

# Bias, precision and heritability of self-reported and clinically measured height in Australian twins

Stuart Macgregor · Belinda K. Cornes ·  
Nicholas G. Martin · Peter M. Visscher

Received: 20 June 2006 / Accepted: 1 August 2006 / Published online: 25 August 2006  
© Springer-Verlag 2006

**Abstract** Many studies of quantitative and disease traits in human genetics rely upon self-reported measures. Such measures are based on questionnaires or interviews and are often cheaper and more readily available than alternatives. However, the precision and potential bias cannot usually be assessed. Here we report a detailed quantitative genetic analysis of stature. We characterise the degree of measurement error by utilising a large sample of Australian twin pairs (857 MZ, 815 DZ) with both clinical and self-reported measures of height. Self-report height measurements are shown to be more variable than clinical measures. This has led to lowered estimates of heritability in many previous studies of stature. In our twin sample the heritability estimate for clinical height exceeded 90%. Repeated measures analysis shows that 2–3 times as many self-report measures are required to recover heritability estimates similar to those obtained from clinical measures. Bivariate genetic repeated measures analysis of self-report and clinical height measures showed an additive genetic correlation  $>0.98$ . We show that the accuracy of self-report height is upwardly biased in older individuals and in individuals of short stature. By comparing clinical and self-report measures we also showed that there was a genetic component to females systematically reporting their height incorrectly; this phenomenon appeared to not be present in males. The

results from the measurement error analysis were subsequently used to assess the effects of error on the power to detect linkage in a genome scan. Moderate reduction in error (through the use of accurate clinical or multiple self-report measures) increased the effective sample size by 22%; elimination of measurement error led to increases in effective sample size of 41%.

## Introduction

Quantitative genetic analysis partitions variation into genetic and environmental components. When the trait in question is measured poorly, the variance due to environment increases, leading to deflated estimates of the heritability. Studies based on self-reported trait measures are common despite the possibility of bias and increase in measurement error. When accurate (clinical) measurements are available, the effect of poor measurement can be calibrated by comparing self-report and clinical measurements. Here we address the issue of measurement error and bias by examining a large stature data set.

Stature has been long the subject of quantitative genetic analysis, beginning with Galton in the 19th century (Galton 1886). In recent years the role of short stature as a risk factor for cardiovascular disease has been hotly debated (Hebert et al. 1993; Kannam et al. 1994; Langenberg et al. 2005; Silventoinen et al. 2006). Height is well known to be highly heritable with increases in mean height in the second half of the 20th century attributable to improved environmental

---

S. Macgregor (✉) · B. K. Cornes · N. G. Martin ·  
P. M. Visscher  
Genetic Epidemiology, Queensland Institute of Medical  
Research, Herston Road, Brisbane 4029, Australia  
e-mail: stuart.macgregor@qimr.edu.au

conditions (Vogel and Motulsky 1997; Fredriks et al. 2000; Padez 2003). Despite increases in mean height throughout the 20th century, the mean difference between European countries has remained stable and trait variance and heritability estimates are broadly similar across countries (Silventoinen et al. 2003b). In modern western societies around 20% of variation in height is attributable to environmental factors (Silventoinen 2003a, b). Estimates of heritability vary according to a number of factors. Estimates derived from twin samples are generally higher than those obtained from general family samples (e.g. parent-offspring, extended families); reasons for this include perfect age matching in twins, greater similarity in environment and differences in the methodologies applied. Furthermore, estimates vary depending upon the degree of measurement error. Characterisation of the measurement error requires multiple measures and/or measurements from both the individual (self-report) and from an independent source (e.g. clinical measures).

Previous studies of human height have addressed bias when individuals report their own height (Himes and Faricy 2001; Rowland 1990). Rowland (1990) describes a bias in self-reported height after age 45 in a United States (US) sample (older individuals overestimate their height) whilst Himes and Faricy (2001) report higher levels of missing data in shorter US adolescents. We were particularly interested in the relative contribution of genes and environment to variation in self-report and clinical height. The study by Eaves et al. (1999) considered self-report height in a large twin family sample from Virginia in the US whilst Silventoinen et al. (2003b) considered (primarily) self-report height measures across eight countries. Height measures have been widely used in genome scans searching for quantitative trait loci underlying the variation in human height. Some genome scans have utilised self-report height measures (Perola et al. 2001) whilst others have had access to clinical measures (Mukhopadhyay et al. 2003).

In this study we characterise the degree of measurement error by utilising a large sample of twins with both clinical and self-reported measures of height. We present an initial analysis which considers mean changes in height, before addressing the sources of variability in the different height measures. A statistical model which partitions the variation into genetic and environment components for a bivariate repeated measures model is described.

## Methods

### Phenotypic data

The data used in this paper were derived from several studies of adult twins recruited from the Australian Twin Registry. Since the sample was from a population register there is no over-representation of individuals affected by a particular disease or phenotype. Twins were selected to ensure each had a least one measure for both clinical and self-reported height. Self-report measures were taken from questionnaires mailed to twins between 1979 and 2004 whilst clinical measures were taken as part of studies conducted between 1992 and 1996.

