Am. J. Hum. Genet. 73:711–719, 2003

INVITED EDITORIAL
"Are We There Yet?": Deciding When One Has Demonstrated Specific Genetic Causation in Complex Diseases and Quantitative Traits

Grier P. Page,[1] Varghese George,[1,2] Rodney C. Go,[3] Patricia Z. Page,[4] and David B. Allison[1,5]

[1]Section on Statistical Genetics, Department of Biostatistics, [2]Division of Rheumatology, Department of Medicine, Departments of [3]Epidemiology and [4]Genetics, and [5]Clinical Nutrition Research Center, University of Alabama at Birmingham, Birmingham

Although mathematical relationships can be proven by deductive logic, biological relationships can only be inferred from empirical observations. This is a distinct disadvantage for those of us who strive to identify the genes involved in complex diseases and quantitative traits. If causation cannot be proven, however, what does constitute sufficient evidence for causation? The philosopher Karl Popper said, "Our belief in a hypothesis can have no stronger basis than our repeated unsuccessful critical attempts to refute it." We believe that to establish causation, as scientists, we must make a serious attempt to refute our own hypotheses and to eliminate all known sources of bias before association becomes causation. In addition, we suggest that investigators must provide sufficient data and evidence of their unsuccessful efforts to find any confounding biases. In this editorial, we discuss what "causation" means in the context of complex diseases and quantitative traits, and we suggest guidelines for steps that may be taken to address possible confounders of association before polymorphisms may be called "causative."

Background

The cystic fibrosis gene was one of the first genes identified by positional cloning (Riordan et al. 1989), and its discovery was heralded with much fanfare. In tribute to the evolving power of modern molecular, genomic, and statistical tools, the identification of genes responsible for Mendelian traits has progressed to such an extent that, in the April 2003 issue of *Nature Genetics,* identification of the genes for no fewer than eight different Mendelian conditions were reported. To date, >1,400 genes for ~1,200 Mendelian traits have been identified.

The tremendous success in identification of genes responsible for Mendelian traits has not been followed, however, by similar successes in the identification of genes responsible for complex diseases or for variation in quantitative traits. Despite >1,300 National Institutes of Health (NIH)–funded studies of complex genetic disease (Computer Retrieval of Information on Scientific Projects) and multiple reports on at least 166 genes, as few as 10–50 causative polymorphisms for complex diseases have been identified in humans (Ioannidis et al. 2003; Lohmueller et al. 2003). It is disturbing that only 16%–30% (Hirschhorn and Altshuler 2002; Ioannidis 2003; Ioannidis et al. 2003; Lohmueller et al. 2003) of initially reported significant associations have been consistently replicated without any evidence of between-study heterogeneity or bias.

Why has the identification of causative genes for complex diseases been so elusive? To a certain extent, their name reveals the problem; complex diseases are complex. They may be affected by both genetic and environment interactions, as well as by gene-by-gene and gene-by-environment factors. Although traits being studied may aggregate in families, they do not segregate in Mendelian fashion, and, even if they did, the late onset of some of these diseases would make family studies difficult. To make matters even more challenging, individual alleles are probably neither necessary nor sufficient to cause the phenotype; thus, the "disease" alleles may be present in the nondiseased population. Perhaps the greatest challenge of all is the fact that we cannot randomly assign people to levels of the independent variable (i.e., genotypes). Most statisticians would agree that random assignment is the sine qua non of the true experiment. This puts us out of the realm of true experimentation (Rubin 1991; Rosenbaum 1995) and leaves open grounds for doubt (Rosenbaum 1995). Our job, then, is providing convincing evidence for a putative causal effect comes from our ability to enumerate and subsequently vitiate each of those grounds for doubt.

Given these challenges, the identification of causative polymorphisms has been difficult, and there is no explicit consensus about what constitutes sufficient evidence to

Received August 5, 2003; accepted for publication August 6, 2003; electronically published September 17, 2003.

Address for correspondence and reprints: Dr. Grier P. Page, RPHB 327, 1530 3rd Avenue South, Birmingham, AL 35294-0022. E-mail: gpage@ms.soph.uab.edu

This article represents the opinions of the authors and has not been peer reviewed.

