



## Proband Concordance vs. Casewise Concordance

Z. Hrubec

*Radiation Epidemiology Branch, National Cancer Institute, Bethesda, Maryland*

I read with enjoyment the excellent article by Hannah et al (*Acta Genet Med Gemelloi* 32:127-137 [3]). The work contributes appreciably to the problem of handling environmental covariables in genetic analyses. It is regrettable that in characterizing the proband concordance rate the authors have helped to spread a common misunderstanding about the rate, reflected in their statement on page 127:

"... to overcome problems associated with the ascertainment of affected individuals it is usual to base inference on the proband concordance rate defined as the proportion of co-twins with the trait for affected individuals independently ascertained. This definition is simple in practice and leads to an estimate of concordance which is independent of the ascertainment probability [11]".

The reference in the quote is to [6].

The proband concordance rate, as Hannah et al define it, can be independent of ascertainment only if it incorporates information on ascertainment completeness obtained through an essentially complete secondary ascertainment process. The description of methods in the article is not explicit on this point. There is no mention of secondary ascertainment, and it is most likely that it was not carried out. The general problem of ascertainment in evaluations of concordance has been formulated [4], discussed [5], and treated in detail [1].

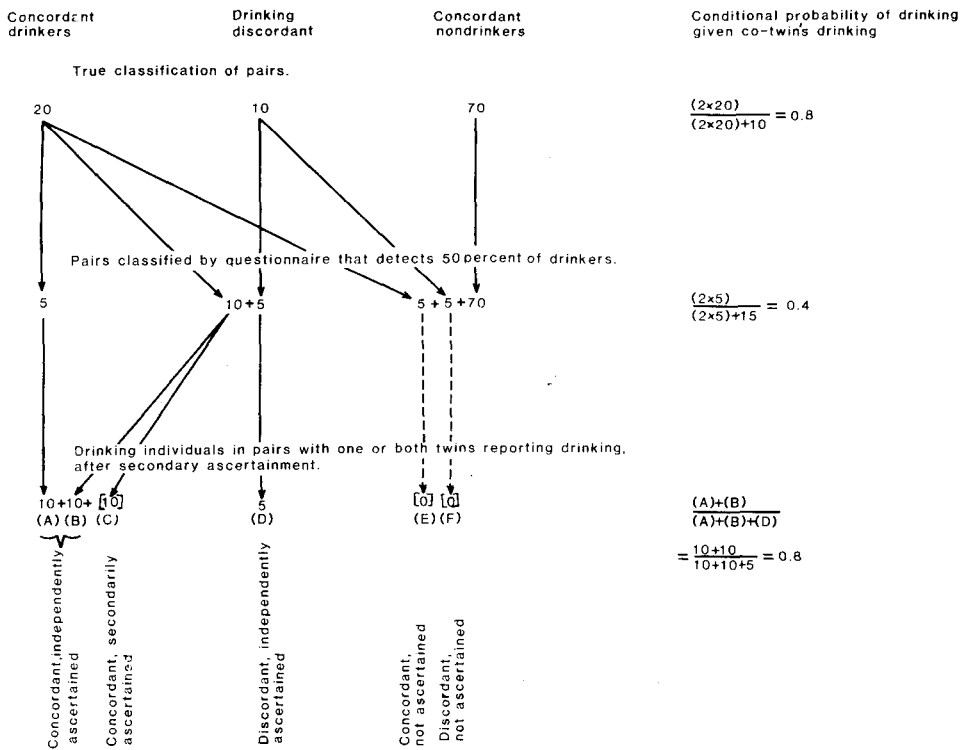
A simple example, expressed in terms of the questionnaire instrument used by Hannah et al and represented in the Figure, illustrates the problem. Suppose that among 100 twin pairs, or 200 individuals, completing the questionnaire, there actually are 20 pairs of concordant drinkers and 10 other drinking individuals whose cotwins are nondrinkers. Thus the prevalence of drinking among individuals is 50/200 and the conditional probability of an individual's drinking, given that his cotwin drinks, is 40/50. The proband concordance rate was designed to estimate this last quantity.

