Sources, uses, strengths and limitations of data collected in primary care in England

Azeem Majeed, Professor of Primary Care
Imperial College London

INTRODUCTION

In this article, which is based on a report prepared for the Department of Health during 2001/2002, I discuss some of the sources, uses, strengths and limitations of the data collected in primary care in England. The report for the Department of Health was compiled through a literature review and by interviews with key informants in the Department of Health, the NHS, and the academic sector. The preliminary findings were then discussed at a workshop organised by the Department of Health in October 2001 and a final report then prepared. This article contains a summary of the report and its findings.

In the UK, general practitioners are responsible for both providing primary care services and for organising referrals for specialist care.[1] Consequently, the medical records held by general practitioners include details of all diagnoses and prescribed drugs, in many cases extending from birth. In an increasing number of general practices, these records are held in electronic format and are potentially available for extraction and analysis. Electronic data from general practice therefore offer unique opportunities to plan and monitor health services, measure the quality of care provided by the National Health Service (NHS), and undertake population-based research (Box 1). There are also a number of other data sets derived from primary care that can be used to support these tasks (Box 2 and Annex Table 1).

The strengths of the data collected from general practice are that they are population-based and not derived from an unrepresentative subset of the population; most contacts with NHS take place in primary care; and they contain information on morbidity, treatment, outcomes and health care utilisation. Furthermore, an increasing number of general practices are computerised and are converting from paper to computerised medical

records. Now that the NHS is proposing to invest heavily in information technology,[2] careful thought needs to be given as to how to improve the quality of electronic health records in general practice and derive maximum benefit from the data they contain.

One current limitation of electronic data from general practice is that some of the most heavily used data sets come from primary care databases derived from volunteer practices, such as those in QRESEARCH, the General Practice Research Database or the IMS Mediplus system. These practices are often larger and better organised than the average practice and hence their data may not give a true picture of the performance of general practice as a whole. The quality and completeness of electronic health records can also vary, and information on socio-economic and ethnic status is generally absent from the data that general practitioners collect.[3]

USES OF GENERAL PRACTICE DATA

Many organisations and individuals make use of data from primary care. The most important use of general practice data will always be in the clinical management of patients by members of the primary health

---

Box one

Main uses of primary care data

- to improve patient care;
- needs assessment;
- monitoring health care utilisation;
- health services planning and disease projections;
- resource allocation;
- clinical governance;
- performance management and measuring quality;
- prescribing policy and pricing;
- monitoring health inequalities;
- health services research;
- epidemiology;
- pharmaco-epidemiology; and
- case-mix measurement and risk-adjustment.

Box two

Sources of primary care data

- Data from computerised general practice records.
- Data derived from projects such as PRIMIS.
- Primary care databases, such as the General Practice Research Database and the IMS Mediplus database.
- Prescribing Analysis and Cost (PACT) data.
- Royal College of General Practitioners' Weekly Returns' Service.
- Data derived from local disease registers.
- Organisations that run morbidity projects; for example the Trent Focus.
- Data derived from individual general practice audits and needs assessments.
- Routine NHS activity data; for example, General Medical Services (GMS) data.
- National Database of Primary Care Groups and Trusts.
- Fourth National Survey of Morbidity in General Practice.

See Annex Table 1 for further details about these sources.

---

care team. The availability of high-quality data about individual patients to the clinician at the point of care ensures that the clinician is fully aware of the patient's medical history and can plan the patient's clinical management appropriately. When combined with computerised prompts and decision support software, the data can be a valuable resource for improving patient care.[4]

However, the data have many uses beyond the clinical management of the individual patient. For these uses, institutions such as the Department of Health, strategic health authorities, and primary care trusts will be the main users of the data. In the near future, the new Commission for Health Audit and Inspection is also likely to become an important user of the data. Inevitably, users will focus on different aspects of the data, depending on their particular needs.

Needs assessment and public health

Data derived from primary care can complement that from other sources for describing the burden of disease in the population. Traditionally, the burden of disease has been measured using mortality data. However, many chronic diseases that contribute significantly to population morbidity are not commonly recorded on death certificates. This has meant that the burden of disease caused by many chronic disorders has not been well measured. Moreover, where information on morbidity is available, it has often been limited in scope and has come from relatively small studies. The samples of patients examined in such studies may not be representative of the population, leading to uncertainty about how accurate estimates of prevalence based on such studies may be. Furthermore, the small number of people studied means that such studies cannot be used to calculate accurate prevalence rates by age group and sex. Finally, single studies can provide information for only one point in time and so cannot provide information on how prevalence rates are changing over time. Using data from primary care can help provide solutions to these problems.[5]

Prevalence data derived from primary care can also be used with population projections to estimate the future number of people suffering from a number of chronic diseases. Such projections show that even if age-specific disease prevalence rates remain constant, the number of people suffering from disorders such as heart failure and atrial fibrillation will increase substantially in future years, because of the increase in the number of people in the population from elderly age groups. This in turn will have major implications for areas such stroke services and anticoagulation clinics. By being aware of such impending changes now, it is possible to plan to ameliorate the effects they have on health services.[6]