0002-9297/2003/7304-0002$15.00

establish causation from association. Several groups have offered thoughts on these matters (Glazier et al. 2002; Sing et al. 2003), but the treatments have either been highly context dependent (Hirschhorn and Altshuler 2002) or heavily focused on situations in which there exist clear animal models of the phenotypes under study (Mackay 2001), which is often not the case for complex traits, such as psychiatric disorders. In this treatise, we advance suggestions for what is sufficient evidence to translate from association to causation.

The Definition of "Causation" in Complex Diseases and Quantitative Traits

"The view [of causation] that we adopt has consequences which reach beyond informal discussion during coffee breaks" (Olsen 1993, p. 205). Causation is an essential concept in the field of genetics. Causal claims, such as "This gene causes such and such disease," have long been in the genetics literature, but it is not clear that geneticists are consistent in using the concept of causation, especially in the context of complex traits.

Many definitions of "cause" have been developed. "The word cause is an abstract noun and, like beauty, will have different meanings in different contexts" (MacMahon and Pugh 1970, p. 396). To use "cause" as a scientific term, it is important that we clarify what is meant in the particular context. If somebody makes a statement that "smoking causes cancer," it could mean "only smoking causes cancer," "all people who smoke will develop cancer," or "at least one smoker will develop cancer," depending on the point of view of the speaker (Kramer and Lane 1992; Parascandola and Weed 2001). On the basis of an extensive review of the epidemiological literature, there are at least five nonmutually exclusive definitions of "causation" (adapted from Parascandola and Weed [2001]):

1. Production: causes are conditions that play an essential part in producing the occurrence of disease.
2. Necessary: a necessary cause is a condition without which the effect cannot occur.
3. Sufficient: when the cause is present, the effect must occur.
4. Probabilistic: a probabilistic cause increases the probability of its effect occurring.
5. Counterfactual: a counterfactual cause makes a difference in the outcome when it is present verses when it is absent.

It is important to note that, when considering the causes of Mendelian traits, we usually talk about "sufficient causes," but, for complex diseases, we are referring to "probabilistic causes," which increase the probability of a disease occurring. For quantitative traits, the concept of "counterfactual cause" is considered. A counterfactual cause makes a difference in the outcome (or probability of an outcome) when it is present, compared with when it is absent, while all else is held constant. Neither probabilistic nor counterfactual causes require the causes to be either necessary or sufficient. (Parascan dola and Weed 2001). The probabilistic and counterfactual uses of "cause" have certain advantages over other definitions, because they are more inclusive and make fewer assumptions, an aim of science. These definitions also provide the theoretical means to construct models of interactions and dose-response relationships through a continuum of probability values, as manifested by the liability for complex traits and the distribution or value of quantitative traits (Falconer and Mackay 1996).

Some History of Guidelines for Establishing Causation

The accepted requirements for demonstrating causation have evolved through time. In the 1880s, Robert Koch established guidelines for determining whether a microorganism is the cause of a disease. Koch (1882) stated that, to establish causation, one had to demonstrate that: (1) the parasite occurs in every case of the disease in question and under circumstances that can account for the pathological changes and the clinical course of the disease; (2) the parasite occurs in no other disease as a fortuitous and nonpathogenic parasite; and (3) after being fully isolated from the body and repeatedly grown in pure culture, the parasite can induce the disease again (Koch 1882, 1891; Rivers 1937; Fredericks and Relman 1996). Although this line of reasoning worked well for tuberculosis, the existence of obligate parasites, such as viruses, made the limitations of this approach for many other diseases so obvious that, by 1936, Thomas Rivers, then President of the Society of American Bacteriologists, wrote "It is unfortunate that so many workers blindly followed the rules, because Koch himself recognized that in certain instances all the conditions could not be met.… Thus in regard to certain diseases, particularly those caused by viruses, blind adherence to Koch's postulates may act as a hindrance instead of an aid." (Rivers 1937, p. 21). Rivers went on to develop a revision for the postulates that later developed into a nine-point list that included epidemiology in the equation (Huebner 1957). Huebner (1957) recognized that setting strict rules could be a hindrance; thus, he proposed his list as guidelines rather than a formal set of rules.

In 1965, Sir Austin Bradford Hill presented guidelines for epidemiological causation in environmental and occupational medicine (Hill 1965):

1. Strength of association
2. Consistency and unbiasedness of association
3. Specificity of the association

4. Temporality
5. Biological gradient
6. Biological plausibility
7. Coherence with previous knowledge
8. Experimental evidence
9. Reasoning by analogy.

