Acta Genet Med Gemellol 34: 125-140 (1985)
© 1985 by The Mendel Institute, Rome

Received 24 June 1985
Final 26 July 1985

Differential Enrollment in Twin Registries: Its Effect on Prevalence and Concordance Rates and Estimates of Genetic Parameters

K.S. Kendler[1], N.V. Holm[2]

[1] *Departments of Psychiatry and Human Genetics, Medical College of Virginia, Virginia Commonwealth University, Richmond;* [2] *University Institute of Medical Genetics and the Danish Twin Register, Odense University Hospital, Denmark*

Abstract. In the NAS-NRC Registry, all major diseases are more common in DZ than in MZ twins. Furthermore, concordance rates for most disorders are lower in the registry than would be expected. In this article we propose a general model which seeks to explain these phenomena. The model explores the impact of traits which increase or decrease the probability of enrollment of individuals given that the registry, like the NAS-NRC, includes only pairs where both members are enrolled. If the trait decreases the probability of selection into the registry, both the prevalence of and concordance for the trait in the registry will be lower than that found in the population. A trait which increases the probability of selection has the opposite effects. However, the magnitude of these effects are a function of the population concordance. If population concordance differs in MZ and DZ twins, the effect of differential enrollment will not be the same for the two zygosity groups. The article examines the impact of differential enrollment on estimates of heritability and common environment and explores ways in which estimates of prevalence and concordance rates can be obtained which are free of the bias introduced by selection.

Key words: Twin registers, Selection, Prevalence, Concordance, Heritability

INTRODUCTION

During an investigation of schizophrenia in the National Academy of Sciences-National Research Council Twin Registry (NAS-NRC) [8], two potentially unusual features of the distribution of this disorder in the registry were found. First, the disorder was more common in dizygotic (DZ) than in monozygotic (MZ) twins. Further investigation revealed that this pattern was not restricted to schizophrenia, but occurred in virtually all other major disorders in the registry including diabetes, hypertension, ischemic heart

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

disease, chronic obstructive pulmonary disease, peptic ulcer disease and neurosis. Furthermore, the mortality rate for MZ twins in the Registry was significantly lower than for DZ twins [5]. Apparently, MZ twins in the NAS-NRC registry were, on average, healthier than DZ twins.

The second unusual feature of schizophrenia in the registry was that concordance rates for the disorder in both MZ and DZ twins were lower than those found in almost all previous studies[7]. There may be many reasons why rates of schizophrenia in MZ twins might be less than in DZ twins and why concordance rates for this disorder in the registry might be low. However, in this report, we suggest that these two findigns can be parsimoniously explained by a single feature of the method of construction of the registry: both members of a twin pair had to pass a health screening for the pair to be included.

In this paper, we first outline an algebric model of this hypothesis, explore some of its implications and then address the question of how the bias suggested by the model can be corrected. This report represents a considerable expansion of an initial brief examination of this issue previously presented by one of us [8]. A key to the main abbreviations used in the text is shown in Table 1.

MODEL

Effect of Selection on Concordance Rates

Twins were identified for the NAS-NRC registry by a two-stage procedure. First, some 54,000 twin births in the years 1917-1927 were identified from birth certificates in 39 of the continental United States [4]. Second, these twins were screened through the Master Index of the United States Veteran Administration to identify the 15,924 twin pairs where *both* members of the pair had served in the US Armed Forces. Induction into the US Armed Forces involves, in addition to motivational and social factors, a health screening. For example, during the years in which most twins in the Registry were inducted, 14.0% of all inductees were rejected for psychiatric reasons alone [6]. The efficacy of this selection is demonstrated by the finding that veteran populations in the US have significantly reduced mortality compared to the general population for nearly all major forms of disease [11]. Therefore, a disease which might be manifest in some form at the age of induction ought to decrease the probability that an individual with that disease would be permitted into the Armed Services. By contrast, it is possible to imagine certain motivational features (eg, patriotism) which might increase the probability of an individual being inducted into the Armed Services.

So, we begin by assuming a population of N twin pairs which correspond to the population from which the NAS-NRC registry was formed. We consider a trait X which will influence the probability that an inductee will be accepted into the Armed Services. Therefore, the population of N twin pairs is divided into those concordant for trait X (C), discordant for trait X (D), and concordant for the absence of trait X (U) (Table 2). We then construct a preliminary twin panel from this population in which the three kinds of twin pairs are divisible into those were both members are enrolled in the panel (C_2, D_2, U_2), one member is enrolled in the panel (C_1, D_1, U_1) and neither member is enrolled in the panel (C_0, D_0, U_0). The D_1 class of twins must be subdivided into those where the twin with trait X is the enrolled twin (D_{1a}) and the twin without the trait is the enrolled twin (D_{1b}).

The probability than an individual twin with trait X will be enrolled in the panel is e_1, while the probability of enrollment in the panel for an individual without X is e_2. The

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

Table 1. Key to Main Abbreviations Used in Text

C	Number of twin pairs concordant for the disease or trait in the total twin population.
D	Number of twin pairs discordant for the disease or trait in the total twin population.
U	Number of twin pairs concordant for the absence of the disease or trait in the total twin population.
P_p	Disease or trait prevalence rate among twin individuals in population.
P_r	Disease or trait prevalence rate among twin individuals in registry.
e_1	Probability that an individual twin with disease or trait will be enrolled in twin panel from which registry is formed.
e_2	Probability that an individual twin without disease or trait will be enrolled in twin panel from which registry is formed.
A	Ratio of enrollment probability for individual twins into the twin panel for those without versus those with the disease or trait ($= e_2/e_1$).
C_{pbp}	Probandwise concordance rate for the disease or trait in the population.
C_{pbr}	Probandwise concordance rate for the disease or trait in the registry.
C_{pwp}	Pairwise concordance rate for the disease or trait in the population.
C_{pwr}	Pairwise concordance rate for the disease or trait in the registry.
Subscript $_M$	refers to monozygotic twins.
Subscript $_D$	refers to dizygotic twins.

