LEADING ARTICLE

Am J Pharmacogenomics 2005; 5 (1): 1-20
1175-2203/05/0001-0001/$34.95/0

The Epidemiologic Approach to Pharmacogenomics

Julian Little,[1] *Linda Sharp*,[2] *Muin J. Khoury*,[3] *Linda Bradley*[3] and *Marta Gwinn*[3]

1 Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada
2 Department of Medicine and Therapeutics, Epidemiology Group, University of Aberdeen, Aberdeen, Scotland
3 Department of Health and Human Services, Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Abstract

The epidemiologic approach enables the systematic evaluation of potential improvements in the safety and efficacy of drug treatment which might result from targeting treatment on the basis of genomic information. The main epidemiologic designs are the randomized control trial, the cohort study, and the case-control study, and derivatives of these proposed for investigating gene-environment interactions. However, no one design is ideal for every situation, and methodological issues, notably selection bias, information bias, confounding and chance, all play a part in determining which study design is best for a given situation. There is also a need to employ a range of different designs to establish a portfolio of evidence about specific gene-drug interactions.

In view of the complexity of gene-drug interactions, pooling of data across studies is likely to be needed in order to have adequate statistical power to test hypotheses. We suggest that there may be opportunities (i) to exploit samples from trials already completed to investigate possible gene-drug interactions; (ii) to consider the use of the case-only design nested within randomized controlled trials as a possible means of reducing genotyping costs when dichotomous outcomes are being investigated; and (iii) to make use of population-based disease registries that can be linked with tissue samples, treatment information and death records, to investigate gene-treatment interactions in survival.

With the completion of the Human Genome Project[1] and advances in technologies for genomic analysis, claims of new gene discoveries affecting disease susceptibility are being hailed as providing the basis for drug discovery to improve the management of disease or prevent it.[2] In theory, human genome discoveries have broad potential applications for improving health and preventing disease. For therapy and tertiary prevention, advances in human genetics could contribute to the development of better drugs and to the tailoring of drug use to the individual's genetic makeup to maximize benefits and minimize harm. For secondary prevention, new screening tests for early disease identification, or modifications of existing ones, might be developed based on stratification by genotype. For primary prevention, an improved understanding of genetic effects and gene-environment interactions in etiology will enable better interventions to be developed, such as chemoprevention and the avoidance of specific exposures, and to identify subgroups of individuals for whom intervention would be of potential value.[3-7]

There are a number of situations in which the targeting of therapy is potentially useful, including when drug-based management is expensive, when an intervention is to be used in otherwise healthy people, or when the value of a therapy is limited by genetic-related toxicity.[8,9] The costs of drug-based management (costs that apply to the health service, the patient, or both) can include the cost of the drug itself or the clinical time needed to optimize the dose. Potentially, targeted therapy could improve the efficacy of the drug, thereby improving outcome, including survival and quality of life. It could minimize adverse drug reactions or toxicity, and this would also lead to improved patient compliance. These effects would be beneficial to the patient and reduce use of healthcare resources.

The discovery of inherited variation in response to pharmaceutical agents[10] has stimulated investigation on gene-environment

interaction in general, for example, the role of variants of genes encoding enzymes involved in xenobiotic metabolism in the etiology of cancer.[11] The term 'pharmacogenetics' was introduced by Vogel[12] to describe the investigation of the genetic basis of variation in response to drugs. More recently, the term 'pharmacogenomics' has been introduced, emphasizing the "development of novel drugs based on newly discovered genes as the entire human genome becomes sequenced".[13] The two terms have also been used interchangeably.[13-15] Tsai and Hoyme[16] define 'pharmacogenomics' as encompassing "all aspects of drug behavior, including absorption …, distribution …, metabolism …, excretion … and receptor-target affinity". Thus, pharmacogenomics would include "comparative genome hybridization, amplification of gene copy number, gene expression microarray analysis of levels of hundreds or thousands of messenger RNAs, and proteomic analysis of the level of expression of very large numbers of proteins".[17] More generally 'genomics' has been defined as "the study not just of single genes, but of the functions and interactions of all the genes in the genome" and 'genetics' as "the study of single genes and their effects".[18] We prefer the broader view that pharmacogenetics is part of pharmacogenomics, so that pharmacogenomics encompasses a continuum from variation in response attributable to variation at a single locus, through variation in response attributable to multiple loci, to variation in expression at multiple loci in somatic cells. Thus, pharmacogenomic data are relevant not only to the investigation of genetic variation in drug response, but also investigation of the effects of exposure (e.g. exposure 'signatures' evidenced by specific somatic mutations, changes in gene expression)[19,20] and definition of outcomes (refinement of disease classification, e.g. defining types of tumor on the basis of gene-expression profiles).[21] In this paper, we focus on drug response differential according to the germline genomics.

It is widely expected that in the future drug treatment will be stratified on the basis of genetic information derived from testing for variants of multiple gene loci. Realization of this promise will depend on the classical public health sciences for precise determination of the effects of genetic variants on drug treatment outcome, and the potential magnitude of impact of drug treatments (and drug-based preventative strategies) on public health,[22,23] as well as for improved laboratory technology, clinical research, evaluation of biologic pathways, and improved drug investigation.[3] Well-conducted epidemiological studies are needed to quantify the impact of gene variants on the efficacy of a drug treatment, and the risk for adverse outcomes. This information is an essential basis for evaluating potential cost-effectiveness of a pharmacogenetic approach to treatment and for decision analysis models.[24] The population focus of epidemiologic research is important in considering the generalizability of findings from investigations of gene-drug interaction. Epidemiologic studies are also required in the process of clinical validation of any new tests that might be developed to stratify treatment on the basis of test results, and to monitor the use of such tests in the populations to which they might be applied.

In this paper, we first outline the biologic basis for gene-treatment interaction. This is followed by a review of each of the main epidemiologic study designs. Randomized control trials, the main design used in experimental epidemiology, can be used to investigate the effects of gene-drug interactions on the efficacy of treatment and chemoprevention. Cohort studies can also be used for these purposes, as well as investigating disease etiology in general. Case-control studies are typically used to investigate adverse drug reactions, and also disease etiology in general. Novel study designs have been developed on the basis of each of these three designs. We discuss methodologic issues in excluding non-causal explanations for gene-drug interactions, some of which are generic issues and some of which affect some designs more than others. Then, issues in epidemiological analysis of gene-drug interactions are considered. Finally, we examine the use of epidemiologic evidence in the development of pharmacogenomic testing.

1. Biologic Basis for Gene-Drug Interaction

There are several processes occurring between the ingestion and excretion of a drug and its metabolites in which interaction with the products of polymorphic genes might be relevant to the safety and effectiveness of the drug.

Firstly, drugs usually undergo oxidation, reduction and hydrolysis (phase I reactions), and conjugation reactions (phase II reactions, e.g. acetylation, sulfation) that convert the drug into metabolites that are more water-soluble and, therefore, more easily excreted.[9,25] These affect response by altering drug concentrations, i.e. a pharmacokinetic effect.[26] Many phase I reactions are catalyzed by members of the cytochrome P450 (CYP) supergene family,[27] many of which are polymorphic.[28] Out of more than 60 genes in this family, *CYP2D6*, *CYP3A4*, *CYP3A5*, *CYP3A6*, *CYP2C9*, *CYP1A2*, *CYP2C19*, and *CYP2E1* account for most of the variation in phase I metabolism of drugs in current use.[14] Examples of polymorphic enzymes affecting phase II reactions include *N*-acetyltransferases, glutathione-*S*-transferases, sulfotransferases, methyltransferases, and uridine 5′-triphosphate glucuronosyltransferases.[9,14]

Secondly, although most drugs or metabolites enter cells by passive diffusion, some are actively transported by transporter proteins, and polymorphisms in genes encoding these proteins, such as *ABCB1*, may influence the effects of certain drugs such as digoxin, anticonvulsants, and protease inhibitors.[14,26,29]

Thirdly, there may be genetically determined variation in drug targets, such as receptors for endogenous regulatory ligands that also serve as receptors for drugs.[14,26,30] This area of investigation (pharmacodynamics) is less established than investigation of genetic variation in drug metabolism.[30] The existence of polymorphic variation in drug targets and different aspects of drug disposition means that the genetic basis of response to drug treatment is likely to be complex. The potential use of genomic information requires rigorous evaluation.

In this paper, the examples relate to polymorphisms at single genetic loci. However, it is likely that multiple genes operating in pathways will determine the response to a drug.[2,31,32] As yet, there are no clear examples of interactions between drug treatment and multiple genes, although it is becoming increasingly feasible to assess variants of multiple genes (single nucleotide polymorphisms [SNPs]) in large population samples. Millions of SNPs have been identified,[33,34] so a challenge is to identify the SNPs that are most likely to be involved in gene-treatment interaction (see section 3.4). Therefore, there has been increasing interest in the potential value of haplotyping for multiple SNPs within candidate genes, since the number of haplotypes within a gene is much smaller than the theoretical number of all possible haplotypes.[35,36] Potentially this means that information on a subset of SNPs could capture most of the information about genetic variation in a stretch of DNA, referred to as a 'block'.[37-40] Such subsets of SNPs are described as haplotype-tagging SNPs. Thus, the aim of haplotype tagging is to reduce the number of SNPs that have to be genotyped, without substantial loss of haplotype diversity, and to maintain statistical power to detect haplotype-disease associations[41] and haplotype-treatment interactions.

A simulation study suggested that selection of SNPs based on maximizing haplotype diversity had greater power to detect haplotype-disease associations than random selection or selection based on pairwise linkage disequilibrium.[42] Kamatani et al.[43] examined 4190 SNPs in 199 genes coding for enzymes involved in drug metabolism and transport in DNA from 752 Japanese subjects. From 3244 common SNPs (allele frequency ≥10%), 1035 (32%) represented most of the major haplotypes within the blocks and hence could tag the haplotype. Almost two-thirds of the uncommon SNPs (allele frequency <10%) were within the blocks, so potentially these 1035 haplotype-tagging SNPs could be used to search for common and uncommon SNPs associated with specific phenotypes. However, a key problem is the definition of haplotype blocks, and differences in this give conflicting results.[44,45]

In an analysis of the published data from two European studies, Zhang et al.[41] found that haplotype-tagging SNPs were generally worse at detecting causal loci than random selection. This is likely due to the fact that SNPs selected by maximizing haplotype diversity tend to be common, and are therefore inefficient in the detection of rarer causal SNPs. An alternative approach of selecting tag SNPs without regard to haplotypes[46] had greater statistical power to detect causal alleles than haplotype-tagging SNPs, but less than randomly selected SNPs.[41] van Hylckama Vlieg et al.[47] recently showed that a strategy based on choosing SNPs spread throughout the gene, for which the rare alleles had non-identical frequencies between 15% and 50% would have led to detection of the Factor V Leiden variant that is now an established risk factor for venous thrombosis. This suggests that a candidate gene approach can be successful in the absence of information about the haplotype structure of the gene.

In the pharmacogenomics area, gene-drug interaction has been considered in two ways. Firstly, the joint effects of genotype and therapy (or chemopreventive agent) on outcome are assessed. For example, Martinez et al.[48] investigated the joint effects of aspirin (acetylsalicylic acid) use and a polymorphism in the ornithine decarboxylase gene (*ODC1*) on the risk for recurrence of colorectal adenomas. Secondly, the relation between genotype and outcome has been assessed only in those receiving a particular therapy. For example, Higashi et al.[49] investigated the relationship between *CYP2C9* variants and serious bleeding events in patients receiving warfarin. Here, there is an assumption that there is no relationship between genotype and outcome in those not receiving a therapy. In theory, for variants of genes coding for enzymes involved in xenobiotic metabolism, such as the CYP enzymes, glutathione-*S*-transferases and *N*-acetyltransferases, a relation between genotype and outcome in the absence of therapy would not be expected. However, these enzymes tend to have broad substrate specificity,[27] so a relationship between genotype and outcome in the absence of therapy may occur because of interaction with other exposures. For example, associations between genetic variants influencing xenobiotic metabolism and several types of cancer have been reported, with suggestions of interaction with exposures to tobacco smoking and the consumption of cooked meats.[50-56] An additional assumption is that among individuals for whom a particular therapy is indicated, the proportion who receive therapy does not vary by genotype.

