

Open Access

Nicole M. Warrington*, Kate Tilling, Laura D. Howe, Lavinia Paternoster,
Craig E. Pennell, Yan Yan Wu and Laurent Briollais

Robustness of the linear mixed effects model to error distribution assumptions and the consequences for genome-wide association studies

Abstract: Genome-wide association studies have been successful in uncovering novel genetic variants that are associated with disease status or cross-sectional phenotypic traits. Researchers are beginning to investigate how genes play a role in the development of a trait over time. Linear mixed effects models (LMM) are commonly used to model longitudinal data; however, it is unclear if the failure to meet the models distributional assumptions will affect the conclusions when conducting a genome-wide association study. In an extensive simulation study, we compare coverage probabilities, bias, type 1 error rates and statistical power when the error of the LMM is either heteroscedastic or has a non-Gaussian distribution. We conclude that the model is robust to misspecification if the same function of age is included in the fixed and random effects. However, type 1 error of the genetic effect over time is inflated, regardless of the model misspecification, if the polynomial function for age in the fixed and random effects differs. In situations where the model will not converge with a high order polynomial function in the random effects, a reduced function can be used but a robust standard error needs to be calculated to avoid inflation of the type 1 error. As an illustration, a LMM was applied to longitudinal body mass index (BMI) data over childhood in the ALSPAC cohort; the results emphasised the need for the robust standard error to ensure correct inference of associations of longitudinal BMI with chromosome 16 single nucleotide polymorphisms.

Keywords: ALSPAC; genome-wide association; longitudinal studies; misspecification; mixed model; robustness.

DOI 10.1515/sagmb-2013-0066

1 Introduction

Over recent years, the study of population genetics has progressed from candidate gene and linkage studies over relatively small regions of the genome to whole genome association analyses. These genome-wide asso-

*Corresponding author: **Nicole M. Warrington**, School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, Australia; and University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia, e-mail: n.warrington@uq.edu.au

Kate Tilling, Laura D. Howe and Lavinia Paternoster: School of Social and Community Medicine, University of Bristol, Bristol, UK; and MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

Craig E. Pennell: School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, Australia
Yan Yan Wu and Laurent Briollais: Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada

ciation studies (GWAS) are designed to search the entire genome for single nucleotide polymorphisms (SNPs) that are associated with a disease or trait of interest. If SNPs are found to be associated, they are then considered to mark a region of the genome that influences the risk of disease or affects the levels of a trait. In general, very small effects are detected so large sample sizes are required. This advance in the scale of genetic analyses has transformed the field from hypothesis driven research to a hypothesis free approach, which has required additional statistical methods to be developed to ensure there is a balance between acceptable levels of power and the chance of inflating the type 1 error. Given the cost of conducting these studies, in terms of both monetary costs for genotyping samples and computational costs for the analysis, it is important that appropriate analyses are conducted from the outset.

To date, most of the GWAS have focused on case/control studies of particular diseases or cross-sectional measurements of phenotypic traits. These study designs typically use relatively simple statistical techniques, such as χ^2 tests or linear (or logistic) regression models, to look at the association between a trait and each of the ~2.5 million SNPs. There are now over 1500 published studies focusing on 250 traits using analyses of this kind (Hindorf et al., 2010). However, researchers are beginning to focus on more complex analyses to uncover additional genetic loci and reduce the currently unexplained heritability of these traits. One area of extension is to use longitudinal studies, with repeated measures on each individual in the study, to understand how SNPs affect changes over time of a particular phenotype (Kerner et al., 2009; Smith et al., 2010; Sikorska et al., 2013). There are several developed statistical methods commonly used for repeated measures data to take into account the non-independence of measurements within an individual. For continuous traits, the most popular statistical method is the linear mixed effects model (LMM) by Laird and Ware (Laird and Ware, 1982). This method can be computationally intensive as the model can account for linear or non-linear trajectories for the outcome of interest over time, correlation between measures at the starting point (intercept) and change over time (slope, or non-linear trajectory) within an individual and adjustment for both time-independent and time-dependent covariates.

In LMMs, the usual assumptions made about the random effects and error distributions include: the random effects and error terms are normally distributed, the random effects are independent of the error term and the error term has homoscedastic variance (Laird and Ware, 1982). In studies utilizing this method to assess the association of a SNP with the trajectory, the fixed effect estimates are often of most interest; the random effects and correlation structure at the individual level are necessary to provide an accurate fit of the model to the data, in addition to providing appropriate test statistics, but are treated as nuisance parameters and are often difficult to interpret. There have been a number of studies investigating whether violations of the assumptions about the random effects and error terms affect the maximum likelihood inference of the fixed effect parameters and their variance estimates; several manuscripts have shown that the fixed effects estimates are robust to non-Gaussian random effects distribution (Verbeke and Lesaffre, 1997; Zhang and Davidian, 2001), non-Gaussian or heteroscedastic error distribution (Jacqmin-Gadda et al., 2007) and that the population fixed effects are robust to misspecified covariance structure (Taylor et al., 1994), but the individual level predictions are not (Taylor and Law, 1998). Jacqmin-Gadda et al. (2007) show the fixed effects estimates are not robust to error variance that is dependent on a covariate in the model that interacts with time. Liang and Zeger (1986) demonstrated that a robust sandwich estimator (Royall, 1986) can correct for biased variance estimates of the fixed effects when the covariance structure is not correctly specified. There has not been any investigation, to our knowledge, into how any of these model misspecifications affect the power and type 1 error in high dimensional studies, for example when running an LMM on a genome-wide scale, and what the value of the robust variance estimator is in this context.

The aim of this study is to assess by simulations whether misspecification of the error term, with either non-Gaussian error distributions or non-constant error variance, in a complex longitudinal model with non-linear trajectories will affect: 1) the coverage probabilities of the 95% confidence interval of the fixed effects parameter estimates; 2) the bias of the fixed effects parameter estimates; 3) the type 1 error of SNP detection in a GWAS; or 4) the statistical power to detect association. We also examined whether our conclusions differ according to minor allele frequency (MAF) for the SNPs or sample size of the investigated cohort.

2 Motivating example

The World Health Organization defines obesity as “abnormal or excessive fat accumulation that presents a risk to health” (World Health Organization, 2012). Obesity is a medical condition which increases an individual’s risk to health problems such as cardiovascular disease, type 2 diabetes and some cancers and therefore reduces life expectancy (Haslam and James, 2005). The prevalence of obesity has been increasing in recent decades in developed countries, particularly in children. Body mass index [BMI; calculated as weight (kg)/height² (m)] is commonly used to define overweight and obesity, with appropriate cut-offs defined for both children (Cole et al., 2000) and adults (WHO, 2000). Childhood obesity is one of the strongest predictors of adult obesity (Serdula et al., 1993; Kindblom et al., 2009). Although the growing prevalence of obesity is most likely to be due to the increasing energy intake and decreasing energy expenditure, twin and adoption studies have provided evidence that BMI is heritable (Maes et al., 1997; Parsons et al., 1999; Haworth et al., 2008; Wardle et al., 2008). Recent GWAS have begun to uncover some plausible genetic loci contributing to higher BMI (Fox et al., 2007; Frayling et al., 2007; Loos et al., 2008; Thorleifsson et al., 2009; Willer et al., 2009; Liu et al., 2010; Speliotes et al., 2010) and obesity in children (Bradfield et al., 2012), with 34 new loci identified. However, none of the studies to date provide information regarding the genetic determinants of the rate of BMI growth over childhood, which leads to obesity.

The Avon Longitudinal Study of Parents and Children (ALSPAC) (Boyd et al., 2013; Fraser et al., 2013) is a birth cohort study; 14,541 pregnant women in the former county of Avon, UK, were recruited into the study if they had an expected delivery date between 1st April 1991 and 31st December 1992. From birth to 5 years, length and weight measurements were extracted from health visitor records, with up to four measurements taken on average at six weeks, 10, 21, and 48 months of age. For a random 10% of the cohort, length and height measurements were taken in eight research clinic visits held between the ages of 4 months and 5 years of age. From age 7 years upwards, all children were invited to annual research clinics from ages 7 to 11 and biannual research clinics thereafter. Details of measuring equipment used in the clinics is described elsewhere (Howe et al., 2010). In addition, parent-reported child height and weight were also available from questionnaires (27% of measurements). Whilst the measurements from routine health care have previously been shown to be accurate in this cohort (Howe et al., 2009), parental report of children’s height tends to be overestimated while weight tends to be under estimated (Dubois and Girard, 2007). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary (<http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>).

A subset of 7916 participants were used for analysis based on the following inclusion criteria: at least one parent of European descent, singleton birth, unrelated to anyone in the sample, genome-wide genotype data, and at least one measure of BMI throughout childhood. Participants have a median of 9 BMI measurements between 1 and 15 years of age (interquartile range 5–12, range 1–29 measurements). Children tend to have rapidly increasing BMI from birth to approximately 9 months of age where they reach their adiposity peak; BMI then decreases until around the age of 5–7 years at adiposity rebound and then steadily increases again until after puberty where it tends to plateau through adulthood. There is a large amount of variability between individuals for both intercept and slope.

The primary research question is to identify SNPs that are associated with average BMI and change in BMI over childhood and adolescence in the ALSPAC data. A LMM was used to appropriately model the longitudinal trajectory over childhood, to account for the large correlation between each of the random effects parameters, to adjust for additional covariates such as the source of the height/weight measurements (clinic or questionnaire) and to allow data to be missing at random across childhood. The general form of the model is as follows:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \quad (1)$$

where Y_i is the response vector for the i th individual, β is the vector of fixed effects and $b_i \sim N(0, \Sigma)$ is the vector of subject specific random effects, X_i and Z_i are the fixed effect and random effect regressor matrices respectively and $\varepsilon_i \sim N(0, \sigma^2)$ is the within subject error vector. When applying this model to the ALSPAC data, the

best model fit included a cubic polynomial of mean centred age (centred at age 8 years) in the fixed effects, a quadratic polynomial of mean centred age in the random effects and a continuous autoregressive correlation structure of order one for the covariance of the within-subject errors. Hence, the final model for both females and males was:

$$\text{BMI}_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + \beta_4 \text{MS}_{ij} + \beta_5 \text{SNP}_i + \beta_6 t_{ij} \text{SNP}_i + \beta_7 t_{ij}^2 \text{SNP}_i + \beta_8 t_{ij}^3 \text{SNP}_i + b_{i0} + b_{i1} t_{ij} + b_{i2} t_{ij}^2 + \varepsilon_{ij} \quad (2)$$

where MS is the measurement source (i.e., clinical visit or questionnaire) of individual i at time j and t_{ij} is the age (centred at 8). Therefore β_0 is the population intercept (i.e., mean BMI at age 8), β_1 – β_3 are the fixed effects for the cubic function of age, β_4 is the measurement source, β_5 is the change in the mean BMI at 8 years of age for each additional copy of the minor allele, β_6 is the SNP by linear age effect, β_7 and β_8 are the SNP by quadratic and cubic effects respectively.

Due to the nature of the data collection, which is often complex in large cohort based studies, we found that the model assumptions were not met due to the following:

1. The questionnaire measures have previously been shown to have greater variability than clinic measured height and weight (Dubois and Girad, 2007); therefore we had variability that was dependent on a covariate in the model.
2. There were only questionnaire measures available around the nadir of the trajectory (also known as the adiposity rebound), which meant we had greater variability around the rebound.
3. The variability within individuals changes over time; particularly with increased variability around puberty and into adolescence.
4. BMI also has a non-Gaussian error distribution. This is in part due to the increasing variability between individuals over time, with some individuals having rapidly increasing BMI while others remain relatively consistent.

In the following, we investigate the robustness of the maximum likelihood inference for the fixed effects, the type 1 error and the power for detecting an association with the SNP when the error distribution is misspecified due to the above intricacies of the data.

