Chinese women in Singapore. *Cancer Epidemiol Biomarkers Prev* 2001; **10:**1063–67.

[11] Spitz MR, Duphorne CM, Detry MA *et al.* Dietary intake of isothiocyanates: Evidence of a joint effect with glutathione S-transferase polymorphisms in lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 2000;**9:**1017–20.

[12] Lewis S, Brennan P, Nyberg F *et al.* Re: Spitz MR, Duphorne CM, Detry MA, Pillow PC, Amos CI, Lei L, de Andrade M, Gu X, Hong WK and Wu X. Dietary intake of isothiocyanates: Evidence of a joint effect with glutathione S-transferase polymorphisms in lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 2000;**9:**1017–20. *Cancer Epidemiol Biomarkers Prev* 2001;**10:**1105–06.

[13] International Agency for Research on Cancer. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Alcohol Drinking.* Lyon: IARC, 1988.

[14] 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* (in press).

[15] Bosron WF, Crabb DW, Li TK. Relationship between kinetics of liver alcohol dehydrogenase and alcohol metabolism. *Pharmacol Biochem Behav* 1983;**18(Suppl.1):**223–27.

[16] Coutelle C, Ward PJ, Fleury B *et al.* Laryngeal and oropharyngeal cancer, and alcohol dehydrogenase 3 and glutathione S-transferase M1 polymorphisms. *Hum Genet* 1997;**99:**319–25.

[17] Harty LC, Caporaso NE, Hayes RB *et al.* Alcohol dehydrogenase 3 genotype and risk of oral cavity and pharyngeal cancers. *J Natl Cancer Inst* 1997;**19:**1698–705.

[18] Olshan AF, Weissler MC, Watson MA, Bell DA. Risk of head and neck cancer and the alcohol dehydrogenase 3 genotype. *Carcinogenesis* 2001;**22:**57–61.

[19] Schwartz SM, Doody DR, Fitzgibbons ED, Ricks S, Porter PL, Chen C. Oral squamous cell cancer risk in relation to alcohol consumption and alcohol dehydrogenase-3 genotypes. *Cancer Epidemiol Biomarkers Prev* 2001;**10:**1137–44.

[20] Zavras AI, Wu T, Laskaris G *et al.* Interaction between a single nucleotide polymorphism in the alcohol dehydrogenase 3 gene, alcohol consumption and oral cancer risk. *Int J Cancer* 2002;**97:** 526–30.

[21] Sturgis EM, Dahlstrom KR, Guan Y *et al.* Alcohol dehydrogenase 3 genotype is not associated with risk of squamous cell carcinoma of the oral cavity and pharynx. *Cancer Epidemiol Biomarkers Prev* 2001; **10:**273–75.

[22] Bouchardy C, Hirvonen A, Coutelle C, Ward PJ, Dayer P, Benhamou S. Role of alcohol dehydrogenase 3 and cytochrome P-4502E1 genotypes in susceptibility to cancers of the upper aerodigestive tract. *Int J Cancer* 2000;**87:**734–40.

[23] Yokoyama A, Muramatsu T, Omori T *et al.* Alcohol and aldehyde dehydrogenase gene polymorphisms and oropharyngolaryngeal, esophageal and stomach cancers in Japanese alcoholics. *Carcinogenesis* 2001;**22:**433–39.

[24] Osier M, Pakstis AJ, Kidd JR *et al.* Linkage disequilibrium at the ADH2 and ADH3 loci and risk of alcoholism. *Am J Hum Genet* 1999;**64:** 1147–57.

[25] Melbye M, Trichopoulos D. Non-Hodgkin's lymphomas. In: Adami HO, Hunter D, Trichopoulos D (eds). *Textbook of Cancer Epidemiology*. New York: Oxford University Press, 2002, pp. 535–55.

[26] Spitz MR, Shi H, Yang F *et al.* Case-control study of the D2 dopamine receptor gene and smoking status in lung cancer patients. *J Natl Cancer Inst* 1998;**90:**358–63.