2. Epidemiologic Study Designs in the Investigation of the Effects of Genetic Variation on Drug Treatment Outcomes

As in any evaluation of the effects of a drug, a range of outcomes may be considered, and study designs differ in their ability to assess effects on these. Treatment outcomes include: (i) response to drug, for example, reduction in tumor mass, change in level of biochemical markers associated with prognosis; (ii) occur-

rence of specific events, both those for which the treatment is intended to provide protection (such as recurrence of disease or the occurrence of hip fracture in persons with osteoporosis), and adverse effects; (iii) quality of life; and (iv) survival. The strengths and weaknesses of the three main epidemiologic designs, namely randomized controlled trials (RCTs), cohort studies, and case-control studies, to investigate gene-drug interactions are summarized in table I. No one design is ideal for every situation. Therefore, there is a need to employ a range of different designs to establish a portfolio of evidence.

2.1 Randomized Controlled Trials

The definitive method of investigating the efficacy of a drug is the randomized controlled trial. The strength of this design is that, provided the trial is sufficiently large, the distribution of potential confounders, known (measured) and unknown (unmeasured), will differ between the group assigned to receive the drug and the control group no more than would be expected by chance. Thus, potentially the RCT could be preferable to observational (i.e. cohort and related) designs for investigating gene-environment interaction in that the exposure (i.e. drug treatment) is well defined, and confounding minimized. A further potential advantage of this approach is that it is possible to have both the clinicians and patients blinded to the intervention they are undergoing, so as to exclude the possibility that knowledge of the drug treatment method could bias the assessment of outcome. If an interaction between the drug treatment and genotype is detected, this implies that testing for the genetic variant could inform drug treatment of the condition. Multiple outcomes can be assessed; this is valuable in determining the balance of benefits to harm and is a strong source of evidence for cost-utility analyses.

Therefore, RCTs offer a powerful potential means of identifying gene-drug interaction (when there is an *a priori* hypothesis). One approach is to collect samples at the time of enrollment. If it is known that the genotype is uncommon, an option at that stage is to stratify randomization by genotype in order to ensure balance between the trial arms. Alternatively, at the end of the trial, the samples can be genotyped. In both situations, interaction between drug treatment and genotype can be assessed. Provided that the trial is of adequate size and analyzed according to intention-to-treat, the distributions of genotype and exposure will be independent. Intention-to-treat is particularly important here as it is possible that a patient would be started on one regimen, tolerate it badly or not respond, and be switched to another. The tolerance/response might be influenced by the genotype(s) under investigation.

As yet, there are relatively few clear examples of gene-drug interactions identified on the basis of RCTs.[58,59] Murphy et al.[60] reported an interaction between antidepressant type (paroxetine versus mirtazapine) and a polymorphism in the serotonin receptor 2A locus (*HTR2A*) in the occurrence of adverse effects and the need to discontinue therapy. In 246 elderly patients with major depression, the proportion discontinuing therapy within 8 weeks was 46.3% in homozygotes for the C allele who were randomized to take paroxetine, compared with 16% in those with other genotypes randomized to take paroxetine, and 15–-16% in those randomized to take mirtazapine (p-value for difference 0.001 for all assessment points in survival analysis using the log-rank method). The severity of adverse effects was also greater in patients who were homozygous for the C allele and assigned to paroxetine therapy.

Although biologic samples are taken and stored in many trials, most have not been designed to test *a priori* hypotheses about gene-treatment interaction and, therefore, may not have adequate statistical power to test specific hypotheses about interaction. For example, in the period 1989–2000, the average number of patients in RCTs of breast cancer therapy was 402 (95% CI 352, 450) and in RCTs of therapy for other types of cancer 213 (95% CI 211, 225).[61] In 77 RCTs initiated by the AIDS Clinical Trials Group in 1986–1996, the largest sample size was <200 subjects for 32 trials, 200–500 for 27 trials, and 500 or more (mean target 1152) for 18 trials.[62] Of these, 48 trials (62%) achieved 80% or more of their target recruitment, 16 (21%) between 50% and 80% and 13 (17%) less than 50%. These sample sizes are small compared with the numbers typically required to detect interaction even in the situation of conducting a nested case-control study (see below) within a trial.[63-67] Therefore, we suggest that it would be worthwhile exploiting samples from trials already completed to investigate possible gene-drug interactions. This could be done as a hypothesis-generating exercise, requiring cautious interpretation. We would emphasize that the findings of such investigation would require replication. Many trials are multicenter in order to have adequate statistical power to detect the main effect of treatment. In consequence, it may be necessary to pool trials in order to have adequate power to detect gene-drug interactions, but this may introduce heterogeneity because of differences in trial design (e.g. eligibility criteria, detail of treatment regimen, method of follow-up). Again, we emphasize the need for cautious interpretation, and replication.

The RCT can be used to evaluate potential interventions for primary prevention. As yet, there have been few such investigations. An example is a study of the interaction between hormone replacement therapy (HRT) and the estrogen receptor (ER)-α (*ESR1*) polymorphism in relation to falls and grip-strength.[68] However, the analysis was not by intention to treat.

Table I. Potential for bias, confounding factors, and ability to investigate different types of outcome of the main epidemiologic study designs in the investigation of gene-drug interactions[a]

Criterion	Study design		
	randomized controlled trial	cohort study	case-control study
Subject group			
Patients with specific condition	Outcomes assessed during follow-up	Outcomes assessed during follow-up	Defines case group; treatment assessed retrospectively
Subjects without disease	Investigation of chemoprevention; disease outcome assessed during follow-up	Investigation of etiology; disease occurrence assessed during follow-up	Sample of these define control group; treatment assessed retrospectively
Selection bias			
Recruitment			
Generalizability[b]	May be limited as result of strict eligibility criteria	Depends on eligibility criteria	Depends on eligibility criteria
Loss to follow-up	✓	May be differential by treatment group	
Control selection			A problem if source population for cases and controls differs
Incomplete ascertainment of outcome	✓	May be differential by treatment group	Incomplete ascertainment of cases
Collection and analysis of DNA			
Source of DNA			Potential problem if differs between cases and controls
Refusal or inability to provide biological specimens	✓	✓	✓
Insufficient amount of DNA limits the number of assays being performed in subsets of subjects	✓	✓	✓
Information bias			
Treatment	Should be of high quality. Unlikely to differ by treatment group	Non-differential error due to problems in assessing treatment information Quality may differ by treatment group	Non-differential error due to problems in assessing treatment information Potential for differential recall between cases and controls
Genetic information	Errors handling specimens such as mislabeling of vials Errors in genotype assays: potential for these to be differential by treatment group if specimen handling or assays not blinded to treatment group	Errors handling specimens such as mislabeling of vials Errors in genotype assays: potential for these to be differential by treatment group if specimen handling or assays not blinded to treatment group	Errors handling specimens such as mislabeling of vials Errors in genotype assays: potential for these to be differential by case-control status if specimen handling or assays not blinded to this

Continued next page

Table I. Contd

Criterion	Study design		
	randomized controlled trial	cohort study	case-control study
Confounding	Minimal potential, provided trial of adequate size (and successful randomization)	Unaccounted factors associated with the outcome and treatment under investigation (that are not an intermediate step between treatment and outcome) Unaccounted alleles associated with the outcome in linkage disequilibrium with the allele under investigation	Unaccounted factors associated with the outcome and treatment under investigation (that are not an intermediate step between treatment and outcome) Unaccounted alleles associated with the outcome in linkage disequilibrium with the allele under investigation
Population stratification	Minimal potential, provided trial of adequate size (and successful randomization)	Unaccounted variation in ethnic backgrounds by treatment group when ethnic groups tend to have different treatments and different frequencies of allelic variants	Unaccounted variation in ethnic backgrounds of cases and controls, when ethnic groups have different rates of outcome and different frequencies of allelic variants
Outcomes	Multiple outcomes can be assessed	Multiple outcomes can be assessed	Single outcome assessed
Response to treatment	✓	✓	Only if defined as dichotomous outcome
Events, including adverse effects	Inefficient for studying rare events Length of follow-up may be too limited to assess occurrences of events with long latent period	Inefficient for studying rare events Length of follow-up may be too limited to assess occurrences of events with long latent period	Potentially highly efficient for rare events, and events with long latent period
Quality of life	✓	Information tends not to be collected	Only if defined as dichotomous outcome; reporting may be influenced by other outcomes of treatment
Survival	✓	✓	Only if defined as dichotomous outcome

a This table includes information from a number of sources, including Garcia-Closas et al.[57]

b In all three designs, genotype frequency results in one geographic area may not be generalizable to another because of geographical variation. Similarly, result of a study done in one ethnic group may not be generalizable to other ethnic groups.

✓ indicates yes.

The main limitation of RCTs is that they are usually carried out on selected groups, and the results may not be generalizable to the population at risk. A particular problem is that subjects who may be particularly vulnerable to the adverse effects of treatment, because of co-morbidity or age, are frequently excluded.[69] In addition, RCTs cannot address all questions about treatment. For example, it would not be possible to initiate a trial of a treatment that was thought to be harmful. It might be unfeasible to conduct a trial in which the outcome of interest occurs a long time after treatment, or for rare outcomes such as certain types of adverse effects (table I).

2.2 Cohort Studies

In a cohort study of treatment outcome in patients with a particular condition, patients who do not have the outcome of interest (e.g. a different condition or an adverse drug reaction) are recruited to participate in the study and are then followed over time to identify who develops the outcome. The approach may also be used to investigate the effect of treatment on disease recurrence. Typically, the rate of occurrence (or recurrence) of the outcome in the group who have been given a drug treatment is compared with the rate in those who have not been given the treatment, or who have been given a different treatment. As in a RCT, multiple outcomes can be assessed. Information (i.e. treatment or drug use, socio-demographic factors, medical history, lifestyle factors) is collected at the beginning of the study prior to the occurrence of the outcome. The cohort design could be regarded as similar to the RCT except that the assignment of drug therapy is subject-driven rather than randomized. Consequently, the design is vulnerable to potential confounding.

An example of the application of cohort studies to the investigation of the joint effects of genotype and drug treatment on recurrence, is a study of aspirin use and a G to A substitution in intron 1 of the ornithine decarboxylase (*ODC1*) gene in the recurrence of colorectal adenoma.[48] 688 persons from whom adenomas had been removed provided data on aspirin use by self-

Table II. Joint effect of aspirin (acetylsalicylic acid) use and ornithine decarboxylase (*ODC1*) IVS1 +317G/A polymorphism genotype on adenoma recurrence[48]a,b

Group	*ODC1* genotype	Aspirin use	Total subjects	Number with recurrence	Incidence ratio (95% CI)
A (reference group)	GG or GA	No	448	237	1.0
B	GG	No	31	13	0.79 (0.52, 1.21)
C	GG or GA	Yes	198	89	0.85 (0.71, 1.01)
D	GG	Yes	11	2	0.34 (0.10, 1.21)

a Overall effect of genotype 0.71 (95% CI 0.47, 1.07); overall effect of aspirin 0.83 (95% CI 0.70, 1.00).

b Deviation from additive model of interaction: D −(B + C −1) = 0.34 −(0.79 + 0.85 −1) = −0.30. Deviation from multiplicative model of interaction: D/(B • C) = 0.34/(0.79 • 0.85) = 0.51.

completed questionnaire and blood samples, and subsequently underwent one or more follow-up colonoscopies. Overall, both aspirin use and homozygosity for the intron 1 variant (IVS1 +317G/A) were associated with a reduced risk of adenoma recurrence (table II). The joint effect of aspirin use and homozygosity for the intron 1 variant was greater than would be expected on the basis either of an additive or multiplicative effect (table II).