Table 2. Number of Twin Pairs in the Population as a Function of Concordance for Trait and Number of Twins Enrolled per Pair

Number of enrolled twins per pair	Concordant for trait		Discordant for trait		Concordant for absence of trait	
	Term	Prev	Term	Prev	Term	Prev
2	C_2	Ce_1^2	D_2	De_1e_2	U_2	Ue_2^2
1	C_1	$2Ce_1(1-e_1)$	D_{1a}	$De_1(1-e_2)$	U_1	$2Ue_2(1-e_2)$
			D_{1b}	$De_2(1-e_1)$		
0	C_0	$C(1-e_1)^2$	D_0	$D(1-e_1)(1-e_2)$	U_0	$U(1-e_2)^2$
Total		C		D		U

$N = C + D + U$

e_1 = probability of enrollment of individual with trait X.

e_2 = probability of enrollment of indiivdual without trait X.

Prev = prevalence.

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

probability of enrollment of a twin is assumed to be independent of the enrollment status of his cotwin. There is only one enrollment procedure; therefore, no attempt is made to enroll an unenrolled cotwin of an enrolled twin. We assume that the risk for the trait is independent of age or sex. As outlined in Table 2, the frequency of the 10 classes of twins in the population can be easily calculated according to the model.

We then form a final twin registry consisting only of twin pairs where both members are enrolled in the panel. Although initially included in the preliminary twin panel, twin pairs with only one enrolled member are now excluded from the registry. Probandwise concordance rate in the population (C_{pbp}) and the registry (C_{pbr}) for trait X will therefore be

$$C_{pbp} = \frac{2C}{2C + D} \tag{1}$$

and

$$C_{pbr} = \frac{2C_2}{2C_2 + D_2} \tag{2}$$

Following Table 2, eq (2) can be re-expressed as

$$C_{pbr} = \frac{2Ce_1^{\,2}}{2Ce_1^{\,2} + De_1e_2} \tag{3}$$

We now define a new term, A, as

$$A = \frac{e_2}{e_1} \tag{4}$$

If $A > 1$, then individuals without trait X are more likely to be enrolled in the registry than those with X. The opposite is true if $A < 1$. Eq (3) can be re-expressed as

$$C_{pbr} = \frac{2C}{2C + DA} \tag{5}$$

We now want to express the probandwise concordance rate in the registry as a function of the probandwise concordance rate in the population. Dividing the numberator and denominator of eq (5) by (2C + D) and simplifying gives

$$C_{pbr} = \frac{C_{pbp}}{C_{pbp} + (1 - C_{pbp})\,A} \tag{6}$$

Eq (6), which shows that probandwise concordance rate in the registry is a function only of probandwise concordance rate in the population and A, can be re-expressed in the following two useful forms

$$C_{pbp} = \frac{A}{\frac{1}{C_{pbr}} - 1 + A} \tag{7}$$

$$\frac{C_{pbr}}{C_{pbp}} = \frac{1}{A + C_{pbp}\,(1 - A)} \tag{8}$$

Figure 1 illustrates the relationship between the probandwise concordance rates in the population and the registry for 13 values of A, ranging from 0.1 to 10. For values of

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

A >1, (ie, trait X decreases the probability of enrollment) the probandwise concordance rate in the registry underestimates the population probandwise concordance rate. For values of A < 1 (ie, trait X increases the probability of enrollment), the probandwise concordance rate in the registry overestimates the population probandwise concordance rate.

A further feature of the relationship of these two concordance rates is seen in eq (8), which shows that the ratio between the probandwise concordance rates in the registry and population is inversely proportional to the population probandwise concordance rates. In other words, for a given value of A, the lower the population concordance rate, the greater is the proportional change between the registry and population probandwise concordance rates.

