



Acta Genet Med Gemellol 38:253-270 (1989)
©1989 by The Mendel Institute, Rome

Received 15 March 1989
Final 24 May 1989

The Estimation of Probandwise Concordance in Twins: The Effect of Unequal Ascertainment

K.S. Kendler^{1,2}, L.J. Eaves²

Departments of ¹Psychiatry and ²Human Genetics, Virginia Commonwealth University/Medical College of Virginia, Richmond, USA

Abstract. This report examines the impact of two major kinds of unequal ascertainment on the estimation of true probandwise concordance (C_{pbt}) in twin studies: 1) *concordance-dependent* – where the ascertainment rate differs in affected members of concordant vs discordant pairs, and 2) *non-independent* – where ascertainment rates differ in affected members of concordant pairs where the cotwin has vs has not been ascertained. Concordance-dependent ascertainment is easily modeled algebraically; non-independent ascertainment is more complex and we here propose a model based on survival analysis. Overall, concordance-dependent ascertainment produces greater bias in estimates of probandwise concordance than does non-independent ascertainment. The bias introduced by concordance-dependent ascertainment is greatest when C_{pbt} is low and/or when the ascertainment rate for twins in concordant pairs is low. The bias introduced by non-independent ascertainment is greatest when C_{pbt} is high and/or when the ascertainment probability for an affected twin in a concordant pair where the cotwin has already been ascertained approaches unity. The impact of concordance-dependent and non-independent ascertainment on estimates of heritability and common environment is examined. Correction terms to estimate C_{pbt} in the presence of concordance-dependent and/or non-independent ascertainment are presented.

Key words: Concordance, Ascertainment, Survival analysis, Twins

In studies of discontinuous traits in twins (and other pairs of relatives), a critical problem is the estimation of the true probability of illness in cotwins of affected

This work was supported in part by National Institutes of Health grants MH-40828 and MH-41953.

twins. In twin studies, the traditional method of estimation of this parameter, initially proposed by Weinberg [3,7], is probandwise concordance [1]. As pointed out by several authors [1-2,4-5,8-9], this method assumes that the probability of an affected twin becoming a proband is independent of both the disease status and the proband status of the cotwin. In other words, the proband method assumes an absence of two potential sources of biased ascertainment: *concordance-dependent* ascertainment – in which the probability of becoming a proband differs in affected members of concordant and discordant pairs – and *non-independent* (or, as termed by some authors, correlated) ascertainment – in which the probability of becoming a proband in a concordant pair depends on the ascertainment status of the affected cotwin.

There are a number of plausible scenarios where the assumption that ascertainment is independent of both the affection status and proband status of the cotwin can be questioned [1,4,9]. For example, non-independent ascertainment may occur in a study where ascertainment takes place at a medical clinic. If one member of a pair of concordant twins attends the clinic, the outcome of that twin's treatment experience may alter the probability that the affected cotwin will attend the clinic, thereby also becoming a proband. Concordance-dependent ascertainment might occur in at least two plausible ways. For example, if ascertainment is by informal contact with medical colleagues, concordant pairs might be preferentially referred and become probands. Or, a family may be able to cope without help with one affected twin, but when confronted with two cases of illness in a twin pair, help-seeking, which will lead to ascertainment, becomes necessary.

This paper has three associated goals: 1) to explore the biases introduced by concordance-dependent and/or non-independent ascertainment on estimates of probandwise concordance; 2) to explore the impact of these same biases on estimates of heritability and common environment; and 3) to provide "correction terms" to allow a more accurate estimate of true probandwise concordance when ascertainment is likely to be influenced by either the affection status or proband status of the cotwin.

MODEL

We assume a population of N twin pairs, C of whom are concordant for the trait in question, D of whom are discordant for the trait and U of whom are concordant for the absence of the trait. To avoid problems associated with age correction, we assume that the trait is fully expressed at all ages. True probandwise concordance rate (C_{pbt}) is therefore $2C/(2C + D)$. (See Table 1 for a listing of the main abbreviations used in this article).

In our discussion of concordance-dependent and non-independent ascertainment, we refer to *primary* ascertainment only. There has been debate about the value of including, in a full model of twin concordance, incomplete *secondary* ascertainment [1,5,6]. For simplicity, in this report, we assume that there is a secondary ascertainment procedure and it is complete (ie, all affected cotwins who are not

probands are detected). Under these circumstances, all twin pairs containing a proband belong to one of three groups: concordant pairs which are doubly ascertained (*DA*) (eg, where both members are probands), concordant pairs which are singly ascertained (*SA*) (eg, where one member is a proband, and the secondary ascertainment procedure determines that the cotwin is affected) and discordant pairs (where the secondary ascertainment procedure determines that the cotwin is unaffected). Thus, if C_2, C_1 , and D_1 equal, respectively, the number of twin pairs in the population that are concordant and doubly ascertained, concordant and singly ascertained and discordant and ascertained, it follows that 1) $C_2 = C \times DA$ and 2) $C_1 = C \times SA$ and 3) $D_1 = D \times m_d$.