With careful reading, it is clear that Hill (1965) intended these as guidelines and not as rigid criteria that must be satisfied. Multiple times, he pointed out that not all of these guidelines will be applicable in all situations and that there may be times when we wish to conclude that a putative cause-effect relationship is real even when some of the criteria are not met. The criteria proposed by Hill (1965) also do not assume a one-to-one relationship between cause and effect, and, in this sense, they are well suited for use with complex diseases and quantitative traits. However, it must be acknowledged that legitimate questions have been raised about the degree to which they are applicable to modern epidemiology, in which the causes under study may have very modest effects (Phillips and Goodman 2001). This criticism extends to the study of genes for complex diseases and quantitative traits a fortiori. Thus, although we think that the modern genetic researcher can benefit from studying Hill's (1965) proposed guidelines and his rationale for them, we believe that they, too, are not entirely satisfactory for our needs.

The genetics literature has been filled with many guidelines. Morton (1955) suggested that a LOD score >3 is required to declare linkage of a Mendelian trait to a locus. There are also well-developed criteria for the identification and reporting of causative polymorphisms for Mendelian traits. Typically, these include: (1) linkage (usually LOD > 3) to a particular region of the human genome; (2) one or more independent mutations that are perfectly concordant with disease status in affected families; (3) defect(s) that lead to macro changes in the protein; (4) putative mutations that are not present in a sample from a control population, or better yet, no macro mutations in the putative gene in the samples from the control populations; and, often, (5) the presence of some other line of biological evidence (expression, protein, knockout, etc. [Glazier et al. 2002]).

Lander and Kruglyak (1995) have advanced a set of criteria for the reporting of linkage of putative loci for complex traits. These guidelines were developed to minimize the rate of type 1 errors in genome scans. Others have suggested false-discovery-rate (FDR) procedures (Benjamini et al. 2001; Storey and Tibshirani 2003). Several groups have advanced criteria for establishing association with complex traits (Lander and Schork 1994; Risch and Merikangas 1996; Phillips 1999; Mackay 2001; Glazier et al. 2002). Risch and Merikangas (1996) have suggested a P value of $.5 \times 10^{-8}$ or smaller, whereas others (Mackay 2001) have suggested more biological lines of evidence, such as complementation or mouse-knockout studies. However, no consensus has yet been reached on what constitutes sufficient evidence for complex diseases or quantitative traits.

From Association to Causation

An association between a polymorphism and a complex disease or quantitative trait can exist for four reasons:

1. The polymorphism is actually causative for the disease or trait.
2. The association is a false positive due to random chance.
3. The polymorphism is in disequilibrium with the true causative allele.
4. The polymorphism is associated because of some systematic bias in the biology, study, samples, or analysis.

We believe that the best method to establish that polymorphisms are causative for complex diseases or quantitative traits is summarized by two quotations. The first is from Sherlock Holmes: "[W]hen you have eliminated all which is impossible, then whatever remains, however improbable, must be the truth" (Doyle 1926, p. 464). Second, a quote from Karl Popper: "Our belief in some hypotheses can have no stronger basis than our repeated unsuccessful critical attempts to refute it" (Popper 1961, p. 100). When an investigator(s) systematically removes any possible sources of random error, bias, confounding factors, or disequilibria, and an association remains, causation between a polymorphism and a trait or disease may be suggested.

Reduction of Association by Chance

It is tempting to set a single, hard P value guideline for causation, such as those proffered by Lander and Kruglyak (1995) and Risch and Merikangas (1996), for they are easy to comply with and are very objective. However, we should not, as Bradford Hill said, "[allow] [t]he glitter of the t table to divert attention from the inadequacies of the fare" (Hill 1965, p. 300). The validity of the P value that results at the end of a study is dependent on the quality of the data put into a study and on the quality of the analyses employed. In other words, "garbage in, garbage out." Setting a high P value only reduces the chance of false-positive associations; it does not deal with the other two sources of spurious associations, disequilibrium and bias (Whitte et al. 1996).

However, there should be some control of errors. A conservative P value—such as a Bonferroni-corrected P value of $P < .05$ or an FDR (Benjamini and Hochberg

1995; Benjamini et al. 2001) of, say, <5%—is useful to reduce the rate of false positives, but this is only one component of a set of criteria for causation.

