Available online at www.sciencedirect.com

Public Health

journal homepage: www.elsevier.com/puhe

Symposium

Biobanking for epidemiology☆

G.A. Zielhuis

Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Accepted 9 December 2011
Available online 10 February 2012

Keywords:
Cohorts
Biorepositories
Biobanks
Epidemiology
Research infrastructure
Personalized medicine

SUMMARY

Biobanks are a key resource in unravelling the association between genetic background, lifestyle and environmental determinants of the incidence, natural course and treatment response for various complex diseases and health traits. Biobanks are goldmines for epidemiological research, provided that they are set up properly, enable multicentre collaboration, and are available for use by all serious epidemiology groups.

Large cohorts and biorepositories form the necessary infrastructure for high-quality epidemiological research. Biobanking – defined as the collection, storage, processing and distribution of human material and associated phenotypic data – is a hot topic in epidemiology at the current time. Biobanks are a key resource in unravelling the association between genetic background, lifestyle and environmental determinants of the incidence, natural course and treatment response for various complex diseases and health traits.

As these types of studies generally lack statistical power, there is an urgent need for collaboration between research groups, and combined efforts to accomplish high-quality large size biobanks as infrastructure for future studies. Several recent initiatives, both at national and international level, have emerged throughout Europe and worldwide.

One example is the Biobanking and Biomedical Resources Infrastructure (BBMRI),[1] a network of 250 established biobanks throughout Europe, which recently became a legal entity (BBMRI-ERIC). BBMRI aims to improve biobanking accessibility and interoperability by harmonizing similar biobanks in different locations, and enriching the genotypic and phenotypic informational content, and to archive using a more even ethical framework. National chapters of BBMRI have emerged in many European countries, with the similar motive to increase the potential for collaborative studies in molecular epidemiology. Funded by national governments, it appears to be possible to create large enriched databases open for public use by epidemiologists with good research plans. The example of BBMRI-NL was given, where infrastructure is being developed through large multicentre collaborations (e.g. the 'Genome of the Netherlands' sequencing the whole genome of 250 country-wide unselected trios (1000 genomes) to increase national rare variant coverage). This will greatly improve resolution in all Genome Wide Association Studies (GWAS) with Dutch samples

☆ Summary report of a minisymposium held on 9 August 2011 at the IEA World Congress of Epidemiology, Edinburgh, UK.

E-mail address: g.zielhuis@ebh.umcn.nl.

0033-3506/$ – see front matter © 2012 The Royal Society for Public Health. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.puhe.2011.12.007

($n \sim 150{,}000$) (*Prof. Gertjan van Ommen, University of Leiden and Director of BBMRI-NL*).

Two types of human biobanks emerge: population biobanks and clinical biobanks. Large-scale prospective studies have several advantages for comprehensive and reliable quantification of the combined effects of lifestyle, environment, genotype and other determinants of disease. In particular, they allow the study of a large range of conditions, avoid recall bias as exposures can be assessed prior to disease development, and allow investigation of factors that may be affected by disease processes and consequent treatment. Prospective studies are also able to assess those conditions that cannot readily be investigated retrospectively, and can include all cases with high fatality rates. The inclusion of 500,000 residents aged 40–69 years in the UK biobank will enable stable estimates of disease risk associations for a whole range of potential risk (and confounding) factors. With enormous effort and support at the national level, it has been shown to be possible to establish a large blood-based population biobank of well-phenotyped individuals who will be followed over the long term to identify a wide range of different health outcomes which will be carefully characterized. With an emphasis on efficient, high-quality procedures, tested in a number of pilot studies,[2] the UK biobank has set the standard for modern population-based biobanking (*Prof. Rory Collins, University of Oxford and Chief Executive of the UK biobank*).

In contrast, clinical biobanks focus on specific disease categories, enabling efficient case–control studies (e.g. for investigating gene–environment interactions) as well as prognostic studies (e.g. for investigating gene–treatment interactions). Again, the scale and standardization of data collection is crucial, suggesting the establishment of prospective collaborative biobanks for patients with specific disease entities.

The eight Dutch university hospitals joined forces to create a national research infrastructure for clinical epidemiology. Initially, this initiative ('String of Pearls'[3]) comprised eight national disease cohorts with highly standardized collection of biomaterial and clinical data, available in one national database. Other patient cohorts will be added. With clinicians and clinical epidemiologists in a leading role, much energy has been spent on standardizing procedures and collection of high-quality material in the setting of routine clinical care. Ethico-legal and quality control procedures were set.