In table II, we have presented genotype as a dichotomous variable. In a two-allele system, genotype is a three-level rather than a two-level variable, and the classification depends on information on the functional consequences of a variant. In this example, among aspirin users, the risk ratio for heterozygotes for the +315 intron 1 variant (GA genotype) was 0.87 (95% CI 0.63, 1.21), and that for those homozygous (AA genotype) was 0.38 (95% CI 0.11, 1.36) compared with those homozygous for the common alleles (GG genotype). In addition, in the interpretation of this study it is important to note that the 688 subjects included were from a total study-base of 1304 participants in a randomized trial of the effect of wheat bran fiber on adenoma recurrence. A large proportion (47%) of the 1304 subjects did not provide a blood sample. If refusal to provide a blood sample were associated with either aspirin use or genotype, this would have biased the results of the study. The investigators compared the characteristics of participants and nonparticipants, but did not observe any statistically significant difference. Thus, it seems unlikely that in this instance nonparticipation affected the internal validity of the study. More generally, nonparticipation is an important consideration in evaluating studies as it may affect both internal validity and generalizability.

A cohort study applied to the investigation of gene-outcome relations in those receiving therapy assumes interaction in the sense that no relation between gene and outcome would be expected in the absence of therapy, and that all those with an indication for therapy would receive it. An example of such a study is the investigation of the *CYP2C9* variants and outcome of warfarin therapy in 185 patients with atrial fibrillation, dilated cardiomyopathy, deep vein thrombosis, pulmonary embolism, or valve replacement.[49] Patients were identified through pharmacist-run anticoagulation clinics that they attended at regular intervals of 2 to 6 weeks. Blood samples from which the DNA was extracted were taken at enrollment and the data were obtained by retrospective chart review – therefore no patients were withdrawn or lost to follow-up. The main endpoints were: (i) serious (requiring treatment or medical evaluation) or life-threatening bleeding events; and (ii) anticoagulation status, as measured by time to therapeutic international normalized ratio (INR; i.e. within optimal range for a given indication), rate of above-range INRs (again defined in relation to the optimal range for a given indication), and time to stable warfarin dosage levels. The incidence of serious or life threatening bleeding events was higher in those with a variant allele than those without (relative risk 2.23, 95% CI 1.05, 4.77). In addition, hazard ratios were calculated, comparing the time to each of the endpoints between patients with at least one variant allele and those who did not have a variant allele. A hazard ratio (HR) of two would indicate that patients with a variant allele would, on average, experience the endpoint in half the time of those who did not have a variant allele, an HR of one would indicate no difference, and an HR of 0.5 would indicate double the time to experience of the endpoint of those who did not have a variant allele. There was no difference between the groups in terms of time to therapeutic INR, but patients with a variant allele experienced an above-range INR sooner (HR 1.40, 95% CI 1.03, 1.90), and required more time to achieve stable warfarin dosage levels (HR 0.65, 95% CI 0.45, 0.94; median of 95 days longer).

In studies of the joint effects of genotype and drug treatment, large numbers of subjects typically have to be enrolled in order to have adequate statistical power to detect gene-drug interaction.[70] In large-scale cohort studies of treatment, it is obviously a challenge to collect data not only on the specific drug used, the dose, when it started to be used, whether and when its use was stopped, and similar details for each period of use, but also on potentially confounding factors, because of the large number of subjects involved. Assessment of medication use at several points in time can be obtained, provided that resources are available. This en-

ables changes in drug use and other relevant exposures to be monitored. This affords the investigator alternative analytic approaches, including analysis focused on use of drugs at the beginning of the study; analysis focused on more recent use; or anaylsis of a summary measure of repeated drugs use over time.

Theoretically, samples for genotyping could be collected at any time during follow-up of the cohort, but bias is least likely to be introduced if this is done at enrollment. If not collected at enrollment, it would have to be assumed that loss to follow-up was not related to genotype. In addition, outcome (e.g. death) might affect the ability to collect the samples.

In a number of countries (notably the North America and Europe), there exist population-based disease registries that can be linked to death records, for example, such registries are well established for cancer, and have been used to investigate variations in cancer survival.[71,72] When survival data can be linked to treatment information, and to tissue samples, there is then the opportunity to investigate gene-treatment interaction in relation to survival. In theory, the linkage could be done without approaching the patient, and so the mechanisms for dealing with data protection issues may differ from other studies. For example, after linkage, the data and samples could be anonymized.[73,74] The possible advantages of this approach include the generalizability of the results, because of the population basis of the register, potential high quality treatment information, and relatively low cost. So far, this opportunity does not appear to have been utilized.

The principles of the cohort design applied to the investigation of gene-drug interaction in disease etiology are the same as those for a cohort study of treatment outcome in patients with a particular condition. There is a need for large studies, and this in turn raises challenges in terms of data and sample collection, which may be addressed, at least in part, by the use of nested case-control or case-cohort studies. However, some of the practicalities are different as, in general, healthy individuals are recruited and followed up in a population setting. A number of large-scale biobanks are being established to investigate interactions between various aspects of lifestyle and genetic factors in the etiology of chronic disease.[75-80] However, they are not primarily aimed at assessing the effects of the long-term use of specific medications. Bias would result if there was loss to follow-up that was differential by exposure or genotype (see below).

2.3 Case-Control Studies

A commonly used design in investigating the occurrence of adverse drug reactions and gene-environment interactions in disease etiology is the case-control study. In a case-control study, individuals who have recently developed an outcome (i.e. specific disease or adverse drug reaction) and a sample of individuals without the outcome being investigated are recruited and information is then collected on potential risk factors during a specified reference period prior to the onset of disease. A DNA sample would be taken at recruitment. The results of a case-control study of genotype-treatment interaction, in their simplest form, can be presented in a two-by-four table. As already noted in relation to the example of the joint effects of *ODC1* genotype and aspirin use, this presentation forces genotype to be classified as a dichotomous variable. The odds ratio closely approximates to the relative risk provided that the cases are newly incident and that the cases and controls are selected from the same source population.[81,82] If the cases are not recent, it is possible that any observed association is in part due to an effect on survival as well as on etiology. In particular, cases with shorter survival times will be under-represented. If duration of survival is related to etiologically relevant factors, the inclusion of prevalent cases will distort the association observed, depending on the nature of the relation.[83]

Hwang et al.[63] have presented calculations showing that when the proportion of subjects receiving a drug and the proportion with a certain genotype lies in the range 30–70%, around 200 cases and 400 controls would be adequate to detect an odds ratio of genotype environment interaction greater than 4 with 80% statistical power. However, misclassification of genotype and exposure reduce the statistical power to detect interaction.[84] This means that larger sample sizes are needed (for further discussion of sample size issues, see section 3.4).

An example of a case-control study is the investigation of the possible interaction between HRT and the Factor V Leiden (*F5*) gene polymorphism, and the risk of venous thrombosis.[85] This study looked at women admitted to hospital in the Oxford area of England with a first episode of deep-vein thrombosis. Up to two control women per case were recruited from women admitted to hospital for diagnoses unrelated to thrombosis and HRT. DNA was obtained from 77 of 80 cases who consented, and 163 of 171 controls. There was an increased risk for venous thrombosis associated with both Factor V Leiden (OR_g = 3.9; 95% CI 1.3, 11.2; adjusted for age and district of admission) and HRT (OR_t = 3.2; 95% CI 1.7, 6.0), and the combined effect (OR_{tg} = 15.5; 95% CI 3.1, 76.7) was greater than the sum of the individual effects.

Prevalence data enable assessment of the potential for preventing disease by targeting drug-based interactions in subgroups of the population defined by the presence of genetic variants. Control series from case-control studies are a potential source of data on the population prevalence of genetic variants affecting response to drugs.[86] However, if there were publication bias in favor of positive associations between specific genetic variants and disease[87] it is possible that genotype frequencies in the general

Table III. Relative risk (RR) and case-only analysis of interaction between the serotonin receptor 2A (*HTR2A*) +102T/C polymorphism genotype and antidepressant therapy in the occurrence of adverse effects and discontinuation of therapy[60] Reference group comprises subjects with genotype postulated not to increase risk (TT or TC) who received standard therapy (mirtazepine)

Treatment	*HTR2A* genotype	Cases[a]	No. of patients	Incidence (%)	RR (95% CI)[b]
Mirtazepine	TT or TC	13	86	15.1	1
	CC	6	38	15.8	R_g = 1.04 (0.43, 2.54)
Paroxetine	TT or TC	13	81	16.0	R_t = 1.06 (0.52, 2.15)
	CC	19	41	46.3	R_{tg} = 3.07 (1.68, 5.58)

a Case only odds ratio = (19 • 13)/(6 • 13) = 3.17 (95% CI 0.83, 12.74).

b $R_{tg}/(R_g \times R_t)$ = 2.78.

R_g = relative risk of adverse effects and discontinuation of therapy in subjects with genotype postulated to increase risk (CC) who received standard therapy, compared with reference group; **R_t** = relative risk of adverse effects and discontinuation of therapy in subjects with genotype postulated not to increase risk (TT or TC) who received novel therapy (paroxetine), compared with reference group; **R_{tg}** = relative risk of adverse effects and discontinuation of therapy in subjects with genotype postulated to increase risk (CC) who received novel therapy (paroxetine), compared with reference group.

population would be under-estimated. In addition, many early studies were based on convenience samples and, not infrequently, little information was given on sample selection.[54,55,88,89]

2.4 Other Designs

Variants of these designs have been proposed to reduce the amount of sample assays needed, or to deal with the potential problem of population stratification. These are listed in this section.

2.4.1 Nested Case-Cohort and Case-Control Study

If the cost of genotyping (or other assay/data collection) is an issue and the outcome is dichotomous, carrying out studies on samples of participants in a cohort study (or a RCT) is a means of cutting the cost of genotyping. In a nested case-cohort study, patients who develop the outcome of interest during the follow-up period are the cases, and controls are sampled to match the cases on a temporal factor, such as age, from patients who do not develop the outcome of interest. Archived specimens are then retrieved for these patients and genotyped. The main comparisons are within time-matched sets of cases and controls.[90]

In a nested case-control study cases are identified in the same way as in a nested case-cohort study, while controls are a random sample of the cohort. Again, archived specimens are analyzed. The effect of age, which is the key time variable, is controlled for in the analysis only. A major advantage of the case-cohort design is that the same comparison group can be used for several different (dichotomous) outcomes.

2.4.2 Case-Only Studies

Another design is the case-only study, which has been proposed as a means of investigating gene-environment interaction in disease etiology.[91-94] In a case-only study nested within a RCT, the outcome of interest would have to be dichotomous, and subjects developing this outcome during follow-up would be defined as cases. The drug under investigation is equivalent to the environmental exposure. Because the treatment to which a patient is allocated and genotype are independent within a RCT, the case-only odds ratio is also a measure of gene treatment interaction. When applying these principles to the study of *HTR2A* genotype and antidepressant therapy,[60] the ratio of relative risks indicating gene-treatment interaction is 2.78 (table III). The case-only odds ratio is 3.17.

