Int. J. Epidemiol. Advance Access published April 17, 2006

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology
doi:10.1093/ije/dyl071

EDITORIAL

Make it HuGE: human genome epidemiology reviews, population health, and the *IJE*

George Davey Smith,[1]* Marta Gwinn,[2] Shah Ebrahim,[3] Lyle J Palmer[4] and Muin J Khoury[2]

The *International Journal of Epidemiology* is concerned with scientific evidence that can ultimately form the basis of strategies for improving population health. Hence, the *IJE* would be expected to remain cautious about the technological advances heralded by the sequencing of the human genome. The classical epidemiological approaches of examining secular trends in disease risk, changes in risk consequent upon migration, and differences in disease rates between populations indicate that little of the global burden of common disease can be attributed to simple differences in genetically determined risk. It is not surprising that many social epidemiologists and public health practitioners (including, in the past, some of the authors of this editorial) have pointed this out. More surprising, perhaps, is that in the spirit of honest accounting, some geneticists and genetic epidemiologists have also punctured the inflated claims of genetic epidemiology by emphasizing that the population-attributable risk of most common genetic variants will be low and that in any case the influence of genetic factors is not reversible through changing genetic make-up. Thus Terwilliger and Weiss[1] point out that alleles identified as increasing the risk of common diseases 'tend to be involved in only a small subset of all cases of such diseases' and that in any case 'while the concept of attributable risk is an important one for evaluating the impact of removable environmental factors, for non-removable genetic risk factors, it is a moot point'.

Bearing these criticisms in mind, should the *IJE* be excited by the growing Human Genome Epidemiology Network (HuGENet)?[2] This global collaboration launched by the Centers for Disease Control and Prevention (CDC) and many partners in 1998, aims to make sense of the implications of gene discovery for epidemiology and public health. The goals of HuGENet are to establish an information exchange network that promotes global collaboration in the development and dissemination of peer-reviewed epidemiological information on human genes; develop an updated and accessible knowledge base on the World Wide Web; and promote the use of this knowledge base for decision making involving the use of genetic information for population health.[2] Since 2001, HuGENet has maintained a searchable, online database of epidemiological studies of genetic variation and disease (www.cdc.gov/genomics/hugenet). By March 1, 2006, the HuGE Published Literature database contained 20 272 reports of research studies involving 2252 genes, 2406 health outcomes, and 743 interacting factors.[2] However, the implications of these articles for population health are unclear. To clarify basic issues, such as the population prevalence of genetic variants, the magnitude of disease risk associated with these variants (in relative and absolute terms), the contribution of these variants to the occurrence of disease in different populations (i.e. attributable risk), the existence of gene–environment and gene–gene interactions, the validity of genetic tests based on such variants in predicting disease risk, and the impact of genetic tests on morbidity, disability, and mortality in different populations, HuGE Net has promoted the completion of reviews (HuGE reviews),[3] which are peer-reviewed, systematic synopses of the epidemiological aspects of variation in particular genetic variants and health outcomes. By March 1, 2006 44 HuGE reviews have been published in various journals, with journal choice partially reflecting the focus of the review and editorial preferences. HuGE reviews are conducted with specific guidelines;[4] the key aspects of which are outlined in Box 1. In addition the first detailed guidance for conducting HuGE Reviews has been published in March 2006 (HuGENet handbook of systematic reviews, available at http://www.cdc.gov/genomics/hugenet/reviews/guidelines.htm).

It would also be worth considering the principles that govern another international group—*The Cochrane Collaboration*—that is attempting to deal with an almost infinite amount of information and assimilating it into systematic reviews of the effects of interventions (see http://www.cochrane.org/docs/descrip.htm). Of particular relevance to genetic epidemiology is the current duplication of effort that is apparent, with several groups conducting meta-analyses of the same associations. In a well-organized collaboration it may be possible to reduce this trend for the common good, allowing resources to be diverted to other areas. Practical issues that are worth considering are the mechanisms for updating of reviews, the standardization of search methods and reporting formats, and the identification and reduction of sources of bias in systematic reviews.

