Behav Genet (2006) 36:935–946
DOI 10.1007/s10519-006-9086-3

ORIGINAL PAPER

Precision and Bias of a Normal Finite Mixture Distribution Model to Analyze Twin Data When Zygosity is Unknown: Simulations and Application to IQ Phenotypes on a Large Sample of Twin Pairs

Beben Benyamin · Ian J. Deary · Peter M. Visscher

Received: 14 November 2005 / Accepted: 18 May 2006 / Published online: 20 June 2006

Abstract The classification of twin pairs based on zygosity into monozygotic (MZ) or dizygotic (DZ) twins is the basis of most twin analyses. When zygosity information is unavailable, a normal finite mixture distribution (mixture distribution) model can be used to estimate components of variation for continuous traits. The main assumption of this model is that the observed phenotypes on a twin pair are bivariately normally distributed. Any deviation from normality, in particular kurtosis, could produce biased estimates. Using computer simulations and analyses of a wide range of phenotypes from the U.K. Twins' Early Developments Study (TEDS), where zygosity is known, properties of the mixture distribution model were assessed. Simulation results showed that, if normality assumptions were satisfied and the sample size was large (e.g., 2,000 pairs), then the variance component estimates from the mixture distribution model were unbiased and the standard deviation of the difference between heritability estimates from known and unknown zygosity in the range of 0.02–0.20. Unexpectedly, the estimates of heritability of 10 variables from TEDS using the mixture distribution model were consistently larger than those from the conventional (known zygosity) model. This discrepancy was due to violation of the bivariate normality assumption. A leptokurtic distribution of pair difference was observed for all traits (except non-verbal ability scores of MZ twins), even when the univariate distribution of the trait was close to normality. From an independent sample of Australian twins, the heritability estimates for IQ variables were also larger for the mixture distribution model in six out of eight traits, consistent with the observed kurtosis of pair difference. While the known zygosity model is quite robust to the violation of the bivariate normality assumption, this novel finding of widespread kurtosis of the pair difference may suggest that this assumption for analysis of quantitative trait in twin studies may be incorrect and needs revisiting. A possible explanation of widespread kurtosis within zygosity groups is heterogeneity of variance, which could be caused by genetic or environmental factors. For the mixture distribution model, violation of the bivariate normality assumption will produce biased estimates.

Keywords Heritability · Intelligence · Kurtosis · Mixture distribution · TEDS · Zygosity

Edited by Dorret Boomsma

B. Benyamin · P. M. Visscher
Institute of Evolutionary Biology, University of Edinburgh,
Ashworth Laboratories, King's Buildings,
West Mains Road, Edinburgh EH9 3JT, UK

B. Benyamin (✉) · P. M. Visscher
Genetic Epidemiology, Queensland Institute of Medical
Research, 300 Herston Road, Brisbane 4029
Queensland, Australia
e-mail: bebenB@qimr.edu.au

I. J. Deary
Department of Psychology, University of Edinburgh,
7 George Square, Edinburgh EH8 9JZ, UK

Introduction

The classical twin design is very useful in partitioning the observed phenotypic variance of complex traits in humans into genetic and environmental components (reviewed by Boomsma et al. 2002). By comparing the resemblance of monozygotic (MZ) twin pairs to that of dizygotic (DZ) twin pairs, twin studies allow the causes of individual differences in complex traits to be quantified. Under the assumption that both types of twins share the same degree of common environmental experiences (the common environment assumption), a larger similarity of MZ pairs compared to DZ pairs indicates that genetic factors influence phenotypic variation (e.g., Evans et al. 2002; Rijsdijk and Sham 2002). The classification of twins based on zygosity is crucial in twin studies. A standard zygosity questionnaire (e.g., Peeters et al. 1998) answered by twins or their parents is usually used to diagnose zygosity. With the advance of molecular genetic markers, such as microsatellites, DNA-based zygosity testing is now widely used and gives a greater accuracy (e.g., Forget-Dubois et al. 2003).

Although zygosity information can now be easily and economically obtained, such information is not always available. Two examples are twin data that were collected before zygosity classification was routine (e.g., the Scottish Mental Surveys 1932 and 1947, Deary et al. 2004; Scottish Council for Research in Education 1949) and data collected from large national studies in the fields of social sciences, economics or education where genetic study was not the main interest (e.g., Scarr-Salapatek 1971). Twins from these studies can be identified by matching a pair with, for example, the same surname, birth date and location such as home address or school, if these identifiers are available. Assuming that an identifier of sex is also available, twin pairs from such studies can only be classified as same sex (SS) or opposite sex (OS) pairs. SS pairs are a mixture of MZ and DZ pairs whereas OS pairs are always DZ. For such studies, the conventional methods which rely on zygosity information cannot be used. Different methods have been proposed to analyze twin data where zygosity information is unavailable. Scarr-Salapatek (1971) estimated the correlations of MZ and DZ pairs by partitioning the z-transformed correlation coefficient of SS twins (Benyamin et al. (2005) describe an analogous method based upon ANOVA). The method, however, assumed that the sample size and correlation of DZ SS twins were the same as those of the observed OS pairs, and is limited to univariate heritability (Neale 2003). The OS correlation can substantially differ from the correlation of DZ SS, for example if the genetic or common environmental covariance is lower in OS pairs.

Neale (2003) proposed a method based upon a normal finite mixture distribution (mixture distribution) to estimate MZ and DZ correlations from SS twins. This method partitions the SS twin distribution into underlying MZ and DZ distributions by maximum likelihood. The estimated proportion of MZ among SS twins (pMZ) is used to weight the likelihood. This method has been applied to analyze individual differences in cognitive ability (the Moray House Test No. 12) from twin data with unknown zygosity of the Scottish Mental Surveys 1932 and 1947 (Benyamin et al. 2005). In addition, Heath et al. (2003) proposed a latent class analysis to diagnose zygosity. This method can be used to analyze discrete data on twins by fitting a 2-class latent class model, which is assumed to correspond to MZ and DZ pairs (Benyamin et al. 2005).

The mixture distribution model of Neale (2003) assumes that the observed phenotypes on a pair follow a bivariate normal distribution in the population. Any deviation from normality, in particular kurtosis, could produce biased estimates because the partitioning of the observed within-pair and between-pair variation is based upon the contrast of the variance and kurtosis (Benyamin et al. 2005).

The purpose of the present study is to quantify the precision and bias of the mixture distribution model in estimating genetic parameters from twin data when zygosity is unknown. Simulation was used to quantify the precision of estimation of the mixture distribution model when the distributional assumptions were met, and to quantify bias when normality assumptions were violated. Finally, we applied the known zygosity and mixture distribution models to a range of IQ phenotypes from the U.K. Twins' Early Development Study (TEDS), a longitudinal study of a representative sample of all twins born in England and Wales between 1994 and 1996. Zygosity information is available on TEDS data. Therefore, the application of the mixture distribution model to these data afforded a check on variance components estimates from our previous application of the mixture distribution model on twins of unknown zygosity of cognitive ability from the Scottish Mental Surveys 1932 and 1947 (Benyamin et al. 2005).

Methods

Simulation Study

MZ and DZ twin data were simulated using a standard ACE model of family resemblance, by sampling additive genetic (A), common environmental (C) and specific environmental (E) effects. No sex effects or other fixed effects were simulated, and for subsequent analyses it was assumed that there were only SS pairs. All simulations were replicated 1,000 times.