In theory, a case-only design could also be nested within a cohort study. This would enable departure from multiplicative effects of gene and drug to be assessed, but the key assumption of independence of the distribution of gene and drug use in patients 'at risk' of the outcome of interest[95] may not be satisfied, in contrast to the situation in a RCT. It is possible that genetic factors influence the ability to tolerate therapy, so independence between genotype and treatment may not be a valid assumption.

If the assumption of independence between genotype and treatment is applied to the general population at risk of the disease, then the gene-treatment interaction could be estimated in any population-based series of cases with a specific disease. This approach would obviate the problems of selecting controls, of potential differential recall of treatment information between cases and controls, and of population stratification, as well as offering advantages in terms of study logistics.

In a case-control study, the odds ratio relating treatment and genotype among case subjects only (OR_D) is a function of the odds ratio for the genotype alone (OR_g), the treatment alone (OR_t) and their joint effects (OR_{tg}). When the assumption that genotype and exposure are independent is valid in the general population at risk of the disease, this would also be the case for any representative sample of controls, and an estimate of $OR_{tg}/(OR_g \times OR_t)$ can be obtained from data in cases only.[91-94] Thus, an estimate of depar-

ture from the multiplicative effects of treatment and genotype can be obtained. However, this design appears to be highly susceptible to the validity of the assumption of independence of genotype and exposure.[95] In the example of Factor V Leiden and HRT in relation with venous thrombosis,[85] the ratio of odds ratios of gene-treatment interaction (crude odds ratio calculated from the data in table II of Rosendaal et al.[85] = 15.5) to the product of the odds ratios of treatment-only (3.2) and genotype-only (3.9) is 1.24, while the case-only odds ratio is 0.97. The control-only odds ratio is 0.78, which is suggestive of absence of independence of genotype and treatment, and could account for the discrepancy between the results of the case-control and case-only analysis.

It seems highly plausible that there would be a relationship between drug use and some genotypes. First, as already noted, there may be an effect of genotype on an individual's tolerance of a drug. Secondly, the genotype may be associated with the absence of symptoms for which a drug is taken; for example, Feigelson et al.[96] observed that women with the *CYP17A2**A2 genotype were only half as likely to use HRT as women with the *CYP17A2**A1 genotype. A possible explanation for this observation is that women with the A2 genotype may have higher estrogen levels prior to the menopause and so may suffer less from menopausal symptoms than women with other genotypes, as they would have had lower estrogen levels initially.[97]

2.4.3 Before-After Comparisons in the Same Patients

Another design option is based on examining the same patients before and after receiving a drug. This design is strongest in the context of a RCT. An example is an investigation of whether ERα (*ESR1*) polymorphisms modified the effects of HRT on high-density lipoprotein (HDL)-cholesterol compared with placebo in 309 women with coronary artery disease.[98] Of the ten polymorphisms examined, homozygosity for the less common alleles of four SNPs within intron 1 of the *ESR1* gene were found to result in higher increases in HDL-cholesterol than for the other genotypes in response to HRT. For example, in women who had the IVS1−401 C/C genotype, the increase in HDL-cholesterol in response to HRT was 13.1 mg/dL, more than twice the 6.0 mg/dL observed in the other women (p-value for interaction 0.004). The before-after comparison approach could also be carried out in the context of a cohort study, but would be more subject to confounding than in a RCT.

2.4.4 Family-Based Case-Control Studies

Concern about the possible effects of population stratification has stimulated development of family-based case-control designs, which essentially eliminate potential confounding from this source.[99,100] The most commonly used examples of such designs involve the use of siblings or parents as controls.

Case-Sibling Control Studies

Sibling controls are derived from the same gene pool as cases. However, selection bias could result because a sibling may not be available for every case – bias would also arise if determinants of availability (e.g. sibship size) were associated with genotype. To investigate gene-treatment interaction in the case-sibling control design, older siblings are preferable as controls so that the comparable reference periods for drug use, such as up to the age of diagnosis of the outcome in the case, or a fixed interval prior to this to allow for latency, can be considered for cases and controls.[101] However, the potential periods during which subjects could have used the drug will be systematically earlier in calendar time for older sibs than for cases, which could lead to confounding as a result of secular trends in drug use or differential recall, with the sib controls having to recall drug use further back in time.

In comparison to a study in which unrelated controls were used, while a study using an equivalent number of sibling controls has less statistical power to detect the main effect of genotype because of over-matching on genotype,[102] it may have greater power to detect gene-drug interaction.[101] This is because the most informative contrast is between genotype-concordant exposure-discordant pairs, and a higher proportion of such pairs occurs in case-sibling control sets than in case-unrelated control sets. However, as the amount of concordance for drug use among siblings increases, the advantage of the case-sibling control design over the case-unrelated control design decreases. The case-sibling design is most efficient when the gene has a dominant mode of inheritance.[103]

Case-Parental Control Studies

In case-parental control study, interactions between genes and drug use of the case, or a parent of the case, can be examined using log-linear modeling[104] or a conditional logistic regression approach.[105] This approach requires an assumption of independence of gene and drug use, conditional on the parents' genotypes.[101] This assumption is less stringent than the assumption of independence of genotype and drug use in the case-only design. The statistical power of case-parental control studies to detect gene-drug interactions generally is greater than that of case-unrelated control studies, particularly when the gene has a recessive mode of inheritance.[103] The need to obtain samples from parents is a practical problem limiting the applicability of the design for diseases of late onset. An additional factor that may limit participation of eligible cases is reluctance to reveal disease status to parents of adult subjects. Adults with conditions such as infertility about which they may feel sensitive may not wish their parents to know that they have the condition.

3. Methodological Issues in Excluding Non-Causal Explanations for Gene-Drug Interactions

Non-causal explanations for observed interactions, including gene-drug interactions and associations, include bias, confounding, and chance. For a single study (as distinct from integrating evidence from multiple studies), the major types of bias are selection bias and information bias.

3.1 Selection Bias

Selection bias occurs when the subjects included in a study are not representative of the source population, and/or when there are selective losses from the study population prior to data analysis.[106] If this were differential by genotype and/or drug exposure, it would distort the observed effect of gene-drug interaction. If this were non-differential by these factors, the internal validity of the estimate of the gene-drug interaction would not be compromised, but the estimate might not accurately predict the results in another group of patients.

Selection bias is one of the main potential biases of the case-control design and may arise from the inappropriate choice of controls or differential participation rates between cases and controls.

In the example relating to the interaction between HRT and the factor V Leiden polymorphism, controls comprised subjects hospitalized for reasons unrelated to thrombosis or HRT, including diseases of the eye, ear, skin, respiratory and alimentary tracts, kidneys, bones, and joints, or trauma.[25] In a hospital based case-control study, Wacholder et al.[107] noted that even if exposure (therapy) or genotype, or both, are associated with the control disease, a departure from multiplicative effects can be estimated without bias. However, this is not the case for departure from additive effects. Although including controls with more than one type of disease might reduce a bias resulting from one disease being associated with exposure, genotype or both, pooling of controls with different diseases can lead to bias in assessing departure from multiplicative interaction, even if there is no such interaction in each individual disease-specific control set.[107] These issues apply more generally to the estimation of both interaction effects and gene-disease associations when controls are not selected from the same source population as the case-subjects.[107,108] An example of the potential problem of selecting controls who do not represent the population from which case-subjects arise is represented by the divergence in odds ratios for the association between colorectal cancer and the *GSTT1* null genotype,[109] when the different control groups were analyzed.[55]

Selection bias may also occur as a result of non-participation that is differential between cases and controls. This would bias the odds ratio for gene-disease association only if the difference were related to genotype, and the measure of gene-drug interaction if the difference were related to genotype, drug use, or both. There has been concern about a decline in participation rates,[110] especially in population-based studies. When participation rates are low, those selected as population controls could be largely those who are likely to be at home for some reason when contacted. Therefore, in studies utilizing population controls, it is critical to demonstrate as large a response as possible from those eligible in the base population. Information on the potential effects of low participation rates is limited.[111]

In cohort studies in general, bias would result if loss to follow-up was differential by exposure or genotype, for example, if both loss to follow-up and drug use varied by socio-economic status, or if loss to follow-up and genotype prevalence varied by ethnic group. Drug use varies by socioeconomic status; for example, cancer patients living in deprived areas in Scotland are less likely to receive chemotherapy than patients from more affluent areas.[112,113] In the US, aspirin and β-blockers are less likely to be administered to poor patients with myocardial infarction,[114] and statin use is less frequent in low-income patients with diabetes mellitus compared with higher-income patients.[115]

In regard to cohort studies of disease etiology, limited data on loss to follow-up tend to be presented. In a longitudinal study of cognitive aging, those who did not return for follow-up had lower educational levels than those who did return.[116] In studies in the US, members of minority groups tend to have higher dropout rates than Whites.[117] In a study of Black women in the US, those who were lost to follow-up tended to be less well-educated than those who remained in the study.[116] A related issue concerns the return of incomplete information during follow-up, i.e. item non-response. This has been shown to be associated with subsequent loss to follow-up.[118]

3.2 Information Bias

Information bias occurs as a consequence of errors in assessing factors of interest, in this case genotype or drug use, or both, or in the assessment of outcome. It is differential when the assessment of the factors of interest is influenced by the outcome under investigation, or vice versa. Both differential and non-differential biases can result in over- or underestimation of an interaction effect.[119] When genotype and drug use are independent in the source population, and the errors in the assessment of each are independent, both differential and non-differential misclassification of a dichotomous factor tend to underestimate departure from a multiplicative gene-drug joint effect.[120] The impact of misclas-

sification on departures from additive effects are difficult to predict.[84]

Factors affecting the potential extent of misclassification of genotype include the types of samples, timing of collection, and the method used for genotyping.[121,122] Quality control procedures are important in assessing the extent of possible misclassification and the extent to which this might be differential (e.g. whether laboratory staff are blinded to drug treatment or outcome). Factors affecting the potential extent of misclassification of drug use include the method of obtaining the information, and its validity and reproducibility.[123] The extent of misclassification of drug use is likely to vary according to the study design. It is most likely to be minimized in RCTs, and less in series of patients with a specific condition under prospective follow-up than in subjects in a population-based study of disease etiology or adverse drug reactions. In the population-based setting, there may be a higher level of non-differential error in a large cohort study than in a case-control study, but a higher level of differential error in the latter.

In regard to cohort studies, a summary measure of drug use over multiple timepoints should be less subject to random misclassification than would a measure at one point in time. However, the interval between repeat assessments is important. People may start and stop taking medications between assessments, and some of these may have stopped taking medications because of adverse effects. This has been suggested as a possible explanation for the discrepancies in results between at least one cohort study of HRT and RCTs.[69] Data collection at multiple timepoints is resource intensive. While the accuracy of drug use as reported by patients may be of concern, it would be a major undertaking to check medical records.

In a study in the Netherlands in elderly people, there was good agreement between patient-interview data and pharmacy records for prescription-only cardiovascular drugs.[124] In a study of HIV-infected patients in the US, there was fair to substantial agreement between patient-interview data, medical records, and planning records for specific mediations, but lower agreement for drug classes.[125] While use of automated databases potentially could address this, such databases do not include over-the-counter drug use.[69] In addition, data on potential confounders in such databases tend to be limited.