Elimination of Association Due to Disequilibrium

Whereas disequilibrium can be used to predict the effect of the true causative loci, as is often done in marker-assisted selection in plants, "the cause of an event in nature is the handle so to speak, by which we can manipulate it" (Collingwood 1940, p. 278). If we do not have the true causative polymorphisms, we have a fragile handle that may not be sufficient for the understanding or manipulation of the disease.

The identification of the true causative polymorphisms in a region is quite involved, owing to the large stretches of DNA over which disequilibrium can exist (Clark et al. 1998; Nickerson et al. 1998, 2000). Careful attention must be paid to identifying all the polymorphisms in a region. Haplotype blocks, allelic heterogeneity, overdominance, and epistasis can all confound the identification of the true causative allele(s) (Culverhouse et al. 2002). This is essentially a problem of multicollinearity, which is best dealt with by experimental assignment designed to break up the collinearity (Glantz and Slinker 1990) or by increasing sample size to compensate for it. The former seems impossible, owing to the inability to assign people to genotypes, and the sample required for the latter may exceed the population of the planet for tightly linked loci.

Eliminating disequilibrium as a source of confounding is an involved process and requires extensive sequencing, genotyping, and statistical testing. Even after great effort, the identification of a true causative polymorphism may still elude us.

Elimination of Errors Due to the Design and Conduct of the Study

Some forms of error or bias are not inherent in the phenomena under study but are the result of methodological artifacts, procedural errors, or (it is hoped, unintentional) investigator biases.

The following list is not meant to be exhaustive but, rather, is a delineation of some of the issues and topics that require proper care and attention during the design and analysis of studies. Although it may go without saying, a refined understanding of the disease process is needed before beginning a study, and this knowledge should be devoted to minimizing the biases and confounding factors, even before the study begins. There are surprisingly few reviews that highlight the issues that need to be considered in the design and conduct of a study; for exceptions, see Terwilliger and Goring (2000) and Ellsworth and Manolio (1999). The phenotype(s) to be studied must be well defined and heritable. Linkage and association, no matter how significant, have no meaning unless there is a genetic contribution to the trait (Ott 1991). Second, all biological assays are imprecise. Genotyping (Gordon et al. 2002, 2003) and phenotyping (Rice et al. 2001; Egan et al. 2003) have error rates. Relatively modest levels of error in either the genotyping or phenotyping will result in significantly diminished power; moreover, some errors, such as null alleles (Ewen et al. 2000), can cause false-positive results. Despite all efforts to the contrary, there are always human errors. Rigorous quality-control checks can be implemented in all studies, to properly account for these types of human errors. Every statistical test is based on certain assumptions about the nature of the data, and, if these assumptions are violated, the significances generated may not be valid. The assumptions that underpin all statistical tests are known and can be verified. The software used can generate different results for essentially the same analysis (Weeks et al. 1995). Finally, there is always some "cleaning" of data before formal analysis, such as correcting systematic errors (e.g., fixing pedigree errors). However, extreme caution must be used in cleaning, for it is possible to alter data to get the answer that one desires. Cleaning should be conducted to get the most accurate data, not the most significant result.

It is here that the role of replication becomes critical. Contrary to popular opinion, replication is not the optimal way to deal with the threat to a conclusion's validity posed by stochastic error. This is best dealt with by seeking an appropriately small frequentist *P* value or FDR or an appropriately high Bayesian posterior probability (Vieland 1998). Instead, replication may be the best way to fend off the threat to a conclusion's validity posed by methodological artifacts, procedural errors, or investigator biases. It is important to note, however, that, if the same experimenter—with the same biases and using the same methods and procedures—repeats a study that led to an erroneous conclusion, then the same erroneous conclusion could be reached again. In this light, David Lykken's classic paper (1968) on replication is instructive. Lykken (1968) distinguished between "operational replication" and "constructive replication." In the former, one strives to repeat the original study exactly, procedure for procedure. In the latter, one strives to evaluate the same conceptual hypothesis but through distinctly different methodology. Most would agree that a constructive replication (e.g., an association observed by a different investigator who uses a different genotyping method, a different phenotyping method, a different sampling scheme, a different statistical procedure, and different data-analytic software) offers far stronger evidence that the original apparent association was not due to investigator bias, systematic genotyping error, sys-

tematic phenotyping error, selection bias, an invalid statistical procedure, or a software bug.