3 Simulation study

We carried out extensive simulations to investigate the effects on the LMM when the error term (also called the level-1 residual, or the occasion-level residual) in the model was non-Gaussian or had a non-constant variance. In each of the simulation scenarios, we set the non-genetic fixed effects parameters [β_0 – β_4 from model (2)] and the variance-covariance matrix similar to those coming from the fitted model for BMI adjusting for the *FTO* rs1121980 SNP in the ALSPAC study; these can be found in Table 1. The measurement source, which is a fixed effect in the LMM and used in the heteroscedastic error simulations, was a randomly generated binary variable for each individual at each time point with distribution throughout the ages similar to the distribution in the ALSPAC cohort (percent questionnaire measurements per follow-up year: year 1=40%, year 2=20%, year 3=40%, year 4=10%, year 5=60%, year 6=99%, year 7=10%, year 8=0%, year 9=0%, year 10=10%, year 11=0%, year 12=0%, year 13=30%, year 14=0%, year 15=0%).

We also investigated the fixed effect estimation for various sample sizes, minor allele frequencies of the SNP and the SNP effect sizes:

1. Sample size: two levels; $n=1000$ and $n=3000$
2. Minor allele frequency: four levels; 0.1, 0.2, 0.3 and 0.4
3. Effect sizes: two combinations; $\beta_5=0.6$, $\beta_6=0.15$, $\beta_7=-0.000752$ and $\beta_8=-0.000380$ (alternative hypothesis) or $\beta_5=\beta_6=\beta_7=\beta_8=0$ (null hypothesis). The alternative hypothesis effect sizes for β_5 and β_6 were chosen to have 80% power to detect with the larger sample size; the effect sizes for β_7 and β_8 were similar to those coming from the fitted model for BMI adjusting for the *FTO* rs1121980 SNP in the ALSPAC study.

Table 1 Parameter estimates from the ALSPAC non-genetic model used to generate the data in the simulation study.

Effect	Parameter	Value
Intercept	β_0	16.534
Age	β_1	0.400
Age2	β_2	0.056
Age3	β_3	-0.003
Source	β_4	-0.153
SD(b_0)	σ_0	2.092
SD(b_1)	σ_1	0.269
SD(b_2)	σ_2	0.0235
Cor(b_0, b_1)	ρ_0	0.820
Cor(b_0, b_2)	ρ_1	-0.389
Cor(b_1, b_2)	ρ_2	-0.092
SD(ε)	σ	1.063
Correlation structure	ρ	0.394

3.1 Sampling designs

As many longitudinal cohorts have different sampling designs, some with variable amounts of missing time points and missing observations at each time point, we investigated five different sampling designs:

1. Sparse complete: $n_i=8$ measures per person with few measures around the adiposity rebound; times of measures are 1, 2, 3, 5, 8, 10, 13, 15
2. Intense complete: $n_i=14$ measures per person with multiple measures around the adiposity rebound; times of measures are 1, 2, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7, 9, 11, 13, 15
3. Equal unbalanced: $n_i=1-15$ measures per person between 1 and 15 years with a mean of 9 measures (proportion of missingness=0.4 across whole age range)
4. Unbalanced with more samples around the adiposity rebound: $n_i=1-15$ measures per person between 1 and 15 years with a mean of 9 measures; proportion of missingness around adiposity rebound of 0.2 and 0.45 outside the 5–7 year age range (average proportion of missingness over whole age range is 0.4)
5. Unbalanced with fewer samples around the adiposity rebound: $n_i=1-15$ measures per person between 1 and 15 years with a mean of 9 measures; proportion of missingness around adiposity rebound of 0.6 and 0.35 outside the 5–7 year age range (average proportion of missingness over whole age range is 0.4)

The first two designs with complete data at each follow-up assume that every individual had the exact same age at follow-up (i.e., came into clinic on their birthday), whereas the other three designs are more representative of longitudinal studies where the actual age of measurement varies between individuals by up to a year (i.e., came into clinic either 6 months before or after a birthday). We assume data is missing completely at random, that is that the probability that an observation is missing for a given individual is independent of all other observed data. The proportion of missingness simulated across the whole range (i.e., 0.4) was equivalent to the amount of missing data observed in the ALSPAC cohort under the assumption that all individuals could have been measured yearly. We used a fully factorial design for the simulations with the 3 data characteristics and the 5 sampling designs.

3.2 Models for data generation

Standard linear mixed model:

Data were generated with Gaussian random effects and error distribution to validate the estimation method.

Non Gaussian error:

Three error structures were investigated:

1. t -distribution: t with 5 degrees of freedom
2. skew-normal distribution: $SN(1.063^2, 40)$
3. Asymmetric mixture of two Gaussian distributions: $0.3N(-0.67, 1^2) + 0.7N(0.5, 0.3^2)$

Heteroscedastic error:

Three cases were studied:

1. Variance dependent on a covariate: $\text{Var}(e_{ij}) = \sigma_e^2 a X_{ij}$
where $\sigma_e^2 = 1.131$, $a = 1.500$ and $X_{ij} = 1$ if measure was from questionnaire and 0 if measure was from a follow-up clinic
2. Variance greater at the adiposity rebound: $\text{Var}(e_{ij}) = \sigma_e^2 a X_{ij}$
where $\sigma_e^2 = 1.131$, $a = 1.500$ and $X_{ij} = 1$ if measure was between 5 and 7 years and 0 if not
3. Variance increasing over time: $\text{Var}(e_{ij}) = \sigma_e^2 a^{t_{ij}}$
where $\sigma_e^2 = 1.131$ and $a = 1.150$

3.3 Data generation

We simulated 1000 datasets under the alternative hypothesis ($\beta_5 = 0.6$ and $\beta_6 = 0.15$) to look at coverage probabilities, bias and power and 5000 datasets under the null hypothesis ($\beta_5 = 0$ and $\beta_6 = 0$) to look at type 1 error at $\alpha = 0.05$. Each SNP (coded as 0, 1, 2) was incorporated into the model assuming an additive genetic model, whereby each additional minor allele increases BMI by an equal amount. We were primarily interested in estimating the SNP main effect, β_5 , which represents the increase on the mean BMI at 8 years of age for each additional copy of the minor allele and the SNP by age effect, β_6 (referred to as the SNP*age interaction), which represents the effect on the mean linear increase of BMI (slope) for each additional minor allele. We calculated a robust standard error for each fixed effect parameter and corresponding p-value; the following formula was used:

$$(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \left(\sum_{i=1}^S \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right) (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$$

Where:

\mathbf{X} is the fixed effect regressor matrix from equation (1)

$\hat{\mathbf{V}}$ is the estimated variance of \mathbf{Y} from equation (1)

$$\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}$$

S is the number of subjects and i is the i th subject

In addition to the fixed effects parameters, we applied a Wald test to assess whether the overall SNP effect was affected by the misspecification. The Wald test was estimated using the General Linear Hypothesis approach (McDonald, 1975). This approach is based on the normal approximation for maximum likelihood estimators using the estimated variance-covariance matrix. The hypothesis can be specified through a constant matrix L to be matched with the fixed effects of the model such that $H_0: L\boldsymbol{\beta} = \mathbf{m}$ where the \mathbf{m} are the hypothesized values. The estimates of the fixed effects, $\boldsymbol{\beta}$, asymptotically follow a multivariate normal distribution $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \text{cov}(\hat{\boldsymbol{\beta}}))$ by the Central Limit Theorem such that the linear form also asymptotically follows a multivariate normal distribution:

$$L\hat{\boldsymbol{\beta}} \sim N(L\boldsymbol{\beta}, L\text{cov}(\hat{\boldsymbol{\beta}})L') \quad (3)$$

Thus the 95% confidence interval and corresponding P-value for the hypothesized value can be obtained accordingly. We tested whether the parameters for the SNP were simultaneously equal to zero. It is computationally intensive to calculate a robust estimate for the Wald test; for example, the robust standard error for the fixed effects takes approximately 7 min for the rs1121980 *FTO* SNP in the ALSPAC data whereas the robust standard error for the global Wald test takes approximately an additional 3 min. These computational

times decrease exponentially as sample size and the number of repeated measures per individual decreases; however, they may not be scalable to a GWAS study. To investigate whether a robust standard error would be beneficial for the global Wald test, we selected the scenario where the inflation was greatest and calculated the robust estimates for all the simulations in this scenario. All analyses were conducted in *R* version 2.12.1 (Ihaka and Gentleman, 1996) using the nlme package.

As it is important to report the uncertainty in any estimates from simulation based studies (Koehler et al., 2009), Monte Carlo error (MCE) was calculated using the joint performance method of $\hat{\beta}$ and s_i outlined in White (2010). A confidence interval for coverage probabilities, bias, type 1 error and power was calculated using the following:

$$\hat{P} \pm 1.96 \sqrt{\left(\frac{P(1-P)}{S} \right)} \quad (4)$$

Where P is the α -level, for example P for coverage estimates is 0.95 and P for type 1 error is 0.05, and S is the number of simulations, for example either 1000 or 5000. The output from the simulations was then assessed as to whether they fell within this confidence interval.

3.4 Results

3.4.1 Coverage probabilities

Coverage probability can indicate whether the confidence interval of the parameter(s) of interest is conservative (i.e., the coverage probability is larger than the nominal confidence interval) or liberal (i.e., the coverage probability is narrower than the nominal confidence interval).

Coverage probabilities for the 95% confidence interval of the fixed effects parameter estimates from each of the simulations are presented in Table 2. No consistent differences were seen across the range of minor allele frequencies, so the results from each of the simulated datasets were combined for ease of presentation; however the coverage probabilities for each of the minor allele frequencies are presented in Supplementary Table 1.

The coverage probabilities of the SNP main effects parameter for all simulations appear to be unaffected by the error misspecifications; only nine of 70 coverage probabilities were significantly different from 95%, that is less than 94.32% or greater than 95.68%, five of which were from the simulations where the error variance increases over time.

Thirty-one of the 70 coverage probabilities (44%) for the SNP*age interaction parameter were significantly different from 95%, with both the non-Gaussian and heteroscedastic error distributions being affected. When the error variance followed a *t* distribution, the coverage probabilities for the confidence interval of the SNP*age interaction parameter are <95% in all designs except the sparse complete scenario. Similarly, the SNP*age interaction parameter had coverage probabilities <95% when the error variance followed a skew-normal distribution, however only in the unbalanced designs where there is missing data. The coverage probabilities were <95% when the error variance was both dependent on a covariate and increased over time, in both the complete and unbalanced designs. All the coverage probabilities that significantly differ from 95% for the SNP*age interaction parameter have underestimated variance estimates and thus confidence intervals that were too narrow, which could lead to test statistics that are too liberal.

3.4.2 Bias

The SNP main effect and the SNP*age interaction parameters are unbiased in the majority of the simulations, indicating that the misspecifications in the error distribution do not affect the estimates of the β 's (Supplementary Tables 2 and 3). Only nine of 140 95% confidence intervals did not cover zero; these nine confidence intervals were across the range of error distributions and designs, showing that no one scenario was particularly biased.

Table 2 Coverage rates of the 95% confidence intervals of the fixed effects; bold and underlined cells are those that are significantly different from the nominal 95% based on 4000 simulations under each design (1000 simulations for each MAF combined into one summary statistic).