Considerable attention has been paid to recall bias in case-control studies. In general, however, assessments of recall of medication use assessed by interviews (compared with medical records) have not shown differences in accuracy of recall between cases and controls.[126-128] It has been suggested that the likelihood of recall bias may be greater when recall is poor in general.[129] However, this was not apparent in a systematic review of empirical studies of recall bias published between 1966 and 1990.[130] In addition, investigations of the theoretical impact of recall bias for dichotomous exposures shows that even severe recall bias causes only weak to moderate spurious associations.[131-133]

3.3 Confounding

Confounding in the investigation of gene-drug interaction could arise (i) because of differences in ethnic origin between cases and controls; (ii) if a gene other than the one of interest were associated with both the gene and the outcome being investigated, i.e. as a result of linkage disequilibrium; (iii) if a factor other than the drug of interest were associated with both the drug and outcome being investigated; and (iv) if the gene of interest affected exposure to factors associated with the outcome being investigated.

3.3.1 Population Stratification

Population stratification is the presence within a population of subgroups between which allele/genotype frequencies and disease risks differ. For example, both allele/genotype frequencies and disease risks may differ by ethnic group within a population. In addition, population stratification can arise because of differences between groups of similar ethnic origin but between which there has been limited admixture, such as in isolated populations. For example, a population might comprise the descendants of waves of immigrants from the same source who differ genetically because of founder effects. The differences may then be apparent because insufficient time has elapsed for mixture between the groups. When the groups compared in the study differ in terms of the proportions of the population subgroups, there is the potential for an association between the genotype and disease being investigated to reflect the fact that genotype is a marker for the population subgroup rather than to be a causal association. Population subgroup is a confounder in this situation as it is associated with both genotype frequency and disease risk.

In the investigation of gene-drug interaction, population subgroups could be a potential problem both within strata of drug use and because drug use may also vary by population subgroup. The most frequently cited examples including the association between type 2 diabetes mellitus and the immunoglobulin allotype *Gm*$^{3;5,13,14}$ haplotype among residents of the Gila River (American) Indian community that was used to present the potential problem,[134] the relationship between the dopamine receptor locus (*DRD2*) A1 allele and alcoholism,[135] and between *CYP3A4* and prostate cancer in African Americans.[136] These examples have helped fuel controversy as to whether population stratification represents a fundamental problem for association studies, or whether it is part of more general issues about rigorous application of epidemiologic study design principles.[137-139]

In an exploration of possible population stratification in US studies of cancer among non-Hispanic Americans of European descent, the effect was considered unlikely to be substantial when epidemiologic principles of study design, conduct, and analysis were rigorously applied.[140] Similar conclusions were reached using data from case-unrelated control studies of non-Hispanic US Whites with hypertension or type 2 diabetes, and Polish subjects with type 2 diabetes.[141] However, these conclusions may not apply to all ethnic groups. For example, variations in the frequency of certain genotypes in African Americans appear to be much wider than those observed in persons of European origin and, therefore, the possibility of stratification may be higher.[142] Evidence of a population stratification effect was weak in data from a case-unrelated control study of hypertension in African Americans, but this was no longer apparent when the study was restricted to persons with US-born parents and grandparents.[141] Millikan[143] reported that bias was minimal in gene-disease associations in studies of African Americans in which differences in ethnic composition were not taken into account, and Wang et al.[144] did not identify substantial bias in hypothetical case-control studies of a candidate gene for prostate cancer in admixed populations such as African Americans.

One of the concerns regarding emphasis on large studies of gene-disease associations and gene-environment interaction is that the potential effect of stratification increases with sample size.[145] Freedman et al.[146] found no empirical evidence of stratification in data on 24–48 SNPs from 11 case-control and cohort studies in the US, Poland, and Portugal. However, when the number of SNPs was quadrupled, and the sample size increased by a factor of 5–6, statistically significant evidence of stratification was found in one of the studies in which a case-cohort design had been used. The effect of the degree of stratification in this study (which was in African Americans) would be to inflate the chi-square statistic for association by a factor of two in a study with 1000 cases and 1000 controls, and by a factor of 2–5 in a study with 2000 cases and 2000 controls. Thus, there is controversy about the potential importance of population stratification for population-based studies of gene-disease association and for studies of gene-environment interaction, and it seems important that potential effects are investigated in a variety of settings.

One approach to minimizing the potential problem of population stratification when unrelated controls are used is to measure and adjust for genetic markers of ethnicity that are not linked to the disease under investigation.[70,147] Such adjustment may not correct for stratification if too few loci are used.[145] This would be expected to control for ethnic variation in disease risk attributable to genetic factors. However, residual confounding from other sources of ethnic variation in disease risk would be a potential issue.[139,148]

3.3.2 Linkage Disequilibrium

Linkage disequilibrium is the tendency for the alleles of two separate but already linked loci on the same chromosome to be found together more than would be expected by chance in the general population. In consequence, when an allele at a specific locus appears to be associated with a disease, the question arises as to whether the allele is causal, or whether the association exists only because the allele is linked to a truly causal allele at another locus. Linkage disequilibrium depends on population history and on the genetic make-up of the founders of that population.[149,150] Linkage disequilibrium varies between populations[150] and, therefore, potentially could be a source of heterogeneity between studies of gene-drug interaction. As discussed in section 1, it may be useful to type several polymorphisms throughout a candidate gene in order to construct haplotypes, which could then be tested for association with the phenotype of interest. The increasing availability of mapped SNP markers offers the opportunity for such an approach and presents methodological challenges.[151-153] A potential limitation of this approach is that the effect of a true functional variant might be diluted when haplotypes rather than loci are the units of analysis. The RCT is just as susceptible to this problem as other study designs. However, confounding by unlinked genetic loci is less likely to occur in RCTs and population-based studies of disease etiology, because of Mendelian randomization.[154,155]

3.3.3 Other Sources of Confounding

Confounding for other sources is potentially minimal in RCTs provided the randomization is successful. For the other designs, it is necessary to collect data on potential confounders and assess these in data analysis.

It may be difficult to determine whether a drug modifies the relationship between a gene variant and a disease, or whether it confounds that relationship. For example, it has been suggested that the polymorphism in the 5′- untranslated region in *CYP17* may be involved in the etiology of breast cancer.[156] If this were so, it might be expected that it would modify the relationship between markers of endogenous hormone levels, for example age at menarche, age at menopause, propensity to use hormone replacement therapy,[96] and the exogenous factors that influence hormone levels, and disease risk. Four studies have investigated possible interactions between post-menopausal hormone use and *CYP17* and breast cancer,[157-160] one of which found evidence suggestive of an interaction.[157] If *CYP17* genotype affects estrogen biosynthesis and hormone levels, it may also affect markers of endogenous hormone levels, including propensity to use HRT, thereby involving these factors in the causal pathway between

CYP17 and breast cancer. Statistically, these could be viewed as confounders of the relationship between *CYP17* and breast cancer. However, it would be inappropriate to consider these as potential modifiers of the relationship between *CYP17* and breast cancer. We stress that this is only an illustrative example and that there is considerable uncertainty about the relation between *CYP17* genotype and use of HRT.

3.4 Chance

Both type I and type II error rates depend on sample size and the critical value (α-level) for the rejection of the null hypothesis. In general, the larger the sample size, the better the precision of the estimate of gene-treatment interaction, and the more likely that an interaction of interest (if real) would be detected. In most of the earliest studies of gene-disease associations, detection of gene-environment interaction was a secondary aim. Most were based on a candidate-gene approach, with strong biologic evidence of the importance of the genes and some evidence about the functional effects of variants of the genes.[161] Most were of modest size, and while their statistical power was adequate for the detection of gene-disease associations, it was inadequate to detect gene-environment interactions. To test for departures from multiplicative effects, it has been noted that study size should be at least four times larger than needed to detect only the main effects of the individual factors.[162] When non-differential misclassification of drug treatment, genotype or both is taken into account, this in turn increases the required study size.[84]

It is now becoming feasible to test several SNPs and haplotypes in hundreds or thousands of genes whose function is unclear or unknown.[161,163] Indeed, it has been suggested that whole-genome scanning may be a means of identifying individuals with a higher risk of experiencing an adverse reaction to a drug.[6] An approach of assessing interaction of every genotype with a drug for a range of interaction models would generate a large number of false-positive results. Increasing the significance level as a safeguard against type I error is unlikely to solve the multiple comparison problem because, in studies of gene-environment interaction in general, there has been limited statistical power to detect interactions because of modest effects and limited sample sizes.[108]

In addition, one of the problems with adjusting for multiple tests is that it is not clear how many hypotheses are being tested.[163,164] Therefore, there is increasing interest in Bayesian approaches. Colhoun et al.[164] suggested that in candidate-gene studies of complex traits, the significance level should be reduced to about 0.00005, on the basis that this would achieve a posterior odds ratio of 20 : 1 that an association was valid. It is not clear whether this should be lower for gene-treatment interaction. On the one hand, many potential interactions could be tested, alternatively there may be more data on the biologic basis for a gene-treatment interaction than for other types of interaction. Another approach is the application of empirical Bayes methods involving the simultaneous conduct of large number of tests.[163,165] Wacholder et al.[161] proposed the assessment of the false-positive report probability, calculated on the basis of the prior probability that a gene-disease association is real, statistical power and the observed p-value. In addition, they suggested that the stringency of this probability should depend in part on the magnitude of the negative consequences of potentially incorrect decisions. For example, it might be less stringent for rare diseases or small initial studies, but more stringent for large studies or pooled analysis that attempted to be more definitive evaluations. One problem with this suggestion is that it would make the integration of evidence from different studies very difficult. Other issues include the problem of false-negatives (Wacholder et al.[161] note that false-negative report probability can be considered) and possible over-emphasis on controlling the false-positive rate.[163]

4. Issues in Epidemiologic Analysis

In studies of gene-drug interactions, many hypotheses of interaction can potentially be tested. There are a range of different potential cross-classifications of drug treatment and genotype. Drug treatment can be classified as 'ever' versus 'never', 'use of drug A' versus 'use of reference drug B', as a continuous variable, and further defined on the basis of period of treatment. Once multiple categories of dose are defined, many different dose-response models can be tested in the data. In a two-allele system, heterozygotes could potentially be considered separately, included in the reference category with homozygotes for the common variant, or grouped with homozygotes for the rarer variant. This is more complex for multi-allelic systems. It may be possible to classify genotypes on the basis of putative functional effects,[26] but this may not be straightforward.

In some studies, genotype has been inferred on the basis of a phenotypic test. A potential advantage of this approach is that it may provide a direct measure of the functional significance of the underlying genetic polymorphism, but such an assay only provides a measure at a single point in time, and potentially may be distorted by systematic influences (e.g. effects of disease stress on metabolism, inducing factors) as well as random measurement error. Contrasting results between studies based on phenotypic and genotype assays have been observed, for example, for the acetylator polymorphism and colorectal cancer.[54] Clearly, model specification becomes more difficult as more environmental factors (and levels of exposure) and genes (and alleles) are included.

Therefore, the distinction between *a priori* hypotheses and hypothesis generation is crucial.

The simplest framework for analysis of gene-drug interaction is the two-by-four table.[166,167] In the examples discussed in this paper, we have focussed on relative risks to show the relationships between the different study designs. However, other measures of risk can be calculated using this framework, including absolute risk (and the related measures of absolute risk reduction and number needed to treat). Absolute risks can be estimated from RCTs and cohort studies. For example, in the RCT of alternative antidepressant drugs[60] described above, when genotype is not taken into account the number of patients needed to be switched from paroxetine to mirtazepine to prevent one discontinuation of therapy resulting from an adverse effect is 9 [1/(0.262 −0.153); see table III]. When *HTR2A* +102T/C genotype is taken into account, the number of patients whose therapy would need to be changed in those homozygous for the C allele is 3, whereas in those with the TT or TC genotype it would be 111. The decision to change therapy for all patients, or only for a genetically defined subset, would depend on other outcomes of therapy, acceptability of genotyping to patients, and, possibly, costs. Absolute risks cannot be estimated in case-control studies unless complementary follow-up data are available.

