

Predicting who will need costly care

HOW BEST TO TARGET PREVENTIVE HEALTH, HOUSING AND SOCIAL PROGRAMMES

Geraint Lewis



© Queen's Printer and Controller of her Majesty's Stationery Office, 2007

Copyright in the typographical arrangement rests with the Crown.

This publication, excluding logos, may be reproduced free of charge in any format or medium for research, private study or for internal circulation within an organisation. This is subject to it being reproduced accurately and not used in a misleading context. The material must be acknowledged as Crown copyright and the title of the publication specified.

Any other use of the contents of this publication would require a copyright licence. Please apply for a Click-Use Licence for core material at www.opsi.gov.uk/click-use/system/online/pLogin.asp, or by writing to the Office of Public Sector Information, Information Policy Team, St Clements House, 2-16 Colegate, Norwich, NR3 1BQ. Fax: 01603 723000 or email: HMSOlicensing@cabinet-office.x.gsi.gov.uk

This is an independent report commissioned by the Department for Communities and Local Government. The study and content therein reflect the views and opinion of Geraint Lewis, not those of the government or government policy.

Produced and printed by the King's Fund.

King's Fund
11-13 Cavendish Square
London W1G 0AN

Contents

About the author	v
A note on the paper	v
Acknowledgements	vi
Summary	vii
Introduction	1
Predictive risk modelling and its relevance to social care	3
Advantages of predictive risk modelling	4
Problems with predictive risk modelling	9
Determining which social care outcomes are important to predict	12
Factors predicting adverse outcomes in social care	13
Predictors of functional decline	13
Predictors of institutionalisation	14
Predictors of admission to nursing home	16
Predictors of need for home care	18
Predictors of high-cost individuals	19
Existing tools that predict risk of adverse outcomes	20
Sources of predictor variables	25
Accessibility of person-level data	26
Feasibility of data linkage	27
Developing risk prediction models	29
Model A (admissions to care homes from hospitals)	29
Model B (admissions to care homes)	33
Model C ('critical' status on FACS)	36
Model D (social care costs)	39
Model E (combined NHS and social care costs)	41
Discussion	45
Next Steps	47
Appendix 1	49
References	50

About the author

Geraint Lewis is a Visiting Fellow at the King's Fund. He also works as a policy advisor in the Cabinet Office and is employed as a public health specialist registrar by the NHS London Deanery; the views expressed in this paper are not necessarily representative of those organisations. Geraint Lewis previously worked at Croydon PCT, where he developed the award-winning 'virtual wards' concept. His background is in acute medicine, and he also works as an accident and emergency doctor at Charing Cross hospital in London. He is spending the 2007/08 academic year as a Harkness Fellow in New York.

A note on the paper

This paper was commissioned by the UK government's Department for Communities and Local Government (CLG) as part of its National Strategy for Housing in an Ageing Society. The brief was to explore the feasibility of developing tools that use routinely collected computerised data to predict which individuals are at risk of needing intensive social care, which could be used to improve the targeting of preventive interventions. Such tools would be analogous to the successful algorithms recently developed for the NHS that forecast which people are at risk of unplanned hospital admission in the forthcoming twelve months.

Acknowledgements

A number of people kindly gave their time providing comments on an earlier draft. I would particularly like to thank the following: John Billings, Niall Dickson, Jennifer Dixon, Deborah Klee, Maggie Ioannou, Rajbant Kaur, Luke O'Shea, Ian Philp, Gerald Pilkington, Guy Robertson and David Walden. Any errors or omissions in the paper are entirely mine.

PROJECT STEERING GROUP

Jennifer Dixon
Luke O'Shea
Guy Robertson
Rajbant Kaur
Alexandra Norrish

KING'S FUND STAFF

Jennifer Dixon
John Billings
Perviz Asaria
Caron Hartely
Natasha Curry

ORGANISATIONS

Cabinet Office
Croydon PCT
Croydon Council
Commission for Social Care Inspection
Communities & Local Government
Department of Health
Health Dialog Analytic Solutions
Health Dialog UK Ltd
Help the Aged
Infoshare Ltd
London Borough of Tower Hamlets
London School of Economics
New York University
Partnerships for Older People Project
Tower Hamlets PCT
University of Kent
University of Sheffield
Warwickshire County Council
Warwickshire PCT

Summary

The costs of social care will become increasingly difficult to sustain as the population ages and more people live with complex conditions. There is a body of research showing that admission to a care home can be delayed or avoided through preventive, upstream interventions. However, the evidence suggests that, increasingly, resources are being concentrated on people with existing high needs, with less investment being made for those people at lower, emerging risk.

‘Upstream’ interventions, which are aimed at maintaining or improving a person’s ability to live independently, are most cost-effective when they are offered to people who would – without intervention – truly go on to require intensive social care. Therefore, if more efficient investment is to be made in preventive interventions, local authorities need ways of identifying individual risk across their population so that they can target effective interventions accurately and proportionately. If reliable predictive tools to do so could be developed that used routine, electronically stored, public sector data, then these would allow local authorities to:

- target preventive interventions to the right groups of vulnerable people; and
- construct ‘business cases’ for doing so.

In combination, these would be powerful incentives for more investment to be made in preventive health, housing and social care.

In the related field of managing long-term medical conditions, unplanned hospital admission can be thought of as being analogous to needing intensive social care. Both situations are typically unwelcome to the individual concerned, they are costly to society and in certain cases, are potentially avoidable. Tools are now being used by the NHS to stratify entire populations according to individual risk of unplanned hospital admission. These tools (such as ‘PARR’ and the ‘Combined Predictive Model’ in England) use patterns in routinely collected data to make predictions about future risk. This feasibility study explores whether similar tools might be constructed to predict the risk of needing intensive social care.

A review of the literature shows that many factors are known to be statistically predictive of a future decline in functioning and the need for intensive social care. Such factors include physical and cognitive disability, sensory impairment, poor housing, poverty and lack of social capital. Certain medical diagnoses are particularly likely to affect independence, for example, cerebrovascular and cardiovascular disease, arthritis, incontinence, dementia and depression. All of these factors are more or less amenable to ‘upstream’, preventive intervention.

Using these known predictive variables, a number of assessment tools have been built to make forecasts of who will need future intensive social care. However, the existing tools

are designed for administration in person with clients (either face-to-face interviews or else by telephone or postal questionnaire), rather than by exploiting existing, computerised datasets alone. Many of the predictive factors used in these existing tools are not recorded in any routine data sources (for example, the opinions of the principal caregiver). However, information on a considerable number of such factors is indeed held in health or social services' computerised records. Examples of such variables include age, number of hospital admissions and specific diagnoses. A new tool that made predictions for individuals based on the person-level data in these datasets may not be as accurate as tailor-made instruments. However, the high cost of social care means that compelling business cases for preventive interventions could still be made on the basis of less accurate predictions.

Making forecasts using routine data would have a number of important advantages:

- by being far less labour intensive to use than face-to-face instruments, such tools could be applied systematically and repeatedly across a population
- they would be free from non-response bias
- they could in theory stratify the whole population, not just those already perceived to be at high risk. This would enable large numbers of people with lower, emerging risk to be identified and offered lower-intensity preventive care.

Other advantages of tools that use routine data include:

- less susceptibility to the inverse care law than a face-to-face tool (since practitioners in deprived areas may have less time to spend completing face-to-face questionnaires)
- avoiding the inverse equity hypothesis (which states that interventions that are targeted geographically to areas of deprivation tend first to be used by relatively less deprived people living in those areas)
- having the potential to be used in resource allocation
- having the potential to be used in performance assessment.

But there are a number of practical and ethical considerations that would have to be taken into account before such tools were built.

- Important issues of confidentiality and consent to consider if named individuals are to be assigned risk scores using routine data.
- Linking data sources at individual level across health and social care is particularly problematic.
- The tools are not 100 per cent accurate and arrangements should be put in place for people whose risk predictions are manifestly inaccurate.
- Data may be missing from routine databases on some individuals or groups (for example, many people are not registered with a GP). Again, arrangements would need to be put in place to support these people.
- The evidence base for 'upstream' interventions needs to be developed.

Five types of predictive models for forecasting the need for intensive social care are presented and appraised (models A–E).

A	Acute model	Predicts admissions direct from acute hospital to care homes
B	Admissions model	Predicts admissions to care homes
C	FACS model	Predicts people whose needs will become 'critical'
D	Social care costs model	Predicts people whose social care costs will escalate
E	Health and social care costs model	Predicts people whose combined health and social care costs will escalate

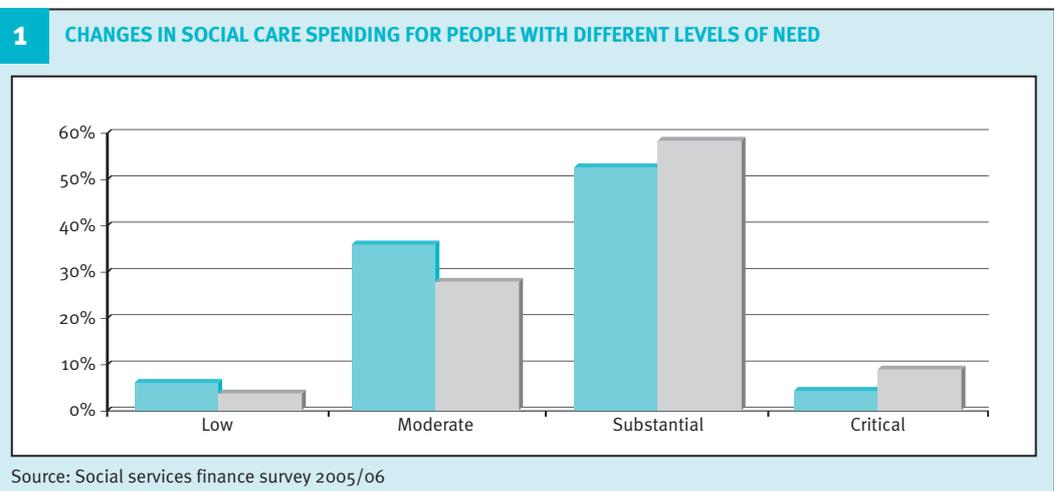
Two models in particular – models A and E – merit further exploration. Model A would be relatively straightforward to construct and use, and would be useful because it predicts a particularly egregious outcome. Model E would encourage health and social care agencies to invest jointly and to work together in the prevention of adverse health and social outcomes.

The next steps for the construction of such models are suggested: one or more tools should be built and tested in a small number of sites across the country, and appraised in terms of accuracy against other methods for identifying those at high risk. Assuming that accurate risk tools can be developed, then the next requirement will be to design and evaluate the effectiveness of 'upstream interventions' that would be offered on the basis of the predictions of the tools according to how they reduce the risk of admission to a care home and their effects on 'downstream' health and social care costs. An analysis of the practical implications for local health and social care agencies will also be required. Then, in a few test sites, the performance of the risk tool(s) and associated interventions should be evaluated in a study akin to that proposed by the Department of Health to evaluate the Whole System Demonstrator sites.¹

Introduction

Being admitted to a care home is a significant and often traumatic event. Not only may it signal the end of independent living outside an institution, but also it frequently entails living at a greater distance from family and friends, and away from a community that has been a key part of life. Financially too it is a very costly event – whether funded by the individual concerned, by his or her family, or by the state. For all of these reasons, supporting older people to continue to live at home where appropriate should be a priority for public policy, as should the development of fairer ways to help individuals and public authorities pay for personal and nursing care.²

In 2000, the United Kingdom spent an estimated £12.9 billion on social care. As the population ages over the coming years, so the amount spent in real terms on care is projected to quadruple to £53.9 billion by 2051.³ In the more immediate future, the number of people aged over 85 is forecast to rise by between 2.2–2.6 per cent each financial year between 2008–09 and 2010–11.⁴ This is already exerting pressure on adult social care budgets and, as a result, already less money is being invested in the social care of people with low to moderate needs. It has been estimated that if current trends were to continue, then spending on such care would end completely by 2009/10 in England.⁵



There is robust evidence that admission to a nursing home can be delayed or avoided by means of preventive, ‘upstream’, interventions. For example, a meta-analysis found that programmes involving a domiciliary multidimensional assessment plus at least nine follow-up visits can reduce admissions to nursing homes by 34 per cent, and reduce deteriorations in functional status by 24 per cent.⁶ Similar benefits are seen even in the case of patients with dementia (a group of incurable conditions characterised by a progressive decline in function). Here, there is high-quality evidence that an ‘upstream’

intervention – in this case the training of carers – delays admission to a nursing home by an average of 20 months.⁷ Likewise, there is evidence that short-term rehabilitation schemes can successfully prevent admission to long-term residential care.^{8,9}

To be cost-effective, ‘upstream’ interventions that aim to maintain or improve a person’s ability to live independently should be offered to people who would, without intervention, truly go on to need intensive social care. Therefore, for more cost-effective investment to be made in preventive interventions, councils with social services responsibilities (CSSRs) need ways of stratifying their population according to individual risk, so that effective interventions can be targeted appropriately. If predictive tools were built that exploited the electronic data routinely collected in the public sector, then these would allow local authorities to offer preventive care to the most vulnerable people, and to construct ‘business cases’ for doing so. This would be a potent incentive to invest more in preventive health, housing and social care.

In the related field of managing long-term medical conditions, unplanned hospital admission is seen as analogous to admission to a care home. Both situations are typically unwelcome to the individual concerned; are costly to society; and – in certain cases – potentially avoidable. Tools are now being used by the NHS to stratify entire populations according to individual risk of unplanned hospital admission. These tools (such as the ‘PARR’ tool^{7,8} and the ‘Combined Predictive Model’¹⁰ being used by primary care trusts (PCTs) in England) use patterns in routine data to make predictions about future risk. PCTs can then offer preventive services that are proportionate to individual risk (for example, intensive case management to those at high risk, and telephone-based coaching to those at lower risk). This feasibility study explores whether similar tools might be constructed to predict the risk of needing intensive social care.

The main objectives of this feasibility study are to:

- summarise general concepts of predictive risk modelling and assess its relevance to social care
- determine what specific risks (outcome or outcomes) could be predicted
- establish what factors are known to be predictive of adverse outcomes in social care
- review what attempts have been made to predict adverse outcomes
- recommend which data sources could be used to develop a model to predict risk of care home admission
- suggest the modelling that would be required
- identify a pilot site (or sites) to work with
- establish the accessibility of individual-level data, and the feasibility of linking it at an individual level, given confidentiality constraints
- estimate the cost of
 - a) developing a model and testing it in other pilot sites
 - b) making it usable by social services staff (for example, using appropriate software and providing the right help and support)
- suggest next steps.

Predictive risk modelling and its relevance to social care

A number of assessment tools have been developed to make forecasts of who will need future intensive social care.^{41,11,12} However, it appears that all current tools are designed to be administered by interviewing clients (either face-to-face, by telephone¹³ or by postal questionnaire¹⁴), rather than by exploiting existing computerised datasets alone. Many of the predictive factors used in existing tools are not recorded in any routine data sources (for example, the opinions of the principal caregiver). However, information on a considerable number of such factors is held in health or social services' computerised records (for example, age, number of hospital admissions and specific diagnoses).

A new tool that made predictions for individuals based solely on their person-level data held in routine datasets might not be as accurate as tailor-made instruments. However, there is evidence that predictive accuracy is similar when comparing tools that predict unplanned hospital admissions.¹⁵ In any case, the high cost of social care means that compelling business cases for preventive programmes may still be made on the basis of less accurate predictions. Making forecasts using routine data would have the following important advantages.

- By being far less labour intensive to use than face-to-face instruments, such tools would be cheaper to administer and could therefore be applied systematically and repeatedly across a population.
- They would be less susceptible to non-response bias.
- They could stratify the whole population, not just those already perceived to be at high risk. This would enable large numbers of people with lower, emerging, risk to be identified and offered lower-intensity preventive care.

Listed below are the other advantages of tools that use routine data.

- Less susceptible to the inverse care law than a face-to-face tool (since practitioners in deprived areas may have less time to spend completing questionnaires with clients).
- Avoid the inverse equity hypothesis (which states that interventions that are targeted geographically to areas of deprivation tend first to be used by relatively less deprived people living in those areas).
- Have the potential to be used in resource allocation: part of the funding that a local authority receives from central government might be linked to the distribution of predicted risk across its local population.
- Have the potential to be used in performance assessment: local authorities could be assessed according to the impact that their preventive services have on the distribution of predicted risk across its population over time.

But there are a number of practical and ethical considerations that would have to be taken into account before such tools were built.