The zygosity of twin pairs was initially determined by twins' responses to standard items about physical similarity and the degree to which others confused them with one another. Many of the twins were subsequently included in studies in which they were typed for large numbers of genetic markers. As a result we expect misclassifications to be very rare in this sample (Cornes et al. 2005).

The data consisted of 1,673 twin pairs (618 MZ females, 239 MZ males, 338 DZ females, 143 DZ males, 334 DZ opposite sex). All individuals had data on at least one clinical and self-report height measure with the average number of measures being 1.2 (range 1–3, standard deviation 0.5) and 2.5 (range 1–6, standard deviation 0.9) for clinical and self-report measures, respectively. It was usual for twins to be measured on the same day. Approximately 10% of the individuals had just 1 self-report measure whilst ~84% of individuals had just 1 clinical measure. Since the self-report measures were collected over a 20 year period the age at which the self-report measures were taken varied; the mean age at first self-report measure was 32 (SD 12) and the mean age at last self-report measure was 44 (SD 13). The clinical height measures were collected over a shorter period; mean age was 44 (SD 12). Prior to analysis individuals were screened to remove outliers; outliers were defined as individuals with self-report mean height differing from clinical mean height by more than 12 cm.

The data were analysed in two stages. First, the data were examined using regressions to examine bias and using twin correlations to evaluate evidence for familiarity. Second, the data were analysed in a variance components framework; this allowed characterisation of the measurement error whilst taking into account fixed effects such as age.

## Exploratory analysis

### Regression analysis

The effect of age on height was considered by regressing height on age, with covariates for sex and year of collection (fitted as a factor for cohort). For regression analysis repeated measures were dealt with by averaging the available data points per individual. The difference between the mean self-report and mean clinical height was then computed. This difference was then regressed on age and on clinical height. Regressions were done using R (R Development Core Team 2004).

### Basic genetic analysis

An initial analysis to evaluate evidence for familiarity was conducted by computing twin pair correlations on the MZ and DZ groups. For this univariate analysis the height measures were based on the first available height after correction for fixed effects. Since we are interested in partitioning the variation in the traits we use a single measure; although the mean (across available measures on an individual) could have been used, it is more satisfactory to deal with the repeated measures using a repeated measures model (see below). Correction for covariates was performed by using the residuals from a linear model with terms for year of collection (cohort), age, age<sup>2</sup>, sex and an age by sex interaction.

### Variance components

#### Univariate genetic analysis

A more comprehensive univariate genetic analysis was conducted by estimating variance components for the additive genetic (A), common environmental (C), dominance genetic (D) and unique environmental (E) effects. Since all of these components cannot be jointly estimated from the available twin data we fitted ACE, ADE, AE, CE and E models.

The main interest was in modelling the variation in height around the mean level. Henceforth we model the mean level in the population by fitting fixed effects in the model for year of collection (i.e. a factor for cohort), age, age<sup>2</sup>, sex and an age by sex interaction. For brevity in the linear models below all fixed effects are denoted by  $\mu$ . The basic linear model for an AE model is

$$y_i = \mu + a_i + e_i$$

where  $\mu$  represents the fixed effects,  $y_i$  is measure on individual  $i$  ( $i = 1, \dots, n$ ),  $a_i$  is the additive genetic effect

and  $e_i$  the random environmental effect. The covariance structure in the AE case is

$$\Omega = A\sigma_a^2 + I_n\sigma_e^2 \quad (1)$$

where  $\sigma_a^2$  and  $\sigma_e^2$  denote the additive genetic and random environmental variances, respectively and  $A$  denotes a matrix which codes the additive genetic coefficients [=2 × coefficient of co-ancestry (Lynch and Walsh 1998)] between individuals. For example, the off-diagonal entries of  $A$  are 0.5 for DZ pairs, 1 for MZ pairs and 0 for unrelated individuals. Models for the other cases (CE, etc.) are similar. Estimation for this mixed model is performed by assuming normality of the height measure and utilising restricted maximum likelihood (REML). REML estimation was performed using ASReml (Gilmour et al. 2002). The necessary coding of the data for twin analysis in ASReml is described in the ASReml manual (Gilmour et al. 2002) and by Visscher et al. (2004).

### Repeated measures models

*Simple measurement error model.* Height is modelled as

$$y_{ij} = \mu + p_i + e_{ij} \quad (2)$$

where  $\mu$  represents the fixed effects,  $y_{ij}$  is measure  $j$  ( $j = 1, \dots, w$ ) on individual  $i$  ( $i = 1, \dots, n$ ),  $p_i$  the permanent effect (i.e. effects present for all height measures on an individual) and  $e_{ij}$  the random environmental effect (i.e. those specific to each measure on an individual). The terms  $p_i$  and  $e_{ij}$  are random effects. Measures are nested within individuals with  $N = nw$  (or with missing data,  $N = n\sum w_i$ ). The overall variance covariance matrix,  $\Omega$ , is hence

$$\Omega = I_n \otimes (1_w 1_w^T \sigma_p^2) + I\sigma_N^2 \quad (3)$$

where  $\sigma_p^2$  and  $\sigma_e^2$  denote the permanent and random environmental variances, respectively.  $I_n$  denotes the  $n$  by  $n$  identity matrix and  $1_w$  is a 1 by  $w$  vector of 1's.  $\otimes$  denotes the direct product of two matrices. Estimation was done by REML, using a linear mixed model formulation as implemented in ASReml. Using REML, instead of say, least squares, estimation ensures missing data are efficiently handled (Lynch and Walsh 1998).