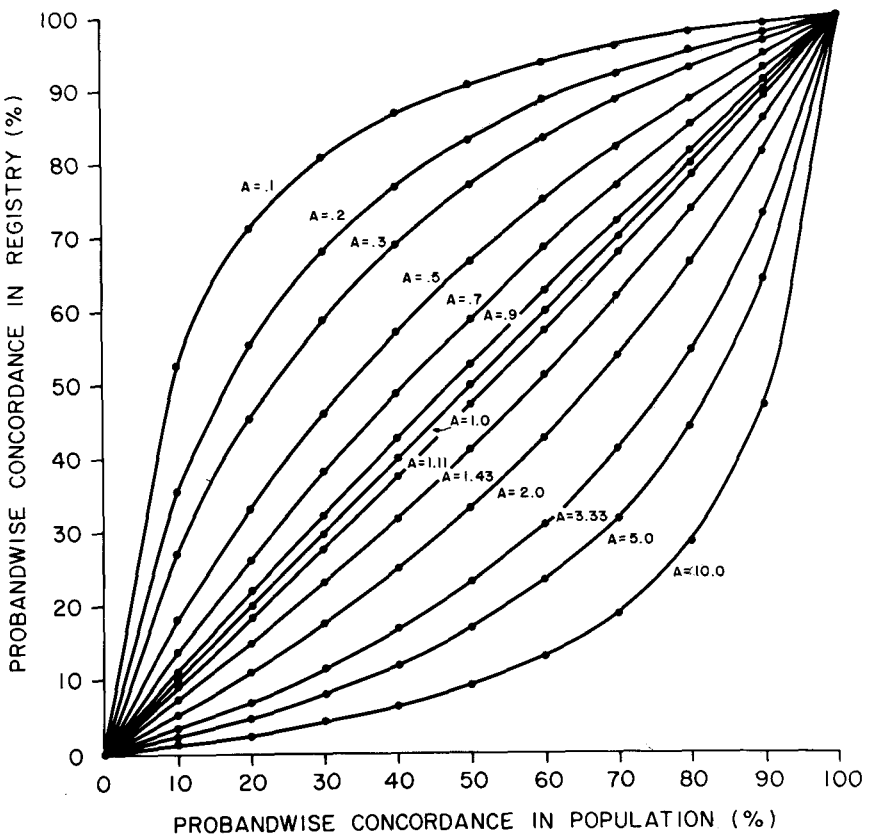

Fig. 1 - The relationship between probandwise concordance for trait X in the population and twin registry as a function of A (the ratio of probability of enrollment for individuals without X and with X). For this and subsequent figures, it is assumed that the registry is formed from pairs where both members have been enrolled in a single enrollment procedure.

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

Pairwise concordance rate in the population (C_{pwp}) and registry (C_{pwr}) are as follows

$$C_{pwp} = \frac{C}{C + D} \tag{9}$$

$$C_{pwr} = \frac{C_2}{C_2 + D_2} \tag{10}$$

From eq (10), following the logic outlined above, the following formulas can be derived:

$$C_{pwr} = \frac{C}{C + DA} \tag{11}$$

$$C_{pwr} = \frac{C_{pwp}}{C + (1 - C_{pwp})\,A} \tag{12}$$

The relationship between pairwise concordance rates in the population and registry is a function of the population pairwise concordance rate and A.

The Effect of Selection on Prevalence Rates

The population prevalence rate of trait X (P_p) is

$$P_p = \frac{2C + D}{2C + 2D + 2U} \tag{13}$$

while the prevalence rate of the trait in the registry (P_r) is

$$P_r = \frac{2C_2 + D_2}{2C_2 + 2D_2 + 2U_2} \tag{14}$$

which can be re-expressed as

$$P_r = \frac{2C + DA}{2C + 2DA + 2UA^2} \tag{15}$$

As outlined in Appendix 1, the following two formulas can be derived from eq (15). They show that the prevalence rate in the registry is a function of A, and the concordance and prevalence rates in the population

$$P_r = \frac{C_{pbp} + A\,(1 - C_{pbp})}{C_{pbp}\,(1 - A)^2 + 2A\,(1 - A) + \dfrac{A^2}{P_p}} \tag{16}$$

and the prevalence rate in the population can be expressed as a function of A and the concordance and prevalence rates in the registry

$$P_p = \frac{P_r A}{1 + 2P_r\,(A - 1) - \dfrac{A + P_r\,(1 - A)^2 - 1}{A + \dfrac{1}{C_{pbr}} - 1}} \tag{17}$$

Figure 2 illustrates the effect of differences in enrollment probability on prevalence of trait X in the twin registry. Here, the prevalence rate in the population has been set at 10%, but it can be shown (details available on request) that a similar qualitative relationship exists regardless of the magnitude of the population prevalence rate. The

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

prevalence rate for the trait in the registry is expressed as a function of the population probandwise concordance rate and A. For any given value of A, the change in prevalence rates from population to registry is greater, the greater the population probandwise concordance rate. In other words, when $A > 1$, given the same population prevalence rate, the registry prevalence rate will be lower, the higher the population probandwise concordance rate. This result is intuitively sensible, because when $A > 1$, chances of enrollment for an individual with trait X is less than that for an individual without the trait. Because inclusion in the registry is by pairs, the selection against entry into the registry is functionally greater for affected members of concordant than for affected members of discordant pairs. Since the probandwise concordance rate is a direct measure of the proportion of affected individuals who are members of affected pairs, the higher such a proportion, the greater the effective selection against individuals with the trait and the lower the registry prevalence for the trait. By similar logic, it can be shown that when $A < 1$, for the same population prevalence, the registry prevalence will be higher, the higher the population probandwise concordance rate.