- Important issues of confidentiality and consent to consider if named individuals are to be assigned risk scores using routine data.

- Linking data sources at individual level across health and social care is particularly problematic since there is no ‘unique identifier’ used across both datasets to identify individuals.
- The tools are not 100 per cent accurate and arrangements should be put in place for people whose risk predictions are manifestly inaccurate.
- Data may be missing from routine databases on some individuals or groups (for example, many people are not registered with a GP). Again, arrangements would need to be put in place to support these people.
- The evidence base needs to be developed for ‘upstream’ interventions.

Advantages of predictive risk modelling

The main advantages of predictive risk modelling are:

- accurate predictions
- frequently refreshed
- whole population
- targeting of preventive interventions
- inverse care law
- inverse equity hypothesis.

ACCURATE PREDICTIONS

In the case of unplanned hospital admissions, the literature suggests that using the opinions of clinicians or applying simple criteria or thresholds do not produce sufficiently accurate predictions of who will benefit from ‘upstream’ care.¹⁶ Evidence suggests that more accurate predictions can be made using predictive risk models. The accuracy of a predictive risk model can be assessed by testing the algorithm against a validation sample that is removed at random from the data before the modelling process begins.

Accuracy is assessed in terms of the sensitivity and the specificity of the predictions. Sensitivity refers to the accuracy of the model insofar as it identifies those who *did* go on to need intensive social care. Specificity refers to the accuracy of the model insofar as it identifies those who *did not* go on to need intensive social care.

The distinction between sensitivity and specificity is important. A tool with high sensitivity identifies a large proportion of the people who will, without preventive intervention, go on to need intensive social care. However, if the specificity of the tool is poor then amongst the group of people that it identifies as high risk, many of these individuals would not, in fact, have gone on to need intensive social care. This means that the resources spent on preventive interventions for these people would be a waste. Therefore a tool with poor specificity results in reduced cost-effectiveness.

For example, for a population of 300,000, a predictive tool might identify 1,000 people as being at very high risk of admission to a nursing home next year. If the tool was not particularly accurate then it may be that only 250 people out of these 1,000 would truly have been admitted to a nursing home the following year without preventive intervention. Therefore, if the model was used to identify people for a preventive intervention, the local authority would have to target the intervention at 1,000 people in order to ensure the 250 people truly at risk were included. These 250 truly high-risk people are referred to as ‘true positives’, and the 750 people who were identified as high risk, but did not in reality go on to be admitted, are known as ‘false positives’.

The higher the false positive rate, the more costly and inefficient the model is to use in practice. A model with a higher specificity, which was better able to identify false positives, would therefore be preferable in this regard. A perfect model would have 100 per cent sensitivity and 100 per cent specificity. In this example, a perfect model would identify the 250 people who will be admitted to a nursing home next year (true positives) and identify the 299,750 people who will not (true negatives).

In reality no predictive model has 100 per cent sensitivity or 100 per cent specificity. Instead there is a trade-off between the two parameters. This can be illustrated graphically by plotting the Receiver Operating Characteristics (ROC) curve of the model. The specificity (false positive rate) is of particular use to commissioners when assessing the business models for preventive interventions for the reasons described above.

FREQUENTLY REFRESHED

Because predictive models use routinely collected data, once they have been set up they are not labour intensive to run. This means that they can be used on large numbers of people (potentially the entire registered population – see below) and can be run frequently, perhaps on a monthly basis or even more frequently. This is in contrast to tools that have to be administered face-to-face, which generally are only used on people perceived as being at high-risk people, and are only used once per person.

The ability to run a predictive tool frequently is useful for case-finding purposes because it provides a snapshot of risk across the population. This means that the people in the population who are at highest risk at that time will be offered the preventive intervention, reflecting the most up-to-date data available. When the tool is refreshed, preventive interventions can be re-targeted to those now at risk.

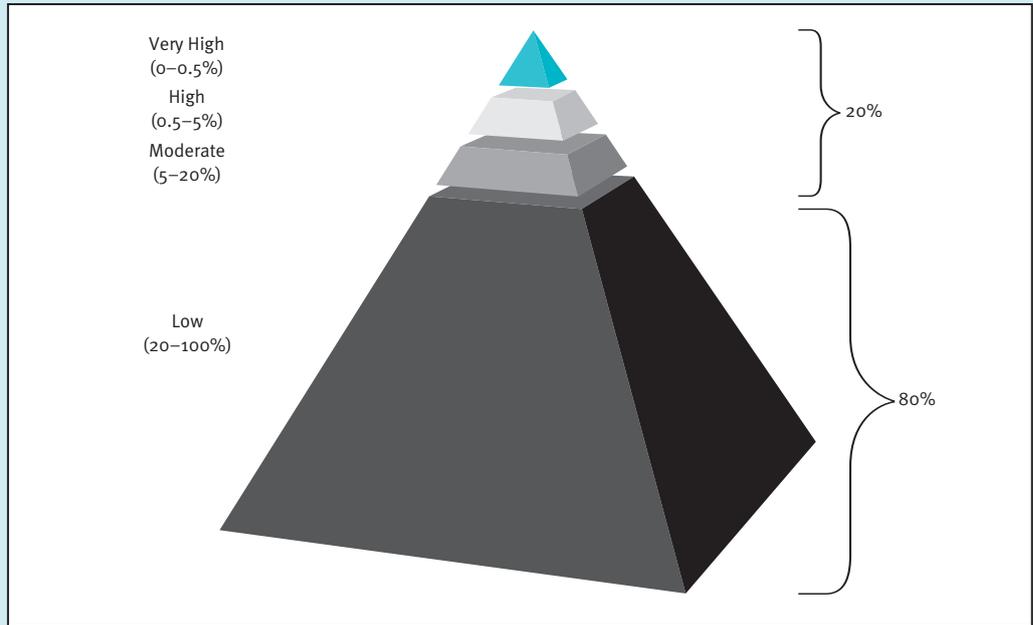
WHOLE POPULATION

The distribution of risk across a population is often represented as a pyramid (known in the case of long-term conditions as the ‘Kaiser Pyramid’, see Figure 2, overleaf). The broad base of the pyramid illustrates the large numbers of people in the population who are at low risk of experiencing the adverse outcome. At the top of the triangle are a much smaller number of people who are at very high risk of unplanned hospital admission.

The pyramid may be divided into four segments, with the top three segments combined representing the top quintile of risk (ie, the top fifth of the population), and the bottom segment representing the bottom four quintiles of risk combined (ie, the bottom four fifths of the population).

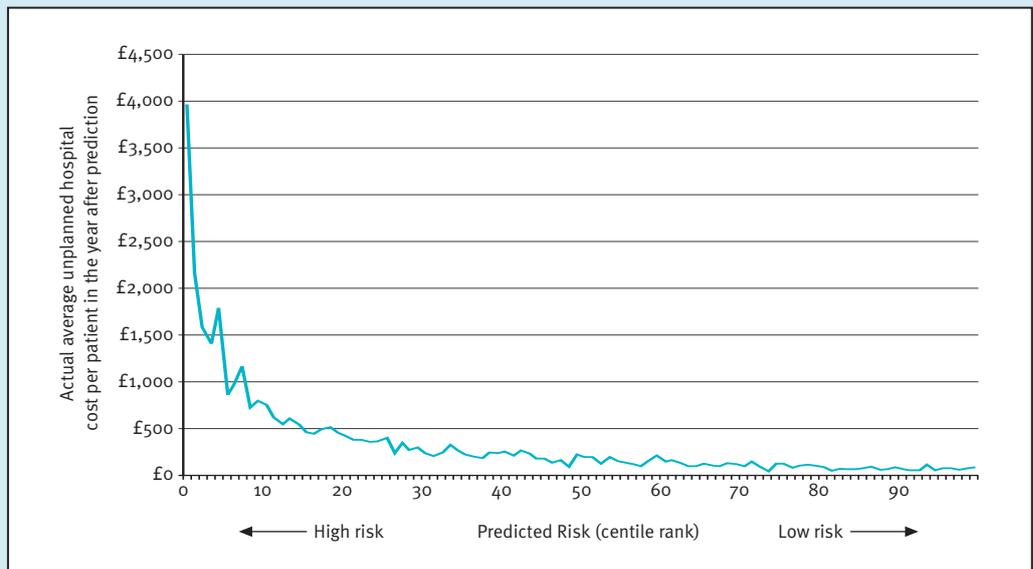
Figure 3 (see overleaf) shows how the actual ‘downstream’ costs of unplanned hospital admission are distributed exponentially across the population. In Figure 3, predicted risk is plotted along the horizontal axis (with patients at high predicted risk towards the left and patients at low predicted risk towards the right) and the vertical axis shows the actual costs that ensued in the twelve months following prediction. As can be seen, the costs for people at very high risk were about twenty times higher than average.

2 RISK PYRAMID



Source: Adapted from Kaiser Permanente

3 DISTRIBUTION OF COSTS ACROSS THE POPULATION AT RISK



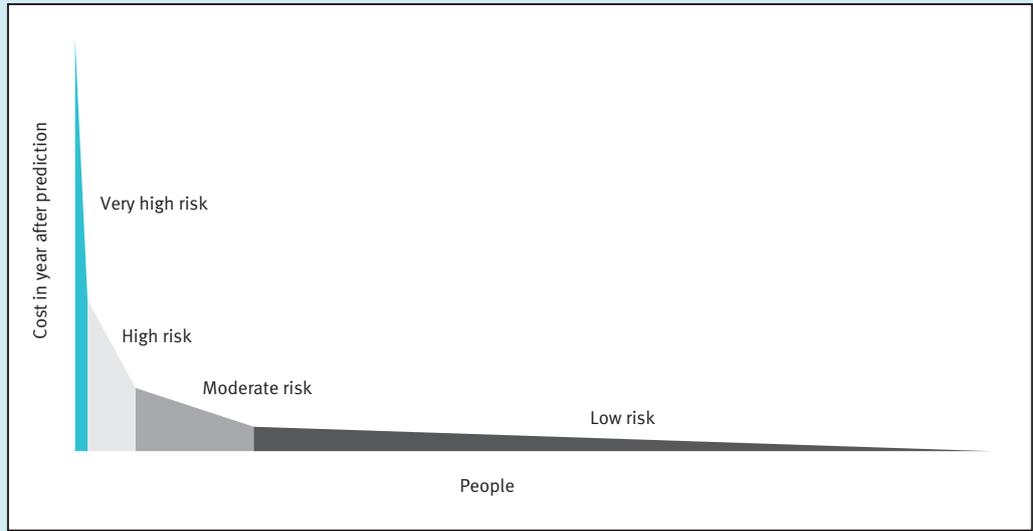
Source: Croydon PCT

The cost experienced by the local primary care trust (PCT) from unplanned hospital admissions is represented by the area under the curve in Figure 2.

Cost to the PCT = Number of People x Individual Cost

Using the same segments as those in Figure 2, the area under the curve (ie cost to the PCT) can be divided as shown opposite (see Figure 4).

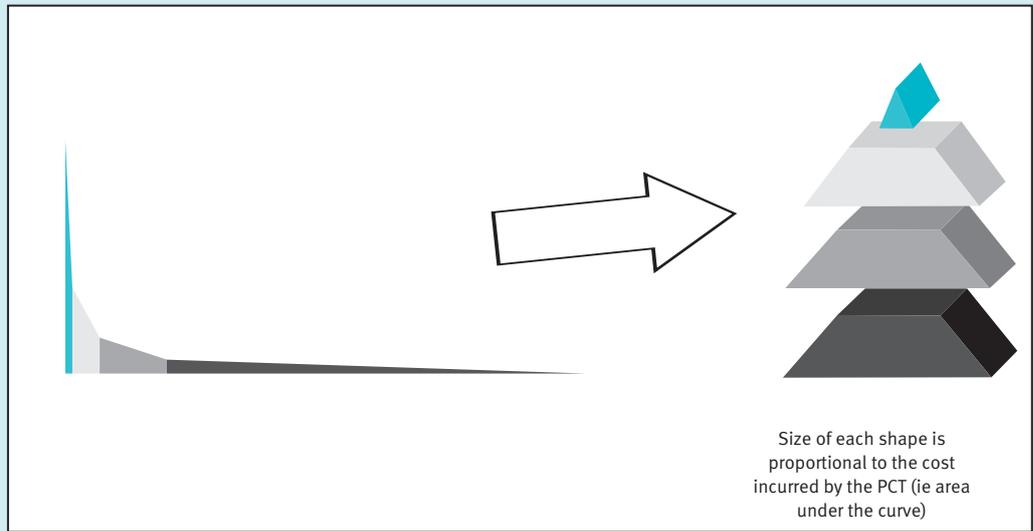
4 DISTRIBUTION OF COSTS ACROSS RISK SEGMENTS



Source: Author's work

The risk pyramid can now be re-drawn to illustrate the distribution of costs across the population. Now the size of each segment of the pyramid is made equal to the area under the curve above.

5 RISK PYRAMID SHOWING PROPORTIONAL DISTRIBUTION OF COSTS



KEY	
■	Very high risk
■	High risk
■	Moderate risk
■	Low risk

Source: Author's work with contribution from Health Dialog UK Ltd

As can be seen, although very high-risk individuals use disproportionately large amounts of resources, the bulk of future utilisation is actually due to people in the rest of the top quintile of risk (ie 0.5–20 per cent). For this reason it is important to ensure that preventive interventions are offered across the top quintile and not concentrated solely on those at very high risk.

For example, those patients who are at very high risk of unplanned hospital admission may be ‘case managed’ by a community matron or admitted to a ‘virtual ward’.¹⁷ People at intermediate risk (who between them represent the largest burden of population risk because there are larger numbers of such people) can be offered less intensive support, perhaps telephone-based health coaching. The bulk of the population, who are at much lower risk can be offered basic support (for example, written material and telephone helplines). By directing preventive interventions in this way, PCTs can offer a progressive universal service, that is to say a service that is available to all but which is offered at intensities proportionate to individual need.

TARGETING OF PREVENTIVE INTERVENTIONS

In terms of preventive interventions for social care, people at very high risk might be offered home adaptations and intermediate care at home to increase independence through rehabilitation. This could be delivered by a multi-disciplinary team. People at high predicted risk might be supported by a case manager or by a health visitor for older people – or they might be offered a package of temporary home care support. Those at intermediate risk might be offered minor home adaptations and tele-care. Community support could be made available to those at lower risk, perhaps delivered by third sector organisations.

One of the key advantages of predictive risk modelling is that it offers the possibility of stratifying the entire population, and of doing so repeatedly. This is in contrast to tools that are administered face-to-face.

INVERSE CARE LAW

By quantifying risk at the individual level it becomes possible to allocate preventive resources systematically, according to individual need. This offers a means of countering the inverse care law¹⁸ which states that the availability of good care typically varies inversely to the need for it across the population served. A predictive risk model (particularly a single, national model) could be used to allocate resources for preventive services according to predicted risk. Once preventive resources were allocated in this way then the tool could be used to assess the impact that preventive services were having on the distribution of risk across a population, that is, as a performance management tool for local authorities and PCTs.

INVERSE EQUITY HYPOTHESIS

Identifying risk at an individual – rather than geographic – level also offers a solution to the inverse equity hypothesis.¹⁹ This is the theory that when services are targeted to areas of high need, they are first used by relatively less needy individuals living in those areas. Therefore geographical targeting (as was used for the Sure Start Children’s Centres and for certain Partnerships for Older People Projects) may fail to reach those who would benefit the most. This is because many of the individuals at high risk can be difficult to engage (e.g. because of a distrust of services, or because of communication difficulties etc). So when resources are targeted geographically these people tend to lose out. In contrast, where risk is quantified at the individual level, individual budgets can be assigned to these people, and commissioners and practitioners can make a concerted effort to ensure their services are as appealing as possible to the high-risk individuals to encourage their engagement.

Problems with predictive risk modelling

Some of the main drawbacks with predictive risk modelling are:

- data quality
- data linkage
- ethics
- incomplete coverage
- low predictive power.

DATA QUALITY

Some of the data held in routine databases are of poor quality. Any systematic trends relating to data quality across a population may lead to biased predictions. However, since both the predictor and outcome variables are held within the routine datasets, the models are, to an extent, internally consistent. That is to say that similar errors will have been incorporated when building the model as when the model is used in practice. For health databases, initiatives such as *Payment by Results* and the *Quality & Outcomes Framework* have led to an improvement in data quality.