Sampling design	Sparse complete		Intense complete		Equal unbalanced		Unbalanced with more samples around the adiposity rebound		Unbalanced with less samples around the adiposity rebound	
	n=1000	n=3000	n=1000	n=3000	n=1000	n=3000	n=1000	n=3000	n=1000	n=3000
Gaussian Distribution										
SNP	95.43	95.03	95.08	95.45	94.83	95.23	95.08	94.70	95.40	94.73
SNP*age	95.00	95.23	94.58	95.13	94.35	94.63	<u>94.30</u>	<u>93.90</u>	94.53	94.35
t-distribution										
SNP	95.45	95.35	<u>95.90</u>	94.55	95.13	94.85	94.65	94.48	95.48	94.95
SNP*age	95.30	94.80	<u>94.05</u>	<u>94.13</u>	94.45	<u>94.10</u>	<u>93.70</u>	<u>94.00</u>	<u>93.33</u>	<u>94.03</u>
Skew-normal Distribution										
SNP	94.90	95.03	95.18	95.10	95.05	<u>94.25</u>	95.43	94.95	94.85	94.75
SNP*age	95.68	95.18	94.63	94.65	<u>93.88</u>	<u>93.73</u>	94.73	<u>94.13</u>	<u>93.90</u>	<u>93.55</u>
Mixture of 2 Gaussian Distributions										
SNP	94.85	94.98	94.83	95.65	95.00	95.08	94.53	95.40	94.33	94.48
SNP*age	95.05	94.78	95.03	94.60	95.20	<u>94.08</u>	94.58	95.20	94.80	<u>94.10</u>
Variance dependent on a covariate										
SNP	94.93	95.05	<u>95.83</u>	95.35	94.43	94.75	94.70	94.93	94.98	94.93
SNP*age	94.93	95.03	94.95	<u>94.15</u>	<u>94.03</u>	<u>94.10</u>	<u>93.75</u>	94.53	<u>93.95</u>	<u>93.93</u>
Variance greater at adiposity rebound										
SNP	94.75	95.23	95.15	95.08	<u>94.25</u>	95.20	95.48	95.35	95.23	94.43
SNP*age	<u>94.05</u>	94.45	95.38	95.38	<u>94.13</u>	94.43	<u>94.00</u>	94.75	94.73	94.60
Variance increasing over time										
SNP	<u>94.20</u>	95.00	<u>94.08</u>	94.38	94.80	94.33	<u>94.30</u>	<u>94.03</u>	<u>95.98</u>	95.48
SNP*age	<u>94.10</u>	94.88	<u>91.78</u>	<u>92.38</u>	94.70	<u>94.23</u>	<u>93.28</u>	<u>93.48</u>	95.65	95.25

No consistent differences were seen in the bias estimates across the range of minor allele frequencies; however, the 95% confidence intervals for the difference between the simulated parameter and the true parameter were tighter as the sample size and minor allele frequency increased (Supplementary Table 4).

3.4.3 Type 1 error

As seen with the coverage probabilities, no consistent differences in type 1 error were evident across the minor allele frequency range, so the results from each of the simulated datasets were combined for ease of presentation (Tables 3 and 4); however the type 1 error for each of the minor allele frequencies tested are given in Supplementary Table 5.

As seen in Table 3, the type 1 error for the complete designs remained within acceptable limits of the nominal alpha level. We observed inflation for the SNP by age interaction parameter in several cases, but this inflation was reduced to nominal levels by using a robust standard error.

Table 4 shows that the type 1 error for the SNP by age interaction was often inflated under the unbalanced designs. However, by using a robust standard error, the inflation seen can be reduced to nominal levels in the majority of cases; approximately 75% of the inflated effects were reduced. The design where the robust standard error didn't seem to have an effect was when the error variance increased over time; only 20% of the estimates were reduced to nominal levels under this design. Interestingly, the robust standard error did not appear to affect the type 1 error for the scenarios that were not originally inflated.

To declare significance in a GWAS, several thresholds are commonly used: suggestive association, significant association and highly significant association. Duggal et al. define suggestive associations as SNPs that reach a P-value threshold under the assumption that one false positive association is expected per GWAS

Table 3 Type 1 error for the complete designs; bold and underlined cells are those that are significantly different from the nominal $\alpha=0.05$ based on 20,000 simulations under each design (5000 simulations for each MAF combined into one summary statistic).

Sampling design	Sparse complete				Intense complete			
	n=1000		n=3000		n=1000		n=3000	
	Standard	Robust	Standard	Robust	Standard	Robust	Standard	Robust
Gaussian Distribution								
SNP	0.0514	0.0528	0.0509	0.0513	0.0502	0.0521	0.0500	0.0510
SNP*age	0.0483	0.0504	0.0483	0.0491	<u>0.0549</u>	0.0486	<u>0.0539</u>	0.0467
Global wald test	0.0497		0.0478		<u>0.0605</u>		<u>0.0620</u>	
t-distribution								
SNP	0.0495	0.0498	0.0489	0.0496	0.0479	0.0510	0.0483	0.0502
SNP*age	0.0521	0.0534	0.0487	0.0492	<u>0.0581</u>	0.0490	<u>0.0563</u>	0.0465
Global wald test	<u>0.0531</u>		0.0508		<u>0.0624</u>		<u>0.0629</u>	
Skew-normal Distribution								
SNP	0.0502	0.0517	0.0524	0.0524	0.0509	0.0526	0.0525	0.0532
SNP*age	0.0503	0.0519	<u>0.0461</u>	0.0474	<u>0.0541</u>	0.0508	0.0529	0.0486
Global wald test	0.0493		0.0488		<u>0.0621</u>		<u>0.0579</u>	
Mixture of 2 Gaussian Distributions								
SNP	0.0498	0.0504	0.0479	0.0479	0.0485	0.0499	0.0510	0.0508
SNP*age	0.0502	0.0510	0.0492	0.0488	0.0528	0.0506	0.0529	0.0495
Global wald test	0.0498		0.0508		<u>0.0615</u>		<u>0.0586</u>	
Variance dependent on a covariate								
SNP	0.0523	0.0527	0.0488	0.0490	0.0485	0.0511	0.0459	0.0485
SNP*age	<u>0.0546</u>	0.0527	<u>0.0531</u>	0.0514	0.0520	0.0493	0.0524	0.0481
Global wald test	0.0515		0.0525		<u>0.0556</u>		<u>0.0546</u>	
Variance greater at adiposity rebound								
SNP	0.0472	0.0478	0.0511	0.0519	0.0477	0.0493	0.0471	0.0490
SNP*age	0.0528	0.0497	0.0570	0.0528	0.0513	0.0513	0.0491	0.0487
Global wald test	0.0527		<u>0.0540</u>		0.0502		0.0478	
Variance increasing over time								
SNP	0.0523	<u>0.0536</u>	0.0471	0.0473	<u>0.0543</u>	0.0513	<u>0.0561</u>	0.0522
SNP*age	<u>0.0564</u>	<u>0.0538</u>	0.0522	0.0491	<u>0.0746</u>	0.0528	<u>0.0746</u>	0.0530
Global wald test	<u>0.0875</u>	0.0549	<u>0.0875</u>	0.0497	<u>0.1667</u>	0.0506	<u>0.1685</u>	0.0506

(Duggal et al., 2008); SNPs reaching this threshold are taken forward for replication. In the context of our simulation study, this definition would equate to a P-value of 0.00005 (1/20,000; where 20,000 is the number of simulations per design and error assumption). The scenario with the highest type 1 error inflation using the classical standard error was for the SNP*age interaction under the intense design where the error variance increased over time (0.0746 for both n=1000 and 3000). In this scenario, 6 SNPs would falsely reach the definition of “suggestive association” for the SNP*age interaction parameter when using the classical standard error with a sample size of 1000 individuals. In contrast, when the model assumptions are met, that is when the error distribution follows a Gaussian distribution with constant variance, only 2 SNPs met the “suggestive association” threshold, indicating an inflation in the type 1 error for the simulations where the variance increased over time due to the misspecification of the error term. When using the robust standard error under the increasing variance over time design, 1 SNP would meet the criteria, showing not only a reduction in the type 1 error from the 7 SNPs seen with the classical standard error, but also a reduction in power in comparison to the model where the assumptions were met.

These results show that there is greater inflation in the type 1 error for the SNP*age interaction in the unbalanced designs than the complete designs. As outlined in Supplementary Figure 1, we simulated additional data to investigate what aspects of the unbalanced design contributed to the inflation. Briefly, the results from these additional simulations are as follows:

Table 4 Type 1 error for the unbalanced designs; bold and underlined cells are those that are significantly different from the nominal $\alpha=0.05$ based on 20,000 simulations under each design (5000 simulations for each MAF combined into one summary statistic).

Sampling design	Equal unbalanced						Unbalanced with more samples around the adiposity rebound						Unbalanced with less samples around the adiposity rebound					
	n=1000			n=3000			n=1000			n=3000			n=1000			n=3000		
	Standard	Robust		Standard	Robust		Standard	Robust		Standard	Robust		Standard	Robust		Standard	Robust	
Gaussian Distribution																		
SNP	0.0518	<u>0.0532</u>		0.0500	0.0508		0.0503	0.0521		0.0478	0.0490		0.0529	<u>0.0550</u>		<u>0.0540</u>	0.0542	
SNP*age	<u>0.0581</u>	0.0526		<u>0.0592</u>	<u>0.0531</u>		<u>0.0566</u>	0.0514		<u>0.0556</u>	0.0496		<u>0.0560</u>	0.0511		<u>0.0575</u>	0.0509	
Global wald test	<u>0.0646</u>			<u>0.0598</u>			<u>0.0601</u>			<u>0.0615</u>			<u>0.0621</u>			<u>0.0609</u>		
t-distribution																		
SNP	0.0510	0.0522		0.0491	0.0497		0.0485	0.0500		0.0495	0.0505		0.0487	0.0499		0.0516	0.0523	
SNP*age	<u>0.0571</u>	0.0487		<u>0.0629</u>	<u>0.0539</u>		<u>0.0596</u>	0.0508		<u>0.0571</u>	0.0475		<u>0.0563</u>	0.0487		<u>0.0577</u>	0.0489	
Global wald test	<u>0.0607</u>			<u>0.0621</u>			<u>0.0620</u>			<u>0.0583</u>			<u>0.0587</u>			<u>0.0605</u>		
Skew-normal Distribution																		
SNP	0.0493	0.0508		0.0495	0.0501		0.0498	0.0517		0.0473	0.0481		0.0512	0.0519		0.0482	0.0484	
SNP*age	<u>0.0548</u>	0.0492		<u>0.0589</u>	0.0526		<u>0.0580</u>	0.0512		<u>0.0571</u>	0.0498		<u>0.0593</u>	<u>0.0532</u>		<u>0.0547</u>	0.0490	
Global wald test	<u>0.0618</u>			<u>0.0582</u>			<u>0.0616</u>			<u>0.0583</u>			<u>0.0632</u>			<u>0.0575</u>		
Mixture of 2 Gaussian Distributions																		
SNP	0.0519	0.0527		0.0490	0.0490		0.0505	0.0510		0.0519	0.0517		0.0510	0.0522		0.0487	0.0483	
SNP*age	<u>0.0534</u>	0.0517		0.0487	<u>0.0459</u>		0.0510	0.0494		<u>0.0538</u>	0.0518		<u>0.0543</u>	0.0511		<u>0.0551</u>	0.0517	
Global wald test	<u>0.0579</u>			<u>0.0581</u>			<u>0.0605</u>			<u>0.0603</u>			<u>0.0589</u>			<u>0.0568</u>		
Variance dependent on a covariate																		
SNP	0.0495	0.0515		0.0482	0.0491		0.0498	0.0513		0.0506	0.0509		0.0512	0.0518		0.0528	0.0502	
SNP*age	<u>0.0586</u>	0.0499		<u>0.0607</u>	0.0514		<u>0.0576</u>	0.0505		<u>0.0588</u>	0.0497		<u>0.0605</u>	0.0507		<u>0.0604</u>	0.0507	
Global wald test	<u>0.0589</u>			<u>0.0611</u>			<u>0.0597</u>			<u>0.0567</u>			<u>0.0620</u>			<u>0.0583</u>		
Variance greater at adiposity rebound																		
SNP	0.0493	0.0504		0.0492	0.0495		0.0486	0.0498		0.0516	0.0526		0.0506	0.0514		0.0496	<u>0.0531</u>	
SNP*age	<u>0.0570</u>	0.0491		<u>0.0563</u>	0.0483		<u>0.0546</u>	0.0482		<u>0.0563</u>	0.0503		<u>0.0561</u>	0.0483		<u>0.0600</u>	0.0505	
Global wald test	<u>0.0572</u>			<u>0.0559</u>			<u>0.0568</u>			<u>0.0541</u>			<u>0.0588</u>			<u>0.0569</u>		
Variance increasing over time																		
SNP	<u>0.0533</u>	<u>0.0545</u>		0.0500	0.0502		<u>0.0564</u>	<u>0.0563</u>		0.0530	0.0520		0.0491	0.0523		0.0500	0.0526	
SNP*age	<u>0.0554</u>	<u>0.0536</u>		<u>0.0571</u>	<u>0.0540</u>		<u>0.0643</u>	<u>0.0570</u>		<u>0.0610</u>	0.0527		0.0497	<u>0.0534</u>		0.0497	0.0513	
Global wald test	<u>0.0911</u>	<u>0.0576</u>		<u>0.0929</u>	0.0529		<u>0.1031</u>	<u>0.0578</u>		<u>0.1011</u>	<u>0.0548</u>		<u>0.0850</u>	<u>0.0559</u>		<u>0.0801</u>	0.0520	

