

The value of linked data for policy development, strategic planning, clinical practice and public health: An international perspective

Michael Goldacre

Professor of Public Health, Unit of Health-Care Epidemiology, Department of Public Health, University of Oxford, Institute of Health Sciences, Old Road, Oxford, OX3 7LF, email michael.goldacre@dphpc.ox.ac.uk

In this paper I discuss the value of using linked data from medical records of illness. The rationale for studies based on medical records is that each person's illness – if recorded, added to others and analysed – has the potential to contribute to the advancement of medical knowledge. There is nothing new about this idea, at least in respect of hospital care. A paper in the *Lancet* in 1841 commented: "When we reflect upon the great number of institutions for the treatment of disease which exist in this country, the gratification arising from the thought of how much suffering might be saved, and affliction relieved by them, is mingled with regret that so vast a source of information on the history of disease and the results of therapeutic treatment should be lost to mankind."¹ The author commented on the potential value of analysing hospital records for among other things, what we would nowadays call outcomes research, audit and public accountability. There is nothing new about the importance of analysing data from medical records. What is fundamentally different between now and 160 years ago, and indeed between now and even five or ten years ago, is the enormous power, at very low cost, of large-scale information processing.

In record-based studies, the added value of data linkage is that two or more items of information about an individual, when recorded at different times and perhaps in different places, may have greater value, if both are considered together, than when either is considered alone. Donald Acheson, the founder of the Oxford record linkage study in England, wrote that when episode-based statistics were first introduced in England in the 19th century – notably statistics based on death registrations – most diseases of public health importance were abrupt, brief, and often fatal.² These characteristics of brevity and fatality meant that, typically, an individual doctor had the opportunity to observe the full circumstances of the disease. Clearly, this is very different with many of the diseases of public health importance today.

I am going to develop my themes on the value of linkage by illustrative examples, drawing, first, on examples from a couple of well-established registration systems; then from work on health inequalities; then studies of treatment rates and outcomes in the English National Health Service (NHS); and finally on some clinically- and epidemiologically-related examples from our recent work.

Cancer and death registration

As we consider novel applications of data linkage, it is also important to acknowledge the value of public health systems that already use data linkage and have done so for decades (though we might not generally think of them in those terms). With these, the value of linkage is already established beyond doubt. Cancer registration is an important example. Typically, to ensure that data capture is as complete as possible, cancer registers receive data about cancers from multiple sources – hospital discharge records, pathology laboratories, death certificates – and often from several sources for the same person and same illness. The data are collated by linkage of one sort or another – historically by drawing data together clerically – to ensure that each individual is counted once, and once only, for each cancer. From these processes, we have a vast body of information, locally, nationally, and internationally, on the incidence of cancer. And, with linkage of cancer registrations to death registrations, we have extensive information about survival rates.

There can be no doubt about the contribution of cancer registration systems to our understanding of the epidemiology and outcomes of cancer. I shall give an example with particular impact on health policy and planning in England. Comparisons of regional and national cancer registry data with those from other countries show that, for some cancers, survival rates in parts of England are worse than those in some continental European countries.³ This contributed to the evidence behind the cancer policies in the government's NHS Plan.⁴ The NHS Plan is the government's key policy document that will guide the development of services over the next few years. The Cancer section of the Plan says that "the government will invest in cancer services to develop services to rival the best in Europe;" to ensure that quality of care is uniformly high across the country; and, to these ends, to develop better professional networks of cancer care" (my italics, to emphasise government's acceptance that in some respects we lag behind). We know about cancer, and about how outcomes compare, because we have information systems, including linkage, to draw on.