DATA LINKAGE

The tools produced as part of the King's Fund project to predict unplanned hospital admissions were built and implemented using encrypted health data. All patient-identifiable fields are first removed (for example, name, address and date of birth) leaving the NHS number and the clinical details. The NHS number (unique for each individual patient) is then encrypted by entering a password into an encrypting device. This converts the NHS number into an apparently meaningless, scrambled number. These records are then termed pseudonymous (ie apparently anonymous) and they can be exported to outside agencies whilst safeguarding confidentiality. When used in practice, the predictive tools assign a risk score to each pseudonymous NHS number. The pseudonymous NHS numbers with risk scores attached are then returned to the patient's GP. Only the patient's own GP, who holds the password used in the encryption process, can decode the pseudonymous NHS number back into a decrypted NHS number – and thence into a name and address. In this way the patient's own GP can then refer the patient to a service offering an appropriate intensity of preventive care.

The original intention for the Combined Predictive Model had been to include social services data as a source of predictor variables. Unfortunately only a very small proportion of social services records contained NHS number or identical demographic details for data linkage. For this reason, social services data were not included in the final version of the Combined Model.

Advances in data linkage technology mean that many records can now be linked successfully even where the demographic details are not identical. This process is called '*fuzzy matching*' and it recognises typical typographic errors, abbreviations and misspellings. A schema for merging health and social services data pseudonymously is shown on page (see page 28). The addition of social services data is likely to add considerable predictive power for forecasting who will have high-level social care needs. Individual social services assessments and self-assessments, although not conducted for all members of the population, could provide a particularly rich source of predictor variables.

ETHICS

As with any form of screening, predictive risk modelling may potentially violate all four of the principles of biomedical ethics.²⁰

ETHICAL ISSUES WITH PREDICTIVE TOOLS

Beneficence	The tool would not be 100 per cent specific, so some people (false positives) who are offered the additional preventive support will not benefit from the intervention they are offered
Non-maleficance	Patients may suffer anxiety from being told that they are at risk of losing their independence
Justice	The tool would not be 100 per cent sensitive so some people not detected by the tool (false negatives) may have benefited more from the preventive care than those who were offered the care
Autonomy	People are not asked to provide explicit consent to the use of their data in a population-based predictive model

Of these, it is the threat to autonomy that is often seen as the most serious ethical problem with predictive risk modelling in that it involves the sharing and processing of data without explicit consent.

When public bodies in England share data they must do so according to the principles established by the Data Protection Act 1998, the Human Rights Act 1998 and the Common Law duty of confidence.²¹ The second principle of the Data Protection Act 1998²² states that personal data should not be processed other than for the original purposes that they were obtained. However, recent guidance from the (now defunct) Department for Constitutional Affairs takes a more flexible stance. It states that,

In our view, the requirement of compatibility does not have to mean ‘identical to’ and provided the further processing is for a purpose that is not contradictory to the original purpose or purposes, it will be consistent with the second principle²¹

The Department of Health’s Patient Information Advisory Group (PIAG) has issued guidance setting out how the Combined Model can be used in an acceptable way without seeking consent from people to share and process their personal data.²³ This guidance stipulates that only GPs or other clinicians already personally known to each patient are entitled to use the output of the predictive tools.

INCOMPLETE COVERAGE

Not all members of the population will be covered by a predictive risk model, and those not covered will therefore be relatively disadvantaged. For example, the PARR algorithm (which makes predictions based on inpatient records) can, by definition, only make forecasts about people who have already had a recent hospital admission. The Combined Model has a broader coverage, but again it cannot make predictions about the significant

number of people who are not registered with a GP.

This problem of incomplete coverage will be important for any predictive model built to forecast the need for intensive social care. This is because not all patients will be registered with a GP, and because most people do not appear in social services records. Additionally, people who pay for their own social care will be disadvantaged because the outcome variable (for example, admission to a private nursing home) may not be recorded in any routine data sources. This is a growing group of people because the next generation of older people will on average be more affluent than previous generations, and so the majority of these people will be home owners who are not currently eligible for free social care.

LOW PREDICTIVE POWER

Since predictive risk modelling relies only on variables that are already collected in routine data, the technique might be expected to generate less accurate predictions than face-to-face tools that specifically enquire about factors that are already known to be predictive. However, the evidence from the literature regarding unplanned hospital admissions suggests that any difference in predictive power tend to be only marginal.

Determining which social care outcomes are important to predict

In the related field of managing long-term conditions, the specific outcome to be predicted is unplanned admission to hospital. This is an unambiguous outcome that is recorded accurately in a single dataset: inpatient data. For social care, the situation is more complex. Here the risk to be avoided is essentially that of functional decline and loss of independence. However, functional decline *per se* is not usually recorded in routine data, although there are many proxies (such as number of hours of care per week). The way in which loss of independence is recorded in routine data is also complex. For some people this event is recorded in several datasets (inpatient, social services, commitment monitoring), but for other people it may not be recorded at all (for example, for people who move electively from private accommodation into a private care home without any assessment or financial contribution by the statutory sector).

There are therefore several outcomes that could potentially be usefully predicted, including:

- direct admission to a care home from an acute hospital
- admission to a care home (nursing homes and/or residential homes)
- care needs as assessed by a standard instrument (for example, Fair Access to Care Services)
- increased social care costs
- increase in combined health and social care costs.

Factors predicting adverse outcomes in social care

A literature search was carried out as outlined in Appendix 1 (see p 49). Many studies were found that have determined which variables are statistically predictive of a range of adverse social outcomes including functional decline, institutionalisation, admission to a nursing home, need for home care and high costs. These studies used several epidemiological techniques, including prospective cohort studies²⁵ (the most commonly used method), retrospective cohort studies²⁴ case-control studies,³⁸ retrospective chart review studies,⁴² prospective observational studies⁴⁵ and analyses of surveys.⁶¹ One problem with these studies is that many of them were conducted on the same small group of databases, which may lead to systematic bias in their findings. Another problem is that many of the studies were based on a sub-section of the US population, so that their findings may not be applicable to the UK population. Noting these potential limitations, the factors that were found to be significantly statistically predictive of the above adverse outcomes are detailed below.

Predictors of functional decline

Several studies examined factors that were predictive of future impairment of an Activity of Daily Living (ADL). These activities include bathing, dressing, eating, transferring (ie moving in and out of a chair or bed), washing and toileting.

Most of these studies considered patients who had been admitted to hospital or were otherwise acutely unwell.^{25,26,27,28,29,30,31} Other studies considered people who were currently well and living in the community.³² Some studies set out to test specific hypotheses, for example, that nutrition,²⁷ depression²⁸ or unsteadiness³¹ were predictive.

The predictive factors found to be statistically significant are summarised in Table 1 overleaf.

TABLE 1: FACTORS STATISTICALLY PREDICTIVE OF FUNCTIONAL DECLINE

Predictive Factor (Functional Decline)	Number of Studies
Age	4
Mini-Mental Status Examination (MMSE) score	3
Prior Instrumental ADL score	3
Falls in the year before hospitalisation	2
Malnutrition	1
Self-reported steadiness	1
Inability to name the vice president	1
Difficulty walking	1
Difficulty bathing or dressing	1
Need for help with personal finances	1
Difficulty lifting 10 pounds	1
Poor self-rated health	1
Use of assistive device indoors	1
Diabetes	1
Timed 'up and go' greater than 40 seconds	1
Low BMI	1

Predictors of institutionalisation

Several studies looked at factors that were predictive of admission to a care home (residential home or nursing home) by monitoring new admissions to homes.^{33,34,35} Other studies examined the factors that predicted which patients, currently receiving home-based care, were at risk of future institutionalisation.^{36,37,38}

One study, which determined care home entry from death certificates, specifically examined how older people's financial resources affect their likelihood of admission to a care home in the United Kingdom. It found that there was no significant effect found for income, but that home ownership reduced the risk of institutionalisation.³⁹

TABLE 2: FACTORS STATISTICALLY PREDICTIVE OF INSTITUTIONALISATION

Predictive Factor (Institutionalisation)	Number of Studies
Age	3
Dementia/Cognitive impairment	2
ADL restriction	2
Number of family members	2
Use of day services	2
Incontinence	1
Co-morbidity	1
Sickness	1
Severe Disability	1
Malignancy	1
Consulting doctors at general hospitals	1
Temporary nursing home assistance	1
Housing conditions	1
Marital status	1
Walking ability	1
Night delirium	1
Mental disorientation	1
Age of primary caregiver	1
Living alone	1
Number of sub-caregivers	1
Number of rooms in house	1
Home ownership	1
Use of home help	1
Self-perceived health	1

Predictors of admission to nursing home

Many studies have examined what factors are predictive of admission to a nursing home *per se* (rather than to a care home – a term that includes both residential homes and nursing homes).

Many studies examined risk factors for nursing home admission amongst people recently discharged from an acute hospital,^{40,41,42,43,44} from a rehabilitation hospital⁴⁵ or from intermediate care.⁴⁶ Other studies examined risk factors predictive of nursing home admission amongst people currently living in the community,^{47,48,49,50,51} receiving home nursing or welfare services,^{52,53,54} those attending outpatients,⁶ or amongst people enrolled on an all-inclusive care programme.⁵⁵

Some studies examined the risk factors for specific groups in the population, for example, people living in public housing,⁵⁶ disabled people,^{57,58} in patients with heart failure,⁴² urinary incontinence⁴⁹ or Alzheimer's disease.⁵⁹

Certain studies compared risk factors for temporary admission to a nursing home versus permanent admission to a nursing home.^{60,47} One paper⁶⁰ found that patients with cognitive impairment (and especially those with a combination of cognitive impairment and functional impairment) were at particularly high risk of permanent admission to a nursing home. Another study⁴⁷ found that patterns of informal help in the community were important: older people who had no children and who lived alone or with unrelated adults were found to be especially prone to permanent admission to a nursing home.

Some studies aimed to test specific hypotheses. One paper examined the effect of lifestyle factors⁶¹ and found that smoking, obesity and inactivity were all predictive of nursing home admissions. Another paper tested the hypothesis that sensorimotor function was predictive⁴⁶ and found that the following were all associated with future admission to a nursing home: tactile sensitivity, ankle dorsiflexion strength and reaction time. Another paper⁴⁹ examined whether place of residence was predictive of nursing home admission in people with urinary incontinence: it found that such people living in rural areas were at particularly high risk.

A systematic review of 4,597 abstracts found 12 data sources that were suitable for meta-analysis.⁶² This found that the variables most strongly predictive of nursing home admission were: the presence of three or more dependencies of activities of daily living; cognitive impairment; and prior nursing home use.

TABLE 3: FACTORS STATISTICALLY PREDICTIVE OF NURSING HOME ADMISSION

Predictive Factor (Nursing home)	Number of Studies
Age	11
Mental impairment/confusion/MMSE/cognitive impairment	7
Living alone	6
Ethnicity	5
Geographic site	5
Dementia	4
Prior admission to a nursing home	4
Number of days in hospital	4
Diabetes	4
Psychosis/Mental disorder	4
Prior ADL dependence	4
Impaired mobility/unsafe gait/difficulty walking	3
General health/social activities limited by health	2
Home ownership	2
Recent fall	2
Combination of cognitive impairment and functional impairment	2
Gender	2
Visual impairment	2
Eight or fewer years of education	2
Urinary incontinence	2
Number of children	2
Caregiver's preference for institutional displacement	1
Impaired sensorimotor function	1
Wandering behaviour	1
Smoking	1
Physician use	1

TABLE 3: CONTD

Predictive Factor (Nursing home)	Number of Studies
Hearing impairment	1
Difficulty with communication	1
Amount of home nursing	1
Bowel incontinence	1
Number of household members	1
Using equipment for bathing	1
Difficulty shopping	1
Inactivity	1
Taking tranquilizers	1
General life satisfaction	1
Obesity	1
Income	1
Type of housing	1
Hypertension	1
Medications	1
Stroke	1
Parkinson's disease	1
Cancer	1
Depression	1

Predictors of Need for Home Care

Two studies were found that identify factors predictive of the need for home-based care.^{63,64}

TABLE 4: FACTORS STATISTICALLY PREDICTIVE OF NEEDING HOME CARE

Predictive Factor (Home-based care)	Number of Studies
Need	1
Low number of household members	1
Low educational attainment	1
Geographical area	1

Predictors of high-cost individuals

A number of tools have been developed to predict which patients are likely to incur high medical expenses in the year following prediction, using variables held in routinely collected data.^{65,66,67} They make predictions on the basis of a large number of variables, including age, number of co-morbid diseases, prescription drug use and use of health services.

These tools use similar data sources, methods and variables to the process of risk adjustment (ie prediction of likely costs in order to determine what health insurance premium should be charged for private health insurance), most of which are not found in the published literature.

Existing tools that predict risk of adverse outcomes

Tools that predict adverse outcomes can be administered either by posing questions to the person concerned, or by analysing factors contained in routinely collected data.

USING FACE-TO-FACE ASSESSMENTS

Using the predictive factors identified above, a number of assessment tools have been constructed that gauge the likelihood of functional decline or admission to a care home. Many such tools are designed to be administered to hospitalised patients to predict how they will fare in the period after discharge. Examples include the *Hospital Admission Risk Profile* (HARP) tool²⁵ and the SHERPA tool.²⁹

- The HARP uses three variables: age, Mini-Mental State Examination (MMSE) and Instrumental Activities of Daily Living (IADL) scores. Using these variables it assigns patients to low, intermediate and high risk categories. Patients in these categories subsequently experience a 19 per cent, 31 per cent or 55 per cent chance respectively of experiencing a functional decline following discharge from hospital.
- The SHERPA tool uses five factors to predict a drop of at least one point on the Activities of Daily Living (ADL) scale or IADL scale three months after discharge from hospital. The sensitivity and specificity of SHERPA are both approximately 70 per cent.

One tool³² predicts ADL dependence in people living in the community who are not acutely unwell. This is constructed using the *Asset and Health Dynamics Among the Oldest Old* (AHEAD) cohort.

- The AHEAD prognostic index is used in people who are currently able to bathe, dress, toilet, transfer and eat independently. It uses nine predictor variables (age >80, diabetes, difficulty walking several blocks, difficulty bathing or dressing, need for help with personal finances, difficulty lifting 10 pounds, inability to name the vice president, history of falling and low Body Mass Index) to assign a risk of losing independence in at least one ADL at two years after prediction. Patients with none of the five risk factors have 0.7 per cent chance of losing independence, compared with 40 per cent of patients with five risk factors.

A number of tools have been built using factors known to be predictive of nursing home admission. These include the Leeds Elderly Assessment Dependency Screening (LEADS) tool, the Medicare Current Beneficiary Survey (MCBS) tool and the Geriatric Postal Screening Survey (GPSS).

- The LEADS tool⁴¹ is a 17-item instrument that is administered to people within two weeks of admission to hospital. It assesses the risk of nursing home placement and its predictions are 88 per cent sensitive and 85 per cent specific.

- The MCBS instrument⁶⁸ uses 14 self-reported characteristics to predict need for long-term care (ie nursing home care or home-based care). Those deemed to be high risk according to the tool are six times more likely to require long-term care than those who are low risk.
- The 10-item GPSS has a sensitivity of around 75 per cent for identifying depression, falls and urinary incontinence. It also makes statistically significant predictions of which people are at high risk of admission to a nursing home in the forthcoming twelve months.

All of the above tools require the specific collection of data from patients or their carers (either face-to-face, over the telephone or by post).

USING ROUTINE DATA

With the exception of tools for predicting high-cost individuals, no tools were found that make quantifiable predictions of social care use in the next year based solely on routinely collected data.

In health care, tools are being used to make quantifiable predictions of risk (the risk of unplanned hospitalisation, for example). These tools are being used to improve the management of people with long-term conditions, for whom unplanned hospital admissions are seen as being undesirable, costly and potentially preventable. A number of 'upstream' interventions have been developed with the aim of preventing these admissions, for example, community matrons, virtual wards and health coaching.

The cost-effectiveness of programmes aimed at preventing unplanned hospital admissions is dependent on their being offered to people who are truly at risk of unplanned hospital admissions ('true positives'). For this reason, in 2004 the Strategic Health Authorities, the Department of Health and the NHS Modernisation Agency invited tenders to produce a risk prediction system for the NHS. The successful bid came from a consortium consisting of the King's Fund, New York University and Health Dialog Data Service (an American company specialising in health data analysis). The King's Fund predictive risk project was divided into three phases, the last of which was completed in December 2006.

