

This is a repository copy of *How do biomarkers and genetics contribute to understanding society?*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/105377/>

Version: Accepted Version

Article:

Benzeval, Michaela, Kumari, Meena and Jones, Andrew Michael
orcid.org/0000-0003-4114-1785 (2016) How do biomarkers and genetics contribute to understanding society? *Health Economics*. pp. 1219-1222. ISSN 1057-9230

<https://doi.org/10.1002/hec.3400>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

How do biomarkers and genetics contribute to Understanding Society?

Michaela Benzeval

Institute for Social and Economic Research, University of Essex

Meena Kumari

Institute for Social and Economic Research, University of Essex
Department of Epidemiology and Public Health, University College London

Andrew M Jones

Department of Economics and Related Studies, University of York
Centre for Health Economics, Monash University
Department of Economics, University of Bergen

Understanding the interaction between social and economic circumstances and health over the life course is important to develop policies not only to improve people's health but also their social and economic capacities. While much research has been produced across different disciplines to examine these issues, they are often limited by their disciplinary base: social science research often treats health as a unitary concept, while biomedical studies generally control for a single measure of socioeconomic status (Herd et al, 2007). However, what is needed are studies that bring together the richness of social science and health data, analysed by researchers with expertise across biomedical and social disciplines, to understand the complex social and biological processes that link different aspects of people's lives and health at different life stages. Including biomarkers in established high quality social science longitudinal surveys is one way to do this. Biomarkers are biological or physiological measures that indicate the presence of a disease or the propensity to develop a disease. They can be used to identify risk factors over the lifecourse and as objective measures of health that avoid contamination by reporting bias (see, e.g., Adda and Cornaglia, 2006; Banks et al., 2006). Such evidence will help to inform policies about how and when to intervene to improve population health and reduce health inequalities.

Longitudinal research shows the importance of biological and psychological development (which can be measured by biomarkers) at each stage of life, beginning *in utero*, for health well-being, behaviours, social and economic status at later ages. Research has attempted to address a wide range of questions; for example, are life course processes a matter of the accumulation of risk factors or do factors at each

stage matter independently or interact in different ways? Or are there 'critical periods' during which certain types of development must take place if they are ever to happen? Do experiences at some points in the life course influence the response to adversities that may happen later? How do early life and socioeconomic circumstances combine to influence adult outcomes? These questions are complex given that societal institutions such as the family, the education system, the structures and conditions of employment, fiscal and welfare policies may produce a variety of life courses regardless of individual biological or psychological differences. These institutions, which have long served to define important life transitions, are in constant change themselves.

Few studies have sufficient environmental, institutional, fiscal, psychological and biological data to attempt to address these questions, one that has developed such a resource is *Understanding Society: the UK Household Longitudinal Study*(UKHLS)

Understanding Society: the UK Household Longitudinal Survey (UKHLS)

UKHLS is one of the largest household panel studies in the world. It is funded by the Economic and Social Research Council with co-funding from a number of Government Departments, and led by the Institute of Social and Economic Research at the University of Essex. Begun in 2009, it has four samples: two of which have received a nurse interview, which collected biomarker data: the general population sample (GPS); and households from the British Household Panel Survey (BHPS), which has been running since 1991.

UKHLS, and BHPS before it, conducts annual interviews covering all members of the household, although in UKHLS fieldwork for each wave takes two calendar years. Information on children under 10 is gathered from the primary carer (generally mothers), those aged 10-15 are given a self-complete questionnaire, while those aged 16 and over complete a face-to-face interview. The questionnaire covers a wide range of questions about people's families, employment and education, housing and neighbourhoods, consumer durables, savings, wealth and income, health, health behaviours and wellbeing, cognition and personality, social support and engagement, health, transport, leisure, environmental and political behaviours. UKHLS also provides researchers with the ability to link to a wide variety of different geographies so that they can add different contextual material, as well as asking consent to link to government administrative records.