Another system of linkage in England which we tend to take for granted, but which we should note as having amply proved its value, is that for the linkage of study cohorts to national death registration data. These are often cohorts that the investi-

gators have established through initial personal contact with the individuals who comprise the study population; from whom the investigators have obtained exposure data; and whom the investigators follow up personally for at least a number of years. The importance of also linking to the national death files is that it increases completeness of ascertainment of death in the study population; it reduces or eliminates biases resulting from loss to personal follow up; and it makes automated long-term follow-up possible after the time when personal follow-up becomes impractical or unaffordable. One example of an English cohort linked to national mortality files is Richard Doll's study of doctors, their smoking histories, and their development of smoking-associated diseases.⁵ Another is Martin Vessey's study of women using different forms of contraception.⁶⁻⁷ Both have benefited enormously from very long-term follow-up. I mention these examples because of the considerable importance of their findings in public health, health policy, and in guiding clinicians' advice to patients in clinical practice. Doll's study is one of the greatest contributions to understanding causes and prevention of disease from any medical discipline. Vessey's study, and several others of similar design on the same topic, are important because the extensive use of widely available, effective contraception – particularly in the past thirty years or so – has been one of the most far-reaching developments in the entire social history of humankind. Accordingly, the issue of safety of contraceptive methods – in particular, the pill – has been one of considerable importance too. Linkage of survey data to national death data has played an important part in understanding both the risks of smoking and (by and large) the safety of oral contraception and of sterilisation for family planning purposes.

Inequalities in health

My next example concerns inequalities in health, as measured by social class. It is well recognised that there are steep social class gradients in mortality in England, with much higher mortality rates in lower than in higher social classes. There has been some uncertainty, however, about how much of the gradient may be attributable to 'drift' down the social class scale of people who are unfit or unwell.

The English national linked Longitudinal Study, which I shall describe more fully in my next presentation, comprises linkage between census records and mortality records. One of its most important uses has been to demonstrate social class gradients in mortality, with the lowest (least advantaged) social classes having the highest mortality rates. By studying the time sequence of events, it demonstrated that 'drift' of ill people down the social scale has only a very small part to play in the cause of the gradient. The importance of this in policy terms is considerable. To quote the NHS Plan again, government in England now gives this full recognition.⁴ It says: "No injustice is greater than the inequalities in health which scar our nation. The life expectancy of a boy born now into the bottom social class is nine years less than a boy born into the most affluent social class." The evidence for the latter statement comes from the linked longitudinal study. The NHS Plan continues: "The

worst health problems in our country will not be solved without dealing with their fundamental causes. This means tackling disadvantage in all its forms – poverty, lack of educational attainment, unemployment and social exclusion."⁴

Hospital admission rates in an English population

My next example illustrates the use of linkage of hospital admission data. In conventional systems of hospital statistics, if a person is admitted to hospital, say, ten times, ten admissions are counted; and there is no way of knowing whether the ten admissions represent one person, or ten people, or any combination in between. Hospital admission rates in England have risen, seemingly inexorably, for many years. A high rate tends to be taken as a mark of high productivity; and the claim is often heard that the National Health Service is "treating more people than ever before". However, it is possible, of course, that an increasing rate may simply reflect higher levels of repeat admissions rather than an increase in the treatment of individual patients; and, without linkage, it is impossible to know. We have studied long-term trends in admission rates in the Oxford Record Linkage Study (ORLS) area. The ORLS comprises linked data on hospital admissions and deaths. The data show that admission rates measured as unlinked episodes have risen impressively over many years. Linkage allows us to identify admission rates measured as "people admitted". These rates, of people admitted, have risen much less sharply. Much of the increase in the episode-based rates is clearly an increase in multiple admissions per person. Much of the increase in the component that represents admissions from the waiting list comprises a rise in admissions for relatively minor procedures, like endoscopies; and there has been a large increase in multiple short-stay admissions per person (e.g. repeat cystoscopies). Admission rates for ordinary elective admissions – admissions from the waiting list which include at least one overnight stay – have actually declined. Once dissected out, it is clear that the scale of the decline in person-based admissions from the waiting list is actually rather striking. Admission rates from the waiting list, counting individual people treated, dropped to a lower level in the late 1990s than at any time since the 1960s. This would not matter if the NHS had succeeded in clearing, or radically shortening, its waiting lists for elective admissions. In fact, however, waiting lists stand at very high levels. At least some of the increase in NHS productivity, measured as a rise in episodes of hospital care, is illusory. We did our studies as part of the present Government's National Beds Inquiry.⁸ Our study is one of a number of pieces of a jigsaw which led the National Beds Inquiry to the conclusion that the NHS has reduced its capacity for elective hospital care too far.⁸

