

INVITED EDITORIAL

'Twixt Cup and Lip: How Intractable Is the Ascertainment Problem?

Robert C. Elston

Department of Biometry and Genetics and the Center for Molecular and Human Genetics,
Louisiana State University Medical Center, New Orleans

"It is a statistical commonplace that the interpretation of a body of data requires a knowledge of how it was obtained." So wrote Fisher (1934, p. 1) in the paper where he introduced the likelihood of the classical ascertainment model for segregation analysis of sibship data. He went on to say, "Nevertheless, in human genetics especially, statistical methods are sometimes put forward, and their respective claims advocated with entire disregard for the conditions of ascertainment." Unfortunately this is still true today, though more in what data analysts write than in what those developing new statistical methods write. Whenever data are analyzed that are not randomly drawn from a well-defined population, the problem of ascertainment can arise. Ascertainment is relevant not only in segregation analyses, but also in the estimation of familial correlations and linkage analyses. It is true that in certain circumstances (indeed, most circumstances in the past) it has been appropriate to formulate linkage analyses in a manner that makes it unnecessary to take ascertainment into account. However, more and more as we use linkage—especially, tight linkage—to unravel the genetics of multifactorial diseases, this will no longer be the case. Therefore, because the data collected for this purpose often comprise ascertained family structures more extensive than full sibships, it is timely that in this issue of the *Journal* Vieland and Hodge (1995) examine a fundamental problem that can arise when one tries to analyze such data.

Before discussing what Vieland and Hodge have shown, a note about terminology will be helpful. They have chosen to follow terminology introduced by Morton (1959), which I think is unfortunate. In the classical model, we assume there is probability π that an affected person becomes a proband, i.e., a person who independently of all

other persons causes a particular sibship to enter the sample. As π tends toward 0, we have what Morton, as well as Vieland and Hodge, call "single selection." I (and others) much prefer the term "single ascertainment," because selection has long had a totally different connotation in genetics. In fact, Vieland and Hodge assume in their paper (for the sake of simplicity only, not because the assumption is necessary for their result) no selection in the usual genetic sense. I (and others) also follow Fisher (1934) in calling the case $\pi = 1$ "complete ascertainment," while Morton called this "truncate selection" and used the term "complete selection" to indicate ascertainment of a sibship through the parents. Although careful use of either terminology should cause no confusion—the same concept may be described by two different terms (complete ascertainment \equiv truncate selection), but two different concepts are not described by the same term—I would make a plea to avoid use of the word "selection" to describe the mode of ascertainment.

What Vieland and Hodge show is that it is not feasible to allow for the mode of ascertainment in an exact fashion, when the pedigree structure sampled is more extensive than a full sibship, except in three situations: (i) the probability that the pedigree is ascertained is not a function of the genetic model parameters; (ii) the particular pedigree structure sampled is not dependent on which members of the pedigree are involved in the ascertainment event(s); and (iii) there is single ascertainment, so that not more than one ascertainment event could lead to the pedigree being sampled. (In the classical model, becoming a proband is the ascertainment event; more generally multiplex ascertainment events, such as the presence of two affected siblings, are possible.) The first situation corresponds to random sampling of pedigrees, and so the ascertainment problem does not arise. It is the last two situations, which are related to each other, that are difficult to achieve.

Elston and Sobel (1979) gave a solution to the ascertainment problem for pedigrees, by assuming (1) that one can identify which members would be eligible, if they had the appropriate phenotype, to become probands; and (2) that, whichever one of these pedigree members were a proband,

Received November 3, 1994; accepted for publication November 3, 1994.

Address for correspondence and reprints: Dr. Robert C. Elston, Department of Biometry and Genetics, LSU Medical Center, 1901 Perdido Street, New Orleans, LA 70112-1393.

© 1995 by The American Society of Human Genetics. All rights reserved.
0002-9297/95/5601-0003\$02.00

exactly the same pedigree structure would be sampled. This, Vieland and Hodge point out, cannot occur if we merely stipulate a simple structure for all the pedigrees to be sampled, unless that structure is the full sibship. Similarly, if one uses the Cannings and Thompson (1977) sequential sampling scheme, there is no way to stop two independently ascertained pedigrees joining up; but the correctness of the likelihood that Cannings and Thompson derived depends critically on the implicit assumption of single ascertainment.

