Behavior Genetics, Vol. 28, No. 6, 1998

Assessing the Effects of Cooperation Bias and Attrition in Behavioral Genetic Research Using Data-Weighting

Andrew C. Heath,[1,3] Pamela A. F. Madden,[1] and Nicholas G. Martin[2]

Received 2 Feb. 1997—Final 25 Sept. 1998

Because twins and adoptees are a rare resource, they are often studied repeatedly over a period of many years. Differential attrition, and in some studies initial cooperation bias, have the potential to lead to serious biases to estimates of genetic and environmental parameters. Since non-response is often influenced by multiple binary or categorical sociodemographic variables, maximum-likelihood methods are not easily adapted to adjust for such effects. In this brief note we illustrate the use of data-weighting to assess the likely effects of cooperation bias or attrition both on measures of mean or prevalence, and on twin pair correlations or concordances, using data from the Australian twin panel 1981 survey and alcohol challenge studies. Participants in the alcohol challenge study were on average younger, more socially nonconforming, heavier drinkers, more likely to be unmarried, and less likely to report their religion as Other Protestant. Reweighting the alcohol challenge sample to have the same distribution on these variables as the Australian twin panel 1981 survey respondents confirmed that individuals who would feel very intoxicated after a challenge dose of alcohol were underrepresented in the study. However, pairwise data-weighting indicated that this cooperation bias was leading to only a slight underestimation of the importance of genetic effects on subjective intoxication.

KEY WORDS: Sampling bias; nonresponse; sampling weights; twin studies.

INTRODUCTION

The potential for biases to estimates of genetic and environmental parameters to occur in behavioral genetic research through differential cooperation or non-response is considerable, and has been widely acknowledged by behavioral geneticists (e.g., Martin and Wilson, 1982; Neale *et al.*, 1989; Neale and Eaves, 1993). Some twin panels such as the Australian Twin Panel (Jardine and Martin, 1984; Heath *et al.*, 1997) or the Maudsley Volunteer Twin Panel (Eaves *et al.* 1989) were formed initially from volunteers recruited through media appeals. Studies of adoptees often begin with indirect approaches to adoptive families by adoption agencies acting on behalf of researchers, with less than optimal response rates (e.g., Scarr *et al.*, 1981). Some twin panels in the U.S. (e.g., Kendler *et al.*, 1992; Hrubec and Neel, 1978; Eisen *et al.*, 1987) and in Scandinavia (e.g., Cederlof *et al.*, 1971; Medlund *et al.*, 1977; Kaprio *et al.*, 1978; Magnus *et al.*, 1983; Holm *et al.*, 1980) have been formed systematically from birth or other official records. In all of these cases, however, because panels of twins or adoptees constitute a rare resource, subjects are assessed repeatedly. Typically researchers

[1] Department of Psychiatry, Washington University School of Medicine, 4940 Children's Place, St. Louis, Missouri 63110.
[2] Division of Epidemiology and Population Health, Queensland Institute of Medical Research, Herston, Brisbane, Australia.
[3] To whom correspondence should be addressed. e-mail: andrew@matlockwustl.edu.

0001-8244/98/1100-0415$15.00/0

have assumed that the effects of cooperation bias or attrition on their data will be small, without presenting empirical data to support this assumption.

In the field of survey research, an extensive literature has developed on correcting for non-response in cross-sectional and longitudinal surveys, using data-weighting, missing data imputation, and methods of data-analysis that correctly handle certain patterns of missing data (Rubin, 1987; Little and Rubin, 1987; Little and Schenker, 1995). Data-analytic approaches have been developed which emphasize the importance of developing and testing models for predicting non-response, using both external comparisons to census data and internal comparisons based on data collected in prior assessments of the same target sample. In the case where relevant variables are all continuous, and non-response is a function of trait values, approaches based on likelihood methods are readily adapted to longitudinal behavioral genetic research (e.g., Neale, 1998). Often, however, the primary variables of interest are binary rather than continuous (e.g., psychiatric disorders, childhood sexual abuse), and major predictors of non-response are categorical sociodemographic variables such as educational level. Since observations on family members are correlated, methods of adjustment for selective non-response cannot always be immediately adapted from standard survey research methods. At the same time, since there will be cases where only some members of a family do not respond, family-structured data have great potential power for obtaining insights about the determinants of non-response.

In this paper, we use data from alcohol-related studies on the Australian National Health and Medical Research Council (NH&MRC) twin panel (Jardine and Martin, 1984; Martin *et al.*, 1985a,b) to explore the use of sampling weights, generated using logistic regression models to predict study participation (cf. Rosenbaum and Rubin, 1983), to evaluate likely biases to estimates of genetic and environmental parameters. During the period 1980–82 some 8183 twins from the Australian twin panel (including 3808 complete pairs) responded to a mailed questionnaire survey. Over the period 1979–81, a much smaller laboratory-based study was conducted using 206 adult twin pairs, who were administered a standard body-weight adjusted dose of alcohol. It is not unreasonable to anticipate that volunteers for such research would not constitute a random sample of the population, an effect that we explore here.

METHODS

Sample

Twin pairs were recruited for the Australian NH&MRC twin panel by extensive appeals through the Australian media, supplemented by systematic recruitment through schools. We focus here on twins from the '1981' cohort, i.e., those pairs from the panel who were aged at least 18 in the period 1979–82 when the first research studies were conducted. Pairs became members of the panel if at least one twin expressed interest in participating in research by mailing back a brief zygosity questionnaire. A small sample of 206 twin pairs participated in an alcohol challenge study (Martin *et al.*, 1985a, b; Heath and Martin, 1992), in which, after completion of baseline questionnaires and measures of psychomotor coordination, subjects were administered a body weight adjusted dose of alcohol (0.75 g/kg body-weight). A much larger sample of 3808 complete twin pairs and 567 single twins (including 132 pairs and 16 single twins from the alcohol challenge sample) responded to a mailed questionnaire survey (1981 survey), conducted in 1980–82, that included assessments of drinking habits, personality [Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975)], attitudes and various sociodemographic and health-related measures such as education, religion, marital status, smoking history, height and weight (Jardine and Martin, 1984; Heath *et al.*, 1989).

