

Predicting social care costs: a feasibility study

The costs of caring for people with complex social care and health care needs are set to rise in the UK over the coming years. As more people live with long-term medical conditions, it will become increasingly important to find ways of helping local councils and health services to take earlier action to support people to remain independent and stay in their own homes.

This briefing describes a study that explored whether statistical models can be used to predict an individual person's future need for intensive social care.¹ Aside from the predictive models we developed, this work generated important lessons about the potential of linked health and social care data to support policy analysis and to guide the planning and commissioning of services.

Key points

- Although health and social care services interact in many ways for millions of people, their information systems tend to be discrete and distinct. This research has shown how it is possible to link routine data from health and social care information systems in a way that protects individuals' identities.
- Within health care, predictive modelling is increasingly used as a strategy to identify people at high risk of future unplanned hospital admission, and so target preventive care. Such approaches have not previously been tested with respect to social care. Predictive models have the potential to provide a better experience for service users and to offer more cost-effective care.
- This project has shown that it is possible to construct predictive models for social care. The next stage will be to see how these models might fit into everyday working practice.
- The predictive accuracy of our models is comparable to some of the models used by the NHS to predict hospital admissions. We suggest that it will be important to pilot and evaluate the use of these tools in practice, across a range of sites.
- Linked person-level information has the potential to improve the quality of care services – whether through improved identification of high-risk individuals, comparative performance measures, service evaluations or budget-setting. At a time when individual budgets and personalisation are seen as important, the need to collate and analyse information of this type seems ever more pressing.
- The quality of data about individual health care use has improved considerably over the past decade. Now a step change is needed to ensure that information about social care services improves in the same way. This will require strategies to improve the coding, collection and sharing of data in ways that protect confidential information.

Background

The costs associated with complex social care and health care needs are set to rise in the UK in the medium term. This is largely due to two linked phenomena: an ageing population and the increasing number of people who will be living with a chronic disease. Moreover, both health and social care costs are highly skewed across the population, with a small number of individuals accounting for most of the expenditure. It would therefore be useful to be able to identify those people most at risk of beginning high-cost care so that they might be offered intensive 'upstream' preventive care and support. In theory, this could potentially promote independent living and, at the same time, yield significant net savings 'downstream'. One way to identify people who are at high risk is with predictive risk models.

Over the last few years, many NHS organisations have begun using predictive models that identify with reasonable accuracy which individual people within a population are at risk of unplanned hospital admission in the forthcoming year. In England⁸ the Department of Health has funded the development of two such models, called 'PARR' (Patients At Risk of Re-hospitalisation) and the 'Combined Model'.^{2,3} The models use pseudonymous

administrative data to make predictions at the individual level. The patient's GP is able to decode an individual patient's risk score so that high-risk patients can be offered an 'upstream' intervention, such as support from a community matron, aimed at mitigating their risk of future hospitalisation.

This study attempted to build a predictive risk model to forecast the future use of social care. Admission to hospital and admission to a care home are analogous in several ways. Both events are typically unwelcome to the person concerned, costly to society, and sometimes avoidable. There is strong evidence that certain interventions designed to maintain people's independence, such as home visits by a multidisciplinary team, can successfully prevent or delay care home admissions.⁴ However, such interventions are typically very costly, so if local councils are to invest more efficiently in preventive interventions, they will need accurate and objective ways of identifying the future risk of each individual in their population.⁵ This study looked at the feasibility of constructing such a model, considering first how to access and link the required data, before testing a number of different statistical models. The work was funded by the Care Services Efficiency Delivery programme (CSED) at the Department of Health.

Box 1: Predictive risk models

Predictive risk models apply statistical techniques such as multiple regression or neural networks to analyse routine electronic data.⁶ These models use historic patterns in the population's data to make predictions at the individual level. The growing use of predictive models in health care over recent years has been made possible by a combination of better access to individual-level electronic data and improvements in computing power. Datasets for large populations, often involving hundreds of millions of observations, can now be analysed according to individual health needs, service use and health outcomes.

The accuracy of a predictive model can be quantified using various measures, based on its performance on a validation sample of data. In this report, we have concentrated on two of these measures:

- **Sensitivity**, which is a measure of how good the model is at detecting people from the population who will experience the specified outcome (such as admission to a care home).