Controlling for Biological Biases

Controlling for Admixture

There are also sources of potential bias that are intrinsic to the nature of phenomena being studied rather than to experimental artifacts. Population admixture is potentially problematic, for it can both lead to false-positive results and mask true association. Opinions regarding the importance of population stratification in association studies vary greatly (Thomas and Witte 2002; Cardon and Palmer 2003; Hoggart et al. 2003). In an effort to make case-control studies valid in the presence of admixture, a wide variety of methods have been developed to address the effects of population stratification (Parra et al. 1998, 2001; Stephens et al. 1998; Pritchard and Rosenberg 1999; McKeigue et al. 2000). Admixture will not only induce false-positive results, but it can also induce false-negative results (Deng 2001; Deng et al. 2001). Some may say that admixture is not an issue, but we believe that, unless it is addressed, it will remain a potential source of bias that detracts from any reported causation. Many of the admixture-controlling methods are dependent on having parental populations; however, when studying an admixed population, there is not one single European, Asian, African, or Native American population, and selecting the incorrect parental population can change the estimates of admixture and thus bias the results. In addition, it is not yet clear which of these methods are most powerful and whether any are valid in practice.

Selecting the Controls

The choice of a control population is probably the most critical factor for the success or failure of a case-control study. A poorly chosen control population can mask true associations (false-negative results) or generate false-positive results (Ellsworth and Manolio 1999). The false-positive results can be detected by replication in other populations, but the false-negative results are very difficult to identify and to eliminate.

Several of the recently reported associations with complex diseases have been in founder populations (Gratacos et al. 2001; Gianfrancesco et al. 2003). Although founder populations can be very powerful for detecting common disease genes, the same founding event that makes them powerful for identifying disease alleles can also cause changes in the allele frequencies for many, if not most, polymorphic sites (de la Chapelle and Wright 1998; Wright et al. 1999). Thus, selecting control samples from even a modest geographic distance away from the founder location can cause spurious associations.

Most complex diseases have an environmental component. Thus, controls should be selected to minimize the confounding factors between genes and environment (Lander and Schork 1994; Glazier et al. 2002; Kaprio et al. 2002). If many traits are to be studied in a population, it may be advantageous to collect a control population that is far larger than the case population and then match on the basis of covariate information.

The use of convenience controls (samples from previous studies, samples from the lab across the hall, and/or random samples from the blood bank) is quite common, but their use can lead to biases. It is often unknown whether the controls have the disease of interest; their phenotypic assessments may have been conducted with different instruments, at different labs, or with different techniques; they may not be matched for admixture; they may have different environmental factors; and there is less ability to recontact, if additional information is needed. It is, thus, usually better to collect a new, well-characterized control population designed to address the disease of interest.

Biological Plausibility

At no point have we talked about the biological plausibility of the polymorphism or about using nonstatistical methods to verify that the polymorphism has some effect. We believe that biological plausibility is useful, but, given that our knowledge of the genome is not complete, biological plausibility may not be apparent. Many of the genes known to harbor variation that predisposes to Mendelian traits were not known to be involved in the biology of the disease until the mutations were discovered. For example, it took several years after the discovery of the *BRCA1* gene to determine how it functions as a tumor suppressor, and some of the functions of *BRCA1* are still not well known. We are not saying that biological plausibility is not important; rather, we do not believe that any one particular biological test is appropriate under all conditions. Some papers suggest complementation—the mating of two homozygous mutants strains to generate or fail to generate a wild-type organism (Mackay 2001; Glazier et al. 2002)—whereas others suggest inducing the mutation in a mouse model. However, in humans, complementation is essentially impossible, and there has been only modest success in replicating human diseases in mice (Ahmad-Annuar et al. 2003). The knockout is the modern interpretation of Koch's hypothesis, useful in some regards but not in others, and is highly limiting if it must be applied in all cases (Thyagarajan et al. 2003). For example, generating noncoding or complex effects, such as those in *CAPN10,* may not be possible. Some of the types of biological proofs that have been offered, such as differential gene expression, reveal only different gene expression; it does not necessarily fol-

low that a disease phenotype will result. Thus, although any causative polymorphisms will have some biological link to the disease or trait, we do not include biological plausibility in our model for probabilistic or counterfactual causation.

"When You Have Eliminated All Which Are Impossible, Then Whatever Remains, However Improbable, Must Be the Truth."