The ratio of between individual (or permanent) variance ( $\sigma_p^2$ ) to the total variance ( $\sigma_p^2 + \sigma_e^2$ ) is termed the repeatability (Falconer and Mackay 1996). Since the repeatability cannot be smaller than the heritability,  $h^2$ , the repeatability is an upper bound for  $h^2$ .

Evaluation of this simple model for the self-report and clinical height measures yields an initial estimate of the relative sources of error due to measurement.

*Repeated measures bivariate genetic model.* The above repeated measures model considered each trait separately. To model the clinical and self-report heights jointly we fitted a bivariate model to the twin data. We also incorporate the A/C/D/E components to allow further partition of the variation. The main change compared with the univariate genetic model is a change in the covariance matrix to incorporate the covariance between traits. This model combines elements of the genetic model with the repeated measures model outlined above. The covariance matrices are changed to reflect the structure of the observations. The overall variance covariance matrix,  $\Omega$ , for an AE model is

$$\Omega = L_A \otimes (A \otimes (1_w 1_w^T)) + L_P \otimes (I_n \otimes (1_w 1_w^T)) + I\sigma_e^2 \quad (4)$$

where  $L_A$  and  $L_P$  denote  $2 \times 2$  matrices of variances and covariances for the two effects (additive genetic and permanent environmental, respectively) for the two traits. Since  $L_A$  and  $L_E$  each contain three distinct elements and there is a single unique environmental error variance, this AE bivariate repeated measures model has seven (co)variance parameters. Models for the other cases (CE, etc.) are similar. Estimation was done using REML.

#### Difference measures

In addition to the analysis of clinical and self-report height, analyses were performed for the difference between self-report height and clinical height and the absolute value of the clinical/self-report difference. To allow the use of all of the data the mean self-report and clinical heights were used. In the case of the absolute value, the variate was log transformed to remove the left skew in the absolute value measure. The data were analysed by computing twin pair correlations on the MZ and DZ groups.

## Results

### Exploratory analysis

#### *Regression analysis*

In males the means (SD) were 176.7 cm (6.5 cm) and 177.2 cm (6.4 cm) for clinical and self-report height,

respectively. In females the means (SD) were 162.1 cm (6.6 cm) and 162.5 cm (6.6 cm) for clinical and self-report height, respectively.

The regression of mean height on mean age (with the effects of sex and cohort removed as they were included as covariates in the model) revealed that whilst both clinical and self-reported height decreased significantly with age ( $P < 10^{-9}$  in both cases), the decline was more marked for clinical height. The decline in clinical height was 0.112 cm/year. This corresponds to a decline in height of 4.5 cm over a 40 year period. The decline in self-report height was 0.065 cm/year. This corresponds to a decline in height of 2.6 cm over a 40-year period. There was no evidence for an interaction between age and sex for either clinical or self-report height. Assuming that the clinical measures show no systematic bias, older individuals tend to (self) report inflated values for their true height; instead many individuals report the height they were in their youth. There was evidence ( $P < 10^{-4}$ ) for a small cohort effect in the self-report sample; the mean height increased 1.06 cm between the period 1979–1984 and the period 1985–1994. The clinical data was not collected over a wide enough range of ages (range was 1993–1996) to reliably detect a cohort effect.

The regression of difference (between self-report and clinical height) on age revealed a similar picture to the result from the height/age regression above. Figures 1 and 2 show difference plotted against age for males and females separately. The regression was significantly different in males and females ( $P < 10^{-4}$  for significance of interaction). The female regression line was

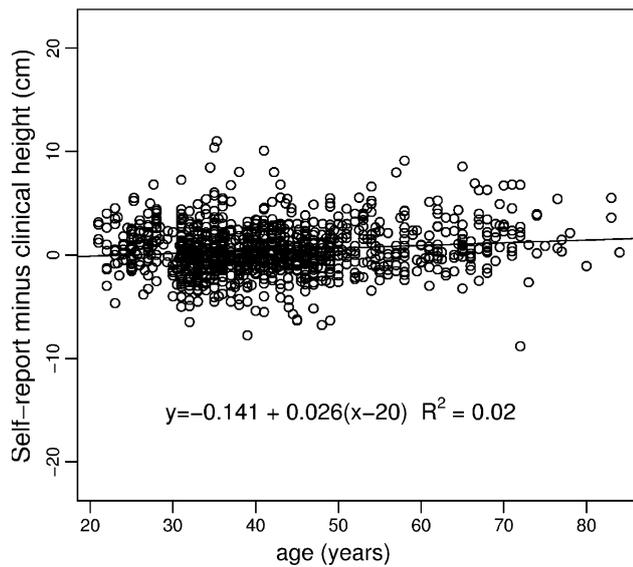
$$\text{difference} = -0.991 + 0.054(\text{age} - 20)$$