Why should the *IJE* be concerned with such a movement? We contend that, contrary to some of the outright dismissals of the public health importance of genetic epidemiology, the field can contribute not just to 'genomic medicine' (which currently we believe has been oversold with respect to common chronic diseases[5]) but also to strengthening causal inferences regarding environmentally modifiable causes of disease. Confirmed associations, which are not distorted by the usual problems of residual confounding and measurement error, between genetic

[1] Department of Social Medicine, University of Bristol, BS8 2PR, UK.
[2] Centers for Disease Control and Prevention, Atlanta, GA, USA.
[3] London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.
[4] Western Australia Institute for Medical Research, Perth, Australia.
* Corresponding author. E-mail: george.davey-smith@bristol.ac.uk

Box 1 Key aspects of HuGE reviews

What is a HuGE review?

Systematic review of epidemiological data on specific human genetic variants at one or more loci in relation to diseases or other health outcomes

What are the types of HuGE reviews?

1. Full review: includes prevalence of genetic variants, disease associations, interactions, implications for population testing and public health impact, knowledge gaps, and research recommendations
2. Gene–disease association review: as above without prevalence review
3. Mini review: review of different outcomes in relation to genetic variant reviewed previously
4. Prevalence review: review limited to prevalence of genetic variants in populations

What are the defining characteristics of a HuGE review?

Systematic, specifying methods used to capture all available data, preferably employing quantitative methods of synthesis (e.g. meta-analysis or pooled analysis of individual-level data)

variants of known function and disease outcomes provide strong evidence of a causal link between the biological processes that the genetic variants are related to and the disease outcomes. Such evidence implies that modification of biological processes by non-genetic means can reduce disease risk and strengthens our conviction that we are indeed focusing environmental intervention efforts on factors that are truly causal. Several existing HuGE reviews illustrate this point. For example, a HuGE review summarized evidence on the association between the *MTHFR* 677*C*→*T* polymorphism and risk of neural tube defects (NTDs).[6] The polymorphism is associated with increased blood levels of homocysteine and lower folate activity; in the HuGE meta-analysis of case–control studies it was shown that infants with the TT vs CC variants have a relative risk of 1.75 (95% CI 1.41–2.18).[6] However, when the parental genotypes were examined, the increased risk of NTD associated with the TT variant was only seen when the mother was the carrier. In this case, it appears that the intra-uterine environment—influenced by maternal TT genotype—rather than the genotype of offspring, increases the risk of NTD, as has been discussed previously in more detail.[7] Thus, this evidence from a systematic HuGE review of genetic association studies provides further evidence that folate levels among mothers influences NTD risk in their offspring. In this case, of course, we have additional evidence from randomized controlled trials that periconceptional maternal folate intake influences NTD risk.[8,9] In other situations, however, robust evidence from genetic associations may point to yet-to-be established causes of disease.

Consider, for example, the controversial idea that infections increase the risk of preterm birth. A HuGE review summarized the evidence on genetic variants related to inflammation[10] and found that polymorphisms that increase the magnitude or duration of the inflammatory response were associated with increased risk of preterm birth. Thus, the evidence from genetic association studies strengthens the inference that maternal infection increases the risk of preterm delivery. In both cases—maternal *MTHFR*, and maternal inflammation-related genetic variants—the implication is not that screening for genetic variants is the key to prevention but that public health initiatives to increase periconceptional maternal folate and to prevent infections during pregnancy would produce health benefits.

Why utilize genetic association studies to strengthen inferences about environmental causes of disease? Is it not more straightforward to study these factors directly using conventional epidemiological techniques? In many cases the answer may be in the negative. As is increasingly evident, observational epidemiological studies have produced seriously misleading findings in areas such as vitamin E supplement use and coronary heart disease (CHD), beta carotene intake and lung cancer, or hormone replacement therapy and CHD.[11,12] In these cases randomized controlled trials (RCTs) demonstrated the fallacious nature of causal inferences drawn from observational studies; in other areas where RCTs have not or cannot be performed it is hardly credible that the success of observational epidemiology has been greater. The considerable advantage of studying genetic variants is that these variants are not generally susceptible to the confounding experienced by dietary patterns or other lifestyle-related risk factors.[13–18] Thus studying genetic variants that provide evidence on environmentally modifiable risk factors—as *MTHFR* does about folate intake and inflammatory variants do about infection in the examples discussed above—can, paradoxically, provide more evidence of the ability to modify disease risk through environmental change than can the direct study of the environmental factors of interest.[7,13]

