View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Crossref

Effects of Censoring on Parameter Estimates and Power in Genetic Modeling

Eske M. Derks,[1] Conor V. Dolan,[2] and Dorret I. Boomsma[1]

[1] *Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands*

[2] *Department of Psychology, University of Amsterdam, the Netherlands*

Genetic and environmental influences on variance in phenotypic traits may be estimated with normal theory Maximum Likelihood (ML). However, when the assumption of multivariate normality is not met, this method may result in biased parameter estimates and incorrect likelihood ratio tests. We simulated multivariate normal distributed twin data under the assumption of three different genetic models. Genetic model fitting was performed in six data sets: multivariate normal data, discrete uncensored data, censored data, square root transformed censored data, normal scores of censored data, and categorical data. Estimates were obtained with normal theory ML (data sets 1–5) and with categorical data analysis (data set 6). Statistical power was examined by fitting reduced models to the data. When fitting an ACE model to censored data, an unbiased estimate of the additive genetic effect was obtained. However, the common environmental effect was underestimated and the unique environmental effect was overestimated. Transformations did not remove this bias. When fitting an ADE model, the additive genetic effect was underestimated while the dominant and unique environmental effects were overestimated. In all models, the correct parameter estimates were recovered with categorical data analysis. However, with categorical data analysis, the statistical power decreased. The analysis of L-shaped distributed data with normal theory ML results in biased parameter estimates. Unbiased parameter estimates are obtained with categorical data analysis, but the power decreases.

Thanks to computational and methodological advances over the last few decades, genetic covariance structure modeling in genetically informative samples is relatively straightforward. Estimates of genetic and environmental variance components may be obtained readily using programs like Mx (Neale, 1997), Lisrel (Jöreskog & Sörbom, 1996a), or Mplus (Muthén & Muthén, 2001; Prescott, 2004). The dominant method of estimation is normal theory Maximum Likelihood (normal theory ML), which is based on the assumption of multivariate normality. Unfortunately, the distribution of phenotypic data may display a large degree of skewness and kurtosis, which renders the choice of normal theory ML to estimate parameters suboptimal. The problem of nonnormality is acute in the study of symptom data, where the distribution of observed symptoms is often L-shaped, due to the fact that the vast majority of subjects display few or no symptoms (Van den Oord et al., 2003). Failure to account for nonnormality may lead to biased parameter estimates and incorrect likelihood ratio tests (Amos, 1994).

There are many possible causational factors for the presence of nonnormality. These can be divided into two categories: a) factors that lead to a nonnormal distribution of the latent trait; and b) factors that lead to a nonnormal distribution of the measured indicators of a normally distributed latent trait. If the distribution of the latent trait is not normal, a possible solution is to adopt a more appropriate distribution (e.g., Poisson). If the latent trait is normally distributed, but the observed trait is not, for example due to censoring, a possible solution is to correct the observed data for the censoring event.

Van den Oord et al. (2003) proposed that the latent distribution of L-shaped behavioral checklist data is normal. They examined this hypothesis by means of Item Response Theory (IRT; Hambleton & Swaminathan, 1985) and found that a model that allowed for nonnormality in the latent distribution did not provide a better fit than a model that did not allow for nonnormality. In other words, they found no evidence against a normal latent distribution. Therefore, we assume that the L-shaped distribution of behavioral checklist data belongs to the second category. This seems plausible because questions in behavior checklists are often developed with the purpose of determining the degree of behavioral dysfunctioning. In the latent normal distribution, children with well-adapted behavior may be found at

Received 3 September, 2004; accepted 13 September, 2004.

Address for correspondence: E. M. Derks, Vrije Universiteit, Department: Biological Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands. E-mail: em.derks@psy.vu.nl

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

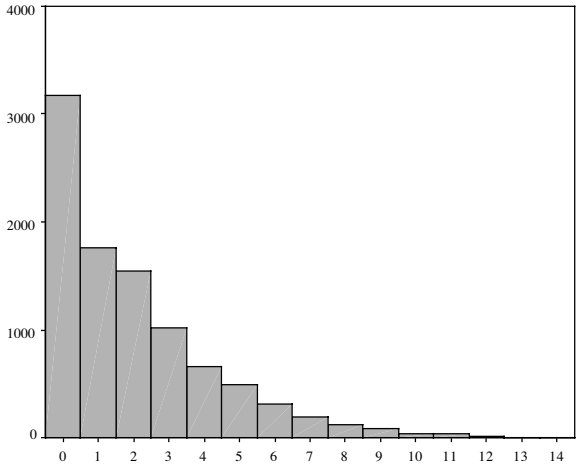


Figure 1

Distribution of maternal reports of sleep problems.

Note: Graph is based on 9415 first-born twins; data of second-born twins are similar.

the left tail of the distribution, and children with dysfunctional behavior may be found at the right tail of the distribution. Because of the focus on behavioral dysfunction, variation in the right tail in the distribution is measured while variation in the left tail of the distribution is not. This results in an L-shaped observed distribution. An example of such a distribution is shown in Figure 1. This figure illustrates the degree of sleep problems in three-year-old children. The distribution clearly is not normally distributed, which is probably caused by censoring.

Censored data arise if values below (or above) a certain threshold y* are observed at y*. As a result, below (or above) this threshold, variation in the distribution of the latent trait is unobserved, and the observed distribution is skewed. The effect of censoring from below is illustrated in Figure 2.