The male regression line was

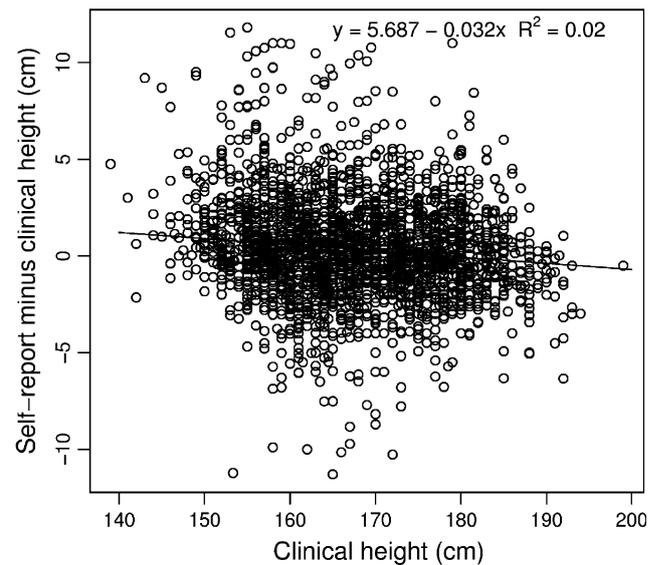
$$\text{difference} = -0.141 + 0.026(\text{age} - 20)$$

These results indicate that the clinical and self-report height measures have similar means early in life but the self-report height measures tend to overestimate “true” height (assuming clinical height is correct) later in life. For example, the average female self-reports her height to be similar to her clinical height at 38 years of age. By 68 years of age, the average female self-reports her height to be 1.7 cm greater than her clinical height.

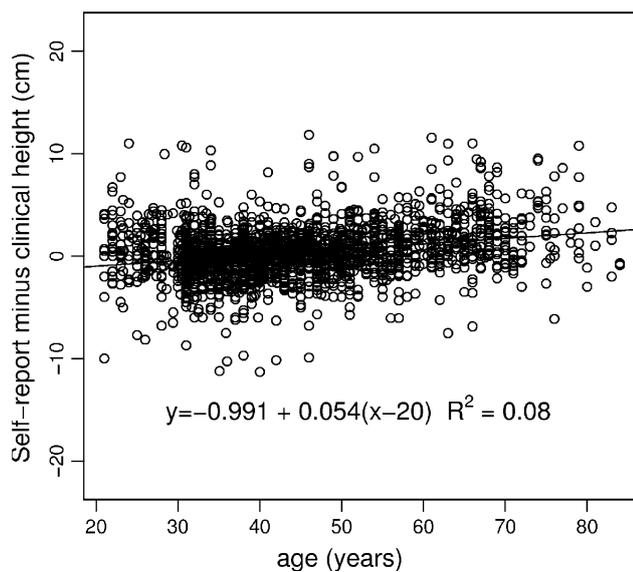
The effect of height upon the difference between self-report and clinical height was examined. Difference versus clinical height is shown in Fig. 3. The linear regression line has equation



**Fig. 1** Self-report minus clinical height versus age: males



**Fig. 3** Self-report minus clinical height versus clinical height



**Fig. 2** Self-report minus clinical height versus age: females

difference =  $5.687 + 0.032 \times \text{clinical height}$

The regression line drops below zero on the y axis at 178 cm. This indicates that, in general, smaller individuals tend to overestimate their height whilst taller individuals tend to underestimate their height. The degree of bias is relatively modest, with the bias difference between individuals of 150 cm and 180 cm being  $30 \times 0.032 = 1.0$  cm. The results with correction for age, sex and cohort were similar (data not shown).

### Basic genetic analysis

The twin correlations for age and sex corrected height (self-report, clinical) are given in Table 1. The clinical height measures show higher MZ correlations than the self-report measures, consistent with there being smaller measurement error in the clinical sample. In both cases the correlations in males and females are very similar and we hence analyse males and females together in the subsequent analysis.

### Variance components

#### Univariate genetic analysis

The results from the variance components analysis for the two traits are given in Table 2. Variance components, log-likelihoods and Akaike's Information Criteria (AIC) are given in each table; in each case the best model by AIC is given in bold. For both height measures the best fitting model is an AE model. The heritability given this model is higher for the clinical height measures than for self-report height. The values given in Table 2 are scaled to sum to 1 but the raw components for the AE model are  $\sigma_A^2 = 37.0$ ,  $\sigma_E^2 = 3.6$  and  $\sigma_A^2 = 36.1$ ,  $\sigma_E^2 = 5.6$ , for the clinical and self-report measures, respectively. Total variance in each case is 40.6 and 41.7.