The above list of potential biases that could be found in a study is by no means complete. It is the responsibility of the investigator(s) to think deeply about all the possible sources of bias and confounding and to remove them. If the biases cannot be removed, it is not appropriate to suggest causation. We believe, in essence, that the investigators must have sufficiently convinced themselves that the reported polymorphism is causative, in that they are ready to stop exploring the region with genetic techniques and either to move on to other unlinked loci or to switch to determining the biological basis for the causative allele. Once this level of belief has been reached, it is time to try to convince others of the validity of the suggested causation.

Dissemination of Ideas

"All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge" (Hill 1965, p. 297). However, at a certain point, knowledge must be disseminated and tested by others.

Information is the key to allow one's peers to make a serious attempt to prove or refute a hypothesis. To facilitate testing the hypothesis of causation, we suggest that all tested phenotypic and genotypic data be reported in online supplements in their original—rather than tabulated—forms, even those results that are nonsignificant. Many reading this editorial will argue that they provide sufficient information or as much information as can be squeezed into a 1,500-word short communication. Although the size of articles may be shrinking, there is essentially unlimited space available for online supplements that can be used to compensate for this shortcoming. If online supplements are used, they must be edited and peer reviewed as well as the main article.

The suggestion that we share a tremendous amount of data may not sit well with investigators. Revealing all this information can lead investigators to feel very vulnerable. They may think, "I collected the data. Why should I let others reap the benefits?" "Did I test everything?" "Will I be embarrassed in public?" We agree that these are genuine concerns. However, there is a movement in the NIH to encourage investigators to report data quickly. By October 2003, all data from new NCI grants with a budget of >$1/2 million per year must put all data in the public domain within a "reasonable" length of time, and many journals have adopted the recommendation of the Microarray Gene Expression Data consortium that all microarray data be put in the public domain before publication (Ball et al. 2002). The push to make public all data collected through publicly funded investigations will only increase, and we should embrace this policy rather than fight it. The goal of this policy is to have science advance faster by having many, rather than a few, working on a problem. As we have seen in the genome-sequencing projects of many organisms, we can reach the goal faster by having researchers at many sites striving toward the same goal and releasing data as they are produced. Similarly, the SNP Consortium (Thorisson and Stein 2003) allows many to work on gene identification while others work on the verification and establish gene frequencies in various populations.

Reporting of this amount and type of data will also allow for better meta-analysis (Goldstein et al. 1999; Guerra et al. 1999; Etzel and Guerra 2002), studies of meta-phenotypes (Uhl et al. 2002), the identification of subgroups, and minimization of publication bias (Lohmueller et al. 2003). For example, the combing studies of alcoholism and nicotine and illegal-drug use could lead to the identification of substance-abuse loci (Uhl et al. 2002). Linkage or association studies of subgroups can be quite powerful (Hall et al. 1990; Hall 2003). Mary Claire King was able to identify linkage to the region of the *BRCA1* locus because she was able to observe a subset of families with very-early–onset breast cancer (Hall et al. 1990).

Attempts to prove or refute a hypothesis of causation should be conducted with as much care to remove biases as was used for the initial study. Biases should be removed from both positive and negative studies, for, although biases can cause false-positive results, they can also cause false-negative results. It would not be fair to attempt to refute a well-designed study with a poorly designed study.

Summary

We believe that definitive proof of causation will remain elusive, owing, in part, to our inability to randomly assign people to genotypes. However, the strongest belief in a causal connection between a polymorphism and a disease or trait will come through repeated, systematic attempts to eliminate any source of bias that could be leading to a false discovery, while maintaining a high level of statistical significance. Since it is always possible that some unmeasured or unimagined effects could be biasing the results, we encourage the reporting of all relevant data so that any biases that could lead to false results can be identified. We recommend reporting in-

formation on the genotypes and phenotypes (including the data that make up composite phenotypes) of all individuals in the study. Only after independent groups have successfully replicated a finding and have been unsuccessful at identifying new sources of systematic bias does it seem reasonable to conclude that causation has been demonstrated with sufficient certainty for everyday life.

Acknowledgments

This work was supported, in part, by NIH grants DK056366, P30DK056336, R01ES09912, P20RR016430, and P01AR049084.