Table 1

<i>C</i>	Number of twin pairs concordant for the disease or trait in the total twin population.
<i>D</i>	Number of twin pairs discordant for the disease or trait in the total twin population.
<i>U</i>	Number of twin pairs concordant for the absence of the disease or trait in the total twin population.
$N =$	$C + D + U$
C_{pbt}	True population probandwise concordance
C_{pbo}	Observed probandwise concordance rate calculated by Weinberg's proband method.
<i>DA</i>	The proportion of concordant pairs that are doubly ascertained.
<i>SA</i>	The proportion of concordant pairs that are singly ascertained.
C_2	The number of doubly ascertained concordant pairs in the twin population.
C_1	The number of singly ascertained concordant pairs in the twin population.
D_1	The number of ascertained discordant pairs in the twin population.
m_{c1}	The probability that an affected member of a concordant pair will be ascertained, given that the cotwin has not been ascertained,
m_{c2}	The probability that an affected member of a concordant pair will be ascertained, given that the cotwin has already been ascertained.
m_d	The probability that the affected member of a discordant pair will be ascertained.
m_c	The probability that an affected member of a concordant pair will be ascertained (used when ascertainment is independent, i.e., $m_c = m_{c1} = m_{c2}$)
$S_1 =$	$1 - m_{c1}$
$S_2 =$	$1 - m_{c2}$
$k =$	$DA + SA/2$
$n =$	m_d/m_c

In the absence of concordance-dependent and non-independent ascertainment, the observed probandwise concordance calculated by Weinberg's proband method (C_{pbo}) is a consistent estimator of C_{pbt} [3]:

$$(1) \quad C_{pbo} = \frac{2C_2 + C_1}{2C_2 + C_1 + D_1}$$

As in much of statistical genetics, the literature on ascertainment is confused by the use of multiple symbols for the same parameters. In most of the genetics

literature; ascertainment probability is symbolized by the Greek letter π [3]. However, in the twin literature, this same parameter has been symbolized by a lower case m [eg, 1,5-6]. In the interest of historical continuity, we also use m and define three specific ascertainment rates:

- m_{c1} = the probability that an affected member of a concordant pair will be ascertained given that the co-twin has not been ascertained;
- m_{c2} = the probability that an affected member of a concordant pair will be ascertained given that the co-twin has already been ascertained, and
- m_d = the probability that an affected member of a discordant pair will be ascertained.

As noted above, the estimation of the effect of unequal ascertainment rates on probandwise concordance is usefully subdivided into two problems: 1) concordance-dependent ascertainment and 2) non-independent ascertainment.

The impact on estimates of probandwise concordance of concordance-dependent ascertained is simply modeled. As shown by Holm [5 - Eq. (10)], if the ascertainment rate in both members of a concordant pair is equal and is expressed as m_c , then

$$(2) \quad C_{pbo} = \frac{2C}{2C + D(m_d/m_c)}$$

Estimating the impact of non-independent ascertainment on estimates of probandwise concordance is more difficult. If ascertainment in concordant pairs is assumed to occur in only two steps, then the process can be simply modeled. In the first step, twin 1 is either ascertained (probability of m_{c1}) or not ascertained (probability of $1 - m_{c1}$). If twin 1 is ascertained, then in the second step, the probability that twin 2 is ascertained is m_{c2} and the probability that twin 2 is not ascertained is $1 - m_{c2}$. If twin 1 is not ascertained, then in the second step, the probabilities that twin 2 is and is not ascertained are, simply, m_{c1} and $1 - m_{c1}$, respectively. Under this model, the distribution of concordant pairs into ascertainment groups is as follows:

$$(3) \quad DA = m_{c2} \times m_{c1}$$

$$(4) \quad SA = m_{c1} \times (2 - m_{c2} - m_{c1})$$

$$(5) \quad UA = (1 - m_{c1})^2$$

If we assume that $m_d = m_{c1}$, then it can be shown that

$$(6) \quad C_{pbo} = \frac{(2 + m_{c2} - m_{c1})C}{(2 + m_{c2} - m_{c1})C + D}$$

The above model of non-independent ascertainment is, however, unsatisfactory because of the unrealistic assumption that ascertainment is a simple two step process. For example, in the event that twin 1 is not ascertained in the first step, but twin 2 is ascertained in the second step, the probability that the first twin would “truly” be ascertained is no longer m_{c1} . The problem with this model can be most clearly seen when m_{c2} equals unity. In the two step model, the probability that a concordant pair would be doubly ascertained would then equal m_{c1} . However, since either member of a twin pair could be ascertained first with a probability of m_{c1} , and if one member was ascertained, the second member would automatically be ascertained, then the true probability that a pair would be doubly ascertained when m_{c2} equals unity should be $[1 - (1 - m_{c1})^2]$.

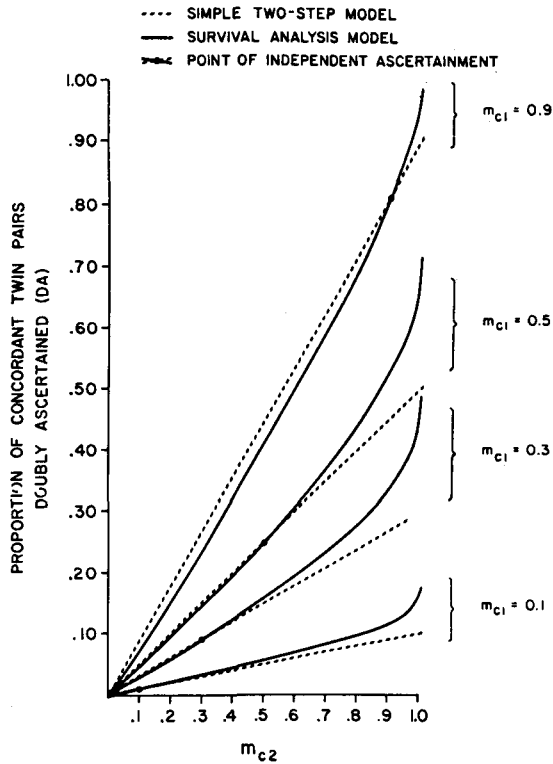


Fig. 1. A comparison of two models for non-independent ascertainment. The proportion of concordant twin pairs in the population which are doubly ascertained as a function of m_{c1} (the probability that an affected member of a concordant pair will be ascertained, given that the cotwin has not been ascertained) and m_{c2} (the probability that an affected member of a concordant pair will be ascertained, given that the cotwin has been ascertained) for the simple “two-step” and the survival based models of non-independent ascertainment. These models are outlined in the text and Appendix. The point of independent ascertainment (where $m_{c1} = m_{c2}$), where the two curves meet, is noted by a black dot.

