

INVITED EDITORIAL

The Emperor's New Methods

M. Anne Spence,¹ David A. Greenberg,² Susan E. Hodge,² and Veronica J. Vieland³

¹Division of Human Genetics, Department of Pediatrics, University of California, Irvine; ²Division of Statistical Genetics, Department of Biostatistics, Mailman School of Public Health, Columbia University, and Clinical-Genetic Epidemiology Unit, New York State Psychiatric Institute, New York, NY; and ³Division of Statistical Genetics, Department of Biostatistics, College of Public Health, and Department of Psychiatry, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City

Grants rejected on the basis of myth? Papers rejected for failing to adhere to dogma? Huge projects launched on the strength of personality cults? This doesn't describe *our* scientific community, does it? Unfortunately, we think that it may and that this situation is damaging both the science and the culture of human genetics research. We fear that some of the "emperor's new methods" are truly without substance. Here we will share our concerns with readers of *The American Journal of Human Genetics*.

Consider the historical example of bilineal pedigrees in gene-mapping studies. In 1989, the National Institute of Mental Health (NIMH) released a Request for Proposal (RFP) for what became known as the "genebank" project, a plan to collect families with three particular psychiatric diseases (Alzheimer disease, schizophrenia, and bipolar disorder). One group preparing to apply for the RFP decided that the best way to collect families for linkage analysis was to exclude bilineal pedigrees—that is, families with the disease on both sides of the family. The next thing we knew, NIMH study sections began requiring exclusion of bilineal pedigrees as a *sine qua non* of gene-mapping studies. The problem was that, at that time, there were *no* actual data in support of this practice: excluding bilineal pedigrees may have seemed commonsensical, but it had no empirical basis.

The issue became so prominent in discussions of study designs in psychiatric genetics that, in the early 1990s, the MacArthur Foundation commissioned a task force to investigate the effect of bilineal pedigrees on linkage analysis. This task force's report (Spence et al. 1993) conclusively dispelled the myth that bilineal pedigrees would lead to erroneous linkage results. As a result, the human gene-mapping community ceased to view bilineality as an impediment to linkage analysis, and the issue is seldom raised in the contemporary literature.

Received January 23, 2003; accepted for publication January 28, 2003; electronically published April 18, 2003.

Address for correspondence and reprints: Dr. M. Anne Spence, Department of Pediatrics, University of California Irvine Medical Center, Building 2, 101 The City Drive, Orange, CA 92868. E-mail: maspence@uci.edu

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

© 2003 by The American Society of Human Genetics. All rights reserved.
0002-9297/2003/7205-0003\$15.00

This story illustrates the virtues of the scientific method: Once put to a rigorous, empirical test, the practice of excluding bilineal pedigrees was shown to be based on myth, and it was discontinued. Unfortunately, recently our field has not been so successful in separating fact from fiction.

Moreover, the story also illustrates what we see as three pernicious themes in current human genetics research (and we will introduce a fourth one below):

1. Any method, whether a data-collection strategy, an analysis method, or a specific computer program, once accepted, is then viewed as the *only* worthwhile method or approach.
2. Research approaches become established in the absence of relevant empirical evidence or even in the face of contradictory evidence. In this way, "myths" become established and are then treated as "facts."
3. New methods and techniques are not rigorously tested before being adopted as the methods of choice.

Theme 1: The Most Popular Approach Being Taken as the Only Acceptable One

This theme persists, though what constitutes the "only" approach can be contradictory in spirit even over short periods of time. For instance, from ~1980 to 1992, the gold-standard paradigm for gene mapping in complex disease was to collect a small number of large, highly multiplex pedigrees. The current view is just the opposite—small families are greatly preferred over larger ones, for many complex disorders. However, the truth is almost certainly that neither large multiplex pedigrees nor small families are ideal for all diseases or even for discovery and characterization of all genes related to any one disease.

On the other hand, the current practice of preferring not just small but *minimal* family structures (such as affected sib pairs [ASPs] or parents plus one-child triads) is almost certainly detrimental. Not only are reviewers reluctant to fund collection of larger pedigrees—as well as what sometimes goes along with that, the careful

phenotypic evaluation of all family members—but the mania for minimal family structures extends even to the analysis of large pedigrees. We know of federally funded projects that collect and extensively phenotype multigenerational pedigrees yet, when it comes time to analyze the data, use only ASP statistics. We are aware of one manuscript that was recommended for rejection by a reviewer on the grounds that ASP methods had not been used to analyze the data—even though the manuscript reported a *LOD score of* >20 in a sample of 97 multigenerational pedigrees! This example is extreme, yet there is no question that the current standards in the field embody an attitude towards ASP designs and analytic techniques that borders on the religious.