*Simple measurement error model.* The repeated measure analysis revealed that the variation attributable to effects that remained the same across repeated height measures (i.e. the permanent effects) was similar in both clinical ( $\sigma_p^2 = 39.1$ ) and self-report measures

**Table 1** Height twin correlations

	Clinical		Self-report	
	N	Correlation (95% CI)	N	Correlation (95% CI)
Female MZ pairs	618	0.92 (0.90,0.93)	618	0.87 (0.85,0.88)
Male MZ pairs	239	0.92 (0.90,0.94)	239	0.87 (0.84,0.90)
Female DZ pairs	338	0.44 (0.35,0.52)	338	0.38 (0.29,0.47)
Male DZ pairs	143	0.39 (0.24,0.52)	143	0.38 (0.23,0.51)
Opposite sex DZ pairs	334	0.42 (0.33,0.51)	334	0.41 (0.32,0.59)

Correlations are based on the first available clinical and self-report height measure. Heights are corrected for year of collection (cohort), age, age<sup>2</sup>, sex and an age by sex interaction

CI confidence interval

**Table 2** Height variance component estimates

Model	$\sigma_A^2/\sigma_T^2$	$\sigma_C^2/\sigma_T^2$	$\sigma_D^2/\sigma_T^2$	$\sigma_E^2/\sigma_T^2$	ln L	Parameters	AIC
Clinical							
ACE	0.911	0		0.089	-7036.20	3	14078.40
ADE	0.851		0.060	0.089	-7036.06	3	14078.12
<b>AE</b>	<b>0.911</b>			<b>0.089</b>	<b>-7036.20</b>	<b>2</b>	<b>14076.40</b>
CE		0.690		0.310	-7360.47	2	14120.70
E				1	-7899.81	1	15801.62
Self-report							
ACE	0.866	0		0.134	-7263.01	3	14532.02
ADE	0.768		0.098	0.134	-7262.63	3	14531.26
<b>AE</b>	<b>0.866</b>			<b>0.134</b>	<b>-7263.01</b>	<b>2</b>	<b>14530.02</b>
CE		0.648		0.352	-7483.85	2	14971.70
E				1	-7937.37	1	15876.74

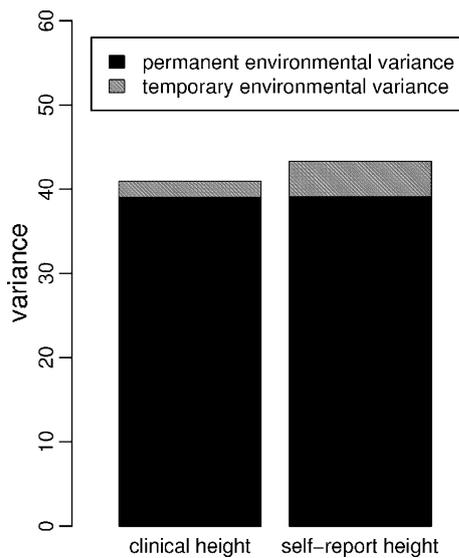
$\sigma_T^2 = \sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2$ . The best model by AIC is in bold

( $\sigma_p^2 = 39.2$ ). In contrast, the temporary environment variance (i.e. the measurement error variance) differed, with the clinical variance ( $\sigma_e^2 = 1.8$ ) substantially less than the self-report variance ( $\sigma_e^2 = 4.1$ ). Since the temporary environmental variance is  $4.1/1.8 = 2.3$  times larger for the self-report measure than the clinical measure, increasing the number of self-report measures (per individual) more than two-fold would be required to account for the increased variability in the self-report measures. That is, with 2–3 times as many self-report measures per individual, the heritability of the self-report measures should be similar to that of the clinical measures.

Total variance in the two cases was 40.9 and 43.3 (see Fig. 4 for summary). The total variance for the whole sample of self-report height ( $\sigma_t^2 = 43.3$ ) was slightly higher than the total variance of the first self-report height ( $\sigma_t^2 = 41.7$ ); there were 2.5 times as many data points in the whole sample analysis. The repeatabilities [i.e. ( $\sigma_p^2/(\sigma_p^2 + \sigma_e^2)$ )] for clinical and self-reported height are 0.96 and 0.90, respectively; these figures demonstrate that a substantial portion of the environmental variance in self-reported measures is attributable solely to measurement error.

*Repeated measures bivariate analysis.* First of all, a non-genetic repeated measures analysis was applied to the self-report and clinical height measures jointly [i.e. Eq. 4 without the  $\mathbf{L}_A \otimes (A \otimes (1_w 1_w^T))$  term]. The permanent environmental variances for clinical height and self-report height were 37.4 and 39.7, respectively; their covariance was 38.2. The residual error was 3.9. Total variance in the two cases was 41.3 and 43.6.

The bivariate repeated measures analysis was then extended to incorporate genetic effects. As expected given the univariate results, none of the alternative models (ADE, ACE, CE, E) gave a better fit to the data than the AE model shown here (data not shown). The additive genetic variance component estimates (i.e. the diagonal elements of the matrix  $\mathbf{L}_A$ ) for the AE model were 37.2 and 36.8 for the clinical and self-report height, respectively. The additive genetic covariance between the clinical and self-report measures was 36.1 (additive genetic correlation = 0.98). With the AE model the permanent environmental variances (i.e. the diagonal elements of the matrix  $\mathbf{L}_P$ ) were estimated to be 0.0 and 2.7 for clinical and self-report height, respectively. This analysis shows that the genetic variances of clinical and self-report height are



**Fig. 4** Variance components from basic repeated measures model

approximately equal. The increased total variance of the self-report measures is primarily attributable to increased measurement error.