We consider here only twin pairs born 1944–1963, the birth years from which alcohol challenge study participants were recruited. For the alcohol challenge study, there were a total of 206 complete pairs [45 monozygotic female (MZF), 42 dizygotic female (DZF), 43 MZM, 37 DZM, and 39 unlike-sex pairs]. This breakdown by zygosity differs from that reported in earlier papers, as a result of the discovery of three cases of inconsistent zygosity assignment between the 1981 mailed questionnaire survey and the alcohol challenge study. When new DNA samples were obtained from these pairs, and genetic markers used to confirm zygosity, in every case it was discovered that lab error had been responsible for the inconsistency with the questionnaire-based zygosity assignment. Respondents to the 1981 survey born 1944–1963 (counting singleton twins as well as twins from complete pairs) included 1532 female MZ twins, 995 female same-sex DZ twins, and 722 female twins from unlike-sex pairs; as well as 773 male MZ twins, 553 male DZ twins and 657 male twins from unlike-sex pairs. For those women and men who participated in both the alcohol challenge study and the 1981 survey, there was an average interval between studies of 9 months and 7.3 months respectively. In what follows we shall assume that the entire sample of twins (including those who participated in only one study) together are representative of adult twins on the twin panel, and constitute the 'baseline sample'.

Assessments

Sociodemographic, personality, and life-style variables were taken from the 1981 survey responses or, for alcohol challenge respondents who did not respond to that survey, from the alcohol challenge baseline questionnaire. Educational level was collapsed to a 4-point scale: (i) early school leavers with no further educational training or apprenticeship; (ii) at least some high school, a diploma, trade certificate or apprenticeship, (iii) technical or teacher's college; or (iv) university education or higher. Religious affiliation was collapsed to a 5-point scale (Church of England, Other Protestant, Catholic, None, or Other/Not given—with the latter category including a small number of Jewish and Greek Orthodox respondents). Church attendance was analyzed as a dichotomous variable (weekly or more often versus less often than that). Church attendance was not assessed in the alcohol challenge questionnaire and therefore was not included in the final logistic regression equations used to correct for differences between volunteers for the alcohol challenge study and the remaining twins of the same birth cohort. Marital status was classified as (i) single or (ii) other, since there were few separated, divorced or remarried participants in the alcohol challenge sample.

The personality assessments comprised all four scales of the full Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975): (i) Extraversion (E), predominantly a measure of sociability and liveliness, which excluded impulsivity items from the earlier Eysenck Personality Inventory; Neuroticism (N), which assessed both anxious and depressive traits; Social Non-conformity (termed the 'Lie' scale by Eysenck), a scale that was designed to detect respondents who were faking good, but in fact appears to assess prosocial versus mild antisocial traits, with items such as "Would you dodge paying taxes if you were sure you could never be found out?" "Have you ever said anything bad or nasty about anyone?"; and Toughmindedness (termed 'Psychoticism' by Eysenck: P), a scale that is factorially and etiologically complex (Heath and Martin, 1990), and includes measures of impulsiveness, risk-taking, and indifference to suffering. The personality scores exhibited good stability across the two assessment situations, with test-retest correlations in women and men who participated in both the alcohol challenge study and the 1981 questionnaire survey of 0.85 and 0.89 for Extraversion, 0.82 and 0.80 for Neuroticism, 0.75 and 0.74 for Social Nonconformity, and 0.73 and 0.69 for Toughmindedness.

Life-style variables used in analyses presented here were measures of smoking history and alcohol consumption. Smoking was defined as a binary variable, whether or not the respondent reported ever smoking cigarettes ('smoking initiation'). Alcohol consumption measures were (i) log transformed average weekly consumption in standard drinks (test-retest correlations of 0.70 in women and 0.81 in men); (ii) reported average number of drinks per drinking occasion, redefined as a 3-level variable (1–2, 3–5, 6 or more); and (iii) frequency of alcohol use (daily, or nearly every day; from once or twice a month to once or twice a week; or less often).

From the alcohol challenge study, we used a single experimental variable: the subject's rating on a 10-point scale of how drunk he or she felt (where 10 indicated the most drunk they had ever been).

Data Analysis

We consider two approaches to the use of data-weighting, firstly for the analysis of twins as individuals, and later the analysis of twin pair data. For analyses of the sample as individuals, i.e., ignoring the twin structure of the sample, we used 'response propensity' analysis to generate individual sampling weights (Rosenbaum and Rubin, 1983; Little and Rubin, 1987). We first analyzed simulated data, where probability of response at follow-up was modeled as a logistic function of three correlated binary variables ('religion,' 'education,' and 'alcoholism'), to ensure that we could indeed reweight the follow-up data to have the same probability distribution as the baseline data. The same procedures were then applied to the analysis of participation in the alcohol challenge sample. Multiple logistic regression was used to identify predictors of whether or not individuals in the baseline sample participated in the alcohol challenge study, separately for males and females. Dummy variables were used to code different levels of sociodemographic variables and categorical lifestyle variables, including interaction effects where necessary. For continuous variables, partial odds ratios were computed for a change in score equal to the interquartile range for the variable (Hosmer and Lemeshow, 1989). In the logistic regression analysis, we did not attempt to adjust for the non-independence of observations on twin pairs. This is a conservative approach, which may have caused us to include unnecessary predictors (since it will cause us to underestimate the standard errors of regression coefficients), but would not have led to the omission of important variables.

