Editorial

Primary care groups as community laboratories

Key points

- Amongst observational strategies, cohort (or longitudinal) studies provide the most reliable evidence of association but prospective studies necessarily entail long waits before the findings emerge.
- ‘Retrospective’ cohort studies are possible, when baseline information on potential cohorts is collected and archived in anticipation of promising future hypotheses.
- With advent of techniques for storing very small samples of biological material and of techniques for their subsequent analysis (e.g. of DNA) and with the data storage capability of computing now available, the archiving of complete potential cohort studies is technically feasible.
- It is proposed that such archiving of data and material might be established formally in ‘primary care groups’ population approximately 100 000) with the active collaboration of both health care professionals and patients.

In spite of the rapid increase in our knowledge of genetics and the genetic contribution to the development of certain human diseases, environmental factors remain important as contributors to the expression of many diseases and conditions. This is probably most neatly illustrated by studies of migrants, whose incidence of (say) cancer or heart disease often assumes rates comparable with those for the host country and over a period of time far shorter than could be explained by genetic changes alone.

Public health and the health services in the developed world are increasingly concerned with the management of chronic disease, including cancer. The ‘gestation’ periods for many of these diseases and conditions are measured over months and years, and the aetiologies are often described as ‘multifactorial’. The environment is becoming more complex. The earth, air and water contain an increasing number of chemicals, produced by innovative industries that are capable of introducing new compounds at rates that exceed the capability of regulatory agencies to check their safety. Our food and drinking water also contain a number of residues that are a cause for concern. Furthermore, the changing nature of employment, characterized by a decline in manufacturing and heavy industry, and in manual labour in those remaining, and a shift to service industries, changes the nature of ‘occupational’ exposure. Therefore, whereas, in the past, external and physical factors (e.g. coal dust exposure in miners) were likely to be the subject of interest, the move to service industries means that it may be more relevant to consider psycho-social variables in respect of work-associated morbidity. Whereas discrete incidents allow the identification of the relevant (exposed) cohort with relative ease, a low-level long-term exposure to certain environmental factors may produce as much or more morbidity but in a more insidious way. Examples where these types of situation might be operating include the hormone-mimicking or antagonistic effects of certain chemicals, declining sperm counts and increases in cancer of the reproductive organs, such as testicular cancer, and breast cancer.

Epidemiological methods for elucidating putative causes or risk factors rely heavily on observational studies. Cohort studies are to be preferred, to reduce sources of bias, give clearer evidence about the chronological sequence of variables and disease (i.e. exposure or development of a ‘risk’ factor precedes the onset of disease) and allow direct calculation of relative risk. However, feasibility, a quick answer and cost usually favour case-control studies. If cohort studies provide ‘better’ data but at the expense of a long wait for any answers, epidemiologists and custodians of the public health should develop more effective and cost-effective cohort studies. A solution, we suggest, lies with the ‘retrospective cohort study’.

The value of existing databases can be illustrated by three examples from literature relating to ischaemic heart disease (IHD) and their ability to support (or refute) certain hypotheses. In 1975, Miller and Miller presented a telling case for reconsidering the role of the high-density lipoprotein (HDL) fraction of circulating lipids as an important risk factor for IHD. The existence of data on HDL, originally collected during 1969–1971 as part of the Framingham study, allowed the analysis of such data with respect to incident cases of IHD in the subsequent period, which was published only 2 years after the Miller paper. A second example relates to the Caerphilly heart disease study, where samples of plasma were stored during cohort recruitment, 1973–1983. In 1988, a report of a case-control study suggested increased concentrations of antibodies to Chlamydia pneumoniae in patients with acute myocardial infarction and chronic IHD. The availability of stored plasma in the Caerphilly study allowed retrospective measurement (in 1998) of antibody levels in men at recruitment (1973–1983) and before the manifestation of acute coronary events, and this allowed examination of both the association and its direction. Third, in 1989, Barker suggested that IHD had its origins in early life, when he presented data relating low birthweights to
increased rates of coronary artery disease in adulthood, and subsequently that these same individuals had a higher prevalence of risk factors, such as hypertension, non-insulin-dependent diabetes and abnormalities in lipid metabolism, than those with normal birthweights. All these early origins of IHD studies depend on the existence of good historical records; in fact, on ‘archived cohorts’.

What we now propose is that a number of population-based cohorts should be identified, established and archived for future use. These cohorts would collect and store personal data, together with biological samples. Clearly, in terms of co-operation and response rates, the less invasive the method of collection of these samples, the better. Thus, samples of blood, plasma, serum, urine and hair and nail clippings could be considered. Freeze-drying of samples may make storage easier. To record the stability or otherwise over time, repeat measurements separated by a number of years could be made. The case for banking human tissue for epidemiological studies has already been made. The technologies for assaying very small quantities of potential hazards (contaminants; e.g. lead in blood or urine) or biological markers (e.g. IgG for Chlamydia pneumoniae) exist, and it can be expected that these technologies will improve and new, more sensitive tests will be introduced to detect contaminants unrecognized at present, infections or biological markers. Only when a clear hypothesis has been made explicit would samples be removed from storage and submitted for measurement of the relevant variables.