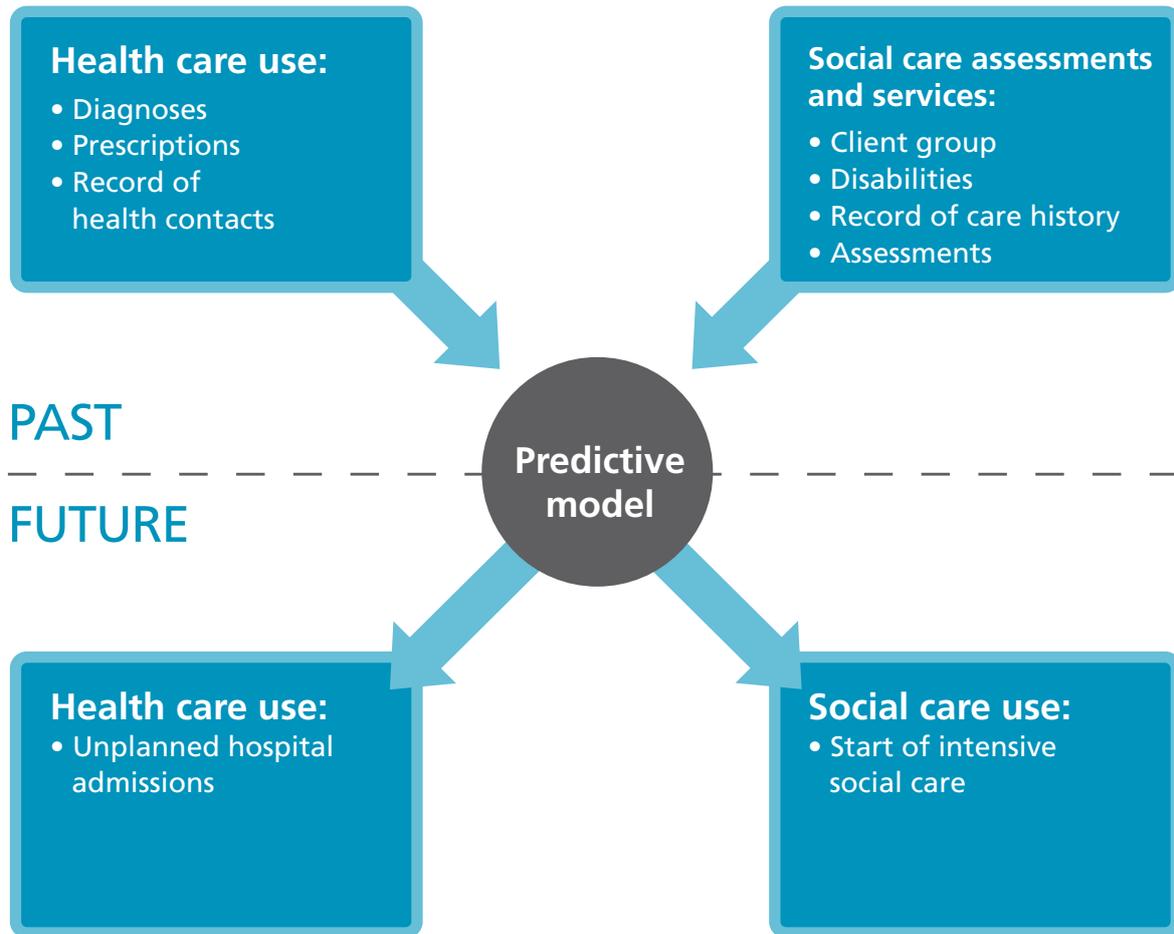
- **Positive predictive value (PPV)**, which measures the reliability of the predictions made by the model: that is, the chance that the people that the model says are at high risk of experiencing the outcome of interest will indeed experience that outcome.

When a predictive model is used in practice, it is applied to the most recent data in order to produce a risk estimate for each individual in the population. Since the uncertainty surrounding the model's predictions is known from its performance on the validation sample, this can help commissioners to build robust business cases for early intervention.

We know from the literature that the predictor variables for care home admission may include, for example, age, sex, ethnicity, deprivation, morbidity, health service use and drugs prescribed, as well as patterns of social care needs and usage.⁵ Since these variables span health and social care records, a complicating yet critical prerequisite for this project was to link health and social care data at the individual level in a way that did not compromise confidentiality.