To assess the precision of estimation of the mixture distribution model, twin data were first simulated under the assumed bivariate normal distribution. Nine different standardized variance component parameters (Table 1) were simulated for different sample sizes (500, 2,000 and 5,000 twin pairs, with equal proportions of MZ and DZ). Each simulated dataset was then analyzed with the conventional (known zygosity) and mixture distribution (Neale 2003) models, using the statistical package Mx (Neale et al. 2002). An overall mean was the only fixed effect fitted in the model.

Although pMZ could in principle be estimated from the data when fitting a mixture distribution model, the estimate is very imprecise (result not shown). Therefore, in the mixture distribution model, an a priori estimate of pMZ is used to weight the likelihood. In a complete population survey, pMZ can be estimated using Weinberg's differential rule as 1–2 × (proportion of OS twin pairs) (Weinberg 1902). This formula assumes that the number of DZ SS twins is the same as DZ OS twins due to the distribution of sexes (Scarr-Salapatek 1971). This proportion may not be accurately estimated in all studies. Therefore, in order to assess whether specifying a wrong proportion in the mixture distribution model has an effect on variance components estimation, different proportions (0.1–0.9) were used in the analyses when the true proportion was 0.5. For this simulation, the standardized *A*, *C* and *E* variance components were $a^2 = 0.50$, $c^2 = 0.25$, $e^2 = 0.25$ and simulations were based on 2,000 twin pairs with an equal proportion of MZ and DZ.

In order to assess the effect of kurtosis on the parameters' estimation of the mixture distribution model, normally distributed twin data were transformed into a distribution with a desired kurtosis value using the Cornish–Fisher expansion (Cornish and Fisher 1937). For each value of an individual (x) drawn for a normal distribution, the transformation is:

$$y = x + \frac{c}{24}(x^3 - 3x),$$

where y is the transformed x with desired kurtosis given by the coefficient c. For positive c smaller than 1, the simulated data has a distribution with the kurtosis value similar to c. For larger positive c, the kurtosis value for the transformed distribution was larger than c. On the other hand, for negative c, the kurtosis value for the transformed distribution was slightly smaller than c. For examples, the corresponding average kurtosis values for c of −2, −1, −0.75, −0.50, −0.25, 0.25, 0.50, 0.75, 1 and 2 were −1.01, −0.70, −0.57, −0.42, −0.23, 0.27, 0.60, 0.99, 1.50 and 4.15, respectively. The exact relationship between the value of c and the kurtosis value of the transformed data is shown in Appendix 1. Data sets for different c (11 different values of c ranging from −2 to 2) were simulated. The standardized variance components were $a^2 = 0.50$, $c^2 = 0.25$, $e^2 = 0.25$ and simulations were based on 2,000 twin pairs with an equal proportion of MZ and DZ pairs. Data were then analyzed using the conventional and mixture distributions models as before.

Parameter estimates obtained from all simulations were further analyzed using the statistical package R (R Development Core Team, 2004).

Data Application

Variables and Zygosity Diagnosis

TEDS is a large scale longitudinal study on language and cognitive developments involving a representative sample of all twins born in England and Wales in 1994–1996 (e.g., Trouton et al. 2002). In the present study, eight variables related to language and cognitive developments of 7-year-old twins were available for analysis. These traits included scores on conceptual grouping, picture completion, similarities, vocabulary and test of word recognition (TOWRE). The composites of a number of the variables, i.e., language (a composite of similarities and vocabulary), non-verbal IQ (a composite of scores for conceptual grouping and picture completion) and general cognitive ability (*g*), which is the composite of language and non-verbal IQ, were also analyzed. The cognitive abilities were measured on each child individually and separately using a telephone interview (Petrill et al. 2002). The complete description and definitions of the IQ variables were presented previously (e.g., Harlaar et al. 2005; Kovas et al. 2005; Price et al. 2004; Spinath et al. 2004). In addition to these variables, height and weight were included in the analyses and these variables were supplied by a parent/guardian of the twins, usually the mother. Parental ratings were used to ascertain the zygosity of SS twin pairs (Kovas et al. 2005). This method has an error rate less than 5%, as validated by

Table 1 Scenarios of simulated variance component proportions

Variance components	I	II	III	IV	V	VI	VII	VIII	IX
a^2	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
c^2	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05
e^2	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05

DNA typing using a multiplexed set of highly polymorphic markers (Harlaar et al. 2005; Kovas et al. 2005; Price et al. 2000).

Samples and Exclusions

Individuals and their co-twin were excluded from the analysis if: (i) their data base entry had missing identifiers (for sex and zygosity), (ii) there were specific medical and genetic conditions recorded (as described by Kovas et al. 2005), (iii) they were of non-white ethnicity, (iv) English is not the language at home (v) either twin had an extreme phenotype (more than three standard deviations from the mean for any variable) and (vi) they were of the opposite sex. The reason for excluding the opposite sex twins from the analyses was to avoid possible (large) biases due to sex-limitation effects. If opposite sex pairs were included in the analysis then the parameter estimates for the DZ SS twins will be centered on the DZ OS twins because the OS intraclass correlation is estimated with more precision than the variance components from the mixture distribution (See Appendix B in Benyamin et al. 2005 for more explanation).

The final data set comprised 3,582 SS twins, 1,904 MZ and 1,678 DZ pairs. The proportion of MZ pairs among all twin pairs in the selected dataset is larger than that of the unselected dataset (0.372 ± 0.007 compared to 0.313 ± 0.009). This difference could be due to a larger participation rate of MZ twins in the cognitive study. However, the proportion of MZ pairs among all pairs in the TEDS twin data is not significantly different from the whole twin population born in England and Wales between 1994 and 1996 (Imaizumi 2003) (0.351 ± 0.005 (TEDS) vs. 0.343 ± 0.003 (population)).

Analysis

Descriptive statistics of the standardized residuals (after a general linear model correction for sex and age effects on all observations) of the IQ variables, height and weight were obtained using SPSS 12.0.2 for Windows (SPSS Inc., 1989–2003). The standardized residuals were then split into MZ and DZ groups. Pearson correlations for MZ and DZ boys and girls were computed after adjustment for age effects. To test for normality, a Kolmogorov–Smirnov normality test (implemented in SPSS) was performed for all phenotypes, after adjustment for sex and age effects. All phenotypes were analyzed using the known zygosity and mixture distribution models. In the mixture distribution model, the observed proportion of MZ among SS twins (pMZ = 0.53) was used to weight the analyses. For all analyses, sex and age were fitted as fixed effects.

Results

Simulation Study

Mixture Distribution Model Under Normality

For normally distributed twin data, heritability (a^2) estimates from the mixture distribution model were compared with that from the known zygosity model. Figure 1 shows the relationship between the estimates from the two models, for the range of heritabilities of 0.1–0.9 (other parameters as in Table 1), for samples of 2,000 twin pairs. For all sets of parameters, the mean estimate of the heritability was very similar for both models, i.e., there was no evidence of a bias in the estimate of heritability, unless the heritability was small ($a^2 < 0.4$). The results showed that the higher the heritability simulated, the more similar the estimates between the two models. For heritabilities ≤ 0.4, although the mean estimates between the two models were similar, the standard deviation of estimates from the mixture distribution was about three times that of the known zygosity model. A similar pattern was also observed for the standardized common environmental variance (c^2), i.e., the larger the heritability simulated, the more similar the c^2 estimates between the two models (Fig. 2).

