The Consequences of Selective Participation on Behavioral-Genetic Findings: Evidence from Simulated and Real Data

Alan Taylor

Institute of Psychiatry, King's College, London, United Kingdom

Nonresponse occurs when individuals either have no chance of being included in a study (noncoverage), refuse to take part (unit nonresponse), or fail to give complete information (item nonresponse). The purpose of this article is to test the possible biasing effects of nonresponse on the results of behavioral-genetic studies. Simulations and a real data 'natural' experiment were used to determine the impact of nonresponse on estimates of additive genetic and environmental effects. The simulations used realistic twin-pair correlations and models of nonresponse derived from prior research. The real data 'natural experiment' used data from a nationally representative birth-cohort twin study (E-Risk Study) and compared model results from families who had responded to a mail survey to those from all study cases. Results showed that the primary influence of nonresponse was to attenuate the effect of the shared environment and to inflate estimates of nonshared environment and additive genetic effects. At high levels of nonresponse a spurious nonadditive genetic effect (suggesting genetic dominance) was also found. Study nonresponse was shown to have the potential to bias the findings of behavioral-genetic research. Design and analysis methods that can be used to alleviate this potentially important biasing effect in behavioral-genetic studies are discussed in light of these findings.

In recent years, behavioral-genetic research has assumed increased importance for two reasons. First, as research on molecular genetics has blossomed due to technical advances (Plomin & Crabbe, 2001), quantitative behavioral-genetic studies have become a necessary preliminary step to identifying heritable phenotypes that can be usefully examined at the molecular-genetic level (Martin et al., 1997). Second, social scientists have also turned their attention to behavioral-genetic methods as a tool for testing causal hypotheses about the effects of measured environmental influences (Rutter et al., 2001). It is therefore important to establish any potential methodological limitations of behavioral-genetic research so that specific research findings can be appropriately evaluated. Many of the basic assumptions of the behavioral-genetic approach have been repeatedly tested (e.g., Kendler et al., 1993, on the equal-environments assumption), so that the substantive results can have maximum credibility. Stoolmiller (1999) has drawn attention to the issue of nonresponse bias in the context of adoption studies. Because of demographic changes, adoption studies are now less frequently used and it is thus important to evaluate nonresponse effects on the primary method of behavioral genetics: the twin study. The goal of this article is therefore to evaluate how sample selection or nonresponse affects estimates of genetic and environmental variation derived from twin studies.

The effects of nonresponse on study estimates are well-documented in survey research (e.g., Brick & Kalton, 1996). Nonresponse is a pervasive problem in all social and medical research and a large literature details the problem and discusses possible remedies (e.g., Brick & Kalton, 1996; Levy & Lemeshow, 1991; Skinner et al., 1989). The effect of nonresponse is to bias study estimates of population characteristics if the sample units which provide valid information are systematically different from those that are missing from the study. For this bias to occur, the probability of responding must vary differentially across the sample. Nonresponse is typically subclassified into three groups. Firstly, *noncoverage* occurs when the list of units to be sampled, which may be at the level of individuals, households or geographic areas, fails to contain all the units in the population of interest. For example, in a study of twins born in a given year, a sample obtained from a register of births would suffer from noncoverage if the registration procedures did not capture illegitimate births. Similarly,

Received 22 January, 2004; accepted 28 July, 2004.

Address for correspondence: Alan Taylor, Social, Genetic and Developmental Psychiatry Centre, Box Number P080, Institute of Psychiatry, King's College London, De Crespigny Park, Denmark Hill, London SE5 8AF, UK. E-mail: A.Taylor@iop.kcl.ac.uk

a population study investigating the prevalence of psychiatric morbidity would suffer from noncoverage if the sample was selected based on a list of private households, as there would be no representation of individuals who were homeless or institutionalised. Secondly, after a sample has been selected, a further type of nonresponse bias can occur when those selected refuse to take part in the study or, as is sometimes the case, cannot be found based on information available on the sampling frame. This type of bias is termed *unit nonresponse,* as no data are collected from certain sample units. Thirdly, nonresponse bias can occur even for sample units that have consented to be included in the study as they may fail to give complete information. This type of nonresponse is termed *item nonresponse*. This is particularly problematic for psychological and psychiatric studies as their substantive focus requires asking highly personal questions about sensitive topics such as substance abuse, antisocial behavior and mental illness. In any given study all of these types of nonresponse biases are liable to appear to varying degrees. To aid discussion we will subsume all types of nonresponse identified above under the generic heading of nonresponse.

The processes generating the biasing effects of nonresponse are of two interrelated types. Firstly, units in the population may have differing probabilities of response to a given study. This may be due to a low probability of inclusion in the sampling frame (noncoverage), a low probability of response to the whole study (unit nonresponse) or a low probability of response to certain items (item nonresponse). If this propensity to respond is positively correlated with one of the measures of interest in the study, be it a risk factor or an outcome, the study sample will underrepresent units with high values of the measure. This type of differential nonresponse is also termed *selection* in the social-science literature (Berk, 1983), and *soft selection* in behavioral genetics (Neale et al., 1989). Secondly, the propensity to respond may be constant over a large range but shift to zero at the upper tail. Again, if the propensity to respond is positively correlated with a measure of interest, individuals scoring at the high end of the measure will fail to respond. This type of nonresponse is sometimes termed *truncation* in the social-science literature (Berk, 1983) and *hard selection* in behavioral genetics (Neale et al., 1989).

As an illustration of the effects of differential nonresponse, consider the twin data in Figure 1. The figure shows the effects of differential nonresponse on twin-pair correlations of .7 (top row) and .5 (bottom row). The first panel in each row gives the true scatter of phenotypic values at these correlations, while the second panel gives the scatter after differential nonresponse. This example was generated by assuming that the probability of response decreases as the phenotypic value increases. For both correlations, differential response reduces the estimate of the population correlation between twins, and the effect is larger for the lower correlation (as can be seen from the regression lines in the final panel). Therefore, differential nonresponse has the potential to bias estimated correlations, which will lead to biased behavioral-genetic estimates obtained from model-fitting.

Previous studies of nonresponse bias in behavioral-genetic research have used mathematical models of the probability of response to investigate the effects of nonresponse on twin correlations (Martin & Wilson, 1982; Neale et al., 1989). Building on these earlier studies, the present article uses two different approaches to investigate the impact of nonresponse. Firstly, simulations are used to explicitly test the substantive impact of particular forms of nonresponse. Secondly, real data are used to confirm the results from simulations using data from a large nationally representative twin study that included a 'natural' experiment in nonresponse. Previous research on nonresponse in behavioral-genetic research has focused on investigating the effects of nonresponse on correlations. In the present study, we sought to further illustrate the effects of nonresponse by fitting standard behavioral-genetic models using Mx (Neale et al., 1999) so that both the attenuation in correlations and the resulting effects on behavioral-genetic model parameters could be investigated and illustrated.

Study 1A: The Consequences of Differential Study Participation on Genetic and Environmental Effects in Twin Studies

Study 1A investigated the effects on study estimates of population parameters when units of a population, such as individuals or twin pairs, have differing probabilities of responding. There are likely to be many substantive instances where the probability of response is related to the actual phenotype under investigation. Certain values of the phenotype are then likely to be underrepresented in the final sample, giving rise to concern over the generality of study results. For example, it is likely that those with high levels of disruptive or aggressive behavior are more at risk of nonresponse than those with lower levels. Therefore, results from studies about the genetic and environmental determinants of human aggression may be affected by differential response probabilities.

Method

Correlation Patterns

The simulations were based on samples of twins. Differing patterns of monozygotic (MZ) and dizygotic (DZ) correlations were simulated so that the effect of differential nonresponse could be investigated over varying twin-pair correlations. S-Plus (S-Plus 2000 User's Guide, 1999) was used to simulate standard normal variables for each twin in a pair so that the expected correlations were of a specified form. The values of the correlations used are given in the first

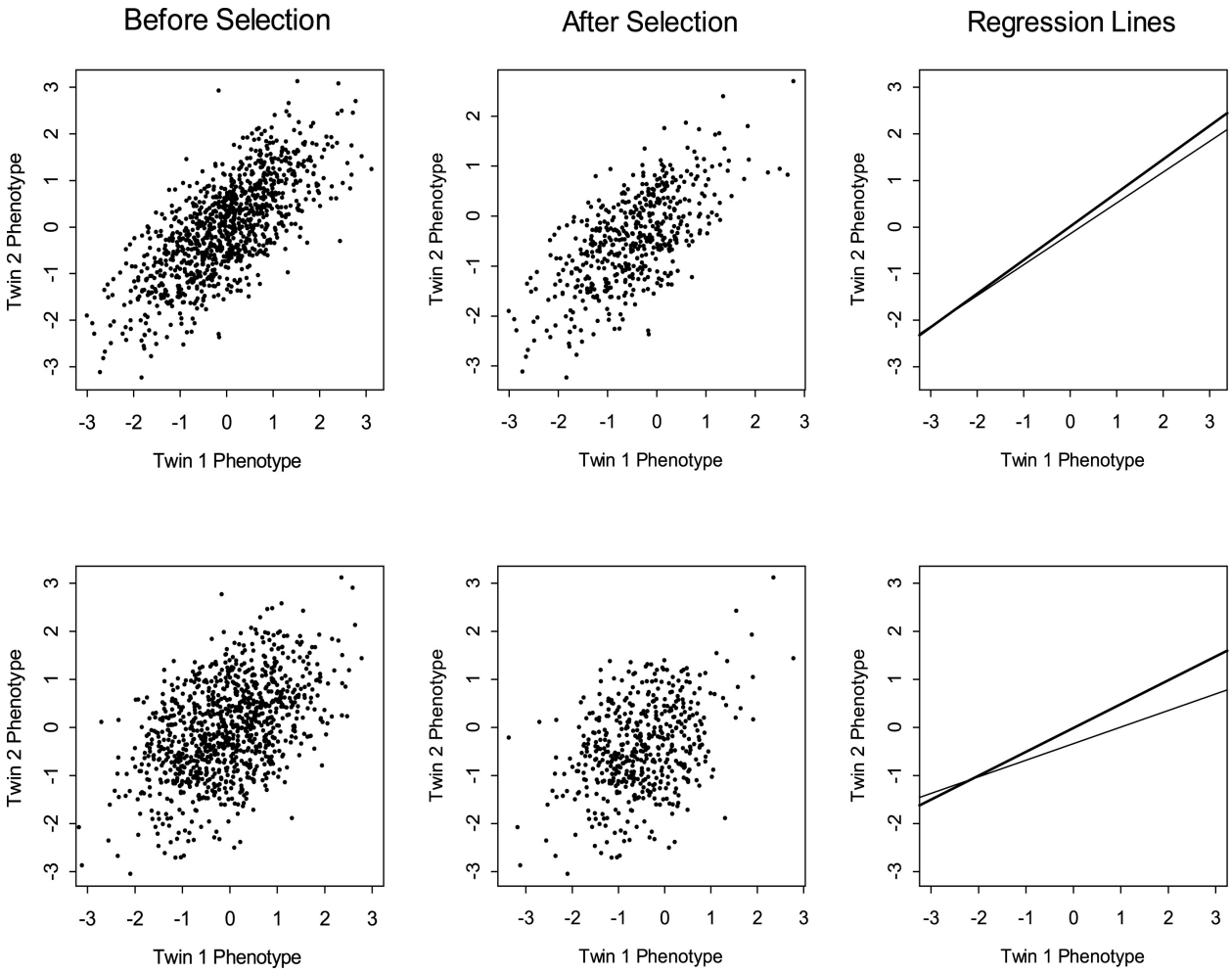

Figure 1

Effect of differential nonresponse on twin correlations of .7 (upper row) and .5 (lower row).