KING'S FUND PREDICTIVE RISK PROJECT

Phase 1

A literature review that examined the evidence for different ways of identifying patients at risk of unplanned hospital admission. Three principal methods used were:

- taking referrals from professionals
- using threshold models, where referral criteria were set and any patients meeting the criteria were offered 'upstream' care
- predictive risk modelling, where patterns in routine data are used to predict who is at risk.

The literature review concluded that the most accurate predictions were made using predictive risk modelling.

Phase 2

This produced a predictive risk algorithm (called the Patients at Risk of Readmission – PARR) which was designed to be straightforward to use. PARR can be downloaded free of charge by primary care trusts (PCTs) from the King's Fund website. PCTs use it by entering inpatient data and census data for their population. The tool then predicts which people are at high risk of another hospital admission. Because it uses inpatient data, PARR can only make predictions about people who have already had a recent hospital admission: the tool cannot make predictions about the 90 per cent or more of the population who have not had a recent hospital admission.

Phase 3

The final phase of the project delivered a more sophisticated predictive tool, known as the Combined Predictive Model. This tool was designed to produce the most accurate predictions possible across the entire population. The Combined Model uses data from general practices, which are collected and collated differently across the country. For this reason the Combined Model has to be adapted to suit local circumstances and is therefore more complicated for PCTs to use than PARR.

Predictive risk modelling offers a way of stratifying a population according to individual risks of experiencing a particular outcome (such as an unplanned hospital admission or admission to a nursing home). The technique uses patterns in historical population data in order to build an algorithm. When this algorithm is then applied to current data, it generates quantified forecasts of future events with a known sensitivity and specificity.

If a predictive risk model were to be built for forecasting people at risk of nursing home admission, then the steps involved in its construction would be as follows:

BUILDING A PREDICTIVE MODEL TO FORECAST NEED FOR SOCIAL CARE

1. Specify the exact outcome that is to be predicted (for example, admission to a nursing home in the next 12 months). This will be called the 'outcome variable'.
2. Identify a source of routine data that records the outcome variable (for example, local authority's commitment monitoring database).
3. Identify sources of routine data that contain a range of potential predictive factors (for example, housing data, GP practice data). These variables are known as the 'predictor variables'.
4. Obtain three years' worth of each of these databases (for example, commitment monitoring, housing and GP databases). The earliest year is called year one, followed by year two and year three (the most recent year).
5. Link the databases together horizontally at person level (ie one record for each person showing all of the commitment monitoring, housing and GP variables) and vertically (ie chronologically over a three-year period: historical years one, two and three).
6. Remove at random the records of 50 per cent of the people. These records are known as the 'validation sample' and will be used later to test the accuracy of the algorithm.
7. In the remaining data (known as the 'development sample'), identify all occurrences of the outcome variable in the most recent year (ie all new admissions to nursing homes that were recorded in the commitment monitoring database for year three).
8. Test a large range of potential predictor variables held in year one and two for these individuals (ie factors in the housing and GP records such as age, diagnoses, type of dwelling). Also test transformed variables (for example, the square root of a predictor variable) and test a range of composite variables (ie a combination of several predictor variables weighted in different ways).
9. Test the statistical significance of these predictor variables (and also composite variables and transformed variables) in years one and two against the outcome variables in year three. The technique normally used is regression analysis: an attempt to fit a model to the observed data in order to quantify the relationship between the predictor variables and the outcome variable.
10. Develop an algorithm that incorporates each of the statistically significant predictor variables, with each variable weighted according to its predictive power. In this way, the model is designed empirically, without any prior assumptions – meaning that any information that helps predict future admissions will be used. This algorithm can then be used to predict future outcomes.
11. Returning to the validation sample that was removed earlier, apply the algorithm to years one and two. Compare the predictions of the algorithm with the outcome variables in historical year three (ie admissions to a care home in the most recent year of data) to quantify the accuracy of the predictions.

BUILDING A PREDICTIVE MODEL TO FORECAST NEED FOR SOCIAL CARE (CONTD)

12. Ensure that the predictor variables did not 'over-fit' to the development sample, ie the predictions need to be generalisable. Quantify this by calculating the following rates in the validation sample:

- True positive (people predicted to have an admission who were in reality admitted to a care home)
- False positive (people predicted to have an admission who were not in reality admitted to a care home)
- True negative (people predicted not to have an admission, who in reality were not admitted to a care home)
- False negative (people predicted not to have an admission, who were in reality admitted to a care home)

13. Start using the algorithm on current data to make predictions about the future.

Sources of predictor variables

Predictive risk modelling requires both the outcome to be predicted (the ‘outcome variable’) and the predictive factors (the ‘predictor variables’) to be held in routinely collected data. In the previous King’s Fund project that predicted unplanned hospital admissions, the outcome variable was found in inpatient data, and the predictor variables were drawn from the following sources:

TABLE 5: SOURCES OF ROUTINE HEALTH DATA

Data Type	Example	Examples of Variables
Geodemographic	Census	Index of Multiple Deprivation
Hospital	Inpatient	Number of hospital admissions Principal diagnoses Length of stay
	Outpatient	Number and date of visits Name of clinic
	A&E	Number and date of attendances Diagnoses Treatment Disposal
Primary Care	GP data	Diagnoses Blood tests Blood pressure Medicines prescribed

It is important that as many of these variables as possible be available at individual rather than area level. This is in order to avoid the ecological fallacy, that is to say the error of assuming that relationships found in area-level data will also be found among individuals.

In the hospital predictions project, it was found that predictive power was maximised by using a combination of demographic, diagnostic, pharmaceutical, clinical and prior utilisation data. Geodemographic data (in this case the Index of Multiple Deprivation⁶⁹) were also added to increase predictive accuracy still further. Other geodemographic data that might potentially be added to new predictive models include the variables in Mosaic United Kingdom^{TM70} (a commercial classification system that assigns each UK household into one of 61 socioeconomic subtypes, based on factors such as income, housing and spending characteristics). However, to be useful they would need to be robust at small area level.

The data sources that might be used for modelling the need for intensive social care are listed below. They are grouped into local authority, health and geodemographic sources. A subjective estimate of the predicted power of their variables is shown: person-level data are more predictive.

TABLE 6: SOURCES OF ROUTINE HEALTH AND SOCIAL DATA

Data Type	Predictors	Outcomes	
Inpatient	2	Models A, E	health data
Outpatient	1	Model E	
A&E	2	Model E	
GP	3		
Commitment monitoring*	2	Models B, D, E	local authority data
Social services*	3	Models C, D, E	
Housing	2	Models D, E	
Census	1		geodemographic data
Mosaic United Kingdom™	2		

*sometimes merged into the same database

Accessibility of person-level data

Individual-level inpatient, outpatient and accident and emergency (A&E) data are made readily available in electronic form by the NHS Care Records Service through its Secondary Uses Service (SUS). SUS was established to provide encrypted patient-level data for purposes other than direct patient care. Such ‘secondary uses’ include health care planning, commissioning, public health, clinical audit, benchmarking, performance improvement, research and clinical governance.⁷¹

Unfortunately, GP data are much less readily available. Although GP records throughout the United Kingdom have been essentially paperless for many years, very few practices export their data routinely. A few primary care trusts (for example, Croydon PCT and the former South Warwickshire PCT) use Miquet to extract primary care data. Miquet is crown copyright software that enables the interrogation and extraction of data from different types of GP Practice systems using a common query language.⁷²

Another potential source of GP data for building a predictive tool is the Qresearch database.⁷³ This consists of the health records for over 9 million patients extracted anonymously from general practices spread across the United Kingdom. The database dates back to the early 1990s making it one of the richest primary care databases in the world.

Local authorities record their expenditure in commitment monitoring databases – either a standalone dataset or else housed within a broader social services database. Housing data is likewise held in electronic databases by local authorities.

Feasibility of data linkage

Health data from different sources can readily be linked because of the presence of a unique identifier, namely the NHS number. Health data from a number of sources (for example, hospital data and GP practice data) can also be linked pseudonymously as follows:

- all patient-identifiable fields (i.e. name, date of birth, address) are removed
- NHS number is encrypted into a pseudonymous number
- the same encryption key is used for all sources, meaning that for all sources of data, any NHS number is transformed into the same encrypted NHS number
- data can now be linked based on encrypted NHS number
- predictive modelling assigns a risk score to each encrypted NHS number
- decryption performed by the patient's own GP.

Most social care records do not contain NHS numbers, and are therefore more problematic to link than health data, or to link with health data. One solution is to use 'fuzzy matching' technology. This links data according to the name, address and date of birth, but makes allowances for slight discrepancies between the datasets. An example is where a person's address, date of birth and first name are identical but the surname is spelt 'Green' in one dataset but 'Greene' in another. Another example is where all demographic details are identical but the first name appears as 'Mohammed' in one dataset and is abbreviated to 'Mo' in another.

If data linkage were to rely on fuzzy matching, then it would be important to audit the rates of true and false positives and negative matches.

The schema overleaf shows how encrypted data from different sectors (for example, GP data and social services data) could be linked pseudonymously using fuzzy matching. As can be seen, the unique identifiers (NHS number or social services number) are encrypted separately, and then the data are split into:

- demographic information (name, address, date of birth)
- predictor variables (blood pressure, social services contacts etc).

The former are sent to a data linkage warehouse, where records are matched by means of fuzzy matching of name, address and date of birth. A unique key is then assigned to each matched record, thereby linking an encrypted NHS number to an encrypted social services number.

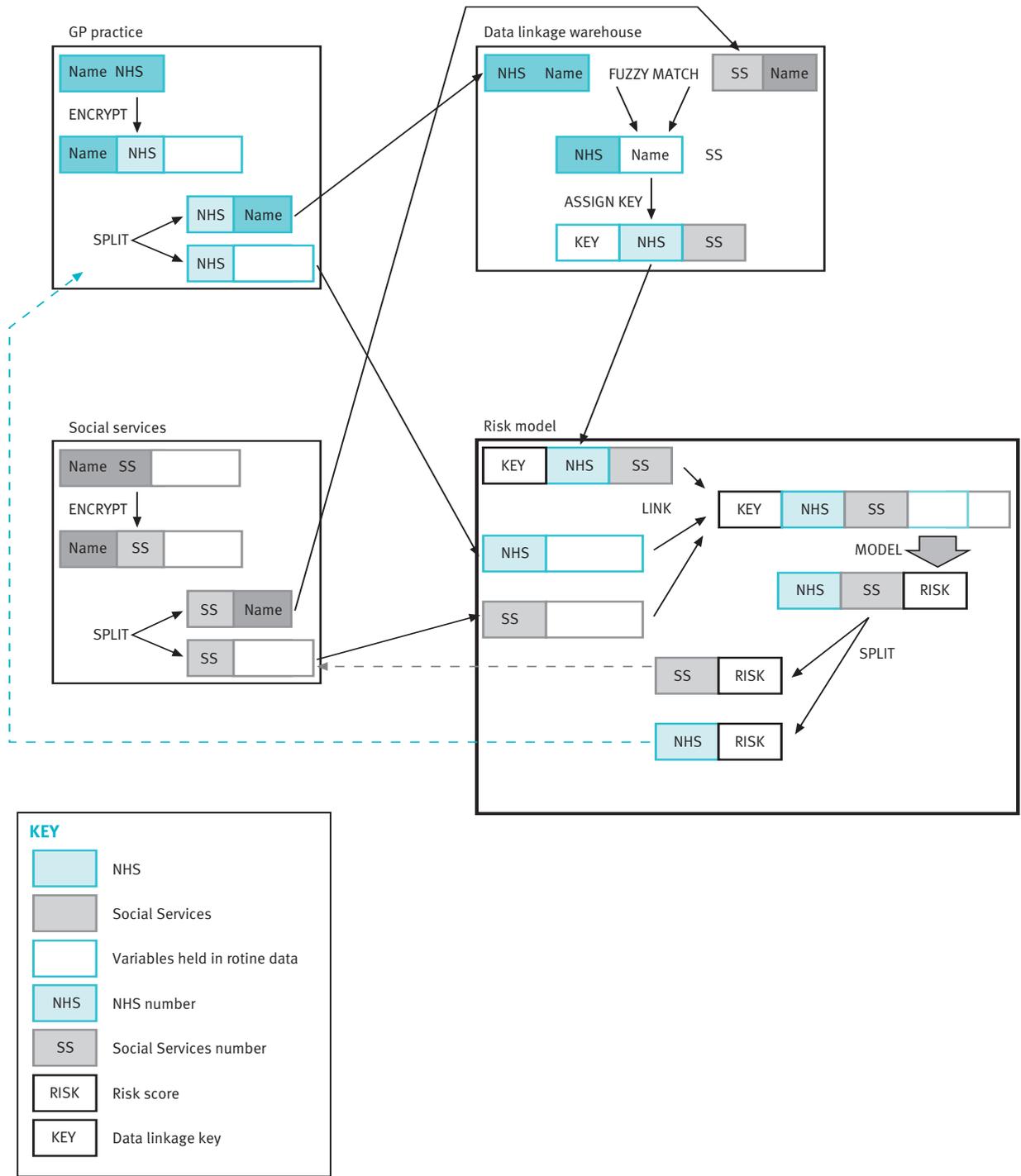
The latter are sent to the predictive risk model, where the unique key from the data linkage warehouse is used to match the variables from both sources.

The output of the predictive risk model is:

- a risk score linked to each encrypted social services number, which is sent back to social services for decryption; and
- the same risk score linked to each encrypted NHS number, which is sent back to the GP for decryption.

The schema ensures that data linkage and analysis only occurs on encrypted data. Risk scores for named individuals are only made available to professionals already known to them. This ensures compliance with the principles of consent that were set out by the Department of Health's Patient Information Advisory Group (PIAG)²³ to determine acceptable usage of the NHS Combined Predictive Model.

SCHEMA FOR MERGING HEALTH AND SOCIAL SERVICES DATA PSUEDONYMOUSLY



Developing risk prediction models

Five models are presented that could potentially be built in order to predict individual-level risk of requiring intensive social care in the year following prediction.

TABLE 7: MODELS FOR PREDICTING NEED FOR SOCIAL CARE

A	Acute model	Predicts admissions direct from acute hospital to care homes
B	Admissions model	Predicts admissions to care homes
C	FACS model	Predicts people whose needs will become 'critical'
D	Social care costs model	Predicts people whose social care costs will escalate
E	Health and social care costs model	Predicts people whose combined health and social care costs will escalate

The advantages and disadvantages of each model are discussed below. Several versions of each model are presented (A1, A2 etc) reflecting the range of potential data sources that might be used to build and run that model. As a general rule, adding more data sources makes the predictions of the model more accurate, but is more complicated logistically. There are particular problems when data sources from more than one sector (ie health and social services) are combined.

If a person moves away from his or her usual abode to take up residence in a care home in another part of the country, or to receive social care at a relative's home, then in certain circumstances it is the original local authority that continues to be responsible for paying the new social care costs. The models below deal appropriately with this scenario because both the predictor variables and the outcome variable will continue to be recorded in the original local authority's databases.

Model A (admissions to care homes from hospitals)

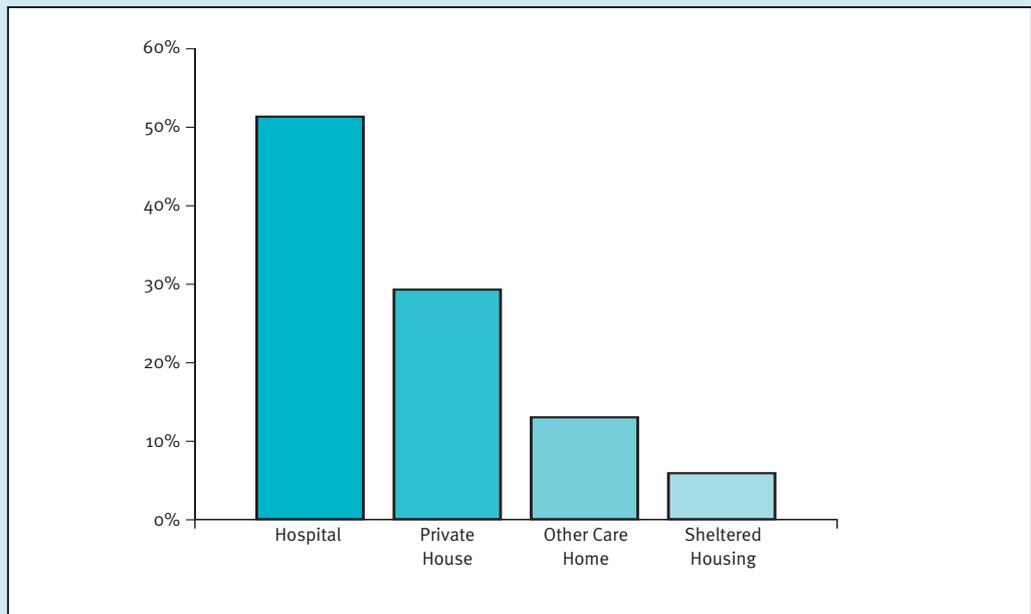
This model is designed to predict admission from an acute hospital admission direct to a nursing home. The model is proposed for two reasons:

1. an admission from an acute hospital direct to a nursing home is seen as a particularly adverse event; and
2. this event should be relatively straightforward to predict in terms of the data sources needed.