From the logistic regression equation, the predicted probability for each subject of participating in the follow-up study, p_i, was computed. For participants in the challenge study, the reciprocal of this number, $x_i = 1/p_i$, was computed: this is the so-called expansion weight, i.e., the weight that would be needed to reweight the alcohol challenge sample back to the frequency distribution of the original sample; and then a sampling weight (or relative weight) was computed as $x_i/\bar{x}$, where $\bar{x}$ is the mean expansion weight for the alcohol challenge sample. The sampling weight was used to reweight the alcohol challenge sample to have (approximately) the same probability distribution on the observed variables as the original sample, while retaining the overall sample size of the alcohol challenge sample. These are standard procedures in the generation of sampling weights in survey research (Lee *et al.*, 1989). Sampling weights were computed separately for males and females.

This approach, which we shall refer to as individual data-weighting, could not be used directly to examine the impact of attrition on measures of twin pair or family resemblance or estimates of genetic and environmental parameters. Since observations on twin pairs are not statistically independent, simply taking the product of individual weights would over-correct the data. However, the extension of this approach to generate pairwise or family weights was straightforward. A stepwise logistic regression analysis was used to identify twin pair variables that were associated with twin pair participation in the alcohol challenge study. Dummy variables were used to code the status of first and second twins with respect to a given predictor variable, e.g., using two dummy variables to distinguish twin pairs where neither twin had completed a university education (0,0), only one twin had completed a university education (1,0) or both twins had completed a university education (0,1). For continuous variables, we used the twin pair mean as the predictor, although with larger sample sizes it would have been feasible to create categorical dummy variables (e.g., for pairs where one twin scored in the highest 25%-ile and the cotwin in the lowest 25%-ile). As before, the logistic regression equation was used to estimate the predicted probability of the twin pair participating in the challenge study, pairwise expansion weights were computed for the challenge study participants as the reciprocal of these predicted probabilities, and pairwise sampling weights were computed as the pairwise expansion weight for the ith pair divided by the mean expansion weight for the challenge sample. A similar procedure would be used in longitudinal follow-up studies, predicting the probability that one or both members of the pair, i.e., at least one member of the pair, participated at follow-up (e.g., Heath *et al.*, 1997).

Sample sizes in the alcohol challenge study, though large by the standards of experimental studies, were nonetheless small by the standards of behavior genetic research. This made it more difficult to conduct pairwise reweighting of the data, since the number of potential dummy variables would have been large compared to the number of observed twin pairs in the smallest group (unlike-sex pairs)! For this reason, we modeled only two-way interaction effects in the pairwise analyses. Likewise, while it would be preferable to estimate logistic regression equations separately for each zygosity group, we were forced to derive separate equations for female like-sex, male like-sex and unlike-sex pairs, because of small sample sizes. We did however compare the results of deriving logistic regression weights separately for MZ and DZ like-sex pairs of a given gender, versus deriving separate logistic regression weights, using the same

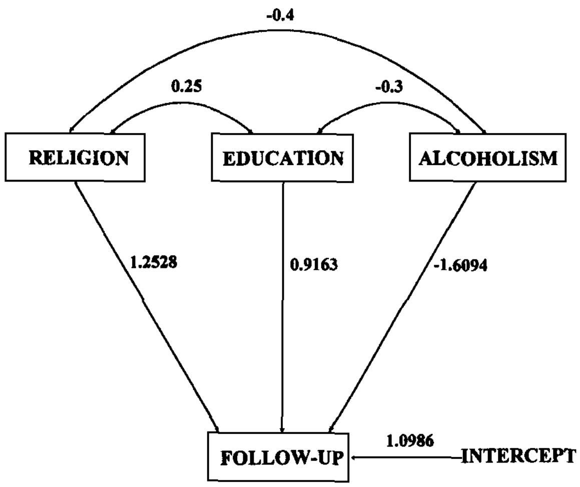

Fig. 1. Logistic regression model used to predict probability of individual retention in a follow-up sample for simulation study of individual sampling weights.

variables identified as significant in analyses pooled across all like-sex pairs of that gender. We were also forced to use pair means for analyses of the unlike-sex pair group, though with larger sample sizes it would have been preferable to include the observed values for male and female cotwins as separate predictors of twin pair participation.

Once again, we used simulated data to ensure that our data-weighting procedure was correctly recovering true parameter values. For the pairwise simulation, probability of response at follow-up was modeled as a logistic function of twin's and cotwin's 'educational status' and 'alcoholism.' The same procedures were then applied to pairwise analyses of the alcohol challenge data. For the alcohol challenge analyses, logistic regression and computation of sampling weights was conducted separately for male like-sex, female like-sex and unlike-sex pairs.

The adequacy of our data-weighting procedure in part depends upon the appropriateness of assuming a logistic regression model to describe the relationship between baseline variables and probability of participating in the challenge study. As an empirical check of the adequacy of our data-weighting procedure, we reran the logistic regression analysis using individual or pairwise weights, and also conducted weighted chisquare tests for any remaining univariate association between baseline variables and participation in the alcohol challenge study. To examine the effects on alcohol challenge performance, we obtained weighted frequency tables by gender for the subjective intoxication variable; and also computed twin pair correlations and covariance matrices by zygosity and used these to estimate genetic and environmental parameters. It should be noted that in order to conduct formal statistical tests using such weighted data, it would be necessary to use boot-strapping (Efron and Tibshirani, 1986) or a sandwich-estimator (StataCorp, 1997) to estimate sampling variances of variables, which will be increased by the weighting procedure compared to unweighted data (Lee *et al.*, 1989). We did not attempt to do so here, since we do not attempt formal hypothesis-testing using the weighted data [but see Heath *et al.* (1997) for an illustration of the use of boot-strapping].