All people aged 16+ years who took part in the relevant main wave(wave 2 for GPS and wave 3 for BHPS) were selected for follow up, provide they had conducted their main interview in English and lived in Great Britain. In the second year of wave 2, in order to accommodate the BHPS sample interviews, a sampling fraction of 0.81 was applied to the GPS participants. In total therefore, across the two samples 35,563 respondents were eligible for the nurse interview, and 20,644 (58%) took part. Of these, approximately 67% have consent to give blood and provide DNA, and samples

were successfully taken on 13,571. Longitudinal weights have been produced to ensure the main nurse measures, and blood related data are representative of the general population. As a result, comprehensive longitudinal socioeconomic data have been linked to biomarkers that include physical measurements and blood analytes (Benzeval *et al.*, 2014).

The interviews, which lasted an hour, took place in respondents' homes approximately 5 months after the main survey. A range of physical measures were taken: including blood pressure, lung function, waist circumference, height, weight and body fat and grip strength. Blood samples were taken and have subsequently been analysed by Newcastle NHS Trust to produce 20 analytes, which capture key chronic diseases, and processes relating to stress and ageing. They include coronary heart disease (blood pressure, body fat, cholesterol and triglycerides); diabetes (HbA1c); liver disease (LFTs); kidney function (creatinine, urea); anaemia and poor nutrition (Hb, ferritin); inflammatory markers (CRP, fibrinogen, CMV); and hormones (testosterone, IGF-1, DHEAs). DNA was extracted and genotyping performed by the Sanger Institute using the Illumina HumanCoreExome chip for approximately 10,500 respondents of white European descent. These measured markers have been employed with the UK10k reference dataset to impute more than 10million common sequence variants. Genetics data alone is available at the EGA and combined with survey data at <https://www.understandingsociety.ac.uk/about/health/data>.

Research potential of biomarker and longitudinal social data

The availability of objectively measured biomarker data has two key advantages over self-report data. First, they provide measures of health that are free of reporting bias; given the intense debate around the extent of socioeconomic-related reporting bias (e.g., Bago d'Uva *et al.*, 2008) this is a significant advantage in research on socioeconomic inequalities in health. For example, biomarker data have been used to analyse the effect of socioeconomic position on the conditional mean of biomarker scores (Banks *et al.*, 2006; Juerges *et al.*, 2013; Muennig *et al.*, 2007; Ploubidis *et al.*, 2014; Dowd and Goldman, 2006). Secondly, biomarkers allow researchers to investigate biological factors that contribute to and interact with health, education, and social conditions. When combined with the longitudinal data, biomarkers can shed light on the complex interplay between biology, behaviour, and environment over the life course for both health and other outcomes.

Data suggest two pressing issues: the ageing population, with its attendant increase in disability and disease, and an increase in psychological distress in recent generations. Indeed depression is currently rated as the second most common cause of disability by WHO. There is a near ubiquitous finding of differences in health and functioning by social position. However, the reasons for these differences are still unclear. The shadow of adversity in childhood and its impact on mental health in later life is well established. However, the specific causal chains and life-course trajectories remain to be elucidated and biological pathways that may mediate these relationships are less

well established. A number of pathways have been implicated, these include the immune system, endocrine function, which is specifically related to gender differences in mental health, and cardiovascular risk, which is implicated in older age depression.

Methodological challenges and opportunities

The appropriate treatment of biomarker and genetic data in statistical analyses plays a vital role in the production of high-quality scientific findings. The biostatistical literature on using such data is more developed than that in social statistics and econometrics, but the basic techniques are very similar at heart, albeit divided by differences in emphasis and terminology. In health economics there is a wide literature on approaches to modelling patient-level cost data and their relative performance (for example Jones *et al.*, 2015a). Much research to date has focused on the prediction of average costs for different types of patient. This can potentially miss important features such as the likelihood of extreme values. For example, Jones *et al.* (2015b) compare methods for fitting the distribution of healthcare costs that go 'beyond the mean'. The distributions of biomarkers share many of the features of health care costs and these methods may be applied to find the model that best fits each outcome. For example, Carrieri & Jones (2016) we have explored the impact of income across the full distribution of the biomarkers measured in the Health Survey for England.

