Recruitment to UK Biobank has begun. The planned recruitment to a cohort of 500 000 men and women aged 40–69 over 5 years remains an ambitious target, but one which demonstrates the scale of study that is required to investigate many emerging research questions. The essential feature of a cohort study, following change over time in exposure and health status at the individual level, is a natural method through which to obtain a complete model of disease causation and public health.1

The value of large scale cohort studies is their ability to address many aetiologic questions definitively, particularly those involving gene/environment interactions. Large numbers also mean that many research questions can be answered in socially and politically acceptable time frames. However, due to inevitably low response rates, their public health utility is limited. Unrepresentativeness leads to biased estimates of incidence rates and biased estimates of the population impact of risk factors. Furthermore, since Doll and Hill established the first aetiological cohort,2 the design has had little public health development.

The high cost of large cohort studies remains an important obstacle to investment in new cohorts. The understandable preference of research funding agencies is to add value to existing cohorts through data sharing. Although data sharing is a helpful way of reducing the influence of study specific bias, many emerging research questions are unlikely to be satisfactorily answered through data collected for other purposes. An alternative economy is to use register-based cohorts. Although enabling very large studies,3 this method is limited to hypotheses involving routinely collected data, and many conditions, such as dementia, will be under diagnosed.4

As the need for longitudinal data increases rather than diminishes, the design and purpose of cohort studies needs to be re-considered. One way to provide greater financial and political incentive to fund new cohorts is to integrate aetiology, public health and health service provision outcomes within the same project. These integrated cohorts may be termed ‘fungible‘ as they trade off benefits between different areas of interest for the greatest public good according to the allocation of resources. For example, over one period of time most resources may be given to addressing aetiologic questions, whilst over a different period the same infrastructure may be used to address largely public health issues. The five essential elements of a fungible cohort are:

(i) A national (or regional) health dedicated database (HDD). Using linked public health registers, and on the basis of general consent, the entire community (the population) is studied. General consent is a matter of public health policy and all individual data are de-identified to protect confidentiality. Although there is little data on the acceptability of such a proposal, it is closely similar to the register-based cohorts of Scandinavia. It is also a necessary pre-condition of efficiently conducting large studies, which provides great population benefit at no individual cost or risk. The HDD may be used for stand alone register-based studies.

(ii) Using the HDD to identify and recruit population samples for nested studies. Nested studies involve hypothesis driven data collection according to specific consent. Specific consent is a matter of individual choice. Nested studies can be used to:

(a) Identify the distributions of risk factors within the population,
(b) Identify the prevalence and incidence of disease within the population,
(c) Investigate aetiological mechanisms,
(d) Investigate the effect of public health practice and policy.

Population samples are recruited according to purpose. For public health studies, including prevalence and incidence studies, the major selection issue is representativeness of the target population to provide generalizable population estimates. For aetiological studies the major sample selection issue is obtaining sufficient range in the exposure or independent variable of interest in order to detect an effect. For health service and policy evaluation the major selection issue is exposure to variations in practice.

(iii) Coordination of de-identified HDD data and nested study data to evaluate systematic and random error. An integrated infrastructure provides the opportunity to systematically study and adjust for sources of error. A systematic programme of error quantification is essential to ensure sufficient precision and validity. Measurement issues of interest include selection and indication bias, regression dilution and inter/intra individual variation in exposure.

(iv) A sufficiently large population to allow definitive conclusions to be drawn from nested studies. Size is relative. A larger population is required for rare conditions than for common ones. A larger sample is required where the range of exposure is small than for when it is large. Given developments in health informatics, regional or national HDDs involving several million people and allowing nested studies involving around 100 000 participants would seem to be a minimum expectation.

(v) A sufficiently diverse community (population) to provide a range of exposure levels in nested studies. Although large numbers will increase the number of persons with extreme exposure levels in nested studies, limitations due to the range of exposure available in the population and the range of exposure likely in a population sample must be taken into account. For some hypotheses the range of available exposure levels within the population will be inadequate. Under these...
circumstances, studies involving more than one population are required.

The public health benefits of a fungible approach stem from the ability to conduct a coordinated programme of nested studies relatively cheaply and quickly once the infrastructure is established. For example, the population impact of aetiological mechanisms which have been identified in large non-representative nested studies may be estimated reasonably and precisely using information on risk factor distributions obtained through smaller studies designed to achieve high response rates in specific localities.

Conducting fungible projects would certainly provide many challenges. At a basic science level, however, the integrated study of entire communities provides a much richer universe of investigation, involving fewer institutional and other artificial boundaries. It is a reasonable anticipation that a fungible project would facilitate multidisciplinary research and provide research impetus generally. For fungibility to succeed, however, political and scientific leadership is paramount. Popular support is required along with substantial community engagement. Issues of governance and data access will be of supreme importance to maintain public confidence, and the co-ordination of a wide variety of academic and clinical interests will be essential. At a technical level we have little idea of the complexity and cost of the management and informatics infrastructure required to run a large fungible project; but the basic bioinformatics and health care informatics technology is available to achieve fungibility with a high degree of validity, precision and confidentiality in many developed and developing economies.

The bottom line, however, is cost and benefit. If large scale aetiologic cohorts are expensive, how much more would be fungible studies? Fungibility brings a substantial increase in benefits rather than reduction in costs. Integrating aetiologic, public health and health services research questions provides a better return from infrastructure investment. An integrated research environment will also facilitate the translation of basic science into improved public health. Nevertheless, it is acknowledged that new methodology is also required to reduce the cost of cohort studies per se. Utilizing information technology for remote recruitment, measurement and follow-up, can dramatically reduce the costs of cohort studies. It may transpire that electronically managed fungible studies are extremely attractive to politicians and research funders alike, if the cost of the cohort design can be reduced and more translatable research delivered.

High quality observational epidemiology, and its translation into evidence based health policy, is more difficult today than was previously. We rarely have the luxury of the homogeneous population and strong perception of community relevance which resulted in the high response rates of Archie Cochrane’s Rhondda studies. In response, fungibility is an idea waiting to happen. It is the practical expression of Kessler’s idea of using entire communities as epidemiologic laboratories and a development of the Nordic register-based cohorts. Now that we have the technology to make it happen, it is time that we think more clearly about what can be achieved and how.

Funding

Funding to pay the Open Access publication charges for this article was provided by the Wales Office for Research and Development.

Conflict of interest: The author is a member of the UK Biobank Steering Group.

References


Received 20 April 2007, accepted 23 July 2007