The population attributable fraction is a measure of the public health consequences of a relation between a risk factor and a disease (including an adverse reaction). It takes account of the strength of association between the risk factor and disease, and the prevalence of the risk factor in the population. It can be estimated from cohort studies and, provided that cases are an unbiased sample of cases in the source population, in case-control studies.[168] In the example of the interaction between HRT and Factor V Leiden, in the etiology of a first episode of deep vein thrombosis,[85] the population attributable fraction for HRT was 16%, for Factor V Leiden 4%, and for their joint effects 1%. Although the interaction effect is large, the public health impact is low because the frequency of the susceptibility genotype is low. For studies of drug use and genotypes common in the population, the population attributable fraction of interaction is large even if the interaction effect is only moderate.[169]

Analytic methods to test for gene-environment interactions are still under development; for example, the application of hierarchical models is being explored.[170,171]Three common methods potentially could be used to assess the statistical significance of gene-drug interactions, defined as departure from multiplicative effects.[108] First, an interaction term could be introduced into a logistic model and then a test (the Wald test) performed to determine whether the linear trend in risk associated with drug exposure, such as dose, frequency, or duration of use, is significantly different between the dichotomous categories of genotype. The utility of this approach depends on whether a linear trend in risk is appropriate. Secondly, a cross-product term for each combination of genotype and drug exposure category (omitting the combination for the reference category) could be introduced into the logistic model and the p-value for the difference in the log-likelihood between this model and the model containing the main effect estimates for the genotype and drug exposure variables could be determined. A potential problem with the likelihood ratio test for interaction is that in situations in which the data depart from an ordered trend, the likelihood ratio test may give a significant result because the cross-product terms improve the fit of the model to the data. Therefore, assessing gene-environment interaction solely by screening for the level of significance of a formal test for interaction should be avoided. Thirdly, estimates of drug effects could be compared between genotype strata. However, the finding of a significant effect in one or more strata but no significance in at least one other stratum does not constitute statistical evidence of interaction. Often such a pattern has been observed when inadequate power exists in one of the strata and a formal test of statistical interaction has been performed to assess the strength of the evidence for interaction. Little work has been done on testing for departures from additive models of genetic and environmental effects.[51,120]

Although most studies have not considered multiple candidate genes,[59] it is unlikely that a single polymorphism will determine the response to a drug.[2,31] As the numbers of factors postulated to interact increases, the number of combinations of factors grows quickly (2^n for n dichotomous factors), and there are likely to be sparse data for many of the combinations. A possible solution is pooled analysis. Pooled analysis of data on individual subjects from multiple studies involves obtaining and re-analyzing the primary data, as distinct from aggregating published information.[172] An additional problem is that interaction between multiple factors may not adequately be described as departures from multiplicative or additive effects.[167] It has been suggested that neural networks may be used to investigate complex interactions because they are less dependent on prior model specification than other methods.[173-175] However, they may be limited in their ability to estimate dependence among risk factors.[173,175]

5. Interpretation

Gene-treatment interaction is complex,[2] but many responses to drugs appear to be simpler traits than are common diseases.[176] As already discussed, it is likely that multiple genes operating in pathways will determine response to a drug. A single gene variant may be associated with an increased risk of one disease, but with

decreased risk for another. For example, the *MTHFR* 667C→T variant is associated with increased risk for coronary heart disease,[177] but a decreased risk for colorectal cancer.[178] Considerable importance is being accorded by many commentators to data on gene function and mechanisms in making causal inference about gene-disease associations and gene-environment interaction.[179-185] On the other hand, there has been concern that mechanistic evidence might be identified selectively, rather than by systematic review, and used to reinforce an assertion of causality.[186] Moreover, Page et al.[185] note that biological plausibility for an association (or interaction) that is valid may not be apparent because of our limited knowledge of gene function and mechanisms. It seems crucially important that connections are made and maintained between research efforts on gene function and mechanisms, and epidemiologic research relevant to gene-treatment interaction. In addition, systematic review principles should be applied to data on gene function and mechanisms, as well as epidemiologic studies.

6. Use of Epidemiological Evidence in the Development of Pharmacogenomic Testing

Haddow and Palomaki[187] have described a framework for evaluating data on emerging genetic tests, based on four components – *a*nalytic validity, *c*linical validity, *c*linical utility and *e*thical, legal and social issues (ACCE). In the context of gene-drug interaction, clinical validity would define the ability of information on the combination of genotype and drug exposure to predict outcome. Clinical utility defines health outcomes (positive and negative) that would be associated with the introduction of a test into practice. Evidence of clinical validity and clinical utility will be derived from well-designed and conducted epidemiological studies, including RCTs. It is important to integrate evidence across studies.

Goldstein[153] has emphasized the importance of making reported associations between genetic variants and drug responses as secure as possible. This requires exclusion of non-causal explanations and determining whether the associations can be replicated. In 2001, an expert panel sponsored by the Centers for Disease Control and Prevention and the National Institutes of Health developed guidelines and recommendations for the reporting, evaluation and integration of data from gene-disease association studies and evaluations of genetic tests. Conclusions and recommendations from this workshop have been published.[108,121,188]

In considering clinical utility, it is important to bring together data on a range of outcomes by methods such as cost-effectiveness and cost-utility analyses.[189] These analyses are highly dependent on the epidemiologic data on gene-drug interaction, gene-disease association, and genotype prevalence. For example, one factor that may have a marked impact on cost-effectiveness and cost-utility is genotype frequency in the patient group or population in which the drug would be applied. In an evaluation of thiopurine *S*-methyltransferase (*TPMT*) genotyping in children with acute lymphoblastic leukemia receiving mercaptourine therapy, when the frequency of the genotype associated with severe hematopoietic toxicity was 0.3%, the incremental cost-effectiveness ratio (ICER) was highly variable and exceeded $US50 000 for many scenarios. In comparison, when the genotype frequency was 1%, there was less variability in the ICER and it was under $ US 50 000 in virtually all scenarios.[24,189]

7. Conclusions

There is a large gap between the promise of what the Human Genome Project could provide and the application of genomic information to improve the use of drugs in the management and prevention of disease. The epidemiologic approach enables the systematic evaluation of the risks and benefits of potential targeting of drug treatment on the basis of genomic information, both in series of patients and in the general population, and high quality epidemiologic data are the key to evaluations of cost-effectiveness and cost-utility analyses.

No one study design is ideal for every situation, in consequence, there is a need to employ a range of different designs to establish a portfolio of evidence and to carefully consider the extent to which non-causal explanations for interactions can be excluded. In view of the likely complexity of gene-drug interaction, pooling of data across studies is likely to be needed to have adequate statistical power to test hypotheses. Integration of evidence is important for replication of observed gene-drug interactions and for cost-effectiveness and cost-utility analyses. Ideally, data on interactions will be available for several outcomes to enable cost-effectiveness and cost-utility analysis as a basis for policy.

Acknowledgements

The authors have provided no information on sources of funding or on conflicts of interest directly relevant to the content of this review.

References

1. Collins FS, Green ED, Guttmacher AE, et al. A vision for the future of genomics research. Nature 2003; 422 (6934): 835-47
2. Nebert DW, Jorge-Nebert L, Vesell ES. Pharmacogenomics and "individualized drug therapy": high expectations and disappointing achievements. Am J Pharmacogenomics 2003; 3 (6): 361-70
3. Collins FS, McKusick VA. Implications of the human genome project for medical science. JAMA 2001; 285 (5): 540-4
4. Emery J, Hayflick S. The challenge of integrating genetic medicine into primary care. BMJ 2001; 322 (7293): 1027-30