⁸ The PRISM (Predictive Risk Stratification Model) is used in Wales, while SPARRA (Scottish Patients at Risk of Readmission and Admission) is used in Scotland.

Figure 1: Using health and social care data to predict health and social care use



Predictive models for health care, such as PARR, use information about past health care use (top left box) to predict future health care use (bottom left box). To predict social care costs, we combined information on both health and social care use and assessments (top two boxes) to predict future social care use (bottom right box).

Extracting and linking data

Predictive models are built on the routine electronic data of very large numbers of people. In order to protect the confidentiality of this information, only pseudonymous data were used (see Box 2). We obtained routine data from five sites in England (four primary care trusts and their local authorities, and one care trust). We extracted individual-level data spanning several years that described the individual health and social care needs of the people living in these areas, and their use of health and social care services.

We linked the pseudonymous health and social care data to a pseudonymous ‘member’ file. This file listed all the people registered with a general practitioner in each site and it acted as an underlying population register for the project. Having a member file enabled us to assess the success of our data linkage attempts and it allowed us to estimate events relative to the denominator populations.

We obtained hospital data from all five sites; this covered inpatient episodes, outpatient visits and A&E attendances. We also obtained social care data from all five sites. However, whereas hospital data are recorded in a standard format nationwide, the detailed social care

data we obtained for this project varied from site to site according to the local IT systems used. In order to make comparisons across sites, we were obliged to group the social care data under broad headings, including social care assessments, social care needs and social care

provision or funding. Our combined health and social care datasets were large and complex. Across the five sites we received pseudonymous data on 578,217 people aged 55 and above (see Table 1).

Table 1: Total numbers of records accessed across the five sites

| | | |
|--------------------|--|-------------|
| Inpatient | Number of records (episodes) | 3,268,439 |
| | Number of spells | 2,831,790 |
| | Number of unique patients | 999,425 |
| Outpatient | Number of records (appointments) | 12,814,644 |
| | Number of actual attendances | 9,093,461 |
| | Number of unique patients | 1,532,360 |
| A&E | Number of records (visits) | 2,126,847 |
| | Number of unique patients | 924,697 |
| GP | Number of registered people aged 55 and over | 578,217 |
| | Number of GP events (2 sites only) | 110,972,879 |
| Social care | Number of events | 445,800 |
| | Number of assessments | 349,700 |
| | Number of unique users | 163,800 |

Box 2: Protecting confidential information

This project required the analysis of very large amounts of routine health and social care data, all linked at the person level. The size of the datasets, which included hundreds of millions of records, meant that it was not feasible to seek individual consent from people to use their data for the project.

Informed consent is the fundamental principle governing the use of person-identifiable information by any part of the NHS or the research community.⁷ While there is an exception to this rule in certain circumstances for essential NHS activities,⁷ we instead chose to use only pseudonymous data for this project. Pseudonymous data have had all person-identifiable fields (such as name, address and date of birth) removed and the unique key (NHS number) replaced by a pseudonym. The Patient Information Advisory

Group (PIAG) at the Department of Health had previously ruled that confidentiality requirements may be met by using pseudonymous data that cannot be identified without a password.

We obtained written confirmation from PIAG that the pseudonymisation process and use of encrypted media that we planned to employ in this project would meet its requirements, and therefore no application was necessary under Section 60 of the Health and Social Care Act 2001, subsequently replaced by section 251 of the NHS Act 2006. (This Section of the Act enables the Secretary of State to require or permit patient information to be shared for medical purposes where this is considered to be in the interests of improving patient care or in the public interest.) The responsibilities of PIAG have since been assumed by the Ethics and Confidentiality Committee of the National Information Governance Board (NIGB).⁸