Monitoring health care utilisation

Most contacts with the NHS take place in primary care. Furthermore, many of the targets in national service frameworks are for treatments provided in primary care.[7] These include treatments such as the use of ACE-inhibitors in people with heart failure, statins in people with ischaemic heart disease, and warfarin in people with atrial fibrillation.[8] Hence, using primary care data to monitor the use of primary care health services will be increasingly important for the NHS. Although these examples are process measures, they are all interventions that lead to better outcomes for patients and hence can be used as markers of the quality of care provided in general practice. The other main source of prescribing data (PACT, Prescribing Analysis and Cost) provides very limited information on prescribing for specific conditions. Hence, for the time being, data collected directly from primary care remains the main method of monitoring prescribing for specific conditions. However, PACT data remain the main tool for monitoring the cost and volume of community prescribing in the NHS.[9]

General practitioners' workload also needs to be monitored. This is important for both general practitioners and the Department of Health. Information on the range of conditions presenting to general practitioners was provided by the Fourth Survey of Morbidity in General Practice but this survey was carried out in 1991–1992 and its results are becoming increasingly out of date.[10] The Royal College of General Practitioners' Weekly Returns Service provides more up to date information on consultation patterns, but the analyses possible with this data source are more limited than with the morbidity survey.

Clinical governance

Many practices use data from their computerised medical records for clinical governance. In some primary care groups, using data in this way has been found to be a powerful incentive for improving the completeness or accuracy of data recording, and for improving the quality of health care. Some clinical systems can now use clinical data to interact with clinical guidelines and decision support software. In the future, such links are likely to become more widespread. This may further help the process of clinical governance and improve the quality of care provided by primary health care teams. In the longer-term, the NHS Information Strategy will encourage high quality data capture in all general practices.

Performance measurement

The monitoring of general practices by primary care trusts has largely been based on information from sources such as PACT and General Medical Services (GMS) data. The 'primary care indicators' produced from such sources are relatively crude and usually provide only proxy measures of performance.[11] In the future, other data collected in from primary care will be used increasingly to monitor the quality of care provided by general practices and primary care trusts. For example, the new general practitioner contract contains many new quality measures, which many practices are likely to derive from their computerised medical records. Some of the results from these analyses are likely to be published, allowing the public access to information on the quality of care provided by their local general practices.

The NHS in England spends about £7 billion per year on community prescribing, more than is spent on all the other aspects of general practice combined. Hence, ensuring that the NHS obtains good value from its community prescribing spending is a key task for both the Department of Health and primary care trusts. PACT data are already used extensively to monitor the implementation of prescribing policies and help ensure cost-effective prescribing in primary care.[12] Other data from primary care will also be used to support work in these areas.

Monitoring health inequalities

Reducing inequalities in health status and the use of health services is one of the government's main priorities for the health service.[13] As an increasing proportion of NHS services will be delivered in primary care, it will be important to monitor socio-economic and ethnic inequalities in the use of primary health care services, to see if these are narrowing over time in line with government targets. Primary health care can also provide information on morbidity that can be used to monitor socio-economic variations in health status, using practice postcodes as a proxy measure of deprivation. Analyses such as these usually show that chronic disease prevalence rates are highest and treatment rates lowest in deprived areas. For example, in a recent study, the prevalence of coronary heart disease was highest in deprived areas and lowest in affluent areas, whereas the use of statins showed the opposite trend.[14]

Health services research

Primary care data are extensively used for health services research.[15] This research includes studies of disease prevalence and treatment, time trends, gender, area and socio-economic variations, and effectiveness studies. Academics who have made substantial use of primary care data in their research identify areas such as greater accessibility of data, opportunities to link different data sets, and the development of indicators of socio-economic status as key areas for the development of primary care data. Another issue they raise is the need for higher quality data and measures of data quality. An increasing amount of research using primary care data has also been commissioned by policy makers to help provide information on public health and NHS priorities. As part of the move towards evidence-based policy, policy-makers and academics are likely to make much greater use of primary care data in the future, particularly if the range of data and its accessibility both increase.

One key requirement will be to develop methodologies for using data from primary care to carry out rapid evaluations of health policy and clinical effectiveness. These types of studies are common in the USA, but carried out less frequently in the UK. Because changes in health policy can be started and discontinued very quickly, methodologies are needed that can evaluate their impact relatively quickly, within the political time frames of central government.

Methodologies are also needed for carrying studies of clinical effectiveness in observational settings.[16,17] The 'gold standard' for such studies is the randomised controlled trial. However, these trials are carried out on carefully selected subsets of patients who may not be typical of patients in the general population. Furthermore, randomised controlled trials usually have more rigorous monitoring and follow-up of patients than might be seen in routine clinical practice. Hence, both policy-makers and clinicians need to know if the benefits and complications of treatment found in trials are also present in routine clinical practice. Using data from primary care may provide one solution to this problem.

Pharmaco-epidemiology

Primary care data are a valuable resource for studies examining the side-effects and complications of drugs.[18,19] Examples of such research include the studies showing the increased risk of thrombo-embolic disease in users of 'third-generation' contraceptive pills. However, these studies also show the problems that can occur in such research. For example, one study reached different conclusions to the other studies in this area, showing how differences in the way conditions are defined and in the methods used can lead to studies reaching very different conclusions.[20] Hence, observational studies about the safety and effectiveness of drugs need to be carried out with care and by researchers familiar with the pitfalls of such research. In particular, researchers need to be aware of the problems of bias and confounding, which can result in researchers drawing erroneous conclusions from studies using data from primary care databases.