There are numerous practical examples of censored distributions in many different fields of inquiry (e.g., economics, medicine and the social sciences). One of the earliest attempts to address censoring is that of Tobin (1958). He studied the demand for various categories of capital goods such as automobiles. Many households report zero expenditures in a given year. Among the households that made an expenditure, there is large variability in amount. The observed demand for capital goods in a given year is therefore censored below. An example from the field of medicine is the assessment of coronary artery calcification (Epstein et al., 2003). Coronary artery calcification is only assumed to be present when it exceeds a certain threshold. Below this threshold, the level is assumed to be zero (Bielak et al., 2001). Finally, censoring is present in behavioral ratings (Nagin & Tremblay, 1999; Rietveld et al., 2003).

There are several methods to correct for nonnormality. First, the data may be transformed in order to achieve normality. Often-applied transformations are the logarithmic and square root transformations (Lynch & Walsh, 1998). Transformations may be substantively motivated, for example, the use of a logarithmic transformation when a trait is measured on a geometric scale instead of an arithmetic scale (Falconer & Mackay, 1996). Examples of such traits are body weight and growth. Alternatively, transfor-

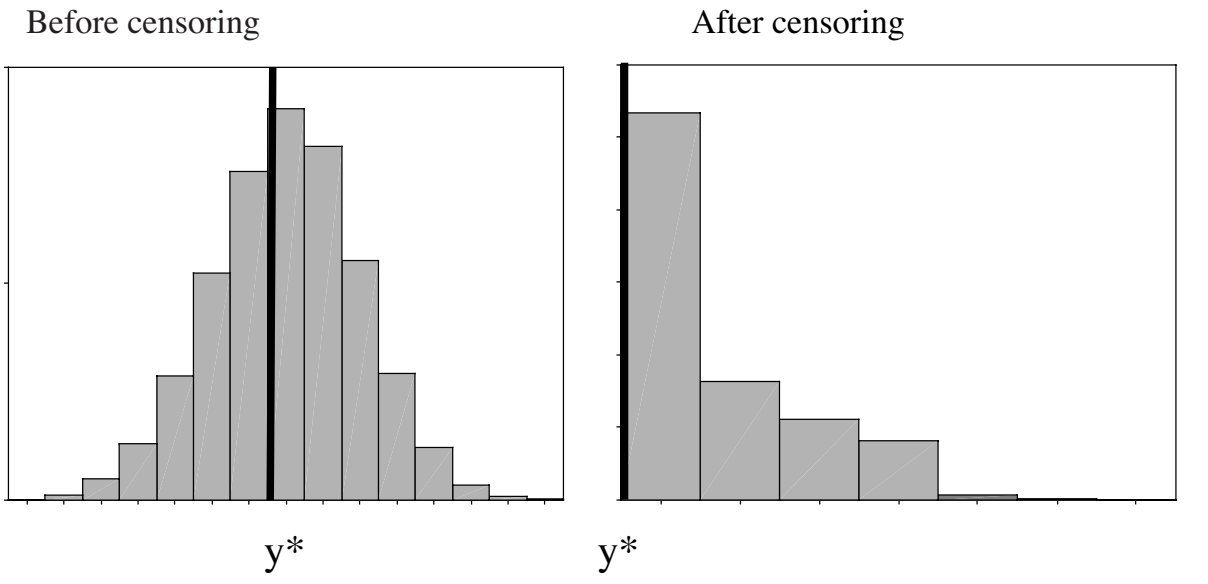


Figure 2

A graphical representation of the censoring process.

Note: Before censoring, the complete normal distribution is shown. After censoring, values below y* are observed at value y*.

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

mations may be ad hoc. An example of this is the single parameter Box–Cox transformation, where a parameter is optimized to achieve normality. In the case of behavioral rating data, the transformation is generally ad hoc. The calculation of so-called normal scores (Jöreskog et al., 1999) may also be viewed as an ad hoc transformation. The aim of this transformation, which is based on the assumption of an underlying normal distribution, is to render the skewness and kurtosis of the data consistent with a normal distribution (i.e., values of approximately 0 and 3, respectively). Unfortunately, given major deviations from normality, transformations may fail to achieve this aim.

Second, we may assume that the observed data is measured on an ordinal scale instead of a continuous scale and adopt a method of estimation which is suitable for ordinal data. Parameter estimates of the genetic model fitting can be obtained under the assumption of an underlying continuous liability distribution that has one or more thresholds that define categories. This technique was independently developed by Crittenden (1961) and Falconer (1965; Lynch & Walsh, 1998, p. 730). The estimates obtained with ordinal data analysis should be unbiased, but the analyses may be computationally more demanding, especially when the number of categories is large. Another disadvantage is the potential presence of empty cells in the contingency tables. For example, the contingency table of a highly heritable trait in MZ twins is likely to have some empty off-diagonal cells.

The purpose of this study is to examine the effect of censoring on the results of genetic modeling. We assume that a latent trait is normally distributed, and that censoring arises due to failure of the measurement instrument to detect values smaller than some general threshold y* (see Figure 2). Three methods which may be used to deal with nonnormal data are compared in a simulation study. Two of these methods concern ad hoc transformations: a square root transformation and the computation of normal scores. The third method is the analysis of categorical data which is based on the liability threshold model. Finally, we apply these methods to real-life data on sleep problems in a large sample of 3-year-old twins.