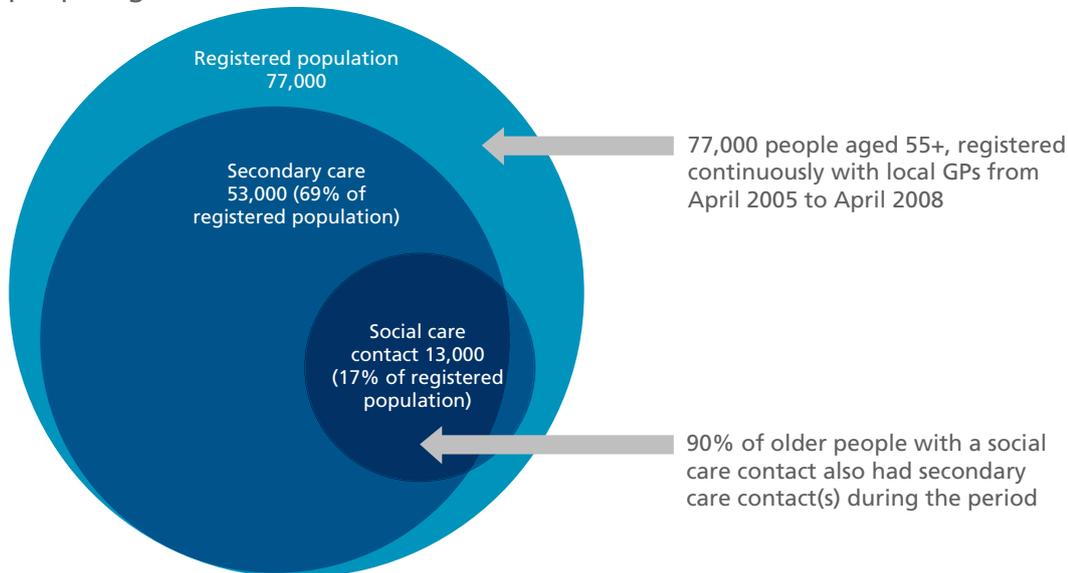
Overlap between health and social care

Although it is widely recognised that health care and social care are strongly interrelated, the availability of linked health and social care data is limited. This project demonstrated that it is possible to extract individual information from health and social care data systems, and to create linked files for large populations. Moreover, by using pseudonymous data we were able to construct individual health and social histories without compromising privacy or confidentiality.

We found some consistent patterns across the five sites.

About 80 per cent of social care users aged 75 and above accessed a hospital service in one year, and around half of these people had a hospital inpatient stay. The frequency of emergency hospital admissions was around three times higher among social care users than among those with no record of social care in that year. When considered over longer time periods, for example over three years, the proportion of social care users who accessed hospitals was higher still. Figure 2 demonstrates the considerable overlap between the use of health and social care services by people aged 55 and over in one site.

Figure 2: Overlap in the use of health and social care in one site over a three-year period for people aged 55+