IJE vol.33 no.1 © International Epidemiological Association 2004; all rights reserved. *International Journal of Epidemiology* 2004;**33**:21–25
DOI: 10.1093/ije/dyh048

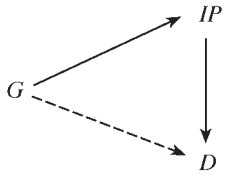
Commentary: The concept of 'Mendelian randomization'

Duncan C Thomas and David V Conti

This issue of the *International Journal of Epidemiology* reprints a seminal letter to the editor by Martijn Katan,[1] which appears to be the first description of the concept of 'Mendelian randomization.' In discussing the controversy over whether the association between low serum cholesterol and cancer is causal or might simply reflect an effect of the disease to lower cholesterol levels ('reverse causation') or confounding by diet or other factors, Katan proposed a test of causality by studying instead the relationship between cancer and a genetic determinant of serum cholesterol, the apolipoprotein A (*APOE*) gene. His rationale was that since alleles are allocated essentially at random, such an association would not be subject to either confounding or reverse causation. Thus, if a causal relationship between *APOE* and serum cholesterol were clearly established, then an association between *APOE* and cancer would provide indirect evidence for the causality of the association between serum cholesterol and cancer. Although Katan did not use the term 'Mendelian randomization', the concept has been attributed to him and subsequently developed by a number of other authors.[2–6] In particular, Davey Smith and Ebrahim[2]

Department of Preventive Medicine, University of Southern California, Los Angeles, CA 9089-9011, USA.

have shown how the magnitude of the estimated effects of a gene (*G*) on an intermediate phenotype (*IP*) and on disease (*D*) can be combined to yield an estimate of the causal effect of the intermediate phenotype on disease, as illustrated in the following figure:



(where the dotted arrow from *G* to *D* represents the indirect association assumed to be mediated entirely through *IP*).

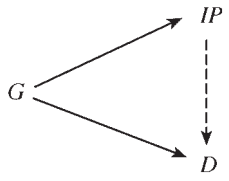
Use of instrumental variables in epidemiology

It may seem perverse to try to study the causality of a relationship between *IP* and *D* through the relationship of each with *G*, but there is merit in the idea. While its application to molecular epidemiology is novel, the idea is more than 70 years old, apparently first introduced into the econometrics literature by Wright[7] and later adopted into the statistical measurement error and causal inference literature under the rubric of 'instrumental variables'.[8–10] The basic idea is that if a causal pathway is correctly specified as in the above Figure (including certain additional assumptions discussed in the Appendix), then the causal effect of *IP* on *D* can be estimated by the ratio of the coefficients for the regression of *D* on *G* and of *IP* on *G*. (An exactly analogous argument applies in randomized controlled trials, where *G* would represent 'intent to treat' and *IP* the treatment actually received: although the *IP–D* association could be biased various ways, the *G–D* association is guaranteed by randomization to be unbiased and can be used to recover an unbiased estimate of the *IP–D* relationship. Similarly, in Berkson error models for measurement error,[11–13] *G* might represent an 'applied exposure' for a group, such as ambient air pollution, and *IP* the unobserved true personal exposure; again, the *IP–D* association can be estimated by observation of the *G–D* association, although here there is no claim that the *G–D* association would be unbiased unless exposure were applied experimentally.) In any of these settings, the precision of this estimate depends strongly upon how well *G* predicts *IP*. Thompson *et al.*[14] have shown that, even if the causal pathway is correctly specified, the statistical uncertainties in the estimates of the *G–IP* and *G–D* associations can combine to yield extremely uncertain estimates of the *IP–D* relationship.

Complications

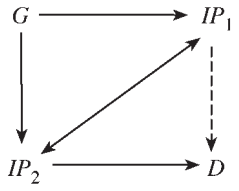
Direct effect of *G* on *D* not mediated through *IP*

One difficulty is that *G* could also have a direct effect on *D*. A minor change in the figure shown above shows how it can be used to represent confounding, by turning the *G–D* arrow into a solid one representing a direct (causal) connection and turning the *IP–D* arrow into a dashed one representing a non-causal connection induced by confounding by *G*:



This would be a case of a 'false-positive' inference—an incorrect conclusion that there is a causal connection between *IP* and *D* when in fact none exists. Of course, negative confounding could also lead to a false-negative conclusion—that there was no association between *IP* and *D* when there really is one.