Methods

Genetic Modeling

Variation in a phenotypic trait can be decomposed into latent genetic and environmental components. The decomposition of variance may be achieved by analyzing data of pairs of individuals who differ in their degree of genetic relatedness. The twin design is a well-known example of this approach. Monozygotic (MZ) twins are genetically identical, while dizygotic (DZ) twins on average share half of their segregating genes. Limiting the genetic decomposition of phenotypic variance to additive genetic (A) effects and dominant genetic (D) effects, the fact that MZ twins are genetically identical implies that they share all the additive and dominant genetic variance. DZ twins on average share half of the additive genetic and one quarter of the dominant genetic variance (Lynch & Walsh, 1998). The environmental phenotypic variance may be decomposed into shared environmental variance and unique environmental variance. The environmental effects shared by two members of a twin pair (C) are by definition perfectly correlated in both MZ and DZ twins. The nonshared environmental effects (E) are by definition uncorrelated between twin pair members. Estimates of the nonshared environmental variance usually include measurement error (Plomin et al., 2001; Neale & Cardon, 1992). In fitting models to twin data, it is not possible to estimate the effects of all components of variance (Va, Vd, Vc and Ve) simultaneously. Specifically, one cannot estimate Vd and Vc simultaneously due to reasons of identification.

Simulation Study

Data were simulated in accordance with three models. In Models 1 and 2, the covariance structure of MZ and DZ twins was attributable to A, C and E. The values of Va, Vc and Ve in Model 1 equaled .50, .20 and .30, respectively. In Model 2, the values of Va, Vc and Ve equaled .20, .50 and .30. In Model 3, the covariances were influenced by A, D and E. The values of the variance components Va, Vd and Ve equaled .45, .25 and .30.

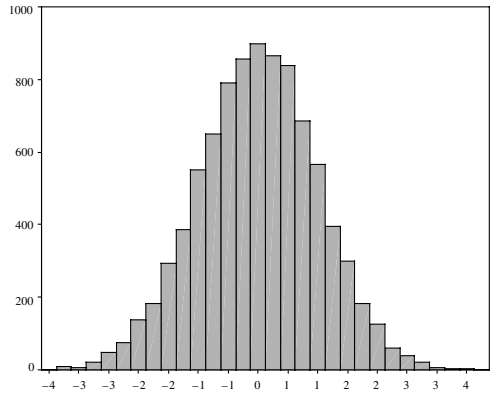
The population covariance matrices of MZ and DZ twins can be calculated under assumption of these three theoretical models. We assumed that there is no assortative mating, epistasis, gene–environment interaction or gene–environment correlation (Lynch & Walsh, 1998). Under these assumptions, the covariances of MZ twins are Va + Vc + Vd, and the covariances of DZ twins are .5*Va + Vc + .25*Vd. The variances equal Va + Vc + Vd + Ve in both MZ and DZ twins. In the simulation study, the covariances of MZ and DZ twins were .70 and .45 (Model 1), .70 and .60 (Model 2), and .70 and .2875 (Model 3), respectively. All variances equaled 1.

In most European countries, the number of DZ twins is larger than the number of MZ twins. For example, in the Netherlands Twin Register, the number of DZ twins is about twice the number of MZ twins. Therefore, the number of DZ twins in the simulation study was also twice the number of MZ twins. We simulated data of 3000 MZ twin pairs and 6000 DZ twin pairs. This sample size is representative for the sample size of twin registers such as the Netherlands Twin Register (Boomsma, 1998; Boomsma et al., 2002). The simulation study comprised 1000 replicates.

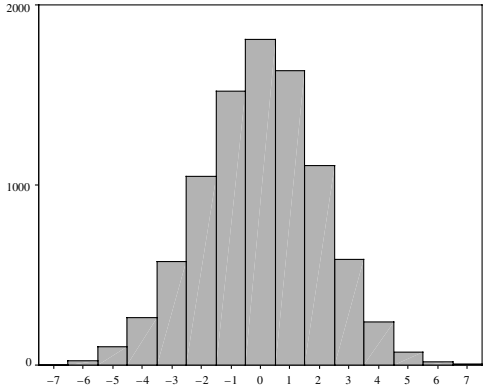
Data simulation was performed in R (Venables et al., 2002). Six different data sets were generated. The distributions of these data sets are shown in Figure 3a to Figure 3f. First, bivariate standard normal distrib-

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

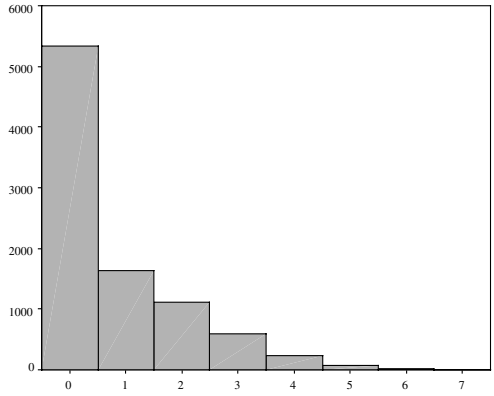
a. Normal distributed scores



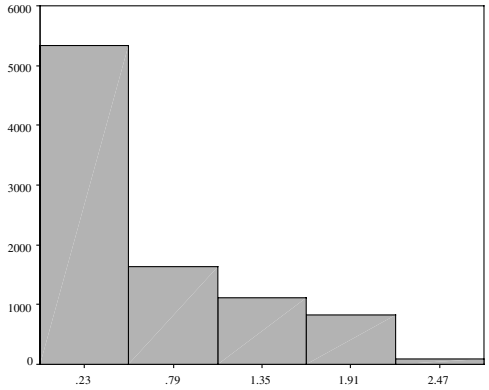
b. Discretized normal distributed scores



