Letters to the Editor

Am. J. Hum. Genet. 68:807–810, 2001

Estimation of Sibling Recurrence-Risk Ratio under Single Ascertainment in Two-Child Families

To the Editor:

Guo (1998) has examined the behavior of the sibling recurrence-risk ratio λ_S (defined as the ratio of disease manifestation, given that one's sibling is affected, compared with disease prevalence in the general population) under ascertainment models. He concludes that, for a fictitious, strictly random (i.e., nongenetic and nonenvironmental) disease in which the sibling risks of disease are identical and independent, estimates of λ_S can be dramatically inflated because of ascertainment bias. One situation that he identifies as being susceptible to bias is the case of single ascertainment. However, we will show here that, at least for two-child families (we do not examine larger families), there is no ascertainment bias under single ascertainment. We argue that Guo finds bias in this situation only because he permits the sibling *who is being conditioned on* (hereafter referred to as the "index subject") to be other than the sibling *through whom the family is ascertained* (the "proband"). This leads to a definition of "sibling recurrence risk in ascertained families" that does not correspond to the definition that researchers in fact use. We do not dispute Guo's larger claim of ascertainment bias in λ_S; we dispute only the particular case of single-ascertainment bias in two-child families.

As stated above, we consider only two-child families ($k = 2$) in the present study. Following Guo, we let X_i represent the random variable denoting the affectedness status of a sibling, where $X_i = 1$ if sibling "i" is affected and $X_i = 0$ if sibling "i" is not affected, for $i = 1, 2$. Without loss of generality, we let $i = 1$ denote the older sibling and let $i = 2$ denote the younger sibling, and we let the sibling recurrence risk represent the risk that the younger sibling will be affected, given that the older sibling is affected. We hereafter use the term "index subject" explicitly to mean the sibling being conditioned on, even though this is not standard terminology.

Following the notation of Sham (1998), we denote this recurrence risk "K_R"; thus, without loss of generality, $K_R = P(X_2 = 1 \mid X_1 = 1)$. Then $\lambda_S = K_R/K$, where K is the population prevalence of the disease. Since the denominator of λ_S—that is, the population prevalence K—is not in question, the real issue is whether the numerator—that is, the sibling recurrence risk K_R itself—is biased under single ascertainment. We show that, under generally accepted definitions of sibling recurrence risk in ascertained families, estimates of this risk are *not* biased. (Throughout, we use "[un]biased" to mean asymptotically [un]biased.)

To begin, we assume *interchangeability* (as does Guo 1998, p. 253), defined here as a lack of birth-order effect. Thus, the marginal probability of being affected is the same for the two siblings:

$$P(X_1 = 1) = P(X_2 = 1) \; . \tag{1}$$

It follows from equation (1) that

$$K_R = P(X_2 = 1 \mid X_1 = 1) = \frac{P(X_2 = 1,\, X_1 = 1)}{P(X_1 = 1)}$$

$$= \frac{P(X_2 = 1,\, X_1 = 1)}{P(X_2 = 1)} = P(X_1 = 1 \mid X_2 = 1) \; .$$

Therefore, sibling recurrence risk may be represented as

$$K_R = P(X_j = 1 \mid X_i = 1),$$
$$\text{for } j = 1, 2;\; i = 1, 2;\; i \neq j \; . \tag{2}$$

So far, we agree with Guo; our equation (2) is the same as the numerator of his equation (1).

By "single ascertainment" we mean that the probability that any affected individual will become a proband is a very small value, denoted by "π," where $0 < \pi \ll 1$. We define A_k to be the event that a family is ascertained through sibling k (i.e., the event that sibling k is a proband), where $k = 1, 2$ and $A = A_1 \cup A_2$, as used by Guo. The distinction between an "index subject" (the sibling being conditioned on) and a "proband" (the sibling through whom the family is ascertained) is important and should be kept in mind in the following discussion. (We frame our discussion in terms of "probands," but our argument in this letter would be equally valid with a proband-free definition of single ascertainment; details are not included here but are available from

the authors; also see Morton 1959; Stene 1977; Ewens and Shute 1986; Hodge and Vieland 1996.)

Our disagreement with Guo concerns the way in which ascertainment is incorporated into the concept of the sibling recurrence risk. To avoid becoming entangled in semantic arguments, we include a diagram (fig. 1), for the sake of clarity. This probability tree illustrates all possible outcomes for two-child families, under the assumptions of single ascertainment. Of the eight possible outcomes, only four are ascertained, and they are numbered "{1}"–"{4}"; thus, {1} indicates that only X_1 is affected and that X_1 is a proband; {2} indicates that only X_2 is affected and that X_2 is a proband; {3} indicates that both siblings are affected and that X_1 is a proband; and {4} indicates that both siblings are affected and that X_2 is a proband. Because of the interchangeability assumption, $P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1)$.