The importance of pharmaco-epidemiology is in the implications it has for public health and patient safety. Data from primary care can provide long-term follow up of new (and established) drugs and help answer questions about the safety of drugs. Furthermore, because these types of studies use data that is already collected, they can be carried out, the results disseminated, and action taken to protect public health, long before any findings would be available from prospective studies. Hence, they provide an important mechanism for protecting the health of the public, both in the UK and elsewhere.

Case-mix measurement and risk-adjustment

Health care systems around the world are all trying to ensure that resources are used appropriately and, as part of this process, they are trying to reduce unacceptable variations between providers. For example, the new NHS plan states that it 'promises better performance and accountability systems to reduce variations in service across England'. In both primary and secondary care, there are large differences in the way in which doctors practise medicine. However, without taking into account differences in case-mix, we do not know whether such differences are justified clinically. Risk adjustment methods, initially developed in the USA, can help correct such variations for underlying differences in population case-mix and thus could lead to performance measures for providers that are fairer and more accurate than the unadjusted measures available now.[5]

IMPROVING THE QUALITY OF DATA

Although the data held by general practitioners on their clinical computer systems could be of great use to clinicians, managers, academics, and policy-makers, a number of issues remain to be resolved before their value can be maximised fully. These issues in this area were identified through the interviews with key informants from the NHS, the Department of Health, and the academic community; and from workshops for people from these groups described earlier in the paper. Some of the issues that arose from these interviews and workshops are listed in Box 3.

Several studies have shown that the quality of data in primary care databases, such as the General Practice Research Database, is good.[21,22] Consequently, they are widely used for health services and epidemiological research.[23] However, these databases are derived from information supplied by volunteer general practices and the quality of data in more representative samples of general practices is much more variable. For example, in a recent study carried out in 47 general practices in one primary care trust in south-west London, there was a

very wide variation in the proportion of cases of ischaemic heart disease that were coded on the practices' computers (Figure 1). No practice had 100 per cent of cases coded (although some achieved coding levels of over 90 per cent) and in some practices, no cases were coded.[24] As this study was population-based, and not based on a sample of volunteer general practices, it illustrates the size of the task facing the NHS if the quality of electronic medical records in primary care is to be improved significantly.

Because of the current variability in the completeness and accuracy of coding, many primary care trusts will find it difficult to examine chronic disease prevalence and treatment rates in their locality. In contrast, it is now possible to get good information on chronic disease management at a national level, using information from primary care databases; for example on atrial fibrillation (Figure 2).[25] The Department of Health has set up a number of initiatives in an attempt to overcome some of the current problems with data quality in primary care (Box 4). These include PRIMIS (Primary Care Information Services) and HIP (Health Information Project). In addition, there are a number of independently-funded data quality projects, such as the Primary Care Data Quality Program. If these initiatives are successful, then by the end of this decade, the NHS will have access to high quality information on morbidity and the use of health services on almost the entire population. However, the size of the task required to achieve this objective should not be under-estimated. Success requires not just accurate and complete entry of data onto general practices' computers, but also consistent and complete Read coding of this information to allow easy extraction and analysis of data.
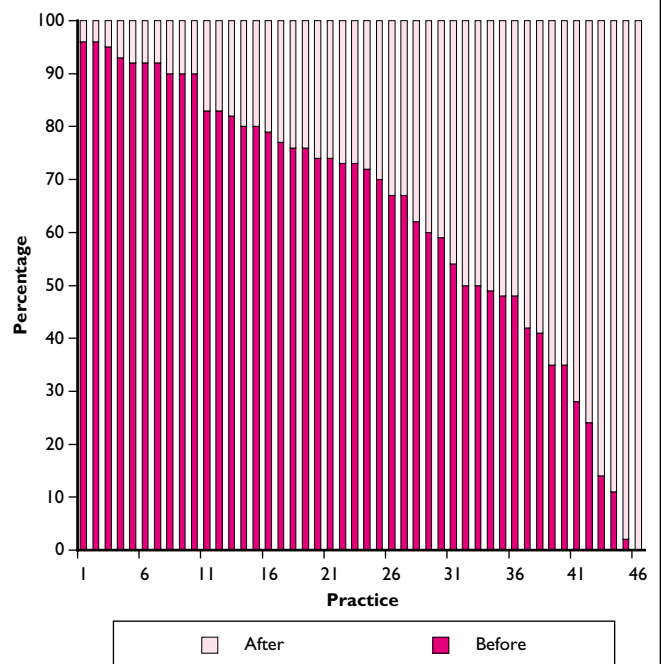
One problem that general practitioners are faced with when trying to improve the quality of their computerised data is that, at present, there are no standards for the recording of morbidity data in primary care. This in turn leads to a wide variation in the coding of clinical information, even among general practices that are fully computerised. One proposal raised in the interviews and workshops to help overcome this problem was the establishment of a 'data accreditation scheme'. Under such a

---

Box three

Key issues about primary care data identified from interviews with informants and from workshops

- How can the quality of primary care data be improved?
- Should there be an accreditation scheme for practices that wish to supply morbidity data for disease registers and for health care planning to ensure that the data they supply reaches specified standards for accuracy and completeness?
- How can awareness and use of primary care data be increased among policymakers, health service planners and commissioners, and health professionals?
- How can a balance be struck between the need for both local and national data?
- How can the NHS make use of the local disease registers that are currently being developed and implemented in many primary care trusts?
- Should the addition of socio-economic and ethnicity variables be seen as a priority for the development of primary care data? If so, what is the best way to achieve this objective?
- How much data derived from primary care should be made available publicly to encourage its wider use, improve accountability, and provide information on the performance of primary care trusts and general practices?