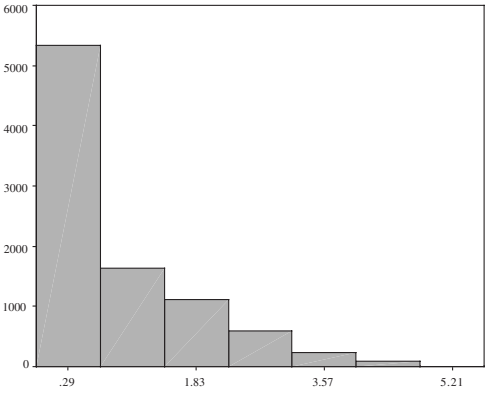
c. Censored scores



d. Square root transformed censored scores



e. Normal scores of censored data



f. Categorical scores of censored data

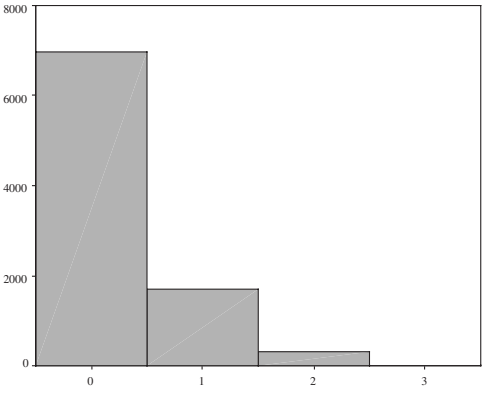


Figure 3

Distributions of six simulated data sets.

Data sets were generated from a bivariate normal distribution (a) and then discretized (b). Next, data were censored (c). These censored data were either transformed by square root transformation (d) or to normal scores (e). Finally, the number of categories of the censored scores was limited to four (f).

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

uted data (Figure 3a) were simulated with the MASS package that is available in the R library. Second, because observational data are usually discrete, the multivariate normal data were discretized (Figure 3b). The number of categories was 15. The values of the categories were chosen arbitrarily: a value of –7 was assigned to the lowest category, and a value of +7 was assigned to the highest category.

Third, the discrete data were censored (Figure 3c). The value of the censor was equal to 0. The values of all data points below 0, which made up 39% of the total data set, were therefore reassigned to 0 in the censored data set. This percentage was chosen because the resulting distribution resembled the distribution of behavioral checklist data in terms of skewness and kurtosis. The fourth and fifth data sets were created by applying two transformations to the censored data. The first transformation was a square root transformation (Figure 3d). The second transformation was the computation of normal scores (Figure 3e). This transformation renders the skewness and kurtosis of the data as close as possible to 0 and 3, the expected values of the skewness and kurtosis when the distribution is normal. The computation of normal scores is implemented in Prelis (Jöreskog & Sörbom, 1996b). For this simulation study, the Prelis procedure was implemented in R. The R syntax is available on request. Finally, the number of categories of the censored data was decreased to four (Figure 3f). In contrast to the five other data sets, where we applied normal theory ML, these data were treated as categorical data. All analyses were performed on raw data.

Genetic Analyses of Sleep Problems

Participants

The participants were all registered at birth with the Netherlands Twin Registry (Boomsma, 1998; Boomsma et al., 2002). In the present study, we have assessed a sample of Dutch twin pairs whose mothers reported on their sleep problems when the twins were three years old. These twins were all born between 1986 and 1997. The sample used for the genetic analyses consisted of 6375 MZ twins and 12,192 DZ twins. Zygosity diagnosis was assessed with the use of a 10-item questionnaire. This procedure allows an accurate determination of zygosity of nearly 95% (Rietveld et al., 2000). For a more detailed description of the sample, see Derks et al. (2004).

Measure

The Child Behavior Checklist (CBCL/2–3) is a standardized questionnaire for parents to report the frequency and intensity of behavioral and emotional problems exhibited by their child in the past six months (Achenbach, 1992). It contains 100 items that measure problem behavior; the items are rated on a 3-point scale ranging from 'not true', 'somewhat or sometimes true' to 'very true or often true'. The CBCL measures the number of symptoms on seven behavioral syndromes, including sleep problems (7 items; Koot et al., 1997). The distribution of sleep problems is shown in Figure 1 (first-borns only to save space).

Results

Simulation Study

The descriptive statistics of the simulated data before and after transformation are reported in Table 1. These descriptives are reported for one replication and a single twin only. As expected, the skewness and kurtosis of the noncensored continuous data did not deviate significantly from the expected values of 0 and 3. After discretization, the variance increased as a result of the larger range of values but the skewness and kurtosis were unaffected. After censoring, the skewness and kurtosis were both positive and deviated significantly from 0 and 3. In addition, the mean of the data increased and the variance decreased in comparison to the noncensored discrete data. Both the square root transformed data and the normal scores showed less skewness and kurtosis than the untransformed censored data, but their values still deviated significantly from 0 and 3.