two columns of Table 1. The last three columns indicate the expected components of variance for each of the parameters of a behavioral-genetic model for these pairs of correlations. We chose this pattern of correlations because the primary effect of differential nonresponse is to reduce the size of correlations, as demonstrated in Figure 1 (Martin & Wilson, 1982); we wanted to investigate the effect of varying the absolute magnitude of the correlations. Under a standard biometric genetic model (Plomin et al., 2001) consisting of the effects of additive genes (A), shared (C), and nonshared (E) environments (ACE model), DZ twin correlations will be smaller than MZ twin correlations unless there is no effect of additive genes (where correlations will be the same for both zygosities). Under an ACE model, this implies that differential nonresponse would have the effect of increasing the difference between the MZ and DZ correlations. As the estimate of the shared environment is determined by the similarity of these correlations, this would imply that differential response would have its largest effect on estimates of the shared environment. Therefore, the DZ correlations were increased by a value of .05 and the MZ correlations adjusted to allow for the investigation of both the effect of the magnitude of the twin correlations and of differing magnitudes of the shared-environment effect. Three values of the shared-environment effect were used that cover the range of effects commonly found in behavioral-genetic research about cognitive abilities, personality, and psychiatric disorders: a small shared-environment effect of 10%; a medium effect of 20%, and a large effect of 30% (Rhee & Waldman, 2002; Nigg & Goldsmith, 1994; Loehlin, 1989).

This profile of correlations provides a realistic basis from which to extrapolate the likely effects of nonresponse on substantive research. The chosen correlations give an adequate distribution of MZ/DZ profiles, while focusing on DZ correlations that cover the lower end of the range. This allowed us to investigate the maximum possible effects of nonresponse

Table 1

Simulated Correlation Profiles

Correlation		Expected ACE components of variance		
MZ	DZ	Genetic (A2)	Shared environment (C2)	Unique environment (E2)
.5	.3	.4	.1	.5
.5	.35	.3	.2	.5
.6	.4	.4	.2	.4
.7	.45	.5	.2	.3
.7	.5	.4	.3	.3
.8	.5	.6	.2	.2

bias. Also, as published meta-analyses (Rhee & Waldman, 2002) indicate that studies show wide variations in estimated MZ/DZ correlations, simulations were focused on the range of correlation profiles that would be most susceptible to the effects of differential nonresponse.

Propensity for Nonresponse

To simulate the effects of nonresponse, a third variable was generated which can be viewed as indexing the propensity for nonresponse for a given twin pair. This variable was simulated to correlate at different levels with the simulated phenotypic variables, and was specified at values from .0 to .8 using increments of .1. This allowed us to investigate the differential effects of nonresponse when the propensity for nonresponse was correlated at different levels with the phenotype under investigation. This propensity variable can be thought of as a latent construct indexing the many individual variables which have been shown to be related to nonresponse (e.g., Rosenthal & Rosnow, 1975; Groves, 1989). This additional complexity was added to simulations to make them more realistic since it is unlikely that in real studies, the propensity for nonresponse would be completely correlated with the phenotypic values, as has been assumed in previous research (e.g., Martin & Wilson, 1982). In order to simplify these simulations, it has been assumed that the value of the propensity for nonresponse variable is constant within a twin pair. This is likely to be a reasonable assumption for twin studies of younger children because guardians make choices about their joint participation, but may become less appropriate with older twins (Lykken et al., 1987) as the propensity for nonresponse is likely to vary between pairs to the extent that MZ twins are more likely to make similar choices than DZ twins (Scarr & McCartney, 1983).

Nonresponse Patterns

To simulate differential nonresponse patterns, a step function was defined relating the propensity variable to a given probability of response. This step function was defined by splitting the nonresponse propensity variable into quintiles and then assuming a constant probability of response within each quintile. This is a simplifying assumption which allowed the use of a discrete-response function rather than applying a more complex continuous function. At the limit, this approach can be made to approximate a continuouslymapped function such as that used in Neale et al. (1989), provided the shape of the function is correctly specified. Two functions were applied in these simulations, which are summarized in Table 2 and illustrated in Figure 2. We have labeled these as the high nonresponse and standard nonresponse profiles. The high nonresponse profile, in which twin pairs with high values of the propensity variable are the least likely to respond, is the more extreme. The overall response rate for this profile is 50%. Response rates of this order are common in mail surveys. For example, Asch, Jedrziewski, and Christakis (1997), reporting on mail survey response rates published in medical journals, give the mean response rate of 236 surveys as 62% ± 21%. The standard nonresponse profile is less extreme, in which only twins with values in the highest quintile of the propensity variable have a marked likelihood of nonresponse. The overall response rate is 70%. This pattern is termed the standard nonresponse profile as response rates of 70% are typically found in large surveys carried out by telephone or face-to-face interviews. For example, the General Social Survey, a large, annual, omnibus personal interview survey in the United States, achieved response rates of approximately 70–80% over a period from 1975 to 1996 (Davis & Smith, 1992; General Social Survey, 2001). Similarly, the British Crime Survey, a large biannual government survey of criminal victimization in England and Wales, achieved response rates of 80% over a 10-year period from 1988 to 1998 (Mirrlees-Black et al., 1998). These two patterns allowed us to specify a realistic approximate lower and upper bound for the effect of differential nonresponse, and therefore gave a context from which to extrapolate to patterns between those two profiles.

Sample Size and Number of Simulations

The sample size for these simulations was 1000 twins of each zygosity for the high nonresponse pattern, resulting in a final sample size of 500 MZ and 500 DZ twin pairs. Compared to the size of existing twin studies this would be considered a relatively large study (e.g., Rhee & Waldman, 2002). For compara-

Table 2

Differential Response Profiles: Probability of Response for Each Quintile of the Response Propensity Variable

	Quintile					Response rate
Profile	1	2	3	4	5	
1 — High nonresponse	.9	.7	.5	.3	.1	50%
2 — Standard nonresponse	.8	.8	.8	.7	.4	70%

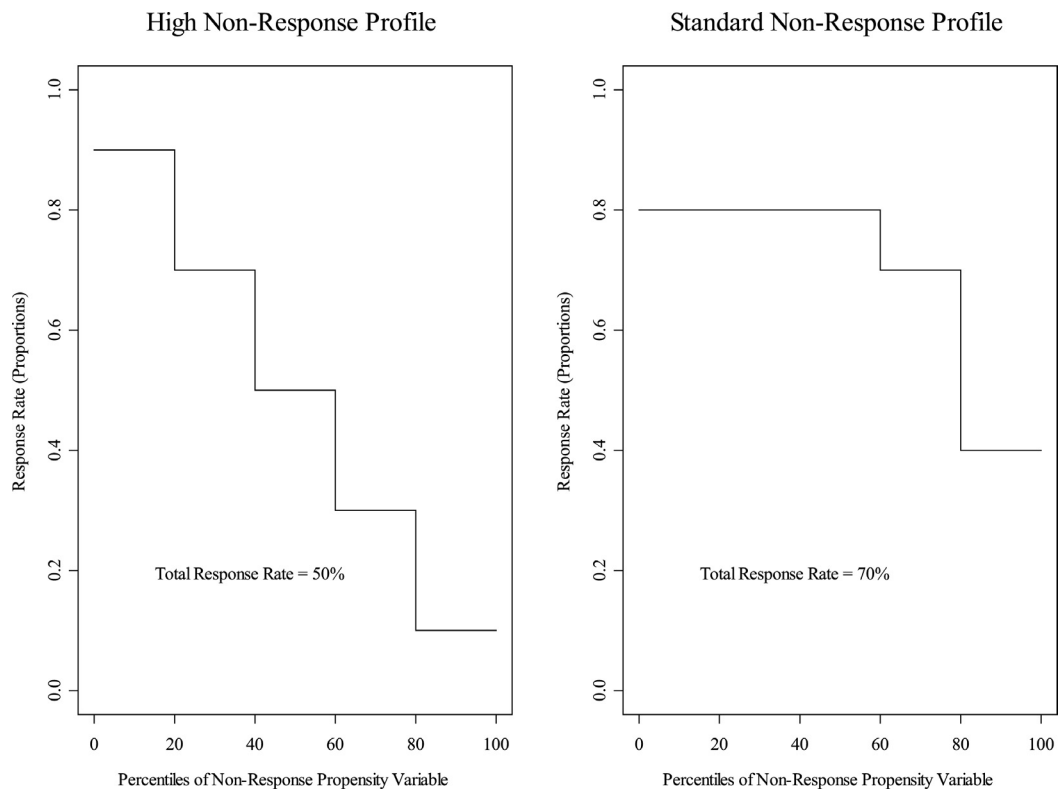

Figure 2

Probability of response by nonresponse propensity variable for high and standard nonresponse profiles.

bility of model results, the standard nonresponse pattern results were specified so that the final sample size of MZ and DZ pairs would be approximately equivalent to that of the high nonresponse profile. This required 716 pairs for each zygosity to be simulated before the application of the nonresponse profile. For each combination of nonresponse pattern and twin-pair correlation, 500 simulations were carried out. The results are presented as mean values over all simulations.

In summary, each simulation consisted of the following steps: first, for each zygosity, standard normal variables were simulated for each twin in a pair to give a sample of the required size. Variables for the phenotype of each twin and a variable for the nonresponse propensity were simulated. These simulations were designed so that the expected values of the correlation matrix for each zygosity would be those defined for a given simulation. For example, where the nonresponse propensity variable was correlated at .8 with the phenotypic variable and the MZ correlation was .5, three variables were simulated for 1000 pairs (under the high nonresponse pattern) to give the expected correlation matrix of .5 between the phenotypic variables and of .8 between these variables and the nonresponse propensity variable. Second, this simulated sample was then reduced in size using the parameters of the specific differential nonresponse model being applied. Third, the resulting matrices by zygosity were output to Mx and a standard ACE model fitted. These three steps were then repeated 500 times for each simulation and the results loaded into a S-Plus data set to calculate summary statistics.