RESULTS

Simulation Study

Figure 1 summarizes the parameters of the logistic regression model that was used to generate baseline and follow-up data for the deterministic simulation for individual data-weighting. The joint distribution of the baseline sample with respect to religion, education and alcoholism was generated under a trivariate normal threshold model, using baseline prevalence estimates of 35% for the religion variable, 25% for the education variable and 15% for the alcoholism variable. Table I illustrates the computation of individual sampling weights from the simulated data. Predicted probabilities of retention in the follow-up sample range from a high of 0.963 (for those positive on Religion and Education and negative on Alcoholism) to a low of 0.375 (for those negative on religion and education and positive for alcoholism), with an overall mean response rate of 77.9%. Because of the non-random attrition, in the simulated follow-up data, the estimated prevalence for alcoholism is reduced from 15% at baseline to 8.5% at follow-up, while prevalence estimates for Religion and Education variables are somewhat increased. The expansion weights are simply the reciprocal of the retention probabilities estimated from the logistic regression analysis of the simulated data; and the relative weights are equal to the corresponding expansion weight divided by the weighted mean expansion

Table I. Use of Individual Sampling Weights: Simulated Data (Figures Are Shown Only to 3 Decimal Places)

Baseline trait value			Baseline probability distribution	Probability of retention at follow-up	Follow-up probability distribution (unweighted)	Estimated expansion weight	Estimated relative weight	Follow-up probability distribution (weighted)
Religion	Education	Alcoholism						
0	0	0	0.403	0.750	0.388	1.333	1.039	0.403
0	0	1	0.115	0.375	0.055	2.667	2.077	0.115
0	1	0	0.118	0.882	0.134	1.133	0.883	0.118
0	1	1	0.014	0.600	0.011	1.666	1.298	0.014
1	0	0	0.215	0.913	0.250	1.095	0.853	0.215
1	0	1	0.017	0.677	0.015	1.476	1.150	0.017
1	1	0	0.114	0.963	0.141	1.038	0.809	0.114
1	1	1	0.004	0.840	0.004	1.191	0.927	0.004
				$\bar{x} = 0.779$		$\bar{x} = 1.2839$		

Prevalence estimate	Baseline (%)	Follow-up (unweighted) (%)	Follow-up (weighted) (%)
Religion	35	41.2	35
Education	25	29.0	25
Alcoholism	15	8.5	15

weight. In the final column, it can be seen that the relative weights do indeed reweight the follow-up sample to have the same distribution as the baseline sample, and allow the baseline prevalence estimates to be recovered.

Figure 2 summarizes the logistic regression model used to simulate twin pair retention in the follow-up sample (defined here as retention of at least one twin in the pair), which now includes only education and alcohol dependence as predictor variables. Data were simulated under the assumptions of 80% heritability of the educational attainment variable, and 60% heritability of alcohol dependence, with the phenotypic correlation of −0.3 entirely explained by shared genetic influences. Table II summarizes the baseline and unweighted follow-up probability distributions, and the estimated expansion and relative (or sampling) weights. The logistic regression analysis, which combined data from both MZ and DZ pairs (since retention of a twin pair in the sample was assumed to be purely a function of individual phenotypic values), yielded odds ratios for inclusion in the follow-up sample of 7.83 for concordant 'educated' pairs, 2.80 for discordant pairs, 0.03 for concordant alcoholic pairs, and 0.17 for discordant pairs. Table III summarizes the estimates of prevalence and of the twin pair tetrachoric correlations for education and alcoholism. Under the severe sample attrition bias (at least with respect to alcoholism) implied by Fig. 2, estimates of the prevalence and heritability of alcoholism decline from the true values of 15% and 60% in the baseline sample to 6.5% MZ prevalence, 7.9% DZ prevalence and 34% heritability in the unweighted follow-up data (not shown). Once again, weighted analysis of the follow-up sample recovered the true parameter values.

Alcohol Challenge Sample

Table IV summarizes the observed frequency distributions for sociodemographic and lifestyle variables for alcohol challenge study participants compared to other Australian twins who responded to the 1981 survey. Mean log-transformed average weekly alcohol consumption, and mean scores on the personality measures of N and L, are also given; no differences were observed for scores on E or P. Compared to other twins from the same age cohort, those who volunteered for the alcohol challenge study were younger, much more likely to be unmarried, were heavier drinkers, were less socially conforming, and less likely to report a religious affiliation of 'Other Protestant.' Women were also more likely to report a University education or to have attended teachers' or technical college, less

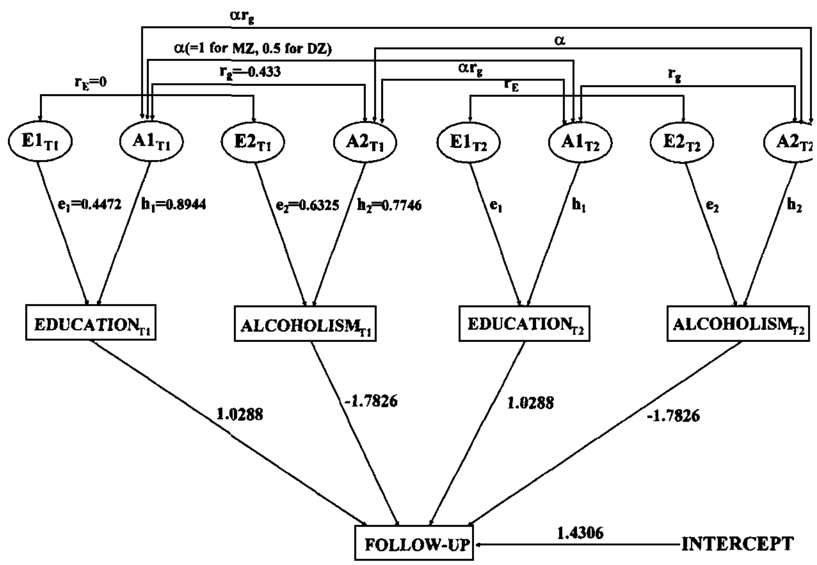

Fig. 2. Logistic regression model used to predict probability of retaining at least one twin from a pair in a follow-up sample for data simulation of pairwise sampling weights.