This approach to utilizing findings from genetic association studies to understand modifiable, environmental causes of disease—a methodology that has been labelled 'Mendelian randomization'[7,19]—has been applied to the study of intermediate phenotypes influenced by genetic variants. Many biological markers have been shown to be related to disease, such as the classic case of blood cholesterol and CHD. In this instance RCTs of cholesterol lowering therapies have shown that the association is causal. In other cases, however, this is less clear. There could be confounding, or the disease could itself influence the circulating levels of the supposed exposure, i.e. reverse causation may be generating the apparent predictive value of the blood borne factor. If the association is causal, then genetic variants that influence the intermediate phenotype should be related to the disease to the degree predicted by the association between the variant and the intermediate phenotype. In the case of cholesterol and CHD, carriers of an apolipoprotein B variant (known as familial defective apo B), who have higher circulating cholesterol levels but are otherwise similar to non-carriers with respect to coronary risk factors, have an increased risk of CHD,[20] as would be anticipated by knowledge of the causal influence of circulating cholesterol. In other cases this is not seen—for example, those with genetically influenced higher fibrinogen level do not have an increased risk of CHD, despite fibrinogen being a predictor of CHD risk.[19,21] This suggests that the observational association between fibrinogen and CHD is non-causal and that lowering fibrinogen level will not, through this means alone, reduce CHD risk.

So, if genetic epidemiology findings are rigorously reviewed and interpreted to make inferences regarding environmentally modifiable risk of disease, does this make it more likely that

the discipline will contribute to the public health enterprise (and thus the goals of the *IJE*)? We feel it does. The overly hasty dismissal of the value of genetic epidemiology that we started with—on the grounds that the population attributable risk of genetic variants is low and that the genetic variants are not modifiable—is rendered moot by considering the potential contributions these genetic associations can make to knowledge of disease aetiology, treatment, and prevention. These contributions are highlighted in the new field of pharmacogenomics and by the growing clinical impact of 'genomic medicine'.[22] The degree to which associations between genetic variants and disease outcomes can demonstrate the importance of environmentally modifiable factors as causes of disease does not depend on, or even relate to, the population attributable risk of the genetic variants themselves. Consider, for example, the case of familial defective apo B. The genetic mutations associated with this condition will only account for a trivial percentage of cases of CHD within the population. However by identifying blood cholesterol levels as a causal factor for CHD the three-way associations between genotype, blood cholesterol, and CHD risk more reliably identify a clearly modifiable factor—circulating cholesterol levels—with a very high population attributable risk. A similar argument could be made with respect to maternal *MTHFR* and NTDs in offspring. Maternal genotype will account for only a small proportion of cases; however the association of maternal genotype and offspring NTDs identifies maternal folate intake as a modifiable influence on NTDs, with a high population attributable risk.

The study of genetic variants and disease can, therefore, add a perhaps surprising amount of information about environmentally modifiable causes of disease, and thus be of major population health importance. At the *IJE* we are very interested in receiving HuGE reviews that follow the HuGE guidelines[4] and explicitly address the implications of the findings of the genetic epidemiological studies for population health. We therefore favour reviews of variants that have functional connotations of relevance to strategies for modifying disease risk. We feel this can be done without abandoning the attractive objectivity of the systematic review, and here we would contrast a HuGE review of several alcohol-metabolism relevant genetic variants and head and neck cancer[23] with a later review that explicitly utilized the association of one particular variant with oesophageal cancer risk to strengthen the evidence base that alcohol intake increases the risk of this cancer.[24] Reviews formulated to provide answers to questions relating to modifiable causes of disease would appear most appropriate to the *IJE*. Thus we call for HuGE reviews to be sent to the *IJE* if they are systematic and attempt to utilize all the data for drawing inferences relevant to health improvement within populations. When appropriate we would like to receive two-stage reviews that include both genotype → intermediate phenotype and genotype → disease associations to formally make inferences about the causal association between intermediate phenotype and diseases, applying appropriate systematic review and statistical methods. HuGENet encourages registration of HuGE reviews with the HuGENet coordinator, and, at the *IJE*, we encourage the submission of protocols for reviews to the journal. Once approved, we guarantee publication of any review that coherently reports and discusses the findings in line with the protocol. The international readership, with an interest in population health, together with rapid turnaround and web-based publication on acceptance should make the *IJE* an ideal vehicle for such reports. Let the submissions begin.