Causality is difficult to ascertain using conventional observational methodologies, given well-known problems of confounding and reverse causality. While longitudinal studies provide a partial solution, but are often limited by poor temporal resolution, unmeasured confounding and limited statistical power. One potential solution to these difficulties lies in the application of the principle of Mendelian randomisation (MR). MR involves the use of specific genetic variants that have been shown to be robustly associated with a specific exposure (e.g., diabetes); these are then used as proxies for the exposure. The principle of MR relies on the basic (but approximate) laws of Mendelian genetics (segregation and independent assortment). If these two laws hold then, at a population level, genetic variants will not be associated with the confounding factors that generally distort conventional observational studies. Whilst it is not possible to fully prove this principle, it has been demonstrated that associations of genetic variants with measured confounders occur at rates consistent with chance, but measured exposures associate with confounding factors much more frequently. In addition, the genes that individuals are born with should not be altered by environmental factors, which removes the issue of reverse causality. Often there are a large number of genetic loci relevant to a particular condition (such as obesity, metabolic markers and diabetes). Research is therefore beginning to employ multi-SNP genetic instrument to increase both the power of the genetic instrument and its specificity. Given the relatively small effects that individual genetic variants exert on exposures, MR analyses generally require large sample sizes, which can only be achieved through collaborative consortia.

Conclusion

Research that combines longitudinal socioeconomic data with biomarkers and genetic markers, using appropriate methodologies, has the potential to add to our understanding of how social and economic circumstances interact with health over the life course, and some of the biological pathways that underpin this. These new data in UKHLS enable researchers to do this, and we are developing both our own research activities as well as capacity building events to support and develop such new research agendas. Such research by us and more broadly will shed light on the complexities of associations between the social and biological for example when inequalities in different dimensions of health emerge and what some of the underlying pathways may be. This will help identify interventions and mechanisms that may help to promote population health and reduce inequalities in health.

References

- Adda J, Cornaglia A. 2006. Taxes, cigarette consumption and smoking intensity. *American Economic Review* **96**: 1013-28.
- Bago d’Uva T, Van Doorslaer E, Lindeboom M, O’Donnell O. 2008. Does reporting heterogeneity bias the measurement of health disparities? *Health Economics* **17**: 351-375.
- Banks J, Marmot M, Oldfield Z, Smith JP. 2006. Disease and disadvantage in the United States and in England. *Journal of American Medical Association* **295**: 2037-45.
- Benzeval M, Davillas A, Kumari M, Lynn P. 2014. Understanding Society: The UK Household Longitudinal Survey biomarker user guide and glossary. Institute for Social and Economic Research, University of Essex.
- Carrieri V, Jones, AM. 2016, The income-health relationship ‘beyond the mean’: new evidence from biomarkers. *Health Economics*, in press.
- Dowd JB, Goldman N. 2006. Do biomarkers of stress mediate the relation between socioeconomic status and health? *Journal of Epidemiology and Community Health* **60**: 633-639.
- Herd, P., Goesling, B. and House, J.S. (2007) 'Socioeconomic Position and Health: The Differential Effects of Education versus Income on the Onset versus Progression of Health Problems' *Journal of Health and Social Behavior* **48**:3: 223-38.
- Jones AM, Lomas J, Moore P, Rice N. 2015a. A quasi-Monte Carlo comparison of recent developments in parametric and semi-parametric regression methods for heavy tailed and non-normal data: with an application to health care costs. *Journal of the Royal Statistical Society Series A*, DOI: 10.1111/rssa.12141.

Jones AM, Lomas J, Rice N. 2015b. Healthcare cost regressions: going beyond the mean to estimate the full distribution. *Health Economics* **24**: 1192-1212.

Juerges H, Kruk E, Reinhold S. 2013. The effect of compulsory schooling on health-evidence from biomarkers. *Journal of Population Economics* **26**: 645-672.

Muennig P, Sohler N, Mahato B. 2007. Socioeconomic status as an independent predictor of physiological biomarkers of cardiovascular disease: evidence from NHANES. *Preventive Medicine* **45**: 35-40.

Ploubidis GB, Benova L, Grundy E, Laydon D, De Stavola B. 2014. Lifelong socioeconomic position and biomarkers of later life health: testing the contribution of competing hypotheses. *Social Science and Medicine* **119**: 258-265.