Now the question is, How does one define "sibling recurrence risk in ascertained families?" Guo gives two possibilities, in his equations (2) and (3), but they prove to be identical when $k = 2$ and when X_1 and X_2 are interchangeable. Thus we discuss only his equation (2):

$$K_R^* = P(X_2 = 1 | X_1 = 1, \mathrm{A}) \ . \qquad (3)$$

(This corresponds to the numerator of his eq. [2] when $k = 2$; and it also equals the numerator of his eq. [3] when $k = 2$, because of eq. [1].) This expression *is* biased, as Guo claims; from figure 1, the definition in equation (3) yields

$$K_R^* = \frac{P(\{3\}) + P(\{4\})}{P(\{1\}) + P(\{3\}) + P(\{4\})}$$
$$= \frac{P(X_1 = 1, X_2 = 1)2\pi}{P(X_1 = 1, X_2 = 0)\pi + P(X_1 = 1, X_2 = 1)2\pi} \ .$$

Straightforward probability calculations reveal that this equals

$$K_R^* = \frac{P(X_1 = 1, X_2 = 1)}{P(X_1 = 1) - (\frac{1}{2})P(X_1 = 1, X_2 = 0)}$$
$$> P(X_2 = 1 | X_1 = 1) = K_R$$

and thus that it is biased.

However, we maintain that this is not the usual way in which sibling recurrence risk in ascertained families is defined; rather, examination of figure 1 reveals that equation (3) would lead to ascertainment of families through *either* affected child but would only express the recurrence risk in terms of conditioning child 2 on child 1 (e.g., always conditioning the younger child on the older, regardless of which child is the proband). In other words, Guo's definition allows for the possibility that one child could be the "proband" *without* being the "index subject."

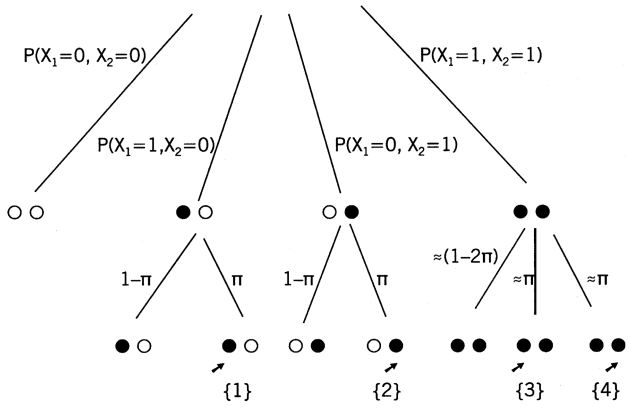

Figure 1 Probability tree illustrating all possible outcomes for a two-child family, under single ascertainment. Numbers along the branches indicate the probability of each outcome. Blackened circles (●) indicate affected children, unblackened circles (○) indicate unaffected children, and arrows (↗) indicate probands. Under the assumptions of single ascertainment, a family can have, at most, one proband. The numbered outcomes, {1}–{4}, indicate the types of ascertained families.

However, all studies known to us condition on the affectedness status *of the proband;* that is, they always treat the proband *as* the "index subject." With this approach, sibling recurrence risk in ascertained families would be defined in one of two ways. The first definition, applicable when the investigator specifies a priori that a particular child (e.g., the older child) must be both the proband and the index subject, is

$$K_R^* = P(X_j = 1 | X_i = 1, \mathrm{A}_i) \ . \qquad (4)$$

The expression in equation (4) could represent either $P(X_2 = 1 | X_1 = 1, \mathrm{A}_1)$ or $P(X_1 = 1 | X_2 = 1, \mathrm{A}_2)$, since they are assumed to be equal. We illustrate with the first expression, $P(X_2 = 1 | X_1 = 1, \mathrm{A}_1)$. From figure 1 this equals

$$\frac{P(\{3\})}{P(\{1\}) + P(\{3\})}$$
$$= \frac{P(X_1 = 1, X_2 = 1)\pi}{P(X_1 = 1, X_2 = 0)\pi + P(X_1 = 1, X_2 = 1)\pi}$$
$$= \frac{P(X_1 = 1, X_2 = 1)}{P(X_1 = 1)}$$
$$= P(X_2 = 1 | X_1 = 1) = K_R$$

and thus is not biased.