Results

Table 3 shows the results of the simulations. To aid the reader this table is also given graphically in Figures 3 and 4. The table gives the results of applying the high and standard nonresponse profiles respectively. The table is divided into 6 sections, where each section corresponds to the correlation profiles outlined in Table 1 (i.e., .5 and .3, .5 and .35, .6 and .4, .7 and .45, .7 and .5, .8 and .5). The base correlations are indicated in bold. Each section of the table gives mean results calculated over all 500 simulations using correlations with the nonresponse propensity variable ranging from .0 to .8. The row where the phenotypic variables have a zero correlation with the nonresponse propensity variable gives a useful baseline for comparison as it shows the results when there is no differential nonresponse. Columns 2–3 show the MZ and DZ correlations after nonresponse; columns 4–6 show mean values for the proportion of phenotypic variance attributable to additive genetic, shared-environment and nonshared-environment effects

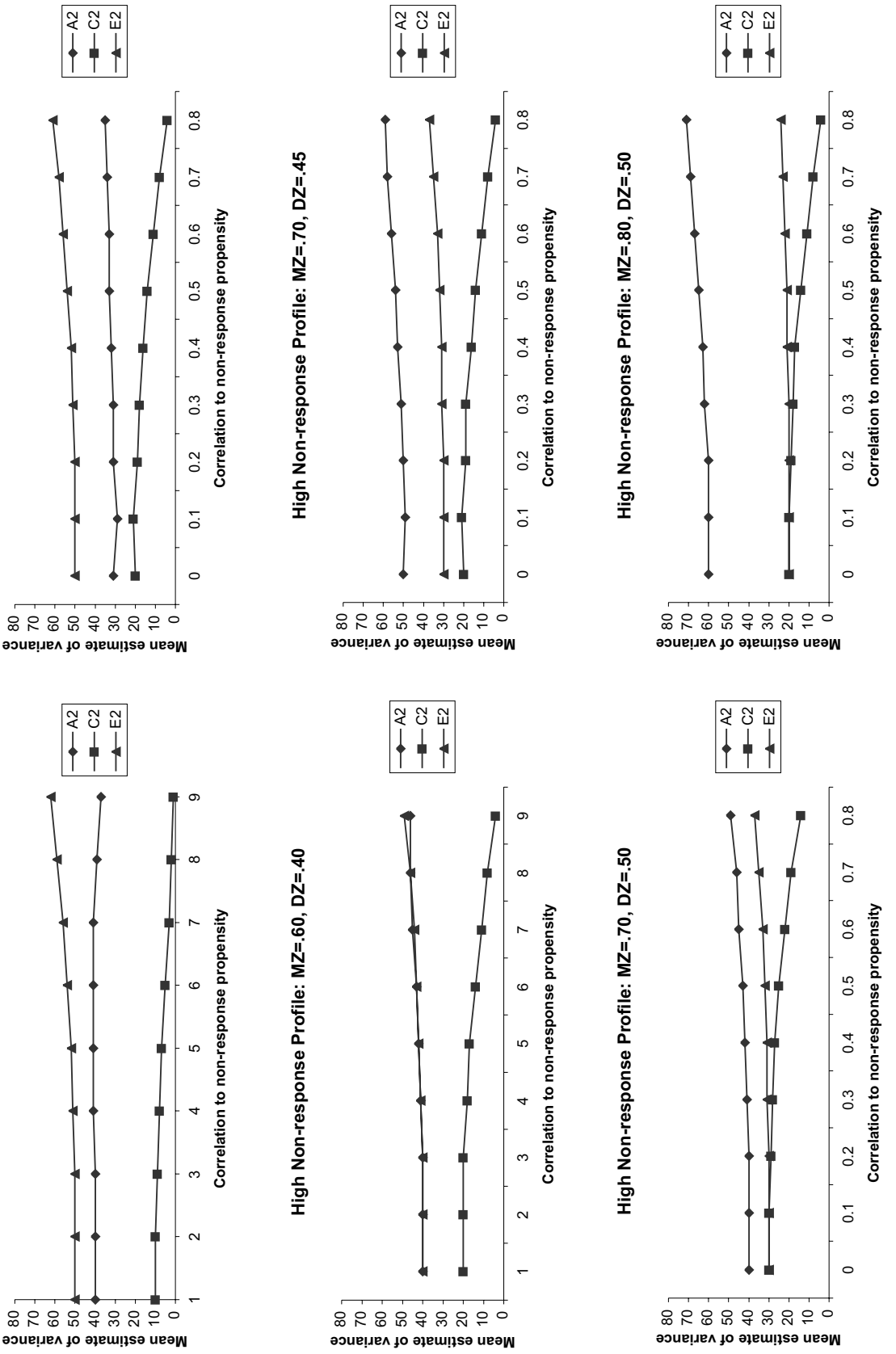

Figure 3
Effect of differential non-response probabilities on behavioural-genetic parameters: High non-response profile.

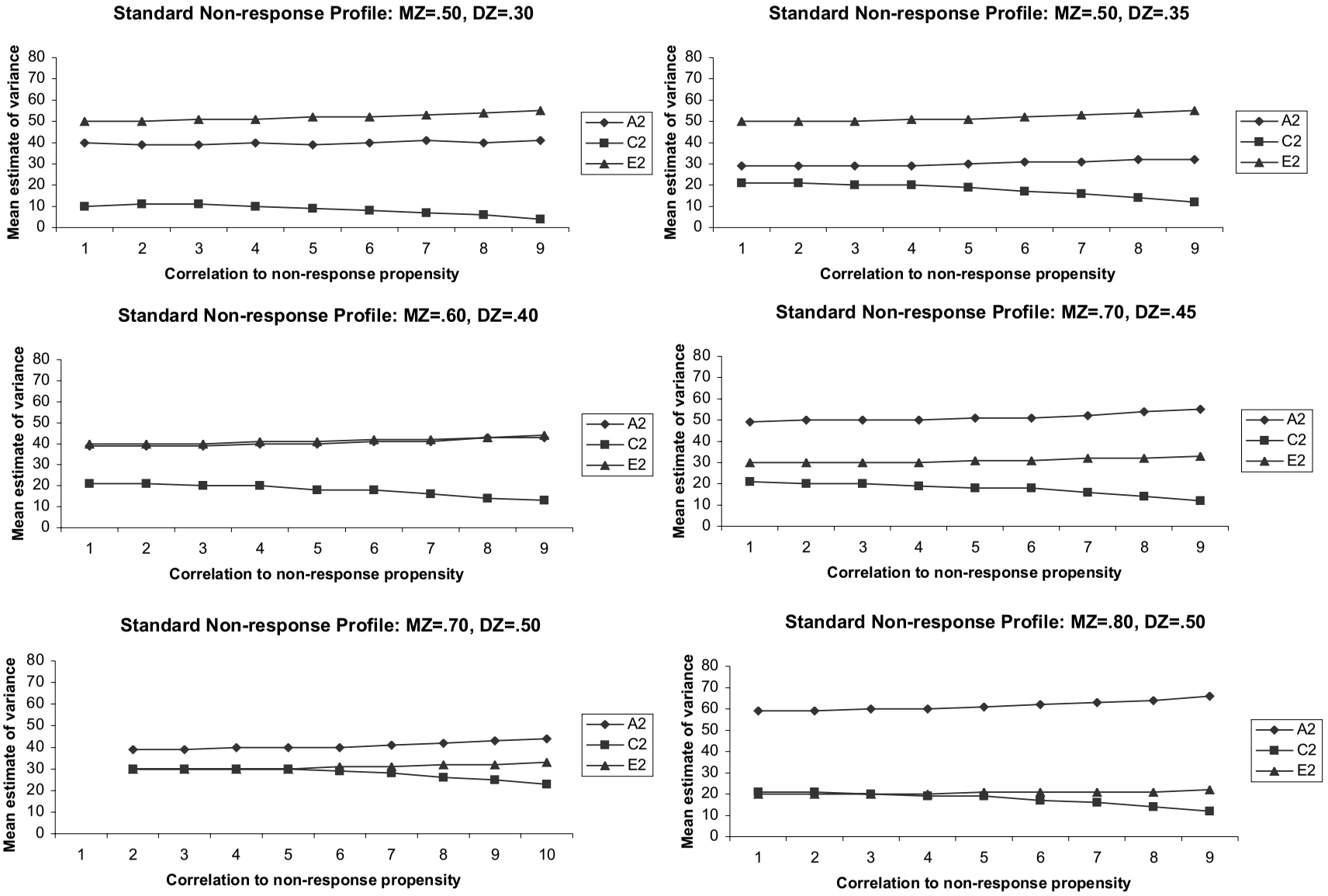

Figure 4

Effect of differential non-response probabilities on behavioural genetic parameters: Standard non-response profile.

Table 3

Effect of Differential Nonresponse Probabilities on Behavioral-genetic Parameters