1. The unbalanced designs differed from the complete designs by including missing data and altering the measurement times so they fell within a range around the scheduled times, both of which are inherent in cohort studies. The additional simulations showed the inflation was greater in the presence of missing data rather than because of the different measurement times between individuals (Supplementary Figure 2).
2. Since the LME is known to be robust to missing data under the missing at random and missing completely at random assumptions, we simulated additional data varying the polynomial function of age in the fixed and random effects. These simulations showed the type 1 error was reduced to nominal levels when the fixed and random effects had the same function of age, i.e., cubic function in both the fixed and random effects (Supplementary Table 6).
3. To determine whether there is remaining inflation in the type 1 error after modelling the same function of age in the fixed and random effects when the error distribution is misspecified we simulated additional data using the equal unbalanced sampling design. These simulations showed that the type 1 error was again reduced to nominal levels when the fixed and random effects had the same function of age regardless of the misspecification in the error distribution (Supplementary Table 7).
4. It is often difficult to estimate higher order terms in the random effects when using real data due to computational and convergence issues. In this case, it is often only possible to fit a lower-order polynomial functions in the random effects than the fixed effects. We simulated additional data where the fixed and random effects included a quadratic function for age but we analyzed the data with a quadratic function in the fixed effects and a linear function in the random effects. In addition, we also simulated data where the fixed effects included a quadratic function for age and the random effects included only a linear function but analyzed the data with a quadratic function in both the fixed and random effects. These simulations showed that the type 1 error was inflated when the analysis model had lower order terms of polynomial function in the random effects compared to the fixed effects terms (Supplementary Table 8).

In summary, it is recommended that one includes the same polynomial function for age in the fixed and random effects to avoid inflation in the type 1 error; however, if this is not possible due to non-convergence of the model then a robust standard error is required to reach nominal levels of type 1 error.

The global Wald test, which is assessing whether there is any genetic effect on the whole BMI growth trajectory, was inflated above the acceptable limits under all error variance misspecifications and even under the Gaussian/constant variance assumption, except under the sparse complete design. The scenario where the error variance increased over time showed the largest inflation; however, using the robust estimates for the Wald test under this scenario were also reduced to nominal levels in most designs; if it wasn't reduced to nominal levels it was dramatically lower than using the classical test (Table 4). Again, having the same structure of fixed and random terms for the age polynomial function would yield nominal type 1 errors.

Given that many researchers investigating GWAS of longitudinal traits are interested in only the SNP main effect and not the SNP*age interaction (Furlotte et al., 2012), we conducted some additional simulations without the SNP*age interaction. Once again, we used the scenario where the error variance increased over time and where there was equal unbalance in the data structure. We found that the type 1 error was within the nominal range for the SNP main effect for both sample sizes ($n=1000$: 0.0506; $n=3000$: 0.0515), where previously we saw inflation for the sample size of 1000 (0.0533 from Table 4). We have no reason to believe that any of the other scenarios would be affected by the misspecifications when the SNP*age interactions are not modelled.

3.4.4 Power

Effect sizes for the alternative hypothesis ($\beta_5=0.6$ and $\beta_6=0.15$) were chosen to have 80% power with a MAF of 0.4 and sample size of 1000 when the error from the fitted LMM follows a Gaussian distribution with constant variance. Therefore, the power for all error distributions and MAFs in the simulations with sample size of 3000 was greater than 80%; so this section will only discuss power for the simulations with a sample size of

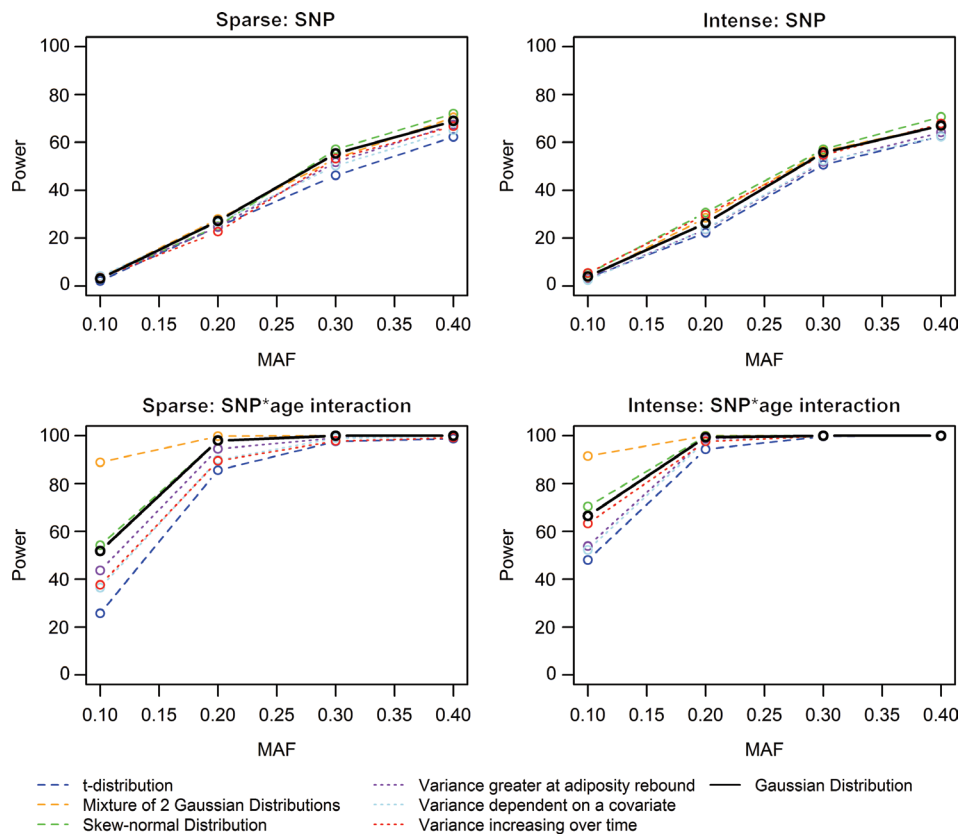


Figure 1 Simulated power of the SNP main effect and SNP*age interaction terms for complete designs. The two plots on the left are for the sparse complete design, while the two plots on the right are from the intense complete design. The solid black line for the Gaussian Distribution is the situation where the model is correctly specified.

1000. Power for the SNP main effect and SNP*age interaction parameters are displayed in Figures 1 (complete designs) and 2 (unbalanced designs).

As expected, the power increases with the MAF. Interestingly, assuming the error distribution has a t-distribution led to lower power for both the SNP main effect and the SNP*age interaction parameters than assuming a Gaussian error distribution. This pattern was consistent across all of the sampling designs; however it appears that the power is slightly closer to that of the error with the Gaussian distribution when there is more data around the adiposity rebound (i.e., the intense complete and unbalanced with more samples around the adiposity rebound). In addition, the simulations where the error distribution follows a skew-normal distribution led to slightly higher power for both the SNP and SNP*age interaction parameters than with the Gaussian error.

When investigating the different error variance structures, the power for the SNP main effect parameter across all MAFs was slightly lower than the power when the constant variance assumption was met. Likewise, for the SNP*age interaction parameter, all of the error variance structures led to lower power than when the constant variance assumption was met. However, simulations under the unbalanced designs where the variance increased over time suffered the most and had notably reduced power until a MAF of approximately 0.3.

3.4.5 Power under the robust standard error

We have shown that using the robust standard error doesn't affect those situations where the type 1 error wasn't initially inflated, however before adopting the robust standard error for a GWAS analysis we also wanted to determine whether using the robust standard error would decrease our power to detect a statistically significant association.

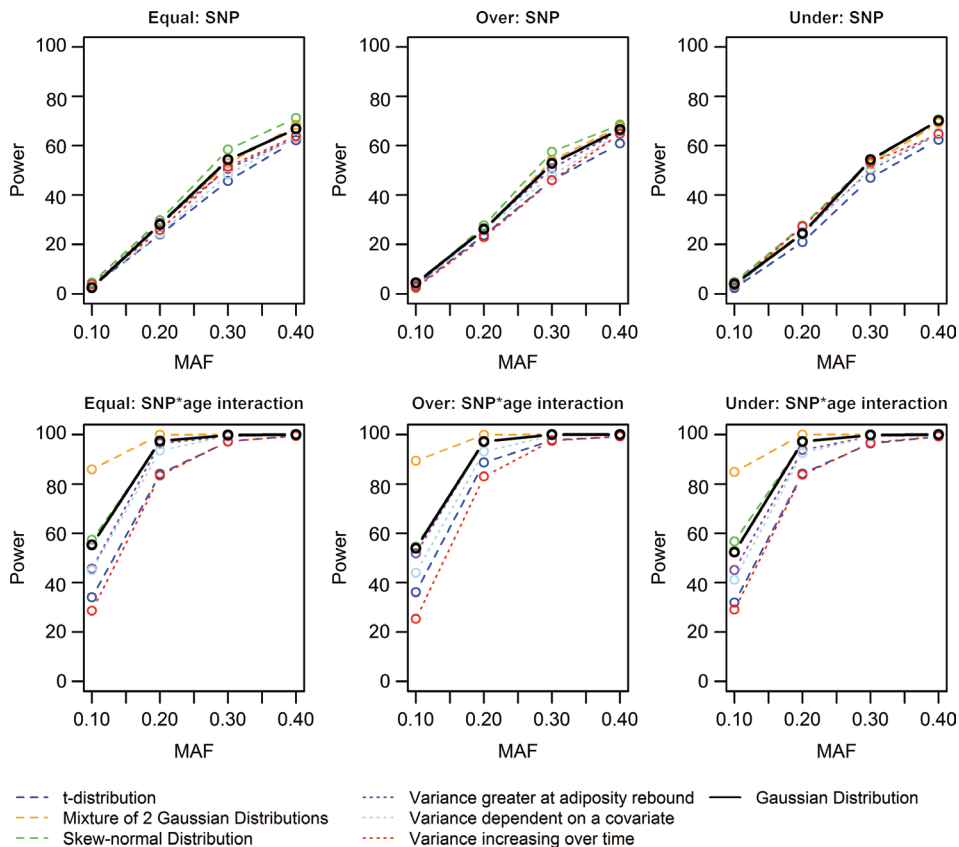


Figure 2 Simulated power of the SNP main effect and SNP*age interaction terms for unbalanced designs. “Equal” is the simulations from the equal unbalanced design, “Over” are the simulations from the unbalanced design with less samples around the adiposity rebound and “Under” are the simulations from the unbalanced design with more samples around the adiposity rebound. The solid black line for the Gaussian Distribution is the situation where the model is correctly specified.