Therefore, in the Appendix, we present a more realistic model for the problem of non-independence of ascertainment based on survival analysis. It is of interest to compare the results of the two models of non-independent ascertainment: the simple "two-step" and the survival analysis based models. As seen in Fig. 1, when values of m_{c2} and m_{c1} are relatively similar, the two models are quite close in their predictions of the proportion of twin pairs that are doubly ascertained. When m_{c2} is much less than m_{c1} , the survival analysis model predicts that a *lower* proportion of concordant pairs will be doubly ascertained than does the two-step model, but the differences are relatively modest. When m_{c2} is much greater than m_{c1} , and particularly when m_{c2} approaches unity, the survival analysis model predicts that a *higher* proportion of twin pairs will be doubly ascertained than does the two-step model, and here the differences can be quite substantial. The survival model correctly predicts that as m_{c2} goes to unity, the proportion of concordant twin pairs that will be doubly ascertained goes to $[1 - (1 - m_{c1})^2]$.

Therefore, to demonstrate the effect of differences in ascertainment rates on estimates of probandwise concordance, we use the expectations from the survival model presented in the Appendix. Eq. (1) can be re-expressed as

$$(7) \quad C_{pbo} = \frac{2C(DA + SA/2)}{2C(DA + SA/2) + Dm_d}$$

We define, for convenience, a variable k which equals $DA + SA/2$. Assuming, per the Appendix, that S_1 and S_2 equal $(1 - m_{c1})$ and $(1 - m_{c2})$, respectively, from Eqs. (A12) and (A13) it can be shown that

$$(8) \quad k = DA + SA/2 = \frac{\ln S_1(2 - S_1^2 - S_2) - \ln S_2(1 - S_1^2)}{2 \ln S_1 - \ln S_2}$$

Values for k are then substituted back into Eq. (7) to derive the formula used for obtaining C_{pbo} .

RESULTS

Concordance-Dependent Ascertainment

The impact of concordance-dependent ascertainment is assessed by assuming independent ascertainment in concordant twin pairs ($m_{c2} = m_{c1} = m_c$) and varying the ascertainment rate in affected members of discordant pairs (m_d). Results are seen in Fig. 2I assuming a high true probandwise concordance rate (0.50) and in Fig. 2II assuming a moderate true probandwise concordance rate (0.10). Four major conclusions can be drawn from these Figures:

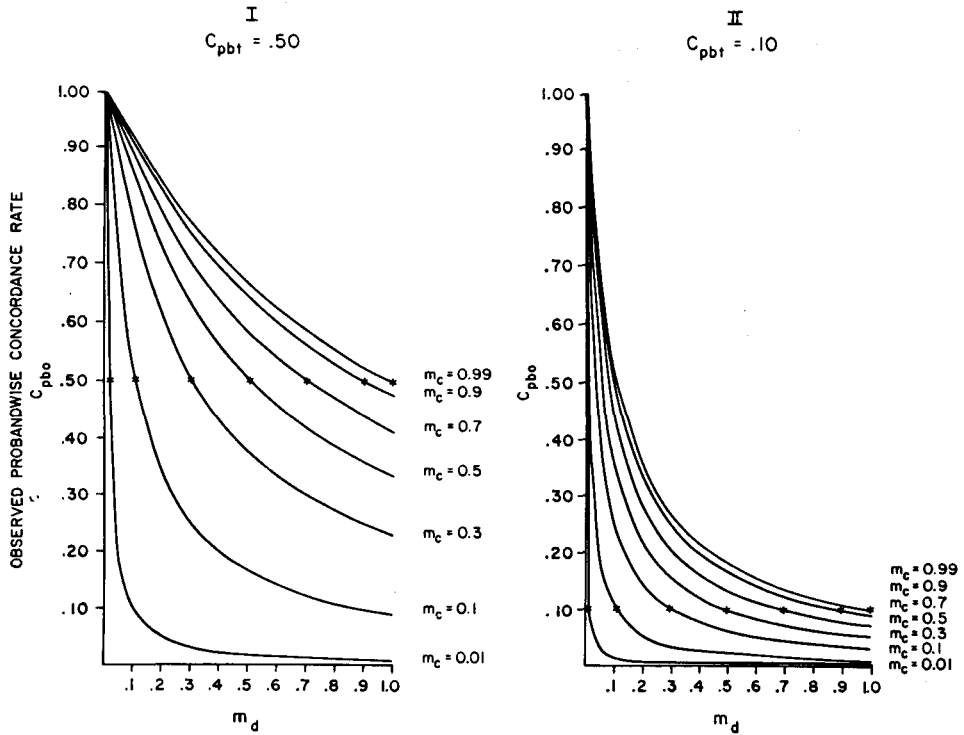


Fig. 2. The effect of concordance dependent ascertainment. The observed probandwise concordance rate (C_{pbo}) in a twin population as a function of values of m_d (the probability that the affected member of a discordant twin pair will be ascertained) and m_c (the probability that an affected member of a concordant twin pair will be ascertained). The left-hand part of the Figure (part I) depicts this relationship when the true population probandwise concordance rate (C_{pbt}) is high (50%), while the right-hand part of the Figure (part II) depicts this relationship when C_{pbt} is low (10%). The point of independent ascertainment (where $m_d = m_c$), where the two curves meet, is noted by an asterisk.