Mortality rates following care

Linkage of hospital data to mortality enables one to study death rates after care, regardless of when and where deaths occur. There is much current interest in mortality rates following care; and, in particular, following surgical care. This partly

follows from the Bristol case of high death rates following paediatric cardiac surgery. But, as well as the occasional extreme example of poor performance, there is important information to be obtained about prognosis, more generally, by analysing patients' records. Information about the outcomes expected after hospital treatment, based on contemporary experience systematically analysed, should be routinely available to doctors and patients. Using unlinked hospital statistics, the only measure that is generally available in England is the so-called 'in-hospital fatality rate'. For example, in-hospital post-surgical fatality rates include only those deaths that occur in the admission in which the surgery was undertaken. Those that occur after transfer to other hospitals, or after discharge, are not identified. We have studied the profile of deaths in the first 30 days after surgery in the Oxford record linkage study area.⁹ In-hospital fatality within 30 days of surgery is a conventionally used statistic. In the 1960s and 1970s, when hospital statistics were first routinely collected in England, and when lengths of hospital stay were much longer than they are now, the great majority of post-surgical deaths within 30 days of operation occurred in the admission episode for the operation. Nowadays, with much shorter lengths of stay and higher transfer rates between hospitals, the proportion of deaths which would be missed by counting in-hospital deaths only is substantial: about 40% of all 30-day deaths would now be missed by unlinked data on in-hospital deaths only.⁹ It is clear that any serious attempt to study post-operative mortality must include deaths beyond those in the admission in which surgery took place.

The English Department of Health has recently publicly published 'hospital league tables' of fatality rates for fractured neck of femur and stroke, comparing hospitals across the country, using routine national unlinked hospital statistics. We have compared fatality rates for each hospital in the ORLS region, using unlinked and linked data, regardless of when and where death occurred, for the two conditions. Comparing hospitals, in-hospital fatality (that without linkage) was a poor predictor of the hospitals' ranking on true fatality rates after fractured neck of femur.¹⁰ This is partly because hospitals have different lengths of stay – some 'miss' more of their own deaths than others – and partly because small but significant differences at 30 days had gone by 90 days. Comparing hospitals, in-hospital fatality after stroke correlated quite well with longer-term fatality rates which did indeed vary between hospitals. The implication of these findings is that, if effort is invested in promulgating death rates, comparing hospitals, based on unlinked data, false alarms will be sparked (as with fractured neck of femur). But, conversely, if death rates are not monitored, because of concerns about the inadequacies of unlinked data, one may be missing the identification of hospitals that have unacceptable results. The important point is that one could not know, either way, whether there are real differences in outcomes unless linked data are used. There is a move towards public provision of hospital death data. It is important to make data as reliable as possible. This, if nothing else, is driving a move towards national record linkage in England.

National record linkage in England

So, a brief mention of national record linkage in England. Our research team now has a commission to do pilot studies of two years' national hospital-to-hospital and hospital-to-death linkage. Its scale is fairly formidable – in two years, 24 million hospital records and 1.6 million death records in a total population of 50 million people – but we are confident that the pilot will be successful.