Table II. Use of Pairwise Sampling Weights: Simulated Data (Figures Are Shown Only to 3 Decimal Places)

Twin A		Twin B		Baseline probability distribution		Pair retention	Follow-up probability distribution (unweighted)[b]		Estimated expansion	Estimated
Education	Alcoholism	Education	Alcoholism	MZ pairs	DZ pairs	probability[a]	MZ pairs	DZ pairs	weight	relative weight
0	0	0	0	0.479	0.417	0.807	0.516	0.447	1.239	0.930
0	1	0	1	0.059	0.034	0.106	0.008	0.005	9.454	7.094
1	0	1	0	0.152	0.095	0.970	0.197	0.122	1.031	0.773
1	1	1	1	0.003	0.001	0.481	0.002	0.001	2.080	1.561
0 0	0 1	0 0	1 0 }	0.132	0.156	0.413	0.073	0.085	2.422	1.817
1 0	0 0	0 1	0 0 }	0.137	0.229	0.921	0.168	0.281	1.085	0.815
0 1	0 1	1 0	1 0 }	0.010	0.016	0.663	0.009	0.014	1.508	1.132
0 1	1 0	1 0	0 1 }	0.010	0.034	0.663	0.008	0.030	1.508	1.132
0 1	1 1	1 0	1 1 }	0.006	0.003	0.249	0.002	0.002	4.022	3.018
1 1	0 1	1 1	1 0 }	0.014	0.012	0.846	0.016	0.007	1.182	0.887
						MZ pairs: 0.748			1.336923	
						DZ pairs: 0.753			1.32837	
								Overall:	1.332633	

[a] Probability that at least one twin will respond at follow-up.
[b] In these simulated data, the weighted follow-up probability distribution was identical to the baseline distribution.

Table III. Estimates of Prevalence and Tetrachoric Correlations from Simulated Data

	Baseline sample		Unweighted follow-up sample		Weighted follow-up sample	
	Prevalence (%)	Tetrachoric correlation ρ	Prevalence (%)	Tetrachoric correlation ρ	Prevalence (%)	Tetrachoric correlation ρ
MZ pairs						
Education	25	0.80	30.9	0.78	25	0.80
Alcoholism	15	0.60	6.5	0.34	15	0.60
DZ pairs						
Education	25	0.40	29.9	0.36	25	0.40
Alcoholism	15	0.30	7.9	0.06	15	0.30

Table IV. Sociodemographic, Lifestyle, and Personality Differences Between Alcohol Challenge Study Participants and 1981 Survey Participants; the Effects of Reweighting the Alcohol Challenge Sample Are Also Shown (Two Rightmost Columns)

	Alcohol challenge study (%)		1981 survey[a] (%)		Association with challenge study participation: partial odds ratio[b]				Weighted challenge data[e] (%)	
	Women	Men	Women	Men	Women		Men			
	(N=213)	(N=199)	(N=3237)	(N=1975)	OR	95% CI[c]	OR	95% CI[c]	Women	Men
0–10 years schooling	21.1	19.1	22.5	12.9^{NS}	1.00	—	1.00	—	21.1	13.3
11–12 years or diploma	39.9	46.7	52.6	53.7^{NS}	0.53	0.38–0.70	0.55	0.39–0.78	59.1	51.2
Technical/teacher's college	24.4	13.6	15.2	11.8^{NS}	1.00	—	1.00	—	13.1	10.2
University education	14.6	20.6	9.7	21.6^{NS}	1.00	—	0.60	0.39–0.91	6.8	25.3
Other Protestant religion	15.5	13.1	30.4	25.0	0.52	0.36–0.77	0.53	0.35–0.82	29.3	24.2
Weekly church attendance[c]	15.4	16.8	22.5	17.8^{NS}	—	—	—	—	20.4	18.7
Never married	65.7	70.4	41.2	57.0	2.10	1.54–2.92	2.02	1.45–2.81	36.1	57.0
Born 1950–63	90.6	85.9	76.5	78.7	2.10	1.27–3.50	—	—	77.9	82.7
Has smoked cigarettes	52.6	54.3	43.7	49.1^{NS}	—	—	—	—	52.6*	45.9
Drinks at least 3–4 times/week	24.4	35.2	13.5	29.4^{NS}	—	—	—	—	14.0	30.6
Drinks at least 1–2 times/month	59.2	52.7	53.6	52.6	—	—	—	—	54.8	54.1
Drinks less often	16.4	12.1	32.9	18.0	—	—	—	—	30.3	16.3
Drinks 1–2 drinks/occasion	39.9	21.1	54.5	33.4	1.00	—	1.00	—	56.0	33.1
Drinks 3–5 drinks/occasion	35.2	36.2	29.4^{NS}	24.6	1.00	—	2.06	1.49–2.84	25.4	24.0
Drinks 6+ drinks/occasion	24.9	42.7	16.1	42.0^{NS}	0.52	0.34–0.78	1.00	—	18.6	42.9
	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$					$\bar{x}$	$\bar{x}$
Social nonconformity[d]	0.65	0.68	0.56	0.62	1.62	1.28–2.05	1.56	1.22–2.01	0.58	0.62
Average weekly alcohol consumption (log)[d]	1.68	2.20	0.98	1.80	2.89	2.15–3.87	1.48	1.18–1.85	0.94	1.84
Neuroticism	0.52	0.42	0.51^{NS}	0.41^{NS}	—	—	—	—	0.47*	0.39

Note: After data-weighting, males from the alcohol challenge sample did not differ significantly from total sample (χ^2_{14} = 6.50, p=0.95); but among women, smokers were over-represented in the alcohol challenge sample (OR = 1.49; 95% CI, 1.13–1.97), and those high on Neuroticism underrepresented (OR = 0.78; 95% CI, 0.65–0.93), compared to the total sample.

[a] Excluding alcohol challenge participants. All differences between challenge and 81 survey samples are significant in univariate comparisons unless indicated by ^{NS}.

[b] Estimated from multiple logistic regression analysis predicting alcohol challenge participation.

[b] Not adjusted for non-independence of observations on twin pairs.

[c] Data from 1981 survey respondents only.

[d] For continuous variables, Odds Ratios are computed for a change in score equal to the interquartile range.

[e] All comparisons between challenge and 1981 survey samples are non-significant unless noted by * (p<.05).