Table 1

Descriptive Statistics of Simulated Data

Data set	Mean	Standard deviation	Skewness	Kurtosis
1. Noncensored continuous data	.000	.996	–.043	2.985
2. Noncensored discrete data (15 categories)	.004	2.012	–.053	2.991
3. Censored discrete data (8 categories)	.790	1.172	1.603	5.293
4. Square root transformed censored data	1.639	.317	1.277	3.779
5. Normal scores of censored data	.786	1.172	.946	2.794
6. Categorical data	Threshold 1 = 0.755			
(4 categories based on censored data)	Threshold 2 = 1.787			
	Threshold 3 = 2.748			

Note: Data sets were generated from a bivariate normal distribution (1) and then discretized (2). Next, data were censored (3). These censored data were either transformed by square root transformation (4) or to normal scores (5). For the 6th data set (four categories based on censored data), the thresholds are given. The number of replications is 1000, but descriptives are given for a single replication and for a single twin only. The number of twins is 9000.

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

Table 2a

Estimates of Standardized Genetic and Environmental Influences in the Six Simulated Data Sets, Averaged Over 1000 Replications

Data set	Method of analysis	Means stand. Va (*SD*)	Mean stand. Vc (*SD*)	Mean stand. Ve (*SD*)
1. Noncensored continuous data	Normal theory	.501	.199	.300
	ML	(.025)	(.021)	(.008)
2. Noncensored discrete data (15 categories)	Normal theory	.491	.195	.314
	ML	(.025)	(.021)	(.008)
3. Censored discrete data (8 categories)	Normal theory	.496	.116	.389
	ML	(.040)	(.031)	(.014)
4. Square root transformed censored data	Normal theory	.488	.122	.390
	ML	(.037)	(.029)	(.014)
5. Normal scores of censored data	Normal theory	.471	.128	.401
	ML	(.036)	(.029)	(.013)
6. Categorical data	Categorical data analysis	.501	.200	.300
(4 categories based on censored data)[a]		(.053)	(.042)	(.019)

Note: True values of Va, Vc and Ve are .50, .20 and .30 respectively.
Stand. = standardized; Va = variance explained by additive genetic effects; Vc = variance explained by common environmental effects; Ve = variance explained by unique environmental effects.
[a] The results of the categorical data analyses are based on 999 replications due to minimalization problems in 1 replication.

Table 2b

Estimates of Standardized Genetic and Environmental Influences in the Six Simulated Data Sets, Averaged Over 1000 Replications

Data set	Method of analysis	Means stand. Va (*SD*)	Mean stand. Vc (*SD*)	Mean stand. Ve (*SD*)
1. Noncensored continuous data	Normal theory	.200	.500	.300
	ML	(.020)	(.016)	(.008)
2. Noncensored discrete data (15 categories)	Normal theory	.196	.490	.314
	ML	(.021)	(.017)	(.009)
3. Censored discrete data (8 categories)	Normal theory	.205	.406	.389
	ML	(.036)	(.027)	(.015)
4. Square root transformed censored data	Normal theory	.203	.407	.390
	ML	(.034)	(.025)	(.014)
5. Normal scores of censored data	Normal theory	.194	.404	.402
	ML	(.034)	(.025)	(.014)
6. Categorical data	Categorical data analysis	.199	.501	.300
(4 categories based on censored data)[a]		(.048)	(.036)	(.018)

Note: True values of Va, Vc and Ve are .20, .50 and .30 respectively.
Stand. = standardized; Va = variance explained by additive genetic effects; Vc = variance explained by common environmental effects; Ve = variance explained by unique environmental effects.
[a] The results of the categorical data analyses are based on 999 replications due to minimalization problems in 1 replication.

Table 2c

Estimates of Relative Genetic and Environmental Influences in the Six Simulated Data Sets, Averaged Over 1000 Replications

Data set	Method of analysis	Means stand. Va (*SD*)	Mean stand. Vc (*SD*)	Mean stand. Ve (*SD*)
1. Noncensored continuous data	Normal theory	.448	.252	.300
	ML	(.047)	(.049)	(.008)
2. Noncensored discrete data (15 categories)	Normal theory	.438	.248	.248
	ML	(.047)	(.049)	(.008)
3. Censored discrete data (8 categories)	Normal theory	.273	.338	.389
	ML	(.059)	(.063)	(.014)
4. Square root transformed censored data	Normal theory	.286	.323	.391
	ML	(.057)	(.061)	(.013)
5. Normal scores of censored data	Normal theory	.299	.300	.402
	ML	(.056)	(.060)	(.013)
6. Categorical data	Categorical data analysis	.446	.254	.300
(4 categories based on censored data)[a]		(.087)	(.093)	(.018)

Note: True values of Va, Vd and Ve are .45, .25 and .30 respectively.
Stand. = standardized; Va = variance explained by additive genetic effects; Vd = variance explained by dominant genetic effects; Ve = variance explained by unique environmental effects.
[a] The results of the categorical data analyses are based on 994 replications due to minimalization problems in 6 replications.