	Response pattern applied											
	High nonresponse profile						Standard nonresponse profile					
Correlation to nonresponse propensity	Mean twin correlations		Mean ACE estimates (standardized)			% C2 estimated as zero	Mean twin correlations		Mean ACE estimates (standardized)			% C2 estimated as zero
	MZ	DZ	A2	C2	E2		MZ	DZ	A2	C2	E2	
.0	**.50**	**.30**	40	10	50	10.0	**.50**	**.30**	40	10	50	9.8
.1	.50	.30	40	10	50	10.8	.50	.30	39	11	50	10.0
.2	.50	.29	40	9	50	14.2	.50	.30	39	11	51	9.0
.3	.49	.28	41	8	51	18.2	.49	.29	40	10	51	16.4
.4	.48	.27	41	7	52	23.6	.49	.30	39	9	52	15.4
.5	.46	.25	41	5	54	37.8	.48	.28	40	8	52	19.2
.6	.44	.22	41	3	56	51.4	.47	.26	41	7	53	25.2
.7	.42	.19	39	2	59	70.6	.46	.25	40	6	54	33.4
.8	.39	.15	37	1	62	86.8	.45	.23	41	4	55	44.0
.0	**.50**	**.35**	31	20	50	0.0	**.50**	**.35**	29	21	50	0.6
.1	.50	.35	29	21	50	0.4	.50	.35	29	21	50	1.2
.2	.50	.34	31	19	50	1.0	.50	.35	29	20	50	1.6
.3	.49	.33	31	18	51	0.2	.49	.35	29	20	51	0.0
.4	.48	.32	32	16	52	5.6	.49	.34	30	19	51	2.6
.5	.46	.30	33	14	54	4.4	.48	.33	31	17	52	1.4
.6	.44	.28	33	11	56	8.4	.47	.32	31	16	53	2.8
.7	.42	.25	34	8	58	23.4	.46	.30	32	14	54	5.2
.8	.39	.20	35	4	61	41.8	.45	.28	32	12	55	8.4
.0	**.60**	**.40**	40	20	40	0.0	**.60**	**.40**	39	21	40	0.2
.1	.60	.40	40	20	40	0.0	.60	.40	39	21	40	0.4
.2	.60	.40	40	20	40	0.6	.60	.40	39	20	40	0.8
.3	.59	.39	41	18	41	1.4	.59	.40	40	20	41	0.6
.4	.58	.37	42	17	42	1.4	.59	.39	40	18	41	0.4
.5	.57	.36	43	14	43	3.4	.58	.38	41	18	42	0.6
.6	.56	.33	45	11	44	8.8	.58	.37	41	16	42	1.2
.7	.54	.30	46	8	46	23.4	.57	.36	43	14	43	3.6
.8	.51	.26	46	4	49	44.2	.56	.34	43	13	44	5.8
.0	**.70**	**.45**	50	20	30	0.0	**.70**	**.45**	49	21	30	0.2
.1	.70	.45	49	21	30	0.0	.70	.45	50	20	30	0.2
.2	.70	.45	50	19	30	0.0	.70	.45	50	20	30	0.4
.3	.69	.44	51	19	31	0.2	.70	.44	50	19	30	0.8
.4	.69	.42	53	16	31	1.4	.69	.44	51	18	31	0.2
.5	.68	.41	54	14	32	4.4	.69	.43	51	18	31	1.0
.6	.67	.38	56	11	33	9.4	.68	.42	52	16	32	1.4
.7	.65	.36	58	8	35	18.6	.68	.41	54	14	32	1.2
.8	.63	.33	59	4	37	40.8	.67	.40	55	12	33	4.2
.0	**.70**	**.50**	40	30	30	0.0	**.70**	**.50**	39	30	30	0.0
.1	.70	.50	40	30	30	0.0	.70	.50	39	30	30	0.0
.2	.70	.50	40	29	30	0.0	.70	.50	40	30	30	0.0
.3	.69	.49	41	28	31	0.0	.70	0.5	40	30	30	0.0
.4	.69	.48	42	27	31	0.0	.69	.49	40	29	31	0.0
.5	.68	.46	43	25	32	0.0	.69	.48	41	28	31	0.0
.6	.67	.44	45	22	33	0.0	.68	.47	42	26	32	0.0
.7	.65	.42	46	19	35	0.6	.68	.46	43	25	32	0.0
.8	.63	.39	49	14	37	3.6	.67	.45	44	23	33	0.0
.0	**.80**	**.50**	60	20	20	0.0	**.80**	**.50**	59	21	20	0.2
.1	.80	.50	60	20	20	0.0	.80	.50	59	21	20	0.0
.2	.80	.50	60	19	20	0.6	.80	.50	60	20	20	0.2
.3	.80	.49	62	18	20	0.0	.80	.50	60	19	20	0.0
.4	.79	.48	63	17	21	1.0	.80	.49	61	19	21	0.6
.5	.79	.46	65	14	21	1.6	.79	.48	62	17	21	0.6
.6	.78	.44	67	11	22	3.4	.79	.47	63	16	21	1.0
.7	.77	.42	69	8	23	15.4	.79	.46	64	14	21	1.6
.8	.76	.39	71	4	24	41.8	.78	.45	66	12	22	2.8

estimated by Mx. Column 7 shows the percentage of the 500 simulations where the value of the shared-environmental variance was estimated as zero. This pattern of results is repeated for the standard nonresponse profile in the left-hand panel of the table.

If zygosity is ignored and one looks at the overall pattern of attenuation for the 8 different correlations (.3, .35, .4, .45, .5, .6, .7 & .8) combined with the propensity for nonresponse (.0 to .8), Table 3 shows that:

- the lower the true correlation, the greater the attenuation (e.g., at a correlation of .8 with the nonresponse propensity variable a correlation of .5 reduces to .39 [see first left-hand panel], compared to a correlation of .8 which only reduces to .76 [see last left-hand panel])
- the higher the correlation of the phenotypic variables with the nonresponse propensity variable (ranging from .0 to .8), the greater the attenuation
- the less severe the selectivity of the response probabilities, the smaller the attenuation. (Compare the high nonresponse profile to the standard nonresponse profile.)

If zygosity is taken into account and the influence of nonresponse on genetic and environmental effects examined, Table 3 reveals two findings. First, under the simulated nonresponse profiles, estimates of the additive effects of genes tended to increase. This is more apparent the higher the correlation between the phenotype and the nonresponse-propensity variable, and for higher MZ correlations. For example, at the highest correlation of .8 with the propensity variable, and with MZ/DZ correlations of .7 and .45, the heritability was overestimated by 18% for the high-nonresponse profile (59% vs. 50%), and by 10% for the standard-nonresponse profile (55% vs. 50%). Similarly, estimates of the nonshared environment effect (which also includes measurement error) increased with higher correlations between the propensity variable and the phenotype and with more severe nonresponse profiles. This is to be expected as the magnitude of the MZ correlation decreases under these conditions and it is this which sets the value of the nonshared environment effect.

Second, Table 3 shows that under the simulated-nonresponse profiles, estimates of shared-environment effects tended to decrease. For example, for the high-nonresponse profile the estimated mean shared-environment effect was reduced by approximately one half of its true value when the phenotypic variables were correlated .6 with the propensity for nonresponse variable. For the standard nonresponse profile, this attenuation was reduced to approximately one-fifth of the true value. The magnitude of this effect was reduced the larger the size of the true shared-environment effect: compare the results for the simulation with the 30% shared-environment effect to that with a 10% effect. Thus, depending on the true value of the shared-environment effect, nonresponse attenuation resulted in a marked reduction in the ability to detect true shared-environment effects. This effect would be exaggerated in studies with sample sizes smaller than those simulated here, as the power to detect the shared-environment effect would decrease, making it more likely that the best-fitting model would not contain this effect. One implication of these findings is that for phenotypes where the true value of shared-environment variation is approximately 20%, studies may fail to detect this effect due to the influence of differential nonresponse alone.

The column indicating the percentage of simulations where the shared environment was estimated as zero provides some useful results on the variability to be expected in behavioral-genetic parameters over a number of replications of a study. For example, for simulated correlations of .5 and .3 for MZ and DZ twins where nonresponse was simulated using the high-nonresponse profile and the correlation to the response propensity was .1, over 10% of the simulations estimated the shared environment as zero. This indicates that even under negligible nonresponse with a true shared-environment effect of 10%, over 10% of studies will fail to establish the significance of this effect. This result holds even when the nonresponse pattern applied is much less severe, as can be seen from the equivalent results under the standard-nonresponse profile. This effect was accentuated as the correlation between the phenotype and the propensity variable increased. Even under the less-severe nonresponse profile, when the correlation between the phenotype and the response propensity was .8, just under half of the simulations estimated the shared-environment effect as zero. Under the high-nonresponse profile, this effect was maintained for DZ correlations up to .45 (which implies a shared-environment effect of 20% in these simulations) but disappeared when correlations implied a shared-environment effect of 30%. Under the less-severe nonresponse profile this effect was not marked for DZ correlations above .3, and by implication, for shared-environment effects of 20% or above. These findings underscore the earlier observation that variability in study results about shared-environment effects may be due to the consequences of moderate to severe nonresponse.

Study 1B: The Consequences of Nonparticipation on Genetic and Environmental Effects in Twin Studies

Study 1B investigated the effects on study estimates of population parameters when sections of a population, such as individuals or twin pairs, do not participate in a study at all. This is equivalent to portions of the population having a zero probability of responding to the study. We were therefore interested in the case where whole segments of the propensity-to-respond distribution do not appear in the final achieved sample

due to this truncation effect. For example, in a study of antisocial behavior in adults the sample would be truncated if all individuals whose phenotypic values were above a given threshold either refused to respond or were not included in the sampling frame. This could possibly occur by not including imprisoned individuals in the sampling frame. This type of nonresponse bias is the most severe as it excludes a whole segment of the population.

Method

Correlation Patterns

Differing patterns of MZ and DZ correlations were simulated so that the effect of truncation could be investigated over varying patterns of twin-pair correlations given in Table 1. S-Plus (*S-Plus 2000 user's guide*, 1999) was used to simulate standard normal variables for each twin in a pair so that the expected correlations were of the required form.

Propensity for Nonresponse

As in Study 1, in order to simulate the effects of nonresponse, a third variable was generated which can be viewed as indexing the propensity for nonresponse for a given twin pair. This variable was simulated to correlate at different levels with the simulated phenotypic variables, and was specified at values of .0 to .8 in increments of .1.

Nonresponse Patterns

To simulate truncation effects, 40th, 60th and 80th percentile values were used to truncate the propensity-variable distribution. This gave a response-rate equal to the percentile of the truncation. Martin and Wilson (1982) term this hard selection. Although their simulations were based on nonresponse being directly related to the phenotypic values rather than having an intervening nonresponse propensity variable, the addition of this step in the present analysis means that the effect of applying truncation to the simulations is to truncate the response-propensity distribution. The level of truncation of the phenotypic variables will therefore depend on the correlation between the nonresponse propensity and the phenotype. A situation equivalent to that investigated by Martin and Wilson (1982) would occur when the nonresponse propensity is exactly correlated with the phenotypic variables. Our model gives a more realistic model of the effect of truncation on population estimates as exclusion from the sample is unlikely to only be determined by phenotypic values.

Sample Size and Number of Simulations

A final sample size of 500 of each type of twin pair was also specified for the truncation simulations, which required the simulated number of pairs to be varied as a function of the truncation percentile. The actual values used were as follows: (1) with truncation at the 80th percentile, 625 twin-pairs of each type were simulated, (2) with truncation at the 60th percentile, 834 pairs of each type, and (3) with truncation at the 40th percentile, 1250 pairs of each type. For each combination of truncation level and twin-pair correlation, 500 simulations were carried out. The results are presented as mean values over all simulations.

In summary, each simulation consisted of the following steps: First, for each zygosity, standard normal variables were simulated for each twin in a pair to give a sample of the required size with variables for the phenotype of each twin and a variable for the nonresponse propensity. Over the whole sample, these three variables would then have an expected correlation matrix as required for each simulation. For example, where the nonresponse-propensity variable was correlated at .8 with the phenotypic variable and the MZ correlation was .5, three variables would be simulated for 1250 pairs (for truncation at the 40th percentile) to give an expected correlation matrix of .5 between the phenotypic variables and of .8 between these variables and the nonresponse-propensity variable. Second, this simulated sample was then reduced in size based on the level of truncation being applied. Third, the resulting matrices by zygosity were then output to Mx and a standard ACE model fitted. These three steps were then repeated 500 times for each simulation and the results loaded into an S-Plus data set to allow for the calculation of summary statistics.