Suicide following discharge from psychiatric care

A specific study undertaken by us on "death after care", which I believe helped influence policy, was one of suicide in patients after discharge from psychiatric in-patient care.¹² We became interested in this because of a smallish number of local high profile cases, reported in the media, of people who had killed themselves shortly after being discharged from care. We wondered whether suicide was truly a rare event following discharge; or whether it was more predictable than that. We showed, on a large-population basis, that suicide rates are indeed substantially elevated shortly after discharge.¹² These findings led, I believe, to a greater general awareness of the vulnerability of some psychiatric patients in the early period after discharge; and of the need for careful post-discharge planning. Our Western Australian colleagues have done similar work on this.¹³

Disease associations

An interesting application of medical data linkage is the study of associations between different clinical conditions.¹⁴ Examples we have studied or are currently studying, using linkage, include the following. Does cholecystectomy predispose to colon cancer?¹⁴ Is trauma an aetiological factor in multiple sclerosis? Does appendectomy in childhood protect against Crohn's disease and ulcerative colitis?¹⁵ Does vasectomy increase the long-term risk of testicular cancer and prostate cancer?¹⁴ Is abortion followed by an increase in long-term risk of breast cancer?¹⁶ As a comment on study design, all these hypotheses have something methodological in common: it is very unlikely that anyone will ever do a randomised controlled trial to test them.

I shall briefly describe one example, the issue of whether termination of pregnancy causes breast cancer. This is a controversial issue, which has been quite extensively but inconclusively studied. Most of the studies have been interview-based case-control studies and some have reported a small elevation of risk of breast cancer in women who have had an abortion. One of the issues in the literature is whether retrospective interview studies of this topic have been prone to responder bias: are women with breast cancer more likely than controls to tell an interviewer if they have had an abortion? Responder bias would be impossible in a randomised controlled trial; but it would be impossible to do one. Responder bias would be impossible in prospective cohort studies or in studies of linkage of independent records. It would be difficult, however, to do cohort studies following women for many years after abortion

to monitor adverse events. Few women would welcome long-term personal follow-up, perhaps for 25 years, after abortion. It is therefore attractive to use data linkage, with linkage of information from independent records. This is what we did; and we found no evidence of an elevation of risk of breast cancer following abortion.¹⁶

There are other data linkage studies on this topic, from New York, Denmark and Sweden.^{17–19} The New York study, the earliest, reported an elevated risk. The Scandinavian studies, like ours, did not confirm any elevation of breast cancer risk. This is a story without a final ending as yet. It would be good if others with longstanding record linkage systems could add more studies to the literature.

Measles-mumps-rubella immunisation (MMR) and meningitis

This is a study that we did a few years ago because of concerns that MMR vaccine might have caused cases of aseptic meningitis. In the Oxford record linkage study area, we linked hospital records of children with meningitis to immunisation records.²⁰ The hypothesis was that, if aseptic meningitis was caused by immunisation, it would have occurred within 15 to 35 days of immunisation. We did indeed find a small cluster of infants with aseptic meningitis within this time-frame.²⁰ In none of the cases had the clinicians considered a link between immunisation and meningitis. This, and other laboratory-based evidence, led the government to change from the Urabe vaccine (the vaccine that had been used in the cases studied at the time) to Jeryl-Lynn vaccine.

Data from general practice

For many purposes, including some of my examples, general practice is obviously the best place to go, if possible, for data.²¹ There is now, in England, a very extensive research dataset from general practice, the General Practice Research Database (GPRD), which contains about 20 million person-years' data. The GP systems which now feed into it were developed at a time when the English Department of Health had no interest in general practice information systems. It was therefore left to individual entrepreneurs, notably Dr Alan Dean who developed what became the basis of the GPRD, to develop GP data commercially for research. It is now being extensively used by those who can afford to use it!^{21,22}

Comment

A few reflections. First, recent advances in data processing should make it possible to realise ambitions with data linkage studies that previous generations of advocates have only dreamt of. Second, partly made possible by this, general practice data are increasingly being computerised. They have enormous potential if they can be accessed for research. Third, it is important to get a wide base of commitment to medical data systems and data linkage. In our experience there is a tendency for health service management to say "this sounds good, but it's obviously research"; and for research bodies to say "this sounds

good, but it's obviously health service management". Fourth, it is clearly important, if possible, to develop systems in large representative populations, because findings from small populations can be questioned, rightly or wrongly, as being unrepresentative. For this reason, in particular, I welcome the prospect of national hospital record linkage in England.