A second example of this theme is provided by how the field reacted to the transmission/disequilibrium test (TDT) of Spielman et al. (1993), modeled on the elegant paradigm of Falk and Rubinstein (1987). This method was originally developed as a way to overcome the problem of population stratification in association studies. Its important contribution was recognized, and more and more variations on the TDT approach were published. A prominent genetics journal even asked its editorial board whether it should stop reviewing papers that did *not* use the TDT and simply reject them out of hand. (Fortunately, the editorial board said “no.”) The drawbacks of the TDT relative to case-control studies, including difficult family-data collection and severely reduced power, were seldom discussed. Furthermore, the situation for which the TDT was designed—that is, population stratification—had never actually been shown to represent a major problem for association studies.

Then, at the ASHG meeting in San Francisco, a workshop on “TDT and Other Tests for Linkage Disequilibrium,” chaired by W. J. Ewens, finally came to the conclusion that case-control studies also had their place and that the hype surrounding the TDT was overblown. Happily, the TDT, a perfectly good method, has now, to some extent, found its place as one appropriate tool we can use to identify disease loci and alleles, but not the only one. However, the damage had already been done. Manuscripts and grants had been rejected on the basis of the semireligious adoration of the TDT combined with the vilification and dismissal of the case-control approach.

Theme 2: Scientific Practice Based on Myth Rather Than Evidence

The bilinear pedigree story recounted above provides one good example of a myth being accepted without evidence to support it. A second example is the persistent belief that the results from two-point linkage analysis are in some sense “not as good” as those from multipoint

linkage analysis. It is apparently not widely understood that the *sole* advantage of multipoint analysis is that combining marker information in the form of haplotypes may increase the information for linkage. The potential disadvantages (e.g., greater dependence on precise and accurate marker locations, more devastating effects of mistyping, etc.) are seldom discussed (but see, e.g., Terwilliger and Göring 2000). Again, we are not attacking multipoint linkage analysis as such but rather criticizing the unfounded belief that multipoint is superior to two-point analysis.

A third example is the myth that the so-called “non-parametric” linkage analysis methods are better than those based on LOD scores, since the former do not *explicitly* assume a mode of inheritance. This myth persists in the face of theoretical work and extensive simulations showing it to be incorrect or, at best, oversimplified (see, e.g., Knapp et al. 1994; Greenberg et al. 1996; Hodge 2001). Closely related to this is a fourth example, the myth of the superiority of ASPs as a sampling unit, referred to above.

Additional current myths that we could nominate include overblown beliefs about (1) the potential of association and linkage-disequilibrium designs to find genes for complex traits, (2) the ability of SNPs to solve all our problems, and (3) the power of “haplotype blocks.” We could go on, but, more to the point, we invite readers to think about this and to add their own myths. Look for beliefs for which there is little or no relevant evidence or, worse, for which the empirical literature actually *contradicts* the belief. The point is not that these beliefs are necessarily false, but that they are widely held despite the lack of supporting evidence.

Theme 3: Willingness to Establish Standards without the Protections of Rigorous Testing

This dangerous theme enables the first two to flourish. Frequently, a new method is published, possibly after being evaluated in only very narrowly circumscribed ways—perhaps with respect to a single simulated data set or with respect to a single real data set for which the truth is unknown. Yet this new method is then applied uncritically under circumstances in which it has not been evaluated and for which it may never have been designed in the first place. New computer programs are adopted without rigorous testing, particularly if they are easy to use, and very little quality control is imposed when transporting programs across platforms or when applying them in novel situations. In addition, end users, in general, know little about whether methods are accurately implemented in new programs or how to recognize when the program has failed to give the correct answer. These shoddy standards for validation and calibration of tools

almost certainly contribute to a climate in which it is extremely difficult to decide which methods are working and which are not. This deprives us in part of the single most important protective facet of empirical work: the proof should be in the pudding! However, what if one has no definition of what constitutes a palatable pudding?

Why Do Bad Things Happen to a Good Field?

What explains the emergence and persistence of these pernicious themes in the field of human genetics? As a recent editorial in *The Journal of the American Medical Association* stated, in a somewhat narrower context, “Finally, the current social context seems to exert a stronger influence on the debate than the scientific arguments...further consideration should be given to how and to why the least evidence-based claims have achieved such impressive changes in funding policy” (Fombonne 2003, pp. 88–89).

In this spirit, we hazard some guesses as to what has happened over the past decade or so. One trend has been an influx of statisticians, drawn to human genetics from their parent disciplines of mathematical statistics and biostatistics (as well as other mathematical areas), because statistical genetics is new and exciting, and it also offers new employment opportunities. Yet, as a general rule, statisticians have, at best, minimal grounding in biology. Rather than understanding the *genetic* basis of the question, they tend to look for applications of the statistical skills in which they are trained. This leads to propagation of new designs and new statistical methods that are poorly adapted to the scientific needs of the field. At the same time, many clinicians and molecular geneticists have only a rudimentary understanding of statistics, and, as a result, are prone either to rely on the simplest methods (which they feel they understand) or to rely on the recommendations of certain experts.