#### Difference measures

Results for the difference measures are given in Table 3. The results in males/opposite sex pairs are difficult to interpret, particularly since the sample size in males is rather smaller than in females. In females, it is striking that the MZ correlation is significantly higher than the DZ correlation for the difference between the clinical and self-report measures. In contrast, when the absolute value of the difference is taken (log transformed to remove skewness) the female MZ correlation equals the DZ correlation. Computation of  $2 \times \text{MZ correlation} - 2 \times \text{DZ correlation}$  yields heritability estimates of 0.36 and 0.00 for the difference and absolute difference, respectively. The absolute value of the difference is a measure of the error that individuals

make in guessing their height whilst the raw difference score measures both the guessing error and a systematic bias. This systematic bias would occur when individuals consistently over (or under) estimate their height. The female only results indicate that whilst “guessing” error is predominantly environmental in origin, systematic error is likely to be influenced by genetic factors.

#### Discussion

We have shown here that height is extremely highly heritable, with these twin sample based estimates ( $h^2 \sim 85\%$  exceeding those generally reported from studies on general (non-twin) family data (Mukhopadhyay et al. 2003; Perola et al. 2001). The heritability of clinical height was shown to be higher than that of self-report height. This discrepancy was largely due to increased measurement error in the self-reported case.

Compared with the clinical measures, the self-reported height measures had a number of failings. Firstly, the clinical measures indicated that the average individual lost 4–5 cm in height over their adult life (possible reasons for this are discussed in Galloway et al. 1990). Although the self-reported measures showed a trend in this direction, the decline was somewhat smaller, with some older individuals appearing to simply (self) report the height they were when they were younger. The self-report measures also appeared to be biased by the height of the respondent. Small individuals tended to overestimate their height, whilst taller individuals were inclined to underestimate their height. Although age has a modest effect on height, for the variance components analysis conducted here we fit covariates for both age and other fixed effects such as cohort; this means that the effects of age and other covariates are removed before we partition the variance into the components of interest.

**Table 3** Twin correlations for differences

	Diff		log <sub>10</sub>  diff	
	N	Correlation (95% CI)	N	Correlation (95% CI)
Female MZ pairs	618	0.36 (0.29,0.43)	618	0.09 (0.02,0.17)
Male MZ pairs	239	0.19 (0.06,0.30)	239	0.03 (-0.10,0.16)
Female DZ pairs	338	0.18 (0.07,0.28)	338	0.10 (0.00,0.21)
Male DZ pairs	143	0.26 (0.10,0.40)	143	0.11 (-0.06,0.27)
Opposite sex DZ pairs	334	0.08 (-0.03,0.18)	334	0.05 (-0.06,0.15)

Correlations are based on the difference (diff) between mean clinical height and mean self-report. Heights are corrected for year of collection (cohort), age, age<sup>2</sup>, sex and an age by sex interaction

CI confidence interval

The results from the repeated measures analysis indicated that with 2–3 times as many self-report measures, the heritability of the self-report measures should be similar to that of the clinical measures. In the sample of self-report data here there were an average of 2.5 self-report measures per individual. We hence calculated the heritability of the mean self-report measure. The value obtained, 0.89, approached the value obtained from the clinical height measured on just one individual (0.91 using just the first available clinical measure, Table 2). It should be pointed out that the value of 2.5 for the average number of self-report measures slightly overestimates the information available because there was considerable variation in the number of measures available. In some cases there were up to six self-report measures per individual; repeated measurements beyond the first three would only marginally decrease the error due to measurement. With smaller variability in the number of measures, a larger estimate of heritability from the mean self-report height would have been obtained.

The two measures of height were used to analyse the discrepancy between self-report and clinical height measures. By assuming that the clinical measures were unbiased and subject to minimal error we were able to show that, in females at least, the difference between the measures and the absolute value of this difference had rather different compositions in terms of their components of variance. Whilst the females ability to guess their own height was largely determined by environmental factors, the differing MZ and DZ correlations suggest a genetic component to females systematically reporting their height incorrectly. This phenomenon was not seen for males although the male sample sizes were substantially smaller. This apparent genetic component could result if female MZ twins conferred when giving their self-reported height measures (or conferring more often than females DZ twins). Resolving this issue for certain would involve utilising an adoption based twin design or a sibling effects model (Eaves 1976; Carey 1986).

To allow ready comparison between non-nested models (e.g. the ACE and ADE models) we used Akaike's AIC. The AIC was designed for the (typical) case where the parameters were not on the edge of the parameter space under the null hypothesis. With testing variance components however, the null is usually that each parameter is zero and negative values are impossible. This means that likelihood ratio tests of the components usually take the form of mixtures (e.g.  $1/2(\chi_1^2):1/2(0)$  for the test of an ADE model versus an AE model, Self and Liang 1987). As a result, the usual AIC is conservative (models with more parameters are

over penalised). An alternative AIC based on say  $-2 \times \log \text{likelihood} + 1 \times \text{degrees of freedom}$  (instead of the usual  $2 \times \text{degrees of freedom}$ ) may well be appropriate here but requires further study.

Silventoinen et al. (2003b) describe a comparison of height heritability estimates across eight countries, based mainly on self-reported height. The heritabilities reported are similar to those found here for self-reported height. The estimates of the unique environmental variance are similar across countries in Silventoinen et al. (2003b) and it seems likely that in these other countries, in line with the results shown here, approximately half the unique environmental variance was directly attributable to measurement error.