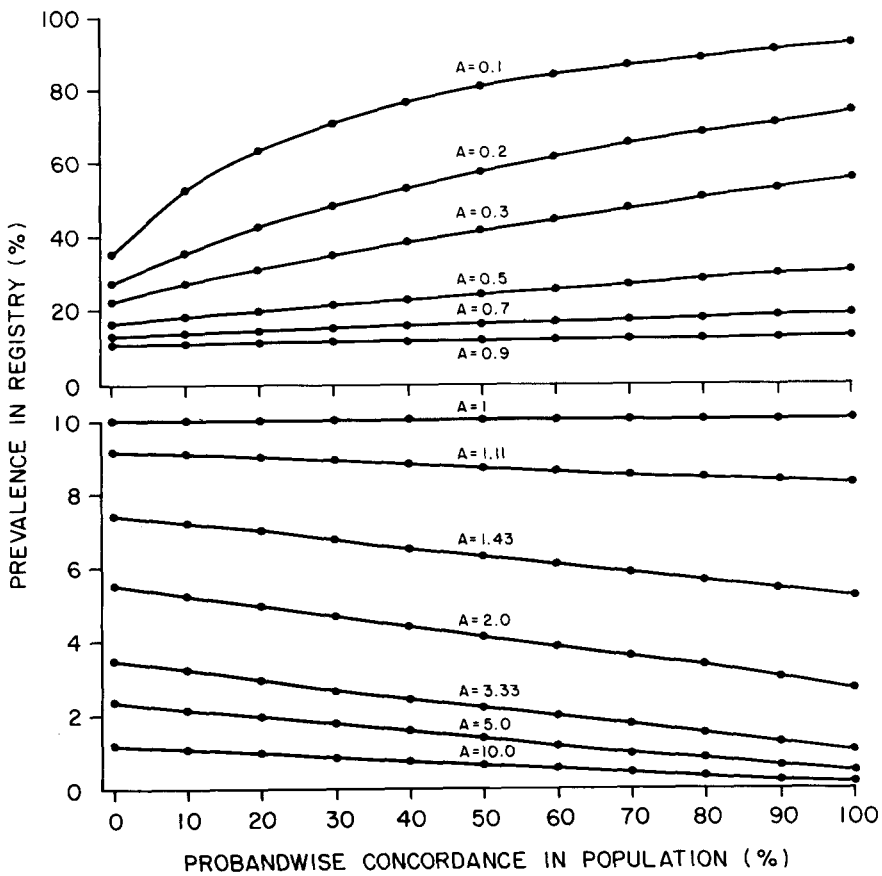

Fig. 2 - The prevalence of trait X in the registry as a function of probandwise concordance for the trait in the population, and the value of A, given that the population prevalence of the trait equals 10.0%. A qualitatively similar relationship is found for all values of population prevalence.

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

Effect of Selection on Concordance and Prevalence Rates in MZ and DZ twins

Until now, twin pairs have been considered regardless of zygosity. Given that genetic factors influence the etiology of trait X, the probandwise concordance rate for X in the population will be greater in MZ than in DZ twins. In Table 3, we show the effect of varying values of A on the prevalence and probandwise concordance rates in the registry for MZ and DZ twins, assuming a population prevalence rate of 1.0% and MZ and DZ probandwise concordance rates of 40 and 10% respectively. For values of A greater than 1, the prevalence rate of X in the registry is lower in MZ than in DZ twins. The opposite occurs when $A < 1$. As values of A change, the absolute change in probandwise concordance rates is greater in MZ twins, while the relative change is greater in DZ twins.

The Effect of Selection on Estimates of Heritability of Liability

Twin studies are frequently used to assess the relative contribution of genetic and environmental factors in the etiology of a trait or disorder. Although numerous statistics have been proposed for this purpose, probably the most helpful and widely used have been estimates for the heritability (h^2), common environment (c^2) and random environment (e^2) based on the correlation of liability [2]. This statistic measures the correlation between relatives for a latent, normally distributed liability to illness. Following the fomula of Smith [12], this correlation (r) is based on both the population risk of illness and the risk of illness in relatives. Using the correlation in liability of MZ (r_{MZ}) and DZ (r_{DZ}) twins, h^2, c^2 and e^2 can be estimated as follows:

$$h^2 = 2(r_{MZ} - r_{DZ}) \tag{18}$$

$$c^2 = 2r_{DZ} - r_{MZ} \tag{19}$$

$$e^2 = 1 - (h^2 + c^2) \tag{20}$$

The effect of differential enrollment on estimates of h^2, c^2 and e^2 for the trait under consideration is also seen in Table 3. When $A > 1$, h^2 and e^2 are underestimated and c^2 is overestimated. When $A < 1$, e^2 is overestimated and h^2 and c^2 are underestimated. Tables 4 and 5 present the effect of differential enrollment on estimates of h^2, c^2 and e^2 for 6 parameter sets including the one presented in Table 3. For each of three population prevalence rates (1, 10 and 25%), one parameter set is for a trait which has a high heritability (0.7-0.8) with only a small contribution from common environment (< 0.10) (Table 4), while the other describes a trait with a modest heritability (0.2-0.3) and a large common environmental component (0.45-0.55) (Table 5). Except for very low values of A, values for c^2 are negatively and those for e^2 are positively correlated with values of A. However, the magnitude of these relationships is greater for rare than for common traits. The effect of A on estimates of h^2 are more complex. For a rare trait, h^2 is maximal when $A = 1$, and decreases when values of A deviate in either direction. For a very common trait (eg, prevalence rate = 25%), values of h^2 are maximal at the lowest values of A and generally decrease as A increases. For intermediate prevalence rates, estimates of h^2 tend to have one peak at low values of A and another at high values of A. In general, the changes in h^2 are not striking, particularly when values of A do not deviate markedly from 1. Changes in c^2 and e^2 are somewhat greater but still not very large as long as values for A are between 0.5 and 2.0.

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

Table 3. The Effect of Varying Values of A on Prevalence, Concordance, Correlation of Liability, and "Genetic" Parameters in MZ and DZ Twin Pairs*

A	Prevalence (%)		Probandwise concordance		Correlation of liability		h^2	c^2	e^2
	MZ	DZ	MZ	DZ	MZ	DZ			
.1	30.58	15.07	87.0	52.6	.978	.699	.558	.420	.022
.2	11.36	6.39	76.9	35.7	.940	.625	.630	.310	.060
.3	6.03	3.91	69.0	27.0	.916	.576	.680	.236	.084
.5	2.73	2.15	57.1	18.2	.879	.513	.732	.147	.121
.7	1.66	1.48	48.8	13.7	.849	.474	.750	.099	.151
.9	1.16	1.12	42.6	11.0	.824	.446	.756	.068	.176
1.0	1.00	1.00	40.0	10.0	.813	.435	.756	.057	.187
1.11	0.87	0.89	37.5	9.1	.801	.424	.754	.047	.199
1.43	0.62	0.68	31.8	7.2	.773	.400	.746	.027	.227
2	0.40	0.48	25.0	5.3	.733	.370	.726	.007	.267
3.33	0.22	0.28	16.7	3.2	.671	.331	.680	−.009	.329
5	0.14	0.19	11.8	2.2	.623	.304	.638	−.015	.377
10	0.06	0.09	6.3	1.1	.548	.265	.566	−.018	.452

* Population prevalence equals 1% and the probandwise concordance in MZ and DZ twins in the population is, respectively, 40% and 10%.