An estimated 51 per cent of older people permanently entering nursing home care, and 43 per cent of those entering residential care homes, come direct from an acute hospital stay.⁷⁴

6

PROVENANCE OF PEOPLE ADMITTED TO A NURSING HOME



Source: Bebbington *et al* 2001⁷⁵

Such admissions are contrary to guidance from the National Service Framework for Older People.⁷⁶

- Patients who are acutely unwell are likely to have little input into the choice of placement. A decision on their long-term care is taken at a time when they have not fully recovered and they may be rushed into making life-changing decisions at a time when they are most vulnerable. Predicting who might be in this position in the future would help the person and their family to plan in advance what they would want to happen if this situation occurred, as well as acting as an incentive to the PCT and local authority to ensure that there are no gaps in the quality of the care that they are currently receiving.
- Patients who are instead offered intermediate care tend to have much improved outcomes.
- ‘...intermediate care should be used as an opportunity to maximise people’s physical functioning, build confidence, re-equip them with the skills they need to live safely and independently at home, and plan any on-going support needed. Rehabilitation reduces the risk of older people being readmitted to hospitals or being placed in long-stay residential care and improves survival rates and physical and cognitive functioning’ (National Service Framework for Older People).⁷⁶

If such a model were to be built then both the outcome variable of interest (ie admission from acute hospital to nursing home) and many of the potential predictor variables would be held in a single data source, namely, the Secondary Uses Service (SUS).⁷⁷ SUS data are collected routinely and can be accessed centrally, making such a model relatively straightforward to run. Geodemographic data (such as Index of Multiple Deprivation or Mosaic™ scores) could be added to improve the predictive accuracy.

The simplest version of this model (model A1) would be closely analogous to the PARR⁷⁸ algorithm. Like PARR, this model would have the advantage of being straightforward to download and use but, again like PARR, it would be disadvantaged in that it could only make predictions about the subset of the population who have had a recent hospital admission.

A complicating issue is that the model should predict new admissions to nursing homes (and not existing nursing home residents who are admitted to hospital and then return to the nursing home). The simplest way to do this would be to use the data fields in the inpatient data which record the provenance of the patient: patients whose provenance was a nursing home could thus be discarded from the predictions. Unfortunately, provenance is poorly recorded in SUS, with the provenance of over 97 per cent of patients simply being recorded as ‘usual place of residence’. For example, an analysis of all hospital admissions for residents of the London Borough of Croydon in the year commencing October 2005 showed that no patients at all were recorded as having come from an NHS-run care home, and that only three patients were recorded as having come from a non-NHS-run care home.⁷⁹

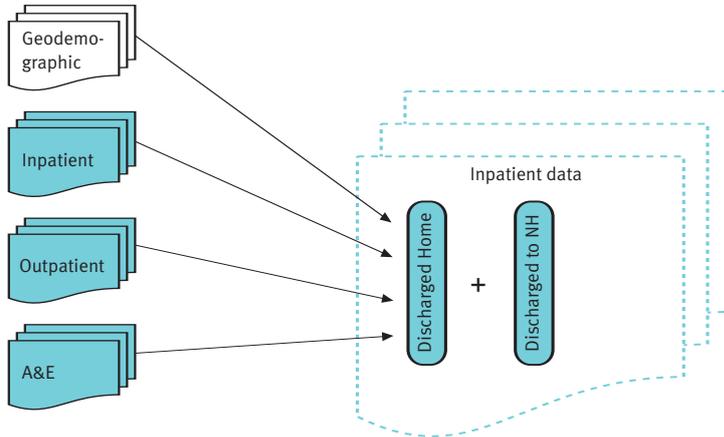
Fortunately, the quality of the discharge destination is much better. The same analysis of Croydon residents showed that 27 patients were discharged from hospital to an NHS-run care home in this time period, and that 357 patients were discharged to a non-NHS-run care home.⁷⁹ This, therefore, offers the possibility of using paired hospital admissions as the event to be predicted, that is, the model would identify admissions where the patient was discharged to the ‘usual place of residence’ (known as a ‘triggering admission’) and then predict the risk of a subsequent hospital admission for these patients where the patient was discharged to a nursing home. This is analogous to the PARR tool which uses any hospital admission as the triggering admission and then predicts the risk of any readmission.

Models A2 and A3 incorporate extra data sources. The addition of GP data (model A2) can only be achieved where GP data are extracted routinely. Currently this is done routinely only in a small number of PCTs. Since GP data have an NHS number attached, it means that they can readily be linked to other NHS data (such as inpatient, accident and emergency (A&E) and outpatient attendances). The addition of social services data (model A3) might be expected to increase predictive accuracy significantly because of the availability of pertinent predictive variables relating to housing, social services etc. However, social services data do not routinely include NHS numbers, that is, there is no unique identifier. As was discussed above, this can potentially be overcome by using ‘fuzzy matching’ techniques.

Advantages	Disadvantages
<ul style="list-style-type: none"> ■ Predicts a particularly egregious outcome and therefore will help guide preventive measures aimed at avoiding this outcome ■ Outcome is recorded in SUS data (which are held centrally and are readily available) 	<ul style="list-style-type: none"> ■ Can only make predictions about the small proportion of people who have had a recent hospital admission ■ As practice changes, the number of people experiencing this event should dwindle rapidly so this model would only be useful in the short term

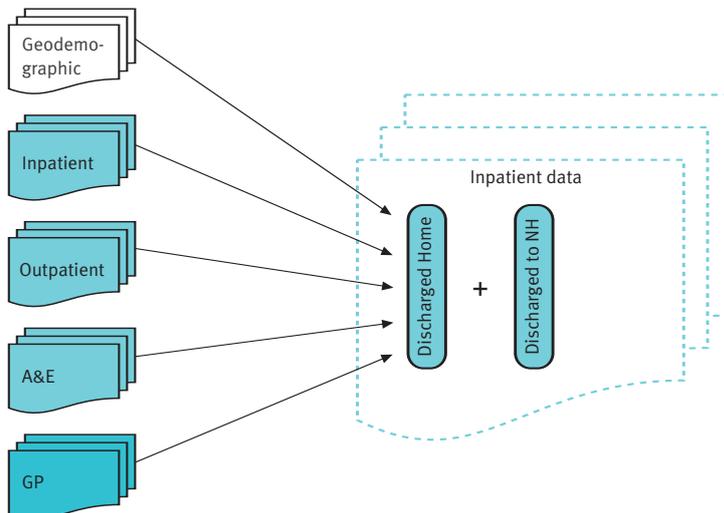
MODEL A1

This model is analogous to PARR in that it only uses SUS and geodemographic data. As discussed above, it uses the following as a triggering event: admission to hospital where the discharge destination was 'usual place of residence'. Then, for each triggering event, it predicts the risk of the following outcome: readmission to hospital where the discharge destination will be a nursing home.



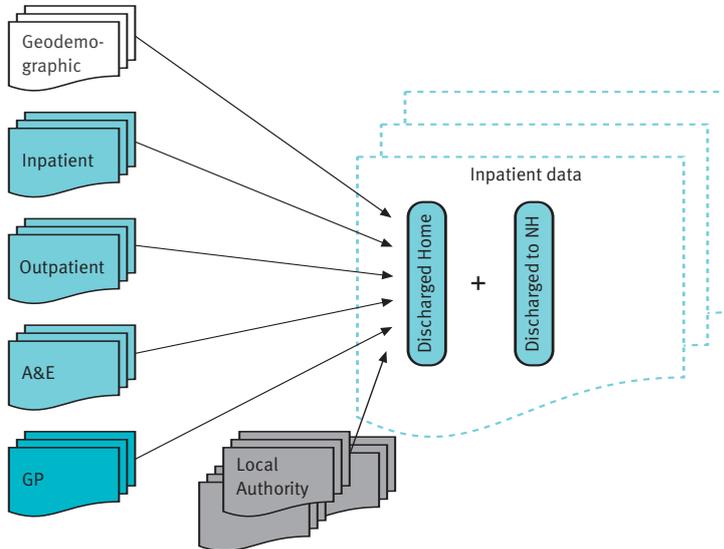
MODEL A2

Model A2 is analogous to the existing Combined Model in that it uses SUS, geodemographic and GP data. Again it uses the following triggering event: admission to hospital where the discharge destination was 'usual place of residence'. For each triggering event the model predicts risk of the following outcome: readmission to hospital where the discharge destination will be a nursing home.



MODEL A₃

This model is the same as A₂ but adds additional predictor variables from local authority databases (for example, social services data and housing data).



Model B (admissions to care homes)

This model is designed to predict admission to a care home. It includes admissions from any provenance (ie not just those people who are admitted to a care home direct from hospital). In terms of predicting loss of independence, model B is therefore the model of choice. However, there are several complicating factors that should be borne in mind if this model were to be built:

- Many packages of home care are now more intensive and more expensive than nursing home care. If the principal outcome of interest is cost then the location where that care is delivered becomes less relevant (see model C).
- It may be preferable for model B to distinguish between residential homes and nursing homes. This is because, overall, the cost of the latter is significantly higher than that of the former. However the difference in cost to local authorities is relatively small now that the NHS should be funding the nursing element of nursing home care.
- The literature suggests that there is a dichotomy between short-term and long-term nursing home patients.⁴⁷ From a costs perspective it would be important for the model to distinguish between the two groups.
- Whilst admission to a care home is not the preferred choice for most people, it does remain the choice for some. Indeed, there is evidence⁸¹ that those people with the most intensive care needs that remain at home often experience significant levels of isolation because they are housebound.

A major problem with model B is that there is currently no overarching database that contains details of all admissions to all care homes. Many self-funding people will not be known to the local authority at all. A tool that aimed to predict loss of independence, but which only used outcome variables held in local authority databases would therefore miss

these people. Nonetheless, for the majority of people for whom the local authority funds all or part of the care, details of their admission to a care home are held in the local authority's commitment monitoring database. This database is either integral to the main social services database, or it is a standalone local authority database that can be readily linked to the social services database.

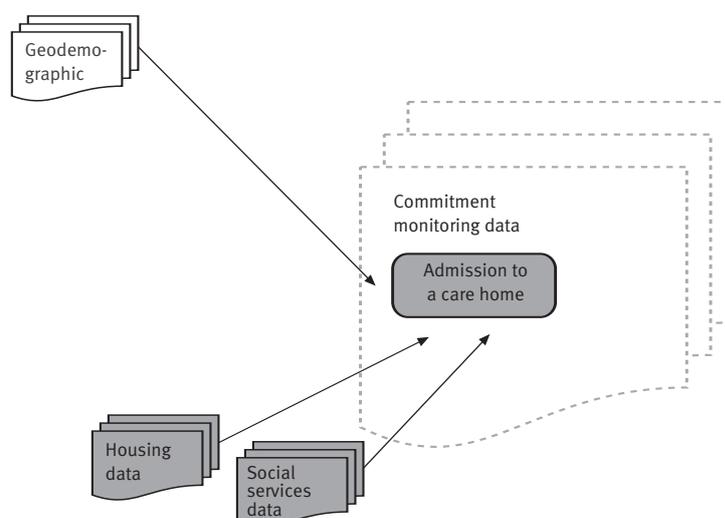
In the simplest version of this model (model B1), only geodemographic data and local authority databases are used. Local authority databases would include the social services database, the commitment monitoring database (if this is separate from the social services database) and the housing database. An advantage of model B1 is that these databases are all held by the local authority, and may indeed already be linked. The disadvantage is that the range of predictor variables will be restricted, particularly for people not living in local authority controlled housing or not previously known to social services.

Models B2 and B3 add health service data. The addition of these data is likely to improve predictive accuracy, but it entails sharing data across organisations – and it will be complicated by the lack of a common identifier. Model B2 adds SUS data (inpatient, outpatient and A&E data); model B3 adds GP data and therefore can only be implemented where GP data are extracted routinely.

Advantages	Disadvantages
<ul style="list-style-type: none"> ■ Predicts all care home admissions, not just the minority that occur direct from hospital ■ Model of choice if the outcome to be prevented is loss of independence 	<ul style="list-style-type: none"> ■ Some packages of intensive home care are more costly than nursing homes, so if the outcome to be prevented is intensive social care then the location of where that care is delivered becomes less important ■ Known dichotomy between short-term and long-term residents of care homes may be difficult to discriminate by the model ■ Misses self-funding people

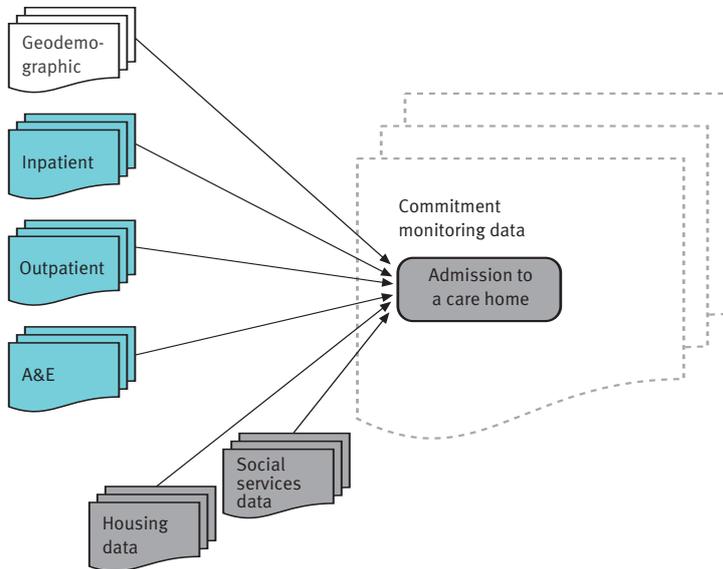
MODEL B1

This model would be relatively straightforward to implement, but it includes only a limited range of predictive variables. It can only make predictions about people who are already known to the housing or social services departments of the local authority.



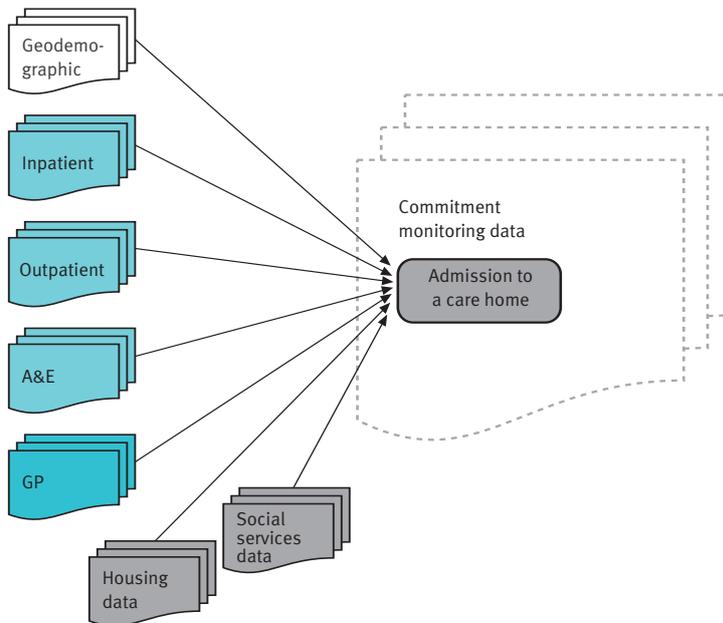
MODEL B₂

Here a richer range of predictive variables is available. SUS data are held centrally and are therefore easy to access. However, model B₂ would involve issues regarding data linkage and data sharing.



MODEL B₃

This version includes a very rich range of predictive variables, but it could only be implemented where GP data are routinely extracted.



Model C ('critical' status on FACS)

This model is designed to predict the need for intensive social care regardless of where that care is provided. The principal advantage of this model is that it would predict need for intensive social care regardless of where that care will be provided. In this way it addresses the fact that certain packages of home care are more intensive and costly than nursing home care. The outcome variable for model C is a *de novo* classification of 'critical' on the Fair Access for Care Services (FACS) scale. FACS is a national benchmark of need for social care, with 'critical' being the highest band.

