

LETTER

# Exposing participants? Population biobanks go geo

*European Journal of Human Genetics* (2016) **24**, 155–156;  
doi:10.1038/ejhg.2015.43; published online 25 March 2015

## POPULATION BIOBANKS MEET GIS

Population biobanks are designed to aid studies into the causes of health and disease by linking nature (genotype), nurture (phenotype) and exposure (environment). To that end, they collect data about the phenotype, genotype, lifestyle and family medical history from large numbers of participants, with longitudinal follow-up.<sup>1</sup> To link the individual phenotypes and genotypes to their past, present and future exposures to their environments, biobanks are increasingly incorporating multiple environmental variables such as air and noise pollution exposure data from traffic-related sources, electromagnetic field exposures data from mobile phone masts or overhead powerlines and exposures related to industrial point sources. A few of these environmental data are available from public sources but many are estimated for specific research projects. Such data can be linked to recorded place of residence, school or workplace through the use of Geographical Information Systems (GIS).

GIS have been defined as computer systems designed to capture, store, process and display spatial or geographical data helping to understand patterns and relationships. GIS was used to link population to exposure data within the recent ESCAPE study where air pollution exposures were combined with population data from over 30 cohorts/biobanks<sup>2</sup> and investigating the effects on a range of health end points.<sup>3,4</sup> Currently the BioSHaRE project is investigating the link between air pollution and noise exposures and health effects within UK Biobank, the Norwegian Hunt Study, the Dutch population biobank Lifelines and EPIC. Within these studies, environmental data were assigned to place of residence, which potentially could help identify the local health hazards of the biobank participants concerned. In addition, in some studies such as LifeLines, occupational addresses of participants are also being collected and geocoded. Address history available from the Municipal Personal Record Database provides insights in residential mobility and length of exposure to different environments, enabling cradle to grave exposure assessment.

## EXPOSING PARTICIPANTS BY PUBLISHING THEIR EXPOSURE

As science dictates, any associations observed should be published, for peer review and verification. As an upside, publication of local health hazards inform public health policies, and perhaps help advance personal geo-medicine.<sup>5</sup> As a downside, these publications might expose the pertinent biobank participants – and non-participants living in the same area – to social and economic risks. If published at a high spatial resolution, using modern day visualization techniques, the findings would enable the rest of the world to, literally, zoom in and out of their local health hazards: simply imagine Google Street View enriched with scientifically published exposure studies. More specifically, published biobank exposure studies could potentially provide

knowledge or information that could be used to deny or condition access of both their participants and non-participant neighbors to private or social services or impact on the value of their private property. Just as the publication of two avian (H5N1) influenza studies recently re-ignited concerns over misuse that could undermine 'bio-security',<sup>6</sup> the publication of biobank geo-exposure studies and the underlying health data, create concern over possible misuse thereby undermining 'social, private and economic security'.

## RISKS REAL OR PERCEIVED?

Presently, it is unknown whether these risks are perceived or real. And even if they are found to be real, these risks should neither override the requirements of science, that findings be published, nor the interests of public health, that epidemiological outbreaks be reported. However, the situation we describe is somewhat different from public health studies of epidemiological outbreaks. Outbreaks of infectious disease are (certainly in the UK and in other European countries) covered by specific legislation that protects public health in which the rights of the individual may be outweighed by needs to protect the health of the population as a whole. Moreover, if registry data (eg, cancer registration) are being explored by public health professionals to consider, for example, reporting clusters of a particular cancer, there will be conditions of use in place that protect inadvertent disclosure through either small numbers in a cell of a table or by mapping. Also, access to the data in both infectious disease outbreak and in clusters of cancers/chronic disease will usually be restricted to approved individuals usually with some health care remit (eg, public health doctors, infection control nurses, etc), that is, who have a 'duty of care' and who are familiar with dealing with identifiable (confidential) patient data.

The situation we describe is also different from that of government agencies who make their spatial data available to the general public.<sup>7</sup> It is generally presumed that exposure data that is publicly available, could be used by anyone with internet access to link exposure to location. But while this is true for some exposures, our own experience suggests that many publicly available environmental data are at a rather coarse spatial resolution, or only available in certain locations or have other limitations that can hinder estimation of community- and individual-level exposures.<sup>8</sup> In contrast, biobank geo-exposure studies actually link and combine genotype, phenotype and exposure data to establish relationships and associations. And the exposure data are collected over the participant's lifespan, from before birth (for three generation studies) until death. Publication of these findings could expose both participants and non-participants to the risk that the public and/or the private sector abuse the refined biobank data as scientifically validated estimates of exposures in the local community concerned.