Electronic-Database Information

The URL for data presented herein is as follows:

Computer Retrieval of Information on Scientific Projects, http://crisp.cit.nih.gov/

References

Ahmad-Annuar A, Tabrizi SJ, Fisher EM (2003) Mouse models as a tool for understanding neurodegenerative diseases. Curr Opin Neurol 16:451–458

Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, Cavalieri D, Gaasterland T, Hingamp P, Holstege F, Ringwald M, Spellman P, Stoeckert CJ Jr, Stewart JE, Taylor R, Brazma A, Quackenbush J (2002) An open letter to the scientific journals. Bioinformatics 18:1409

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125:279–284

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361:598–604

Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengård J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF (1998) Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am J Hum Genet 63:595–612

Collingwood R (1940) Causation in practical natural science: an essay in metaphysics. Clarendon Press, Oxford

Culverhouse R, Suarez BK, Lin J, Reich T (2002) A perspective on epistasis: limits of models displaying no main effect. Am J Hum Genet 70:461–471

de la Chapelle A, Wright F (1998) Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. Proc Natl Acad Sci USA 95:12416–12423

Deng HW (2001) Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. Genetics 159:1319–1323

Deng HW, Chen WM, Recker RR (2001) Population admixture: detection by Hardy-Weinberg test and its quantitative effects on linkage-disequilibrium methods for localizing genes underlying complex traits. Genetics 157:885–897

Doyle AC (1926) The adventure of the blanched soldier. Reprinted in: Doyle AC (1998) The complete Sherlock Holmes: all four novels and 56 short stories. Bantam Classic and Loveswept, New York

Egan S, Nathan P, Lumley M (2003) Diagnostic concordance of ICD-10 personality and comorbid disorders: a comparison of standard clinical assessment and structured interviews in a clinical setting. Aust N Z J Psychiatry 37:484–491

Ellsworth DL, Manolio TA (1999) The emerging importance of genetics in epidemiologic research II: issues in study design and gene mapping. Ann Epidemiol 9:75–90

Etzel CJ, Guerra R (2002) Meta-analysis of genetic-linkage analysis of quantitative-trait loci. Am J Hum Genet 71:56–65

Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW, Foote SJ (2000) Identification and analysis of error types in high-throughput genotyping. Am J Hum Genet 67:727–736

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th ed. Addison-Wesley, New York

Fredericks DN, Relman DA (1996) Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clin Microbiol Rev 9:18–33

Gianfrancesco F, Esposito T, Ombra MN, Forabosco P, Maninchedda G, Fattorini M, Casula S, Vaccargiu S, Casu G, Cardia F, Deiana I, Melis P, Falchi M, Pirastu M (2003) Identification of a novel gene and a common variant associated with uric acid nephrolithiasis in a Sardinian genetic isolate. Am J Hum Genet 72:1479–1491

Glantz S, Slinker B (1990) Primer of applied regression and analysis of variance. McGraw-Hill, New York

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science 298:2345–2349

Goldstein DR, Sain SR, Guerra R, Etzel CJ (1999) Meta-analysis by combining parameter estimates: simulated linkage studies. Genet Epidemiol 17:S581–S586

Gordon D, Finch SJ, Nothnagel M, Ott J (2002) Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. Hum Hered 54:22–33

Gordon D, Levenstien MA, Finch SJ, Ott J (2003) Errors and linkage disequilibrium interact multiplicatively when computing sample sizes for genetic case-control association studies. Pac Symp Biocomput 490–501

Gratacos M, Nadal M, Martin-Santos R, Pujana MA, Gago J, Peral B, Armengol L, Ponsa I, Miro R, Bulbena A, Estivill X (2001) A polymorphic genomic duplication on human chromosome 15 is a susceptibility factor for panic and phobic disorders. Cell 106:367–379

Guerra R, Etzel CJ, Goldstein DR, Sain SR (1999) Meta-analysis by combining p-values: simulated linkage studies. Genet Epidemiol 17:S605–S609

Hall JG (2003) A clinician's plea. Nat Genet 33:440–442

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–1689

Hill A (1965) The environment and disease: association or causation? Proc R Soc Med 58:295–300

Hirschhorn JN, Altshuler D (2002) Once and again—issues surrounding replication in genetic association studies. J Clin Endocrinol Metab 87:4438–4441

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM (2003) Control of confounding of genetic associations in stratified populations. Am J Hum Genet 72:1492–1504

Huebner R (1957) The virologist's dilemma. Ann NY Acad Sci 67:430–438

Ioannidis JP (2003) Genetic associations: false or true? Trends Mol Med 9:135–138

Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG (2003) Genetic associations in large versus small studies: an empirical assessment. Lancet 361:567–571

Kaprio J, Pulkkinen L, Rose RJ (2002) Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. Twin Res 5:366–371

Koch R (1882) Die Aetiologie der Tuberculose. Reprinted in Clark D (ed) (1942) Source book of medical history. Dover Publications, New York

——— (1891) Uber bakteriologische Forschung Verhandlung des X Internationalen Medichinischen Congresses, Berlin, 1890. Xth International Congress on Medicine, Berlin

Kramer MS, Lane DA (1992) Causal propositions in clinical research and practice. J Clin Epidemiol 45:639–649

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177–182

Lykken D (1968) Statistical significance in psychological research. Psychol Bull 70:151–159

Mackay TF (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35:303–339

MacMahon B, Pugh T (1970) Epidemiology: principles and methods. Little Brown, Boston

McKeigue PM, Carpenter J, Parra E, Shriver MD (2000) Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Ann Hum Genet 64:171–186

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318

Nickerson DA, Taylor SL, Fullerton SM, Weiss KM, Clark AG, Stengard JH, Salomaa V, Boerwinkle E, Sing CF (2000) Sequence diversity and large-scale typing of SNPs in the human apolipoprotein E gene. Genome Res 10:1532–1545

Nickerson DA, Taylor SL, Weiss KM, Clark AH, Hutchinson RG, Stengård J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF (1998) DNA sequence diversity in a 9.7 kb region of the human lipoprotein lipase gene. Nat Genet 19:233–240

Olsen J (1993) Some consequences of adopting a conditional deterministic causal model in epidemiology. Eur J Public Health 3:204–209

Ott J (1991) Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore

Parascandola M, Weed DL (2001) Causation in epidemiology. J Epidemiol Community Health 55:905–912

Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, Jin L, McKeigue PM, Kamboh MI, Ferrell RE, Pollitzer WS, Shriver MD (2001) Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. Am J Phys Anthropol 114:18–29

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD (1998) Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 63: 1839–1851

Phillips CV, Goodman, KJ (2001) The messed lessons of Sir Austin Bradford Hill. Congress of Epidemiology, Toronto, p 765

Phillips PC (1999) From complex traits to complex alleles. Trends Genet 15:6–8

Popper K (1961) The logic of scientific discovery. Science Editions, New York

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220–228

Rice J, Saccone N, Rasmusssen E (2001) Definition of the phenotype. In: Rao D, Province M (eds) Genetic dissection of complex traits. Academic Press, San Diego, pp 31–36

Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, Drumm M, Iannuzzi MC, Collins FS, Tsui LC (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245:1066–1073

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Rivers T (1937) Viruses and Koch's postulates. J Bacteriol 33: 1–12

Rosenbaum P (1995) Observational studies. Springer-Verlag, New York

Rubin D (1991) Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. Biometrics 47:1213–1234

Sing CF, Stengard JH, Kardia SL (2003) Genes, environment, and cardiovascular disease. Arterioscler Thromb Vasc Biol 23: 1190–1196

Stephens JC, Briscoe D, O'Brien SJ (1998) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet 55:809–824

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100:9440–9445

Terwilliger JD, Goring HH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 72:63–132

Thomas DC, Witte JS (2002) Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev 11:505–512

Thorisson GA, Stein LD (2003) The SNP Consortium website: past, present and future. Nucleic Acids Res 31:124–127

Thyagarajan T, Totey S, Danton MJ, Kulkarni AB (2003) Ge-

netically altered mouse models: the good, the bad, and the ugly. Crit Rev Oral Biol Med 14:154–174

Uhl GR, Liu QR, Naiman D (2002) Substance abuse vulnerability loci: converging genome scanning data. Trends Genet 18:420–425

Vieland VJ (1998) Bayesian linkage analysis, or: how I learned to stop worrying and love the posterior probability of linkage. Am J Hum Genet 63:947–954

Weeks DE, Sobel E, O'Connell JR, Lange K (1995) Computer programs for multilocus haplotyping of general pedigrees. Am J Hum Genet 56:1506–1507

Whitte JS, Elston ROC, Schork NJ (1996) Genetic dissection of complex traits. Nat Genet 12:355–357

Wright AF, Carothers AD, Pirastu M (1999) Population choice in mapping genes for complex diseases. Nat Genet 23:397–404