Suppose that a questionnaire has been used which is imperfect, and detects only one half of the drinkers. To keep things simple also suppose that the questionnaire does not classify any nondrinkers as drinkers, is administered to members of twin pairs independent-

ly, and responses are uncorrelated within twin pairs. Of the 200 individual respondents, 25 admit drinking on the questionnaire (prevalence 25/200). Among the 20 drinking concordant pairs, in 5 pairs both admit drinking and are correctly classified as concordant drinkers; in 10 pairs one twin but not the other admits drinking; and in 5 pairs neither admits drinking. Thus, the 20 truly concordant pairs contribute 20 individuals who admit drinking, 10 of them in admission concordant and 10 in admission discordant pairs. Of the 10 drinkers originally in discordant pairs, 5 admit drinking. Altogether, we now have 5 pairs who concordantly admit drinking and 10 + 5 other admitted drinkers. The conditional probability of admitted drinking given cotwin's admitted drinking then appears to be 10/25, or one half of the correct value.

Suppose that we subsequently reevaluate the 20 pairs *in which one or both twins admitted drinking* with an infallible laboratory test of weekly alcohol use. We will find that among the 15 pairs who reported discordant drinking on the questionnaire there are in fact 10 concordant pairs for whom drinking was "independently ascertained" by questionnaire in only one twin. The proband concordance is thus computed as (10 + 10)/(10 + 10 + 5) which now corresponds to the original 40/50.

Classification of twins in computing concordance with 50% ascertainment completeness



Figure

Several comments are in order.

- 1) In the above example, as is typical, secondary ascertainment, was applied only to pairs with drinkers detected in the primary ascertainment by questionnaire. The secondary ascertainment found 10 pairs, originally classified as discordant who are in fact concordant. The 10 independently ascertained individuals in them, who reported drinking initially, provide an estimate of ascertainment completeness that makes it possible to correct for the incomplete primary ascertainment. Without this secondary ascertainment information, the estimate obtained is proportional to the ascertainment probability.
- 2) The term proband concordance implies that appropriate information from secondary ascertainment has been used. If such information is not available, or has not been correctly employed, the designation is misleading. The term casewise concordance has been developed to provide a distinction between the corrected and uncorrected estimates [2]. It is an artificial and somewhat awkward construction, but its consistent use would avoid the confusion to which Hannah et al fell prey.
- 3) The 50% ascertainment probability in the example was chosen for simplicity, and it is not meant to imply anything about the accuracy of the instrument Hannah et al administered. Conceivably, their instrument detects all drinkers as they define them. In that situation, the formulas for the proband and the casewise rate are algebraic identities. Even then, it may be useful to highlight the absence of secondary ascertainment by referring to the estimated conditional probability as a casewise rate.
- 4) In the absence of secondary ascertainment, there is no distinction between the casewise and the proband rate, and the question of the correct designation is only one of semantics. If other prerequisite assumptions are not met, designating the index as a “proband rate” confers no additional advantage, the “proband” and the “casewise” rates will be equally in error. The simple model exemplified above is easily modified to require less restrictive assumptions [1].

As a minor point, there are inaccuracies in the estimates of the  $r \pm \sqrt{V_r}$  for the threshold liability model shown in Table 5. According to the data in Table 2, these should be, respectively for MZ and DZ females:  $0.67 \pm 0.31$ , and  $0.54 \pm 0.27$ . These  $r$  values, although still greater, are appreciably closer to the corresponding MLEs in Table 4 than those published in Table 5. Their correct standard errors are also greater than the standard errors of the MLEs, but these correct standard errors are also closer to those of the MLEs than the values shown in Table 5.

## REFERENCES

1. Allen G, Hrubec Z (1979): Twin concordance a more general model. *Acta Genet Med Gemellol* 28:3-13.
2. Gottesman II, Shields J (1972): *Schizophrenia and genetics*. New York, Academic Press.
3. Hannah MC, Hooper JL, Mathews JD (1983): Twin concordance for a binary trait. 1. Statistical models illustrated with data on drinking status. *Acta Genet Med Gemellol* 32:127-137.
4. Hrubec Z (1973): The effect of diagnostic ascertainment in twins on the assessment of the genetic factor in disease etiology. *Am J Hum Genet* 25:15-28.
5. Hrubec Z, Allen G (1975): Letter to editor, Methods and interpretation of twin concordance data. *Am J Hum Genet* 27:808-809.
6. Smith C (1974): Concordance in twins: methods and interpretations. *Am J Hum Genet* 26:454-466.

**Correspondence:** Dr. Z. Hrubec, Radiation Epidemiology Branch, National Cancer Institute, Bethesda, MD 20205, USA.