Model results and sensitivity analyses

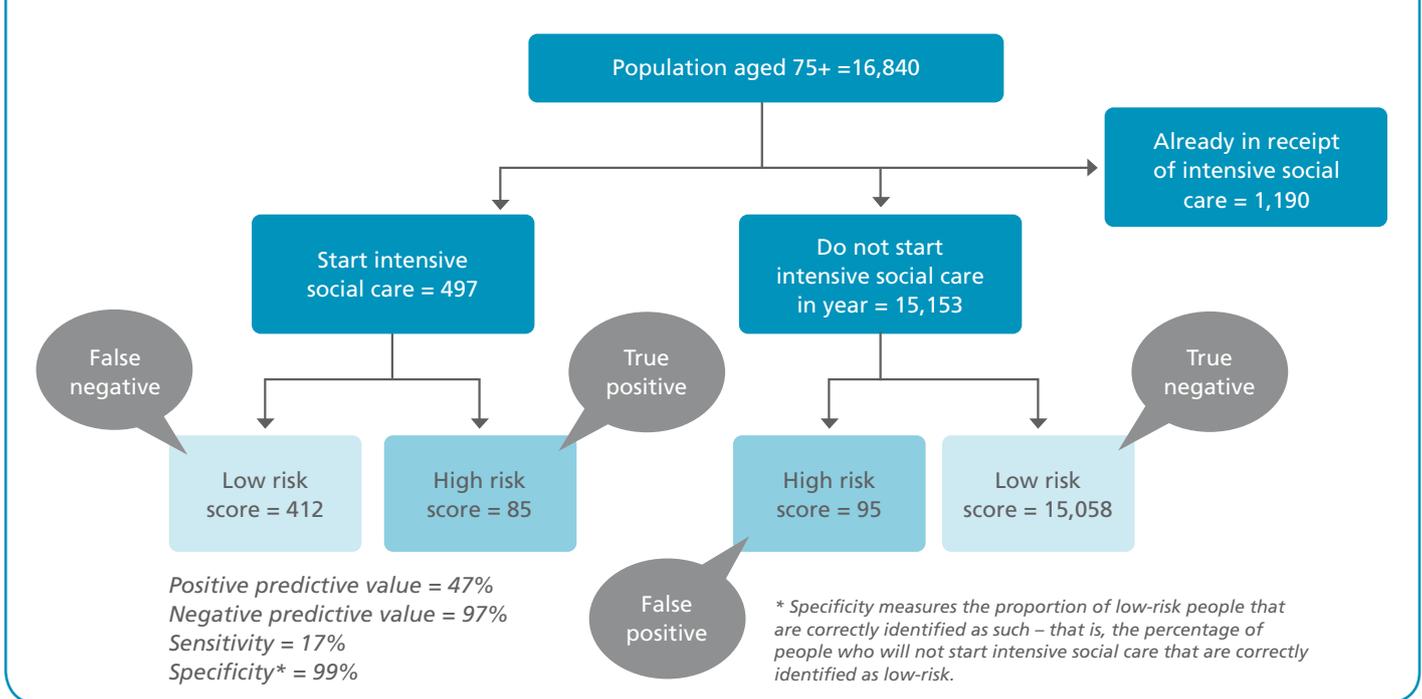
Over the course of this study we tested many different predictive risk models for social care. Our initial models examined whether prior health and social care information could be used to predict the start of 'intensive social care' funded by the council. This was defined as a move into a care home, the start of ten or more hours of home care per week, or an increase in annualised social care costs of over £5,000 per year. In four out of the five sites we were able to construct stable models to predict the start of intensive social care. The accuracy of these models, as measured by their positive predictive value (PPV), was satisfactory if not remarkable, with PPVs ranging from 35 to 50 per cent. However, the models were relatively insensitive: that is, they only detected a small proportion of the people across the population who did start intensive social care. In practical terms, this means that these models

would identify a relatively small number of high-risk individuals (low sensitivity) but these individuals were truly at a high risk of beginning intensive social care (adequate PPV); see Figure 3.*

We believe that one of the reasons for the rather modest accuracy of these models is that the outcome we were trying to predict was relatively rare. It was certainly far rarer than emergency hospital admission, which is the outcome predicted by PARR and the Combined Model. Consequently, when we loosened the definition of intensive social care, we found that the performance of the models improved. For example, when we included an increase in annualised social care costs of above £3,000, or of above £1,000, in the definition of intensive social care, the models generated considerably more accurate predictions.

* Note, however, that PPV and sensitivity can be traded off against each other.

Figure 3: Predictive value, sensitivity and specificity of the model incorporating a £5,000 threshold



As expected, higher age was significantly and consistently associated with higher risk scores. We also found that information regarding prior social care use and social care needs contained important predictor variables. Models built from social care data alone performed roughly as well as those that contained health and social care data. Nevertheless, certain health variables were significantly predictive of future social care costs.

All of the models performed best when calibrated to local datasets (in other words, when the variables were weighted specifically in ways that fitted local data). However, we were able to construct models based on pooled data (that is, data from more than one site) and we achieved results that were almost as good.

We tested a wide range of variants of the models but none of them produced dramatic improvements. The variants we tried included:

- use of the Index of Multiple Deprivation (IMD) as a predictor variable
- constructing models based only on residents living in 'deprived' areas
- addition of clinical GP data (in two sites)
- use of community health care data (district nursing, community physiotherapy etc.)
- classifying people according to Adjusted Clinical Groups (ACG™) and related groupings.

Other uses of linked health and social care information

This project required us to collate and analyse linked routine health and social care datasets from nine different organisations. Pseudonymous data linkage at an individual level has rarely been attempted before on this scale. We believe that the individual descriptive information contained in this collation of data has considerable value, quite apart from the predictive models themselves.