Table 4. The Effect of Varying Values of A on Estimates of Genetic Parameters for a Trait with a High Population Heritability

Pop Prev	0.01			0.10			0.25		
Pop MZ Conc	0.40			0.60			0.70		
Pop DZ Conc	0.10			0.30			0.45		
A	h^2	c^2	e^2	h^2	c^2	e^2	h^2	c^2	e^2
0.1	.558	.420	.022	.780	.157	.063	.894	−.050	.156
0.2	.630	.310	.060	.702	.212	.086	.828	.031	.141
0.3	.680	.236	.084	.686	.211	.104	.792	.066	.142
0.5	.732	.147	.121	.700	.173	.127	.762	.086	.152
0.7	.750	.099	.151	.716	.136	.148	.762	.075	.163
0.9	.756	.068	.176	.728	.106	.166	.758	.067	.175
1.0	.756	.057	.187	.732	.093	.175	.762	.058	.180
1.11	.754	.047	.199	.736	.080	.184	.762	.051	.187
1.43	.746	.027	.227	.742	.051	.207	.764	.032	.204
2.0	.726	.007	.267	.738	.019	.243	.764	.030	.233
3.33	.680	−.009	.329	.714	−.018	.304	.746	−.033	.287
5.0	.638	−.015	.377	.678	−.034	.356	.718	−.054	.336
10.0	.566	−.018	.452	.604	−.046	.442	.648	−.070	.424

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

Table 5. The Effect of Varying Values of A on Estimates of Genetic Parameters for a Trait with a Low Population Heritability

Pop Prev		0.01			0.10			0.25	
Pop MZ Conc		0.30			0.50			0.60	
Pop DZ Conc		0.20			0.40			0.55	
A	h^2	c^2	e^2	h^2	c^2	e^2	h^2	c^2	e^2
0.1	.146	.785	.069	.250	.569	.181	.174	.481	.345
0.2	.168	.722	.110	.222	.598	.180	.160	.539	.301
0.3	.188	.671	.141	.218	.593	.189	.156	.558	.286
0.5	.214	.597	.189	.224	.563	.213	.150	.567	.283
0.7	.224	.550	.226	.232	.531	.237	.150	.558	.292
0.9	.228	.517	.255	.238	.503	.259	.150	.546	.304
1.0	.230	.502	.268	.240	.490	.270	.152	.538	.310
1.11	.228	.490	.282	.242	.478	.280	.152	.531	.317
1.43	.228	.459	.313	.244	.448	.308	.152	.510	.338
2.0	.224	.421	.355	.244	.407	.349	.154	.476	.370
3.33	.210	.374	.416	.236	.351	.413	.150	.421	.429
5.0	.198	.341	.461	.224	.313	.463	.144	.378	.478
10.0	.176	.296	.528	.198	.260	.542	.130	.313	.557

Effect of Selection on Frequency of MZ and DZ Twins in a Twin Registry

If probandwise concordance rates for trait X differ in MZ and DZ twins, differential enrollment will change not only the prevalence rate of the trait in the registry, but also the relative frequency of MZ and DZ twins in the registry. We initially assume, for simplicity sake, that the ratio of MZ and DZ twins in the population is 1. Let P_{pM} and P_{pD} equal the population prevalence rates and P_{rM} and P_{rD} the registry prevalence rates for the trait in MZ and DZ twins, respectively. C_{pbpM} and C_{pbpD} are the population probandwise concordance rates in MZ and DZ twins, respectively. Assuming that $P_{pM} = P_{pD} = P_p$, it can be shown (see Appendix 2) that

$$\frac{P_{rM}}{P_{rD}} = \frac{C_{pbpM}(1-A)^2 + 2A(1-A) + \dfrac{A}{2P_p}}{C_{pbpD}(1-A)^2 + 2A(1-A) + \dfrac{A}{2P_p}} \tag{21}$$

Clearly, this ratio can now be extrapolated back to any original ratio of MZ and DZ twins in the population. A surprising result which is obvious on inspection of eq (21), is that the ratio of MZ to DZ twins in the registry will always exceed that found in the population as long as the population probandwise concordance rate for the trait under consideration is greater in MZ than in DZ twins. Intuitively, this can be understood as resulting from the fact that, under such circumstances, the proportion of pairs that are concordant either for the trait or for its absence is always greater in MZ than in DZ twins. Therefore,

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

regardless of whether A is less than 1 (and enrollment is highest for pairs concordant for the trait) or greater than 1 (and enrollment is highest for pairs concordant for the absence of the trait), net enrollment will always be higher for MZ than for DZ twins. The magnitude of the effect of the trait on the overall proportion of MZ and DZ twins in the registry is a function of its population frequency, the magnitude of the difference in population probandwise concordance rates in the two zygosity groups, and the impact of the trait on enrollment. Large effects on the ratio of the zygosity groups is only seen in traits that are common, that substantially increase the probability of enrollment and that have, in the population, a much higher concordance rate in MZ than in DZ twins. While rare traits alone have little impact on the ratio of zygosity groups, many such traits with effects in the same direction could summate to have a considerable effect on the overall proportion of MZ and DZ twins in the registry.