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1375/twin.7.6.659

Table 3

A Comparison of Statistical Power in the Six Simulated Data Sets

Data set	Model 1		Model 2		Model 3
	ACE–AE	ACE–CE	ACE–AE	ACE–CE	ADE–AE
	Mean –2LL (*SD*)	Mean –2LL (*SD*)	Mean –2LL (*SD*)	Mean –2LL(*SD*)	Mean –2LL (*SD*)
Theoretical population covariance matrices	83.347 (18.204)	365.304 (38.200)	625.616 (50.005)	83.620 (18.234)	28.053 (10.498)
1. Noncensored continuous data	83.583 (17.712)	367.988 (35.632)	627.045 (45.754)	84.245 (17.342)	29.506 (10.787)
2. Noncensored discrete data (15 categories)	78.391 (17.017)	335.681 (33.895)	582.241 (44.798)	75.763 (16.447)	27.927 (10.486)
3. Censored discrete data (8 categories)	25.436 (12.881)	259.637 (42.549)	336.466 (47.447)	59.437 (20.792)	45.865 (16.588)
4. Square root transformed censored data	27.583 (12.751)	250.526 (39.616)	335.946 (43.172)	57.857 (19.375)	41.761 (15.283)
5. Normal scores of censored data	29.410 (12.735)	228.906 (35.763)	324.412 (41.969)	51.337 (17.443)	35.741 (13.786)
6. Categorical data (4 categories based on censored data)[a]	23.298 (9.545)	84.15322 (18.951)	167.982 (27.309)	17.661 (11.028)	8.705 (5.707)

Note: The theoretical values of –2LL are based on analysis of the theoretical population covariance matrices. The number of twin pairs is 9000, and the number of replications is 1000.

–2LL = minus 2 log likelihood.

The true model parameters in Model 1: Va = .50, Vc = .20, Ve = .30; Model 2: Va2 = .20, Vc = .50, Ve = .30; Model 3: Va = .45, Vd = .25, Ve = .30

[a] In the categorical data analysis, the number of replications was 999, 999 and 996 for model 1 to 3 respectively.

The Parameter Estimates of the ACE Models

Table 2a and Table 2b show the mean point estimates of the standardized variance components and their standard deviations in the 1000 replications. The mean point estimate of the categorical data analyses was based on slightly fewer than 1000 replications, because the minimalization of the likelihood failed in one of the replications.

As expected, the analysis of noncensored continuous data produced the correct mean parameter estimates in both models. Discretization did not affect the mean or standard deviation of the standardized parameter estimates. After censoring, the estimate of Va was unbiased but Vc was underestimated and Ve was overestimated. A square root transformation or a transformation to normal scores did not improve the parameter estimates. In contrast, when the categorical data were analyzed using the threshold model, the correct parameter estimates were recovered. However, as is to be expected given the reduced amount of information, the standard errors of the parameter estimates increased which results in wider confidence intervals and less precise estimates.

The Parameter Estimates of the ADE Model

The results of the ADE model (Table 2c) are in agreement with the results of the ACE models. The noncensored continuous data and the noncensored discrete data both recovered the correct parameter estimates. The analyses of the censored untransformed data, the square root transformed data, and the normal scores lead to biased parameter estimates. Va was underestimated, while Vd and Ve were both overestimated. When the categorical data option in Mx was used, the unbiased parameter estimates were obtained, but again with increased standard errors. The mean point estimate of the categorical data analyses was again based on slightly less than 1000 replications, because the minimalization of the likelihood failed in six of the replications.

The underestimation of Va in the ADE model was large compared to the other deviations. While the underestimation of Vc in the ACE model and the overestimation of Vd and Ve in the ADE model varied between 5% and 10% of the variance, the underestimation of Va in the ADE model was about 20%.

Power Analyses

One of the desirable features of an estimation method is that it should produce unbiased parameter estimates. Another important feature is statistical power. In this section, we compare the power of the different methods. To this end, we compared the fit of the true ACE model to the fit of an AE model and the fit of a CE model. We also compared the fit of the ADE model to the fit of an AE model. We did not compare the fit of the ADE model to the fit of a DE model because the presence of dominant genetic influences in the absence of additive genetic influences is biologically implausible (Falconer & Mackay, 1996).

Table 3 shows the results of the power analyses. We have used a type-I error rate of .05. Because of the large sample size we are not interested in power per se, but in the effect of the estimation method on power. In Table 3, we first report the theoretical value of the difference in –2LL and its standard deviation.

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

Table 4

Descriptive Statistics of Maternal Child Behavior Checklist Reports on Sleep Problems in 9415 Three-Year-Old Dutch Twins (First-Borns Only)

Data set	Mean	Standard Deviation	Skewness	Kurtosis
Raw data (15 categories)	1.983	2.229	1.444	5.220
Square root transformed data	1.083	.900	.165	1.990
Normal scores	2.074	2.053	.540	2.613

The theoretical value of the difference in –2LL was determined by analyzing the population covariance matrices in Mx. It is equal to the number of degrees of freedom (*df*) plus the noncentrality parameter (λ). The standard deviation was calculated with the following formula: $SD = (2\ (df + 2 * \lambda))^{0.5}$. As can be seen in Table 3, the mean –2LL of the continuous data analyses was quite similar to the theoretical value of –2LL.

It is important to realize that the values of the noncentrality parameter can only be interpreted in terms of null and nonnull distributions of the likelihood ratio test in the case of the normally distributed data (continuous or 15-point scale), and in the case of the categorical data estimator. For example, the results observed in the case of Model 3 seem to suggest that the power increased after censoring (e.g., from 28.053 to 45.865). However, this is due to the fact that the test statistics do not follow their expected noncentral and central chi-square distributions. This is also true in case of the transformed censored data.

After discretization, the mean difference in –2LL decreased slightly. This is a reflection of the decreased power due to a loss of information after discretization. In all three models, the power was lowest when the categorical data were analyzed. This is evident in the low mean difference in –2LL. In addition, when we look at the categorical data analyses of the ADE model, the drop of the D parameter did not lead to a significantly worse fit in 20% of the cases, although this parameter explained 25% of the variance. In other words, the power to detect a dominant genetic parameter that explains 25% of the variance is 80%. In comparison, the power is 100% when any of the other methods of analysis is chosen.