References

[1] Terwilliger JD, Weiss WM. Confounding, ascertainment bias, and the blind quest for a genetic 'fountain of youth'. *Ann Med* 2003;**35:**532–44.

[2] Human Genome Epidemiology Network (HuGENet). Available at: http://www.cdc.gov/genomics/hugenet/default.htm (Accessed February 6, 2006).

[3] Little J, Khoury MJ, Bradley L *et al.* The human genome project is complete: how do we develop a handle for the pump? *Am J Epidemiol* 2003;**157:**667–73.

[4] HuGENet guidelines for HuGE reviews. Available at: http://www.cdc.gov/genomics/hugenet/reviews/guidelines2.htm (Accessed February 6, 2006).

[5] Davey Smith G, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PJ. Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* 2005;**366:**1484–98.

[6] Botto LD, Yang Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE Review. *Am J Epidemiol* 2000;**151:**862–77.

[7] Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32:**1–22.

[8] MRC Vitamin Study Research Group. Prevention of neural tube defects: Results of the Medical Research Council vitamin study. *Lancet* 1991;**338:**131–37.

[9] Czeizel AE, Dudás I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *New Engl J Med* 1992;**327:**1832–35.

[10] Crider KS, Whitehead N, Buus RM. Genetic variation associated with preterm birth: A HuGe review. *Genet Med* 2005;**7:**593–604.

[11] Davey Smith G, Ebrahim S. Data dredging, bias, or confounding. *BMJ* 2002;**325:**1437–38.

[12] Davey Smith G, Ebrahim G. Folate Supplementation and cardiovascular disease. *Lancet* 2005;**366:**1679–81.

[13] Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;**33:**30–42.

[14] Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Int J Epidemiol* 2004;**33:**9.

[15] Katan MB. Commentary: mendelian randomization, 18 years on. *Int J Epidemiol* 2004;**33:**10–11.

[16] Keavney B. Commentary: Katan's remarkable foresight: genes and causality 18 years on. *Int J Epidemiol* 2004;**33:**11–14.

[17] Tobin MD, Minelli C, Burton PR, Thompson JR. Commentary: Development of Mendelian randomization: from hypothesis test to 'Mendelian deconfounding'. *Int J Epidemiol* 2004;**33:**26–29.

[18] Davey Smith G, Lawlor D, Harbord R, Timpson N, Rumley A, Lowe G *et al.* Association of C-reactive protein with blood pressure and hypertension: lifecourse confounding and Mendelian randomisation tests of causality. *Arterioscler Thromb Vasc Biol* 2005;**25:**1051–56.

[19] Youngman LD, Keavney BD, Palmer A *et al.* Plasma fibrinogen and fibrinogen genotypes of 4685 cases of myocardial infarction and in 6002 controls: test of causality by 'Mendelian randomization'. *Circulation* 2000;**102:**31–32.

[20] Tybjærg-Hansen A, Steffenson R, Meinertz H, Schnohr P, Nordestgaard BG. Association of mutations in the apolipoprotein B gene with hypercholesterolemia and the risk of ischemic heart disease. *N Engl J Med* 1998;**338:**1577–84.

[21] Davey Smith G, Harbord R, Milton J, Ebrahim S, Sterne JAC. Does elevated plasma fibrinogen increase the risk of coronary heart disease?: evidence from a meta-analysis of genetic association studies. *Arterioscler Thromb Vasc Biol* 2005;**25:**2228–33.

[22] Willard HF, Angrist M, Ginsburg GS. Genomic medicine: genetic variation and its impact on the future of health care. *Philos Trans R Soc Lond B Biol Sci* 2005;**360:**1543–50.

[23] Brennan P, Lewis S, Hashibe M *et al.* Pooled analysis of alcohol dehydrogenase genotypes and head and neck cancer: a HuGE review. *Am J Epidemiol* 2004;**159:**1–16.

[24] Lewis S, Davey Smith G. Alcohol, ALDH2 and esophageal cancer: a meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. *Cancer Epidemiol Biomarkers Prev* 2005;**14:**1967–71.