When the estimate of the heritability is unbiased, a useful criterion for precision of estimation of the mixture approach is the standard deviation of the difference in the estimate of the heritability between the two models. Figure 3 shows this standard deviation for a range of sample sizes from 500 to 5,000 pairs, for the range of population parameters as given in Table 1. As expected, the larger the sample size, the smaller the standard deviations of the difference between the two estimates. For a sample size as large as 5,000 pairs, the maximum standard deviation of the difference was 0.15 (for a low heritability) and for a large heritability little information is lost by not knowing zygosity. However, when the sample size is quite small (e.g., 500 pairs), the standard deviations of the difference between the two estimates of heritability were quite large even for heritability as high as 0.6.

The effect on bias in the estimates of variance components when specifying a wrong pMZ in the mixture distribution model is presented in Fig. 4, when $a^2 = 0.50$, $c^2 = 0.25$ and $e^2 = 0.25$. The magnitude and direction of the effects on each parameter estimate

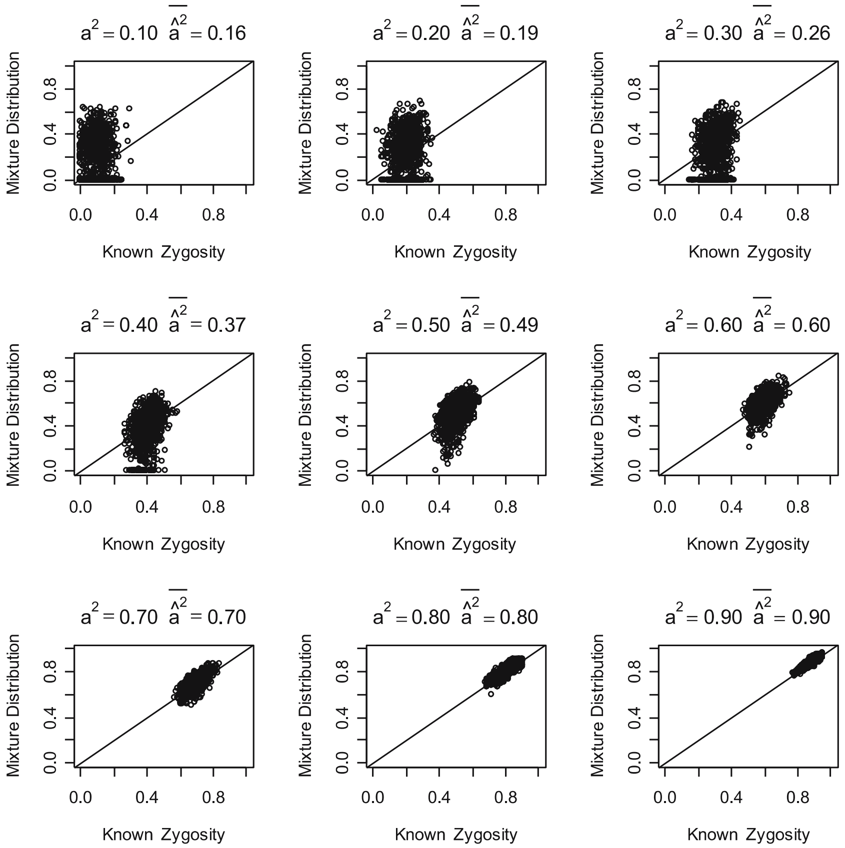

Fig. 1 Heritability (a^2) estimates from the mixture distribution and known zygosity models for different values of simulated heritability (Table 1) under an ACE model. The results were based on 2,000 twin pairs (1,000 MZs and 1,000 DZs) and 1,000 replicates. $\overline{\hat{a}^2}$ is the mean heritability estimate from the mixture distribution model

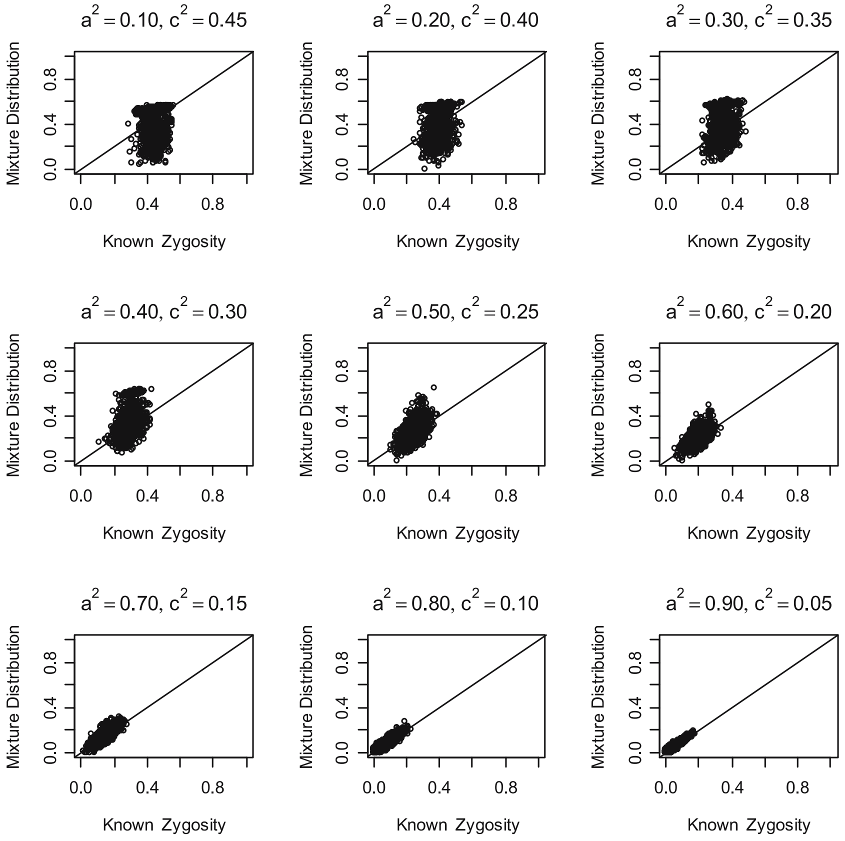

Fig. 2 Standardized common environmental variance (c^2) estimates from the mixture distribution and known zygosity models for different values of standardized variance components (Table 1) under an ACE model. The results were based on 2,000 twin pairs (1,000 MZs and 1,000 DZs) and 1,000 replicates

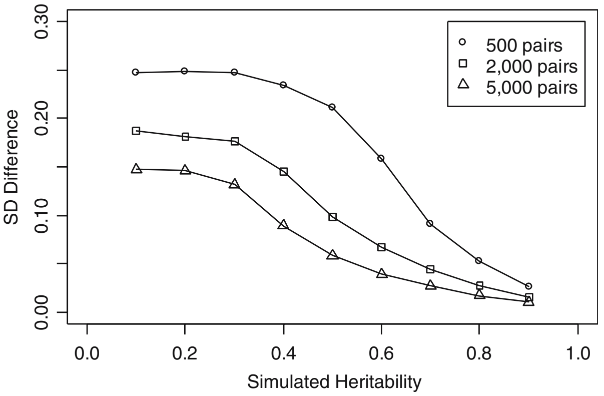

Fig. 3 Standard deviation of the difference in heritability estimates between the mixture distribution and known zygosity models for different sample sizes

were different. The effect on heritability and error variance estimates were small. For example, when the actual pMZ is 0.5 and we specify it as 0.6 or 0.4, the mean bias of heritability estimates from the mixture distribution was less than 2% . The effects were slightly larger on the common environmental variance estimates, especially when pMZ was underestimated.