- 1) The observed probandwise concordance rate varies inversely with values for m_d . If m_d is lower than m_c , C_{pbo} overestimates C_{pbt} . If m_d is higher than m_c , C_{pbo} underestimates C_{pbt} .
- 2) The slope of the function is exponential-like and relatively steep. That is, modest changes in m_d can produce substantial changes in C_{pbo} .
- 3) The smaller the value m_c , the steeper is the function relating C_{pbo} to m_d . That is, the same absolute change in values for m_d will produce a much greater change in C_{pbo} if m_c is small (eg, < 0.20) than if it is large (eg, > 0.50).
- 4) The smaller the value of C_{pbt} , the steeper is the function relating C_{pbo} to m_d .

Non-Independent Ascertainment

Non-independent ascertainment is modeled by assuming an equal ascertainment rate in members of concordant pairs where the cotwin is unascertained (m_{c1}) and in affected members of discordant pairs (m_d) and varying values for the rate of ascertainment in members of concordant pairs where the cotwin *has been* ascertained (m_{c2}) (Fig. 3). The following major conclusions can be drawn from the Figure:

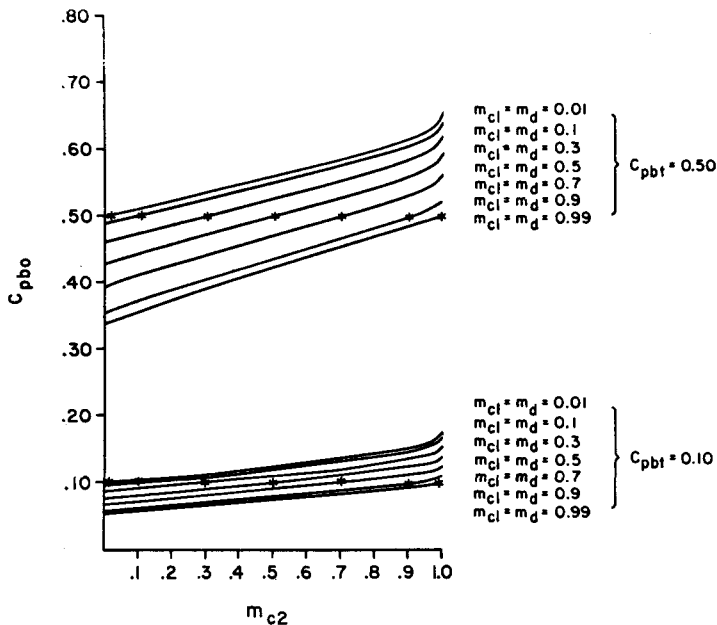


Fig. 3. The effect of non-independent ascertainment. The observed probandwise concordance rate (C_{pbo}) in a twin population as predicted by the survival analysis model as a function of values of m_{c2} (the probability that an affected member of a concordant twin pair will be ascertained, given that the cotwin has already been ascertained) and $m_{c1} = m_d$ (the probability of ascertainment for the affected member of a concordant twin pair, given that the cotwin is not ascertained, and the affected member of a discordant twin pair). The upper part of the Figure represents this relationship when the true population probandwise concordance rate (C_{pbt}) is high (50%), while the lower part of the Figure depicts this relationship when C_{pbt} is low (10%). The point of independent ascertainment (where $m_{c2} = m_{c1}$) is noted by an asterisk.

- 1) The observed probandwise concordance rate varies directly with values of m_{c2} . If m_{c2} is higher than $m_{c1} = m_d$, then C_{pbo} overestimates C_{pbt} . If m_{c2} is lower than $m_{c1} = m_d$, then C_{pbo} underestimates C_{pbt} .
- 2) The slope of the function is nearly linear and relatively flat. That is, modest changes in m_{c2} produce only small changes in C_{pbo} . The major non-linearity of the function occurs as m_{c2} approaches unity.
- 3) The smaller the value of C_{pbt} , the flatter is the function relating C_{pbo} to m_{c2} .

Concordance-Dependent and Non-Independent Ascertainment

Fig. 4 examines the impact of *both* concordant-dependent and non-independent ascertainment on the observed probandwise concordance rate. This is done by allowing both m_{c2} and m_d to vary for a given value of m_{c1} . As would be expected from the previous Figures:

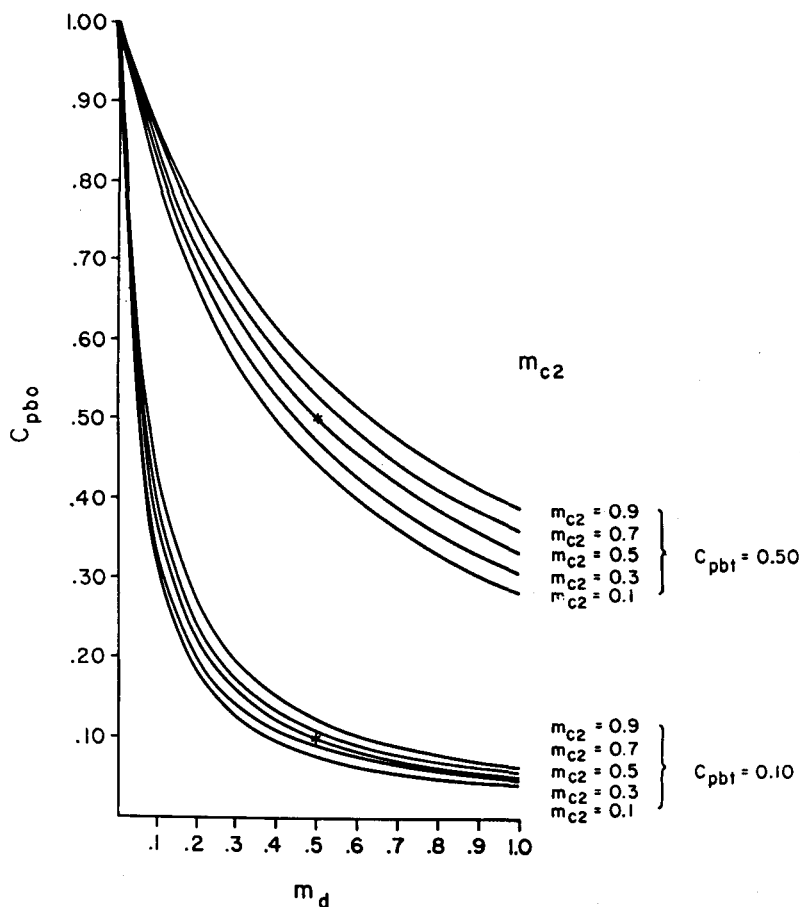


Fig. 4. The effect of both concordance-dependent and non-independent ascertainment. The observed probandwise concordance rate (C_{pbo}) in a twin population as a function of values of m_{c2} (the probability that an affected member of a concordant twin pair will be ascertained, given that the cotwin has already been ascertained) and m_d (the probability of ascertainment for the affected member of a discordant twin pair), given that m_{c1} (the probability of ascertainment for an affected member of a concordant pair, given that the cotwin is not ascertained) is set at 0.50. The upper part of the Figure represents this relationship when the true population probandwise concordance rate (C_{pbt}) is high (50%), while the lower part of the Figure depicts this relationship when C_{pbt} is low (10%). The points of independent ascertainment (where $m_{c2} = m_{c1} = m_d = 0.5$) are noted by an asterisk.

- 1) Values for C_{pbo} are, in general, more sensitive to changes in m_d than to changes in m_{c2} .
- 2) The relative change in C_{pbo} resulting from changes in m_d and m_{c2} is a function of the value of C_{pbt} . That is, when C_{pbt} is small, C_{pbo} is *much* more sensitive to changes in m_d than to changes m_{c2} . When C_{pbt} is relatively large, C_{pbo} is still somewhat more sensitive to changes in m_d than to changes in m_{c2} , but the difference is much less.

Effect of Unequal Ascertainment on Estimates of Heritability

In Tables 2 and 3, we address the following question: if concordant-dependent and non-independent ascertainment are present, but are assumed to be absent, what effect will they have on estimates of heritability of liability (h^2) and common environment (c^2)? These calculations, which use Smith's formula [9] to calculate the correlation in liability to illness (r) in twins, assume a normally distributed liability to illness with a single threshold of manifestation. A rough estimate of heritability of liability (h^2), common environment (c^2) and random environment (e^2) can be obtained from the formulas: $h^2 = 2(r_{MZ} - r_{DZ})$, $c^2 = r_{DZ} - r_{MZ}$ and $e^2 = 1 - (h^2 + c^2)$. We assume that m_{c1} equals 0.5 and that rates of ascertainment are equal in MZ and DZ twins.

Table 2 - The effect of concordance-dependent ascertainment on estimates of heritability and common environment

K_p (%)	m_d	DZ Twins		MZ Twins		h^2	c^2	e^2
		Con(%)	r	Con(%)	r			
0.01	0.1	4.30	0.584	44.4	0.924	0.680	0.244	0.076
	0.3	1.50	0.445	21.1	0.817	0.744	0.073	0.183
	0.5	0.90	0.386	13.8	0.753	0.734	0.019	0.247
	0.7	0.65	0.349	10.3	0.709	0.720	-0.011	0.291
	0.9	0.50	0.322	8.2	0.675	0.706	-0.031	0.325
1.00	0.1	35.70	0.781	83.3	0.996	0.430	0.566	0.004
	0.3	15.60	0.549	62.5	0.931	0.764	0.167	0.057
	0.5	10.00	0.434	50.0	0.873	0.878	-0.010	0.132
	0.7	7.40	0.363	41.7	0.824	0.922	-0.098	0.143
	0.9	5.80	0.309	35.7	0.781	0.944	-0.163	0.219
10.00	0.1	57.90	0.807	76.2	0.940	0.266	0.674	0.060
	0.3	31.50	0.484	51.6	0.747	0.526	0.221	0.253
	0.5	21.60	0.300	39.0	0.597	0.594	0.003	0.403
	0.7	16.40	0.181	31.4	0.482	0.602	-0.120	0.518
	0.9	13.30	0.099	26.2	0.391	0.584	-0.193	0.609

Assuming $m_{c1} = m_{c2} = 0.50$
 K_p = population risk; r = correlation in liability; Con = concordance, probandwise.

Table 2 examines the impact of concordance-dependent ascertainment on estimates of heritability and common environment for a rare (K_p , or population risk,

equal to 0.01%), uncommon ($K_p = 1.0\%$) and common ($K_p = 10.0\%$) trait with a relatively high heritability and little or no common environment. m_{c1} and m_{c2} are both fixed at equal to 0.5 and m_d is varied. Consistently, across all population rates, estimates of c^2 are inversely and estimates of e^2 are directly related to m_d . That is, if rates of ascertainment are lower in discordant than in concordant pairs, concordance in both twin types and c^2 will be overestimated, but e^2 will be underestimated. However, if rates of ascertainment are higher in discordant than in concordant pairs, concordance in both twin types and c^2 will be underestimated, but e^2 will be overestimated. The effect of concordance-dependent ascertainment on estimates of h^2 are more complex. When the trait is either uncommon or common, estimates of h^2 have a strong positive correlation with values of m_d . However, when the trait is rare, estimates of h^2 appear to have a weak negative correlation with m_d .