References

- 1 Anonymous. Importance of case-books in hospitals. *Lancet* 1841, July 10: 545–6.
- 2 Acheson ED. *Medical Record Linkage*. Oxford: Oxford University Press 1967.
- 3 International Agency for Research on Cancer. *Survival of Cancer Patients in Europe: The Eurocare Study* (eds. Berrino F et al). Lyon: IARC Scientific Publications no. 132, 1995.
- 4 Department of Health. *The NHS Plan*. London: Stationery Office 2000.
- 5 Doll R, Peto R, Wheatley K et al. Mortality in relation to smoking: 40 years' observations on male British doctors. *British Medical Journal* 1994; 309: 901–911.
- 6 Vessey M, Doll R, Peto R et al. A long-term follow up study of women using different methods of contraception. *J Biosoc Sci* 1976; 8 : 373–427.
- 7 Goldacre MJ. The role of cohort studies in medical research. *Pharmacoepidemiology and Drug Safety* 2001; 10: 5–11.
- 8 <http://www.doh.gov.uk/nationalbeds.htm>
- 9 Goldacre MJ, Griffith M, Gill L, Mackintosh A. In-hospital deaths as a fraction of all deaths within 30 days of admission for surgery: analysis of routine statistics. *British Medical Journal* 2002; 324: 1069–1070.
- 10 Goldacre MJ, Roberts SE, Yeates D. Mortality following admission to hospital with fractured neck of femur: database study. *British Medical Journal* 2002; 325: 868–869.
- 11 Roberts SE, Goldacre MJ. Case-fatality rates after admission to hospital with stroke: linked database study. *British Medical Journal* 2003; 326: 193–194.
- 12 Goldacre M, Seagroatt V, Hawton K. Suicide after discharge from psychiatric inpatient care. *Lancet* 1993; 342: 283–286.
- 13 Lawrence D, Holman CD, Jablensky AV et al. Increasing rates of suicide in Western Australian psychiatric patients: a record linkage study. *Acta Psychiatrica Scandinavica* 2001; 104: 443–451.
- 14 Goldacre M, Kurina L, Yeates D, Seagroatt V. Use of large medical databases to study associations between diseases *Quarterly Journal of Medicine*, 2000; 93: 669–675.
- 15 Kurina LM, Goldacre MJ, Yeates D, Seagroatt V. Appendectomy, tonsillectomy and inflammatory bowel disease: a case-control record linkage study. *J Epidemiol Community Health* 2002; 56: 551–554.
- 16 Goldacre MJ, Kurina LM, Seagroatt V, Yeates D et al. Abortion and breast cancer: a case-control record-linkage study. *J Epidemiol Community Health* 2001; 55: 336–337.
- 17 Howe HL, Senie RT, Bzduch H, Herzfeld P. Early abortion and breast cancer risk among women under age 40. *Int J Epidemiology* 1989; 18: 300–304.
- 18 Harris BM, Eklund G, Meirik O et al. Risk of cancer of the breast after legal abortion in the first trimester: a Swedish register study. *British Medical Journal* 1989; 299: 1430–1432.
- 19 Melbye M, Wohlfahrt J, Olsen JH. Induced abortion and the risk of breast cancer. *New England Journal of Medicine* 1997; 336: 81–85.
- 20 Miller E, Goldacre MJ, Pugh S et al. Risk of aseptic meningitis after mumps vaccine in children. *Lancet* 1993; 341: 979–982.
- 21 Lis Y, Mann R. The VAMP Research multi-purpose database in the UK. *J Clin Epidemiol* 1995; 48: 431–433.
- 22 Wood L, Coulson R. Revitalising the General Practice Research Database: plans, challenges and opportunities. *Pharmacoepidemiology and Drug Safety* 2001; 10: 379–383.