Mixture Distribution Model When Twin Data is Kurtotic

For twin data with a kurtotic distribution of the phenotype, estimates from the known zygosity model were unbiased (results not shown). Figure 5 presents the mean difference of parameter estimates between the mixture distribution and known zygosity models for platykurtic (negative kurtosis) and leptokurtic (positive kurtosis) distributions. As indicated from Fig. 5, the mixture distribution model resulted in larger heritability and smaller c^2 and e^2 estimates on simulated data with a kurtotic distribution compared to the known zygosity model. However, for smaller kurtosis values ($-0.5 < k < 1$), the mean parameter estimates from the mixture distribution model did not differ substantially from that of the conventional analysis.

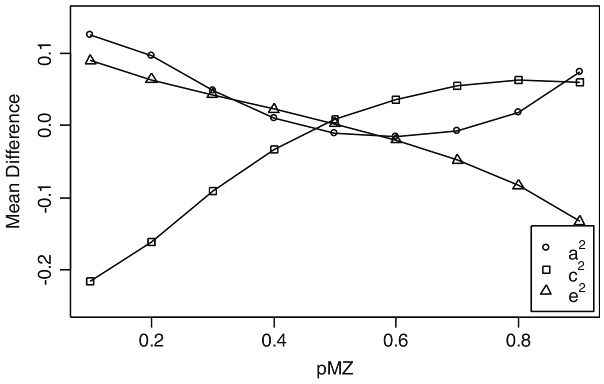

Fig. 4 Mean difference of standardized variance component estimates between the mixture distribution and known zygosity models when the incorrect mixture proportion was assumed (true proportion = 0.5). The simulated parameters are: a^2 = 0.50, c^2 = 0.25, e^2 = 0.25 and simulations are based on 2,000 twin pairs (1,000 MZs and 1,000 DZs) and 1,000 replicates

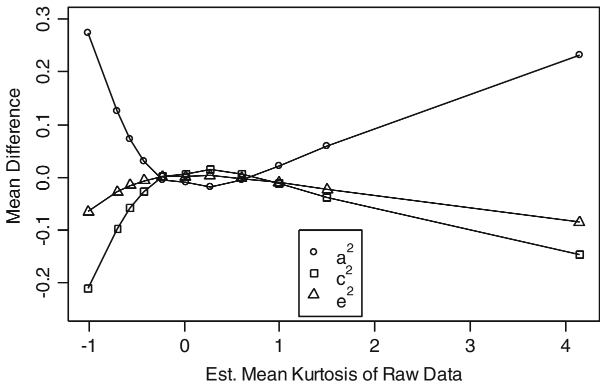

Fig. 5 Mean difference of standardized variance component estimates between the mixture distribution and known zygosity models for a given estimated mean kurtosis. The simulated parameters are: a^2 = 0.50, c^2 = 0.25, e^2 = 0.25 and simulations are based on 2,000 twin pairs (1,000 MZs and 1,000 DZs) and 1,000 replicates

Analyses of TEDS Data

Descriptive Statistics and Phenotypic Distribution

Descriptive statistics of the data after exclusions are presented in Table 2. Between 2,279 and 2,545 pairs for which both twins had a phenotype on any variable from a total of 3,582 pairs were available for analysis. The main reason for a considerable missing data is that not all twins were tested/interviewed at age 7. The age of the twins when the parents' booklet was returned, which was used as a covariate in the genetic analysis, had a mean and SD of 7.05 and 0.25, respectively. Although the distribution of the phenotypes appeared normal, the Kolmogorov–Smirnov normality test showed that the trait distribution was significantly different from normality for all traits (except language (MZ, DZ) and *g* (DZ)). The skewness and kurtosis values ranged from −0.38 to 0.48 and −0.75 to 1.12, respectively. Note that, with these kurtosis values, the simulations showed that the bias in the estimate of heritability of the mixture distribution model was less than 0.1 for a heritability of 0.5.

Twin Correlations

Twin correlations for all phenotypes, after adjustment for age, are presented in Table 3. The MZ and DZ twin

Table 2 Descriptive statistics of standardized residuals of TEDS variables after adjustment for sex and age effects

Variables	Zygosity	Sample size[a]	Mean (SD)	Skewness (SE)	Kurtosis (SE)	Skewness of pair difference (SE)	Kurtosis of pair difference (SE)
Age (years)	MZ	1,823	7.05 (0.25)				
	DZ	1,580	7.06 (0.25)				
Weight (kg)	MZ	1360	−0.04 (0.99)	0.45 (0.05)	0.26 (0.09)	0.21(0.07)	3.45 (0.13)
	DZ	1,185	0.05 (1.01)	0.48 (0.05)	0.35 (0.10)	0.03 (0.07)	0.74 (0.14)
Height (cm)	MZ	1,337	−0.02 (0.96)	0.00 (0.05)	0.59 (0.09)	−0.04 (0.07)	6.79 (0.13)
	DZ	1,162	0.03 (1.04)	−0.24 (0.05)	1.12 (0.10)	0.17 (0.07)	1.54 (0.14)
TOWRE	MZ	1,255	−0.02 (1.02)	0.10 (0.05)	−0.75 (0.10)	0.23 (0.07)	1.20 (0.14)
	DZ	1,133	0.03 (0.97)	0.05 (0.05)	−0.65 (0.10)	0.01 (0.07)	0.26 (0.15)
Conceptual grouping	MZ	1,290	−0.01 (1.00)	−0.27 (0.05)	−0.67(0.10)	0.06 (0.07)	0.16 (0.14)
	DZ	1,155	0.01 (1.00)	−0.34 (0.05)	−0.61 (0.10)	0.02 (0.07)	0.20 (0.15)
Similarities	MZ	1,281	−0.03 (1.01)	0.02 (0.05)	0.19 (0.10)	−0.12 (0.07)	0.65 (0.14)
	DZ	1,145	0.04 (0.98)	0.01 (0.05)	0.37 (0.10)	−0.19 (0.07)	1.12 (0.15)
Vocabulary	MZ	1,284	−0.02 (1.00)	−0.01 (0.05)	−0.15 (0.10)	0.10 (0.07)	0.61 (0.14)
	DZ	1,149	0.02 (1.00)	0.02 (0.05)	−0.09 (0.10)	−0.11 (0.07)	0.47 (0.15)
Picture completion	MZ	1,291	−0.04 (0.98)	−0.29 (0.05)	−0.07 (0.10)	0.00 (0.07)	0.44 (0.14)
	DZ	1,153	0.05 (1.02)	−0.38 (0.05)	0.06 (0.10)	0.11 (0.07)	0.71 (0.15)
g	MZ	1,270	−0.04 (1.00)	−0.10 (0.05)	−0.19 (0.10)	−0.01(0.07)	0.30 (0.14)
	DZ	1,137	0.04 (1.00)	−0.13 (0.05)	−0.16 (0.10)	−0.04 (0.07)	0.70 (0.15)
Language	MZ	1,274	−0.03 (1.01)	0.01 (0.05)	−0.20 (0.10)	0.07(0.07)	0.59 (0.14)
	DZ	1,141	0.03 (0.99)	0.06 (0.05)	−0.09 (0.10)	−0.11 (0.07)	0.58 (0.15)
Non-verbal	MZ	1,128	−0.03 (0.99)	−0.20 (0.05)	−0.31 (0.10)	−0.01 (0.07)	−0.10 (0.14)
	DZ	1,151	0.04 (1.01)	−0.28 (0.05)	−0.24 (0.10)	0.06 (0.07)	0.33 (0.15)

Note: The Kolmogorov–Smirnov normality test showed that for most variables, both the trait distribution (except language (MZ, DZ) and *g* (DZ)) and the distribution of the pair difference (except language (MZ), and *g* (DZ), non-verbal (DZ), TOWRE (DZ)) showed significant deviation from normality

[a] Number of pairs for which both twins had a phenotype

correlations were similar across sexes. MZ correlations were consistently higher than DZ correlations. Results in Table 3 indicate strongly that genetic factors play an important role in explaining phenotypic variance in most of the traits.