Figure 1 **Percentage of coronary heart disease patients correctly identified on practice computers in 47 medical practices after searching paper medical records in 2001**

Box four

Initiatives aimed at improving the quality of primary care data

PRIMIS (Primary Care Information Services)
http://www.primis.nhs.uk/Default.htm
PRIMIS is a no-charge support service designed to help primary care organisations improve patient care through the effective use of their clinical computer systems.

HIP for CHD (Health Information Project for Coronary Heart Disease)
http://www.hipforchd.org.uk
HIP for CHD is Department of Health funded initiative to develop methods that will help general practices and primary care trusts measure their progress in meeting the standards set out in the National Service Framework for Coronary Heart Disease.

MIQUEST
http://www.nhsia.nhs.uk/nhsia/pages/products/vaprod/miquest/
MIQUEST is a computer program designed to extract data held within practice clinical systems for analysis.

Primary Care Data Quality
http://www.pcdq.org
PCDQ is an educational intervention that aims to build upon existing skills and knowledge for recording data within primary care and to maximise the use of technology to monitor and assess improvements in the quality of data, resulting in better care for patients.

Figure 2   **Percentage of men with atrial fibrillation prescribed oral anticoagulants in 1994 and 1998**
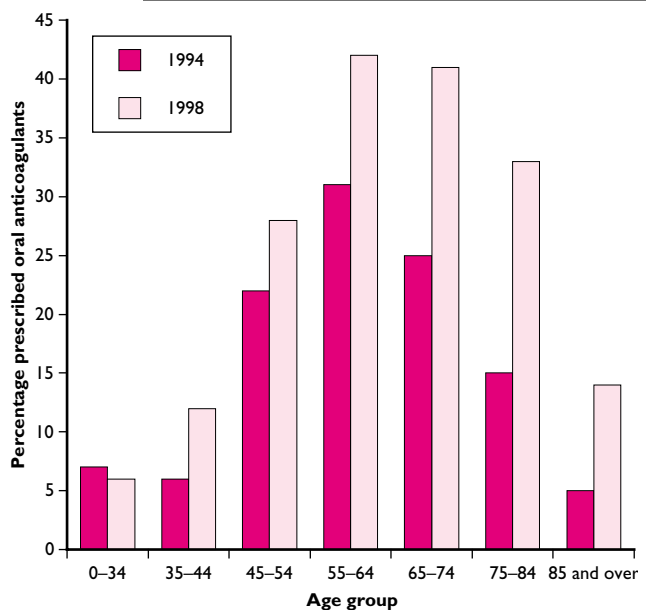
scheme, standards would be set for the completeness and accuracy of data recording in general practice. The scheme could also help ensure that diagnoses were recorded using similar Read codes, which would facilitate comparisons of general practices and primary care trusts, as well as allowing aggregation of data to produce national estimates. Some guidelines already exist for practices contributing data to primary care databases and the Weekly Returns Service, and these could form the basis of a new accreditation scheme.

Finally, a new general practitioner contract will be implemented by the Department of Health in April 2004. The contract contains a quality framework with clinical quality measures from 10 clinical domains. It is likely that it will act as a powerful driver to improve the accuracy of computerised medical records, but this will have to be confirmed in future research studies.

INCREASED AWARENESS OF PRIMARY CARE DATA

Another issue lies around raising awareness of the sources, uses, strengths and weaknesses of primary care data among potential users. This is particularly important for primary care trusts, as these organisations will need to use such data to help them plan and monitor services, measure progress towards national targets, and improve the quality of the services they deliver. Some progress has been made in this area with the publication of documents such as *Key Health Statistics from General Practice*, which contain national estimates of chronic disease prevalence and management,[26] and through educational initiatives such as PRIMIS.

As well as raising awareness of primary care data, the NHS also needs to develop expertise its analysis. This is a complex area and it is sometimes possible to reach opposing conclusions when research groups use different methods to analyse primary care data from the same source. This will require support from the Department of Health and research councils for units that can develop such expertise and disseminate it more widely, both in the NHS and among the academic community. This would allow much more rapid appraisal of changes in health policy and the configuration of health services, as well as promoting greater analysis of primary care data to support public health activities.[27] Such analyses are more common in the USA, partly driven by the more widespread availability of data from administrative databases and surveys, and help to support and evaluate new initiatives in primary care.[28]

PRODUCTION OF LOCAL AND NATIONAL STATISTICS

In the last few years, many general practices and primary care trusts have started to produce local disease registers.[29] They have often been produced to meet the requirements of national service frameworks, most commonly for coronary heart disease. However, using these registers to compare the management of chronic diseases among general practices and primary care trusts has been difficult. In many cases, the method used to produce such registers, including case definitions, was left to individual practices. This meant that it was often not possible to compare practices, even within the same primary care trust, because of the variability in their construction. Trying to compare different primary care trusts is often even more difficult using information from these local disease registers.

Tools such as MIQUEST (a computer program designed to extract information from general practitioners' computer systems in a standardised format) could facilitate trying to obtain data from practices in a standardised way. However, only around 85 per cent of clinical systems are currently compatible with MIQUEST. Hence, some work needs to be done by the NHS Information Authority to overcome some of the problems that limit wider take-up of MIQUEST and ensure that the remaining systems also become compatible with it. MIQUEST can also be difficult to use and in many localities there is lack of expertise in writing MIQUEST queries.