It is important to understand the framework upon which Vieland and Hodge build their proof. They start by assuming that there exists a population of discrete pedigrees (after all, they argue, if discrete pedigree structures do not exist, the problem will only be made worse) and then mimic the ascertainment process by sampling only *parts* of these pedigrees. Thus, if we can in fact define a population of discrete pedigrees, and if we can be sure that, each time we ascertain a pedigree, the whole pedigree—not just a part of it—enters the sample, we will be in situation (ii), and the problem will disappear. With a little ingenuity, this is not impossible. If we stipulate that we want a sample of nuclear families and try to achieve this by sampling probands, their sibs, and their parents, there is the possibility that two such nuclear families will join up to form a three-generation pedigree; in this case we would not be in situation (ii). But the possibility of two families joining up can be avoided by requiring that all probands be born within a specified decade. As another example, suppose that our proband sampling frame (i.e., the set of all persons who would be eligible, with the appropriate phenotype, to become probands) consists of female persons 45–50 years old and that we sample pedigrees each consisting of the proband, every full sib and maternal half sib of the proband, the children of all these persons, and the proband's mother; then every proband belongs to exactly one such pedigree structure, and any two probands in the same pedigree would lead to exactly the same pedigree being ascertained. In a population in which there is lifetime monogamy, so that half-sibs do not exist, it is possible to define even larger discrete pedigree structures. Provided that a population of such pedigree structures exists, we can make inferences about it, using the Elston and Sobel approach.

Vieland and Hodge also state that the ascertainment-assumption-free approach of conditioning the likelihood on that part of the data that is “relevant to ascertainment” (Ewens and Shute 1986; Shute and Ewens 1988) is correct only in situation (ii), because otherwise it would be necessary to condition the likelihood on unobserved phenotypes—in particular, the phenotypes of the pedigree members not sampled under their framework. Here again, the problem may be overcome with a little ingenuity. If we define a small enough proband sampling frame—say, all

the students in a particular school who undergo some kind of screening—there is no problem in conditioning the likelihood on the phenotypes of all members of the proband sampling frame; and we are then back in situation (ii).

The above examples show that we must be cautious in concluding that the ascertainment problem is inherently intractable for pedigree data. Indeed, Vieland and Hodge themselves point out that one way of overcoming the problem is to include in the final sample every person in the proband sampling frame (they use the phrase “a well-delineated catchment area”). There is to my mind no doubt about the main result shown in their paper, which I have summarized above. These results imply that exact ascertainment correction will be difficult, if not impossible, for any situation in which separately sampled pedigrees can join up. But some of the other conclusions drawn are less convincing. Vieland and Hodge conclude, for example, that the cause of the problem is the fact that all humans are related. I do not believe that this is so. I do not see any need for the sampling units in a population to be independent when one is formulating a relevant likelihood. Furthermore, if we sample nuclear families as indicated above, so that no two nuclear families can join up, the fact that the sample may then contain cousins does not lead to an inherent ascertainment problem. The real difficulty is how to formulate a relevant likelihood when there is no well-defined population of discrete sampling units. One possibility is to sample fixed pedigree structures via probands and (wrongly) to assume independence of these structures when one is formulating the likelihood. In the case of nuclear families, for example, the likelihood for every nuclear family is entered in its entirety as a factor in the overall likelihood, and the fact that some individuals are thus entered twice is ignored. Such a likelihood cannot be used in a simple manner to test hypotheses or to derive standard errors of estimates, but it will nevertheless lead to consistent estimates if the appropriate ascertainment correction (for probands in sibships) is included.

Of course, I do not believe that the ascertainment problem for pedigrees will always be tractable, and I agree with Vieland and Hodge that it will be necessary to study the utility of various approximations. It has been clear for some time that the “bias due to ascertainment” decreases as the number of probands in the sample, expressed as a proportion of those with similar phenotype, decreases (Elston 1979). The importance of this new theoretical framework described by Vieland and Hodge lies not so much in all the conclusions that they draw from it as in the insights that it leads to. By shedding light on just what breaks down when we ascertain large pedigrees, their main result can guide us toward sampling and analytical techniques that can minimize the problem. There's many a slip, a Greek bard wrote, 'twixt cup and lip. But that does not mean that there is no value to the contents of the cup.

Acknowledgments

This work was supported in part by U.S. Public Health Services research grant GM 28356 from the National Institute of General Medical Sciences. I thank Veronica Vieland and Susan Hodge for helpful comments on an early draft of this editorial.

References

- Cannings C, Thompson EA (1977) Ascertainment in the sequential sampling of pedigrees. *Clin Genet* 12:208–212
- Elston RC (1979) Likelihood models in human quantitative genetics. In: Sing CF, Skolnick M (eds) *Genetic analysis of common diseases: applications to predictive factors in coronary heart disease*. Alan R Liss, New York, pp 391–405
- Elston RC, Sobel E (1979) Sampling considerations in the gathering and analysis of pedigree data. *Am J Hum Genet* 31:62–69
- Ewens WJ, Shute NCE (1986) A resolution of the ascertainment sampling problem. *Theor Popul Biol* 30:388–412
- Fisher RA (1934) The effect of methods of ascertainment upon the estimation of frequencies. *Ann Eugenics* 6:13–25
- Morton NE (1959) Genetic tests under incomplete ascertainment. *Am J Hum Genet* 11:1–16
- Shute NCE, Ewens WJ (1988) A resolution of the ascertainment sampling problem. III. Pedigrees. *Am J Hum Genet* 43:387–395
- Vieland VJ, Hodge SE (1995) Inherent intractability of the ascertainment problem for pedigree data: a general likelihood framework. *Am J Hum Genet* 56:33–43 (in this issue)