Results

Table 4 gives the results for the truncation simulations. As per Table 3, these tables give the results in mean values over 500 simulations for the simulated twin-correlation patterns. For each specific pattern of correlations, the results are given for *truncation values* using the 40th, 60th and 80th percentiles of the nonresponse-propensity distribution. The row where the phenotypic variables have a zero correlation with the nonresponse-propensity variable gives a useful baseline for comparison as it shows the simulation results where the nonresponse-propensity variable in uncorrelated with the phenotypes. As in Table 3, the first column gives the correlation between the nonresponse-propensity variable and the phenotypes. For each segment of the table, the first two columns show the MZ and DZ correlations after nonresponse; columns 3–5 show mean values for the proportion of phenotypic variance attributable to additive genetic, shared-environment and nonshared-environment effects estimated by Mx, and column 6 shows the percentage of the 500 simulations where the value of the shared-environmental variance was estimated as zero.

Ignoring zygosity, and focusing on the 8 individual correlations values (.3, .35, .4, .45, .5, .6, .7 & .8), the table shows that first, the attenuating effect of truncation on twin correlations was larger the lower the true correlation. For example, for truncation at the 80th percentile and a correlation between the

propensity for nonresponse and the phenotype of .6, a correlation of .5 falls by 20% (to .41) while a correlation of .35 falls by 30% (to .24). Second, the impact of truncation increased with the severity of truncation. Third, the magnitude of these effects depended on the correlation between the phenotype and the nonresponse propensity variable, such that the higher the correlations with the propensity for nonresponse the larger the attenuating effect of truncation.

If zygosity is taken into account and the influence of truncation nonresponse on genetic and environmental effects examined, Table 4 reveals two findings. First, when the nonresponse variable is highly correlated with the phenotype, shared-environment effects disappear and genetic effects are overestimated. This situation is problematic when the true MZ/DZ correlations are high. For example, for truncation at the 80th percentile with an MZ/DZ correlation of .8 and .5, the true shared environment variance component of 20% was estimated as 1% when the phenotype was correlated .8 with the propensity for nonresponse. Second, when the true MZ/DZ correlations are low and the nonresponse propensity is highly correlated with the phenotype, the twin-correlation structure degraded. This degradation produced two interesting effects. First, in some instances, a true positive DZ correlation became negative. Second, a true MZ/DZ-correlation structure that pointed to additive genetic effects yielded genetic dominance, as the DZ correlation was less than half the MZ correlation (Plomin et al., 2001). For example, for truncation at the 40th percentile and MZ/DZ-correlation of .5 and .3, the correlations became .11 and –.25 when the phenotype was correlated at .8 with the nonresponse propensity variable. Negative DZ correlations and DZ correlations of less than half the MZ correlation have been noted in numerous behavioral-genetic studies of personality development (Plomin & Caspi, 1999). They are usually ascribed a substantive explanation (e.g., epistatic interaction) or attributed to measurement problems (e.g., contrast effects in comparing twins). The present results suggest that some of the peculiarities of behavioral-genetic findings, such as negative DZ correlations, may, in fact, be the result of the effect of nonresponse.

Study 2: The Effect of Nonresponse on Genetic and Environmental Effects in Twin Studies: Results from a 'Natural' Experiment

Study 2 investigated whether the effects of nonresponse found in the simulations of Study 1 are found in real data. For a nationally representative twin sample visited at home, a pilot mail survey was also carried out for which there was only a limited response. Descriptive statistics could then be compared and results modeled for the full sample and for those who only responded to the mail survey in a 'natural' experiment. As all families had completed a face-to-face interview when the twins were 5 years of age, we had a wide range of measures on which to compare the full sample with the subsample who also returned the mail survey (i.e., excluding nonresponders). It was expected that mean levels would be attenuated for variables correlated with the liability to respond to the mail survey, such as socioeconomic deprivation (Thornberry et al., 1993). Based on the simulations, estimates of the shared environment were expected to be attenuated, and where the size of the shared environment effect was negligible, nonresponse was expected to induce artifactual nonadditive genetic effects. Investigation of model-fitting results was made using the 'Aggression' and 'Delinquency' scales from the Child Behavior Checklist (CBCL; Achenbach, 1991a;1991b). These measures were chosen as twin and adoption studies of these scales report higher heritability for aggression (around 60%) than delinquency (around 30–40%), while the shared environment effect is significant only for the delinquency scale (also around 30–40%; e.g., Deater-Deckard & Plomin, 1999; Edelbrock et al., 1995; Eley et al., 1999). This differential pattern of expected behavioral-genetic estimates allowed the investigation of the two types of nonresponse effect found in the simulations.

Method

The E-Risk Study Sample

Participants are members of the Environmental Risk (E-Risk) Longitudinal Twin Study, which investigates how genetic and environmental factors shape children's development. The E-Risk sampling frame was two consecutive birth cohorts (1994 & 1995) in a birth register of twins born in England and Wales (Trouton et al., 2002). Of the 15,906 twin pairs born in these two years, 71% joined the register.

The E-Risk Study probability sample was drawn using a high-risk stratification sampling frame. High-risk families were those in which the mother had her first birth when she was 20 years of age or younger. We used this sampling (1) to replace high-risk families who were selectively lost to the register via nonresponse and (2) to ensure sufficient base-rates of problem behavior given the low base-rates expected for 5-year-old children. Age at first childbearing was used as the risk-stratification variable because it was present for virtually all families in the register, it is relatively free of measurement error, and early childbearing is a known risk factor for children's problem behaviors (Maynard, 1997; Moffitt & The E-Risk Study Team, 2002). The high-risk sampling strategy resulted in a final sample in which one-third of study mothers constitute a 160% oversample of mothers who were at high risk based on their young age at first birth (15–20 years), while the other two-thirds of study mothers accurately represent all mothers in the general population (aged 15–48) in England and Wales in 1994–95 (estimates derived from the General Household Survey; Bennett, et al., 1996). To

Table 4

Effect of Truncation on Behavioral Genetic Parameters

Correlation to nonresponse propensity	Mean twin correlations		Mean ACE estimates (standardized)			% C2 estimated as zero	Mean twin correlations		Mean ACE estimates (standardized)			% C2 estimated as zero	Mean twin correlations		Mean ACE estimates (standardized)			% C2 estimated as zero
	MZ	DZ	A2	C2	E2		MZ	DZ	A2	C2	E2		MZ	DZ	A2	C2	E2	
	Truncation value: 40th percentile						Truncation value: 60th percentile						Truncation value: 80th percentile					
.0	**.50**	**.30**	39	11	50	11.4	**.50**	**.30**	39	11	50	10.6	**.50**	**.30**	39	11	50	12.0
.1	.50	.30	40	10	50	13.4	.50	.30	40	10	50	12.6	.50	.30	39	10	50	14.0
.2	.48	.28	40	9	52	18.6	.49	.28	40	9	51	16.1	.49	.29	40	9	51	17.2
.3	.47	.25	41	6	53	33.6	.47	.26	40	7	53	27.2	.48	.27	40	8	52	22.2
.4	.44	.21	40	3	57	57.2	.45	.23	40	4	56	45.6	.50	.25	40	6	54	31.8
.5	.40	.15	38	1	62	83.0	.41	.18	39	2	60	70.8	.44	.22	40	4	56	51.2
.6	.34	.07	30	0	70	98.6	.37	.12	34	0	66	92.0	.41	.18	38	2	60	73.4
.7	.25	–.06	18	0	82	100.0	.30	.03	26	0	74	99.2	.40	.12	35	0	65	89.6
.8	.11	–.25	1	0	99	100.0	.20	–.11	12	0	88	100.0	.32	.05	28	0	72	98.8
	Truncation value: 40th percentile						Truncation value: 60th percentile						Truncation value: 80th percentile					
.0	**.50**	**.35**	30	20	50	1.0	**.50**	**.35**	30	20	50	1.0	**.50**	**.35**	30	20	50	0.8
.1	.49	.35	29	20	51	0.6	.50	.35	30	20	50	1.4	.50	.35	29	20	50	0.8
.2	.49	.33	31	18	51	1.6	.49	.34	30	18	51	1.2	.49	.34	29	19	51	0.8
.3	.47	.31	32	15	53	4.6	.47	.32	31	16	53	4.8	.48	.33	30	18	52	1.4
.4	.44	.27	33	10	56	11.2	.45	.28	33	12	55	8.8	.46	.31	31	15	54	5.0
.5	.40	.22	34	5	60	33.8	.42	.24	34	7	59	23.0	.44	.28	33	11	56	12.0
.6	.33	.13	31	1	68	77.0	.37	.18	33	3	64	57.6	.41	.24	33	8	59	26.2
.7	.24	.02	20	0	79	98.4	.30	.09	28	0	72	91.4	.37	.18	33	4	64	50.0
.8	.11	–.16	3	0	97	100.0	.21	–.03	15	0	85	99.6	.31	.11	29	1	70	77.8
	Truncation value: 40th percentile						Truncation value: 60th percentile						Truncation value: 80th percentile					
.0	**.60**	**.40**	40	20	40	0.8	**.60**	**.40**	40	20	40	0.6	**.60**	**.40**	40	20	40	0.6
.1	.60	.40	40	19	40	0.8	.60	.40	40	20	40	1.0	.60	.40	40	20	40	0.6
.2	.59	.38	41	18	41	1.2	.59	.39	41	18	41	1.0	.59	.39	41	19	41	1.0
.3	.57	.36	46	15	43	3.2	.58	.37	42	16	42	2.2	.58	.38	41	17	42	1.4
.4	.55	.32	45	10	45	10.8	.56	.34	44	12	44	4.6	.57	.36	42	15	43	3.2
.5	.52	.27	46	5	49	33.8	.53	.30	45	8	47	20.8	.55	.33	44	11	45	9.6
.6	.47	.20	45	1	54	77.4	.49	.24	46	3	51	53.4	.53	.29	45	7	47	22.2
.7	.40	.09	37	0	63	98.8	.44	.16	42	0	58	88.2	.50	.25	46	3	51	51.6
.8	.29	–.07	21	0	79	100.0	.36	.05	32	0	68	99.2	.45	.18	43	1	56	81.4

Table 4 (continued)