Genetic Analyses of Sleep Problems

To illustrate the previous findings, we analyzed data on sleep problems in 6375 MZ twins and 12,192 DZ twins. The descriptive statistics are summarized in Table 4. These descriptives are only reported for the first-borns to save space; the descriptive statistics of the second-borns are similar. The skewness and kurtosis of the raw scores are similar to the skewness and kurtosis after censoring in the simulation study. A square root transformation and the computation of normal scores both reduced the skewness and kurtosis.

The correlations were computed in four different ways, namely the Pearson product moment correlation (ppmc) of the untransformed raw scores, the square root transformed scores, and the normal scores. In addition, polychoric correlations of the categorical data were computed. The estimates are shown in Table 5. The ppmc's were quite similar, but the polychoric correlations were somewhat higher in both MZ and DZ twins.

Based on the correlations, an ACE model seemed to be most plausible. The MZ correlation was slightly less than twice the DZ correlation, which implies the absence of Vd and a small contribution of Vc. Genetic model fitting analyses of untransformed, transformed and categorical data showed that the influences of A and C were both significant. Table 6 shows the point estimates and the confidence intervals of the standardized variance components (Va, Vc and Ve) on sleep problems. The estimate of Va was similar across methods, which was to be expected in the light of the results of the simulation study. The estimate of Vc ranged from .055 to .081 when normal theory ML was used. The estimate was .116 in the categorical data analysis. In contrast, the estimate of Ve was lower in the categorical data analyses. As expected, the categorical data analyses showed wider confi-

Table 5

Twin Correlations for Maternal Child Behavior Checklist Reports on Sleep Problems

Data set	Correlation MZ (*N* = 3162 complete pairs)	Correlation DZ (*N* = 6053 complete pairs)
Raw scores (ppmc)	.745	.384
Square root transformed data (ppmc)	.741	.408
Normal scores (ppmc)	.748	.406
Categorical data (pc)	.786	.451

Note: ppmc = Pearson product moment correlation; pc = polychoric correlation

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

Table 6

Estimates and 95% Confidence Intervals of Standardized Estimates of Genes and Environment on Maternal Child Behavior Checklist Reports of Sleep Problems

Data set	Va (low–high)	Vc (low–high)	Ve (low–high)
Raw scores	.676 (.628–.724)	.055 (.011–.097)	.270 (.256–.284)
Square root transformed data	.658 (.611–.705)	.081 (.038–.122)	.262 (.248–.276)
Normal scores	.665 (.618–.712)	.077 (.034–.118)	.258 (.245–.272)
Categorical data	.675 (.620–.733)	.116 (.064–.166)	.210 (.194–.227)

Note: Va = proportion of variance explained by additive genetic effects; Vc = proportion of variance explained by common environmental effects; Ve = proportion of variance explained by unique environmental effects.

dence intervals than the analyses based on normal theory ML.

Discussion

This paper deals with the effects of censoring on parameter estimates and statistical power in genetic analyses of quantitative traits. The censoring of normal distributed data results in data with an L-shaped distribution. The distribution resembles the distribution of most behavioral checklist data. This paper looks at the effects of censoring through a series of simulations. Data were simulated in accordance with three genetic models: two ACE models with different factor loadings of A, C and E, and one ADE model.

Multivariate normal data were simulated and discretized because behavior checklist data are usually discrete. We replicated the finding of Dolan (1994) that discretization of normal distributed data does not lead to biased parameter estimates when the number of categories is seven or more and when the distribution is symmetric. Next, the simulated data were censored, which resulted in L-shaped distributions. When analyzing the censored ACE data with normal theory ML, the common environmental component was underestimated while the unique environmental component was overestimated. Transformation of the data did not eliminate this bias, although the skewness and kurtosis decreased. Interestingly, a common finding in behavioral genetic studies is a small influence of shared environment on individual differences in behavior. This may partly be due to the fact that the influence of this component is underestimated when L-shaped data are analyzed with normal theory ML. However, the underestimation of the relative influence of the additive genetic component was only 8% to 10%. When analyzing the ADE data, a quite large underestimation of about 20% of the additive genetic component was found and both D and E were overestimated by about 10%.

In order to examine if these results hold when a smaller percentage of the data is censored, we examined the amount of bias in parameter estimates when 10% of the data set was censored instead of 39%. In this situation, the bias decreased and ranged from 3% in the ACE model to 5% in the ADE model (data not shown). Thus, depending on the level of censoring, the results of twin studies which have used normal theory ML to analyze L-shaped distributed data may be biased.

The bias in parameter estimates may be avoided by using categorical data analysis. However, this method has three disadvantages. First, the statistical power is reduced. This result is in agreement with the results of the simulation study of Neale et al. (2004) who found that in categorical data analysis approximately three times the sample size was needed for equivalent power to continuous data analysis. In our study, the decrease in power was most apparent when the simulated ADE data were analyzed. Even with 9000 twin pairs, the power to detect a dominant genetic component that explains 25% of the variation, decreased with 20% (from 100% to 80%). However, one should realize that the type-II error rate may be lower when censored data are analyzed with normal theory ML compared to categorical data analysis, but that the actual type-I error rate may be higher than the hypothesized value of .05. One way to deal with the low power, is to choose a type-I error rate of .10 or .15 instead of .05. Second, the analyses are computationally more demanding. This problem may be solved by using Weighted Least Squares (WLS) in Lisrel (Jöreskog & Sörbom, 1996a). However, this method has the disadvantage that missing data are excluded which can be a problem when dealing with incomplete twin data or with longitudinal data in which observations may be missing at some time-points. A third disadvantage is that the contingency tables may have empty cells. One remedy to the presence of empty cells is to decrease the number of categories.