Variance Component Estimation

Initially, the variance component estimation using the known zygosity model was performed with separate variance components for boys and girls. However, for most variables there was no significant difference between variance component estimates in boys and girls, except for weight, TOWRE, similarities and picture completion (results not shown). The pooled (boys and girls) estimates from the known zygosity and mixture distribution models are presented in Table 4.

Known zygosity With the exception of TOWRE score for which a large heritability was estimated (about 0.6), the heritability estimates of other IQ phenotypes were small to moderate, ranging from 0.16 to 0.33. Shared environmental variance accounted for 19–37% of the phenotypic variance of IQ variables. Thus, most of the phenotypic variation in IQ related variables (except for TOWRE) was specific to individuals.

Genetic factors constituted a large proportion of the phenotypic variation in weight and height. About 70% of the total phenotypic variance in weight and height were attributed to genetic factors. These findings are similar to previous studies on heritabilities of weight (reviewed by Pietilainen et al. 2002) and height (reviewed by Silventoinen 2003).

Table 3 Twin correlations and their standard errors after adjustment for age effects

Variables	MZ—Boy	MZ—Girl	DZ—Boy	DZ—Girl
Weight	0.84 (0.02)	0.85 (0.02)	0.47 (0.03)	0.52 (0.03)
Height	0.92 (0.02)	0.94 (0.02)	0.56 (0.03)	0.64 (0.03)
TOWRE	0.85 (0.02)	0.84 (0.02)	0.51 (0.03)	0.50 (0.04)
Conceptual grouping	0.38 (0.04)	0.32 (0.04)	0.24 (0.04)	0.29 (0.04)
Similarities	0.51 (0.04)	0.43 (0.04)	0.37 (0.04)	0.35 (0.04)
Vocabulary	0.63 (0.03)	0.57 (0.04)	0.49 (0.03)	0.47 (0.04)
Picture Completion	0.47 (0.04)	0.48 (0.04)	0.39 (0.04)	0.40 (0.04)
g	0.68 (0.03)	0.61 (0.03)	0.49 (0.03)	0.48 (0.04)
Language	0.67 (0.03)	0.61 (0.03)	0.50 (0.03)	0.49 (0.04)
Non-verbal	0.45 (0.04)	0.45 (0.04)	0.39 (0.04)	0.38 (0.04)

Table 4 Standardized variance component estimates from the known zygosity and mixture distribution models (boys and girls are pooled)

Variables	Models	a^2 (95%CI)	c^2 (95%CI)	e^2 (95%CI)
Weight	Known	0.71 (0.63–0.79)	0.14 (0.06–0.22)	0.15 (0.14–0.17)
	Mixture	0.96 (0.93–0.97)	0.00 (0.00–0.03)	0.04 (0.03–0.05)
Height	Known	0.69 (0.63–0.76)	0.24 (0.17–0.31)	0.07 (0.06–0.07)
	Mixture	0.82 (0.75–0.91)	0.14 (0.05–0.22)	0.04 (0.03–0.05)
TOWRE	Known	0.63 (0.55–0.71)	0.21 (0.12–0.29)	0.16 (0.15–0.18)
	Mixture	0.88 (0.77–0.91)	0.01 (0.00–0.11)	0.11 (0.09–0.13)
Conceptual grouping	Known	0.16 (0.02–0.30)	0.19 (0.07 – 0.30)	0.65 (0.61–0.70)
	Mixture	0.41 (0.13–0.46)	0.00 (0.00–0.21)	0.59 (0.54–0.66)
Similarities	Known	0.19 (0.06–0.31)	0.27 (0.18–0.39)	0.54 (0.50–0.58)
	Mixture	0.59 (0.50–0.63)	0.00 (0.00–0.06)	0.41 (0.37–0.46)
Vocabulary	Known	0.26 (0.16–0.36)	0.35 (0.25–0.43)	0.40 (0.37–0.43)
	Mixture	0.72 (0.55–0.75)	0.00 (0.00–0.14)	0.28 (0.25–0.33)
Picture completion	Known	0.20 (0.08–0.32)	0.29 (0.19–0.39)	0.51 (0.47–0.56)
	Mixture	0.61 (0.54–0.65)	0.00 (0.00–0.04)	0.39 (0.35–0.44)
g	Known	0.33 (0.23–0.43)	0.32 (0.23–0.41)	0.35 (0.32–0.38)
	Mixture	0.74 (0.56–0.78)	0.01 (0.00–0.16)	0.26 (0.22–0.30)
Language	Known	0.26 (0.16–0.36)	0.37 (0.28–0.46)	0.37 (0.34–0.40)
	Mixture	0.71 (0.53–0.78)	0.04 (0.00–0.18)	0.26 (0.22–0.30)
Non-verbal	Known	0.16 (0.04–0.29)	0.30 (0.20–0.40)	0.54 (0.50–0.58)
	Mixture	0.55 (0.27–0.60)	0.00 (0.00–0.22)	0.45 (0.40–0.52)

Note: Known and mixture are the known zygosity and mixture distribution models, respectively.

Mixture distribution The heritability estimates from the mixture distribution were consistently larger than those from the conventional model for all variables (Table 4). This observation was unexpected. The confidence intervals of the estimates of heritability from the two models did not overlap for most traits. In addition, with the exception of height, the estimate of common environmental variance was zero (or close to zero) for all variables. For IQ phenotypes (except TOWRE), the mean difference of heritability estimates from the mixture distribution compared to the known zygosity model was 0.40. The difference in the average estimate of common environment variance was 0.29. However, the sum of the proportion of variance due to additive genetic and common environmental effects (giving the repeatability, the proportion of phenotypic variance of single measurements due to the effects of genetic and permanent environmental factors (Falconer and Mackay 1996)) was similar between the two models (Fig. 6).

Discussion

Neale (2003) has shown that a mixture distribution model can be used to analyze twin data when zygosity information is incomplete or unavailable with little bias. His simulations were from (bivariate) normal distributions, an assumption of the mixture distribution model. In addition, Neale (2003) only simulated a single set of standardized variance components (i.e., $a^2 = 0.6$, $c^2 = 0.2$, $e^2 = 0.2$) for twins without zygosity information. Our study further explored the properties of this model for a wider range of parameters and different scenarios. It includes assessing the mixture distribution model when the data is not normally distributed by simulating platykurtic and leptokurtic distributions. Different (incorrect) proportions of MZ among twins were also used to weight the analysis in the mixture distribution for a given true proportion, to assess the bias introduced by misspecification of this parameter. The simulation results suggested that, if the normality assumption was satisfied and the sample size was large, then the variance component estimates from the mixture distribution are unbiased and accurate for analysing twin data where zygosity information is unavailable. However, if the heritability is small ($a^2 < 0.4$), then the estimates are imprecise.