Effect of Truncation on Behavioral Genetic Parameters

Correlation to nonresponse propensity	Mean twin correlations		Mean ACE estimates (standardized)			% C2 estimated as zero	Mean twin correlations		Mean ACE estimates (standardized)			% C2 estimated as zero	Mean twin correlations		Mean ACE estimates (standardized)			% C2 estimated as zero
	MZ	DZ	A2	C2	E2		MZ	DZ	A2	C2	E2		MZ	DZ	A2	C2	E2	
	Truncation value: 40th percentile						Truncation value: 60th percentile						Truncation value: 80th percentile					
.0	**.70**	**.45**	50	20	30	0.8	**.70**	**.45**	50	20	30	0.4	**.70**	**.45**	50	20	30	0.4
.1	.70	.45	50	19	30	0.4	.70	.45	50	20	30	0.4	.70	.45	50	20	30	0.6
.2	.69	.43	52	17	31	0.8	.69	.44	51	18	31	0.8	.69	.44	51	19	31	0.8
.3	.68	.38	57	11	33	8.6	.68	.42	53	16	32	1.8	.69	.43	52	17	31	1.0
.4	.66	.38	56	10	34	9.0	.67	.39	55	12	33	5.6	.68	.41	53	14	32	2.4
.5	.64	.33	59	5	36	33.6	.65	.36	57	8	35	18.8	.67	.39	55	11	34	8.2
.6	.60	.27	59	1	40	78.6	.62	.31	59	3	38	53.2	.65	.35	57	7	35	21.4
.7	.55	.17	53	0	47	99.4	.58	.23	56	0	43	90.4	.62	.31	59	3	38	51.4
.8	.46	.02	40	0	60	100.0	.52	.13	49	0	51	99.8	.59	.25	57	1	42	84.8
	Truncation value: 40th percentile						Truncation value: 60th percentile						Truncation value: 80th percentile					
.0	**.70**	**.50**	40	30	30	0.0	**.70**	**.50**	40	30	30	0.0	**.70**	**.50**	40	30	30	0.0
.1	.70	.50	40	30	30	0.0	.70	.50	40	30	30	0.0	.70	.50	40	30	30	0.0
.2	.69	.48	41	28	31	0.0	.69	.48	41	28	31	0.4	.69	.49	41	29	31	0.0
.3	.68	.46	43	25	32	0.0	.68	.47	42	26	32	0.0	.69	.49	40	29	31	0.0
.4	.66	.44	45	21	34	0.0	.67	.45	44	23	33	0.2	.68	.47	43	25	32	0.0
.5	.64	.39	49	15	36	2.8	.65	.42	46	19	35	0.2	.67	.44	44	22	33	0.0
.6	.60	.34	52	8	40	16.8	.62	.37	50	12	38	5.2	.65	.41	47	18	35	1.4
.7	.55	.25	53	1	46	74.8	.58	.30	53	5	42	38.2	.62	.37	50	13	38	7.0
.8	.47	.11	43	0	57	100.0	.52	.21	51	1	49	88.0	.59	.32	53	6	41	29.6
	Truncation value: 40th percentile						Truncation value: 60th percentile						Truncation value: 80th percentile					
.0	**.80**	**.50**	60	20	20	0.8	**.80**	**.50**	60	20	20	0.4	**.80**	**.50**	60	20	20	0.2
.1	.80	.50	61	19	20	0.2	.80	.50	60	20	20	0.2	.80	.50	60	20	20	0.6
.2	.79	.49	62	18	21	0.4	.80	.49	61	18	20	0.4	.80	.49	61	19	20	0.4
.3	.79	.47	64	15	21	1.8	.79	.47	63	16	21	1.0	.79	.41	69	9	22	0.6
.4	.77	.44	67	10	23	7.8	.78	.45	66	12	22	4.0	.79	.47	64	15	21	2.2
.5	.76	.39	71	5	24	34.4	.77	.42	69	8	23	14.2	.78	.44	66	11	22	7.2
.6	.73	.34	72	1	27	81.4	.75	.37	72	3	25	84.2	.76	.41	69	7	24	18.6
.7	.70	.25	69	0	31	99.4	.72	.30	71	0	28	92.0	.75	.37	72	3	25	53.4
.8	.64	.11	61	0	39	100.0	.68	.21	67	0	33	100.0	.73	.32	72	1	28	86.6

provide unbiased statistical estimates from the whole sample that can be generalized to the population of British families with children born in the 1990s, the data reported in this article were corrected with weighting to represent the proportion of maternal ages in that population.

The study sought a sample size of 1100 families to allow for attrition in future years of the longitudinal study while retaining statistical power. An initial list of families who had same-sex twins was drawn from the register to target for home visits, with a 10% oversample to allow for nonparticipation. Of the 1203 families from the initial list who were eligible for inclusion, 1116 (93%) participated in home-visit assessments when the twins were 5 years old (forming the base sample for the study), 4% of families refused, and 3% were lost to tracing or could not be reached after many attempts. With parents' permission, questionnaires were posted to the children's teachers, and teachers returned questionnaires for 94% of cohort children. Written informed consent was obtained from mothers. The E-Risk Study has received ethical approval from the Maudsley Hospital Ethics Committee.

Zygosity was determined using a standard zygosity questionnaire, which has been shown to have 95% accuracy (Price et al., 2000). Ambiguous cases were zygosity-typed using DNA. The sample includes 56% monozygotic (MZ) and 44% dizygotic (DZ) twin pairs. Sex is evenly distributed within zygosity (49% male).

Postal Questionnaire Study

A short questionnaire was mailed to all study families before the start of the main age-5 interview in order to introduce the study, and to establish the efficacy of this low-cost mode of data collection for a high-risk sample. The response rate was very poor (43%, 477/1116), although acceptable compared to other mail surveys (Asch et al., 1997), and led to the abandonment of data-collection using this method from the E-Risk study families.

Childhood Measures

Children's antisocial behavior. Children's antisocial behavior at the age of 5 was assessed using the Achenbach family of instruments (Achenbach, 1991a; 1991b). The Aggressive (e.g., physically attacks people, has temper tantrums or hot temper) and Delinquent (e.g., lying or cheating, swearing or bad language) Antisocial Behavior Scales were derived by summing items from mother and teacher reports of the Child Behavior Checklist and the Teacher Report Form. Mother and teacher reports of aggression and delinquency correlated .28 ($p < .001$) and .20 ($p < .001$) respectively, which is typical of inter-rater agreement about behavioral problems (Achenbach et al., 1987). Scores for the Aggressive Scale ranged from 0 to 81 ($M = 14.24$, $SD = 10.84$) and for the Delinquent Scale scores ranged from 0 to 22 ($M = 2.60$, $SD = 2.66$). The internal consistency was .95 for the Aggressive Scale and .94 for the Delinquent Scale.

Family Characteristics

Maternal antisocial behavior. Mothers reported their own histories of antisocial behavior using the Young Adult Self Report (YASR; Achenbach, 1997), modified to obtain lifetime data. We report scores on the externalizing syndrome. Mothers rated each behavior as being (0) 'not true', (1) 'somewhat true', or (2) 'very often true'. Scores ranged from 0 to 60 ($M = 11.25$, $SD = 9.71$) and the internal consistency was .90.

Biological father's antisocial behavior. Mothers also reported the biological father's lifetime history of antisocial behavior, using the Young Adult Behavior Checklist (YABCL; Achenbach, 1997), modified to obtain lifetime data. We report scores on the externalizing syndrome as for mothers. Scores ranged from 0 to 88 ($M = 14.76$, $SD = 16.29$) and the internal consistency was .95. A methodological study of mother–father agreement attests to the reliability of these women's reports about men's problem behaviors; mothers' reports account for more than 75% of the variance in men's self-reports on these scales (Caspi et al., 2001).

Family SES disadvantage. An index was created summing binary indicators of seven aspects of SES disadvantage: 1) head of household has no educational qualifications; 2) head of household is employed in an unskilled occupation or is not in the labor force; 3) total household gross annual income is less than £10,000; 4) family receives at least one government benefit, excluding disability benefit; 5) family housing is government-subsidized; 6) family has no access to a vehicle, and 7) family lives in the poorest of six neighborhood categories (CACI Information Services, 1993) in an area dominated by government-subsidized housing, low incomes, high unemployment, and single-parent families. These 7 measures showed strong intercorrelations with a coefficient alpha of .79. Forty-five per cent of study families experienced at least one SES disadvantage.

Partnership violence. Adult domestic violence was assessed by inquiring about 12 acts of physical violence (pushed/grabbed/shoved; slapped; shook; threw an object; kicked/bit/hit with fist; hit with something; twisted arm; threw bodily; beat up; choked/strangled; threatened with knife/gun; used knife/gun). Mothers were asked about their own violence toward any partner and about any partners' violence toward them. Responses were *not true* (coded 0) or *true* (coded 2). Another response option (coded 1) was available for women who felt uncertain about their responses, but this was virtually unused. The measure represents the variety of acts of violence mothers experienced as both victims and perpetrators. Scores were summed (range 0–40; $M = 2.75$, $SD = 5.67$). The internal consistency reliability of the physical-abuse

scale was .89. Interpartner-agreement reliability for this measure is very high (latent correlation .77; Moffitt et al., 1997). Moreover, this scale is a strong predictor of which couples in the general population experience clinically significant violence, involving injury and intervention by official agencies (Moffitt et al., 2001), and high-scorers on this scale experience domestic violence that is more chronic (lasts for more months with more incidents per month) than low-scorers (Ehrensaft et al., in press).

Statistical Methods

Analysis of means was carried out using multiple regression. Regression results are based on the sandwich or Huber/White variance estimator (Rogers, 1993; Williams, 2000), a method available in STATA 8.0 (StataCorp, 2003) which adjusts estimated standard errors to account for the dependence in the data due to analyzing sets of twins and provides results that are robust to model assumptions (Lumley et al., 2002).

As with the simulations, the statistical package Mx (Neale et al., 1999) was used to conduct the quantitative genetic analyses. The ACE, ADE and AE models were fitted to the data. Because E includes measurement error, it is not usually dropped in univariate analyses. The AE model is nested within the full ACE and ADE model (i.e., AE models are identical to the ACE or ADE model with the exception of constraints placed on the submodel; for detailed explanations of the statistical methods that are applied to operationalize the logic behind behavior-genetic designs, see Plomin et al., 2001 & Carey, 2003). In comparing the fit of different models, four model-selection statistics were used. The first was the χ^2 goodness-of-fit statistic. Large values compared to model degrees of freedom indicate poor model-fit to the observed covariance structure. When two models are nested (i.e., identical with the exception of constraints placed on the submodel), the difference in fit between them can be evaluated with the χ^2 difference, using as its degrees of freedom the *df*-difference from the two models. When the χ^2 difference is not statistically significant the more parsimonious model is selected, as the test indicates that additional constraints do not decrease the model fit. The second model-selection statistic was Akaike's Information Criterion (AIC; Burnham & Anderson, 1998). When comparing two models, the model with the lowest AIC value is selected as the best-fitting model. The third model-selection statistic was the Bayesian Information Criterion (BIC), where increasingly negative values correspond to increasingly better-fitting models. In comparing two models, differences of BIC between 6 and 10 give strong evidence in favor of the model with the smaller value (Raftery, 1995). The fourth model-selection statistic was the Root Mean Square Error of Approximation (RMSEA), which is an index of the model discrepancy per degree of freedom, from the observed covariance structure (MacCallum et al., 1996). A RMSEA of less than or equal to .06 indicates a good-fitting model (Hu & Bentler, 1999). To ensure that any differences between the full E-Risk sample and the subsample responding to a mail survey were not due to differing samples sizes, the models were estimated with the sample sizes set in Mx to be identical to the full E-Risk sample (i.e., $N = 1116$).