Table 3 - The effect of non-independent ascertainment on estimates of heritability and common environment

K_p (%)	m_d	DZ Twins		MZ Twins		h^2	c^2	e^2
		Con(%)	r	Con(%)	r			
0.01	0.1	0.72	0.361	11.3	0.723	0.724	-0.001	0.277
	0.3	0.81	0.373	12.5	0.738	0.730	0.008	0.278
	0.5	0.90	0.386	13.8	0.753	0.734	0.019	0.247
	0.7	1.01	0.399	15.2	0.767	0.736	0.031	0.233
	0.9	1.14	0.413	16.9	0.784	0.742	0.084	0.174
1.00	0.1	8.10	0.385	44.3	0.841	0.912	-0.071	0.159
	0.3	9.00	0.409	47.2	0.856	0.894	-0.038	0.144
	0.5	10.00	0.434	50.0	0.873	0.878	-0.010	0.132
	0.7	11.10	0.461	52.8	0.888	0.854	0.068	0.078
	0.9	12.40	0.489	56.0	0.903	0.828	0.075	0.097
10.00	0.1	18.00	0.220	33.7	0.520	0.600	-0.100	0.500
	0.3	19.80	0.261	36.4	0.560	0.598	-0.038	0.440
	0.5	21.60	0.300	39.0	0.597	0.594	0.003	0.403
	0.7	23.60	0.341	41.7	0.634	0.586	0.048	0.366
	0.9	26.00	0.388	44.9	0.672	0.568	0.208	0.224

Assuming $m_{c1} = m_d = 0.50$

K_p = population risk; r = correlation in liability; Con = concordance, probandwise.

Table 3 examines the impact of non-independent ascertainment on estimates of heritability and common environment for the same traits examined in Table 2. m_{c1} and m_d are both fixed at equal to 0.5 and m_{c2} is varied. As expected, the impact of non-independent ascertainment on both the concordance rates and the estimates of heritability and common environment are considerably less marked than those of concordance-dependent ascertainment. Consistently, across all population rates, estimates of c^2 are directly and estimates of e^2 are inversely related to m_{c2} . When the trait is either uncommon or common, estimates of h^2 have a modest negative correlation with values of m_{c2} . However, when the trait is rare, estimates of h^2 appear to have a weak positive correlation with m_{c2} .

The impact of concordance-dependent and non-independent ascertainment on estimates of heritability and common environment has also been examined for traits with modest heritabilities and a substantial common environmental component (results available on request). The impact on estimates of c^2 and e^2 are consistent with those noted above, while the effect on estimates of h^2 are more variable.

Correction Terms

Up until now, we have been interested in estimating the impact of concordance-dependent and non-independent ascertainment on the observed concordance rate calculated using Weinberg’s method assuming that we knew the true probandwise concordance rate. In research, the opposite problem is usually confronted. That is, one begins with the observed concordance rate and, hopefully, some idea of the magnitude of the concordance-dependent and/or non-independent ascertainment and wants to estimate the true population probandwise concordance. Therefore, we here present the following “correction terms”. For concordance-dependent ascertainment, assuming that $m_{c2} = m_{c1} = m_c$, it can be shown that

$$(9) \quad C_{pbt} = \frac{(m_d/m_c)C_{pbo}}{C_{pbo}[(m_d/m_c) - 1] + 1}$$

Note that for estimating the true from the observed probandwise concordance given concordance-dependent ascertainment, it is only the ratio of m_d to m_c that is important, not their absolute values. Thus, if we define a variable $n = m_d/m_c$, Eq. (9) reduces to

$$(10) \quad C_{pbt} = \frac{nC_{pbo}}{C_{pbo}(n - 1) + 1}$$

For non-independent ascertainment, we first present a correction term based on the “two-step” model of ascertainment in concordant pairs outlined above. This simple formula will be reasonably accurate as long as m_{c1} and m_{c2} are not too different and $m_{c2} < 0.8$:

$$(11) \quad C_{pbt} = \frac{2C_{pbo}}{2 + (m_{c2} - m_{c1})(1 - C_{pbo})}$$

Again note that using Eq. (11) only the difference in m_{c1} and m_{c2} , not their absolute values, is used.

A more definitive correction term for non-independent ascertainment is based on the survival model presented in the Appendix. Recalling parameter k from Eq. (8) (which is derived from the values of m_{c1} and m_{c2}), and assuming that $m_d = m_{c1}$, C_{pbt} can be shown to equal

$$(12) \quad C_{pbt} = \frac{C_{pbo}}{C_{pbo} + [k(1 - C_{pbo})]/m_{c1}}$$

For both concordance-dependent and non-independent ascertainment, Eq. (12) needs to be only slightly modified to produce the following expression for C_{pbt}

$$(13) \quad C_{pbt} = \frac{C_{pbo}}{C_{pbo} + [k(1 - C_{pbo})]/m_d}$$

COMMENT

The goal of this report was to examine the impact of unequal ascertainment rates on estimates of population probandwise concordance rates in twin pairs. The problem was divided into two parts: concordance-dependent ascertainment (ie, different ascertainment rates in affected members of discordant and concordant pairs), and non-independent ascertainment within concordant pairs. As has been shown by others, the impact of concordance-dependent ascertainment is easily modeled. The impact of non-independent ascertainment is more difficult and we present here two models: a simple approximate algebraic model which assumes “two steps” in the ascertainment of concordant pairs, and a more complex and accurate survival analysis based model.

The major results can be easily summarized. First, the observed probandwise concordance rate (C_{pbo}) is, in general, more sensitive to the effects of concordant-dependent ascertainment than to the effect of non-independent ascertainment. Second, the impact on C_{pbo} of the two kinds of differential ascertainment varies in opposite ways as a function of the true population probandwise concordance (C_{pbt}). As C_{pbt} becomes smaller, the impact of concordance-dependent ascertainment increases while the impact of non-independent ascertainment decreases.