If the distribution of the phenotypes is kurtotic then the mixture distribution produced biased estimates. However, this bias was small for kurtosis values in the range of −0.5 and 1. Specifying a wrong mixture proportion in the analysis had small impact, in terms of bias, on the estimates of variance components, unless the difference between the true and estimated proportion was very large (e.g., 0.5 and 0.2). For a population survey, the estimated proportion of MZ among SS twins can usually be estimated accurately. For the TEDS data, the estimated proportion of MZ twins among SS twins using Weinberg's differential rule (Weinberg 1902) was very similar to the observed proportion, 0.523 and 0.519, respectively.

Post-hoc Analyses

The analyses of IQ phenotypes, height and weight from the TEDS data showed that the estimate of heritabil-

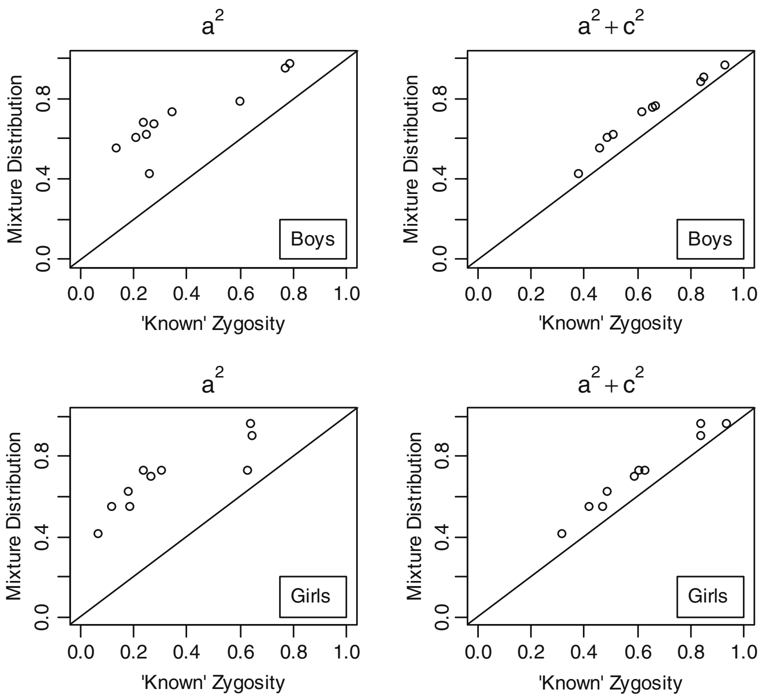

Fig. 6 Heritability (a^2) and repeatability (defined as the sum of a^2+c^2) estimates of nine IQ phenotypes, height and weight of TEDS data from the known zygosity and mixture distribution models for boys and girls. Sex and age were fitted as covariates

ities from the mixture distribution were consistently larger than those from the conventional model. These results are inconsistent with those from all of our simulations and demand an explanation. For the observed kurtosis values of the TEDS variables (in the range of −0.75–1.12, see Table 2), the observed differences of variance component estimates between the two models were considerably larger than those from simulations. Thus, the observed differences could not be attributed to the kurtosis of the trait distributions. However, a further detailed dissection of the phenotypic distributions has shown that the distributions of *pair difference* were all leptokurtic (except non-verbal ability scores of MZ twins (Table 2), even for traits where the univariate (single twin) distribution was close to normality. This finding was unexpected and implies a violation of the usual assumption of bivariate normality of twins' phenotypes (e.g., Huggins et al. 1998; Neale 2003; Rijsdjik and Sham 2002).

To verify that kurtosis of pair difference was the cause of the observed discrepancy, twin data that mimic the average parameter estimates of IQ phenotypes (except TOWRE) from the known zygosity model were simulated as an example (i.e., a^2 = 0.22, c^2 = 0.30, e^2 = 0.48 with a kurtosis of pair difference of 0.48). Phenotypes (y_1 and y_2) were simulated for a twin pair from a normal distribution, their difference ($D = y_1 - y_2$) was transformed (to D^*) using the previously described Cornish-Fisher transformation, and finally individual observations were backtransformed to $y_i^* = y_i \times D^*/D$. This transformation was made to keep the means and variances of the individual observations approximately the same whilst creating kurtosis of the pair difference. The results (Table 5) clearly showed that the variance component estimates of the simulated data from the mixture distribution model resembled those of the IQ phenotypes of the TEDS study: the average estimates for the simulated data were a^2 = 0.58, c^2 = 0.01, e^2 = 0.41 (Table 5), whereas the estimates of the IQ phenotypes (except TOWRE) were a^2 = 0.62, c^2 = 0.01, e^2 = 0.38. The discrepancy on variance component estimates between the normal and transformed (kurtosis) data using the known zygosity model (Table 5) was a direct result of the transformation.

To assess further the effects of kurtosis of pair difference on the mixture distribution model, another simulation was carried out by simulating different values of kurtosis on pair difference. The results showed that kurtosis on pair difference had considerable effect on heritability estimations using the mixture distribution model (Fig. 7). It can be seen clearly from the figure that the mixture distribution produced biased estimates even for small kurtosis values of the pair difference ($-0.5 < k < 0.5$). For a leptokurtic distribution of pair difference, the mixture distribution overestimated the heritability compared to known zygosity models. Even for a small positive kurtosis value ($k < 0.5$), the overestimation was not trivial (i.e., about 40%). On the other hand, the mixture distribution underestimated heritability if the distribution of

Table 5 Mean estimates (SE) from 1,000 simulated twin data sets with parameters that mimic the average estimates of standardized variance components from the TEDS data (i.e., $a^2 = 0.22$, $c^2 = 0.30$, $e^2 = 0.48$)

Models	a^2 (SE)	c^2 (SE)	e^2 (SE)
Known (normal[a])	0.222 (0.002)	0.297 (0.002)	0.481 (0.001)
Mixture (normal[a])	0.211 (0.007)	0.306 (0.005)	0.483 (0.002)
Known (kurtosis[b])	0.257 (0.002)	0.236 (0.002)	0.507 (0.001)
Mixture (kurtosis[b])	0.581 (0.002)	0.005 (0.001)	0.413 (0.001)

Note: Known and mixture are the known zygosity and mixture distribution models, respectively

[a] Normally distributed twin data

[b] Transformed twin data with a kurtosis value of pair difference of 0.48

pair difference was platykurtic. The bias produced by this type distribution was even larger than the bias from a leptokurtic distribution. These results are consistent with the observation in Benyamin et al. (2005) that the information to separate the two mixtures from the mixture distribution model comes from the difference in the squared variance and kurtosis. Hence, if the pair difference within zygosity class is kurtotic, the mixture distribution will produce biased estimates because the model assumes that the only source of kurtosis is the mixture of two normal distributions. For the known zygosity model, the differences between the variance component estimates from the normal and transformed data are a direct result of the transformation of the data which changed the MZ and DZ correlations (results not shown), and merely show that the correlations depend on the scale of the observations.