One way such a situation could come about is when a single gene has pleiotropic effects. Suppose, for argument sake, that the true causal picture were as follows:



where the solid arrows indicate causal connections and the dashed arrow indicates a non-causal association induced by the other associations.

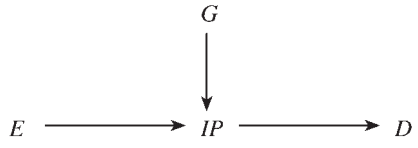
For example, Davey Smith and Ebrahim[2] provide an interesting discussion of the role of folate, homocysteine, and the methylenetetrahydrofolate reductase (*MTHFR*) gene in the aetiology of coronary heart disease (CHD) and neural tube defects (NTD). This is a very complex pathway, involving several feedback loops. For CHD, we agree with their assessment that the similarity of the direct estimate of the association between homocysteine and CHD and the indirect estimate based on the associations of each with *MTHFR* supports a causal interpretation. For NTD, on the other hand, they find a similar concordance of the estimates, but a causal interpretation seems less appropriate. We think it more likely that the second picture applies here, where IP_1 might represent homocysteine and IP_2 folate availability.[6,15,16] Nevertheless, whether homocysteine or serum folate is the proximal cause of NTD, an intervention to increase dietary folate could be an effective preventive measure.

Although *G* may be the ultimate determinant of *IP*, many other factors can induce expression of *G*, so that associations between *IP* and *D* could better reflect that proximal causal relationship than the more distant *G–D* association. Davey Smith and Ibrahim discuss the complications posed by the phenomenon of 'canalization,' the buffering of effects of genetic or environmental influences to maintain homeostatic equilibrium, via such mechanisms as alternative metabolic pathways, possibly regulated by different genes. On the other hand, *G* remains constant over time and is generally measured with a high degree of accuracy, whereas *IP* varies throughout the aetiologically relevant period and a measurement at a single point in time may subject to a large amount of

measurement error (or even bias in the case of reverse causation). These are well-known advantages of the instrumental variables approach, which apply equally to Mendelian randomization.

Gene—environment interactions

The diagrams we have considered so far do not include any external environmental factors or gene—environment (G × E) interactions. Such a model might be represented schematically as follows:



where *E* represents exposure (e.g. dietary folate) and the two arrows converging on *IP* could represent independent main effects or a $G \times E$ interaction (different genetic sensitivities to *E* or induction of the expression of *G* by *E*). In many, but not all, circumstances, it may be reasonable to assume that *G* and *E* are independently distributed in the population at risk, i.e. the gene does not predispose one to become exposed. (Obvious counterexamples might be a gene for addictive behaviour, where *E* is the substance to which the gene makes one addicted, or where a non-causal association between *G* and *E* is induced by some confounding factor such as population stratification by ethnicity or use of oral contraceptives being influenced by family history of breast or ovarian cancer.) However, if *G* and *E* *are* independently distributed in the source population and each makes independent (additive) contributions to *IP*, then *E* can be ignored in a linear model and the marginal *G–D* and *G–IP* associations can still be used to estimate the effect of *IP* on *D*. This may not apply in non-linear models, however (Appendix). Furthermore, if the two factors are associated or have interactive effects, this distortion could be quite severe and could lead to either false-positive or false-negative inferences. This picture can become even more complicated when the architecture of competing pathways evolves over time in response to developmental influences or exposures via adaptive mechanisms (canalization).[2]

Even in linear models, it seems a stretch to conclude that:

> the association of genotype with NTD risk … demonstrates that an environmental intervention may benefit the whole population, independently of the genotype of individuals receiving the intervention[2]

—at least without good observational evidence about the association of exposure and disease *within* genotype. One would also want to see evidence that changes in exposure actually lead to changes in disease risk, particularly in complex systems where there are multiple points at which different genetic and environmental perturbations may lead to various phenotypic outcomes.[6]