This logic of these results can be appreciated if we combine the effect of concordance-dependent ascertainment (Eq. (2)) and the simple “two-step” model of non-independent ascertainment (Eq. (6)) into a single formula:

$$(14) \quad C_{pbo} = \frac{(2 + m_{c2} - m_{c1})C}{(2 + m_{c2} - m_{c1})C + (m_d/m_{c1})D}$$

While Eq. (14) will not be accurate when m_{c2} and m_{c1} differ substantially or when m_{c2} approaches unity, it is nonetheless helpful in clarifying the nature of the results obtained. Two major points can be deduced from Eq. (14). First, the impact of concordance-dependent ascertainment is a function of the *ratio* of m_d to m_{c1} while the impact of non-independent ascertainment is a function of the *difference* between m_{c2} and m_{c1} . This explains why, when m_{c1} and m_{c2} are equal, C_{pbo} is approximately an exponential function of m_d (Fig. 2), yet when m_{c1} and m_d are equal, C_{pbo} is nearly a linear function of m_{c2} (Fig. 3). Eq. (14) also clarifies why the same absolute change in m_d produces a much larger change in C_{pbo} when m_{c1} is small than when m_{c1} is large; yet the change produced in C_{pbo} by a given change in m_{c2} is approximately independent of the value of m_{c1} .

The second major point that can be deduced from Eq. (14) is that the impact of concordant-dependent ascertainment is a function of D , the number of discordant pairs in the population, while the impact of non-independent ascertainment is a function of C , the number of concordant pairs in the population. This explains why changes in the true population probandwise concordance (C_{pbt}) have opposite effects on the impact of concordance-dependent and non-independent ascertainment on C_{pbo} . That is, when C_{pbt} is low (ie, $D \gg C$), C_{pbo} will be more affected by changes in concordance-dependent ascertainment than by changes in non-independent ascertainment. This makes sense because changes in the independence of ascertainment will only affect concordant pairs. By contrast, when C_{pbt} is high (ie, $C \geq D$), the relative impact on C_{pbo} of changes in concordant-dependent ascertainment will be less while the relative impact of non-independent ascertainment will increase.

Next, we illustrated the possible impact of concordance-dependent and non-independent ascertainment on estimates of heritability and common environment. Because of the number of parameters involved, it was not possible to thoroughly explore this problem. Nonetheless, the example presented demonstrates that ignoring concordance-dependent and to a lesser extent non-independent ascertainment when it is present can lead to serious biases in estimates of genetic parameters.

Finally, we presented "correction terms" for the use of investigators who suspect the presence of concordance-dependent and/or non-independent ascertainment. These formulas will allow them to calculate probandwise concordance by the standard Weinberg's formula (Eq. (1)) and then correct this "observed" concordance for the effects of unequal ascertainment rates to obtain an estimate of the true population probandwise concordance.

Our models and results for non-independent ascertainment proposed differ considerably from that of Allen and Hrubec [1], who, on the basis of a "two-step, sequential" model, conclude that when secondary ascertainment is complete, non-independent ascertainment introduces no bias in estimates of the true probandwise concordance rate. Space prohibits a detailed critique of their model. However, the key difference with our "two-step" models appears to be in how singly ascertained concordant pairs are modeled (as, using our terminology, both two-step models assume that doubly ascertained are proportional to $m_{c1} \times m_{c2}$). In any event, as we outlined above, we would conclude that any "two-stage" model is less realistic than the survival analysis based model that we propose and use in this report. In addition, our results are consistent with those of Rice et al [8], who, without formalizing a model, suggest that non-independent ascertainment will change estimates of probandwise rates in twins and their relatives, but under most circumstances, these changes will be quite small.

The models and results presented in this report should be interpreted in the context of several potentially important limitations. First, no attempt was made to deal with the probabilistic nature of the parameters used and estimated. For example, the correction terms presented in Eqs. (9-13) are not "true numbers" but rather are estimates with confidence intervals attached to them. In fact, these confidence intervals may, under certain circumstances, be quite large because of

the number of individual parameters in the formulas. Second, while we examined the impact of the affection and proband status of the cotwin on the probability of ascertainment, we did not explore the impact of several other potentially realistic factors such as severity of illness and/or age at onset. Since such factors might be correlated in twins and/or may be indices of disease liability, their impact on ascertainment, and hence on estimation of concordance rates and heritability, could be relatively complex. Third, our examination of the effect of concordance-dependent and non-independent ascertainment on estimates of heritability assumed that no difference existed in ascertainment rates across zygosity groups. For example, if there is concordance-dependent ascertainment, it is perfectly plausible that the ascertainment rate might be higher for concordant MZ twins than for concordant DZ twins.

Finally, the model for non-independent ascertainment based on survival analysis itself made three significant simplifying assumptions. First, it assumed an ascertainment period which is fixed and identical for both affected twins. This is a reasonable assumption when, for a population of pairs of twins both of whom are already affected, an investigator defines an arbitrary ascertainment period. This may be an unrealistic assumption in other situations. For example, if the ascertainment period begins at disease onset, the model assumes that twins become affected at the same time.

A second assumption of this model is that, conditional on the ascertainment status of the cotwin, the probability of an unascertained affected twin becoming ascertained is constant over time. This is probably a valid assumption if the factors leading an affected individual to ascertainment are largely random. However, if ascertainment is tied to medical treatment this assumption may not be valid. For example, the probability of a previously untreated individual seeking care for a disease is unlikely to be constant over the course of the illness; rather it will probably be high at the onset of symptoms and decline rapidly thereafter.

The third important assumption of the survival analysis model is that ascertainment is a "one-hit" process. This may be realistic for a wide range of situations. Using gamma functions, the model could be expanded to consider ascertainment as a "multiple hit" process.

While the models presented in this report, and the results obtained therefrom, are certainly not definitive, they do represent tractable and potentially useful ways to model the impact of unequal ascertainment rates on the results of twin studies.

REFERENCES

1. Allen G, Harvald B, Shields J (1967): Measures of twin concordance. *Acta Genet* 17:475-481.
2. Allen G, Hrubec Z (1979): Twin concordance: A more general model. *Acta Genet Med Gemellol* 28:3-13.
3. Crow JF (1965): Problems of ascertainment in the analysis of family data. In Neel JV, Shaw MW, Schull WJ (eds): *Genetics and the Epidemiology of Chronic Diseases*. Washington, DC: US Dept of Health, Education & Welfare, pp 23-44.