The second definition, which is applicable when the

investigator allows *either* child to be both the proband and the index subject, is

$$K_R^* = P[X_1 = 1,\ X_2 = 1 \,|\, (X_1 = 1,\ A_1) \cup (X_2 = 1,\ A_2)]\ . \tag{5}$$

From figure 1, this equals

$$\frac{P(\{3\}) + P(\{4\})}{P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\})}$$

$$= \frac{P(X_1=1,X_2=1)2\pi}{P(X_1=1,X_2=0)\pi + P(X_1=0,X_2=1)\pi + P(X_1=1,X_2=1)2\pi}$$

$$= P(X_2=1\,|\,X_1=1) = K_R\ ,$$

which also is unbiased. Thus, both formulations of K_R^*—our equations (4) and (5)—equal K_R, that is, $P(X_2 = 1 \,|\, X_1 = 1)$; and both are *unbiased*. Hereafter, we discuss only equation (5).

To make the differences between Guo's definition, in our equation (3), and our definition in equation (5) concrete, consider a numerical example in which, again, X_1 always represents the older child and X_2 represents the younger child. In the example, we set $p = .2$, where p represents the population's disease frequency as defined by Guo. Thus, $P(X_1 = 1,\ X_2 = 0) = P(X_1 = 0,\ X_2 = 1) = .16$, and $P(X_1 = 1,\ X_2 = 1) = .04$. To illustrate, we consider "perfect samples" from this numerical example, and we show the results that are obtained when our definition of K_R^* in equation (5) is used, compared with the results that are obtained when Guo's definition, in equation (3), is used. (We encourage readers, in order to fully understand the differences between the two definitions, to work through these examples for themselves.)

1. Using our definition of K_R^*, in equation (5), one would ascertain $.40\pi$ families: in half of those families (i.e., $.20\pi$) the older child would be both the proband and the index subject, and in 20% of that half (i.e., $.04\pi$) the younger child would also be affected; in the other half the younger child would be both the proband and the index subject, and in 20% of that half the older child would also be affected. Thus, the observed proportion of 20% sibling recurrence risk would be unbiased.
2. However, using Guo's definition of K_R^*, in our equation (3), one would ascertain *all* families in which X_1 is affected, regardless of whether X_1 is the proband; that is, one would ascertain $.32\pi$ families, corresponding to outcomes {1}, {3}, and {4} in the figure. Nevertheless, one would always ask whether the younger child were also affected—not only in outcomes {1} and {3}, in which X_1 is the proband, but also in outcome {4}, in which X_2 is the proband and would, of course, be affected. Thus, in some of the families (those corresponding to outcome {4}), X_2 would be the *proband*, whereas X_1 would be the *index subject*. This would yield $.08\pi$ families, or a proportion of 25%, a result that would be inflated or biased.

Note that our demonstration that both equation (4) and equation (5) are unbiased does not actually require Guo's assumption of a strictly random disease with population prevalence p. We did use that model in the numerical example, but the derivations that show equations (4) and (5) to be unbiased require only the assumption of interchangeability in equation (1).

The critical difference between our definition in equation (5) and Guo's definition, in equation (3), is that Guo seems to assume that the affected "index subject" on whom he is conditioning is independent of the ascertainment process—that is, that *the sibling through whom the family is ascertained is not necessarily the sibling who is considered to be the index subject.*

Thus, as shown in our equation (3), Guo conditions only one specified sibling on the other sibling (i.e., he allows only one sibling to be the "index subject"), but he allows ascertainment through *either* sibling (i.e., he allows either sibling to be the "proband"). It is this inconsistency that leads to the unnecessary bias in his formula. In contrast, in our formulation of K_R^*, given in equation (4) or equation (5), the index subject and the proband are always the same.

Note, too, that if one did want to relax the interchangeability assumption and to consider the situation in which $P(X_2 = 1) \neq P(X_1 = 1)$, then one would use the K_R^* definition that is given in our equation (4). *Only* families that are ascertained through X_1 would be considered—that is, only X_1 would be considered as the proband/index subject—and, again, there would be no bias.

Guo mentions several studies in support of his definition of K_R^*. He cites two genetic epidemiological studies of homosexuality (Pillard and Weinrich 1986; Bailey and Benishay 1993) as using the definition in his equation (2) and two studies of prostate cancer (Monroe et al. 1995; Narod et al. 1995) as using the definition in his equation (3). However, when we read them, it seems clear to us that all four studies are using the definition in our equation (5) and are not using either of Guo's formulations. Moreover, it is difficult for us to imagine a genetic or family study that *would* use Guo's definition.

In conclusion, in two-child families, the designation "index subject" for the sibling on whose affected status we are conditioning does not give rise to bias in K_R when the mode of ascertainment is single; rather, the cause of the bias observed by Guo is the fact that he allows for

an "index subject" who is not simultaneously a "proband." Moreover, our conclusions hold regardless of whether the marginal probabilities of the two siblings being affected are equal: if they are equal, then conditioning on the proband will give an unbiased estimate (our eq. [5]); if they are not equal, then conditioning on the proband who has been defined as such prior to ascertainment will *also* give an unbiased estimate (our eq. [4]).

