



Multilevel Twin Models: Geographical Region as a Third Level Variable

Z. Tamimy¹ · S. T. Kevenaar¹ · J. J. Hottenga^{1,2} · M. D. Hunter³ · E. L. de Zeeuw¹ · M. C. Neale⁴ · C. E. M. van Beijsterveldt¹ · C. V. Dolan^{1,2} · Elsje van Bergen¹ · D. I. Boomsma^{1,2}

Received: 6 August 2020 / Accepted: 21 January 2021 / Published online: 27 February 2021
© The Author(s) 2021

Abstract

The classical twin model can be reparametrized as an equivalent multilevel model. The multilevel parameterization has underexplored advantages, such as the possibility to include higher-level clustering variables in which lower levels are nested. When this higher-level clustering is not modeled, its variance is captured by the common environmental variance component. In this paper we illustrate the application of a 3-level multilevel model to twin data by analyzing the regional clustering of 7-year-old children's height in the Netherlands. Our findings show that 1.8%, of the phenotypic variance in children's height is attributable to regional clustering, which is 7% of the variance explained by between-family or common environmental components. Since regional clustering may represent ancestry, we also investigate the effect of region after correcting for genetic principal components, in a subsample of participants with genome-wide SNP data. After correction, region no longer explained variation in height. Our results suggest that the phenotypic variance explained by region might represent ancestry effects on height.

Keywords Multilevel model · Classical twin design · OpenMx · Region · Ancestry · Height

Introduction

The classical twin model (CTM) is often approached from a structural equation modeling (SEM) framework (Bentler and Stein 1992; Boomsma and Molenaar 1986; Heath et al. 1989; Neale and Cardon 1992; Rijdsdijk and Sham 2002). In this framework, it is a one-level model with family as level

one sampling unit. The analysis of twin data can, however, also be approached from a multilevel model (MLM) perspective. MLMs were developed specifically for the analysis of clustered data (Goldstein 2011; Laird and Ware 1982; Longford 1993; Paterson and Goldstein 1991). Classical examples are children (level 1 units), who are clustered in classes (level 2) within schools (level 3; Sellström and Bremberg 2006). Other examples are fMRI measures (level 1) that are clustered in individuals (level 2), who are clustered in scanner type (level 3; Chen et al. 2012), or biomarker data (level 1) that are clustered in measurement batches (level 2; Scharpf et al. 2011). The classical twin design is based on data that also have natural clustering, namely, twins are clustered within pairs. For this reason, the MLM framework can accommodate the CTM (Guo and Wang 2002; McArdle and Prescott 2005; Rabe-Hesketh et al. 2008; Van den Oord 2001). Hunter (2020) provides a detailed account of the CTM in the MLM framework with example code and several extensions. While the MLM specification of the CTM is equivalent to the SEM approach, it also has some interesting, yet underexplored, advantages. In this paper we aim to elaborate on these advantages, and to provide an empirical illustration of a multilevel twin model, where we study the clustering of children's height in geographical

Edited by Elizabeth Prom-Wormley.

Z. Tamimy and S.T. Kevenaar have contributed to the work equally.

✉ Z. Tamimy
z.tamimy@vu.nl

- ¹ Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit, Van der Boechorststraat 7-9, 1081 BT Amsterdam, The Netherlands
- ² Amsterdam Public Health (APH) and Amsterdam Reproduction and Development Research Institutes, Amsterdam, The Netherlands
- ³ School of Psychology, Georgia Institute of Technology, 654 Cherry Street NW, Atlanta, GA 30332-0170, USA
- ⁴ Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 1-156, P.O. Box 980126, Richmond, VA 23298-0126, USA

regions in the Netherlands, and consider the role therein of genetic ancestry.

In the SEM approach to the CTM, the covariance structure of twin-pairs is modelled to decompose phenotypic variance into multiple components that represent genetic and non-genetic influences. Given the biometrical underpinning of the twin model (Eaves et al. 1978; Falconer and MacKay 1996; Fisher 1918), the phenotypic variance can be decomposed into additive genetic variance (A), non-additive or dominance genetic variance (D), common environmental variance (C), and unique environmental variance (E) components. Variance decomposition is based on the premise that monozygotic (MZ) twins share 100% of their DNA and dizygotic (DZ) twins share on average 50% of their segregating genes. Hence, additive and non-additive genetic variance is fully shared by MZ twins, whereas additive and non-additive variance components are shared for 50% and 25% by DZ twins. In the CTM, all influences that are not captured by segregating genetic variants are labeled as “environment”. These influences can be categorized as common environment (i.e., shared by twins from the same family) or unique or unshared environment (i.e., creating variation among members from the same family). These are also referred to as between and within family environmental influences. The full ACDE model is not identified when analyzing one phenotype per twin, and only three of the four components can be simultaneously estimated. In this SEM approach to modeling twin data, the variance decomposition is based on the bivariate data observed in twin pairs (i.e., one phenotype for twin 1, and one for twin 2, which are both level 1 units).