Clayton and McKeigue[3] have argued that:

> Despite current enthusiasm for study of gene—environment interactions, the closely related issue of how to define and interpret interaction between environmental factors remains unresolved after two decades of debate. … We suggest that epidemiologists should focus instead on use of genetic associations to test hypotheses about causal pathways amenable to intervention. … In this example [*NAT* and heterocyclic amines in cooked meat], as with the *MTHFR* gene, there is a possible biological interaction between genotype and dietary intake, but testing for statistical interactions between genotype and dietary intake would not contribute much to our understanding of these biological interactions or to our ability to exploit them in disease prevention. … The prospects for epidemiology in the post-genome era depend on understanding how to use genetic associations to test hypotheses about causal pathways, rather than on modeling the joint effects of genotype and environment.

Part of their argument relies on the observation that power to test main effects will often be much better than for interactions, although there are exceptions.[17] Hence the opportunity to exploit Mendelian randomization to assess causality is a great advantage of tests of pure genetic main effects. Indeed, the track record of replication of reports of $G \times E$ interactions seems to be even more dismal than for main effects of gene associations,[18–20] perhaps in large part because such studies are frequently underpowered, involve some data dredging, and are subject to publication bias. We generally agree with their conclusion that:

> A case-control study of the relation between the TT genotype [of *MTHFR*] and risk of neural tube defect *can be interpreted as equivalent to a randomized trial* of the effect on disease risk of alteration of the availability of folate[3] [emphasis added].

Gene—gene interactions

The same picture might apply if one were to replace *E* by another gene, say *H*. It is quite conceivable that a second causal variant may exist within the same candidate gene region and be in linkage disequilibrium with *G*. The lack of independence between *H* and *G* may lead to substantial bias in the estimation of the *G–D* association.[6] Furthermore, by the same line of argument as above for non-linear models, if *IP* were determined by two genes (either independently or in some interactive manner), but one only assessed *G*, then the association between *IP* and *D* estimated from the associations of each with *G* would also be biased. In particular, a false-negative conclusion could be reached if *H* were really the more relevant determinant of *IP* and failure to account for it led to a null result for the *G–D* association. As with $G \times E$, failure to account for $G \times G$ interactions could also lead to either false-positive or false-negative inferences.

Population stratification

A reservation about the broad conclusion Mendelian randomization is equivalent to a randomized trial is that *G–D* associations from case-control studies are susceptible to distortion by population stratification.[6] Not only substantial genetic differences in populations, but more subtle clustering of genetically similar individuals within the population, can bias a test of the *G–D* association.[21] Although some have argued that population stratification may not be a serious concern, at least in Caucasian populations of European descent,[22,23] this problem can be overcome by appropriate design or analysis. The low power of Mendelian randomization compared with direct tests of association implies that very large sample sizes will be required.

Unfortunately, the problem of inflation of Type I Error rates by population stratification will only increase with increasing sample size, as smaller and smaller biases will become significant.

To fully exploit the power of Mendelian randomization, one should consider using the case-parent-triad design that is based on the random transmission of alleles from parents to offspring and is therefore robust to population stratification.[24,25] Similar properties are shared by other family-based association tests (FBAT), such as a sib case-control design and those that exploit both parents and siblings or even extended pedigrees.[26–29]

Conclusions

We conclude that the validity of the Mendelian randomization approach to evaluating the causality of an association between *IP* and *D* depends upon the correct specification of the causal model. If *G* has multiple effects, at least one of which has a causal effect on *D* through some pathway not involving *IP*, or if the association between *G* and *D* is confounded by population stratification or other genes it is in linkage disequilibrium with, then the estimated association between *IP* and *D* will be distorted. The method will be most efficient when the connection between *G* and *IP* is strong, as noted by Davey Smith and Ibrahim[2] in comparing the usefulness of the beta-fibrinogen and haptoglobin polymorphisms as predictors of plasma fibrinogen and vitamin C respectively.