5. Ginsburg GS, McCarthy JJ. Personalized medicine: revolutionizing drug discovery and patient care. Trends Biotechnol 2001; 19 (12): 491-6
6. Roses AD. Pharmacogenetics. Hum Mol Genet 2001; 10 (20): 2261-7
7. Khoury MJ, McCabe LL, McCabe ER. Population screening in the age of genomic medicine. N Engl J Med 2003; 348 (1): 50-8
8. Melzer D, Raven A, Detmer DE, et al. My very own medicine: what must I know? Information policy for pharmacogenetics. Cambridge: University of Cambridge, 2003
9. Guttmacher AS, Collins FS. Inheritance and drug response. N Engl J Med 2003; 348 (6): 529-37
10. Motulsky AG. Drug reactions, enzymes and biochemical genetics. JAMA 1957; 165 (7): 835-7
11. Vineis P, Malats N, Lang M, et al. Metabolic polymorphisms and susceptibility to cancer. Lyon: International Agency for Research on Cancer, 1999
12. Vogel F. Moderne probleme der humangenetik. Ergeb Inn Med Kinderheilkd 1959; 12: 52-125
13. Nebert DW. Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? Clin Genet 1999; 56 (4): 247-58
14. Fagerlund TH, Braaten O. No pain relief from codeine?: an introduction to pharmacogenomics. Acta Anaesthesiol Scand 2001; 45 (2): 140-9
15. Nebert DW, Bingham E. Pharmacogenomics: out of the lab and into the community. Trends Biotechnol 2001; 19 (12): 519-23
16. Tsai YJ, Hoyme HE. Pharmacogenomics: the future of drug therapy. Clin Genet 2002; 62: 257-64
17. Omenn GS, Motulsky AG. Integration of pharmacogenomics into medical practice. In: Rothstein MA, editor. Pharmacogenomics: social, ethical, and clinical dimensions. Hoboken (NJ): John Wiley & Sons Ltd, 2003
18. Guttmacher AE, Collins FS. Genomic medicine: a primer. N Engl J Med 2002; 347 (19): 1512-20
19. Vahakangas K. TP53 mutations in workers exposed to occupational carcinogens. Hum Mutat 2003; 21: 240-51
20. Hamadeh HK, Bushel PR, Jayadev S, et al. Prediction of compound signature using high density gene expression profiling. Toxicol Sci 2002; 67: 232-40
21. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002; 347: 1999-2009
22. Khoury MJ, Beskow L, Gwinn ML. Translation of genomic research into health care. JAMA 2001; 285 (19): 2447-8
23. Khoury MJ, Little J, Burke W. Human genome epidemiology: scope and strategies. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology. New York: Oxford University Press, 2004: 3-16
24. Veenstra DL. The interface between epidemiology and pharmacogenomics. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology. New York: Oxford University Press, 2004: 234-46
25. Daly E, Vessey MP, Hawkins MM, et al. Risk of venous thromboembolism in users of hormone replacement therapy. Lancet 1996; 348 (9033): 977-80
26. Evans WE, McLeod HL. Pharmacogenomics: drug disposition, drug targets, and side effects. N Engl J Med 2003; 348 (6): 538-49
27. Fryer AA, Jones PW. Interactions between detoxifying enzyme polymorphisms and susceptibility to cancer. In: Vineis P, Malats N, Lang M, et al., editors. Metabolic polymorphisms and susceptibility to cancer. Lyon: IARC, 1999: 303-22
28. Gemignani F, Landi S, Vivant F, et al. A catalogue of polymorphisms related to xenobiotic metabolism and cancer susceptibility. Pharmacogenetics 2002; 12: 459-63
29. Johnson JA. Pharmacogenetics: potential for individualized drug therapy through genetics. Trends Genet 2003; 19 (11): 660-6
30. Johnson JA, Lima JJ. Drug receptor/effector polymorphisms and pharmacogenetics: current status and challenges. Pharmacogenetics 2003; 13 (9): 525-34
31. Sing CF, Stengard JH, Kardia SLR. Genes, environment, and cardiovascular disease. Arterioscler Thromb Vasc Biol 2003; 23 (10): 1190-6
32. Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. Nature 2004; 429: 464-8
33. Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet 2001; 27: 234-6
34. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 2003; 33 Suppl.: 228-37
35. International HapMap Consortium. The international HapMap project. Nature 2003; 426: 789-96
36. Zhao LP, Li SS, Khalid N. A method for the assessment of disease associations with single-nucleotide polymorphism haplotypes and environmental variables in case-control studies. Am J Hum Genet 2003; 72: 1231-50
37. Daly MJ, Rioux JD, Schaffner SF, et al. High-resolution haplotype structure in the human genome. Nat Genet 2001; 29 (2): 229-32
38. Patil N, Berno AJ, Hinds DA, et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 2001; 294 (5547): 1719-23
39. Johnson GCL, Esposito L, Barratt BJ, et al. Haplotype tagging for the identification of common disease genes. Nat Genet 2001; 29: 233-7
40. Sebastiani P, Lazarus R, Weiss ST, et al. Minimal haplotype tagging. Proc Natl Acad Sci U S A 2003; 100 (17): 9900-5
41. Zhang W, Collins A, Morton NE. Does haplotype diversity predict power for association mapping of disease susceptibility? Hum Genet 2004; 115: 157-64
42. Huang Q, Fu YX, Boerwinkle E. Comparison of strategies for selecting single nucleotide polymorphisms for case/control association studies. Hum Genet 2003; 113: 253-7
43. Kamatani N, Sekine A, Kitamoto T, et al. Large-scale single-nucleotide polymorphism (SNP) and haplotype analyses, using dense SNP maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNPs within haplotype blocks and the haplotypes constructed with haplotype-tagging SNPs. Am J Hum Genet 2004; 75: 190-203
44. Cardon LR, Abecasis GR. Using haplotype blocks to map human complex triat loci. Trends Genet 2003; 19: 135-40
45. Ke X, Hunt S, Tapper W, et al. The impact of SNP density on fine-scale patterns of linkage disequilibrium. Hum Mol Genet 2004; 13: 577-88
46. Carlson CS, Eberle MA, Rieder MJ, et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analysis using linkage disequilibrium. Am J Hum Genet 2004; 74: 106-20
47. van Hylckama Vlieg A, Sandkuijl LA, Rosendaal FR, et al. Candidate gene approach in association studies: would the factor V Leiden mutation have been found by this approach? Eur J Hum Genet 2004; 12: 478-82
48. Martinez ME, O'Brien TG, Fultz KE, et al. Pronounced reduction in adenoma recurrence associated with aspirin use and a polymorphism in the ornithine decarboxylase gene. Proc Natl Acad Sci U S A 2003; 100 (13): 7859-64
49. Higashi MK, Veenstra DL, Kondo LM, et al. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. JAMA 2002; 287 (13): 1690-8
50. Bartsch H, Nair U, Risch A, et al. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. Cancer Epidemiol Biomarkers Prev 2000; 9: 3-28
51. Engel LS, Taioli E, Pfeiffer R, et al. Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: a HuGE review. Am J Epidemiol 2002; 156 (2): 95-109
52. Marcus PM, Hayes RB, Vineis P, et al. Cigarette smoking, N-acetyltransferase 2 acetylation status, and bladder cancer risk: a case-series meta-analysis of a gene-environment interaction. Cancer Epidemiol Biomarkers Prev 2000; 9 (5): 461-7
53. Benhamou S, Lee WJ, Alexandrie AK, et al. Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. Carcinogenesis 2002; 23: 1343-50
54. Brockton N, Little J, Sharp L, et al. N-acetyltransferase polymorphisms and colorectal cancer: a review. Am J Epidemiol 2000; 151 (9): 846-61
55. Cotton SC, Sharp L, Little J, et al. Glutathione S-transferase polymorphisms and colorectal cancer. Am J Epidemiol 2000; 151 (1): 7-32
56. Ye Z, Parry JM. Meta-analysis of 20 case-control studies on the N-acetyltransferase 2 acetylation status and colorectal cancer risk. Med Sci Monit 2002; 8 (8): CR558-65
57. Garcia-Closas M, Wacholder S, Caporaso N, et al. Inference issues in cohort and case-control studies of genetic effects and gene-environment interactions. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press, 2004: 127-44
58. Ioannidis JPA, Lau J. Integrating genetics into randomized controlled trials. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology: a

scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press, 2004: 247-63
59. Koopmans RP, Insel PA, Michel MC. Pharmacogenetics of hypertension treatment: a structured review. Pharmacogenetics 2003; 13 (12): 705-13
60. Murphy Jr GM, Kremer C, Rodrigues HE, et al. Pharmacogenetics of antidepressant medication intolerance. Am J Psychiatry 2003; 160 (10): 1830-5
61. Hillner BE. Trends in clinical trials reports in common cancers between 1989 and 2000. J Clin Oncol 2003; 21 (9): 1850-8
62. Haidich AB, Ioannidis JPA. Patterns of patient enrollment in randomized controlled trials. J Clin Epidemiol 2001; 54: 877-83
63. Hwang SJ, Beaty TH, Liang KY, et al. Minimum sample size estimation to detect gene-environment interaction in case-control designs. Am J Epidemiol 1994; 140 (11): 1029-37
64. Foppa I, Spiegelman D. Power and sample size calculations for case-control studies of gene-environment interactions with a polytomous exposure variable. Am J Epidemiol 1997; 146 (7): 596-604
65. Goldstein AM, Falk RT, Korczak JF, et al. Detecting gene-environment interactions using a case-control design. Genet Epidemiol 1997; 14 (6): 1085-9
66. Garcia-Closas M, Lubin JH. Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. Am J Epidemiol 1999; 149 (8): 689-92
67. Luan JA, Wong MY, Day NE, et al. Sample size determination for studies of gene-environment interaction. Int J Epidemiol 2001; 30: 1035-40
68. Salmen T, Heikkinen AM, Mahonen A, et al. Relation of estrogen receptor-alpha gene polymorphism and hormone replacement therapy to fall risk and muscle strength in early postmenopausal women. Ann Med 2002; 34 (1): 64-72
69. Ray WA. Population-based studies of adverse drug effects. N Engl J Med 2003; 349 (17): 1592-4
70. Thomas DC. Statistical issues in the design and analysis of gene-disease association studies. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press, 2004: 92-110
71. Singh GK, Miller BA, Hankey BF, et al. Area socioeconomic variations in US cancer incidence, mortality, stage, treatment, and survival, 1975-1999. NCI Cancer Surveillance Monograph Series, Number 4. Bethesda (MD): National Cancer Institute, 2003. NIH Publication No. 03-5417
72. Berrino F, Capocaccia R, Esteve J, et al. Survival of cancer patients in Europe: the EUROCARE-2 study. IARC Scientific Publications No. 151. Lyon: IARC, 1999
73. Hair JF, McNicol AM, Gusterson BA. Is research on human tissues at a crossroads? Eur J Cancer 2003; 39 (16): 2253-5
74. McQuillan GM, Porter KS, Agelli M, et al. Consent for genetic research in a general population: the NHANES experience. Genet Med 2003; 5 (1): 35-42
75. Riboli E, Hunt KJ, Slimani N, et al. European prospective investigation into cancer and nutrition (EPIC): study populations and data collection. Public Health Nutr 2002; 5 (6B): 1113-24
76. Hakonarson H, Gulcher JR, Stefansson K. deCODE genetics, Inc. Pharmacogenetics 2003; 4 (2): 209-15
77. Wright AF, Carothers AD, Campbell H. Gene-environment interactions: the BioBank UK study. Pharmacogenomics J 2002; 2 (2): 75-82
78. Branum AM, Collman GW, Correa A, et al. The national children's study of environmental effects on child health and development. Environ Health Perspect 2003; 111 (4): 642-6
79. Hirtzlin I, Dubreuil C, Preaubert N, et al. An empirical survey on biobanking of human genetic material and data in six EU countries. Eur J Hum Genet 2003; 11 (6): 475-88
80. Pembrey M. Genetic epidemiology: some special contributions of birth cohorts. Paediatr Perinat Epidemiol 2004; 18 (1): 3-7
81. Breslow NE, Day NE. Statistical methods in cancer research. IARC Scientific Publication No. 32 ed. Lyon, France: International Agency for Research on Cancer, 1980
82. Wacholder S, Hartge P. Case-control study. In: Gail MH, Benichou J, editors. Encyclopedia of epidemiologic methods. Chichester: John Wiley & Sons Ltd, 2000: 145-55
83. Lang JM. Case-control study, prevalent. In: Gail MH, Benichou J, editors. Encyclopedia of epidemiologic methods. Chichester: John Wiley & Sons Ltd, 2000: 160-9
84. Garcia-Closas M, Rothman N, Lubin J. Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. Cancer Epidemiol Biomarkers Prev 1999; 8 (12): 1043-50
85. Rosendaal FR, Vessey M, Rumley A, et al. Hormonal replacement therapy, prothrombotic mutations and the risk of venous thrombosis. Br J Haematol 2002; 116 (4): 851-4
86. 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 (8): 667-73
87. Ioannidis JPA, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. Nat Genet 2001; 29 (3): 306-9
88. Botto LD, Yang Q. 5,10 methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. Am J Epidemiol 2000; 151 (9): 862-77
89. Wang SS, Fernhoff PM, Hannon WH, et al. Medium chain acyl-CoA dehydrogenase deficiency human genome epidemiology review. Genet Med 1999; 1 (7): 332-9
90. Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control designs. Epidemiology 1991; 2 (2): 155-8
91. Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing case-series. Cancer Epidemiol Biomarkers Prev 1994; 3 (2): 173-5
92. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. Stat Med 1994; 13 (2): 153-62
93. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls. Am J Epidemiol 1996; 144 (3): 207-13
94. Yang Q, Khoury MJ. Evolving methods in genetic epidemiology: III. Gene-environment interaction in epidemiologic research. Epidemiol Rev 1997; 19 (1): 33-43
95. Albert PS, Ratnasinghe D, Tangrea J, et al. Limitations of the case-only design for identifying gene-environment interactions. Am J Epidemiol 2001; 154 (8): 687-93
96. Feigelson HS, McKean-Cowdin R, Pike MC, et al. Cytochrome P450c17alpha gene (CYP17) polymorphism predicts use of hormone replacement therapy. Cancer Res 1999; 59 (16): 3908-10
97. Sharp L, Cardy AH, Cotton SC, et al. CYP17 gene polymorphisms: prevalence and associations with hormone levels and related factors. A HuGE review. Am J Epidemiol 2004 Oct 15; 160 (8): 729-40
98. Herrington DM, Howard TD, Hawkins GA, et al. Estrogen-receptor polymorphisms and effects of estrogen replacement on high-density lipoprotein cholesterol in women with coronary disease. N Engl J Med 2002; 346 (13): 967-74
99. Risch N, Teng J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases 1. DNA pooling. Genome Res 1998; 8 (12): 1273-88
100. Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. Am J Epidemiol 1999; 149 (8): 693-705
101. Thomas DC. Statistical methods in genetic epidemiology. New York: Oxford University Press, 2004
102. Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. J Natl Cancer Inst Monogr 1999; 26: 31-7
103. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. Stat Med 2002; 21 (1): 35-50
104. Umbach DM, Weinberg CR. The use of case-parent triads to study joint effects of genotype and exposure. Am J Hum Genet 2000; 66: 251-61
105. Self SG, Longton G, Kopecky KJ, et al. On estimating HLA/disease association with application to a study of aplastic anemia. Biometrics 1991; 47: 53-61
106. Hill HA, Kleinbaum DG. Bias in observational studies. In: Gail MH, Benichou J, editors. Encyclopedia of epidemiologic methods. Chichester: John Wiley & Sons Ltd, 2000: 94-100
107. Wacholder S, Chatterjee N, Hartge P. Joint effects of genes and environment distorted by selection biases: implications for hospital-based case-control studies. Cancer Epidemiol Biomarkers Prev 2002; 11 (9): 885-9
108. Little J. Reporting and review of human genome epidemiology studies. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press, 2004: 168-92