In contrast to primary care, NHS hospitals have a well-developed and standardised system (hospital episode statistics) for collecting information on their workload. The collection of a minimum data set in each hospital ensures that information can be aggregated to produce regional and national statistics, as well as statistics on each individual hospital. It has also allowed the production of more clinically focused performance indicators than for primary care. Now that the NHS is shifting to providing a greater proportion of its services in primary care and community settings, the performance indicators published by the Department of Health need to reflect this shift and start to exploit the rich data sources available in primary care.

More valid performance measurement

The Department of Health and the Commission for Health Audit & Inspection are committed to producing a range of indicators that can be used to measure the performance of primary care trusts in meeting their three key tasks: commissioning specialist health services, providing primary care and community health services, and improving health. The new performance framework will be used to measure how well primary care trusts are carrying out these key tasks. Furthermore, based on these performance measures, some primary care trusts will be rewarded for their performance (for example, by being given greater autonomy), whereas some other primary care trusts could face sanctions. Hence, it is important that any performance indicators that are developed for primary care trusts provide an accurate and valid assessment of the quality of care in key areas.[30]

A second issue arises around measuring the performance of individual general practices. For most patients, primary health care teams are their first point of contact with the NHS. Hence, the care they receive has a major impact on health services utilisation (for example, on prescribing, investigations, referrals, and hospital admissions) and on their views on the quality of care that the NHS provides. An overall performance measure for a primary care trust will mask the significant variations in performance that we know are present among general practices.[31] Hence, it is also important to develop performance measures for general practices. During the 1990s, many health authorities developed performance indicator packages in an attempt to try to measure the performance on general practices and also to give primary health care teams information they could find useful in improving the services they provided.[32] However, there were no national standards for these packages and hence there was a wide variation in their contents and the ways in which they were used. The increasing computerisation of general practices offers an opportunity to produce more clinically focused and meaningful performance measures, if some of the current problems about data quality and access to data can be overcome.

Collection of socio-economic and ethnic data

One of the key priorities for the NHS is reducing inequalities in access to health services and in health status. Most health services are delivered in primary care and primary health care teams have an important role in reducing health inequalities.[33] However, there is currently almost no monitoring of primary care services to determine if these are being used equitably by different socio-economic and ethnic groups. This is partly because socio-economic and ethnic data have not been routinely collected in primary care. Some pilot projects have now been set up to assess the feasibility of collecting data on ethnicity to allow limited ethnic monitoring. However, it remains unclear whether all general practices will have the ability or time to collect accurate socio-economic and ethnic data on their patients. For example, coding of information on ethnicity on hospital admission records in England remains very poor even though this has been mandatory since 1995. An alternative approach would be to provide support, training and funding to a sample of practices to allow them to collect these data accurately and for their entire population, as was done with the Fourth Survey of Morbidity in General Practice.[10]

Wider release of data

Assuming that the NHS is successful in improving the quality and completeness of data collected in primary care, it then needs to make these data available to a wider audience, including health professionals, managers, academics, and the public. The type and format of data people from these groups will want will vary. For example, academics may want person-based datasets that they can analyse for their own research projects. Clinicians and managers are more likely to want information that has already been analysed and presented in a user-friendly format that supports them in their day to day work. Members of the public will need information presented in a non-technical manner that allows them to learn more about the range of services offered and the quality of care in their local general practices, and how they compare with practices in other localities.[34]

Conclusions

The most important reason for developing high-quality computerised medical records in primary care is to improve the quality of patient care. However, the successful achievement of this objective offers the opportunity of providing information for many other important health service and public health activities. The NHS is proposing to invest considerable sums in information technology over the next few years. Developing high-quality clinical information systems in primary care and ensuring the data they produce is widely disseminated and used should be seen as a key priority for this NHS Information Technology strategy.

Key points

- The medical records held by general practitioners include details of all diagnoses and prescribed drugs, in many cases extending from birth.
- In an increasing number of general practices, these records are held in electronic format and are potentially available for extraction and analysis.
- Electronic data from general practice offer unique opportunities to plan and monitor health services, measure the quality of care provided by the NHS, and undertake population-based research.
- Although the data held by general practitioners on their clinical computer systems could be of great value, a number of issues remain to be resolved before their value can be maximised.
- One of the most important tasks to achieve will be to improve the quality of electronic medical records in primary care.
- A second key task will be to raise awareness of the sources, uses, strengths and weaknesses of primary care data among potential users.

CORRESPONDENCE TO:

Professor Azeem Majeed
Department of Primary Health Care & General Practice
Imperial College London
Charing Cross Campus
Reynold's Building
St Dunstan's Road
London W6 8RP
Email a.majeed@ic.ac.uk
Internet www.azmaj.org