To illustrate the results of the simulation study, we analyzed real-life data on sleep problems. The skewness and kurtosis were similar to the skewness and kurtosis of the simulated data. In this example, the small common environmental influence explained enough variance to be detected in the categorical data analysis. The heritability was quite stable and ranged from 66 to 68%. The estimate of the common environmental influence was somewhat higher in the categorical data analysis (12%) than in the other analyses (6% to 8%). Based on the results of the sim-

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

ulation study, we can conclude that the estimate of 12% in the categorical analysis, is the correct estimate. The unique environmental influence ranged from 21% to 27%. In conclusion, sleep problems are, like other behavioral problems in young children, explained by large genetic influences and moderate environmental influences. The latter include shared environmental influences.

The main question that we addressed was: What is the best approach when analyzing L-shaped distributed phenotypic data? The results of the simulation study show that the analysis of L-shaped distributed data with normal theory ML is not advisable when the data show high skewness and kurtosis. Transformations may reduce the skewness and kurtosis but do not eliminate the bias in parameter estimates. Categorical data analysis is a better option, because this is the only method with which unbiased parameter estimates are obtained. Because this estimation method has its own limitations, the best option would be to develop checklists that measure variation in the whole latent distribution of behavior. To this end, items should reflect both well-adapted and dysfunctional behavior.

Acknowledgments

This work was supported by NWO Grant numbers 575–25–006 and 904–57–94 (Boomsma, P. I.), and by NIMH Grant number MH58799 (Hudziak, P. I.).

References

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2–3 and 1992 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Amos, C. I. (1994). Robust Variance–Components approach for assessing genetic linkage in pedigrees. *American Journal of Human Genetics, 54*, 535–543.

Bielak, L. F., Sheedy, P. F., II, & Peyser, P. A. (2001). Coronary artery calcifications measured at electron-beam CT: Agreement in dual scan runs and change over time. *Radiology*, *218*, 224–229.

Boomsma, D. I. (1998). Twin registers in Europe: An overview. *Twin Research, 1*, 34–51.

Boomsma, D. I., Vink, J. M., Van Beijsterveldt, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M., Derks, E. M., Riese, H., Willemsen, A. H. M., Bartels, M., Van den Berg, M., Kupper, H. M., Polderman, J. C., Posthuma, D., Rietveld, M. J. H., Stubbe, J. H., Knol, L. I., Stroet, T., & Van Baal, G. C. M. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research, 5*, 401–406.

Crittenden, L. B. (1961). An interpretation of familial aggregation based on multiple genetic and environmental factors. *Annals of the New York Academy of Sciences, 91*, 769–780.

Derks, E. M., Hudziak, J. J., Van Beijsterveldt, C. E. M., Dolan, C. V., & Boomsma, D. I. (2004). A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. *Behavior Genetics, 34*, in press.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimates using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*, 309–326.

Epstein, M. P., Lin, X., & Boehnke, M. (2003). A Tobit Variance–Component method for linkage analysis of censored trait data. *American Journal of Human Genetics, 72*, 611–620.

Falconer, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics, 29*, 51–71.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Harlow, UK: Pearson Education.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff.

Jöreskog, K. G., Sörbom, D., Du Toit, S., & Du Toit, M. (1999). *LISREL 8: New statistical features*. Chicago, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1996a). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1996b). *PRELIS 2 User's reference guide*. Chicago, IL: Scientific Software International.

Koot, H. M., van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1997). Behavioural and emotional problems in young preschoolers: Cross-cultural testing of the validity of the Child Behavior Checklist/2–3. *Journal of Abnormal Child Psychology, 25*, 183–196.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

Muthén, B., & Muthén, L. (2001). *Mplus user's guide*. Los Angeles, CA: Muthén and Muthén.

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70*, 1181–1196.

Neale, M. C. (1997). *Mx: Statistical modeling*. Richmond, VA: Department of Psychiatry, Medical College of Virginia.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. North Atlantic Treaty Organization, Scientific Affairs Division. Dordrecht, the Netherlands: Kluwer Academic.

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659

Neale, M.C., Eaves, L. J., & Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics, 24,* 239–258.

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

Prescott, C. A. (2004). Using the Mplus computer program to estimate models for continuous and categorical data from twins. *Behavior Genetics, 34,* 17–40.

Rietveld, M. J. H., Hudziak, J. J., Bartels, M., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2003). Heritability of attention problems in children: I. Cross-sectional results from a study of twins, age 3–12 years. *American Journal of Medical Genetics, 117B,* 102–113.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Journal of the Econometric Society, 26,* 24–36.

van den Oord, E. J. C. G., Pickles, A., & Waldman, I. (2003). Normal variation and abnormality: An empirical study of the liability distributions underlying depression and delinquency. *Journal of Child Psychology & Psychiatry, 44,* 180–192.

Venables, W. N., Smith, D. M., & the R Development Core Team. (2002). *An introduction to R*. Bristol, UK: Network Theory.

Downloaded from https://www.cambridge.org/core. Open University Library, on 27 Jan 2020 at 16:00:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms . https://doi.org/10.1375/twin.7.6.659