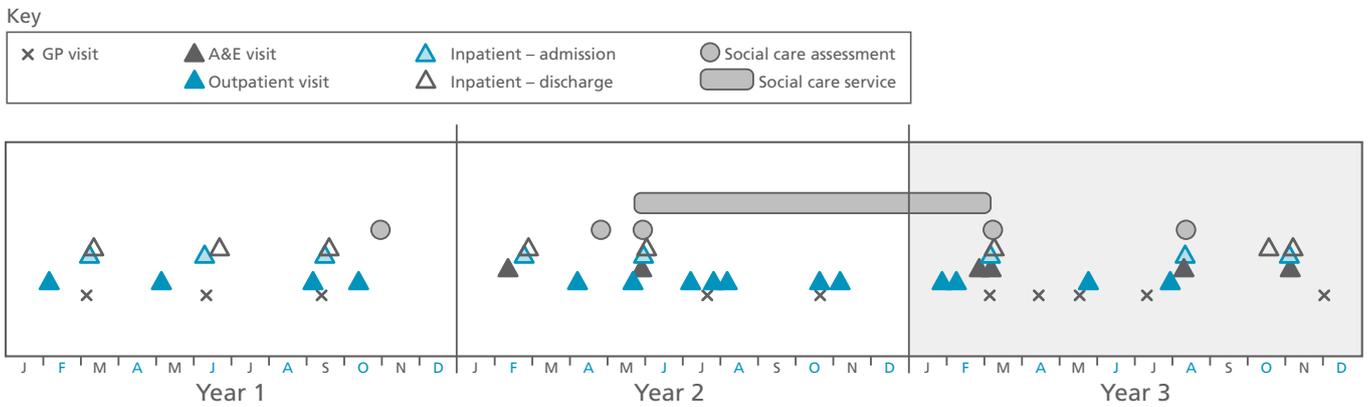
Some of the other potential applications of pseudonymous, linked health and social care data include:

- the development of shared information profiles for use by teams of professionals or clinicians, service users and their carers (see Figure 4)
- analyses of gaps in care (differences between the care received and the optimal care as defined by evidence-based guidelines)
- planning tools that assess the financial impact of preventive health and social care interventions
- tools for the evaluation, performance management and regulation of health and social care
- improved approaches to performance monitoring and benchmarking across pathways of care.

Figure 4 shows all contacts that one individual person had with all health and social care services over a three-year period. Presenting this information graphically risks allowing individual identities to be deduced, so we have introduced a degree of random error in order to

protect the identity of the individual. These charts are an extremely powerful way to convey large amounts of information in meaningful ways. In practice, they should only be constructed on demand and with the service user's explicit consent, and destroyed when the user logs off.

Figure 4: Individual health and social care event timeline over a three-year period



The chart shows one individual's health and social care history over a period of three years (the data have been part-randomised to protect anonymity). In the first year there were four outpatient attendances and three hospital admissions, as well as some GP visits. A social care assessment was carried out towards the end of the year but this did not result in any services being provided. With more hospital contacts in the following year, two social care assessments took place and a low-intensity package of home care was instigated. In the early part of the third year an unplanned hospital admission occurred, the provision of home care ceased and a social care assessment was conducted. During the remainder of the year, the person had another two unplanned hospital admissions and a further social care assessment was conducted.

Key issues raised by the research

By linking pseudonymous health and social care data at the individual level, we have been able to conduct a range of novel and sophisticated analyses. We believe that the findings from such analyses will be potentially valuable to service users, professionals and commissioners alike. This is a rich seam of information whose potential has not previously been fully exploited.

We suggest the following specific recommendations to help further develop this promising agenda:

1. The predictive accuracy of our models is comparable to some of the models used by the NHS to predict hospital admissions. However, the ways that our models might be used in practice are less clear. We suggest that it will be important to pilot and evaluate the use of these tools in practice in a range of sites.
2. Linking health and social care data raises a number of concerns relating to information governance. If the huge potential of linked data is to be exploited, clearer guidance will be required over what is and is not permissible. We suggest that clear protocols should be agreed by the National Information Governance Board, the Information Commissioner's Office and the NHS Information Centre, and that these protocols should be widely disseminated.
3. We believe that it will be helpful to establish an experimental dataset that includes linked pseudonymous health and social care data from several sites. This dataset will be useful for evaluating a range of initiatives. For example, it could be used to test the effects of 're-ablement' (that is, an intervention designed to help people regain their independence after a spell in hospital) according to the risk profiles of users; for setting personal budgets; and for promoting the integration and coordination of care.
4. Social care data systems would benefit from greater consistency in how they record and code information. Some work on this theme is underway at the NHS Information Centre and at the Department of Health. A notable example is the development of the Tools for Rapid Integration of Public Submissions (TRIPS) system.⁹ We believe that there is a need for a national group to agree a common coding approach for social care data. This is a prerequisite for making detailed, meaningful comparisons across areas. In turn, we expect that this will lead to improvements in coding practices and to higher-quality social care data, especially if a secondary uses service for social care were to be established analogous to the existing