Table V. Effects of Marital Status, Smoking Status, and Religious Affiliation on Willingness to Participate in Alcohol Challenge Research

				Women			Men		
				1981 Survey[a] (%)	Alcohol challenge sample Unweighted (%)	Alcohol challenge sample Weighted[b] (%)	1981 Survey[a] (%)	Alcohol challenge sample Unweighted (%)	Alcohol challenge sample Weighted[b] (%)
Group	Marital status single?	Other Protestant?	Ever smoked?						
1	✗	✗	✗	19.5	10.3	20.3	12.2	8.5	10.9
2	✓	✗	✗	17.0	32.4	20.1	23.4	31.7	28.1
3	✗	✓	✗	12.4	3.3	9.3	6.0	3.5	4.9
4	✗	✗	✓	20.1	16.9	22.0	18.9	15.1	18.6
5	✓	✓	✗	7.4	1.4	4.6	9.5	2.0	12.0
6	✗	✓	✓	6.8	3.8	10.1	5.7	2.5	5.0
7	✓	✗	✓	13.0	24.9	11.1	20.6	31.7	17.9
8	✓	✓	✓	3.9	7.0	2.6	3.8	5.0	2.7

[a] Excluding alcohol challenge participants.
[b] Weights derived from a logistic regression equation incorporating dummy variables to contrast to the reference group (group 1): (i) group 3, (ii) group 4, (iii) group 5, (iv) group 6, and (v) groups 2, 7, and 8, in addition to other predictors from Table IV multiple logistic regression equations.

likely to have a high-school education or post-school diploma, less likely to report weekly church attendance, and more likely to have been a cigarette smoker.

Also given in Table IV are partial odds ratios estimated from the multiple logistic regression analysis predicting alcohol challenge study participation. Major predictors of alcohol challenge study participation were average weekly alcohol consumption, drinking 3–5 drinks per occasion (in men only), being single, being from the youngest birth cohort (women only), not having a religious affiliation of Other Protestant, and not having just a high school education (or other diploma). Reported drinking of 6 or more drinks per drinking occasion in women was actually associated with reduced likelihood of alcohol challenge study participation, once total weekly consumption was controlled for, suggesting that women participants tended to be frequent drinkers rather than binge drinkers. University-educated respondents were found to be less likely to participate - the apparent inconsistency with the significant univariate association in women is most probably explained by the higher rates of University education in the youngest birth cohort, in which alcohol challenge participants were overrepresented.

Finally, the two right-hand columns of Table IV present the weighted frequency distribution for the alcohol challenge sample when individual response propensity weights were used. In men, data-weighting removed the differences between the alcohol challenge sample and the baseline sample. In women, differences were removed for most variables, including church attendance, which had not been included in the multiple logistic regression analysis. However, a significant difference in Neuroticism was introduced, with women in the alcohol challenge sample having significantly (albeit modestly) lower Neuroticism scores. This effect was easily removed by including Neuroticism score in the logistic regression equation predicting challenge study participation (not shown). Women from the alcohol challenge sample were still significantly more likely to be smokers, and this remained true even when smoking history was included in the logistic regression equation.

One possible explanation of the failure to remove the association between smoking and alcohol challenge performance in women was that a basic assumption of the model predicting volunteering, that there were no significant interactions between predictor variables, was incorrect. Closer examination of the data confirmed this (see Table V). Specifically, we observed a three-way interaction between marital status, religious affiliation and smoking history. Women who were unmarried were much more likely to participate in the alcohol

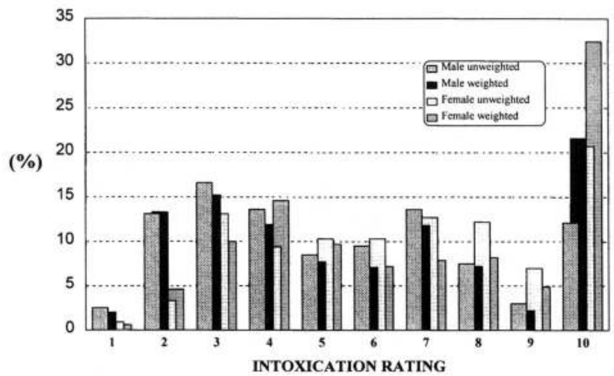

Fig. 3. Frequency distribution of subjective intoxication ratings based on unweighted data versus data weighted using individual weights.

challenge study unless they reported a religious affiliation of Other Protestant, in which case they had a reduced probability of participation—unless they were also smokers, in which case they were still more likely to participate! Exactly the same trend was observable in men. (With hindsight, this pattern is easily explained. The category 'Other Protestant' includes both religious denominations which disapprove of alcohol use, and others which do not. Presumably the smokers come disproportionately from the latter denominations).

When the logistic regression equations of Table IV were modified by the use of five dummy variables to parameterize the interaction of marital status, religion and smoking history (see Table V), instead of 3 parameters for the main effects of these variables only, the resulting data weights did indeed successfully reweight the female alcohol challenge study participants to have the same distribution as the entire baseline sample. This was true both with respect to the measures of marital status, other Protestant religion and smoking status (see two right-most columns of Table V), and with respect to all the other predictor variables from Table IV (not shown). The weighted estimates of the prevalence of lifetime smoking in women, based on the alcohol challenge sample, was 45.7%, compared to 43.7% in the remaining 1981 cohort sample.

Figure 3 compares the frequency distribution of subjective intoxication ratings for the alcohol challenge sample based on unweighted versus weighted data, using weights based on logistic regression equations that took into account the 3-way interaction of marital status, religion and smoking history. The weighted data give an approximation to what would have been observed if we had been able to test a random sample from the twin panel. There is a very striking increase in the weighted data in the proportion of subjects, both female and male, reporting they felt the most drunk they had ever been, confirming that it was specifically those individuals who were most sensitive to the effects of alcohol, who were presumably also the lightest drinkers, who were being undersampled.

