risk calculated by SPARRA will be an underestimate of the patient’s true risk; however they are on the list.

3. **The patient has been not admitted to hospital.**
   In this case the true level of risk in July 2006 will be lower than the level of risk calculated as of February 1st 2006. This will be particularly true of patients admitted in the period immediately before the 'index date' for prediction.

Overall, the use of historic data which does not include the most recent admissions will affect the predicted probabilities compared with the probabilities which would be generated if we had real-time data running up to the moment of case-finding. If however, the aim is to identify a pool of older people who are out in the community at a continuing high-level of risk of emergency admission, then the level of distortion is unlikely to be major.
12. Further development.

The further development of SPARRA is likely to proceed in two main directions. The first is the incorporation of more up-to-date admissions data. The second will be to work with local systems to incorporate primary care data into the risk prediction algorithm and to incorporate predicted admission probabilities into local primary care systems.

More up-to-date (‘real-time’) admissions data.

As pointed out in the previous section, if the primary aim of case finding is to identify those patients in the community who are at long-term elevated risk of emergency admission then SPARRA, based on historic data, will provide an adequate initial starting point.

It is clear however that we need to move as quickly as possible towards the incorporation of up-to-date real-time data into the system. This will be especially important if local initiatives are aimed at intervening while patients are in hospital – so as to reduce the likelihood of subsequent avoidable readmissions.

This can be achieved by:

a) linking the historic SMR01 data to real-time data received via the SystemWatch process at ISD Scotland;

b) linking the historic SMR data to more up-to-date data available locally;

c) using local historic and real-time data.

The SPARRA team is currently open to any of these possibilities.

SystemWatch based possibilities.

SystemWatch was originally developed as a response to increasing winter pressures in the form of surges in emergency admissions in Scotland as an aid to demand prediction and bed-management. SystemWatch receives virtually real-time input of hospital admission records for the vast bulk of admissions in Scotland. The incoming records are extracts from hospital Patient Administration Systems containing information such as date and type of admission, patient age and gender and, in some cases, postcode and CHI number.

SystemWatch automatically carries out predictive statistical modelling on the admissions data and generates predicted levels of admissions and occupied beds at hospital level for forthcoming weeks. Trends and predictions are automatically output to the web. SystemWatch also acts as a vehicle for putting trends in a range of other information onto the web, examples being ambulance callouts and NHS 24 contacts.
12. Further development (continued).

Simple record linkage based on hospital case reference number is carried out within SystemWatch. This enables the identification of patients with multiple admissions. A small number of NHS Boards receive notification by email of any patients who have been admitted for the third time in a year. NHS Lothian for example uses this feedback as an initial basis for identifying high-risk patients.

SystemWatch is currently being redeveloped as a production system and this has involved a freeze on additional functionality over the last year. The redeveloped version will however become progressively available over the summer of 2006.

In order to improve the quality of record linkage within SystemWatch and the potential for linkage with historic data and primary care data, it is important that SystemWatch records contain the CHI number. Some NHS Boards already submit the CHI number as part of SystemWatch. As highlighted in the box however, when the redeveloped version goes live in August 2006, having the CHI number on new SystemWatch records will be extremely helpful.

If you would like SystemWatch to become a vehicle for real-time feedback of predicted readmission probabilities for admitted patients –

Please ENSURE THAT CHI NUMBER IS INCLUDED IN THE ADMISSION DATA SUBMITTED TO SYSTEMWATCH as soon as possible.

Other possibilities for more up-to-date admissions data.

It will take time for the redeveloped SystemWatch to come fully on-stream and for CHI to be applied to historic SMR01 records.

In the interim, the SPARRA team is keen to take advantage of local opportunities to modify the risk prediction algorithm to take advantage of more recent data that might not, for example, contain information on diagnoses.

An initial opportunity is likely to be afforded by the data set developed in NHS Forth Valley, which contains both historic and up-to-date admissions data.
12. **Further development (continued).**

**Primary care data and the risk prediction agenda.**

Improved management of long-term conditions is an agenda which is being implemented mainly in primary care.

From this point of view it is something of a paradox that the most powerful predictors of future admission are in the short-term to be found in history of previous admission to the acute sector.