Theme 4: The Unfortunate Development of a “Cult of Personality”

By this last theme we do not mean reasonable, rational reliance on experts, which is indeed the basis of successful interdisciplinary collaboration. Rather, a small number of opinion makers have arisen who seem to have the ear of people in power. When these individuals give an opinion, it is as if they are speaking *ex cathedra*. Policy is determined, funds are allocated, and new directions are set for our field, with little or no open public discussion and sometimes in the complete absence of any empirical evidence. The cult of personality may be our

field’s dirty little secret; it is awkward to discuss, but we must find a way to do it, despite our discomfort. Reliance of an entire field on the recommendations or prejudices of a handful of individuals has, in the history of science as a whole, proved to be a very poor method of moving closer to the truth.

The field is faced with a situation in which its scientific goals require enormous specialization—in the molecular, the clinical, and the statistical subsciences—as well as intense collaboration. Perhaps what has been lost over time is a common core grounding in *human genetics* itself. Without that grounding, it may be inevitable that the field is drifting away from the mooring of rigorous science toward a world in which major decisions are based on myth, conformity, and allegiance to leaders, rather than the facts.

If the bleak outlook we are voicing here has substance, what can be done? In writing this opinion piece, we hope to spark discussion and action, and we by no means claim to have all the answers ourselves. However, we end with some of our own recommendations.

First, some things we can do as individuals:

1. As reviewers of grants and manuscripts, be open to innovation and designs that vary from standard or fashionable protocols, as long as the science is well defended. There is no “sole true path.”
2. “Believe none of what you hear and only half of what you read”; retain healthy skepticism about all unproven (or even “proven”) assertions.
3. Insist that all new methods be tested before you adopt them. (One valuable resource is the Genetic Analysis Workshops, or GAWs, which have accumulated >12 simulated data sets—a tremendous aid for evaluation of new methods; see, e.g., Wijsman et al. [2001].)
4. Minimalism in music may be interesting, but minimalism in genetic data collection is disastrous. Collect as much genetic *and phenotypic* information on families as possible.

Finally, some things we should perhaps be doing as a field:

1. Design and implement rigorous training programs appropriate to the needs of human genetics in its contemporary form, addressing statistics, epidemiology, and clinical issues, as well as molecular genetics. This approach must also extend to the programs that self identify as “genome,” since they are heavily integrated with the human genetics community.
2. Devise mechanisms for the efficient evaluation and comparison of new methods, specific programs, and the performance of programs on different platforms. Two mechanisms are already in place:
 - a. The GAWs were originally intended to provide a mechanism for resolving some thorny argu-

ments regarding segregation analysis, and they were so productive that they have continued through 13 workshops. Thus, the GAW itself might provide a venue for addressing some of the problems in the field. However, GAWs occur only once every 2 years, with participants working on problems that have been distributed well in advance, thus limiting their effectiveness for answering urgent questions.

- b. Another possibility is a model used in the clinical arena—consensus conferences, convened when major points of clinical controversy arise. Some version of this mechanism might work in human genetics as well. (Indeed, the MacArthur Foundation task force mentioned above had much of this flavor to it.) Such conferences need not achieve consensus but rather summarize the evidence for and against different approaches to complex human genetics.

In conclusion, we reiterate that we are not criticizing particular methods or approaches *per se*; rather, we are criticizing *how* methods become accepted and *how* decisions are made. Our suggestions above may not be optimal, and they are certainly not exhaustive. However, if they serve as a starting point for discussions that will lead to improvements in research protocols in human genetics, we will have accomplished our immediate goal. Once the emperor's "nakedness" is pointed out, there is at least the hope of getting him some real clothes.

References

- Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* 51:227–233
- Fombonne E (2003) The prevalence of autism. *JAMA* 289: 87–89
- Greenberg DA, Hodge SE, Vieland VJ, Spence MA (1996) Affecteds-only methods are not a panacea. *Am J Hum Genet* 58:892–895
- Hodge SE (2001) Model-free vs. model-based linkage analysis: a false dichotomy? *Am J Med Genet* 105:62–64
- Knapp M, Seuchter SA, Baur MP (1994) Linkage analysis in nuclear families. 2: Relationship between affected sib-pair tests and lod score analysis. *Hum Hered* 44:44–51
- Spence MA, Bishop DT, Boehnke M, Elston RC, Falk C, Hodge SE, Ott J, Rice J, Merikangas K, Kupfer D (1993) Methodological issues in linkage analyses for psychiatric disorders: secular trends, assortative mating, bilineal pedigrees. Report of the MacArthur Foundation Network I Task Force on Methodological Issues. *Hum Hered* 43:166–172
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516
- Terwilliger JD, Göring HHH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol* 72:63–132
- Wijsman EM, Almasy L, Amos CI, Borecki IB, Falk CT, King TM, Martinez MM, Meyers DA, Neuman RJ, Olson JM, Rich SS, Spence MA, Thomas DC, Vieland VJ, Witte JS, MacCluer JW (2001) Genetic Analysis Workshop 12: analysis of complex genetic traits: applications to asthma and simulated data. *Genet Epidemiol Suppl* 21