Pairwise Weights

Table VI summarizes the significant pairwise predictors of twin pair participation in the alcohol challenge study. Across all groups, alcohol challenge participation was significantly more likely in pairs where both twins scored high on social nonconformity, and significantly less likely in pairs where both twins had only a high school education. In like-sex pairs, participation was more likely in pairs with higher mean reported average weekly alcohol consumption, and where both twins were unmarried. Finally, female like-sex pairs were much less likely to be from pairs who were concordant both for being non-smokers and for Other Protestant religious affiliation, while unlike-sex pairs were much more likely to participate if they were concordant smokers and neither had an Other Protestant religious affiliation.

The pairwise weights generated using the logistic regression equations of Table VI removed the association between alcohol challenge participation and personality and average weekly alcohol consumption for both like-sex and unlike-sex twin pairs, and removed the association between other predictor variables and alcohol challenge participation for male like-sex and unlike-sex pairs. However, for female pairs significant associations were created between alcohol challenge participation and being from pairs that were (i) concordant for having completed a University education, (ii) discordant for marital status with one twin unmarried, (iii) discordant for both ever smoking and other protestant religion, and the association between reporting a high school education and alcohol challenge study participation was not completely removed. Including additional variables for (i)-(iii)

Table VI. Twin Pair Variables that Predict Participation in the Alcohol Challenge Study

	Female like-sex pairs		Male like-sex pairs		Unlike-sex pairs	
	Partial OR	95% CI	Partial OR	95% CI	Partial OR	95% CI
Concordant other Protestant & non-smokers	0.22	0.05–0.93	—	—	—	—
Concordant smokers, neither other Protestant	—	—	—	—	2.28	1.14–4.53
Concordant unmarried	2.66	1.65–4.28	2.16	1.31–3.57	—	—
Concordant high school	0.42	0.24–0.71	0.57	0.34–0.97	0.24	0.09–0.62
Pair mean social non-conformity score	1.83	1.27–2.63	1.63	1.09–2.45	2.09	1.17–3.73
Pair mean weekly alcohol consumption	2.54	1.66–3.88	1.76	1.22–2.54	—	—

in the logistic regression equation predicting female like-sex pair alcohol challenge participation removed these associations, but left pairs concordant for being Other Protestant and non-smokers significantly underweighted. It is likely that this difficulty reflects the fact that we were unable to model the three-way interactions between smoking status, marital status, and Other Protestant religious affiliation in pairwise analyses, because of small numbers of twin pairs.

Table VII summarizes unweighted and weighted estimates (using weights derived from regression equations of Table VI, with the addition of terms for female like-sex pairs as described above) of product-moment correlations for subjective intoxication, for each twin pair zygosity group. With such small sample sizes, it is difficult to reach firm conclusions. Nonetheless data-weighting has somewhat increased the monozygotic correlations, particularly in females, and has increased rather than diminished the differences between monozygotic and like-sex dizygotic correlations. When regression weights were derived separately for MZ and DZ like-sex pairs, and zygosity-specific sampling weights estimated, weighted correlations differed very little from those in Table 7 (0.51, 0.51, −0.30, −0.19, 0.23). Apparently any cooperation bias effects have led us to underestimate rather than overestimate the importance of genetic influences on subjective intoxication after alcohol challenge. Model-fitting to the weighted covariance matrices yielded a broad heritability estimate of 44%, compared to an estimate of 42% for the unweighted data. Data weighting thus suggests that volunteer bias, at least with respect to the personality, lifestyle and sociodemographic variables that we have examined, has had a minimal effect on our estimates of the genetic contribution to differences in subjective intoxication after alcohol challenge.

Table VII. Unweighted and Weighted Estimates of Twin Pair Correlations for Subjective Intoxication After Alcohol Challenge

	MZF	MZM	DZF	DZM	DZFM
N pairs	45	43	42	37	39
Unweighted *r*	0.44	0.48	−0.14	0.12	0.08
Weighted *r*	0.53	0.49	−0.29	−0.14	0.23

CONCLUSIONS

In this brief methodologic paper, we have illustrated how logistic regression analysis can be used to generate 'response propensity' weights (Rosenbaum and Rubin, 1983) in order to assess the effects of selective attrition on estimates of prevalence and of genetic and environmental variances. In the case under examination, willingness to volunteer for an alcohol challenge study, while there appeared to be quite marked undersampling of light drinkers, effects on estimates of genetic and environmental parameters were rather slight, leading to a modest under rather than over-estimation of the importance of genetic factors. The approach that we have adapted is made more complicated when there are subtle interactions between variables, as we observed for the interaction of religious affiliation, marital status and smoking status. We conclude, however, that it is likely to be of some utility in behavioral genetic analyses of longitudinal data on psychiatric disorders or other binary outcomes, as well as of continuous data when major predictors of non-response and covariates are categorical (e.g. sociodemographic) variables.

APPENDIX

Example SAS Program (SAS, 1990) for Generating Individual Sampling Weights

The script for generating pairwise weights would differ from this principally in the use of twin pair variables predictors in the logistic regression equation, and the use of SAS Proc Corr, in conjunction with a weight statement, to obtain the twin pair variance-covariance matrix. An example script for this purpose is available from the authors.

```
libname work 'input.dat';
***file with alcohol challenge (alcchall = 1) plus 1981 survey only sample (alcchall = 0);
data all; set work.all;
*logistic regression analysis to compute predicted probabilities of participation in study from baseline variables;
proc logistic des; model alcchall = var1-var10/rl;
output out = outfile p = p;
***alcohol challenge respondents only selected, mean expansion weight computed;
data chall; set outfile; if alcchall = 1;
*pinv is the expansion weight;
pinv = 1/p;
proc univariate; var pinv;
data work; set chall;
*sampling ('relative') weight computed;
wt = pinv/pbar; *pbar is numerical mean value of pinv obtained from proc univariate;
*Get weighted frequency of subjective intoxication variable;
proc freq; tables twnsadsp; weight wt;
proc sort; by idnumber;
proc sort data = all; by idnumber;
data updated; update all chall; by idnumber;
*Update original file with weights for challenge subsample, set weights to unity for remaining sample, and check associations have been removed in weighted data;
if alcchall eq 0 then wt = 1;
proc freq; tables alcchall*(var1-var10)/all; weight wt;
```

ACKNOWLEDGMENTS

This work was supported by NIH Grants AA07535, AA07728, AA09022, AA10249, AA10339, MH31307, MH52813 (A.C.H.), and DA00273 (P.A.F.M.) and by grants from the Alcohol Beverage Medical Research Foundation (P.A.F.M.), the Australian National Health and Medical Research Council (N.G.M.), and the Templeton Foundation (A.C.H.).