The results presented here can be used to assess the effects of measurement error upon a linkage (quantitative trait locus or QTL) analysis searching for genes underlying the trait of interest. If measurement error can be reduced this will decrease the residual variance, increasing the power to detect linkage in a genome scan. Take for example the scenario where the variance components underlying a trait are as follows: QTL variance 20, polygenic residual variance (genetic variance for all genomic regions unlinked to the QTL) 60 and error variance 20. This may be a reasonable scenario for a relatively large QTL underlying a trait such as height. Assume for now the trait is derived from a single self report measure. In this scenario the non centrality parameter (NCP, in this context a measure of how much information there is for linkage) with a marker completely linked to the QTL is  $\text{NCP} = 0.0083$  per sib pair (Genetic Power Calculator: Purcell et al. 2003). As an example of how this translates into power, with 2,000 sib pairs the power to detect the QTL at the 0.0001 level (i.e. approximately  $\text{LOD} = 3$ ) is 0.64.

Suppose we could reduce the error attributable to inaccurate measurement from the error variance. The results above suggest that approximately half of the error variance is attributable to measurement error. If this can be reduced  $\sim 2.5$ -fold (by either using clinical measures or multiple self-report measures, as discussed previously) so that instead of the residual error variance comprising 10 units measurement error and 10 units environmental effect error we have 4 units measurement error and 10 units environmental effect error. The variance components would hence become QTL variance 20, polygenic variance 60 and error variance 14. With this scenario (i.e. QTL explains 21.3% of variance with this reduction in error variance) the NCP per sib pair is 0.0101. This means that by reducing the error in measurement, the effective sample size for the QTL analysis has effectively increased by 22%

(0.0101/0.0083 = 1.22). If there were more measurements such that the measurement error was further reduced to effectively 0 (with 10 units environmental error remaining) and the computations re-calculated, the NCP increases to 0.00117. This represents an increase in sample size of 41% compared with the original case (0.0117/0.0083 = 1.41). With 2,000 sib pairs, the partially and fully reduced measurement error scenarios correspond to power to detect the QTL at the 0.0001 level of 0.78 and 0.86, respectively. Repeating the above power calculations for a wide range of QTL effect sizes (1–30% of the variance) yields very similar results to those shown above for a QTL explaining 20% of the variance. The above calculations were done using parameters that were realistic for a height QTL analysis but similar results would be obtained for a large number of other traits currently being scanned for QTL. Following a successful linkage analysis, a common next step is association analysis. Decreased measurement error is likely to increase the power of any association analysis to detect QTL although the effective increase in sample size is likely to be less marked than is the case for linkage analysis. This is because in linkage analysis the QTL and shared environmental and/or polygenic factors are explicitly modelled, typically leaving a relatively small residual error variance. This means that decreases in measurement error variance (which form a proportion of the residual error variance) have a relatively large impact on power. In contrast, for association analysis, shared environmental and/or polygenic factors are not modelled; this means a reduction in measurement error will have a smaller effect on power.

When faced with possibly inaccurate and biased variables, researchers should take steps to identify the extent of the problem. If there are multiple self-report and clinical measures, even if only in a subset of the available data, then repeated measure analysis (as outlined in the methods section) will allow the relative magnitudes of the measurement error to be assessed. Failing that, a pilot study may be performed to allow quantification of the measurement error. Since moderate reduction in measurement error will lead to reasonably large increases in effective sample size for linkage analysis (as shown above), investment in a pilot study to help inform phenotype choice may be cost effective in cases where linkage analyses are planned. A suitable pilot study may also be invaluable in assessing any systematic bias in the variables studied; such bias will often take the form of the age-related bias in height we describe here. In the case of height a simple regression based correction for age can be used to remove the effects of age (e.g. we report here that

females overestimate their true height by 1.7 cm by the age of 68). Similarly, other traits can be corrected by considering possible sources of bias in the pilot sample. It is worthwhile stressing, however, that the effects of measured covariates (such as age) can be accounted for in any genetic analysis (e.g. heritability calculation or linkage analysis) by fitting covariates in the analysis. There may be circumstances in which there is clinical data available for some subjects and self-report for others. This situation can be dealt with by fitting an appropriate quantitative genetic model. For example, for a heritability calculation the analysis can be done using a heterogeneous variances model that explicitly takes into account that some measures have more variance than others (for details of the implementation of this see the ASReml manual, Gilmour et al. 2002). This is analogous to fitting an appropriate model that takes into account that observations on males and females differ for traits where this is the case (Neale and Cardon 1992). For linkage analysis, the presence of both clinical and self-report measures presents no analytical problems. However, the contribution of the more accurately measured individuals (in terms of say, NCP per sib pair) will be higher than for the poorly measured individuals.