REFERENCES

1. De Maeseneer J, Hjortdahl P and Starfield B (2000) Fix what's wrong, not what's right, with general practice in Britain. *BMJ* **320**, 1616–1617.
2. Delivering IT in the NHS. Department of Health: London, 2002. Available at http://www.doh.gov.uk/ipu/whatnew/deliveringit/index.htm.
3. Smeeth L and Heath I (1999) Tackling health inequalities in primary care. *BMJ* **318**, 1020–1021.
4. de Lusignan S, Thiru K, Wells S, Dobson M and Howitt A (2000) Overcoming the constraints to becoming paperless. *Br J Gen Pract* **50**, 504–505.
5. Majeed A, Bindman A B and Weiner J P (2001) Use of risk adjustment in budget setting and performance measurement in primary care. Paper I. How it works. *BMJ* **323**, 604–607.
6. Ellis C, Gnani S and Majeed A (2001) Prevalence and management of heart failure in general practice in England and Wales, 1994–1998. *Health Statistics Quarterly* **11**, 17–24.
7. Hippisley–Cox J and Pringle M (2001) General practice workload implications of the national service framework for coronary heart disease: cross sectional survey. *BMJ* **323**, 269–270.
8. Department of Health. National Service Framework for Coronary Heart Disease. London: Department of Health, 2000. Available at: http://www.doh.gov.uk/heart/index.htm.
9. Majeed A, Evans N and Head P (1997) What can PACT tell us about prescribing in general practice? *BMJ* **315**, 1515–1519.
10. McCormick A, Fleming D and Charlton J (1995) *Morbidity Statistics from General Practice.* Fourth National Study 1991–1992. OPCS Series MB5 No 3. HMSO: London.
11. Marshall MN, Hiscock J and Sibbald B (2002) Attitudes to the public release of comparative information on the quality of general practice care: qualitative study. *BMJ* **325**, 1278–1281.
12. Ashworth M, Golding S and Majeed A (2002) Prescribing indicators and their use by Primary Care Groups to influence prescribing. *Journal of Clinical Pharmacology & Therapeutics* **27**, 197–204.
13. HM Treasury and Department of Health (2002) *Tackling health inequalities*, Department of Health: London. Available at: http://www.doh.gov.uk/healthinequalities/ccsrsummaryreport.htm.
14. Ryan R and Majeed A (2001) Prevalence of ischaemic heart disease and its management with statins and aspirin in general practice in England and Wales, 1994–1998. *Health Statistics Quarterly* **12**, 34–39.
15. Hippisley–Cox J, Pringle M, Coupland C, Hammersley V and Wilson A (2001) Do single handed practices offer poorer care? Cross sectional survey of processes and outcomes. *BMJ* **323**, 320–323.
16. Pocock S J and Elbourne D R (2000) Randomized trials or observational tribulations? *N Engl J Med* **342**, 1907–1909.
17. McKee M, Britton A, Black N, McPherson K, Sanderson C and Bain C (1999) Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* **319**, 312–315.
18. Jick H, Garcia Rodriguez L A and Perez-Gutthann S (1998) Principles of epidemiological research on adverse and beneficial drug effects. *Lancet* **352**, 1767–1770.
19. Lawrenson R A, Williams T and Farmer R D (1999) Clinical information for research; the use of general practice databases. *Journal of Public Health Medicine* **21**, 299–304.
20. Skegg D C G (2002) Oral contraceptives, venous thromboembolism, and the courts. *BMJ* **325**, 504–505.
21. Hollowell J (1997) The General Practice Research Database: quality of morbidity data. *Population Trends* **87**, 36–40.
22. Lawrenson R, Todd J C, Leydon G M, Williams T J and Farmer R D (2000) Validation of the diagnosis of venous thromboembolism in general practice database studies. *Br J Clin Pharmacol* **49**, 591–596.
23. Hammersley V, Hippisley-Cox J, Wilson A and Pringle M (2002) A comparison of research general practices and their patients with other practices – a cross-sectional survey in Trent. *Br J Gen Pract.* **52**, 463–468
24. Gray J, Ekins M, Scammell A, Carroll K and Majeed A (2002) Workload implications of identifying patients with ischaemic heart disease in primary care: population-based study. *Journal of Public Health Medicine* **25**, 223–227.
25. Majeed A, Moser K and Carroll K (2001) Trends in the prevalence and management of atrial fibrillation in general practice in England and Wales, 1994–1998: analysis of data from the General Practice Research Database. *Heart* **86**, 284–288.
26. Office for National Statistics (2000) *Key Health Statistics from General Practice 1998*. Series MB6 No.2, Office for National Statistics: London. Available at: http://www.statistics.gov.uk/statbase/Product.asp?vlnk=4863.
27. Parente S T, Weiner J P, Garnick D W, Richards T M, Fowles J, Lawthers A G et al. Developing a quality improvement database using health insurance data: a guided tour with application to Medicare's National Claims History file. *Am J Med Qual* **10**, 162–176.
28. Cherry D K and Woodwell D A (2002) National Ambulatory Medical Care Survey: 2000. Advance data from vital and health statistics; No.328. Hyattsville,Maryland:National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/data/ad/ad328.pdf.
29. Carroll K, Majeed A, Firth C and Gray J (2003) Prevalence and management of coronary heart disease in primary care: population-based cross-sectional study using a disease register. *Journal of Public Health Medicine* **25**, 29–35.
30. Bindman A B (1999) Can physician profiles be trusted? *JAMA* **281**, 2142–2143.
31. Khunti K, Ganguli S, Baker R and Lowy A (2001) Features of primary care associated with variations in process and outcome of care of people with diabetes. *Br J Gen Pract* **51**, 356–360.
32. Aveyard P (1997) Monitoring the performance of general practices. *J Eval Clin Pract*. **3**, 275–281.
33. Shi L, Starfield B, Politzer R and Regan J (2002) Primary care, self-rated health, and reductions in social disparities in health. *Health Serv Res.* **37**, 529–550.
34. Davies H T, Washington A E and Bindman A B (2002) Health care report cards: implications for vulnerable patient groups and the organizations providing them care. *J Health Polit Policy Law* **27**, 379–399.