REFERENCES

Cederlof, R., Epstein, F. H., Friberg, L. T., Hrubec, Z., and Radford, E. P. (1971). Twin registries in the study of chronic disease (with particular reference to the relation of smoking to cardiovascular and pulmonary diseases). *Acta Med. Scand. Suppl.* **523**:1–40.

Eaves, L. J., Eysenck, H. J., and Martin, N. G. (1989). *Genes, Culture, and Personality: An Empirical Approach*, Academic Press, London.

Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1**:54–77.

Eisen, S. A., True, W., Goldberg, J., Henderson, W., and Robinette, C. D. (1987). The Vietnam Era Twin (VET) Registry: Method of construction. *Acta Genet. Med. Gemellol.* **36**:61–66.

Eysenck, H. J., and Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*, Hodder & Stoughton, London.

Heath, A. C., and Martin, N. G. (1990). Psychoticism as a dimension of personality: A multivariate genetic test of Eysenck and Eysenck's psychoticism construct. *J. Pers. Soc. Psychol.* **58**:1–11.

Heath, A. C., and Martin, N. G. (1992). Genetic differences in psychomotor performance decrement after alcohol: A multivariate analysis. *J. Stud. Alcohol* **53**:262–271.

Heath, A. C., Jardine, R., and Martin, N. G. (1989). Interactive effects of genotype and social environment on alcohol consumption in female twins. *J. Stud. Alcohol* **50**:38–48.

Heath, A. C., Bucholz, K. K., Madden, P. A. F., Dinwiddie, S. H., Slutske, W. S., Statham, D. J., Dunne, M. P., Whitfield, J., and Martin, N. G. (1997). Genetic and environmental contributions to alcohol dependence risk in a national twin sample: Consistency of findings in women and men. *Psychol. Med.* **27**:1381–1396.

Holm, N. V., Hauge, M., and Harvald, B. (1980). Etiologic factors of breast cancer elucidated by a study of unselected twins. *J. Natl. Cancer Inst.* **65**(2):285–298.

Hosmer, D. W., and Lemeshow, S. (1989). *Applied Logistic Regression*, John Wiley & Sons, New York.

Hrubec, Z., and Neel, J. V. (1978). The National Academy of Sciences—National Research Council Twin Registry: Ten years of operation. In Nance, W. E. (ed.), *Twin Research, Part B, Biology and Epidemiology*, Alan R. Liss, New York, pp. 154–172.

Jardine, R., and Martin, N. G. (1984). Causes of variation in drinking habits in a large twin sample. *Acta Genet. Med. Gemellol.* **33**:435–450.

Kaprio, J., Sarna, S., and Rantasalo, I. (1978). The Finnish Twin Registry: Formation and compilation, questionnaire study, zygosity determination procedures, and research program. *Prog. Clin. Biol. Res.* **24**:179–184.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., and Eaves, L. J. (1992). A population based twin study of major depression in women: The impact of varying definitions of illness. *Arch. Gen Psychiatry* **49**:257–266.

Lee, E. S., Forthofer, R. N., and Lorimor, R. J. (1989). *Analyzing Complex Survey Data*, Sage, Newbury Park.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, New York, John Wiley.

Little, R. J. A., and Schenker, N. (1995). Missing data. In Arminger, G., Clogg, C. C., and Sobel, M. E. (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Plenum, New York, pp. 39–69.

Magnus, P., Berg, K., and Nance, W. E. (1983). Predicting zygosity in Norwegian twin pairs born 1915-1960. *Clin. Genet.* **24**:103–112.

Martin, N. G., and Wilson, S. R. (1982). Bias in the estimation of heritability from truncated samples of twins. *Behav. Genet.* **12**:467–472.

Martin, N. G., Oakeshott, J. G., Gibson, J. B., Starmer, G. A., Perl, J., and Wilks, A. V. (1985a). A twin study of psychomotor and physiological responses to an acute dose of alcohol. *Behav. Genet.* **15**:305–347.

Martin, N. G., Perl, J., Oakeshott, J. G., Gibson, J. B., Starmer, G. A., and Wilks, A. V. (1985b). A twin study of ethanol metabolism. *Behav. Genet.* **15**:93–109.

Medlund, P., Cederlof, R., Floderus-Myrhed, B., and Sorensen, S. (1977). A new Swedish twin registry. *Acta Med. Scand. Suppl.* 600.

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

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

Neale, M. C. (1998). *Mx: Statistical Modeling*, Department of Psychiatry, Virginia Commonwealth University.

Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the proensity score in observational studies for causal effects. *Biometrika* **70**:41–55.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley, New York.

SAS Institute (1990) *SAS/STAT User's Guide, Version 6, 4th ed.*, SAS Institute, Cary, NC.

Scarr, S., Webber, P. L., Weinberg, R. A., and Wittig, M. A. (1981). Personality resemblance among adolescents and their parents in biologically related and adoptive families. In Gedda, L., Parisi, P., and Nance, W. E. (eds.), *Twin Research 3: Intelligence, Personality, and Development*, Alan R. Liss, New York, pp. 99–120.

StataCorp (1997). *Stata Statistical Software: Release 5.0*, Stata Corporation, College Station, TX.

Edited by Stacey Chesny