109. Chenevix-Trench G, Young J, Coggan M, et al. Glutathione S-transferase M1 and T1 polymorphisms: susceptibility to colon cancer and age of onset. Carcinogenesis 1995; 16 (7): 1655-7
110. Olson SH. Reported participation in case-control studies: changes over time. Am J Epidemiol 2001; 154 (6): 574-81
111. Madigan MP, Troisi R, Potischman N, et al. Characteristics of respondents and non-respondents from a case-control study of breast cancer in younger women. Int J Epidemiol 2000; 29 (5): 793-8
112. McLeod A. Variation in the provision of chemotherapy for colorectal cancer. J Epidemiol Community Health 1999; 53 (12): 775-81
113. Campbell NC, Elliott AM, Sharp L, et al. Impact of deprivation and rural residence on treatment of colorectal and lung cancer. Br J Cancer 2002; 87 (6): 585-90
114. Rathore SS, Berger AK, Weinfurt KP, et al. Race, sex, poverty, and the medical treatment of acute myocardial infarction in the elderly. Circulation 2000; 102 (6): 642-8
115. Brown AF, Gross AG, Gutierrez PR, et al. Income-related differences in the use of evidence-based therapies in older persons with diabetes mellitus in for-profit managed care. J Am Geriatr Soc 2003; 51 (5): 665-70
116. Van Beijsterveldt CE, van Boxtel MP, Bosma H, et al. Predictors of attrition in a longitudinal cognitive aging study: the Maastricht aging study (MAAS). J Clin Epidemiol 2002; 55 (3): 216-23
117. Russell C, Palmer JR, Adams-Campbell LL, et al. Follow-up of a large cohort of black women. Am J Epidemiol 2001; 154 (9): 845-53
118. Deeg DJ, van Tilburg T, Smith JH, et al. Attrition in the longitudinal aging study Amsterdam: the effect of differential inclusion in side studies. J Clin Epidemiol 2002; 55 (4): 319-28
119. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol 1980; 112 (4): 564-9
120. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet 2001; 358 (9290): 1356-60
121. Little J, Bradley L, Bray MS, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am J Epidemiol 2002; 156 (4): 300-10
122. Steinberg K, Gallagher M. Assessing genotypes in human genome epidemiology studies. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press, 2004: 79-91
123. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press, 1992
124. Sjahid SI, van der Linden PD, Stricker BH. Agreement between the pharmacy medication history and patient interview for cardiovascular drugs: the Rotterdam elderly study. Br J Clin Pharmacol 1998; 45 (6): 591-5
125. Korthuis PT, Asch S, Mancewicz M, et al. Measuring medication: do interviews agree with medical record and pharmacy data? Med Care 2002; 40 (12): 1270-82
126. Boudreau DM, Daling JR, Malone KE, et al. A validation study of patient interview data and pharmacy records for antihypertensive, statin, and antidepressant medication use among older women. Am J Epidemiol 2004; 159 (3): 308-17
127. Cotterchio M, Kreiger N, Darlington G, et al. Comparison of self-reported and physician-reported antidepressant medication use. Ann Epidemiol 1999; 9 (5): 283-9
128. Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. Am J Epidemiol 1982; 116 (1): 114-22
129. Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol 1990; 43 (1): 87-91
130. Chouinard E, Walter S. Recall bias in case-control studies: an empirical analysis and theoretical framework. J Clin Epidemiol 1995; 48 (2): 245-54
131. Drews CD, Greenland S. The impact of differential recall on the results of case-control studies. Int J Epidemiol 1990; 19 (4): 1107-12
132. Khoury MJ. Case-parental control method in the search for disease-susceptibility genes. Am J Hum Genet 1994; 55 (2): 414-5
133. Swan SH, Shaw GM, Schulman J. Reporting and selection bias in case-control studies of congenital malformations. Epidemiology 1992; 3 (4): 356-63
134. Knowler WC, Williams RC, Pettitt DJ, et al. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 1988; 43 (4): 520-6
135. Gelernter J, Goldman D, Risch N. The A1 allele at the D2 dopamine receptor gene and alcoholism: a reappraisal. JAMA 1993; 269 (13): 1673-7
136. Kittles RA, Chen W, Panguluri RK, et al. CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? Hum Genet 2002; 110 (6): 553-60
137. Thomas DC, Witte JS. Point: population stratification: a problem for case control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev 2002; 11 (6): 505-12
138. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet 2003; 361 (9357): 598-604
139. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiology studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev 2002; 11 (6): 513-20
140. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2000; 92 (14): 1151-8
141. Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case-control studies. Am J Hum Genet 2002; 71 (2): 304-11
142. Garte S. The role of ethnicity in cancer susceptibility gene polymorphisms: the example of CYP1A1. Carcinogenesis 1998; 19 (8): 1329-32
143. Millikan RC. Re: population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2001; 93 (2): 156-7
144. Wang Y, Localio R, Rebbeck TR. Evaluating bias due to population stratification in case-control association studies of admixed populations. Genet Epidemiol 2004; 27: 14-20
145. Marchini J, Cardon LR, Phillips MS, et al. The effects of human population structure on large genetic association studies. Nat Genet 2004; 36 (5): 512-7
146. Freedman ML, Reich D, Penney KL, et al. Assessing the impact of population stratification on genetic association studies. Nat Genet 2004; 36 (4): 388-93
147. Hoggart CJ, Parra EJ, Shriver MD, et al. Control of confounding of genetic associations in stratified populations. Am J Hum Genet 2003; 72 (6): 1492-504
148. Lee WC. Case-control association studies with matching and genomic controlling. Genet Epidemiol 2004; 27: 1-13
149. Hirschhorn JN, Lohmueller K, Byrne E, et al. A comprehensive review of genetic association studies. Genet Med 2002; 4 (2): 45-61
150. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 2002; 3 (4): 299-309
151. Schork NJ, Fallin D, Lanchbury JS. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin Genet 2000; 58 (4): 250-64
152. Fallin D, Cohen A, Essioux L, et al. Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. Genome Res 2001; 11 (1): 143-51
153. Goldstein DB. Pharmacogenetics in the laboratory and the clinic. N Engl J Med 2003; 348 (6): 553-6
154. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003; 32: 1-22
155. Little J, Khoury MJ. Mendelian randomisation: a new spin or real progress? Lancet 2003; 362: 930-1
156. Feigelson HS, Coetzee GA, Kolonel LN, et al. A polymorphism in the CYP17 gene increases the risk of breast cancer. Cancer Res 1997; 57: 1063-5
157. Helzlsouer KJ, Huang HY, Strickland PT, et al. Association between CYP17 polymorphisms and the development of breast cancer. Cancer Epidemiol Biomarkers Prev 1998; 7 (10): 945-9
158. Haiman CA, Hankinson SE, Spiegelman D, et al. The relationship between a polymorphism in CYP17 with plasma hormone levels and breast cancer. Cancer Res 1999; 59 (5): 1015-20
159. Mitrunen K, Jourenkova N, Kataja V, et al. Steroid metabolism gene CYP17 polymorphism and the development of breast cancer. Cancer Epidemiol Biomarkers Prev 2000; 9: 1343-8
160. Ambrosone CB, Moysich KB, Furberg H, et al. CYP17 genetic polymorphism, breast cancer, and breast cancer risk factors. Breast Cancer Res 2003; 5: R45-51
161. Wacholder S, Chanock S, Garcia-Closas M, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004; 96 (6): 434-42

162. Smith PG, Day NE. The design of case-control studies: the influence of confounding and interaction effects. Int J Epidemiol 1984; 13 (3): 356-65
163. Thomas DC, Clayton DG. Betting odds and genetic associations. J Natl Cancer Inst 2004; 96 (6): 421-3
164. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet 2003; 361: 865-72
165. Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. Genetics 2003; 164 (2): 829-33
166. Botto LD, Khoury MJ. Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. Am J Epidemiol 2001; 153 (10): 1016-20
167. Botto LD, Khoury MJ. Facing the challenge of complex genotypes and gene-environment interaction: the basic epidemiologic units in case-control and case-only designs. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology. New York: Oxford University Press, 2004: 111-26
168. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol 1974; 99 (5): 325-32
169. Yang Q, Khoury MJ, Botto L, et al. Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes. Am J Hum Genet 2003; 72 (3): 636-49
170. Aragaki CC, Greenland S, Probst-Hensch N, et al. Hierarchical modeling of gene-environment interactions: estimating NAT2 genotype-specific dietary effects on adenomatous polyps. Cancer Epidemiol Biomarkers Prev 1997; 6 (5): 307-14
171. Witte JS. Genetic analysis with hierarchical models. Genet Epidemiol 1997; 14 (6): 1137-42
172. Ioannidis JPA, Rosenberg PS, Goedert JJ, et al. Commentary: meta-analysis of individual participants' data in genetic epidemiology. Am J Epidemiol 2002; 156: 204-10
173. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 1996; 49 (11): 1225-31
174. Warner B, Misra M. Understanding neural networks as statistical tools. Am Stat 1996; 50 (4): 284-93
175. Duh MS, Walker AM, Ayanian JZ. Epidemiologic interpretation of artificial neural networks. Am J Epidemiol 1998; 147 (12): 1112-22
176. Goldstein DB, Tate SK, Sisodiya SM. Pharmacogenetics goes genomic. Genetics 2003; 4: 937-47
177. Klerk M, Verhoef P, Clarke R, et al. MTHFR 677C->T polymorphism and risk of coronary heart disease. JAMA 2002; 288 (16): 2023-31
178. Sharp L, Little J. Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review. Am J Epidemiol 2004; 159 (5): 423-43
179. Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? Nat Genet 2000; 26: 151-7
180. Challenges for the 21st century. Nat Genet 2001; 29: 353-4
181. Cloninger CR. The discovery of susceptibility genes for mental disorders. Proc Natl Acad Sci U S A 2002; 99: 13365-7
182. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. Science 2002; 298: 2345-9
183. Khoury MJ. Commentary: epidemiology and the continuum from genetic research to genetic testing. Am J Epidemiol 2002; 156 (4): 297-9
184. Harrison PJ, Owen MJ. Genes for schizophrenia?: recent findings and their pathophysiological implications. Lancet 2003; 361: 417-9
185. Page GP, George V, Go RC, et al. "Are we there yet?": deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. Am J Hum Genet 2003; 73: 711-9
186. Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001; 2 (2): 91-9
187. Haddow JE, Palomaki GE. ACCE: a model process for evaluating data on emerging genetic tests. In: Khoury MJ, Little J, Burke W, editors. Human genome epidemiology. New York: Oxford University Press, 2004: 217-33
188. Burke W, Atkins D, Gwinn M, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. Am J Epidemiol 2002; 156 (4): 311-8
189. Veenstra DL, Higashi MK, Phillips KA. Assessing the cost-effectiveness of pharmacogenetics. AAPS Pharm Sci 2000; 2 (3): E29

Correspondence and offprints: Dr *Julian Little*, Department of Epidemiology and Community Medicine, University of Ottawa, 451 Smyth Rd., Ottawa, Ontario K1H 8M5 , Canada.
E-mail: jlittle@uottawa.ca