POLICY

UK Biobank Data: Come and Get It

BIO MEDICAL SCIENCE NOW FACES THE DAUNTING CHALLENGE OF DECIPHERING THE genetic and environmental determinants that contribute to common life-threatening and disabling diseases (1). To meet this challenge, the UK Medical Research Council (MRC) and Wellcome Trust established UK Biobank, a population-based prospective cohort of 500,000 men and women in the UK. Extensive and reliable measurement of a wide range of exposures together with rigorous follow-up of health outcomes in these individuals will allow detailed investigation into the combined effects of genetic and environmental determinants for a wide range of diseases of middle and old age (2). In March 2014, UK Biobank will begin a multimodal imaging study of the brain, heart, and abdomen, together with an ultrasound of the carotid arteries, and whole-body dual-energy x-ray absorptiometry (DXA) of the bones and joints in 100,000 participants. Here, we describe the UK Biobank and encourage biomedical and health researchers to make use of the data in this open-access resource.

The prospective nature of the UK Biobank study is important because the effects of risk factors can be assessed before a disease or its treatment affects a participant. The study design also allows for a wide range of conditions to be investigated, including those that are difficult if not impossible to study retrospectively (for example, dementia and rapidly fatal conditions such as pancreatic or lung cancer). Moreover, both the beneficial and adverse effects of a specific factor on the risk of developing disease can be considered simultaneously, and this information can then be used to provide evidence-based public health guidance for primary prevention. Prospective studies must, however, involve large numbers of participants because only a relatively small proportion of the cohort will develop any particular condition, and the effect of any one risk factor on overall disease risk is likely to be small.

HUMAN RESOURCES

UK Biobank recruited 500,000 participants aged 40 to 69 years who were registered with the National Health Service (NHS) and lived within reasonable traveling distance of a total of 22 assessment centers across the UK between 2007 and 2010. Volunteers gave informed consent and completed questionnaires on their lifestyle, family, and medical history; underwent assessments of cognitive function and occupational history. Tri-axial accelerometers are also being mailed to a large subset of the cohort in order to obtain objective measures of physical activity (such as intensity and duration), which can be used to supplement the baseline questionnaire data. These devices can also provide more reliable assessment of other aspects of normal daily living, such as sleep patterns.

To further enhance the value of the UK Biobank resource, we will measure, in baseline samples from all 500,000 participants, a wide range of biochemical markers known to be relevant for specific diseases (for example, lipids for cardiovascular disease) or of high diagnostic value (for example, hemoglobin A1c for diabetes) or that characterize phenotypes not otherwise well assessed (for example, measures of liver and kidney function). The genotypes of the

Citation:
10.1126/scitranslmed.3008601
entire cohort are being determined by using a bespoke genome-wide DNA microarray that contains ~850,000 genetic variants; these have been selected for the analysis of rare, intermediate, and common genetic variants (including a panel for use in imputation) that include single-nucleotide polymorphisms, insertions or deletions (indels), and gene copy-number variants. On the basis of a high-throughput approach, we anticipate that both the biochemical and genotyping data for the full cohort will be available for use by researchers in 2015.

A CLOSER LOOK
UK Biobank also has ambitious plans for a multimodal imaging study of unprecedented size, to include magnetic resonance imaging of the brain, heart, and abdomen; ultrasound of the carotid arteries; and whole-body DXA of the bones and joints in 100,000 participants. This pilot study, which will begin in March 2014 at UK Biobank’s Co-ordinating Centre, aims to demonstrate the feasibility of the project and to evaluate the approach for dealing with incidental findings. If the pilot is successful and the remaining funding is released, dedicated imaging centers will be established at other sites across the UK so as to complete the imaging study over the following 5 years. Not only will these imaging measures provide quantitatively defined phenotypes that are closely related to disease, they will enable the calibration of some of the less precise measurements made on all participants at baseline (for example, body fat by impedance and bone density by calcaneal ultrasound). The availability of multimodality imaging data on large numbers of participants, together with the vast amount of other data available in UK Biobank, will provide a singular resource for investigating the causes of and biological mechanisms that underlie disease. For example, researchers will be able to examine the extent to which dementia is related to imaging measures as well as lifestyle factors and biochemical and genetic markers.

The grand scale of this endeavor is a testament to the commitment of the funders and the study participants who are willing to help create a resource of global importance for the health of future generations.

The value of UK Biobank depends not only on its ability to obtain detailed exposure data but also on achieving detailed follow-up of the health of participants, which is made possible through linkage to routine data available from the NHS, including data on deaths, cancer registrations, hospital admissions, primary care, and a range of other specialized data sets. Supplementary information will also be sought directly from participants about conditions that are typically underreported in medical records (such as depression and cognitive decline). With linkage to health-outcome data now in place, opportunities exist for research on some of the more common incident conditions (for example, more than 2000 incident breast cancers and 900 deaths from ischemic heart disease have already been registered). Over the next few years, UK Biobank will become sufficiently mature to allow reliable investigation of an increasingly wide range of conditions. Considerable effort is now being directed toward characterizing and confirming health outcomes; this will be done by cross-referencing diagnoses from multiple sources of information, starting with lower-cost electronic sources and then incorporating data acquired by more resource-intensive methods for the confirmation of the initial diagnosis and further subclassification of disease.

UK Biobank is an open-access resource that encourages researchers from around the world—including those from the academic, nonprofit, public, and commercial sectors—to access the data and biological samples for any health-related research that is in the public interest (www.ukbiobank.ac.uk). Since its launch in April 2012, more than 900 researchers have registered to use the resource, and more than 100 research applications have been submitted. The online application process enables researchers to select data fields specific to their research proposal and is linked to automated systems that can retrieve biological samples, if required (5). Robust safeguards are in place to help ensure anonymity and confidentiality of participants’ data and samples. UK Biobank is a registered charitable company, and as such, researchers are required to pay only for access to the resource, so that UK Biobank can recover its costs. Research results from studies that use UK Biobank data should be incorporated back into the resource so that others can benefit from the findings.

UK Biobank has shown that it is possible to establish a population-based prospective health study on a large scale and to make the resulting resource available to investigators. Our long-term goal is to help scientists worldwide to conduct research that will lead to new strategies for the prevention, diagnosis, and treatment of a wide range of conditions.

– Naomi E. Allen, Cathie Sudlow, Tim Peakman, and Rory Collins, on behalf of UK Biobank
3. UK Biobank, Protocol for a large-scale prospective epidemiological resource; available at www.ukbiobank.ac.uk/resources.
5. UK Biobank access procedures: Application and review procedures for access to the UK Biobank resource; available at www.ukbiobank.ac.uk/resources.

ACKNOWLEDGMENTS: UK Biobank is funded by the MRC, Wellcome Trust, Department of Health, British Heart Foundation, Northwest Regional Development Agency, Scottish government, and Welsh Assembly government. The views expressed herein are those of the authors and not necessarily those of the UK NHS, National Institute for Health Research, or Department of Health.