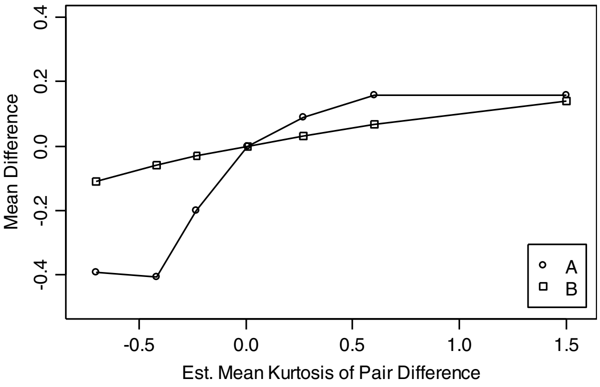

Fig. 7 Mean differences of heritability estimates between the mixture distribution and known zygosity models for a given estimated mean kurtosis of the pair difference. The simulated parameters are: $a^2 = 0.50$, $c^2 = 0.25$, $e^2 = 0.25$ and simulations are based on 2,000 twin pairs (1,000 MZs and 1,000 DZs) and 1,000 replicates. **(A)** is the difference between the mixture distribution and known zygosity models for transformed data with specific kurtosis; **(B)** is the difference of the known zygosity model between transformed data with specific kurtosis and normally distributed data

Are these results particular to the TEDS data? To explore this possibility, we analyzed a number of IQ variables from an independent smaller sample of 272 MZ and 191 SS DZ twins with known zygosity and an average age of 16 years-old from the ongoing Brisbane Memory, Attention, and Problem-Solving (MAPS) twin study (Luciano et al. 2003; Wright et al. 2001). Eight IQ measures, namely information, arithmetic, vocabulary, verbal IQ, spatial, object assembly, performance IQ and full scale IQ assessed with the Multidimensional Aptitude Battery II (MAB-II) (Jackson, 1998) were analyzed using the known zygosity and mixture distribution model. Individuals with more than three standard deviations from the mean were excluded from the analysis, and sex and age were fitted as fixed effects. As with the TEDS data, an ACE model was fitted. The kurtosis of the pair difference of variables from the MAPS study were not different from zero for most traits but the SE was relatively large, ranging from 0.29 to 0.35. Heritability estimates ranged from 0.39 to 0.68 for the known zygosity analysis, and 0.01 to 0.85 for the mixture distribution model analysis. For six out of the eight traits, the estimate of the heritability from the mixture distribution model was larger than the estimate from the known zygosity model, consistent with an observed leptokurtic distribution of the pair difference, averaged over MZ and DZ pairs. For the other two traits, the lower estimate of the heritability from the mixture distribution model was consistent with the observed platykurtic distribution of the pair difference. For these traits, the average kurtosis of the pair difference from MZ and DZ pairs was −0.36 and −0.09, respectively. Although both the estimates of the heritability and their standard errors are larger in the MAPS study, making exact comparisons difficult, the results are qualitatively similar to those from the TEDS study, in that the difference in parameter estimates between the two models are consistent with the observed kurtosis of the pair difference.

What could be the cause of the observed kurtosis on the pair difference, and what are the consequences for twin studies in general? Kurtosis on the pair difference when there is no kurtosis in the population could be due to a ‘known’ zygosity group itself being a mixture with respect to within-family variances. This could be the case for example if MZ are ‘contaminated’ with DZ pairs, and vice versa. This is not likely to be an explanation for the data analyzed, because the zygosity protocol is well-established and ambiguities about zygosity were resolved by DNA typing. For $a^2 = 0.4$, $c^2 = 0.2$, $e^2 = 0.4$ and a bivariate normal distribution within zygosity group, a 5% error rate would create a kurtosis value of the pair difference of 0.034 and 0.016 within assigned MZ and DZ groups, respectively (Benyamin et al. 2005). These predicted values are below what was observed from the TEDS data (Table 2). Heterogeneity of within-family variance could be due to many factors, including heterogeneity of environmental variance (both MZ and DZ) and heterogeneity of within-family genetic variance (DZ). One speculative biological cause of heterogeneity of variance for MZ pairs is that such pairs vary in the amount of genome-wide methylation or placentation effects that are shared.

The way in which data are collected or scored can also cause the observed kurtosis. For example, sum scores collected from questionnaires may not be multivariate normally distributed. For the TEDS data, the pair difference was extremely kurtotic for height and weight, traits that were reported by parents, and a histogram of the pair difference showed a huge peak at zero, both for MZ and DZ (results not shown). This suggests that the parents may report the average of their twins' height and weight correctly but not their difference. If this reporting bias is stronger in MZ than in DZ then parameter estimates will also be biased using the standard model with known zygosity. Although this may be an explanation for the height and weight data from the TEDS study, it is unlikely to be an explanation for the IQ phenotypes, which were measured on each child individually and separately using a telephone interview and material sent by post, presumably independently of parental input.

Although twin researchers may check normality assumptions of the data before embarking on a maximum likelihood analysis that assumes normality, it is unusual to check for the assumption of bivariate normality in zygosity groups. Our results suggest that while the known zygosity model is quite robust to the violation of bivariate normality assumption, a re-examination of bivariate normality for existing data may be prudent. For unknown zygosity data, consistency of the estimates of variance components with those from the known zygosity pairs should be checked. Finally, we suggest that the possibility of extensive heterogeneity of within-family variance needs further attention.

Acknowledgments We thank Allan McRae and Ian M. White for advice on the simulation studies; Bill Hill, Nick Martin and Sarah Medland for discussions and useful suggestions; Robert Plomin for kindly providing the TEDS data and helpful comments; Nicole Harlaar for providing information regarding the TEDS data; Michelle Luciano for helping to extract the MAPS data and helpful feedback. Beben Benyamin was funded by a School of Biological Sciences Studentship of the University of Edinburgh, and by an Overseas Research Students Award. Ian Deary is the recipient of a Royal Society-Wolfson Research Merit Award. The MAPS study has been funded by ARC grants A79600334, A79906588, A79801419, DP0212016 and DP0343921. We also thank the editor and three anonymous reviewers for helpful comments.

Appendix 1: The Expected Kurtosis Value of Transformed Data for Given *c*

Let $x \sim N(0,1)$. The expected kurtosis value of the transformed variate y for given c is derived from the moments of y,

$$
\begin{aligned}
y &= x + \frac{c}{24}\left(x^3 - 3x\right)\\
y^2 &= \left(\frac{8-c}{8}\right)^2 x^2 + \left(\frac{8c-c^2}{96}\right)x^4 + \frac{c^2}{576}x^6\\
y^4 &= \left(\frac{8-c}{8}\right)^4 x^4 + 2\left(\frac{8-c}{8}\right)^2\left(\frac{8c-c^2}{96}\right)x^6\\
&\quad + \left[2\left(\frac{8-c}{8}\right)^2\left(\frac{c^2}{576}\right) + \left(\frac{8c-c^2}{96}\right)^2\right]x^8\\
&\quad + 2\left(\frac{8c-c^2}{96}\right)\left(\frac{c^2}{576}\right)x^{10} + \left(\frac{c^2}{576}\right)^2 x^{12}
\end{aligned}
$$

Following Kendall and Stuart (1947), the expected value of 2*r*th moment of the normal distribution is:

$$
\begin{aligned}
&E\left(x^{2r}\right) = \frac{(2r)!}{2^r r!}\sigma^{2r}, \quad E\left(x^{2r+1}\right) = 0, r \geq 1;\\
&E\left(x^2\right) = \sigma^2; \quad E\left(x^4\right) = 3\left(\sigma^2\right)^2,\\
&E\left(x^6\right) = 15\left(\sigma^2\right)^3, \quad E\left(x^8\right) = 105\left(\sigma^2\right)^4;\\
&E\left(x^{10}\right) = 945\left(\sigma^2\right)^5; \quad E\left(x^{12}\right) = 10{,}395\left(\sigma^2\right)^6
\end{aligned}
$$