ANNEX TABLE 1 SOURCES OF PRIMARY CARE DATA

Source and Purpose	Population covered	Strengths	Weaknesses	Areas for development
QRESEARCH QRESEARCH is currently a pilot database containing data from 43 English general medical practices (family medicine practices). The QRESEARCH team has now recruited 500 general practices spread throughout England, Wales, Northern Ireland and Scotland covering every Strategic Health Authority Area. See www.qresearch.org for further detail.	When it is fully established, QRESEARCH will contain the records of approximately 3.5 million currently registered patients from around 500 general practices	1. Fairly complete recording of morbidity information 2. Includes information on age and sex, and on preventive care 3. Good information on prescribing 4. Now becoming to be used for research 5. Provides some small area socio-economic data 6. Can be used to provide information on current standards of care and the inter-practice variation among contributing practices, for use in 'benchmarking' 7. Can provide time series' data for monitoring purposes 8. Provides some feedback of data to contributing practices	1. Still being developed (but this should be complete soon). 2. There are some costs for using the data 3. All projects have to approved by a Scientific Advisory Group 4. No direct link between diagnosis and prescribing (indirect links only)	1. Reduce the costs of accessing the data by the acquisition 2. Increased use for public health and policy driven research
General Practice Research Database (GPRD) Initially set up as a research tool, principally to carry out pharmaco-epidemiological studies (mainly drug side-effects). Now increasingly used for health services research, e.g., time trends in disease prevalence and treatment, area & inter-practice variations. Costs of data collection are £2–3 million. This has to be paid by researchers and data users via a data charge for using the data. Other primary care databases include THIN, DIN, and the IMS Mediplus system.	1.4–3.5 million people (depending on selection criteria for inclusion of practices). Number of practices included varies, depending on inclusion criteria, from between 200–400.	1. Fairly complete recording of morbidity information. 2. Includes information on age and sex, and on preventive care. 3. Good information on prescribing. 4. Widely used for research (mainly pharmaco-epidemiology). 5. Although the GPRD can not provide information on all practices in England, it can be used to provide information on current standards of care and the inter-practice variation among contributing practices, for use in 'benchmarking'. 6. Can provide time series' data for monitoring purposes.	1. Initially established as a research tool, rather than a statistical database 2. Requires high degree of technical expertise to use 3. Costs of using the data can be high (typically, £25,000 per project) 4. All projects have to approved by a Scientific & Ethical Advisory Group, slowing down access to the database 5. Has a new interface that researchers currently have little experience of 6. Can not be used to provide information smaller areas, such as PCG/Ts 7. No direct link between diagnosis and prescribing (indirect links only) 8. Projects can take a long time to set up and complete. This can often be incompatible with Department of Health timescales and priorities	1. Reduce the costs of accessing the data by the acquisition of a DH/NHS licence. 2. Addition of socio-economic information – initially by the use of practice postcodes, later by patient postcodes. 3. Development of a method for linking mothers' & children's' records 4. Increased use for policy driven research 5. Could be complemented by using information from other primary care databases
Prescribing Analysis and Cost (PACT) data Initially developed as a by-product of the system for reimbursing pharmaceutical contractors. Used extensively for monitoring general practice prescribing and setting prescribing budgets. Also now increasingly used for research. Prescription Pricing Authority's administrative costs are about £62 million per year. Costs of accessing the data are relatively low.	Entire population of England (about 49 million people). Information available on all general practices in England (about 9,000), primary care trusts and health authorities.	1. Used extensively by different organisations, hence considerable experience of how it can be used to monitor prescribing and inform policy. 2. Provides an accurate and complete record of the cost and volume of GP prescribing. 3. Can provide information at general practice, PCT, and national levels. 4. Costs of accessing the data are low, both at national or regional level.	1. Can not be linked to patients – hence, no information on prescribing by age and sex. 2. Can not be linked to diagnostic data- hence, limited information on prescribing for specific disorders. 3. Many of the suggested enhancements to PACT may be unfeasible for technical, financial and political reasons.	1. Linkage of data to information on patients' age and sex. 2. Addition of diagnostic data to information collected on prescriptions. 3. Availability of longer time series data at general practice and local levels (three years data available currently). Much longer term time series data already available at national level, through the Department of Health's PCA system.
Fourth Survey of Morbidity in General Practice (MSGP4) The survey examined the patterns of disease in general practice, in relation to the age, sex, and socio-economic status of patients Cost about £6 million pounds to complete. Costs of a repeat survey are expected to be considerably lower. Costs of accessing the data are very low.	502,000 people from 60 volunteer practices in England and Wales	1. Provides a complete record of consultations with GPs. 2. Good recording of primary care morbidity data. 3. Contains individual level demographic, socio-economic and ethnic data.	1. Carried out in 1991–92, so now 10 years out of date. 2. No information on prescribing. 3. Limited information on the use of secondary care.	1. More up to date information on GP workload and socio-economic variations is needed. 2. Need to include information on prescribing, use of community services, and hospital referrals in a new survey.