Correction on Enrollment Bias

Given that an investigator is working with a twin population which is organized like the NAS-NRC registry, the bias introduced by differential enrollment can be dealt with in two ways. The first method is to correct the prevalence and concordance rates obtained in the registry for the effects of enrollment bias. Given a value for A and values for the prevalence and concordance rates in the registry, the population probandwise concordance and prevalence rates for the trait of interest can be estimated from eqs (7) and (17), respectively. However, estimates for A might not be always readily available. If we assume that the prevalence rate for the trait in the population is the same in MZ and DZ twins, A can be estimated entirely from data obtained from the registry. An approximate estimate for A can be obtained from the following formula:

$$A = \frac{(1 - C_{pbrD})\,P_{rD} - (1 - C_{pbrM})\,P_{rM}}{C_{pbrM}\,P_{rM} - C_{pbrD}\,P_{rD}} \tag{22}$$

where C_{pbrM} and C_{pbrD} are probandwise concordance rates and P_{rM} and P_{rD} are prevalence rates for the trait in the registry for MZ and DZ twins, respectively. For a trait with a population prevalence rate of 1%, this estimate is accurate to within 10% for values of A from 0.2 to 5. For a trait with a population prevalence rate of 10%, this estimate is accurate to within 10% for values of A from 0.5 to 5. For more common traits, the estimate is progressively less accurate. We were unable to derive a simple and more accurate formula for A. Greater accuracy of estimation, however, can be obtained using eq (17) in an iterative fashion. By entering registry prevalence and probandwise concordance rates for MZ and DZ twins, values of A can be altered until the predicted population prevalence in MZ and DZ twins is the same to any required degree of accuracy.

The above method for correcting for the effects of differential enrollment will be illustrated using data on ischemic heart disease (IHD) (ICDA 410-414) from the NAS-NRC twin registry [8]. As of 10/81, the prevalence rate for this disorder in the registry was 6.31% in MZ and 6.62% in DZ twins. The probandwise concordance rate was 29.1% in MZ and 18.3% in DZ twins. Putting these values into eq (22) produces an estimate for A of 1.4925. Putting these values into eq (17) produces estimates for the population prevalence of 10.10% in MZ and 10.08% in DZ twins. By iteration, a value of A of 1.476 produced the same population estimate for IHD in MZ and DZ twins to 4 significant places (9.967%). Putting this value of A, and values of the registry probandwise concord-

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

ance rates into eq (7) predicts population probandwise concordance rates for IHD of 37.7% in MZ and 24.8% in DZ twins. Using the registry data, the correlation of liability for MZ and DZ twins were, respectively, 0.535 and 0.332, yielding estimates of $h^2 = 0.406$, $c^2 = 0.129$, and $e^2 = 0.465$. Using the estimated population data, the correlation of liability for IHD in MZ and DZ twins is 0.579 and 0.365, respectively. These correlations yield estimates of $h^2 = 0.428$, $c^2 = 0.151$, and $e^2 = 0.421$. In accord with the conclusions noted above, distortions in estimates of h^2, c^2 and e^2 for values of A between 0.5 and 2.0 are not great.

Of interest, using eq (22), estimates of A were obtained from the NAS-NRC twin registry for the following disorders: schizophrenia, 1.52; hypertension, 1.31; and peptic ulcer, 1.29.

A second, but more difficult, way is available to avoid the bias associated with differential enrollment. This would involve the selection of the registry on the basis of individuals and not pairs and a second enrollment procedure in which unenrolled cotwins of affected twins in the registry were followed up and examined. Under these circumstances, the probandwise concordance rate in the registry (C'_{pbr}) would equal

$$C'_{pbr} = \frac{2C_2 + C_1}{2C_2 + C_1 + D_2 + D_{1a}} \tag{23}$$

which can be simplified to

$$C'_{pbr} = \frac{2C}{2C + D} \tag{24}$$

This, of course, is an unbiased estimate of the true population probandwise concordance rate. This problem is analagous to the problem of incomplete ascertainment [1,3]. By selecting the registry by individuals and not pairs and by following up unenrolled cotwins of affected twins, an accurate estimate of probandwise concordance can be obtained regardless of enrollment bias.

However, the prevalence rate will still be biased. If only pairs where both members are enrolled in the registry in the first enrollment procedure are used for the calculation of prevalence rate, then the results will be the same as those outlined above (see eqs 14-17). If pairs with unenrolled cotwins found in the first enrollment procedure are included, the prevalence rate in the registry (P'_r) will equal

$$P'_r = \frac{2C_2 + D_2 + C_1 + D_{1a}}{2C_2 + 2D_2 + 2U_2 + C_1 + D_{1a} + D_{1b} + U_1} \tag{25}$$

which simplifies to

$$P'_r = \frac{2C + D}{2C + D + A(D + 2U)} \tag{26}$$

Comparing this to eq (13), it can be seen that when A exceeds 1, then the prevalence rate in the registry will be less than the prevalence rate in the population and the opposite will be seen when A is less than 1. Including twins evaluated in the secondary enrollment in the calculation of prevalence will make the situation even more complex. Further expressions can be developed for the prevalence rate in the registry under these new enrollment conditions, but they will not be presented here. However, it should be noted that since calculation of the correlation of liability depends on accurate estimates of both probandwise concordance and population prevalence rates, these correlations and the

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

estimates of h^2, c^2 and e^2 derived from them, will not be accurately obtained by this method unless the prevalence rates in the registry are corrected for the effect of differential enrollment.