'The [FACS] framework is based on individuals' needs and associated risks to independence, and includes four eligibility bands – critical, substantial, moderate and low... At the heart of the guidance is the principle that councils should operate just one eligibility decision for all adults seeking social care support; that is, should people be helped or not?' (Department of Health 2003)

A person is assigned to the highest band ('critical') on FACS when:

- 'life is, or will be, threatened; and/or
 - significant health problems have developed or will develop; and/or
 - there is, or will be, little or no choice and control over vital aspects of the immediate environment; and/or
 - serious abuse or neglect has occurred or will occur; and/or
 - there is, or will be, an inability to carry out vital personal care or domestic routines; and/or
 - vital involvement in work, education or learning cannot or will not be sustained; and/or
 - vital social support systems and relationships cannot or will not be sustained; and/or
 - vital family and other social roles and responsibilities cannot or will not be undertaken.'
- (Fair Access to Care Services, Department of Health 2003)

As the FACS is a national classification, it means that this model could potentially be used across England, and therefore might be used in future for resource allocation and performance management uses.

TABLE 8: UNIT COST OF SERVICES FOR OLDER PEOPLE (PER WEEK, HEALTH AND SOCIAL CARE COSTS COMBINED)

Private nursing home (nursing costs excluded)		£605
Residential home	Private	£430
	Voluntary sector	£426
	Local authority	£749
Sheltered housing (comprehensive package)	Local authority	£234
	Housing authority	£293
Community care package	Very low intensity	£333
	Low intensity	£348
	Median intensity	£589
	High intensity	£993
	Very high intensity	£1,201

Source: Unit Costs of Health and Social Care (2006)⁸²

However, there is a perception that there are major inconsistencies in how FACS is applied across the country from one local authority to another, for different age groups and for different conditions.^{83,84} Furthermore, the Inverse Care Law suggests that practitioners in deprived areas will have fewer resources to apply the FACS methodically and that this model might therefore lead to a worsening of inequalities if it were used for that purpose.

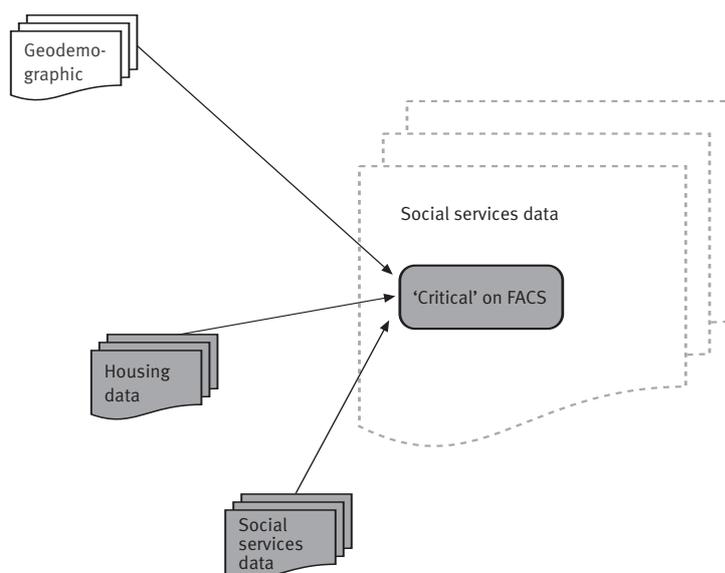
Since some people who self-fund their own care may not undergo a FACS assessment, this model cannot make predictions for these people.

As with model B, the simplest version of this model (model C₁) uses only geodemographic data and local authority databases for predictor and outcome variables (FACS is recorded within the social services database). Models C₂ and C₃ add increasing amounts of health service data. Again, the addition of these data is likely to be helpful in terms of predictive accuracy, but it entails data sharing across organisations and will be complicated by the lack of a common identifier in the databases. Model C₃ can only be used where GP data are extracted routinely.

Advantages	Disadvantages
<ul style="list-style-type: none"> ■ Predicts need for intensive social care, regardless of where that care is delivered 	<ul style="list-style-type: none"> ■ FACS is applied inconsistently across the country ■ Does not predict loss of independence <i>per se</i> ■ Discriminates against self-funding people ■ Potential to exacerbate the <i>Inverse Care Law</i>

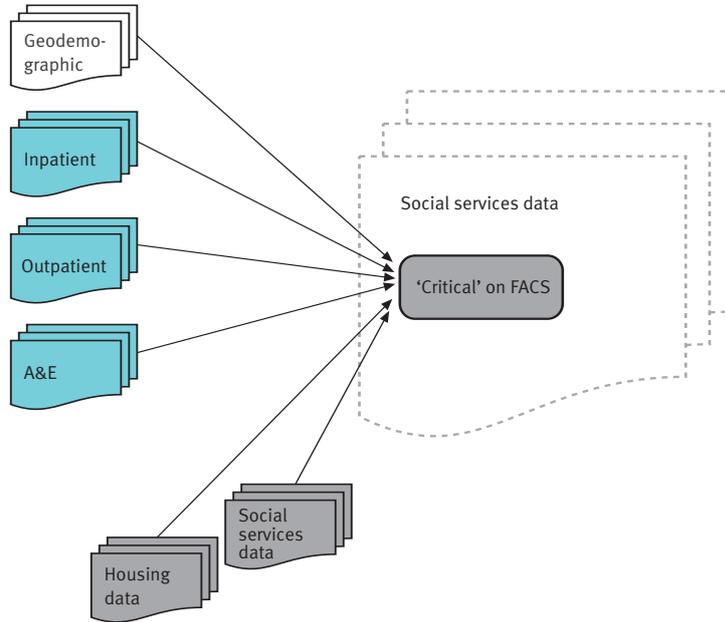
MODEL C₁

As with B₁, this model would be relatively straightforward to implement, but has a limited range of predictive variables and can only make predictions about people already known to the local authority's housing or social services departments.



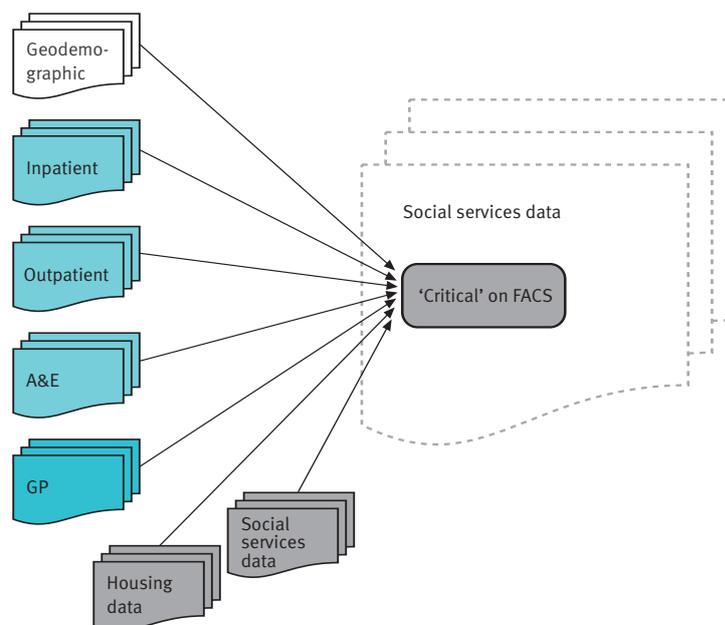
MODEL C₂

This model, which adds SUS data, would be more complex to implement but is likely to be more accurate than model C₁.



MODEL C₃

This model could only be used in areas where GP data were routinely extracted, but is likely to be significantly more accurate than models C₁ and C₂.



Model D (social care costs)

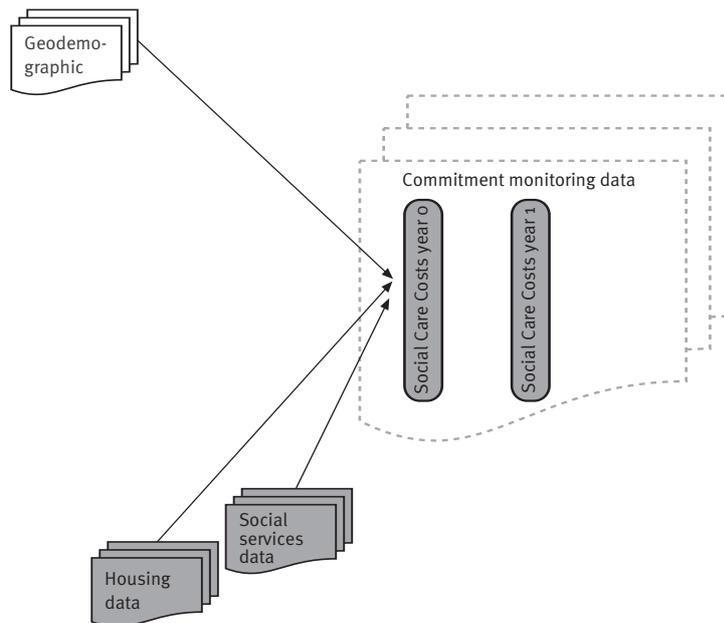
This model predicts which people are at high risk of experiencing a large rise in the cost of their social care (from the perspective of the local authority) in the coming year. As with model C, this model predicts the need for intensive social care regardless of where that care is delivered. This model avoids the potential problems of inconsistency relating to the FACS. Predictions instead would be based on rises in social care costs, where such a rise acts as a proxy for a rise in need.

Rises in costs can be considered in relative or absolute terms (ie percentage change or rise in currency terms). The latter is likely to be more pertinent to commissioners of preventive services. Since this model only considers social care costs and not health care costs, it risks creating perverse incentives to shift costs from local authorities to the NHS (in the same way that critics warn that the PARR tool and Combined Model risk shifting costs in the opposite direction).

Advantages	Disadvantages
<ul style="list-style-type: none"> ■ Predicts need for intensive social care, regardless of where that care is delivered ■ Presents information in stark monetary form thereby promoting business case development ■ Avoids problem of inconsistent application of FACS 	<ul style="list-style-type: none"> ■ Does not predict loss of independence <i>per se</i> ■ Discriminates against self-funding people ■ Potential problems with public perception ■ Risk of cost-shifting between health and social care

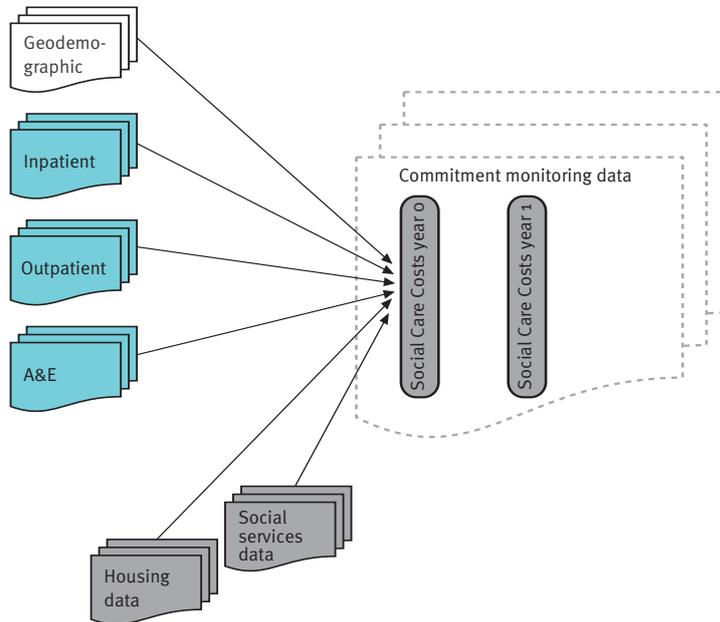
MODEL D₁

This is the simplest model, using local authority and geodemographic data only to predict a large absolute rise in social care costs in the year after prediction.



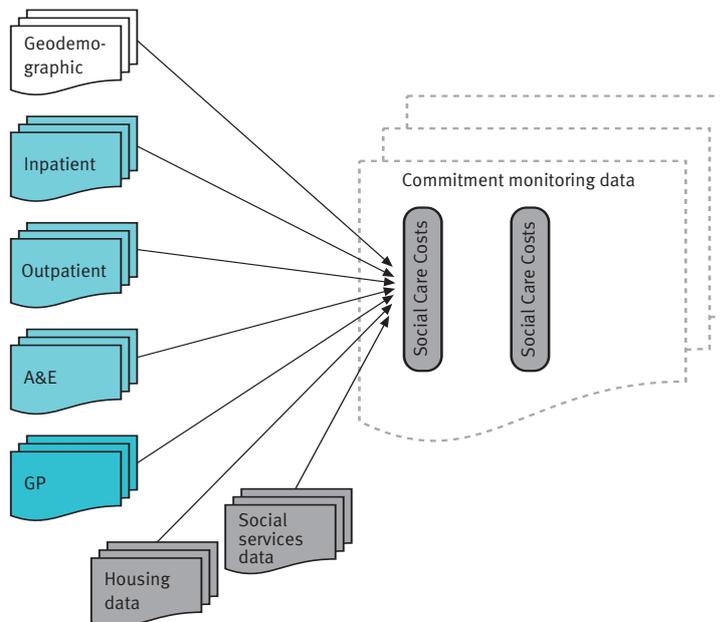
MODEL D₂

This model adds SUS data.



MODEL D₃

This is the most complex model, using local authority, geodemographic, SUS and GP data. It could only be used in areas that extract GP data routinely.



Model E (combined NHS and social care costs)

This model predicts which people are at high risk of experiencing a large absolute rise in their combined health and social care costs (from the perspective of the local authority and the local PCT) in the coming year. As with model C, this model predicts the need for intensive social care, regardless of where that care is delivered. This model avoids the potential problems of inconsistency relating to the FACS and avoids the potential for cost-shifting between health and social services. It is likely to lead to increased joint investment, and increased co-operation between health and social services in the provision of preventive care.

Predictions are based on absolute rises in health and social care costs, where a rise in cost is taken as a proxy for a rise in need.

Advantages	Disadvantages
<ul style="list-style-type: none"> ■ Predicts need for intensive social care, regardless of where that care is delivered ■ Presents information in stark monetary form thereby promoting business case development ■ Avoids problem of inconsistent application of FACS ■ Avoids risk of cost-shifting between health and social care ■ Promotes joint investment between health and social services ■ Promotes co-operation between health and social services in the delivery of preventive services 	<ul style="list-style-type: none"> ■ Does not predict loss of independence <i>per se</i> ■ Misses self-funding people

Figure 7, overleaf, courtesy of Health Dialog UK Ltd, shows how PCTs can build business models for preventive care interventions based on the predictions of the NHS Combined Predictive Model. The bar in Figure 7 allows commissioners to select any segment of the population according to risk of unplanned hospital admission. In the figure, this is currently set at 0–5 per cent (ie the top 5 per cent of the population at highest predicted risk). In the figure this equates to the 16,125 people at highest risk out of the population of 322,518.

Commissioners then set how much money they wish to invest in preventive care for these 16,125 people. In Figure 8 this is currently set at £100, that is a total of £1.6 million to be invested in these 16,125 people. The panel at the bottom of the figure contains details of four adverse outcomes that the preventive intervention is designed to avoid. In the case of the Combined Model these outcomes are:

- inpatient emergency admissions
- Inpatient other hospital admissions
- A&E attendances
- outpatient visits.

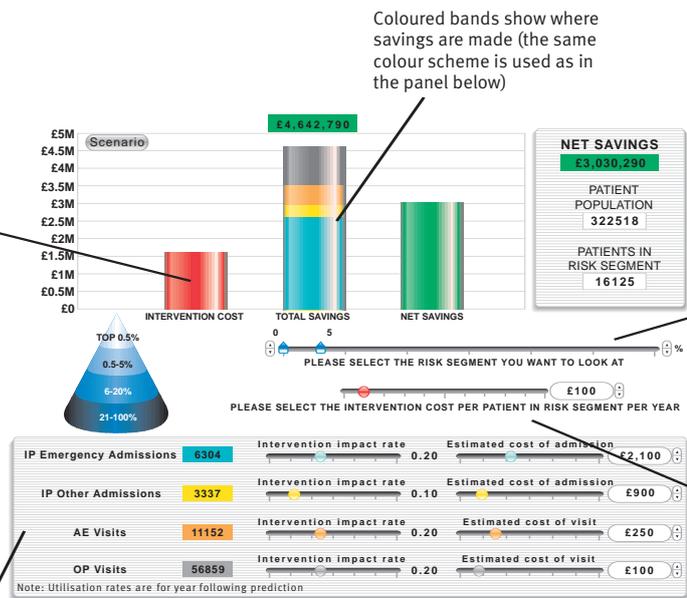
Commissioners set the cost of these adverse outcomes (for example, £2,100 for an emergency inpatient admission), and they set the estimated effect that the intervention will have on avoiding this outcome (for example, expected 20 per cent drop in emergency inpatient admissions as a result of the intervention).