The power for the SNP main effect parameter remains almost unchanged when using the robust standard error rather than the normal standard error in all scenarios and under all model misspecifications (Figures 3 and 4). The only scenario where the power decreased for the SNP main effect parameter by using the robust standard error was where there was increasing variance over time under the intense complete scenario. Given that the type 1 error was not inflated using either standard error estimate, there appears to be no harm in using a robust standard error for estimation even when not required.

The power for the SNP*age interaction parameter, particularly for low MAF, is much more variable. Under the sparse complete design, where there was no inflation in the type 1 error, the power remains about the same using either the classical or robust standard error. For the other designs, the power for the SNP*age interaction parameter decreases using the robust standard error, but only by 5% or less for most error misspecifications, when the MAF was 0.2 or greater. Assuming a t-distribution for the error led to a decrease of about 5–10% power using the robust standard error when the MAF 0.1 or 0.2; this might be due to the substantial reduction in type 1 error. The power also decreases by greater than 5% when the variance is greater at the adiposity rebound and the variance is dependent on a covariate, for values of MAF around 0.1 in our scenarios.

4 Analysis of chromosome-wide body-mass-index data

Given our simulation results, in particular the need for a robust standard error to ensure accurate inference for the SNP*age interaction where the type 1 error is inflated, we wanted to investigate the impact of the dis-

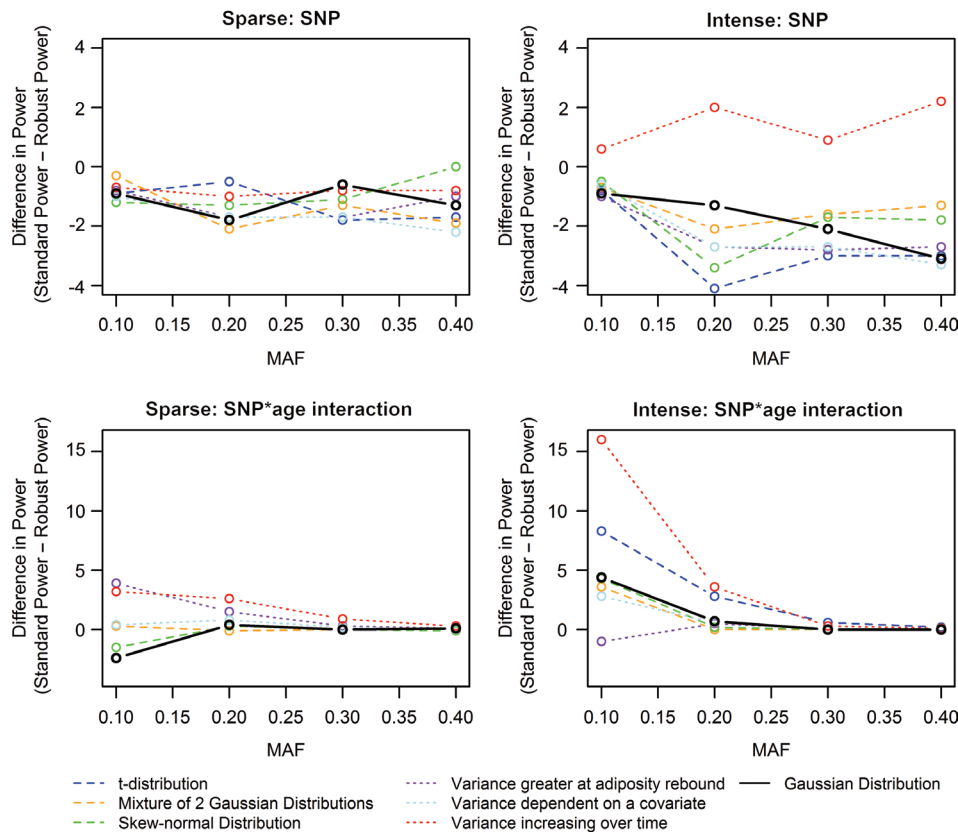


Figure 3 Difference in power based on a normal standard error vs. a robust standard error for the complete designs.

A positive value indicates the power using the normal standard error is greater than the power using the robust standard error. The two plots on the left are for the sparse complete design, while the two plots on the right are from the intense complete design. The solid black line for the Gaussian Distribution is the situation where the model is correctly specified.

tribution assumption problems in a real data application. GWAS analysis of multiple cohorts would be ideal to observe the effect of the different error term misspecifications; however this would require a large amount of computing time and was thus determined to be prohibitive. Instead, we chose to conduct analysis using chromosome 16 in the ALSPAC data as the most replicated gene for BMI to date, the fat mass and obesity gene (*FTO*), is located on this chromosome and we therefore hypothesised that we would detect some significant loci on this chromosome as well as many non-associated SNPs. We used the same LMM model as in equation 2, with the inclusion of an age*sex interaction in the fixed effects for all the age components (i.e., $\beta_9 \text{sex}_1 + \beta_{10} t_{ij} \text{sex}_1 + \beta_{11} t_{ij}^2 \text{sex}_1 + \beta_{10} t_{ij}^3 \text{sex}_1$) to account for the differences in growth between males and females (Warrington et al., 2013). There were 14,875 SNPs genotyped on chromosome 16, all of which had a MAF greater than 1%; GWAS are designed to look at common SNPs, so it is a common strategy to exclude SNPs with MAF <1%. Each SNP was incorporated into the model assuming an additive genetic model.

As expected, SNPs in the *FTO* gene were highly significant for the global tests as well as the SNP main effect and SNP*age interactions. It is common to display GWAS analysis as a QQ plot of the observed $-\log_{10}(P)$ with the expected $-\log_{10}(P)$ under the null distribution. Figure 5 displays a QQ plot from the chromosome 16 analysis in ALSPAC for each of the parameters which displayed inflated levels of type 1 error in the simulation study. As we believe SNPs within the *FTO* region to be true positives, we also display the QQ plots excluding SNPs from this region (Figure 5C and D). Lambda (λ) values are also commonly calculated for GWAS analyses, which is the ratio of the median of the empirically observed distribution of the test statistic to the expected median. The λ quantifies the extent of the excess false positive rate, with values close to 1 indicating no inflation and values deviating from 1 indicating increasing levels of false positives. The lambda values corresponding to each QQ plot were calculated using the `estlambda` function in the GenABEL software (Aulchenko

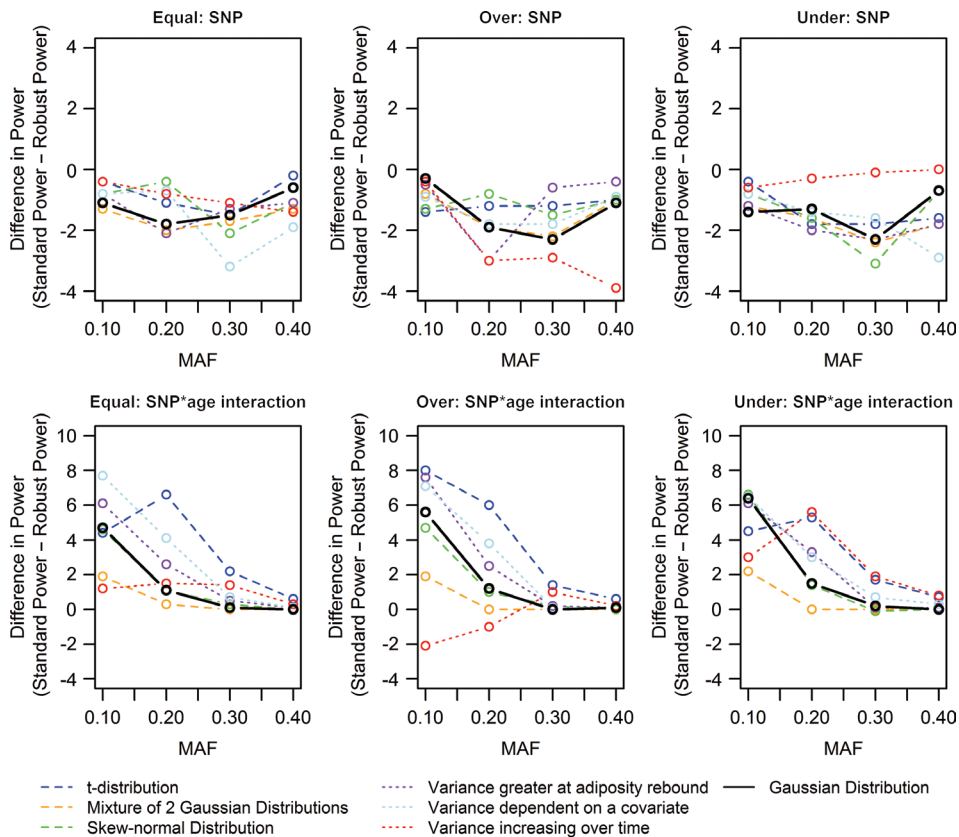


Figure 4 Difference in power based on a normal standard error vs. a robust standard error for the unbalanced designs. A positive value indicates the power using the normal standard error is greater than the power using the robust standard error. Here, “Equal” is the simulations from the equal unbalanced design, “Over” are the simulations from the unbalanced design with less samples around the adiposity rebound and “Under” are the simulations from the unbalanced design with more samples around the adiposity rebound. The solid black line for the Gaussian Distribution is the situation where the model is correctly specified.

et al., 2007); the “median” method was used with one degree of freedom for the fixed effects SNP terms and four degrees of freedom for the Wald test of the overall SNP effect. These QQ plots and lambda statistics clearly show that where the parameters have lambda values greater than one using the classical test, the robust test reduces this to nominal levels. When using the robust tests, we were still able to detect an association with SNPs in the *FTO* gene.

In the chromosome wide analysis, the P-value to declare “suggestive significance” would be 0.000067 (1/14,875). Using this threshold, 57 SNPs would reach suggestive significance for the SNP by age interaction using the classical standard error in comparison to only 16 SNPs using the robust standard error. Six of these 16 SNPs were in the *FTO* gene, four of which would reach the significant threshold.