DISCUSSION

This report began with the observation that schizophrenia in the NAS-NRC registry was less common in MZ than in DZ twins and had concordance rates lower than would be expected from the rest of the world's literature [7,8]. We proposed a model which sought to explain these two features as a result of the effect of differential probability of enrollment of individuals in a twin registry which, like the NAS-NRC registry, includes only pairs where both members are enrolled [4]. In this analysis, it was found useful to introduce the parameter "A", which is the ratio of the probability of enrollment of an individual without a given trait (X) to the probability of enrollment of an individual with that trait. When A is greater than 1, which means that the trait diminishers the probability of enrollment, both the prevalence and the concordance rate in the registry are lower than that found in the population. The magnitude of this effect increases as A increases. The opposite is seen when A is less than 1 (ie, trait X increases the probability of enrollment). Under this circumstance, both the prevalence and concordance rates for the disorder are higher in the registry than in the population, and this effect increases as A approaches 0. As the population probandwise concordance rate increases, the effect of differential enrollment on the concordance rate decreases, but its effect on the prevalence rate increases.

If we assume that genetic factors influence the probability of manifesting trait X, then the population probandwise concordance rate for the disorder in MZ twins will exceed that found in DZ twins. Therefore, if A exceeds 1 and the population prevalence rate for the disorder in MZ and DZ twins are the same, the prevalence rates for the trait in the registry will be lower in MZ than in DZ twins. If under the same conditions, A is less than 1, then the prevalence rate in the registry will be greater in MZ than in DZ twins.

Since comparisons of MZ and DZ twins are often used to infer causes of variation for human traits, we were interested in determining the effect of differential enrollment on estimates of heritability (h^2) and common and random environment (c^2 and e^2). Under almost all circumstances for a given population prevalence and probandwise concordance rate, the correlation of liability in the registry was negatively correlated with the value of A (result not shown). That is, the lower the value of A, the higher the correlation of liability was found to be. In most circumstances, values of A of less than 1 lead to an overestimation of c^2 and an underestimation of e^2. The opposite is seen when A is greter than 1. This effect was greater, the rarer the trait. The effect of differential enrollment on estimates of h^2 was more complex. Depending on the population prevalence rate for the trait, differential enrollment could either increase or decrease estimates for h^2.

Differential enrollment will not only alter the prevalence rate for the trait influencing enrollment, but it will also influence the ratio of MZ to DZ twins in the registry. Although this effect will be small with rare traits, common traits could produce a substantial excess of MZ twins in the registry if the trait markedly increases the probability of enrollment and the population probandwise concordance rate for the trait is much greater in MZ than in DZ twins.

Given that enrollment by pairs is a frequent method of construction of twin registries

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

and that the biases introduced by differential enrollment are often not trivial, the practical issue arises of how this bias should be dealt with. Two strategies are possible. The first is to accept the method of construction of the registry and correct for the effect of differential enrollment. To do this accurately, an estimate of the magnitude of the differential enrollment (A) is necessary. It is conceivable such an estimate could be obtained from knowledge of the construction of the twin panel. If such is not available, A can be estimated from registry data on MZ and DZ twins, assuming that the population prevalence rate for the disorder is the same in both zygosity groups. Once an estimate of A is available, then the population prevalence rate, the population concordance rates and the true correlations of liability for MZ and DZ twins can all be calculated based only on information from the registry. Estimates of h^2, c^2 and e^2 can then be obtained from these population estimates.

The second way of dealing with the bias of differential enrollment is to change the construction and enrollment system of the twin registry. If twins are accepted into the registry as individuals and unenrolled cotwins of affected enrolled twins are subject to secondary enrollment, then an unbiased estimate of the population probandwise concordance rate can be obtained regardless of enrollment effects. This effect is analagous to the way in which under incomplete ascertainment, a secondary enrollment procedure can result in an unbiased estimate of the true probandwise concordance rate. However, estimates of the prevalence rate for the trait in the registry are still biased.

To what extent are available observations consistent with the predictions of this model? In the NAS-NRC registry differential enrollment would bias against entry of individuals with illness, and hence values of A would be greater than 1. As predicted by the model, the prevalence rate of all major medical and psychiatric disorders [8] as well as mortality [5] is lower in the MZ than in the DZ twins. By selecting for pairs, for any disorder with a genetic component, the MZ twins in the NAS-NRC registry constitute a healthier population than the DZ pairs. At least for schizophrenia, concordance rates in the NAS-NRC registry are lower than found in other studies. When these figures are corrected for the effects of differential enrollment, the results are much more in line with other investigations [7]. When values of A are estimated for several common diseases in the NAS-NRC registry, reasonable values are obtained which suggest that selection against individuals prone to schizophrenia is somewhat greater than found for individuals prone to hypertension.

If a heritable discrete trait existed which substantially influenced the probability of volunteering for a twin registry, this model would predict that such registries should have an excess of MZ twins. This has indeed been consistently observed [9].

Several limitations of this model are worth outlining. First, no attempt was made to deal with the probabilistic nature of the variables used and estimated. Estimates for the population prevalence and concordance rates in the population obtained from observations made in the registry will have confidence intervals attached to them, and they may be fairly large because of the number of individual parameters required to estimate them. Second, we have not considered the longitudinal issues that differential enrollment raise. In the NAS-NRC registry, we are now examining a population 40 years after the selection took place. How traits selected against in 1942 will influence the prevalence rates of disorders in 1982 is obviously a complex matter. Third, we only examine the impact of differential enrollment on the trait which itself influences enrollment. If traits A and B are correlated in the population and only trait A influences selection, prevalence and concordance rates in the registry should differ from their population counterparts not only for trait A, but also for trait B.