## ASSESSING THE RISKS

As population biobanks have a legal obligation to inform their participants about all risks of their participation as well as to promote responsible publishing, we recommend that they identify, assess and mitigate the risks posed by the performance and/or the publication of GIS-enabled exposure studies to the legitimate interests of their participants and the public. Although we did not perform an exhaustive survey, it seems that policies to deal with these issues vary considerably. In the ESCAPE study, for example, some cohorts have very strict policies where no outputs with mapping below region level

are permitted and geocodes should be separated from health data, whereas some other cohorts have no restrictions at all.

To that end, we submit that population biobanks and/or their supervising review boards subject the collection and/or publication of biobank geo-exposure data and studies to a social and economic human rights impact assessment to:

- Analyze whether geo-referencing of address data of biobank participants, (even when encrypted with restricted researcher access), is covered by their consent or by IRB approval.
- Examine whether statutory protections against misapplication of (personal) health and genetic data extend to biobank exposure findings and data.
- Determine whether the risks described above are real and, if so, to what extent these studies indeed qualify as 'dual use research of concern', that is, research conducted for legitimate purposes that can be utilized for benevolent or harmful purposes.<sup>6</sup>
- Undertake a risk–benefit assessment, to balance the divergent arguments in favor of (unfettered) publication – the potential to damage the social and economic human rights and interests of both biobank participants and non-participants and against publication – progress of science and promotion of public health.
- If the arguments in favor of publication prevail, does the published matter need to be sensitive to non-identifiability concerns (eg, publish an overall dose–response association but not maps of rates for areas with small numbers of individuals).

As an additional benefit, such an impact assessment would help cultivate a sense of responsibility for potential misapplications of population biobank GIS publications among population biobankers and scientists.

At present, while individual scientists might consider the social and economic human rights implications of their publications as a matter of responsible reporting, there seems to be no systematic approach for a thorough assessment of impact, at least, not for population biobanks. This gap should be filled, either by a set of 'points to consider' as part of the review of research protocol and publication, a Code of Conduct or an institutional oversight model, as appropriate. For the latter, the US Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern to Bio Security could serve as an example.<sup>9</sup>

## CONFLICT OF INTEREST

The authors declare no conflict of interests.

## ACKNOWLEDGEMENTS

The research leading to this paper has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 261433: BioSHaRE and from the Canada Research Chair in Law and Medicine.

Jasper Adriaan Bovenberg<sup>\*,1</sup>, Bartha Maria Knoppers<sup>2</sup>,  
Anna Hansell<sup>3</sup> and Kees de Hoogh<sup>4</sup>

<sup>1</sup>*Institute for Bio-Law, Aerdenhout, The Netherlands;*

<sup>2</sup>*Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, QC, Canada;*

<sup>3</sup>*Small Area Health Statistics Unit (SAHSU), MRC-PHE Centre for Environment and Health, Public Health and Primary Care, Imperial College Healthcare NHS Trust, London, UK;*

<sup>4</sup>*Environmental Exposure and Health Unit, Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland*

*E-mail: jabovenberg@xs4all.nl*

1 Wichmann HE, Kuhn KA, Waldenberger M *et al*: Comprehensive catalogue of european biobanks. *Nat Biotechnol* 2011; **29**: 795–797.

2 Beelen R, Hoek G, Vienneau D *et al*: Development of NO<sub>2</sub> and NO<sub>x</sub> land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project. *Atmos Environ* 2013; **72**: 10–23.

3 Cai Y, Schikowski T, Adam M *et al*: Cross-sectional associations between air pollution and chronic bronchitis: an ESCAPE meta-analysis across five cohorts. *Thorax* 2014; **69**: 1005–1014.

4 Adam M, Schikowski T, Carsin AE *et al*: Adult lung function and long-term air pollution exposure. ESCAPE: a multicentre cohort study and meta-analysis. *Eur Respir J* 2015; **45**: 38–50.

5 Your\_health\_depends\_on\_where\_you\_live. Available at [http://www.ted.com/talks/bill\\_davenhall](http://www.ted.com/talks/bill_davenhall).

6 Wolinetz CD: Implementing the new U.S. dual-use policy. *Science* 2012; **336**: 1525–1527.

7 What's in Your Backyard? Available at <http://apps.environment-agency.gov.uk/wiyby/default.aspx>.

8 Hansell AL, Beale LA: *The Environment and Health Atlas for England and Wales*. Oxford, UK: Oxford University Press. 2014, ISBN 978-0-19-870694-6.

9 United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern. Available at: <http://www.phe.gov/s3/dualuse/Documents/oversight-durc.pdf>.