ANNEX TABLE 1 continued SOURCES OF PRIMARY CARE DATA

Source and Purpose	Population covered	Strengths	Weaknesses	Areas for development
Weekly Returns Service Publishes weekly incidence for selected respiratory and communicable diseases compared to the previous week and the mean weekly incidence of influenza and flu-like illness. Costs about £400,000 per year to run. Only limited analyses are possible but these are available freely.	600,000 – 700,000 people depending on number of practices contributing data at any one time. Number of practices varies from 60–70.	1. Provides timely information on the number of people presenting to GPs with specified conditions 2. Practices that contribute are experienced in collecting morbidity data and many took part in the Fourth Annual Morbidity Survey 3. Provides information on GP workload	1. Only tabulated data available – not possible to link data as no individual level data. 2. No information on prescribing or community and hospital care.	1. Annual prevalence rates, as well as weekly consultation rates, will be available soon. 2. Provide additional resources to allow data to be linked to individual patients. 3. Include data on treatment. 4. Could be used as the basis of a new morbidity survey.
Local PCG disease registers Established to meet requirements of National Service Frameworks. Do not generally receive special funding in many PCG/Ts. Difficult to access in PCG/Ts where the data have not been collated centrally or where not accessible using MIQUEST. Where this has been done, the costs of access are low.	Typically, one primary care group or trust (100,000–250,000 people). Number of general practices varies considerably depending on PCG area and population.	1. Provides local data on morbidity and sometimes on treatment. 2. Can provide information to help improve the quality of care and the accuracy of data recording.	1. Only produced for a limited number of conditions (most often coronary heart disease). 2. Wide variations in the accuracy of data between practices.	1. Develop guidelines for data recording and an accreditation scheme. 2. Demonstrate benefits to GPs of improving the quality of their data. 3. Develop capability to compare registers in different areas to allow comparisons of PCG/Ts.
General Medical Services (GMS) data Collected through the system of fees and allowances paid to general practices. Used to provide comparative information on general practices and PCTs. Costs of collecting data difficult to measure, as this is part of the administrative function of health authorities and PCTs. Costs of accessing the data, both locally and nationally, are low.	Entire population of England (about 49 million people). Information available on all general practices in England (about 9,000).	1. Provides local and national level data. 2. Commonly used in local primary care indicator packages. 3. Straightforward to access at both local & national level.	1. Largely derived from claims and registration data. 2. No clinical information. 3. No information on process of clinical care.	1. Also include information from local audits and disease registers. 2. Publish information on practice and PCT performance to allow greater public scrutiny.
National Database of Primary Care Groups and Trusts Established to provide baseline information on PCG/Ts, for use in monitoring and research. Funded by the Department of Health. Access to the database is free for registered users.	All primary care groups and trusts in England (about 49 million people) Information is not available at general practice level.	Provides population, socio-economic and GMS data on all PCG/Ts in England. Updated regularly.	1. Does not yet contain any information on prescribing, hospital referrals or hospital admissions. 2. No information on the supply of secondary care services.	1. Addition of information on prescribing and use of hospital care, as well as on waiting lists and times. 2. Addition of supply. variables, such as specialists or hospital beds per capita.

ANNEX TABLE 1 continued Sources of primary care data

Source and Purpose	Population covered	Strengths	Weaknesses	Areas for development
TWO NETWORKS (PRIMIS AND TRENT FOCUS) THAT USE MIQUEST **PRIMIS (P)** Main purpose to educate practices through local information facilitators in data quality and information management skills. Data quality assessment and feedback integral part of the process. Also, clinically-focused data extraction done regularly on a variety of relevant topics. Costs about £1.2m per year (for training, data quality work and CAS). **TRENT FOCUS (TF)** Network of 60 practices recruited from the whole of Trent. Minimum data quality standards. Patients representative of other patients in Trent. Paid £1,000 a year linked to participation in research studies. Track record of research successfully completed and published. The CRN model could be extended nationally via PRIMIS.	**PRIMIS** 20 million patients from 3,000 practices spread across 150 PCTs; and potentially double that by March 2003. **TRENT FOCUS** 800,000 patients from 60 practices spread among 51 PCTs.	**THE NETWORKS** 1. Representative sample of practices (TF). 2. National sample (P). 3. Accessible data, low cost. 4. Validated databases of proven quality (TF and increasingly P). 5. Continuous monitoring of data quality and morbidity (P). 6. Includes information on age and sex, and on preventive care (TF and P). 7. Good information on prescribing (TF and P). 8. Can obtain questionnaires from patients at cost. **MIQUEST** 1. Compatible with European laws for confidentiality and consent. 2. Can be used to write flexible queries tailored to evolving needs. 3. Has been used for successfully for research including monitoring NSFs and inequalities and family studies. 4. Can be used to provide summary information on smaller areas such as PCGs/T and report on the inter-practice variation among contributing practices, for use in 'benchmarking'. 5. Can provide time series data for monitoring purposes. 6. Interview data from staff.	**PRIMIS** 1. Dependent on PCTs investing in facilitator. 2. Dependent on quality and abilities of facilitator. 3. Dependent on willingness of practices to change. 4. Data quality improvement is a slow process. **TRENT FOCUS** Need to validate recording of death data. **MIQUEST** 1. Requires high level of expertise to specify and write queries. 2. Dependent on quality of system interpreters.	1. Detailed strategy for improving MIQUEST under discussion: • System suppliers to update. MIQUEST interpreters. • MIQUEST language needs updating. • Needs network option to be activated by suppliers to speed up response time and make process less people-intensive. 2. Add socio-economic data on all patients in selected practices by linking postcodes and then survey data. 3. Use of household ID to enable linking of families for genetic studies. 4. Increased use for policy driven research. 5. Could be complemented by using information from other primary care databases. 6. Already developing web based feedback to practices.