Biological pathways are extremely complex, so a simple triangulation picture will almost certainly be wrong in most situations. However, our understanding of these pathways will doubtless continue to improve (and hence the pictures will get more and more complicated), but on the other hand, prospects for overcoming confounding and reverse causation in traditional observational studies of the *IP–D* association are very limited. In the long run, the concept of Mendelian randomization may prove to be a valuable way for epidemiology to move 'beyond its limits'. Thus, the conditions for its validity deserve careful consideration.

Appendix

Validity and efficiency of the 'instrumental variables' approach

Suppose we wish to estimate the slope of a regression of *D* on *IP* and we have available a surrogate variable *G* for *IP*—that is, *G* is a determinant of *IP* but is conditionally independent of *D* given *IP*. We begin by assuming all the relationships are linear, before turning our attention to the additional complications that arise in non-linear models. Thus, we assume

$$\mathrm{E}(D|IP) = \beta_0 + \beta_1 IP, \qquad \mathrm{var}(D|IP) = \sigma^2$$
$$\mathrm{E}(IP|G) = \alpha_0 + \alpha_1 G, \qquad \mathrm{var}(IP|G) = \tau^2$$

Then,

$$\mathrm{E}(D|G) = \gamma_0 + \gamma_1 G, \qquad \mathrm{var}(D|G) = \omega^2$$

where $\gamma_0 = \beta_0 + \beta_1\alpha_0$, $\gamma_1 = \beta_1\alpha_1$, and $\omega^2 = \sigma^2 + \beta_1{}^2\tau^2$. Thus, if one had unbiased estimators of γ_1 and α_1, then γ_1/α_1 becomes an unbiased estimator of β_1 but its variance is complex (see ref. 14 for a derivation), and can be infinite in the event that the variance of α_1 is large in relation to its true value. Note also that the parameter of interest, β_1, is involved in var($D|G$) and thus ignoring this information will lead to a less than fully efficient estimator. In particular, if var($IP|G$) were not constant, then an appropriately weighted estimator of γ would be required. In any event, the variance of a ratio can be quite unstable, depending strongly upon τ^2. Thus, if *G* were not a good predictor of *IP*, then var(γ_1/α_1) will be very large.

The derivation above assumes that the model is correctly specified. In the main body of our article, we describe several ways the model could be misspecified and the implications for bias. For example, suppose *G* has a direct effect on *D*, independent of *IP*, say $\mathrm{E}(D|IP) = \beta_0 + \beta_1 IP + \beta_2 G$. Then it is easy to see that γ_1/α_1 estimates $\beta_1 + \beta_2/\alpha_1$, and thus will yield a biased estimate of the causal effect of *IP* on *D*. Alternatively, suppose there is another factor that influences *IP*, say *H*, which could be another gene or some environmental factor. Suppose first that *H* has no direct effect on *D* other than through its influence on *IP*, and *G* and *H* are independent and contribute additively to *IP*, that is, $\mathrm{E}(IP|G,H) = \alpha_0 + \alpha_1 G + \alpha_2 H$. Then even if *H* is ignored, γ_1/α_1 still estimates β_1, although its variance will be increased. However, if *G* and *H* are associated in the population or if they have an interactive effect on *IP*, then both γ_1 and α_1 will be biased, but to the same extent, so their ratio γ_1/α_1 turns out to be a consistent estimator of β_1 (assuming there is no direct effect of *G* or *H* on *D* except through *IP*). To see this, suppose $\mathrm{E}(H|G) = \eta_0 + \eta_1 G$. Then if $\mathrm{E}(IP|G,H) = \alpha_0 + \alpha_1 G + \alpha_2 H$, then $\mathrm{E}(IP|G) = \alpha_0 + \alpha_1 G + \alpha_2 \mathrm{E}(H|G) = \alpha_0{}^* + \alpha_1{}^* \mathrm{G}$, where $\alpha_0{}^* = \alpha_0 + \alpha_2\eta_0$ and $\alpha_1{}^* = \alpha_1 + \alpha_2\eta_1$. Likewise, if $\mathrm{E}(D|IP,G,H) = \beta_0 + \beta_1 IP$, then $\mathrm{E}(D|G) = \beta_0 + \beta_1 \mathrm{E}[IP|G,\mathrm{E}(H|G)] = \gamma_0{}^* + \gamma_1{}^* G$, where $\gamma_0{}^* = \beta_0 + \beta_1(\alpha_0 + \alpha_2\eta_0)$ and $\gamma_1{}^* = \beta_1(\alpha_1 + \alpha_2\eta_1)$. Thus $\gamma_1{}^*/\alpha_1{}^* = \beta_1(\alpha_1 + \alpha_2\eta_1)/(\alpha_1 + \alpha_2\eta_1) = \beta_1$. This also applies if *G* and *H* have an interactive effect on *IP* (but no direct effects on *D*), provided the estimates of the *GD* and *GIP* associations derive from the same dataset or studies with the same joint distribution of *G* and *H*.