4. Greenberg DA (1986): The effect of proband designation on segregation analysis. *Am J Hum Genet* 39:329-339.
5. Holm NV (1983): A note on ascertainment probability in the Allen/Hrubec twin model. *Acta Genet Med Gemellol* 32:37-47.
6. 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.
7. Nichols FW (1982): Simple segregation analysis: A review of its history and terminology. *J Hered* 73:444-450.
8. Rice J, Gottesman II, Suarez BK, O'Rourke DH, Reich T (1982): Ascertainment bias for non-twin relatives in twin proband studies. *Hum Hered* 32:202-207.
9. Smith C (1974) Concordance in twins: Methods and interpretation. *Am J Hum Genet* 26:454-466.

Correspondence: Professor K.S. Kendler, Box 710 MCV Station, Richmond, VA 23298-0710, USA.

Appendix

The model begins with two twins who, at time 0, are both affected and both unascertained. Ascertainment for both twins can occur from time 0 to time T . We define the following two ascertainment rates:

m_{c1} = the probability of ascertainment of a sibling from time 0 to time T given that the cotwin remains unascertained during the entire time period.

m_{c2} = the probability of ascertainment of a sibling from time 0 to T given that the cotwin was ascertained at time 0.

That is, the model assumes that the probability of ascertainment is constant over the time period 0 to T , *except* that the ascertainment rate changes when the cotwin is ascertained.

Using the terminology of survival analysis, the model assumes that ascertainment is equivalent to failure. Therefore, we define two analogous "survival" terms:
 $S_1 = 1 - m_{c1}$ = the probability that a twin "survives" unascertained from time zero to time T given that the cotwin remains unascertained during the entire time period.

$S_2 = 1 - m_{c2}$ = the probability that a twin "survives" unascertained from time 0 to time T given that the cotwin was ascertained at time 0.

Now, we define λ to be equal to the ascertainment rate for an affected twin given that the affected cotwin is unascertained and $\lambda + d$ to equal the ascertainment rate for an affected twin when the cotwin has been ascertained. d may be positive or negative and will equal 0 when ascertainment is independent.

Given that the ascertainment rate is constant, we then have

(A1)
$$S_1 = e^{-\lambda T}$$

(A2)
$$S_2 = e^{-(\lambda+d)T}$$

As noted above, there are three possible ascertainment outcomes for a concordant pair of twins: double ascertainment (DA), single ascertainment (SA) and unascertained (UA).

The probability of UA , which is independent of m_{c2} and S_2 , can be simply expressed:

(A3)
$$UA = S_1^2 = e^{-2\lambda T}$$

To derive the expected value of SA , we begin by expressing the probability that *either* the first *or* second twin will be ascertained at time u *and* that the cotwin will survive until time u without being ascertained:

(A4)
$$2\lambda e^{-\lambda u} e^{-\lambda u} = 2\lambda e^{-2\lambda u}$$

After u , the ascertainment rate for the second twin becomes $\lambda + d$, so the probability that the second twin will survive until time T (eg, the period $T - u$) unascertained is

(A5)
$$e^{-(\lambda+d)(T-u)}$$

Therefore, the probability that one twin will be ascertained at time u and the cotwin will survive till time T without being ascertained equals

$$(A6) \quad 2\lambda e^{-2\lambda u} e^{-(\lambda+d)(T-u)} = e^{-(\lambda+d)T} 2\lambda e^{-(\lambda-d)u}$$

To express the probability that one twin will be ascertained at any time u from 0 to T and that the cotwin will remain unascertained till time T , we need to integrate the expressions in Eq. (A6) for values of u from 0 to T . Thus

$$(A7) \quad SA = 2\lambda e^{-(\lambda+d)T} \int_0^T e^{-(\lambda-d)u} du$$

which can be shown to be equal to

$$(A8) \quad SA = \frac{2\lambda}{\lambda-d} e^{-(\lambda+d)T} \left[1 - e^{-(\lambda-d)T} \right]$$

Now, it is helpful to convert Eq. (A8) into a more readily usable form by scaling time so that $T = 1$. Then

$$(A9) \quad \lambda = -\ln S_1$$

$$(A10) \quad d = \ln S_1 - \ln S_2$$

and

$$(A11) \quad S_1^2/S_2 = e^{-(\lambda-d)}$$

Substituting Eqs. (A9-A11) back into Eq. (A8) and simplifying produces

$$(A12) \quad SA = \frac{2(S_1^2 - S_2) \ln S_1}{\ln S_2 - 2 \ln S_1}$$

From the function $DA = 1 - SA - UA$, the value of DA can, from Eqs. (A3) and (A12), be shown to be

$$(A13) \quad DA = \frac{(1 - S_1^2) \ln S_2 - 2(1 - S_2) \ln S_1}{\ln S_2 - 2 \ln S_1}$$

Eqs. (A12) and (A13) become formally undefined under 3 conditions: 1) $S_1 = 0$, 2) $S_2 = 0$, and 3) $\ln S_2 = 2 \ln S_1$ (or $S_1^2 = S_2$). We see no simple way to apply our model to condition 1, where the probability of ascertainment of the first twin is unity. Condition 2, by contrast, is readily interpretable. Under these circumstances (eg, the second twin is automatically ascertained given that the first twin is ascertained), no twin pairs will be singly ascertained ($SA = 0$), and DA will simply equal $1 - UA$. Condition 3, where $\lambda = d$, makes sense in the limit as the value of λ approaches that of d .

The analytic solutions presented by this model have been tested by large-sample stochastic simulations and close agreement has been found (results available on request).