5 Discussion

In this article, we simulated longitudinal data that mimicked childhood BMI to explore the coverage probability, bias, type 1 error and power for association with a SNP when the linear mixed effects model is misspecified with either a non-Gaussian error distribution or heteroscedastic error. We have shown that the type 1 error for the SNP*age interaction terms in a genetic association study has no inflation if the same function of age is included in both the fixed and random effects. However, type 1 error is inflated, regardless of the model misspecification, if the age function in the fixed and random effects differs. In situations where the model is

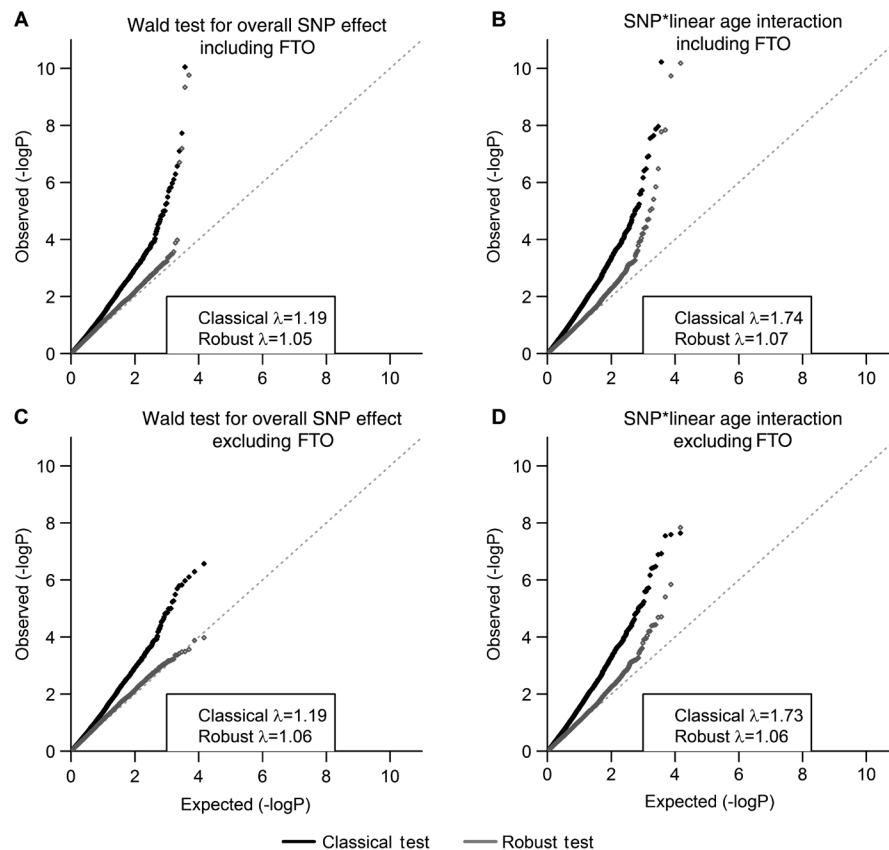


Figure 5 QQ plots of the chromosome 16 analysis in the ALSPAC cohort.

These plots are the observed $-\log_{10}(P)$ against the expected $-\log_{10}(P)$ under the null hypothesis for each SNP on chromosome 16. P-Values deviating from the dotted $x=y$ line indicate significant findings, whether they be false (i.e., inflation in type 1 error) or true.

too complex and will not converge with a high order polynomial function in the random effects, an appropriate way to deflate the type 1 error to nominal levels is to use a robust standard error for the fixed effects parameters. Although robust standard errors have been previously used in a wide range of statistical applications, LMM's are only just beginning to be utilized in GWAS and therefore guidance on their application was warranted. Given that QQ plots in GWAS are an important diagnostic to rule out the possibility of population stratification, it is essential to generate standard errors that perform well under the null hypothesis so that any remaining inflation is not due to the model fitting. Similar to the conclusions by Gurka et al. (2011) and Verbeke and Molenberghs (2000) using other applications, the sandwich estimator is a valid alternative in GWAS when the model assumptions are misspecified, however it is less efficient than using the correct covariance model.

Similar to Jacqmin-Gadda et al. (2007), we have shown that estimates of differences in slope by the number of copies of minor allele are sensitive to heterogeneous error variance particularly when the error variance depends on a covariate or increases over time. The variance of the estimates is underestimated and therefore the confidence interval is too narrow; this is consistent with the inflated type 1 error under these misspecified model assumptions.

Of all the misspecifications investigated, the situation where the error variance increases over time and is not accounted for in the modelling has poor parameter estimates, low power and the most inflation of the type 1 error, particularly for the SNP*age interaction terms. It also appears that by using the robust standard error, the inflation in the type 1 error is reduced to the nominal level in only some of the scenarios. It is therefore imperative that some adjustment is made in the modelling to account for this increasing variance over time. In the ALSPAC BMI data, the variance stays relatively constant until around the age of four years where

it rapidly increases until around 11 years of age where it plateaus again. This is due to the different growth rates between individuals through the adiposity rebound and puberty. Increasing variability over time can be seen with many other phenotypes both in childhood and adulthood; for example lung function in an elderly population can decrease due to the rate at which individuals are diagnosed with diseases such as chronic obstructive pulmonary disease, while other individuals remain healthy. Variance functions for modelling heteroscedasticity in mixed effects models have been studied in detail by Davidian and Giltinan (1995) and can be implemented using the `varFunc` classes in the `nlme` package in R (Pinheiro and Bates, 2000). There are also equivalent functions in alternative statistical packages such as `MLwiN` (Rasbash et al., 2012). The use of these variance functions could be recommended in the context of GWAS, if there is remaining heteroscedasticity in the residuals after appropriately modelling the fixed and random effects; however further studies are needed to assess their properties in this context.

When looking at SNPs with low minor allele frequencies, we have seen that by using the robust standard error we reduce our power by approximately 5%. To counteract this reduction, we can increase the sample size though the use of meta-analysis of multiple cohort studies as is commonly done in GWAS analyses. However, several manuscripts have previously discussed the extended computational time for longitudinal GWAS in comparison to GWAS of cross-sectional phenotypes, so it is recommended that large computing clusters are available to those cohort studies conducting analyses. The longitudinal GWAS of cardiovascular risk factors presented in Smith et al. (2010) took approximately 3 h on 64 processors of a compute cluster for 600,000 tests in 525 individuals. Sikorska et al. (2013) illustrated that the analysis of 2.5 million SNPs using the `LME` function in the `nlme` package of R would take 3500 h for a sample size of 3000 individuals on a desktop computer (Intel(R) Core(TM) 2 Duo CPU, 3.00 GHz). These times are consistent with those in this study; the chromosome 16 analysis of 14,875 SNPs in the 7916 ALSPAC individuals took approximately 125 h on 32 processors of a compute cluster (BlueCrystal Phase 2 cluster with each node having four 2×2.8 GHz core processors and 8 GB of RAM).

It has been suggested that the genome-wide significance threshold be set at 5×10^{-8} (Dudbridge and Gusnanto, 2008; Risch and Merikangas, 1996). In addition, Duggal et al. (2008) established an appropriate *p*-value threshold based on the number of independent SNP tests in a GWAS. If study data is imputed against the HapMap CEU population, they suggest a threshold of $p < 6.09 \times 10^{-6}$ be used to select SNPs with suggestive evidence for follow-up. Many cross-sectional GWAS studies use thresholds around this, generally ranging from $p < 5 \times 10^{-6}$ (Speliotes et al., 2010) to $p < 10^{-5}$ (Thorleifsson et al., 2009), to select SNPs for replication. In longitudinal genetic association studies, particularly those with complex, non-linear trajectories, controlling the type 1 error of the many parameters involving SNP effects, can be quite challenging. This would be the case when using for example smoothed spline functions and those functions could interact with the SNP effects. Providing robust standard errors in this context can be difficult. As an alternative, it may be plausible to use genomic control procedures to reduce a possible inflation in the type 1 error for the parameters involving the SNP effects (Devlin and Roeder, 1999; Dadd et al., 2009). Genomic control is typically used in genetic association studies to account for the potential confounding due to cryptic relatedness, and makes the assumption that the inflation in type 1 error is constant across all markers in the genome; this is plausible in the context of cryptic relatedness as the inflation is due to the kinship coefficients which are unrelated to the individual loci, however in the context of LMM's one would need to show that the inflation was uniform across the genome or genetic region of interest. Benke et al. (2013) suggested using a joint test of all SNP effects, similar to the global Wald test used in the current study, as an optimal way to control the type 1 error and increase power. However, caution needs to be applied when utilizing this method for complex traits, such as BMI trajectories over childhood, and a genome-wide significance threshold should only be used if there is no inflation detected in the type 1 error. Benke et al. (2013) used a trait with a linear decrease over time and low correlation between the intercept and slope parameters; in contrast, in this study we have a complex trajectory over time with high correlation between the intercept and slope parameters, which indicated that the joint test has inflated type 1 error and can only be reduced using a robust estimate in some scenarios.

In summary, based on our simulation results, we strongly suggest fitting the same function of age in the fixed and random effect to avoid inflation of the type 1 error of the SNP*age interaction terms. If this is not

possible due to convergence issues, then we suggest using a robust standard error for the SNP by age interaction terms to reduce the type 1 error inflation in GWAS, regardless of whether the error term of the model correctly follows the model assumptions or not. If no inflation in the type 1 error is detected for a particular parameter of interest, then the classical standard error should be used; for example, for the SNP main effect parameter in this study.

Acknowledgments: We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 092731) and the University of Bristol provide core support for ALSPAC. NM Warrington is funded by an Australian Postgraduate Award from the Australian Government of Innovation, Industry, Science and Research and a Raine Study PhD Top-Up Scholarship. LD Howe is funded by a UK Medical Research Council Population Health Scientist fellowship (G1002375). L Paternoster is funded by a UK Medical Research Council Population Health Scientist fellowship (MR/J012165/1). K Tilling, LD Howe and L Paternoster work in a Unit that receives core funding from the University of Bristol and the UK Medical Research Council (Grant ref: MC_UU_12013/9). The UK Medical Research Council also supports K Tilling's research (G1000726/1).

References

- Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn (2007): "GenABEL: an R library for genome-wide association analysis," *Bioinformatics*, 23, 1294–1296.
- Benke, K. S., Y. Wu, D. M. Fallin, B. Maher and L. J. Palmer (2013) "Strategy to control type I error increases power to identify genetic variation using the full biological trajectory," *Genet. Epidemiol.*, 37, 419–430.
- Boyd, A., J. Golding, J. Macleod, D. A. Lawlor, A. Fraser, J. Henderson, L. Molloy, A. Ness, S. Ring and G. Davey Smith (2013) "Cohort Profile: the children of the 90s—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.*, 42, 111–127.
- Bradfield, J. P., H. R. Taal, N. J. Timpson, A. Scherag, C. Lecoeur, N. M. Warrington, E. Hypponen, C. Holst, B. Valcarcel, E. Thiering, R. M. Salem, F. R. Schumacher, D. L. Cousminer, P. M. Sleiman, J. Zhao, R. I. Berkowitz, K. S. Vimalaswaran, I. Jarick, C. E. Pennell, D. M. Evans, B. St Pourcain, D. J. Berry, D. O. Mook-Kanamori, A. Hofman, F. Rivadeneira, A. G. Uitterlinden, C. M. van Duijn, R. J. van der Valk, J. C. de Jongste, D. S. Postma, D. I. Boomsma, W. J. Gauderman, M. T. Hassanein, C. M. Lindgren, R. Magi, C. A. Boreham, C. E. Neville, L. A. Moreno, P. Elliott, A. Pouta, A. L. Hartikainen, M. Li, O. Raitakari, T. Lehtimäki, J. G. Eriksson, A. Palotie, J. Dallongeville, S. Das, P. Deloukas, G. McMahon, S. M. Ring, J. P. Kemp, J. L. Buxton, A. I. Blakemore, M. Bustamante, M. Guxens, J. N. Hirschhorn, M. W. Gillman, E. Kreiner-Moller, H. Bisgaard, F. D. Gilliland, J. Heinrich, E. Wheeler, I. Barroso, S. O'Rahilly, A. Meirhaeghe, T. I. Sorensen, C. Power, L. J. Palmer, A. Hinney, E. Widen, I. S. Farooqi, M. I. McCarthy, P. Froguel, D. Meyre, J. Hebebrand, M. R. Jarvelin, V. W. Jaddoe, G. D. Smith, H. Hakonarson and S. F. Grant (2012) "A genome-wide association meta-analysis identifies new childhood obesity loci," *Nat. Genet.*, 44, 526–531.
- Cole, T. J., M. C. Bellizzi, K. M. Flegal and W. H. Dietz (2000) "Establishing a standard definition for child overweight and obesity worldwide: international survey," *Br. Med. J.*, 320, 1240–1243.
- Dadd, T., M. E. Weale and C. M. Lewis (2009) "A critical evaluation of genomic control methods for genetic association studies," *Genet. Epidemiol.*, 33, 290–298.
- Davidian, M. and D. M. Giltinan (1995): *Nonlinear models for repeated measurement data*, Monographs on statistics and applied probability; 62. London: Chapman & Hall.
- Devlin, B. and K. Roeder (1999): "Genomic control for association studies," *Biometrics*, 55, 997–1004.
- Dubois, L. and M. Girard (2007): "Accuracy of maternal reports of pre-schoolers' weights and heights as estimates of BMI values," *Int. J. Epidemiol.*, 36, 132–138.
- Dudbridge, F. and A. Gusnanto (2008): "Estimation of significance thresholds for genomewide association scans," *Genet. Epidemiol.*, 32, 227–234.
- Duggal, P., E. M. Gillanders, T. N. Holmes and J. E. Bailey-Wilson (2008): "Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies," *BMC Genomics*, 9, 516.
- Fox, C. S., N. Heard-Costa, L. A. Cupples, J. Dupuis, R. S. Vasan and L. D. Atwood (2007): "Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100 K project," *BMC Med. Genet.*, 8(Suppl 1), S18.
- Fraser, A., C. Macdonald-Wallis, K. Tilling, A. Boyd, J. Golding, G. Davey Smith, J. Henderson, J. Macleod, L. Molloy, A. Ness, S. Ring, S. M. Nelson and D. A. Lawlor (2013): "Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort," *Int. J. Epidemiol.*, 42, 97–110.