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

A fourth limitation of the current treatment is that it deals only with discontinuous traits. Interestingly, the single previous examination of the problem of differential enrollment of which we are aware dealt only with a normally distributed quantitative trait [10]. Martin and Wilson examined the effect of truncate selection in the formation of a twin registry on the correlation between twins for that trait. The underlying assumptions of their treatment are so different from those used here that the results of the two models are difficult to compare. This is particularly true because correlation coefficients are quite sensitive to the reduction in total variance and range produced by truncation, a phenomenon which has no parallel in our treatment. Although the precise nature of their conclusions differed from those reached here, both suggest that differential enrollment can significantly alter the results of twin studies and can be ignored only at the investigator's peril.

Acknowledgments. The data from the National Academy of Sciences-National Research Council Twin Registry used in this report was kindly supplied by Dr. C. Dennis Robinette. This work was supported in part by funds from the Department of Mental Health and Mental Retardation of the Commonwealth of Virginia (to KSK), a research fellowship (to NVH) from the Danish Cancer Society and a travel grant (to NVH) from the Union of Scandinavian Cancer Societies.

REFERENCES

1. Allen G, Hrubec Z (1979): Twin concordance. A more general model. Acta Genet Med Gemellol 28:3-13.
2. Falconer DS (1965): The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet 29:51-76.
3. Holm NV (1983): A note on ascertainment probability in the Allen/Hrubec twin model. Acta Genet Med Gemellol 32:37-47.
4. Hrubec Z, Neel JV (1978): The National Academy of Sciences-National Research Council Twin Registry: Ten years of operation, In Nance WE et al (eds): Twin Research, Part B: Biology and Epidemiology: New York: Alan R. Liss, pp. 153-172.
5. Hrubec Z, Neel JV (1981): Familial factors in early deaths: Twins followed 30 years to ages 51-61 in 1978. Hum Genet 59:39-46.
6. Hyde RW, Chisholm RM (1944): Studies in medical sociology III. The relation of mental disorders to race and nationality. New England J Med 231:613-619.
7. Kendler KS (1983): Overview: A current perspective on twin studies of schizophrenia. Am J Psychiatry 140:1413-1425.
8. Kendler KS, Robinette CD (1983): Schizophrenia in the National Academy of Sciences-National Research Council Twin Registry: A 16-year update. Am J Psychiatry 140:1551-1563.
9. Lykken DT, Tellegen A, DeRubies R (1978): Volunteer bias in twin research: The rule of two thirds. Soc Biol 25:1-9.
10. Martin NG, Wilson RS (1982): Bias in the estimation of heritability from truncated samples of twins. Behav Genet 12:467-472.
11. Seltzer CC, Jablon S (1974): Effects of selection on mortality. Am J Epidemiol 100:367-372.
12. Smith C (1974): Concordance in twins: Methods and interpretation. Am J Hum Genet 26:454-466.

Correspondence: Kenneth S. Kendler, MD, Dept Psychiatry, MCV, PO Box 710, Richmond VA 23298, USA.

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press

APPENDIX 1

From Table 2, we derive the following expressions for C, D and U in terms of N, P_p and C_{pbp}:

$$C = C_{pbp} NP_p \tag{A1}$$

$$D = 2NP_p (1 - C_{pbp}) \tag{A2}$$

$$U = N [1 - P_p (2 - C_{pbp})] \tag{A3}$$

These equations are now substituted into eq (15) and, after dividing numerator and denominator by 2N, the following is obtained:

$$P_r = \frac{P_p C_{pbp} + P_p A (1 - C_{pbp})}{P_p C_{pbp} + 2P_p A (1 - C_{pbp}) + A^2 [1 - P_p (2 - C_{pbp})]} \tag{A4}$$

Dividing numerator and denominator by P_p, and rearranging yields eq (16). If both sides of eq (16) are now multiplied by the denominator of the right side of eq (16) and rearranged, we obtain

$$\frac{P_r}{P_p} = \frac{C_{pbp} (1 - A) + A - P_r [C_{pbp} (1 - A)^2 + 2A (1 - A)]}{A^2} \tag{A5}$$

Inverting eq (A5) and multiplying both sides by P_r, we obtain

$$P_p = \frac{P_r A^2}{C_{pbp} (1 - A) + A - P_r [C_{pbp} (1 - A)^2 + 2A (1 - A)]} \tag{A6}$$

Eq (17) is obtained by substituting eq (7) into eq (A6), dividing the numerator and denominator by A, and rearranging.

APPENDIX 2

We begin by obtaining from Table 2 the registry prevalence for MZ and DZ twins, where the subscript M indicates MZ and D DZ:

$$P_{rM} = C_M e_1^2 + D_M e_1 e_2 + U_M e_2^2 \tag{A7}$$

$$P_{rD} = C_D e_1^2 + D_D e_1 e_2 + U_D e_2^2 \tag{A8}$$

We divided eq (A7) by eq (A8), substitute eqs (A1-A3) into the result and then divide numerator and denominator by Ne_1^2 to obtain

$$\frac{P_{rM}}{P_{rD}} = \frac{C_{pbpM} P_{pM} + 2P_{pM} A (1 - C_{pbpM}) + A^2 [1 - P_{pM} (2 - C_{pbpM})]}{C_{pbpD} P_{pD} + 2P_{pD} A (1 - C_{pbpD}) + A^2 [1 - P_{pD} (2 - C_{pbpD})]} \tag{A9}$$

Assuming that $P_{pM} = P_{pD} = P_p$, this can be simplified to eq (21).

https://doi.org/10.1017/S0001566000004657 Published online by Cambridge University Press