For dichotomous disease traits, the derivation is somewhat more complex and the conditions for validity are more restrictive. The most tractable situation is when $IP \sim N(\alpha_0 + \alpha_1 G, \tau^2)$ and $\ln[\Pr(D = 1|IP)] = \beta_0 + \beta_1 IP$ for a rare disease. Then it is easily shown that $\ln[\Pr(D = 1|G)] = \gamma_0 + \gamma_1 G$, where $\gamma_0 = \beta_0 + \beta_1\alpha_0 + \beta_1{}^2\tau^2/2$ and $\gamma_1 = \alpha_1\beta_1$, so γ_1/α_1 is a consistent estimator of β_1, just as in the linear model. For a probit link, the corresponding expression is $\beta_1 = \gamma_1/\sqrt{(\alpha_1{}^2 - \gamma_1{}^2\tau^2)}$, without the need for a rare disease assumption, but now the ratio γ_1/α_1 is only an approximate estimator of β_1. Closed-form solutions are not available for the logistic model, but qualitatively the behaviour is similar.[14] As before, a direct effect of *G* on *D* will yield a biased estimator.

Unlike the linear model, however, if there is another factor *H* influencing *IP*, then if *G* and *H* are not independent, the estimators γ_1 and α_1 are both biased, but these biases may no longer cancel out exactly. Suppose that $\ln[\Pr(D = 1|IP)] = \beta_0 + \beta_1 IP$ and $IP \sim \mathrm{N}(\alpha_0 + \alpha_1 G + \alpha_2 H,\ \tau^2)$. If $H \sim \mathrm{N}(\eta_0 + \eta_1 G,\ \omega^2)$ and *H* is ignored, then $IP \sim \mathrm{N}(\alpha_0{}^* + \alpha_1{}^* G,\ \tau^2 + \alpha_2{}^2\omega^2)$, where $\alpha_1{}^* = \alpha_1 + \alpha_2\eta_1$, and $\ln[\Pr(D = 1|G)] = \gamma_0{}^* + \gamma_1{}^* G$, where $\gamma_1{}^* = \beta_1(\alpha_1 + \alpha_2\eta_1)$, so in this case γ_1/α_1 is indeed a consistent estimator of β_1. But now suppose instead that *H* were dichotomous, with $\Pr(H = 1|G) = p_G$. Then $\gamma_1{}^* = \beta_1\alpha_1 + \ln[1 + p_1\ \exp(\beta_1\alpha_2)] - \ln[1 + p_0\ \exp(\beta_1\alpha_2)]$ and $\alpha_1{}^* = \alpha_1 +$

$\alpha_2(p_1 - p_0)$. Thus γ_1^*/α_1^* will not estimate β_1 unless $p_1 = p_0$ or $\alpha_2 = 0$ or $\beta_2 = 0$.