We suggest that a suitable population base for these cohorts, in the United Kingdom, would be the primary care groups (PCGs), as proposed in the White Paper The new NHS. These PCGs are based on populations of 100 000 or so and bring together the general practitioners (GPs) and other health care professionals who serve that same population. The intention is to create a new structure within the NHS with, over time, an increasing autonomy over the planning and delivery of primary care, the commissioning of secondary care, and continuing professional development including ‘clinical governance’. Given the forward-looking direction of the White Paper, we are presented as a joint partnership between researchers and individuals, where those taking part are seen as participants rather than subjects; such an approach is gaining ground in the context of randomized trials. A sense of belonging is important also in prospective cohort studies. In terms of the central coordinating function, it would clearly be important to include lay representation in any group, having the responsibility for formulating and prioritizing questions to be addressed, and the capacity for more formal population-based studies. Furthermore, even in partnerships of GPs, the base population (or the relevant subset) is likely to be too small to produce sufficient cases for study to answer many current questions. Primary care groups of 100 000 or so provide ‘useful’ numbers of cases for many conditions. For studying conditions with a low incidence rate, several PCGs may need to be involved. The advantages of using the PCG structure, intended primarily for health care delivery, as the basis for studies are clear: most individuals in the UK register with a GP, and combined lists of GPs are the nearest approximation to a UK population register. Other practical reasons for basing these potential cohorts on PCGs include the fact that patients visit their GP regularly – 66 per cent of men and 79 per cent of women are seen each year by their GP – which allows the opportunistic capture of data. In case this is seen as adding another burden to GPs’ already full workloads, much of the information is already collected routinely and computing systems are currently available that can abstract and store the relevant data. Much of the extra data collection could be carried out by specially trained practice nurses.

Data linkage of individuals who moved to an area covered by a different PCG will be facilitated, in theory, by the introduction of the more user-friendly NHS number.

As described so far, this exercise may sound the equivalent of a ‘fishing trip’ amassing large amounts of data, for which no precise hypotheses or questions have been made explicit. Normally, these questions once formulated serve to provide the basis for collecting such data prospectively. Given the long lead time for the manifestation of much chronic disease, the difficulty lies in anticipating what the question(s) should be. By the time putative associations can be postulated, setting-up cohorts anew means a long period of follow-up before a sufficient number of incident cases is reached.

These proposals for ‘latent cohorts’ and archiving of information and biological material will require a clearing house to provide a necessary structure and co-ordination. It would be necessary also to establish protocols for assessing the priority of questions, for which the samples would provide the relevant data. Parallels already exist in the research field in the United Kingdom, where areas for priority are determined centrally.

To make this approach work two further important conditions need to be met: first, a degree of co-operation between doctors and professionals from many disciplines; second, and perhaps more important, the co-operation and informed consent of individuals to answer questionnaires and provide samples. This would be facilitated if the approach were presented as a joint partnership between researchers and individuals, where those taking part are seen as participants rather than subjects; such an approach is gaining ground in the context of randomized trials. A sense of belonging is important also in prospective cohort studies. In terms of the central coordinating function, it would clearly be important to include lay representation in any group, having the responsibility for formulating and prioritizing questions to be addressed, and
therefore, the samples to be used, given that each question answered carries a number of opportunity costs.

In the United Kingdom, the Data Protection Act 1998 may provide a number of difficulties to this approach, as it will require the aggregation of data at the individual level, and this may involve the linkage of a number of datasets. However, Section 33 of the Act allows for certain exemptions, including research where this is the prime purpose for the collection and storage of data. It would be good practice, anyway, to make clear to participants the purpose of collecting the samples and data when seeking their consent at the outset. Furthermore, participants would be given the guarantee of personal anonymity and confidentiality, that no information would be published that would permit the identification of individuals.

The ideas described in this paper came about in response to the difficulties of orthodox approaches to research and its constituent methods to deal with problems, or questions, where relatively long time spans may separate exposure and the manifestation of disease. Methods relying on case–control studies may introduce a number of biases, which make drawing of conclusions difficult. Prospective cohort studies may take a considerable period of time to produce relevant data. All of these factors have profound implications for the advancement of scientific knowledge, health care and public health. With the proposed approach, a framework is introduced that allows cohort studies to be undertaken 'retrospectively'.

David Bainton
Senior Lecturer (Retired)
Robert West
Reader in Epidemiology
University of Wales College of Medicine, Cardiff
E-mail: WestRR@cardiff.ac.uk

References


Accepted on 26 July 2001