- Frayling, T. M., N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, J. R. Perry, K. S. Elliott, H. Lango, N. W. Rayner, B. Shields, L. W. Harries, J. C. Barrett, S. Ellard, C. J. Groves, B. Knight, A. M. Patch, A. R. Ness, S. Ebrahim, D. A. Lawlor, S. M. Ring, Y. Ben-Shlomo, M. R. Jarvelin, U. Sovio, A. J. Bennett, D. Melzer, L. Ferrucci, R. J. Loos, I. Barroso, N. J. Wareham, F. Karpe, K. R. Owen, L. R. Cardon, M. Walker, G. A. Hitman, C. N. Palmer, A. S. Doney, A. D. Morris, G. D. Smith, A. T. Hattersley and M. I. McCarthy (2007): "A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity," *Science*, 316, 889–894.
- Furlotte, N. A., E. Eskin and S. Eyheramendy (2012): "Genome-wide association mapping with longitudinal data," *Genet. Epidemiol.*, 36, 463–471.
- Gurka, M. J., L. J. Edwards and K. E. Muller (2011): "Avoiding bias in mixed model inference for fixed effects," *Stat. Med.*, 30, 2696–2707.
- Haslam, D. W. and W. P. James (2005): "Obesity," *Lancet*, 366, 1197–1209.
- Haworth, C. M., S. Carnell, E. L. Meaburn, O. S. Davis, R. Plomin and J. Wardle (2008): "Increasing heritability of BMI and stronger associations with the FTO gene over childhood," *Obesity (Silver Spring)*, 16, 2663–2668.
- Hindorf, L. A., J. MacArthur, A. Wise, H. A. Junkins, P. N. Hall, A. K. Klemm and T. A. Manolio (2010): A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed: 1 October 2013.
- Howe, L. D., K. Tilling, L. Benfield, J. Logue, N. Sattar, A. R. Ness, G. D. Smith and D. A. Lawlor (2010): "Changes in ponderal index and body mass index across childhood and their associations with fat mass and cardiovascular risk factors at age 15," *PLoS One* 5, e15186.
- Howe, L. D., K. Tilling and D. A. Lawlor (2009): "Accuracy of height and weight data from child health records," *Arch. Dis. Child.*, 94, 950–954.
- Ihaka, R. and R. Gentleman (1996): "R: a language for data analysis and graphics," *J. Comput. Graph. Stat.*, 5, 299–314.
- Jacqmin-Gadda, H., S. Sibillot, C. Proust, J. M. Molina and R. Thiébaud (2007): "Robustness of the linear mixed model to misspecified error distribution," *Comput. Stat. Data An.*, 51, 5142–5154.
- Kerner, B., K. E. North and M. D. Fallin (2009): "Use of longitudinal data in genetic studies in the genome-wide association studies era: summary of Group 14," *Genet. Epidemiol.*, 33(Suppl 1), S93–S98.
- Kindblom, J. M., M. Lorentzon, A. Hellqvist, L. Lonn, J. Brandberg, S. Nilsson, E. Norjavaara and C. Ohlsson (2009): "BMI changes during childhood and adolescence as predictors of amount of adult subcutaneous and visceral adipose tissue in men: the GOOD Study," *Diabetes*, 58, 867–874.
- Koehler, E., E. Brown and S. J. Haneuse (2009): "On the assessment of Monte Carlo error in simulation-based statistical analyses," *Am. Stat.*, 63, 155–162.
- Laird, N. M. and J. H. Ware (1982): "Random-effects models for longitudinal data," *Biometrics*, 38, 963–974.
- Liang, K. Y. and S. L. Zeger (1986): "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22.
- Liu, J. Z., S. E. Medland, M. J. Wright, A. K. Henders, A. C. Heath, P. A. Madden, A. Duncan, G. W. Montgomery, N. G. Martin and A. F. McRae (2010): "Genome-wide association study of height and body mass index in Australian twin families," *Twin. Res. Hum. Genet.*, 13, 179–193.
- Loos, R. J., C. M. Lindgren, S. Li, E. Wheeler, J. H. Zhao, I. Prokopenko, M. Inouye, R. M. Freathy, A. P. Attwood, J. S. Beckmann, S. I. Berndt, K. B. Jacobs, S. J. Chanock, R. B. Hayes, S. Bergmann, A. J. Bennett, S. A. Bingham, M. Bochud, M. Brown, S. Cauchi, J. M. Connell, C. Cooper, G. D. Smith, I. Day, C. Dina, S. De, E. T. Dermizakis, A. S. Doney, K. S. Elliott, P. Elliott, D. M. Evans, I. Sadaf Farooqi, P. Froguel, J. Ghorri, C. J. Groves, R. Gwilliam, D. Hadley, A. S. Hall, A. T. Hattersley, J. Hebebrand, I. M. Heid, C. Lamina, C. Gieger, T. Illig, T. Meitinger, H. E. Wichmann, B. Herrera, A. Hinney, S. E. Hunt, M. R. Jarvelin, T. Johnson, J. D. Jolley, F. Karpe, A. Keniry, K. T. Khaw, R. N. Luben, M. Mangino, J. Marchini, W. L. McArdle, R. McGinnis, D. Meyre, P. B. Munroe, A. D. Morris, A. R. Ness, M. J. Neville, A. C. Nica, K. K. Ong, S. O'Rahilly, K. R. Owen, C. N. Palmer, K. Papadakis, S. Potter, A. Pouta, L. Qi, J. C. Randall, N. W. Rayner, S. M. Ring, M. S. Sandhu, A. Scherag, M. A. Sims, K. Song, N. Soranzo, E. K. Speliotes, H. E. Syddall, S. A. Teichmann, N. J. Timpson, J. H. Tobias, M. Uda, C. I. Vogel, C. Wallace, D. M. Waterworth, M. N. Weedon, C. J. Willer, W. Raithe, X. Yuan, E. Zeggini, J. N. Hirschhorn, D. P. Strachan, W. H. Ouwehand, M. J. Caulfield, N. J. Samani, T. M. Frayling, P. Vollenweider, G. Waeber, V. Mooser, P. Deloukas, M. I. McCarthy, N. J. Wareham, I. Barroso, P. Kraft, S. E. Hankinson, D. J. Hunter, F. B. Hu, H. N. Lyon, B. F. Voight, M. Ridderstrale, L. Groop, P. Scheet, S. Sanna, G. R. Abecasis, G. Albaladejo, R. Nagaraja, D. Schlessinger, A. U. Jackson, J. Tuomilehto, F. S. Collins, M. Boehnke and K. L. Mohlke (2008): "Common variants near MC4R are associated with fat mass, weight and risk of obesity," *Nat. Genet.*, 40, 768–775.
- Maes, H. H., M. C. Neale and L. J. Eaves (1997): "Genetic and environmental factors in relative body weight and human adiposity," *Behav. Genet.*, 27, 325–351.
- McDonald, L. (1975): "Tests for the general linear hypothesis under the multiple design multivariate linear model," *An. Stat.*, 3, 461–466.
- Parsons, T. J., C. Power, S. Logan and C. D. Summerbell (1999): "Childhood predictors of adult obesity: a systematic review," *Int. J. Obes. Relat. Metab. Disord.*, 23(Suppl 8), S1–S107.
- Pinheiro, J. and D. Bates (2000): *Mixed effects models in S and S-Plus*. Springer: New York, NY, USA.
- Rasbash, J., F. Steele, W. J. Browne and H. Goldstein (2012): *A user's guide to MLwiN, v2.26*. centre for multilevel modelling, University of Bristol: UK.
- Risch, N. and K. Merikangas (1996): "The future of genetic studies of complex human diseases," *Science*, 273, 1516–1517.