With enormous effort and support at the national level, it has been shown to be possible to establish a large blood-based clinical collaborative biobank of well-phenotyped patients who will be followed on a regular basis for natural course, health outcomes, morbidity and mortality. Data and material will be made available for every investigator with a good research plan (*Prof. Jacqueline Dekker, VU University Medical Centre Amsterdam and Co-ordinator of the Diabetes cohort in the String of Pearls Initiative*).

The technical challenges of connecting the research infrastructures are increasingly being met by new information technology solutions. Harmonization tools, such as developed in the Public Population Project in Genomics (P3G),[4] enhance the capacity to synthesize data with enormous potential for epidemiological research. The issue of 'access' is central to this. Researchers who target a set of interesting material and data must still overcome a series of ethical, legal and other conditions to access these. A more 'open access' culture for biobank data requires that the biobankers properly manage the massive amount of potentially sensitive information in a way that will be of optimal use for the scientific community, while also ensuring proper protection and respect for the privacy and confidentiality of the donors. Different rules and regulations between countries and centres may create other hurdles to international collaborative use of biobank material and data. P3G has developed a catalogue of these international differences in ethical, legal and social issues, with potential innovative solutions being proposed to meet some of these challenges (*Dr. Mylene Deschenes, P3G Montreal*).

One of these innovations is DataSHIELD, an approach to analyzing pooled data that circumvents the fundamental conflict between two competing public goods: the need for combining individual-level data from different sources in a single large dataset in one hand, and valid ethico-legal constraints prohibiting sharing individual data in the other. DataSHIELD (Data Aggregation Through Anonymous Summary-statistics from Harmonized Individual-levEL Databased)[5] uses modern distributed computing, and takes advantage of the properties of an algorithm that iterately updates parameter estimates in general linear modelling. Results have been shown to be identical with traditional analyses, and pilot studies have started to learn more about its opportunities and challenges in actual epidemiological research (*Prof. Paul Burton, University of Leicester*).

The use of biobanks for epidemiological research does not only depend on the availability and quality of the biomaterials, but also on the collection of associated clinical and personal characteristics. For longitudinal designs to study the course of disease, the requirements for obtaining time-specific phenotypic data are even stronger. The development of precisely defined clinical data elements (CDEs) may help to ensure that clinical relevant data are collected at each time interval. Experiences to date (e.g. with the biobanking models of the Coriell Institute for Medical Research) have shown that it is no trivial task to define a CDE and correlate its potential relevance for epidemiological studies. Once the definitions are set, much emphasis must be placed on maintaining standardization across sites and over time. Information technology solutions may serve these processes.

In addition to serving epidemiological studies, high-quality biobanks may also serve initiatives for personalized medicine. Data that are medically relevant for the individual donor (i.e. informative according to medical experts about increased, amendable individual risk) may be reported back to that individual. The Coriell Personalized Medicine Collaborative Model has demonstrated its potential for this purpose (*Joe Mintzer, Coriell Institute for Medical Research, Camden, NJ*).

In conclusion, biobanks are goldmines for epidemiological research, provided that they are set up properly, enable multicentre collaboration, and are available for use by all serious epidemiology groups.

Acknowledgements

Ethical approval

None.

Funding

BBMRI-NL (www.bbmri.nl), UK Biobank (www.ukbiobank.ac.uk), String of Pearls Initiative (www.string-of-pearls.org), P3G (www.p3g.org) and Coriell Institute for Medical Research (www.coriell.org).

Competing interest

None.

REFERENCES

1. Biobanking and Biomedical Resources Infrastructure. Available at http://www.bbmri.eu [accessed 07.11.11].
2. Elliott P, Peakman TC. The UK Biobank sample: handling and storage validation studies. *Int J Epidemiol* 2008;**37**(Suppl. 1):1–64.
3. The String of Pearls Initiative. Available at: http://www.string-of-pearls.org [accessed 07.11.11].
4. Public Population Project in Genomics (P3G). Available at: http://www.p3g.org [accessed 07.11.11].
5. Wolfson M, Wallace SE, Masca N, Rowe G, Sheehan NA, Ferretti V, et al. DataSHIELD: resolving a conflict in contemporary bioscience – performing a pooled analysis of individual-level data without sharing the data. *Int J Epidemiol* 2010;**39**:1372–82.