As has already been mentioned in this report, it will be crucial for the full development of the risk-prediction agenda to dovetail admissions based models with the wide range of data available in primary care settings.

This will be a two way process.

On the one hand we need to work towards incorporating variables derived from primary care data sources into the risk prediction algorithm. Important data items are likely to include information on long-term conditions generated as a by-product of the Quality and Outcomes Framework (QOF); prescribing information; information on clinical risk factors; and information on GP contact patterns.

The extent to which this range of information is accessible on a systematic basis from primary care information systems varies a great deal by locality. NHS Ayrshire and Arran for example have developed a rich data set incorporating this type of information.

On the other hand, because primary care teams will be responsible for managing the long-term conditions that underlie the high levels of risk of emergency admission in the types of patients being identified, it will be important to embed information on predicted risks into primary care information systems.

Thus, as well as information on long-term conditions, risk factors, prescribing etc., a practice information system would contain the likelihood (if things stay the same) that each patient will be admitted to hospital over the coming year.

In moving towards incorporating primary care data we will build on the pioneering work carried out in NHS Highland based on linkage between primary care data and hospital admissions for one practice, as well as the work on linked hospital admission and pharmacy data carried out for NHS Tayside at the University of Dundee. The gold standard for risk-prediction based on primary care and admissions data is likely to be defined by the research analysis to be carried out by ISD Scotland and the Universities of Edinburgh, Tayside and Aberdeen of linked primary care and hospital admissions data.
12. Further development (continued).

Extending the feedback from admissions data.

At first, ISD Scotland will feedback basic information about patients at high-risk of future admission: identifying information plus a range of the independent variables generated for risk prediction. For example, this could be diagnosis at most recent admission and number of emergency admissions.

Since the data for modelling are derived from a comprehensive history of hospital admissions, it should be possible to provide a more comprehensive portrait of each high-risk patient’s recent hospital experience.

Economic analysis.

The data generated by risk prediction modelling are a rich source for economic impact analyses. For example the report on the King’s Fund model presents extensive scenarios showing the economic impact of intensive case management under various assumptions. (King’s Fund, 2006) NHS Highland has carried out a pioneering exercise based on a linkage between hospital admissions and primary care data, which is beginning to explore these issues in a Scottish context.
13. Conclusion

This report, along with the work of others engaged in the development of risk prediction in Scotland, represents the beginning of a journey. The shift from reactive care based on crisis management to anticipatory care based on prevention depends on our being able to anticipate better, to predict better.

SPARRA Version 1 is a first contribution to the initial risk stratification of the population with long-term conditions in Scotland. It allows us to start to populate the tip of the Kaiser-Permanente pyramid. However, assignment of sections of the population to a static position in the risk-stratification pyramid is only a first phase.

True population management of long-term conditions will involve dynamic risk-prediction embedded into constantly updated electronic patient records. As well as incorporating data based on hospital admissions and data of the type currently recorded on primary care systems, these will involve real-time information on clinical risk factors and other warning signs such as falls and bereavement.

Although these future steps might seem a long way off, it is important to make continuous small steps forward based on existing forms of data.

The SPARRA team is keen to move forward by working with local teams as part of a developing national agenda. In the next year or two we might just surprise ourselves.
Acknowledgements.

We would like to thank the developers of the King’s Fund tool and especially John Billings for sharing their experience. Peter Donnan of Dundee University has been equally kind in sharing his insights.

In Scotland, discussions with the following have been particularly valuable:

NHS Lanarkshire – Anne Hendry, Kate Bell and Trudi Marshall;
Ayrshire and Arran – Lyall Cameron, Paul Ardin, Richard Walker and Helen Mosson;
Highland – Simon Steer and Paul Leak;

and many others involved in the development of new approaches for high-risk patients.

Within ISD, thanks to Bill Boyd and Richard Lawder and especially to Andrew Elders for providing data from, and advice relating to the 2003 Scottish Health Survey and SMR01 linked data set.
References


King’s Fund (2005) Predictive Risk project: Literature Review:


University of Birmingham Health Services Management Centre (2006) Improving Care for People with Long-term Conditions: a review of UK and international frameworks. NHS Institute for Innovation and Improvement.