Once all of these assumptions have been set, the tool uses the known sensitivity and specificity of the Combined Model to calculate the savings that the intervention will yield next year. For the example given the gross savings will be £4.6 million, giving a net saving of £3 million when the upfront costs of the intervention are subtracted. The coloured bands indicate where the savings will be made (for example, £2.6 million will be released from reductions in emergency inpatient admissions).

If model E were to be built, then a similar commissioning tool could be constructed. Here the ‘downstream’ savings would need to be grouped into savings from health and savings from social care. The relative savings to the local authority and to the PCT would then be apparent, meaning that the upfront investment costs could then be divided pro rata by the two organisations (see Figure 8, opposite). These relative proportions would vary across different segments of risk.

7 COMMISSIONING ‘UPSTREAM’ INTERVENTIONS

If model E were built, this bar would be subdivided into health and social care costs, in proportion to the savings made by each sector. These relative proportions would vary across risk segments of the population. In this way, health and social care would invest amounts that were proportionate to the savings released



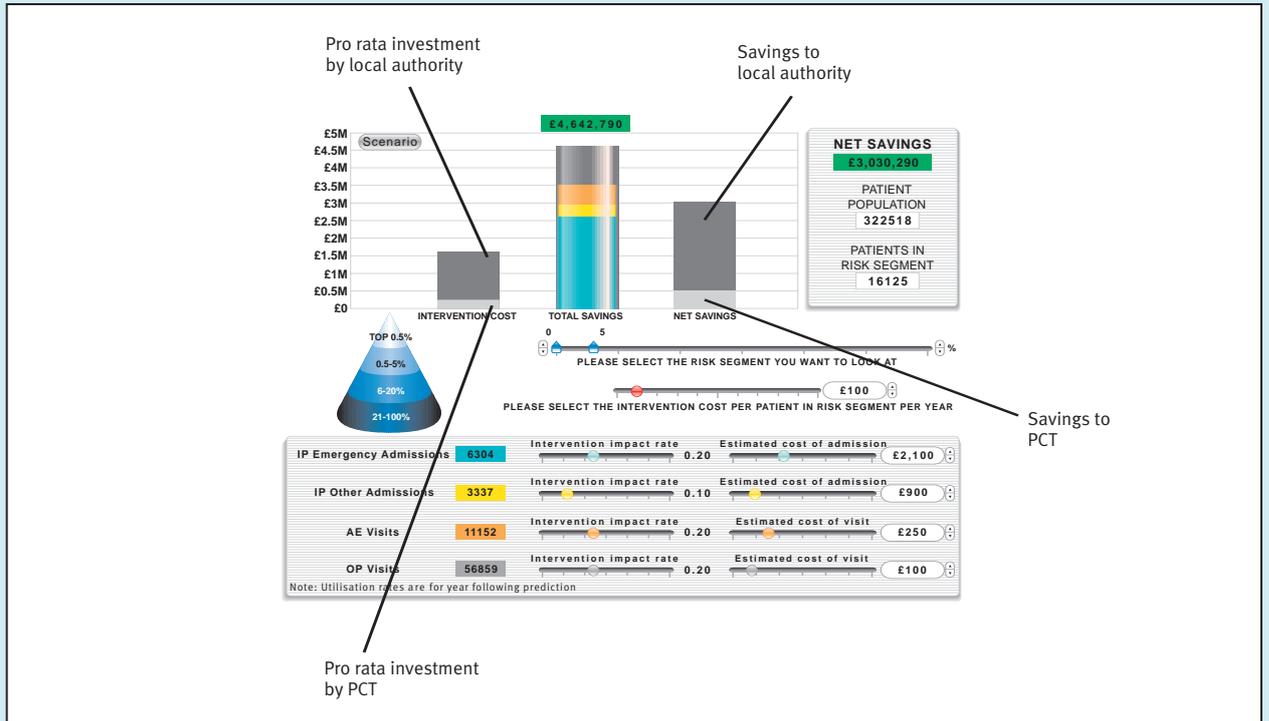
Coloured bands show where savings are made (the same colour scheme is used as in the panel below)

Bar represents the whole population, ranked according to risk

The intervention cost is set here (it can be varied so that a lower cost is set for lower risk segments of the population)

This panel shows four adverse outcomes that the intervention aims to avoid. If model E were built, then as well as the four hospital outcomes shown, this panel would also show costly social care outcomes (for example, admissions to a nursing home, admission to a residential home, and packages of intensive home care)

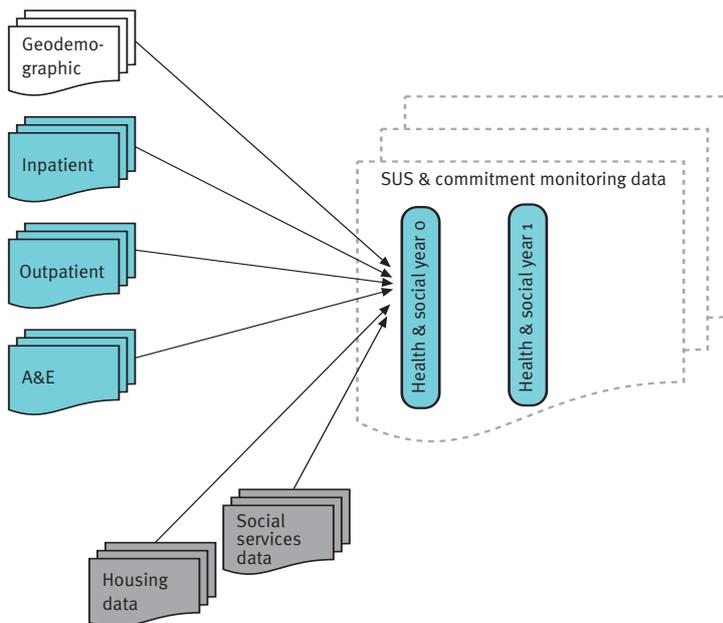
Source: Health Dialog UK Ltd



Source: Adapted from © Health Dialog UK Ltd

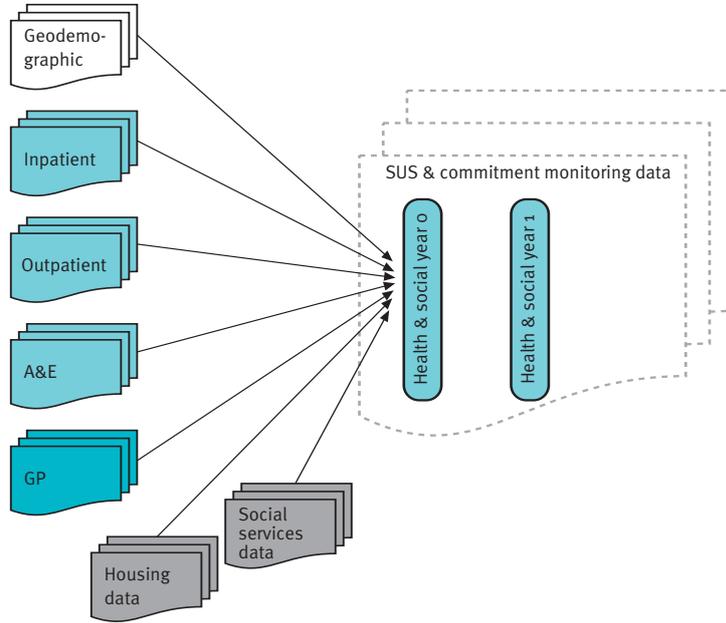
MODEL E1

This is the simpler of two models, as it uses local authority, SUS and Geodemographic data only.



MODEL E2

This is the more complex model, adding GP data. It could only be used in areas where GP data are extracted routinely.



Discussion

It is becoming clear that the costs of social care will rise dramatically in the United Kingdom over coming years as the population ages and more people live with complex, long-term medical conditions. This trend is already exerting pressure on local authorities to divert funds away from low-intensity care in favour of more intensive care. There is evidence that interventions for people at emerging risk can delay or prevent admission to a care home. Such interventions themselves are costly, and therefore in order for sustainable business cases to be built, accurate predictions are needed of which people would, without intervention, go on to require intensive social care.

A great deal is known about factors that are predictive of future need for intensive social care, of functional decline and admission to a care home. Many tools have been built that use these factors to make forecasts of future social care needs. However, the face-to-face nature of these instruments mean that they are laborious to complete, and cannot be administered to entire populations. Moreover, those people who are at highest risk may be least likely to participate (so-called participation bias).

An alternative approach might be to use patterns in routine, computerised, administrative data in order to make predictions. This approach is already being used by the NHS to predict unplanned hospital admissions. The technique raises important ethical and legal issues, and there are many practical and technological barriers that would need to be overcome. However, the technique offers a number of important benefits.

We suggest a number of potential models that could conceivably be constructed.

- Model A predicts admission direct from hospital to a nursing home. This model is attractive because it is (a) straightforward to predict as outcome is recorded in SUS data, and (b) particularly egregious therefore useful to predict in advance.
- Model B predicts nursing home admissions, and would therefore be the model of choice if the aim was solely to predict loss of independence. However, many packages of home care are more intensive (and therefore costly) than nursing homes and so may not predict those with the highest social care needs.
- Model C predicts an increase in social care needs according to the FACS classification (Fair Access to Care Services). Whilst this is appealing theoretically, there is a widespread perception that FACS criteria are applied unevenly.
- Model D predicts a rise in social care costs, and would use cost as a proxy of intensity of need. As discussed above, intensive packages of home care often exceed the cost of a care home admission, so local authorities are interested in predicting those people

who will become users of intensive services (and not merely those predicted to require long-term admission to a care home).

- Model E goes one step further by including health costs also. This model would enable a local health and social care economy to identify which people in its area were at risk of becoming high-cost users the subsequent year. The primary care trust (PCT) and local authority could then construct business cases for offering preventive support to these vulnerable people. The potential savings to be made by the PCT and the local authority respectively could be calculated so that the each organisation would contribute a pro rata amount to the investment. This model would help counter the perceived tendency to shift cost from one area to the other, and would encourage a more integrated approach.

It will be crucial to ensure that all interventions offered on the basis of any of these interventions were firmly evidence-based. It is also important to recognise that a shift from reactive to preventive services will involve an element of ‘double running’ in the short term.⁸⁵

Next steps

Following publication of this feasibility study, the King's Fund will work to publicise the findings amongst relevant stakeholders, some of which are listed below.

- Communities and Local Government (including Supporting People, National Housing Strategy for an Ageing Society, Digital Inclusion Unit and Neighbourhood Renewal Unit).
- Department of Health (including the Care Services Improvement Partnership, Social Care Directorate, Partnerships for Older People Projects and Public Health).
- Department for Work and Pensions (including Ageing Society Division and Link-Age Plus).
- Cabinet Office (Transformational government, Social Exclusion Task Force, Ministerial Committee on Data Sharing (MISC31) and the Prime Minister's Strategy Unit).

If funding were to be made available to continue with this work then, as a first step, it is proposed that one or two pilot sites that extract GP data routinely be identified. Two such sites are Croydon (Croydon Council and Croydon PCT) and Warwickshire (Warwickshire County Council and NHS Warwickshire). Between them, these sites offer a suitable range of deprivation and urban/rural mixture. Both sites have coterminous health and social services, both collect GP data routinely, and both sites have experience of providing anonymous and pseudonymous data for modelling purposes. Although, potentially, the Qresearch database⁷³ might be an alternative source of GP data for building a predictive tool, the participating practices are distributed widely across the country. This means that a large number of local authorities would have to be approached if linked social services data were to be obtained for these patients.

Work would then need to be conducted with the pilot site(s) to establish the data sources available to build the necessary model or models; to identify the accessibility of individual-level data; and to assess the feasibility of data linkage given confidentiality constraints. Research proposal forms would need to be submitted to each of the relevant bodies in order to gain permission to access their encrypted data.

For health and social services data to be linked, data-sharing software (such as those produced by companies such as Geometric⁸⁶, Visionware⁸⁷ and Infoshare⁸⁸) would need to be evaluated in order to assess how accurately and completely they were able to link health and social services on the basis of name, address and date of birth. Access to the best-performing software would need to be purchased.

An important lesson to learn from the previous King's Fund project on predicting unplanned hospital admissions is the importance of early 'branding' of the predictive tools that are built. The name 'PARR' (Patients at Risk of Readmission) has a strong brand identity, but there is confusion between the different versions of PARR (PARR 1; PARR 2; PARR+; PARR++). The name 'Combined Model' is less memorable, and it is often confused with the versions of PARR.

Another lesson from PARR is that it is not enough just to develop a tool that is freely available to local authorities and make it widely available (for example, downloadable from a website). The model will also need to be updated regularly. It is estimated that the cost of developing a model and testing it would be approximately £300,000, and that the cost of making the tool usable by local staff (for example, using appropriate software, providing help and support, and updating the model to reflect changes in social patterns) would be approximately £100,000 per annum.

If such tools could be developed and were tested successfully in the field, then the subsequent step would be to use them to target a range of appropriate, evidence-based interventions. This stage could again involve two or three pilot sites. This stage would be a development and research project on a much larger scale – similar to the Department of Health’s Whole System Demonstrator (WSD) pilots.⁸⁹ The development arm of the WSD study is costing £12 million, with the cost of the evaluation likely to be £1–2 million.

Appendix 1

The following databases were searched for studies examining what factors are predictive of the need for intensive social care: PubMed, Social Care Online, Social SciSearch, Scopus and the ASSIA databases. Using the Medical Subject Headings (MeSH) below, studies were identified that addressed the risk factors for functional decline, institutionalisation (admission to a residential home or nursing home), admission to a nursing home, and need for home care. Related studies were then sought.

MESH HEADINGS

- Algorithms
- Forecasting
- Risk factors
- Mass screening
- Risk assessment
- Frail elderly/statistics and numerical data
- Geriatric assessment/methods
- Logistic models
- Nursing homes/utilisation
- Patient discharge
- Severity of illness index
- Geriatric assessment
- Skilled nursing facilities
- Nursing homes
- Long-term care
- Home care services
- Residential facilities
- Patient admission
- Activities of daily living

Some studies were restricted to specific subgroups of the population (for example, disabled people, people with urinary incontinence, or people living in public housing). Other studies tested specific hypotheses (for example, whether nutritional status, depression or unsteadiness was predictive of nursing home residence).

References

- ¹ www.dh.gov.uk/en/PolicyAndGuidance/HealthAndSocialCareTopics/LongTermConditions/DH_4140328
- ² Wanless D, Forder J, Fernández JL, Poole T, Beesley L, Henwood M and Moscone F (2006). *Wanless Social Care Review, Securing Good Care for Older People, Taking a long-term view*. Report to King's Fund. London: King's Fund.
- ³ Joseph Rowntree Foundation (2006). *Foundations: Paying for long-term care*. York: Joseph Rowntree Foundation.
- ⁴ Local Government Association (2006). *Without a care*. London: Local Government Association.
- ⁵ Local Government Association (2006). *Meeting the Challenges Ahead*. London: Local Government Association.
- ⁶ Stuck AE, Egger M, Hammer A, Minder CE, Beck JC (2002). 'Home visits to prevent nursing home admission and functional decline in elderly people: systematic review and meta-regression analysis'. *Journal of the American Medical Association*, vol 287, pp 1022–8.
- ⁷ Brodaty H, Gresham M, Luscombe G (1997). 'The Prince Henry Hospital dementia caregivers' training programme', *International Journal Geriatric Psychiatry*, Dementia study, vol 12, pp 183–92.
- ⁸ Devon Social Services (1999). *The Exebank Handbook*. Exmouth, Devon: Exebank Rehabilitation Unit.
- ⁹ Herbert G and Townsend S (1999). *Rehabilitation Pathways for Older People*. Yorkshire: Nuffield Institute for Health, University of Leeds and YHEC, University of York.
- ¹⁰ www.kingsfund.org.uk/current_projects/predictive_risk/combined.html
- ¹¹ Goodlin S, Boulton C, Bubolz T and Chiang L (2004). 'Who will need long-term care?: Creation and validation of an instrument that identifies older people at risk'. *Disease Management*, vol 7, no 4, pp 267–74.
- ¹² Covinsky H E, Hilton J, Lindquist K and Dudley RA (2006). 'Development and validation of an index to predict activity of daily living dependence in community-dwelling elders'. *Medical Care*, vol 44, no 2, pp 149–57.
- ¹³ Fries BE, James M (2003). 'Identifying 'appropriate' applicants for home and community based services: the MI Choice screening system'. *Policy Brief* (Center for Home Care Policy Research), no 13, pp 1–6.
- ¹⁴ Alessi, CA, Josephson KR, Harker JO, Pietruszka FM, Hoyl MT and Rubenstein LZ (2003). 'The yield, reliability, and validity of a postal survey for screening community-dwelling older people'. *Journal of the American Geriatrics Society*, vol 51, no 2, pp 194–202.
- ¹⁵ Coleman EA, Wagner EH, Grothaus LC, Hecht J, Savarino J and Buchner DM (1998). 'Predicting hospitalization and functional decline in older health plan enrollees: Are administrative data as accurate as self-report?' *Journal of the American Geriatrics Society*, vol 46, no 4, pp 419–25.
- ¹⁶ Curry N, Billings J, Darin B, Dixon J, Williams M and Wennberg D (2005). *Predictive Risk Project Literature Review*. London: King's Fund.