In general, the validity of Mendelian randomization lies in the equivalency of $\alpha_1\beta_1 = \gamma_1$. That is, the association between *GIP* and *IPD* is assumed to be equivalent to the *GD* relation. If $g(D|\cdot)$ gives the functional relation between an exposure and the disease outcome and $h(IP|G)$ gives the relation between the gene variant and the intermediate phenotype, then for Mendelian randomization estimates to be valid, it must be possible to write $g(D|G,\gamma_1) = \int g(D|IP,\beta_1)\ h(IP|G,\alpha_1)\ d\ IP$ as $g(D|G,\alpha_1\beta_1)$. This holds if $h(\cdot)$ is conjugate to $g(\cdot)$.

References

[1] Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986;**i:**507–08.

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

[3] Clayton DG, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; **358:**1357–60.

[4] Keavney B. Genetic epidemiological studies of coronary heart disease. *Int J Epidemiol* 2002;**31:**730–36.

[5] Youngman L, Keavney B, Palmer A *et al*. Plasma fibrinogen and fibrinogen genotypes in 4685 cases of myocardial infarction and in 6002 controls: test of causality by 'Mendelian randomization'. *Circulation* 2000;**102(Suppl.2):**31–32.

[6] Little J, Khoury MJ. Mendelian randomization: a new spin or real progress? *Lancet* 2003;**362:**930–31.

[7] Wright S. Appendix. In: Wright PG (ed.). *The Tariff on Animal and Vegetable Oils*. New York: Macmillan, 1928.

[8] Buzas JS, Stefanski LA. Instrumental variable estimation in generalized linear measurement error models. *J Am Statist Assoc* 1996; **91:**999–1006.

[9] Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;**29:**722–29.

[10] Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2000.

[11] Berkson. Are there two regressions? *J Am Statist Assoc* 1950;**45:** 164–80.

[12] Thomas DC, Stram D, Dwyer J. Exposure measurement error: Influence on exposure-disease relationships and methods of correction. *Ann Rev Public Health* 1993;**14:**69–93.

[13] Fuller W. *Measurement Error Models*. New York: Wiley, 1987.

[14] Thompson JR, Tobin MD, Minelli C. On the accuracy of estimates of the effect of phenotype on disease derived from Mendelian randomisation studies. Technical report 2003-GE1. Leicester: University of Leicester (available from http://www.prw.le.ac.uk/ research/HCG/ research.html). 2003.

[15] Friso S, Choi SW, Girelli D *et al*. A common mutation in the 5, 10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. *Proc Natl Acad Sci USA* 2002;**99:**5606–11.

16 Yamada K, Chen Z, Rozen R, Matthews RG. Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. *Proc Natl Acad Sci USA* 2001;**98:** 14853–58.

[17] Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med* 2002;**21:** 35–50.

[18] Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;**361:** 865–72.

[19] Rothman N, Wacholder S, Caporaso NE, Garcia-Closas M, Buetow K, Fraumeni JF Jr. The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochimica Biophysica Acta* 2001;**1471:**C1–10.

[20] Brennan P. Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* 2002; **23:**381–87.

[21] Thomas DC, Witte JS. Point: Population stratification: A problem for case-control studies of candidate gene associations? *Cancer Epidemiol Biomarkers Prev* 2002;**11:**505–12.

[22] Wacholder S, Rothman N, Caporaso N. Counterpoint: Bias from population stratification is not a major threat to the validity of conclusions from epidemiologic studies of common polymorphisms and cancer. *Cancer Epidemiol Biomarkers Prev* 2002;**11:** 513–20.

[23] Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;**361:**598–604.

[24] Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;**52:**506–16.

[25] Self SG, Longton G, Kopecky KJ, Liang KY. On estimating HLA/ disease association with application to a study of aplastic anemia. *Biometrics* 1991;**47:**53–61.

[26] Curtis D. Use of siblings as controls in case-control association studies. *Ann Hum Genet* 1997;**61:**319–33.

[27] Lunetta KL, Faraone SV, Biederman J, Laird NM. Family-based tests of association and linkage that use unaffected sibs, covariates, and interactions. *Am J Hum Genet* 2000;**66:**605–14.

[28] Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered* 2000;**50:**211–23.

[29] Shih MC, Whittemore AS. Tests for genetic association using family data. *Genet Epidemiol* 2002;**22:**128–45.