- Royall, R. M. (1986): “Model robust confidence intervals using maximum likelihood estimators,” *Int. Stat. Rev./Revue Internationale de Statistique*, 54, 221–226.
- Serdula, M. K., D. Ivery, R. J. Coates, D. S. Freedman, D. F. Williamson and T. Byers (1993): “Do obese children become obese adults? A review of the literature,” *Prev. Med.*, 22, 167–177.
- Sikorska, K., F. Rivadeneira, P. J. Groenen, A. Hofman, A. G. Uitterlinden, P. H. Eilers and E. Lesaffre (2013): “Fast linear mixed model computations for genome-wide association studies with longitudinal data,” *Stat. Med.*, 32, 165–180.
- Smith, E. N., W. Chen, M. Kahonen, J. Kettunen, T. Lehtimäki, L. Peltonen, O. T. Raitakari, R. M. Salem, N. J. Schork, M. Shaw, S. R. Srinivasan, E. J. Topol, J. S. Viikari, G. S. Berenson and S. S. Murray (2010): “Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study,” *PLoS Genet.*, 6, e1001094.
- Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. L. Allen, C. M. Lindgren, J. Luan, R. Magi, J. C. Randall, S. Vedantam, T. W. Winkler, L. Qi, T. Workalemahu, I. M. Heid, V. Steinthorsdottir, H. M. Stringham, M. N. Weedon, E. Wheeler, A. R. Wood, T. Ferreira, R. J. Weyant, A. V. Segre, K. Estrada, L. Liang, J. Nemesh, J. H. Park, S. Gustafsson, T. O. Kilpeläinen, J. Yang, N. Bouatia-Naji, T. Esko, M. F. Feitosa, Z. Kutalik, M. Mangino, S. Raychaudhuri, A. Scherag, A. V. Smith, R. Welch, J. H. Zhao, K. K. Aben, D. M. Absher, N. Amin, A. L. Dixon, E. Fisher, N. L. Glazer, M. E. Goddard, N. L. Heard-Costa, V. Hoesel, J. J. Hottenga, A. Johansson, T. Johnson, S. Ketkar, C. Lamina, S. Li, M. F. Moffatt, R. H. Myers, N. Narisu, J. R. Perry, M. J. Peters, M. Preuss, S. Ripatti, F. Rivadeneira, C. Sandholt, L. J. Scott, N. J. Timpson, J. P. Tyrer, S. van Wingerden, R. M. Watanabe, C. C. White, F. Wiklund, C. Barlassina, D. I. Chasman, M. N. Cooper, J. O. Jansson, R. W. Lawrence, N. Pellikka, I. Prokopenko, J. Shi, E. Thiering, H. Alavere, M. T. Alibrandi, P. Almgren, A. M. Arnold, T. Aspelund, L. D. Atwood, B. Balkau, A. J. Balmforth, A. J. Bennett, Y. Ben-Shlomo, R. N. Bergman, S. Bergmann, H. Biebertmann, A. I. Blakemore, T. Boes, L. L. Bonnycastle, S. R. Bornstein, M. J. Brown, T. A. Buchanan, F. Busonero, H. Campbell, F. P. Cappuccino, C. Cavalcanti-Proenca, Y. D. Chen, C. M. Chen, P. S. Chines, R. Clarke, L. Coin, J. Connell, I. N. Day, M. Heijer, J. Duan, S. Ebrahim, P. Elliott, R. Elosua, G. Eiriksdottir, M. R. Erdos, J. G. Eriksson, M. F. Facheris, S. B. Felix, P. Fischer-Posovszky, A. R. Folsom, N. Friedrich, N. B. Freimer, M. Fu, S. Gaget, P. V. Gejman, E. J. Geus, C. Gieger, A. P. Gjesing, A. Goel, P. Goyette, H. Grallert, J. Grassler, D. M. Greenawalt, C. J. Groves, V. Gudnason, C. Guiducci, A. L. Hartikainen, N. Hassanal, A. S. Hall, A. S. Havulinna, C. Hayward, A. C. Heath, C. Hengstenberg, A. A. Hicks, A. Hinney, A. Hofman, G. Homuth, J. Hui, W. Igl, C. Iribarren, B. Isomaa, K. B. Jacobs, I. Jarick, E. Jewell, U. John, T. Jorgensen, P. Jousilahti, A. Jula, M. Kaakinen, E. Kajantie, L. M. Kaplan, S. Kathiresan, J. Kettunen, L. Kinnunen, J. W. Knowles, I. Kolcic, I. R. König, S. Koskinen, P. Kovacs, J. Kuusisto, P. Kraft, K. Kvaloy, J. Laitinen, O. Lantieri, C. Lanzani, L. J. Launer, C. Lecoeur, T. Lehtimäki, G. Lettre, J. Liu, M. L. Lokki, M. Lorentzon, R. N. Luben, B. Ludwig, P. Manunta, D. Marek, M. Marre, N. G. Martin, W. L. McArdle, A. McCarthy, B. McKnight, T. Meitinger, O. Melander, D. Meyre, K. Midthjell, G. W. Montgomery, M. A. Morken, A. P. Morris, R. Mulic, J. S. Ngwa, M. Nelis, M. J. Neville, D. R. Nyholt, C. J. O'Donnell, S. O'Rahilly, K. K. Ong, B. Oostra, G. Pare, A. N. Parker, M. Perola, I. Pichler, K. H. Pietiläinen, C. G. Platou, O. Polasek, A. Pouta, S. Rafelt, O. Raitakari, N. W. Rayner, M. Ridderstrale, W. Rief, A. Ruokonen, N. R. Robertson, P. Rzehak, V. Salomaa, A. R. Sanders, M. S. Sandhu, S. Sanna, J. Saramies, M. J. Savolainen, S. Scherag, S. Schipf, S. Schreiber, H. Schunkert, K. Silander, J. Sinisalo, D. S. Siscovick, J. H. Smit, N. Soranzo, U. Sovio, J. Stephens, I. Surakka, A. J. Swift, M. L. Tammesoo, J. C. Tardif, M. Teder-Laving, T. M. Teslovich, J. R. Thompson, B. Thomson, A. Tonjes, T. Tuomi, J. B. van Meurs, G. J. van Ommen, V. Vatin, J. Viikari, S. Visvikis-Siest, V. Vitart, C. I. Vogel, B. F. Voight, L. L. Waite, H. Wallaschofski, G. B. Walters, E. Widen, S. Wiegand, S. H. Wild, G. Willemsen, D. R. Witte, J. C. Witterman, J. Xu, Q. Zhang, L. Zgaga, A. Ziegler, P. Zitting, J. P. Beilby, I. S. Farooqi, J. Hebebrand, H. V. Huikuri, A. L. James, M. Kahonen, D. F. Levinson, F. Macciardi, M. S. Nieminen, C. Ohlsson, L. J. Palmer, P. M. Ridker, M. Stumvoll, J. S. Beckmann, H. Boeing, E. Boerwinkle, D. I. Boomsma, M. J. Caulfield, S. J. Chanock, F. S. Collins, L. A. Cupples, G. D. Smith, J. Erdmann, P. Froguel, H. Gronberg, U. Gyllenstein, P. Hall, T. Hansen, T. B. Harris, A. T. Hattersley, R. B. Hayes, J. Heinrich, F. B. Hu, K. Hveem, T. Illig, M. R. Jarvelin, J. Kaprio, F. Karpe, K. T. Khaw, L. A. Kiemeny, H. Krude, M. Laakso, D. A. Lawlor, A. Metspalu, P. B. Munroe, W. H. Ouwehand, O. Pedersen, B. W. Penninx, A. Peters, P. P. Pramstaller, T. Quertermous, T. Reinehr, A. Rissanen, I. Rudan, N. J. Samani, P. E. Schwarz, A. R. Shuldiner, T. D. Spector, J. Tuomilehto, M. Uda, A. Uitterlinden, T. T. Valle, M. Wabitsch, G. Waeber, N. J. Wareham, H. Watkins, J. F. Wilson, A. F. Wright, M. C. Zillikens, N. Chatterjee, S. A. McCarroll, S. Purcell, E. E. Schadt, P. M. Visscher, T. L. Assimes, I. B. Borecki, P. Deloukas, C. S. Fox, L. C. Groop, T. Haritunians, D. J. Hunter, R. C. Kaplan, K. L. Mohlke, J. R. O'Connell, L. Peltonen, D. Schlessinger, D. P. Strachan, C. M. van Duijn, H. E. Wichmann, T. M. Frayling, U. Thorsteinsdottir, G. R. Abecasis, I. Barroso, M. Boehnke, K. Stefansson, K. E. North, M. I. McCarthy, J. N. Hirschhorn, E. Ingelsson and R. J. Loos (2010): “Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index,” *Nat. Genet.*, 42, 937–948.
- Taylor, J. M. G., W. G. Cumberland and J. P. Sy (1994): “A Stochastic Model for Analysis of Longitudinal AIDS Data,” *J. Am. Stat. Assoc.*, 89, 727–736.
- Taylor, J. M. and N. Law (1998): “Does the covariance structure matter in longitudinal modelling for the prediction of future CD4 counts?” *Stat. Med.*, 17, 2381–2394.
- Thorleifsson, G., G. B. Walters, D. F. Gudbjartsson, V. Steinthorsdottir, P. Sulem, A. Helgadóttir, U. Styrkarsdóttir, S. Gretarsdóttir, S. Thordarson, I. Jonsdóttir, T. Jonsdóttir, E. J. Olafsdóttir, G. H. Olafsdóttir, T. Jonsson, F. Jonsson, K. Borch-Johnsen, T. Hansen, G. Andersen, T. Jorgensen, T. Lauritzen, K. K. Aben, A. L. Verbeek, N. Roeleveld, E. Kampman, L. R. Yanek, L. C. Becker, L. Tryggvadóttir, T. Rafnar, D. M. Becker, J. Gulcher, L. A. Kiemeny, O. Pedersen, A. Kong, U. Thorsteinsdóttir and K. Stefansson (2009): “Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity,” *Nat. Genet.*, 41, 18–24.

- Verbeke, G. and E. Lesaffre (1997): "The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data," *Comput. Stat. Data An.*, 23, 541–556.
- Verbeke, G. and G. Molenberghs (2000): *Linear mixed models for longitudinal data: Springer Series in Statistics*, New York: Springer-Verlag.
- Wardle, J., S. Carnell, C. M. Haworth and R. Plomin (2008): "Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment," *Am. J. Clin. Nutr.*, 87, 398–404.
- Warrington, N. M., Y. Y. Wu, C. E. Pennell, J. A. Marsh, L. J. Beilin, L. J. Palmer, S. J. Lye and L. Briollais (2013): "Modelling BMI trajectories in children for genetic association studies," *PLoS One*, 8, e53897.
- White, I. (2010): "simsum: Analysis of simulation studies including Monte Carlo error," *The Stata Journal*, 10, 369–385.
- WHO. (2000): "Obesity: preventing and managing the global epidemic. Report of a WHO Consultation. WHO Technical Report Series 894. Geneva: World Health Organization, 2000.
- Willer, C. J., E. K. Speliotes, R. J. Loos, S. Li, C. M. Lindgren, I. M. Heid, S. I. Berndt, A. L. Elliott, A. U. Jackson, C. Lamina, G. Lettre, N. Lim, H. N. Lyon, S. A. McCarrroll, K. Papadakis, L. Qi, J. C. Randall, R. M. Roccasecca, S. Sanna, P. Scheet, M. N. Weedon, E. Wheeler, J. H. Zhao, L. C. Jacobs, I. Prokopenko, N. Soranzo, T. Tanaka, N. J. Timpson, P. Almgren, A. Bennett, R. N. Bergman, S. A. Bingham, L. L. Bonnycastle, M. Brown, N. P. Burt, P. Chines, L. Coin, F. S. Collins, J. M. Connell, C. Cooper, G. D. Smith, E. M. Dennison, P. Deodhar, P. Elliott, M. R. Erdos, K. Estrada, D. M. Evans, L. Gianniny, C. Gieger, C. J. Gillson, C. Guiducci, R. Hackett, D. Hadley, A. S. Hall, A. S. Havulinna, J. Hebebrand, A. Hofman, B. Isomaa, K. B. Jacobs, T. Johnson, P. Jousilahti, Z. Jovanovic, K. T. Khaw, P. Kraft, M. Kuokkanen, J. Kuusisto, J. Laitinen, E. G. Lakatta, J. Luan, R. N. Luben, M. Mangino, W. L. McArdle, T. Meitinger, A. Mulas, P. B. Munroe, N. Narisu, A. R. Ness, K. Northstone, S. O'Rahilly, C. Purmann, M. G. Rees, M. Ridderstrale, S. M. Ring, F. Rivadeneira, A. Ruokonen, M. S. Sandhu, J. Saramies, L. J. Scott, A. Scuteri, K. Silander, M. A. Sims, K. Song, J. Stephens, S. Stevens, H. M. Stringham, Y. C. Tung, T. T. Valle, C. M. Van Duijn, K. S. Vimalaswaran, P. Vollenweider, G. Waeber, C. Wallace, R. M. Watanabe, D. M. Waterworth, N. Watkins, J. C. Witteman, E. Zeggini, G. Zhai, M. C. Zillikens, D. Altshuler, M. J. Caulfield, S. J. Chanock, I. S. Farooqi, L. Ferrucci, J. M. Guralnik, A. T. Hattersley, F. B. Hu, M. R. Jarvelin, M. Laakso, V. Mooser, K. K. Ong, W. H. Ouwehand, V. Salomaa, N. J. Samani, T. D. Spector, T. Tuomi, J. Tuomilehto, M. Uda, A. G. Uitterlinden, N. J. Wareham, P. Deloukas, T. M. Frayling, L. C. Groop, R. B. Hayes, D. J. Hunter, K. L. Mohlke, L. Peltonen, D. Schlessinger, D. P. Strachan, H. E. Wichmann, M. I. McCarthy, M. Boehnke, I. Barroso, G. R. Abecasis and J. N. Hirschhorn (2009): "Six new loci associated with body mass index highlight a neuronal influence on body weight regulation," *Nat. Genet.*, 41, 25–34.
- World Health Organization. Obesity and Overweight Fact Sheet (No 311), May 2012 2012 [cited 4 September 2012. Available from <http://www.who.int/mediacentre/factsheets/fs311/en/index.html>.
- Zhang, D. and M. Davidian (2001): "Linear mixed models with flexible distributions of random effects for longitudinal data," *Biometrics*, 57, 795–802.

Supplemental Material: The online version of this article (DOI: 10.1515/sagmb-2013-0066) offers supplementary material, available to authorized users.