- ¹⁷ Croydon PCT Case Study. King's Fund website. Available at: www.kingsfund.org.uk/document.rm?id=6746 (accessed 3 July 2007).
- ¹⁸ Tudor Hart J (1971). 'The inverse care law'. *Lancet*, vol 1, pp 405–12.
- ¹⁹ Victora C (2004). 'Achieving universal coverage with health interventions'. *Lancet*, 2004, vol 364, pp 1541–48.
- ²⁰ Beauchamp, TL, & Childress, JF (2002). *Principles of Biomedical Ethics*. 5th ed. Oxford: OUP.
- ²¹ Data Sharing – Frequently Asked Questions. Website of the Department for Constitutional Affairs. Available at: www.foi.gov.uk/sharing/faqs.htm#part6 (accessed 29 May 2007).
- ²² Data Protection Act 1998. Available at: www.opsi.gov.uk/acts/acts1998/19980029.htm (accessed 25 May 2007).
- ²³ Patient Information Advisory Group (2006). *The use of patient information in the Long Term Conditions programme*. Department of Health Advisory Bodies. Available at: www.advisorybodies.doh.gov.uk/piag/piag-ltc-nov2006.pdf (accessed 29 May 2007)
- ²⁴ Flacker JM, Kiely DK (2003). 'Mortality-related factors and 1-year survival in nursing home residents'. *Journal of the American Geriatrics Society*, vol 51, no 2, pp 213–21.
- ²⁵ Sager MA, Rudberg MA, Jalauddin M, Franke T, Inouye SK, Ladefeld CS, Siebens H, Winograd CH (1996). 'Hospital admission risk profile (HARP): identifying older patients at risk for functional decline following acute medical illness and hospitalization'. *Journal of the American Geriatrics Society*, vol 44, no 3, pp 251–7.
- ²⁶ Covinsky KE, Palmer RM, Fortinsky RH, Counsell SR, Stewart AL, Kresveic D, Burant CJ, Landefeld CS (2003). 'Loss of Independence in activities of daily living in older adults hospitalized with medical illnesses: increased vulnerability with age'. *Journal of the American Geriatrics Society*, vol 51, no 4, pp 451–8.
- ²⁷ Covinsky KE, Martin GE, Beyth RJ, Justice AC, Sehgal AR, Landefeld CS (1999). 'The relationship between clinical assessments of nutritional status and adverse outcomes in older hospitalized medical patients'. *Journal of the American Geriatrics Society*, vol 47, no 5, pp 532–8.
- ²⁸ Covinsky, KE, Fortinsky RH, Palmer RM, Kresveic DM, Landefeld CS (1997). 'Relation between symptoms of depression and health status outcomes in acutely ill hospitalized older persons'. *Annals of Internal Medicine*, vol 126, no 6, pp 417–25.
- ²⁹ Cornette P, Swine C, Malhomme B, Gillet JB, Meert P and D'Hoore W (2006). 'Early evaluation of the risk of functional decline following hospitalization of older patients: development of a predictive tool'. *The European Journal of Public Health*, vol 16, no 2, pp 203–8.
- ³⁰ Hansen K, Mahoney J, Palta M (1999). 'Risk factors for lack of recovery of ADL independence after hospital discharge'. *Journal of the American Geriatrics Society*, vol 47, no 3, pp 360–5.
- ³¹ Lindeberger EC, Landefeld CS, Sands LP, Counsell SR, Fortinsky RH, Palmer RM, Kresvic DM, Covinsky KE, (2003). 'Unsteadiness reported by older hospitalized patients predicts functional decline'. *Journal of the American Geriatrics Society*, vol 51, no 5, pp 621–6.
- ³² Covinsky KE, Hilton J, Lindquist K, Dudley RA (2006). 'Development and validation of an index to predict activity of daily living dependence in community-dwelling elders'. *Medical Care*, vol 44, no 2, pp 149–57.
- ³³ Welz-Barth A, Garcia-Schurman C, Fusgen I (1998). 'Incontinence, dementia and multiple morbidity – predictive factors for nursing care requirement and nursing home admission'. *Wiener medizinische Wochenschrift*, vol 148, no 13, pp 305–8.
- ³⁴ Klein T, Salaske I (1994). 'Determinants of nursing home admission of elderly patients and chances for prevention'. A longitudinal study in Germany. *Zeitschrift für Gerontologie und Geriatrie*, vol 27, no 6, pp 442–55.

- ³⁵ Jette AM, Branch LG (1983). 'Targeting community services to high-risk elders: toward preventing long-term care institutionalization'. *Prevention in Human Services*, vol 3, no 1, pp 53–69.
- ³⁶ Oura A, Washino M, Wada J, Arai Y, Mori M (2006). 'Factors related to insititutionalization among the frail elderly with home-visiting service in Japan'. *Gerontology*, vol 52, no 1, pp 66–8.
- ³⁷ Matsuu K, Washio M, Arai Y, Mori M, Ide S (2002). 'Factors related to admission or institution among the frail elderly with a home-visit nursing service'. *Japanese Journal of Public Health*, vol 49, no 10, pp 1107–16.
- ³⁸ Fujiita T, Mori H, Osaki Y, Minowa M, Ishihara S, Masuda N, Hanzawa Y, Mori C, Nanba T, Ohta H *et al* (1992). 'A case-control study on factors relating to discontinuation of domiciliary care for the bedridden elderly in a metropolitan area'. *Japanese Journal of Public Health*, vol 39, no 9, pp 687–95.
- ³⁹ Hancock R, Arthur A, Jagger C, Matthews R (2002). 'The effect of older people's economic resources on care home entry under the United Kingdom's long-term care financing system'. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol 57, no 5, S 2, pp 85–93.
- ⁴⁰ Shuy YI, Lee HC (2002). 'Predictors of nursing home placement and home nursing services utilization by elderly patients after hospital discharge in Taiwan'. *Journal of Advanced Nursing*, vol 38, no 4, pp 398–406.
- ⁴¹ Slade A, Fear J, Tennant A (2006). 'Identifying patients at risk of nursing home admission: The Leeds Elderly Assessment Dependency Screening tool (LEADS)'. *BMC Health Services Research*, vol 13, no 6, S 31.
- ⁴² Ahmed A, Allman RM, DeLong JF (2003). 'Predictors of nursing home admission for older adults hospitalized with heart failure'. *Archives of Gerontology and Geriatrics*, vol 36, no 2, pp 117–26.
- ⁴³ Rudberg MA, Sager MA, Zhang J (1996). Risk factors for nursing home use after hospitalization for medical illness. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol 1, no 5, M189–94.
- ⁴⁴ Fogel JF, Hyman RB, Rock B, Wolf-Klein G (2000). 'Predictors of hospital length of stay and nursing home placement in an elderly medical population'. *Journal of the American Medical Directors Association*, vol 1, no 5, pp 202–10.
- ⁴⁵ Aditya BS, Sharma JC, Allen SC, Vassallo M (2003). 'Predictors of a nursing home placement from a non-acute geriatric hospital'. *Clinical Rehabilitation*, vol 17, no 1, pp 108–13.
- ⁴⁶ Lord SR (1994). 'Predictors of nursing home placement and mortality of residents in intermediate care'. *Age and Ageing*, vol 23, no 6, pp 499–504.
- ⁴⁷ Boaz RF, Muller CF (1994). 'Predicting the risk of "permanent" nursing home residence: the role of community help as indicated by family helpers and prior living arrangements'. *BMC Health Services Research*, vol 29, no 4, pp 391–414.
- ⁴⁸ Greene VL, Ondrich JI (1990). 'Risk factors for nursing home admissions and eits: a discrete-time hazard function approach'. *Journal of Gerontology*, vol 45, no 6, S2, pp 50–8.
- ⁴⁹ Coward RT, Horne C, Peek CW (1995). 'Predicting nursing home admissions among incontinent older adults: a comparison of residential differences across six years'. *Gerontologist*, vol 35, no 6, pp 732–43.
- ⁵⁰ Finlyson M (2002). 'Changes predicting long-term care use among the oldest-old'. *Gerontologist*, vol 42, no 4, pp 443–53.
- ⁵¹ Pan SM, Yang JT, Chen CC (1998). 'The predictors of long-term care service utilization among older Americans. *The Kaohsiung Journal of Medical Sciences*, vol 14, no 4, pp 226–33.
- ⁵² Nygaard HA, Albreksten G (1992). 'Risk factors for admission to a nursing home. A study of elderly people receiving home nursing'. *Scandinavian Journal of Primary Health Care*, vol 10, no 2, pp 128–33.

- ⁵³ Watanabe M, Kono K, Kawamura K, Matsuura T (1998). 'Factors influencing admission to nursing homes among frail and bedridden elderly persons receiving public health and welfare services'. *Japanese Journal of Public Health*, vol 45, no 1, pp 45–55.
- ⁵⁴ Bauer EJ (1996). 'Transitions from home to nursing home in a capitated long-term care program: the role of individual support systems'. *Health Services Research*, vol 31, no 3, pp 309–26.
- ⁵⁵ Friedman SM, Steinwachs DM, Rathouz PZ, Burton LC, Mukamel DB (2005). 'Characteristics predicting nursing home admission in the program of all-inclusive care for elderly people'. *Gerontologist*, vol 45, no 2, pp 157–66.
- ⁵⁶ Black BS, Rabins PV, German PS (1999). 'Predictors of nursing home placement among elderly public housing residents'. *Gerontologist*, vol 39, no 5, pp 559–68.
- ⁵⁷ Hanley RJ, Alexih LM, Wiener JM, Kennell DL (1990). 'Predicting elderly nursing home admissions'. Results from the 1982-1984 National Long Term Care Survey. *Research on Ageing*, vol 12, no 2, pp 199–228.
- ⁵⁸ Liu K, Coughlin T, McBride T (1991). 'Predicting nursing-home admission and length of stay'. A duration analysis. *Medical Care*, vol 29, no 2, pp 125–41.
- ⁵⁹ Stern Y, Tang MX, Albert MS, Brandt J, Jaobs DM, Bell K, Marder K, Sano M, Devanand D, Albert SM, Bylsma F, Tsai WY (1997). 'Predicting time to nursing home care and death in individuals with Alzheimer's disease'. *Journal of the American Medical Association*, vol 277, no 10, pp 806–12.
- ⁶⁰ Coughlin TA, McBride TD, Liu K (1990). 'Determinants of transitory and permanent nursing home admissions'. *Medical Care*, vol 28, no 7, pp 616–31.
- ⁶¹ Valiyeva E, Russell LB, Miller JE, Safford MM (2006). 'Lifestyle-related factors and risk of future nursing home admission'. *Archives of Internal Medicine*, vol 166, no 9, pp 985–90.
- ⁶² Gaugler JE, Duval S, Anderson KA and Kane RL (2007). 'Predicting nursing home admissions in the US: a meta-analysis'. *BMC Geriatrics* 2007, vol 7, no 13.
- ⁶³ Pan SM, Yang JT, Chen CC (1998). 'The predictors of long-term care service utilization among older Americans'. *The Kaohsiung Journal of Medical Sciences*, vol 14, no 4, pp 226–33.
- ⁶⁴ Rabiner DJ (1995). 'Patterns and predictors of non-institutional health care utilization by older adults in rural and urban America. *Journal of Rural Health*, vol 11, no 4, pp 259–73.
- ⁶⁵ Dove HG, Duncan I and Robb A (2003). 'A prediction model for targeting low-cost, high-risk members of managed care organisations'. *The American Journal of Managed Care*, vol 9, no 5, pp 381–9.
- ⁶⁶ LoBianco RT, Mills ME and Moore HW (1996). 'A model for case-management of high-cost Medicaid users'. *Nursing Economics*, vol 14, pp 303–307, 314.
- ⁶⁷ Meenan RT, O'Keefe-Rosetti, Hornbrook MC, Bachman DJ, Goodman MJ, Fishman PA, Hurtado AV (1999). 'The sensitivity and specificity of forecasting high-cost users of medical care'. *Medical Care*, vol 37, pp 815–823.
- ⁶⁸ Goodlin S, Boulton C, Bubolz T, Chiang L (2004). 'Who will need long term care? Creation and validation of an instrument that identifies older people at risk'. *Disease Management*, vol 7, no 4, pp 267–73.
- ⁶⁹ www.communities.gov.uk/index.asp?id=1128440
- ⁷⁰ www.experianbs.com/Content.asp?ArticleID=566
- ⁷¹ www.connectingforhealth.nhs.uk/systemsandservices/sus/
- ⁷² www.connectingforhealth.nhs.uk/systemsandservices/ssd/products_and_services/vaprodmiquest

- ⁷³ www.qresearch.org/default.aspx
- ⁷⁴ Audit commission (1997). *The Coming of Age: Improving care services for older people*. London: Audit Commission.
- ⁷⁵ Bebbington A, Darton R and Netten A (2001). *Care Homes for Older People*, Vol 2. Admissions, needs and outcomes. Canterbury: Personal Social Services Research Unit, University of Kent.
- ⁷⁶ Department of Health (2001). *National Service Framework for Older People*. London: Department of Health.
- ⁷⁷ Secondary Uses Service. Connecting for Health website. Available at: www.connectingforhealth.nhs.uk/systemsandservices/sus/ (accessed 14 June 2007).
- ⁷⁸ www.kingsfund.org.uk/current_projects/predictive_risk/patients_at_risk.html
- ⁷⁹ Personal communication from David Osborne, Croydon PCT using 2005/06 NHS-Wide Clearing System (NWCS) data for Croydon.
- ⁸⁰ Boaz RF, Muller CF (1994). 'Predicting the risk of "permanent" nursing home residence: the role of community help as indicated by family helpers and prior living arrangements'. *BMC Health Services Research*, vol 29, no 4, pp 391–414.
- ⁸¹ Netten A, Forder J and Shaprio J (2006). *Measuring Personal Social Services Outputs for National Accounts: Services for Older People*. Discussion Paper 2267/3. January 2006. Canterbury: Personal Social Services Research Unit, University of Kent.
- ⁸² Curtis L and Netten A (2006). *Unit Costs of Health & Social Care 2006*. Canterbury: Personal Social Services Research Unit, University of Kent.
- ⁸³ Cestari L (MA), Munroe M (MSc), Evans S (PhD), Smith A (BSc), Huxley P (PhD) (2006). 'Fair Access to Care Services (FACS): implementation in the mental health context of the UK'. *Health and Social Care in the Community*, vol 14, no 6, pp 474–81.
- ⁸⁴ Centre for Policy on Aging (2005). *What Price Care in Old Age?* Available at: www.cpa.org.uk/cpa/cpa_what_price_care.pdf (accessed 19 June 2007).
- ⁸⁵ Local Government Association (2005). *Briefing on Independence, Well-Being and Choice*. London: Local Government Association.
- ⁸⁶ www.geometricltd.com
- ⁸⁷ www.visionwareplc.com/page.cfm?pageid=171
- ⁸⁸ www.infoshare-is.com
- ⁸⁹ White Paper pilots: whole system long term conditions (LTC) demonstrator programme. Department of Health Website. Available at: www.dh.gov.uk/en/PolicyAndGuidance/HealthAndSocialCareTopics/LongTermConditions/DH_4140328 (accessed 11 June 2007).