The genetic component of height was found to be largely additive with estimates for the additive genetic component exceeding 90%. There was some indication of a dominance component to the variation in height but this was not significant despite our large sample size. The large twin sample analysed by Eaves et al. (1999) report a broadly similar conclusion; in their study (focusing solely on the self-report measure of height), the genetic variation was largely additive. Their extended twin design sample allowed them to partition the remaining genetic variation into small components attributable to dominance and to assortative mating. The data here (MZ and DZ twins) only allows estimation of ADE/ACE models and there is no scope for estimation of further components of variance. In particular, there is no scope for the present data set to test for variance associated with assortative mating or epistatic variance components; these components are completely confounded (Eaves et al. 1999). Consistent with the results shown here, the “total genetic” component reported by Eaves et al. (1999) explained ~85% of the variance (based on self-reported height); we predict that this figure would have increased to >90% if clinical measures of height were used.

**Acknowledgments** This research was supported in part by grants from the National Health and Medical Research Council

of Australia (389892 and 389891-PMV, NGM), the United States National Institutes of Health (AA13326-01, AA13446-03, MH66206-01A1 and AA007728, PMV), the Australian Research Council, and by the GenomEUtwin project which is supported by the European Union contract number QLRT-2001-01254. We thank Bert Klei for posing the original question which prompted this work.

## References

- Carey G (1986) Sibling imitation and contrast effects. *Behav Genet* 16(3):319–341
- Cornes BK, Medland SE, Ferreira MA, Morley KI, Duffy DL, Heijmans BT, Montgomery GW et al (2005) Sex-limited genome-wide linkage scan for body mass index in an unselected sample of 933 Australian twin families. *Twin Res Hum Genet* 8(6):616–632
- Eaves L (1976) A model for sibling effects in man. *Heredity* 36(2):205–214
- Eaves L, Heath A, Martin N, Neale M, Meyer J, Silberg J, Corey LA, Truett M et al (1999) Biological and cultural inheritance of stature and attitudes. In: Cloninger C (Ed) *Personality and psychopathology*, American Psychopathological Society London
- Falconer DS, Mackay TFC (1996) *Introduction to quantitative genetics*, 4th edn. Longman, Essex, United Kingdom
- Fredriks AM, Van Buuren S, Burgmeijer RJF, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM (2000) Continuing positive secular growth change in the Netherlands 1955–1997. *Pediatric Res* 47:316–323
- Galloway A, Stini WA, Fox SC, Stein P (1990) Stature loss among an older United-States population and its relation to bone-mineral status. *Am J Phys Anthropol* 83:467–476
- Galton F (1886) Hereditary stature. *Nature* 34:295–298
- Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2002) *ASREML User Guide Release 1.0*. VSN International Ltd, Hemel Hempstead
- Hebert PR, Richedwards JW, Manson JE, Ridker PM, Cook NR, Oconnor GT, Buring JE et al (1993) Height and incidence of cardiovascular-disease in male physicians. *Circulation* 88:1437–1443
- Himes JH, Faricy A (2001) Validity and reliability of self-reported stature and weight of US adolescents. *Am J Hum Biol* 13:255–260
- Kannam JP, Levy D, Larson M, Wilson PWF (1994) Short stature and risk for mortality and cardiovascular-disease events - the Framingham Heart-Study. *Circulation* 90:2241–2247
- Langenberg C, Hardy R, Breeze E, Kuh D, Wadsworth MEJ (2005) Influence of short stature on the change in pulse pressure, systolic and diastolic blood pressure from age 36 to 53 years: an analysis using multilevel models. *Int J Epidemiol* 34:905–913
- Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, USA
- Mukhopadhyay N, Finegold DN, Larson MG, Cupples LA, Myers RH, Weeks DE (2003) A genome-wide scan for loci affecting normal adult height in the Framingham Heart Study. *Hum Hered* 55:191–201
- Neale MC, Cardon LR (1992) *Methodology for genetic studies of twins and families*. Kluwer Academic, Dordrecht
- Padez C (2003) Secular trend in stature in the Portuguese population (1904–2000). *Ann Hum Biol* 30:262–278
- Perola M, Ohman M, Hiekkalinna T, Leppavuori J, Pajukanta P, Wessman M, Koskenvuo M et al (2001) Quantitative-trait-locus analysis of body-mass index and of stature, by combined analysis of genome scans of five Finnish study groups. *Am J Hum Genet* 69:117–123
- Purcell S, Cherny SS, Sham PC (2003) Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149–150
- R Development Core Team (2004) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Austria. ISBN 3-900051-00-3
- Rowland ML (1990) Self-reported weight and height. *Am J Clin Nutr* 52:1125–1133
- Self SG, Liang KY (1987) Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under non-standard conditions. *J Am Stat Assoc* 82:605–610
- Silventoinen K (2003a) Determinants of variation in adult body height. *J Biosoc Sci* 35:263–285
- Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L et al (2003b) Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* 6:399–408
- Silventoinen K, Zdravkovic S, Skytthe A, McCarron P, Herskind AM, Koskenvuo M, de Faire U, Pedersen N, Christensen K, Kaprio J (2006) Association between height and coronary heart disease mortality: a prospective study of 35,000 twin pairs. *Am J Epidemiol* 163:615–621
- Visscher P, Benyamin D, White I (2004) The use of linear mixed models to estimate variance components from data on twin pairs by maximum likelihood. *Twin Res* 7:670–674
- Vogel F, Motulsky A (1997) *Human genetics*. Springer, Berlin Heidelberg New York