The expected value of y, y^2 and y^4 are therefore:

$$
\begin{aligned}
E(y) &= 0\\
E\left(y^2\right) &= \frac{\sigma^2}{576}\left[9(8-c)^2 + 18c(8-c)\sigma^2 + 15c^2\sigma^2\right]\\
E\left(y^4\right) &= 3\left(\frac{8-c}{8}\right)^4\left(\sigma^2\right)^2 + 30\left(\frac{8-c}{8}\right)^2\left(\frac{8c-c^2}{96}\right)\left(\sigma^2\right)^3\\
&\quad + \left[210\left(\frac{8-c}{8}\right)^2\left(\frac{c^2}{576}\right) + 105\left(\frac{8c-c^2}{96}\right)^2\right]\left(\sigma^2\right)^4\\
&\quad + 1890\left(\frac{8c-c^2}{96}\right)\left(\frac{c^2}{576}\right)\left(\sigma^2\right)^5\\
&\quad + 10{,}395\left(\frac{c^2}{576}\right)^2\left(\sigma^2\right)^6
\end{aligned}
$$

Then, the expected kurtosis of y is:

$$k(y) = \frac{E(y - E(y))^4}{\left(E(y^2)\right)^2} - 3$$

$$\begin{aligned} k(y) = {} & \left\{ 3\left(\frac{8-c}{8}\right)^4 \left(\sigma^2\right)^2 + 30\left(\frac{8-c}{8}\right)^2 \left(\frac{8c-c^2}{96}\right)\left(\sigma^2\right)^3 \right. \\ & + \left[210\left(\frac{8-c}{8}\right)^2 \left(\frac{c^2}{576}\right) + 105\left(\frac{8c-c^2}{96}\right)^2 \right]\left(\sigma^2\right)^4 \\ & + 1890\left(\frac{8c-c^2}{96}\right)\left(\frac{c^2}{576}\right)\left(\sigma^2\right)^5 \\ & \left. + 10,395\left(\frac{c^2}{576}\right)^2 \left(\sigma^2\right)^6 \right\} \Big/ \\ & \left\{ \frac{\sigma^2}{576}\left[9(8-c)^2 + 18c(8-c)\sigma^2 + 15c^2\sigma^2 \right] \right\}^2 - 3 \end{aligned}$$

For a small value of c, $k(y) \approx c$.

References

Benyamin B, Wilson V, Whalley LJ, Visscher PM, Deary IJ (2005) Large, consistent estimates of the heritability of the cognitive ability in two entire populations of 11-year-old twins from Scottish Mental Surveys of 1932 and 1947. Behav Genet 35:525–534

Boomsma D, Busjahn A, Peltonen L (2002) Classical twin studies and beyond. Nat Rev Genet 3:872–882

Cornish EA, Fisher RA (1937) Moments and cumulants in the specification of distributions. Revue de L'Institut 5:307–320

Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC (2004) The impact of childhood intelligence on later life: following up the Scottish Mental Surveys of 1932 and 1947. J Pers Soc Psychol 86:130–147

Evans DM, Gillespie NA, Martin NG (2002) Biometrical genetics. Biol Psychol 61:33–51

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman Group Ltd, Essex CM20 2JE, England

Forget-Dubois N, Perusse D, Turecki G, Girard A, Billette JM, Rouleau G, Boivin M, Malo J, Tremblay RE (2003) Diagnosing zygosity in infant twins: physical similarity, genotyping, and chorionicity. Twin Res 6:479–485

Harlaar N, Spinath FM, Dale PS, Plomin R (2005) Genetic influences on early word recognition abilities and disabilities: a study of 7-year-old twins. J Child Psychol Psyc 46:373–384

Heath AC, Nyholt DR, Neuman R, Madden AF, Bucholz KK, Todd RD, Nelson EC, Montgomery GW, Martin NG (2003) Zygosity diagnosis in the absence of genotypic data: an approach using latent class analysis. Twin Res 6:22–26

Huggins RM, Loesch DZ, Hoang NH (1998) A comparison of methods of fitting models to twin data. Austral NZ J Statist 40:129–140

Imaizumi Y (2003) A comparative study of zygotic twinning and triplet rate in eight countries, 1972–1999. J Biosoc Sci 35:287–302

Jackson D (1998) Multidimensional aptitude battery II—manual. Sigma Assessment Systems, Port Huron, MI

Kendall MG, Stuart A (1947) The advanced theory of statistics, Vol 1, 3rd edn, Charles Griffin & Company Limited, London

Kovas Y, Harlaar N, Petrill SA, Plomin R (2005) 'Generalist genes' and mathematics in 7-year-old twins. Intelligence 33:473–489

Luciano M, Wright MJ, Geffen GM, Geffen LB, Smith GA, Evans DM, Martin NG (2003) A genetic two-factor model of the covariation among a subset of multidimensional aptitude battery and Wechsler adult intelligence scale—revised subtests. Intelligence 31:589–605

Neale MC, Boker SM, Xie G, Maes HH (2002) Mx: statistical modelling, 6th edn, Department of Psychiatry, Richmond, VA 23298

Neale MC (2003) A finite mixture distribution model for data collected from twins. Twin Res 6:235–239

Peeters H, Van Gestel S, Vlietinck R, Derom C, Derom R (1998) Validation of a telephone zygosity questionnaire in twins of known zygosity. Behav Genet 28:159–163

Petrill SA, Rempell J, Oliver B, Plomin B (2002) Testing cognitive abilities by telephone in a sample of 6- to 8-year-olds. Intelligence 30:353–360

Pietilainen KH, Kaprio J, Rasanen M, Rissanen A, Rose RJ (2002) Genetic and environmental influences on the tracking of body size from birth to early adulthood. Obes Res 10:875–884

Price TS, Freeman B, Craig I, Petrill SA, Ebersole L, Plomin R (2000) Infant zygosity can be assigned by parental report questionnaire data. Twin Res 3:129–133

Price TS, Dale PS, Plomin R (2004) A longitudinal genetic analysis of low verbal and nonverbal cognitive abilities in early childhood. Twin Res 7:139–148

R Development Core Team (2004) R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org)

Rijsdijk F, Sham PC (2002) Analytic approaches to twin data using structural equation models. Brief Bioinform 3:119–133

Scarr-Salapatek S (1971) Race, social class, and IQ. Science 174:1285–1295

Scottish Council for Research in Education (1949) The trend of Scottish Intelligence. University of London Press, London

Silventoinen K (2003) Determinants of variation in adult body height. J Biosoc Sci 35:263–285

Spinath FM, Price TS, Dale PS, Plomin R (2004) The genetic and environmental origins of language disability and ability. Child Dev 75:445–454

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

Weinberg W (1902) Beiträge zur physiologie und pathologie der meshrlingsgeburten beim menschen. Archiv Gesamte Physiol Menschen Tiere 88:346–430

Wright M, De Geus E, Ando J, Luciano M, Posthuma D, Ono Y, Hansell N, Van Baal C, Hiraishi K, Hasegawa T, Smith G, Geffen G, Geffen L, Kanba S, Miyake A, Martin N, Boomsma D (2001) Genetics of cognition: outline of a collaborative twin study. Twin Res 4:48–56