Results

Effects of Nonresponse on Study Variables

The effects of nonresponse on mean levels of crucial study variables are given in Table 5. Three broad conclusions can be drawn from these results: first, nonresponse to the mail survey was significantly related to the mean level of variables indexing both parental and child antisocial behavior. Children and parents in nonresponding families were significantly more antisocial. Second, nonresponse to the mail survey was also significantly related to the mean level of variables indexing social deprivation and environmental stressors. Nonresponding families were more disadvantaged and their children were exposed to

Table 5

Mean Scores on Study Variables as a Function of Nonresponse to a Mailed Questionnaire

Measure	Response to mailed questionnaire: Yes (*N* = 477) Mean (*SD*)		Response to mailed questionnaire: No (*N* = 639) Mean (*SD*)		*t(df),p*	Effect size[1]
Child's aggressive behaviour	13.1	10.0	15.3	11.4	3.81 (1115), $p < .001$	0.21
Child's delinquent behaviour	2.3	2.4	2.9	2.9	3.82 (1115), $p < .001$	0.20
Biological mother's antisocial behaviour	9.8	8.2	12.6	10.7	4.93 (1112), $p < .001$	0.30
Biological father's antisocial behaviour	11.4	13.3	17.8	18.1	6.68 (1108), $p < .001$	0.40
Social disadvantage	0.9	1.4	1.5	1.9	6.38 (1115), $p < .001$	0.37
Partnership violence	2.0	4.6	3.5	6.4	4.70 (1094), $p < .001$	0.27

Note: [1]Effect size of the group differences were derived using the following formula:

$d = (M_r - M_{nr}) / sd$

where M_r is the mean for the group responding to the mail survey and M_{nr} is the mean for the nonrespondents and *sd* is the standard deviation taken over the whole sample. Following from Cohen (1992), $d = .2$ is a small effect size, $d = .5$ is a medium effect size and $d = .8$ is a large effect size.

more domestic violence. Third, the effect of nonresponse was most marked for measures proximal to the adults responsible for responding to the mail survey, and less so — albeit still significantly — for measures of child-behavior problems. For example, nonresponse had a moderate effect on the mean level of father's antisocial behavior (effect size = .4) but a small effect on the mean level of children's aggressive behavior (effect size = .2).

Effects of Nonresponse on Behavioural-genetic Parameters

Tables 6 and 7 give the model-fitting results for ACE models of children's aggression and delinquency, respectively. The top panel in each table contains the model-fitting results obtained when data were used from the full E-Risk sample. The bottom panel in each table contains the model-fitting results obtained when using only data from families that also responded to the mail survey. The best-fitting model in each panel is indicated by bold text.

In the full E-Risk sample the MZ-twin correlation for aggressive behavior was .73 and the DZ-twin correlation was .24. The ADE and the AE models fit the data well, with the BIC favoring the more parsimonious AE model, indicating that aggressive behavior was influenced by additive genetic factors and nonshared environmental factors. In the subsample with a mail survey response, the MZ twin correlation was .74, but the DZ twin correlation was .12. The ADE model was unequivocally the best fit to these data, indicating that aggressive behavior was influenced by nonshared environmental factors and also by nonadditive genetic factors.

In the full E-Risk sample, the MZ-twin correlation for delinquent behavior was .72 and the DZ twin correlation was .43. The ACE-model fit the data well, indicating that delinquent behavior was influenced by additive genetic factors and shared and nonshared environmental factors. In the subsample with a mail-survey response, the MZ-twin correlation was .76, but the DZ-twin correlation was .31. The AE model fit these data best, indicating that delinquent behavior was influenced by additive genetic factors and nonshared environmental factors only. In the subsample the 'true' estimate of the shared environment effect of 22% was no longer detectable.

These results are in line with both the simulations from Study 1 and prior research. They show that nonresponse can attenuate means and correlations and affect the conclusions of behavioral-genetic models by eliminating the effects of the shared environment and inducing artificial nonadditive genetic effects.

Discussion

This article has shown that nonresponse can have three effects on the results of behavioral-genetic studies of twins. First, nonresponse can result in failure to detect significant shared-environment effects. This finding is consistent with other research which shows that shared-environment effects are reduced when censored variables are analysed (van den Oord & Rowe, 1997). Second, nonresponse can lead to the spurious identification of nonadditive genetic effects which researchers may erroneously interpret as evidence of contrast effects, biased measurement or genetic dominance. Third, nonresponse can also inflate estimates of additive genetic and nonshared environment effects. These findings were found using data simulations and a 'natural' experiment from a large representative twin study. Additionally, data simulations indicate, as expected, that these errors are most marked in the most severe type of nonresponse due to nonparticipation (or truncation), but they are also found in milder cases of differential nonresponse.

An important implication of these findings is that nonresponse may contribute to the emergence of contradictory findings in etiological research. Researchers should therefore be aware of these dangers when designing, analysing and evaluating behavioral-genetic studies. It is also important to be able to evaluate a study for all three types of nonresponse. The rates of unit and item nonresponse in a study should be explicitly noted in research reports, so that the potential impact of these is easy to evaluate. It is much more difficult to evaluate the extent of noncoverage, although study designs that do not attempt to sample from all units in a defined population (e.g., volunteer samples) are common and are prone to this type of nonresponse.

Given the impact of nonresponse bias, studies should be designed to maximize participant response and to minimize the effects of differential nonresponse (e.g., Stouthamer-Loeber & Van Kammen, 1995). Where this is not possible, and differential nonresponse is unavoidable, it may be necessary to consider alternative research strategies. For example, one design that can obviate the effects of nonresponse is the use of a high-risk, stratified sample design which attempts to capture the full range of phenotypes and risk factors by oversampling subjects who are least likely to respond to a normal survey.

Methods that make adjustments for the effects of nonresponse should also be considered. One simple approach is to define weights that adjust the sample distribution of key variables to that of the population. The resulting weights can then be used to obtain appropriately adjusted model estimates (Heath et al., 1998; Kaplan & Ferguson, 1999), although this method is not applicable for adjusting for the effects of truncation. Mplus (Muthén & Muthén, 1998) now provides capabilities for appropriately analysing data obtained from complex samples, including the use of weights, although correct inference in the presence of weights has still to be implemented in most structural equation–modelling software. Imputation methods should also be considered, particularly for item nonresponse. These methods provide well-validated

Table 6

Univariate Estimates of Genetic and Environmental Contributions to Aggressive Behavior, as a Function of Nonresponse to a Mailed Questionnaire

Model	Variance components			Overall model fit						Model difference test		
	Genetic	Environmental										
		Shared	Nonshared									
	a2 (95%CI)	c2 or d2 (95%CI)	e2 (95%CI)	χ^2	*df*	*p*	AIC	RMSEA	BIC	$\Delta\chi^2$	Δdf	*p*
Aggressive behavior (*N* = 1116): *Full E-Risk sample*												
ACE	0.71 (0.65–0.75)	0.00 (0.00–0.06)	0.29 (0.25-0.32)	16.837	3	<.001	10.837	0.091	–4.216			
AE	**0.72 (0.68–0.75)**		**0.29 (0.25–0.32)**	**16.837**	**4**	**<.001**	**8.837**	**0.076**	**–11.23**	**0.000**	**1**	**———***
ADE	**0.33 (0.00–0.66)**	**0.39 (0.06–0.73)**	**0.28 (0.25–0.32)**	**11.286**	**3**	**<.010**	**5.286**	**0.070**	**–9.767**	**5.551**	**1**	**< .018**
Aggressive behavior (*N* = 1116): *Subsample responding to mail survey*												
ACE	0.72 (0.68–0.75)	0.00 (0.00–0.03)	0.28 (0.25–0.32)	42.582	3	<.001	36.582	0.135	21.529			
AE	0.72 (0.68–0.75)		0.28 (0.25–0.32)	42.582	4	<.001	34.582	0.114	14.512	0.000	1	———*
ADE	**0.00 (0.00–0.20)**	**0.72 (0.53–0.76)**	**0.27 (0.24–0.31)**	**13.778**	**3**	**<.001**	**7.778**	**0.076**	**–7.275**	**28.804**	**1**	**< .001**

Note: * *p* for the chi-square difference could not be calculated because its value is 0.
AIC — Akaike's Information Criterion; RMSEA — Root Mean Square Error of Approximation ; BIC — Bayesian Information Criterion.
N refers to the number of twin pairs with data as input into Mx. The *N* for the subsample responding to the mail survey was increased from its true value to be identical to the full E-Risk sample so that differences in results would not be due to differences in sample size.
Model Comparison $\Delta\chi^2$: ACE vs AE and ADE vs AE.
Bold indicates the best fitting model.

Table 7

Univariate Estimates of Genetic and Environmental Contributions to Delinquent Behavior, as a Function of Nonresponse to a Mailed Questionnaire

Model	Variance components			Overall model fit						Model difference test		
	Genetic	Environmental										
		Shared	Nonshared									
	a2 (95%CI)	c2 or d2 (95%CI)	e2 (95%CI)	χ^2	*df*	*p*	AIC	RMSEA	BIC	$\Delta\chi^2$	Δdf	*p*
Delinquent Behavior (*N* = 1116): *Full E-Risk sample*												
ACE	**0.48 (0.34–0.64)**	**0.22 (0.07–0.35a)**	**0.30 (0.27–0.34)**	**5.836**	**3**	**< .120**	**–0.164**	**0.041**	**–15.22**			
AE	0.70 (0.67–0.74)		0.30 (0.26–0.33)	13.619	4	< .009	5.619	0.065	–14.45	7.783	1	< .005
ADE	0.70 (0.60–0.74)	0.00 (0.00–0.11)	0.30 (0.26–0.33)	13.619	3	< .001	7.619	0.079	–7.434			
Delinquent Behavior (*N* = 1116): *Subsample responding to mail survey*												
ACE	0.74 (0.64–0.77)	0.00 (0.00–0.10)	0.26 (0.23–0.29)	11.885	3	< .008	5.885	0.073	–9.168			
AE	**0.74 (0.71–0.77)**		**0.26 (0.23–0.29)**	**11.885**	**4**	**< .018**	**3.885**	**0.060**	**–16.19**	**0.000**	**1**	**———***
ADE	0.59 (0.24–0.77)	0.15 (0.00–0.50)	0.26 (0.23–0.29)	11.038	3	< .012	5.038	0.070	–10.02	0.847	1	.357

Note: * *p* for the chi-square difference could not be calculated because its value is 0.
AIC — Akaike's Information Criterion; RMSEA — Root Mean Square Error of Approximation ; BIC — Bayesian Information Criterion.
N refers to the number of twin pairs with as input into Mx. The *N* for the subsample responding to the mail survey was increased from its true value to be identical to the full E-Risk sample so that differences in results would not be due to differences in sample size.
Model Comparisons $\Delta\chi^2$: ACE vs AE and ADE vs AE.
Bold indicates the best fitting model.

strategies to 'fill in' missing values on specific variables (Brick & Kalton, 1996; Little, 1988). In recent years there have been many new developments to aid the applied researcher in dealing with missing-data issues such as multiple-imputation and full-information maximum-likelihood methods (Schafer & Graham, 2002), which are now becoming more easily available in statistical-analysis software. A further approach for detecting and adjusting for volunteer bias from twin pairs discordant for participation has also been developed by Neale and Eaves (1993).