Finally, to end on a more positive note, we look at the broader context of Guo's paper. We have investigated only two-child families ($k = 2$), and we have investigated those selected under single ascertainment only. Guo may be correct in asserting that there is ascertainment bias in larger families. Moreover, we are certainly not defending the use of λ_S, since we suspect that this measure probably *is* subject to ascertainment bias when ascertainment is other than single (also see Olson and Cordell 2000). We applaud Guo's work on this subject but believe that it will be more useful if he clarifies the definition of sibling recurrence risk in ascertained families.

Acknowledgments

This work was supported by National Institutes of Health grants MH-57813 (to P.J.W.) and MH-48858 and DK-31813 (both to S.E.H.).

PRIYA J. WICKRAMARATNE AND SUSAN E. HODGE
Department of Psychiatry, College of Physicians and Surgeons, and Division of Biostatistics, Mailman School of Public Health, Columbia University, and Division of Clinical and Genetic Epidemiology, New York State Psychiatric Institute, New York

References

Bailey JM, Benishay DS (1993) Familial aggregation of female sexual orientation. Am J Psychiatry 150:272–277

Ewens WJ, Shute NCE (1986) The limits of ascertainment. Ann Hum Genet 50:399–402

Guo SW (1998) Inflation of sibling recurrence-risk ratio, due to ascertainment bias and/or overreporting. Am J Hum Genet 63:252–258

Hodge SE, Vieland VJ (1996) The essence of single ascertainment. Genetics 144:1215–1223

Monroe KR, Yu MC, Kolonel LN, Coetzee GA, Wilkens LR, Ross RK, Henderson BE (1995) Evidence of an X-linked or recessive genetic prostate cancer risk. Nat Med 1:827–829

Morton NE (1959) Genetic tests under incomplete ascertainment. Am J Hum Genet 11:1–16

Narod SA, Dupont A, Cusan L, Diamond P, Gomez J-L, Suburu R, Labrie F (1995) The impact of family history on early detection of prostate cancer. Nat Med 1:99–101

Olson JM, Cordell HJ (2000) Ascertainment bias in the estimation of sibling genetic risk parameters. Genet Epidemiol 18:217–235

Pillard RC, Weinrich JD (1986) Evidence of familial nature of male homosexuality. Arch Gen Psychiatry 43:808–812

Sham P (1998) Statistics in human genetics. Arnold, London

Stene J (1977) Assumptions for different ascertainment models in human genetics. Biometrics 33:523–527

Address for correspondence and reprints: Dr. Priya Wickramaratne, NYSPI—Unit 24, 1051 Riverside Drive, New York, NY 10032. E-mail: wickramp@child.cpmc.columbia.edu

0002-9297/2001/6803-0028$02.00

Am. J. Hum. Genet. 68:810–812, 2001

Reply to Wickramaratne and Hodge

To the Editor:

Sibling recurrence-risk ratio is perhaps the most widely used measure for familial aggregation of complex diseases and is often used as a measure of genetic effect. It is frequently used in power calculations in gene-mapping studies and in exclusion mapping. It is little known among human geneticists, however, that familial clustering of risk factors of an environmental nature also can elevate sibling recurrence-risk ratio, even in the complete absence of any genetic (hereditary) factors (Guo 2000*a*). In addition, ascertainment bias in estimation of sibling recurrence-risk ratio is frequently overlooked or simply ignored in genetic epidemiological studies (Guo 1998). I am very pleased to see the letter of Wickramaratne and Hodge (henceforth, "W&H"), which attempts to further take up this issue.

As a measure of familial aggregation, the original definition of sibling recurrence-risk ratio (see, e.g., Risch 1990) is very intuitive and appealing. My paper (Guo 1998) demonstrates that, when the actual use of this measure deviates from its original definition (i.e., definitions 2 and 3 in Guo 1998), the estimation of sibling recurrence-risk ratio can be artificially inflated if there is ascertainment bias and/or overreporting under single and multiple ascertainment schemes. It should be noted that I did not define sibling recurrence risk in ascertained families, as claimed by W&H, but, rather, that I pointed out the consequence of "misunderstanding of the original definition of λ_S" (Guo 1998).

W&H assert that Guo (1998) finds bias "only because he permits the sibling *who is being conditioned on*… to be other than the sibling *through whom the family is ascertained,*" and that my definition "does not correspond to the definition that researchers in fact use," at least for two-child families. Furthermore, they claim that, under their definition, when the sibship size (k) is two, there is no ascertainment bias under single ascertainment in the estimation.