Hospital Episode Statistics (HES) and the Secondary Uses Service (SUS) for health care.

Finally, we believe that data linkage should be promoted not only for predictive modelling purposes, but more widely for the commissioning of integrated health and social care. The benefits of linking health and social care data include:

- Information for professional and clinical staff – potentially displaying graphically an entire history of an individual's health and social care contacts. This can provide an extremely rich picture of a person's interaction with services.
- The development of tools for planning and commissioning care services, with the potential to deliver benefits to both health and social care services.
- Evaluation, benchmarking and performance management to improve the nature of comparative information and to test for potential trade-offs between the two care sectors. The Nuffield Trust is currently undertaking some formal evaluations using these techniques.

This research, which was originally published in the journal *Age and Ageing*, is reported here by permission of the British Geriatrics Society:

Bardsley M, Billings J, Dixon J, Georghiou T, Lewis GH and Steventon A (2011) 'Predicting who will use intensive social care: case finding tools based on linked health and social care data', *Age and Ageing* 40(02).

References

1. Bardsley M, Billings J, Chassin L, Dixon J, Eastmure E, Georghiou T, Lewis G and Steventon A (2011) *Predicting Social Care Costs: A feasibility study*. London: Nuffield Trust.
2. Billings J, Dixon J, Mijanovich T and Wennberg D (2006) 'Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients', *BMJ* 333(7563), 327.
3. Wennberg D, Siegel M, Darin B, Filipova N, Russell R, Kenney L, Steinort K, Park T, Cakmakci G, Dixon J, Curry N and Billings J (2006) *Combined Predictive Model: final report and technical documentation*. Available from www.kingsfund.org.uk/document.rm?id=6745.
4. Stuck AE, Egger M, Hammer A, Minder CE and Beck JC (2002) 'Home visits to prevent nursing home admission and functional decline in elderly people: systematic review and meta-regression analysis', *Journal of the American Medical Association* 287, 1022–8.
5. Lewis, GH (2007) *Predicting Who Will Need Costly Care: How best to target preventive health, housing and social programmes*. London: The King's Fund.
6. Cousins MS, Shickle LM and Bander JA (2002) 'An introduction to predictive modeling for disease management risk stratification', *Disease Management* 5(3), 157–67.
7. www.dh.gov.uk/ab/Archive/PIAG/index.htm
8. www.nigh.nhs.uk/ecc
9. www.cipfa.org.uk/socialcare/download/SCP_summer2010.pdf, p.8

About the Nuffield Trust

The Nuffield Trust is a charitable trust carrying out research and health policy analysis on health services. Our focus is on the reform of health services to improve the efficiency, effectiveness, equity and responsiveness of care.

Key current work themes include:

- new forms of care provision
- commissioning
- efficiency
- national and international comparisons
- competition policy.

This briefing forms part of the Trust's work on the commissioning of health care. The full report *Predicting Social Care Costs: A feasibility study* is available at: www.nuffieldtrust.org.uk/publications

The report authors are Martin Bardsley, John Billings, Ludovic Jean Chassin, Jennifer Dixon, Elizabeth Eastmure, Theo Georghiou, Geraint Lewis and Adam Steventon.

For more information on our work, and to receive updates via our regular e-newsletter, visit www.nuffieldtrust.org.uk



The Nuffield Trust
FOR RESEARCH AND POLICY
STUDIES IN HEALTH SERVICES

The Nuffield Trust
59 New Cavendish Street
London W1G 7LP
Tel: 020 7631 8450
Email: info@nuffieldtrust.org.uk
Fax: 020 7631 8451

© Nuffield Trust 2011

Not to be reproduced
without prior written
permission.