In the MLM framework the phenotypic variance can be decomposed into a within-pair (level 1) and a between-pair (or family; level 2) components. This requires reparameterization of the model into level 1 and level 2 variance components. Because the E component captures variance that is not shared by twins, this component is an individual level 1 variance component. The C component is by definition shared by twins, regardless of zygosity, and is a family level 2 variance component. The A component, however, is more complicated, as it is a level 2 component in MZ twin pairs, but both a level 1 and a level 2 component in DZ twin pairs. To account for this, the A-component is divided into two orthogonal components, unique additive (A_U) and common additive (A_C). Here, A_U is a first-level component representing the A variance at the individual level (within pairs or within families), while A_C is a second-level component (between pairs or between families), representing the A variance at the twin-pair level. These definitions are consistent with the classical notations in which A_C refers to within family genetic variance known as A_1 (Boomsma and Molenaar 1986; Martin and Eaves 1977), or the average breeding value variance (Barton et al. 2017), while A_U refers to the between family genetic variance known as A_2

(Boomsma and Molenaar 1986; Martin and Eaves 1977), or the segregating genetic variance (Barton et al. 2017). In MZs, the A_U variance component is 0, since all the variance explained by A is shared by both twins from a pair. For DZ twins, the variance of both A_C and A_U are constrained to equal 0.5, since on average 50% of the A variance is shared by the individuals and 50% of the A variance is unique for the individual.

An important, yet underexplored, advantage of the MLM approach, is the possibility to include higher-level variables in which lower-levels are nested. By including these higher-level variables, we can identify variance components which are attributable to higher-level clustering. Such clusters may be a consequence of data acquisition or design, e.g., clustering of biomarker data that are measured in batches, or clustering of brain imaging data by fMRI scanner type. They may also occur naturally, for example, families in regions, neighborhoods or schools. If the higher-level variable is not included in the variance decomposition models, the variance that it explains will be captured as part of the C-component, since both twins, regardless of zygosity, share the higher-level variable (i.e., the twin pair is nested in the higher-level variable).

Within the SEM framework, higher-level variables can be included in the model as a fixed effect on the individual level (i.e., covariate) by means of (linear) regression. For nominal covariates (i.e., factors in the ANOVA sense), this approach requires the variable to be dummy coded, which may be impractical, for example when the number of assays for a biomarker or the number of schools that twins are enrolled in is large. In the MLM framework, however, the higher-level variable is treated as a random rather than a fixed effect, and this reduces the number of parameters to one single variance component. That is, given a factor with L categories, the fixed effects approach requires L-1 additional parameters, whereas the random effects approach requires one additional parameter (a variance component). In addition, the MLM approach is more suitable than the SEM framework in dealing with unequal group sizes (Gelman 2005). Finally, an MLM approach allows us to evaluate the contribution of the higher-level component to the C-component, as estimated in the standard twin model. This can be achieved by comparing the C-component estimate of the two-level model (i.e., the standard twin model) to the estimate of the three-level model.

In this paper, we illustrate the use of multilevel twin models by investigating the regional clustering of children’s height with twin data from the Netherlands Twin Register (Boomsma et al. 1992; Ligthart et al. 2019). Height serves as an indicator of the general development of a country, and is known to decrease in times of scarcity and increase in times of prosperity (Baten and Blum 2014; Baten and Komlos 1998). Also, children’s height is an indicator of overall

development, where height is associated with cognitive development and school achievement (Karp et al. 1992; Spears 2012). In 7-year-old children, resemblance between family members for height is explained by additive genetic (approximately 60%) and common environmental (approximately 20%) factors (Jelenkovic et al. 2016; Silventoinen et al. 2004, 2007).

In the Netherlands, the association between height and geographical region is well established (Abdellaoui et al. 2013), which makes this a clustering variable of interest. Inhabitants of different geographical region may display genetic and environmental differences. Location is associated with genetic differences (e.g. Abdellaoui et al. 2019) and differences in social and cultural traditions, diet, socioeconomic status, and living circumstances (e.g., rural vs urban, e.g. Colodro-Conde et al. 2018). By analyzing height and geographical region data in a three-level MLM, we can determine whether variance in children's height is associated with geographic region, and estimate the proportion of the common environmental or between-family variance that can be explained by these regional effects.