A primary assumption in this paper should be noted here: if the phenotype of interest is not correlated with the propensity of an individual to respond to a study, there will be no effect of nonresponse on study estimates. This will be the case even at excessive levels of nonresponse. This also implies that a low level of study response is not necessarily indicative of the presence of nonresponse bias. For example, Gerrits et al. (2001) found no consistent evidence of nonresponse bias in their studies even though some of their nonresponse rates were of the order of 50%. It is therefore important to determine whether nonresponse is likely to be correlated with the variables of interest. Previous work on nonresponse (e.g., Groves, 1989) has indicated that characteristics such as low income and low education are related to lower response rates. Of particular note to behavioral-genetic studies is that these characteristics are highly correlated with certain phenotypes, such as intelligence, personality traits and psychiatric disorder. It is likely that many behavioral-genetic studies are susceptible to the effects of differential nonresponse, although more research is needed to determine the actual correlation between the propensity for nonresponse and the phenotypes of interest in behavioral-genetic research (Heath et al., 2001). In addition, the present article has examined the effects of nonresponse in relation to standard univariate twin models. Although these findings raise serious concerns, their generality needs to be evaluated in relation to more complex behavioral-genetic analyses (e.g., to multivariate analyses) and to extended twin designs (e.g., to twin-family and to 'twins-on-top' designs [Jacob et al., 2001]).

The purpose of this paper is to provide a clarion call to greater attention to sampling considerations in the behavioral sciences. More research resources need to be directed to improve sample recruitment and retention, and funding agencies should be alerted to this need. It must be emphasized that the effects of nonresponse are not specific to behavioral-genetic studies, but as the tools of quantitative genetics research are increasingly more widely used in etiological studies, greater attention to sampling will prove indispensable to generating accurate findings.

Acknowledgments

This research was supported by the Medical Research Council. I wish to thank Avshalom Caspi, Michael Rutter, Andrew Pickles, Terrie Moffitt, Louise Arseneault, Sara Jaffee and Robert Krueger for their helpful comments. I am grateful to the study mothers and fathers, the twins, and the twins' teachers for their participation. Thanks to Robert Plomin, to Thomas Achenbach for kind permission to adapt the CBCL, to Hallmark Cards, and to members of the E-Risk team for their dedication, hard work and insights. The E-Risk Study is funded by the Medical Research Council (Grant code: UK-MRC Grant G9806489).

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1997). *Manual for the Young Adult Self-Report and Young Adult Behavior Checklist*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213–232.

Asch, D. A., Jedrziewski, M. K., & Christakis, N. A. (1997). Response rates to mail surveys published in medical journals. *Journal of Clinical Epidemiology, 50*, 1129–1136.

Bennett, N., Jarvis, L., Rowlands, O., Singleton, N., & Haselden, L. (1996). *Living in Britain: Results from the general household survey*. London: HMSO.

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review, 48*, 386–398.

Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research, 5*, 215–238.

Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference: A practical information theoretic approach*. New York: Springer-Verlag.

CACI Information Services (1993). *ACORN user guide*. London: CACI.

Carey, G. (2003). *Human genetics for the social sciences*. Thousand Oaks, CA: Sage.

Caspi, A., Taylor, A., Jackson, J., Tagami, S., & Moffitt, T. E. (2001). Can women provide reliable information about their children's fathers? Cross-informant agreement about men's antisocial behavior. *Journal of Child Psychology & Psychiatry, 42*, 915–920.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

Davis, J. A., & Smith, T. W. (1992). *The NORC general social survey: A user's guide*. Newbury Park, CA: Sage.

Deater-Deckard, K., & Plomin, R. (1999). An adoption study of the etiology of teacher and parent reports of externalising behavior problems in middle childhood. *Child Development, 70*, 144–154.

Edelbrock, C., Rende, R., Plomin, R., & Thompson, L. A. (1995). A twin study of competence and problem behavior in childhood and early adolescence. *Journal of Child Psychology and Psychiatry, 36*, 775–785.

Ehrensaft, M. K., Moffitt, T. E., & Caspi, A. (in press). Clinically abusive relationships and their developmental antecedents in an unselected birth cohort. *Journal of Abnormal Psychology.*

Eley, T. C., Lichtenstein, P., & Stevenson, J. (1999). Sex differences in the etiology of aggressive and nonaggressive antisocial behavior: Results from two twin studies. *Child Development*, *70*, 155–168.

General Social Survey. (2001). *Sampling design & weighting*. Retrieved August 29, 2001, from http://www.icpsr.umich.edu/GSS//rnd1998/appendix/apdx_a.htm

Gerrits, M. H., van den Oord, E. J. C. G., & Voogt, R. (2001). An evaluation of nonresponse bias in peer, self, and teacher ratings of children's psychosocial adjustment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*, 593–602.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley.

Heath, A. C., Howells, W., Kirk, K. M., Madden, P. A. F., Bucholz, K. K., Nelson, E. C., Slutske, W.S., Statham, D.J., & Martin, N.G. (2001). Predictors of nonresponse to a questionnaire survey of a volunteer twin panel: Findings from the Australian 1989 twin cohort. *Twin Research, 4*, 73–80.

Heath, A. C., Madden, P. A. F., & Martin, G. (1998). Assessing the effects of cooperation bias and attrition in behavioral-genetic research using data-weighting. *Behavior Genetics, 28*, 415–427.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling, 6*, 1–55.

Jacob, T., Sher, K. J., Bucholz, K. K., True, W. T., Sirevaag, E. J., Rohrbaugh, J., Nelson, E., Neuman, R. J., Todd, R. D., Slutske, W. S., Whitfield, J. B., Kirk, K. M., Martin, N. G., Madden, P. A., & Heath, A. C. (2001). An integrative approach for studying the etiology of alcoholism and other addictions. *Twin Research, 4*, 103–118.

Kaplan, D., & Ferguson, A. J. (1999). On the utilization of sample weights in latent variable models. *Structural Equation Modelling, 6*, 305–321.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric Illness. *Behavior Genetics, 23*, 21–28.

Levy, P. S., & Lemeshow, S. (1991). *Sampling of populations: Methods and applications*. New York: John Wiley.

Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics, 6*, 287–297.

Loehlin, J. C. (1989). *Genes and environment in personality development*. London: Sage.

Lumley, Y., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health, 23*, 151–169.

Lykken, D. T., McGue, M., & Tellegen, A. (1987). Recruitment bias in twin research: The rule of two-thirds reconsidered. *Behavior Genetics, 17*, 343–362.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

Martin, N., Boomsma, D., & Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genetics, 17*, 387–392.

Martin, N. G., & Wilson, S. R. (1982). Bias in the estimation of heritability from truncated samples of twins. *Behavior Genetics, 12*, 467–472.

Maynard, R. A. (1997). *Kids having kids: Economic costs and social consequences of teen pregnancy*. Washington DC: Urban Institute Press.

Mirrlees-Black, C., Budd, T., Partridge, S., & Mayhew, P. (1998). *The 1998 British crime survey, England and Wales, Home Office statistical bulletin, 21/98*. London: Home Office.

Moffitt, T. E., Caspi, A., Krueger, R. F., Magdol, L., Margolin, G., Silva, P. A., & Sydney, R. (1997). Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner agreement. *Psychological Assessment*, *9*, 47–56.

Moffitt, T. E., Robins, R. W., & Caspi, A. (2001). A couples analysis of partner abuse with implications for abuse prevention. *Criminology and Public Policy, 1*, 5–36.

Moffitt, T. E., & The E-Risk Study Team. (2002). Teenaged mothers in contemporary Britain. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 43*, 727–742.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (1999). *Mx: Statistical modelling*. (5th ed.). Richmond, VA: Department of Psychiatry.

Neale, M. C., & Eaves, L. J. (1993). Estimating and controlling for the effects of volunteer bias with pairs of relatives. *Behavior Genetics, 23*, 271–277.

Neale, M. C., Eaves, L. J., Kendler, K. S., & Hewitt, J. K. (1989). Bias in correlations from selected samples of relatives: The effects of soft selection. *Behavior Genetics*, *19*, 163–169.

Nigg, J. T., & Goldsmith, H. H. (1994). Genetics of personality disorders: Perspectives from personality and psychopathology research. *Psychological Bulletin, 115*, 346–380.

Plomin, R., & Caspi, A. (1999). Behavioral genetics and personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 251–276). New York: Guilford Press.

Plomin, R., & Crabbe, J. (2001). DNA. *Psychological Bulletin, 126*, 802–828.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral Genetics* (4th ed.). New York: Worth.

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3*, 129–133.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Oxford: Blackwells.

Rhee, S. H., & Waldman, I. D. (2002). Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. *Psychological Bulletin*, *128*, 490–529.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin, 13,* 19–23. (Reprinted in *Stata Technical Bulletin Reprints, 3,* 88–94).

Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject*. New York: John Wiley.

Rutter, M., Pickles, A., Murray, R., & Eaves, L. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin, 127*, 291–324.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype environment effects. *Child Development, 54*, 424–435.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Skinner, C. J., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex samples*. New York: John Wiley.

StataCorp. (2003). Stata Statistical Software (Release 8.0) [Computer software]. College Station, TX: Stat Corporation.

S-Plus 2000 user's guide (1999). Seattle, WA: MathSoft.

Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior-genetic adoption studies. *Psychological Bulletin, 125*, 392–409.

Stouthamer-Loeber, M., & Van Kammen, W. (1995). *Data collection and management: A practical guide*. Thousand Oaks, CA: Sage.

Thornberry, T. P., Bjerregaard, B., & Miles, W. (1993). The consequences of respondent attrition in panel studies: A simulation based on the Rochester Youth Development Study. *Journal of Quantitative Criminology, 9*, 127–158.

Trouton, A., Spinath, F. M., & Plomin, R. (2002). Twins Early Development Study (TEDS): A Multivariate, longitudinal genetic investigation of language, cognition, and behavior problems in childhood. *Twin Research*, *5*, 444–448.

van den Oord, E. J. C. G., & Rowe, D. C. (1997). Effects of censored variables on family studies. *Behavior Genetics, 27*, 99–112.

Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics, 56,* 645–646.