In a subsample of 7-year-old participants, we investigated the extent to which regional clustering may be due to genetic ancestry by including the first three genetic principal components (PCs; Hotelling 1933). The genetic PCs are obtained through principal component analysis of the covariance matrix of the genotype Single Nucleotide Polymorphism (SNP) data (Reich et al. 2008). In the Netherlands, the first genetic PC is associated with a north–south height gradient (Abdellaoui et al. 2013; Boomsma et al. 2014). This gradient is likely a result of social, geographical and historical divisions between the north and the south. Southern regions were conquered by the Roman empire, adopted Catholicism, and were geographically separated from the northern regions by five large rivers in the Netherlands (Schalekamp 2009). This first Dutch PC also shows a strong correlation with the European PC that differentiates northern from southern European populations (1000 Genomes PC4; Abdellaoui et al. 2013). The second PC is associated with the east–west division of the Netherlands. This PC may reflect differences between rural and urban environments, since the east of the Netherlands is characterized by less populous and rural areas, while the west includes the largest concentration of urban areas in the Netherlands. Alternatively, it could also be a result of geographical separation by the IJssel river or the Veluwe hillridge. The third PC is associated with the more central regions of the country (Abdellaoui et al. 2013). By adding the PCs to our models, we assessed the role of genetic ancestry of individuals between regions.

In this paper, we first considered regional clustering of children's height in a large data set of MZ and DZ twins ($N = 7436$). Secondly, we considered the model within a subgroup of children who were genotyped on genome-wide

SNP arrays ($N = 1375$). Subsequently, we determined whether the region effects represent genetic ancestry. And finally, we analyzed the relationship between the three PCs and height in 7-year-old children, and included the genetic PCs that show an association as an individual level (level 1) covariate in the model.

Methods

Participants and procedure

The data were obtained from the Netherlands Twin Register (NTR), which has collected data on multiple-births and their family members since 1987 (Ligthart et al. 2019). In the longitudinal NTR surveys of phenotypes in children, parents were asked to complete questionnaires on their children's health, growth, and behavior with intervals of approximately two years.

For the present study, we included data on 6- and 7-year-old twin children (range 6 years and 0 months to 7 years and 11 months). The sample included 7346 twin children (50.3% girls) in 3724 families. The twins were 7.4 (SD 0.3) years old on average, when their mothers reported their height. Of these children, 1375 (18.7% of total) were genotyped. Genotyping largely took place independent of phenotype criteria. The 1375 genotyped individuals were from 714 families, 52.4% of this subsample were girls and the average age was 7.4 (SD 0.3).

We included data from 2002 onwards, as that was when active collection of postal code data began. In approximately 1% of the questionnaires that were sent out after 2002, postal code was missing and approximately 20% of the parents did not report their children's height at age 7. We only included participants with both height and postal code information at age 7 in our initial selection. Next, children with severe handicaps were excluded, as were multiple twin pairs per family, twins born before 34 weeks of gestation, and twins outside the 6–8 age range. A flowchart outlining the sample size after every step of exclusion is displayed in Fig. 1. Zygosity was determined by DNA polymorphisms or by a parent-reported zygosity questionnaire on twin similarity. The zygosity determination by questionnaire has an accuracy of over 95% (Ligthart et al. 2019). Table 1 displays the descriptive statistics of the phenotypic data by zygosity for the total and for the genotyped sample.

Measures

Height

Mothers reported child height in centimeters and the date of measurement. Estourgie-van Burk et al. (2006) demonstrated

Fig. 1 Flowchart containing sample size for the total sample (upper row) and the genotyped subsample (lower row) after every step of exclusion

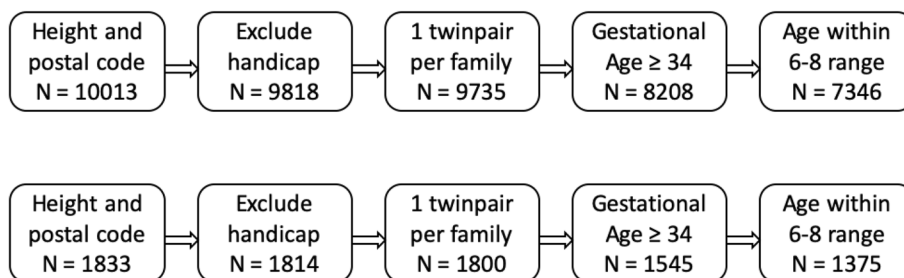


Table 1 The number of twins, the mean, standard deviation and the twin correlation per zygosity group for the total sample and the genotyped subsample

	MZm	DZm	MZf	DZf	DZmf	DZfm
Total sample						
N (individuals)	1283	1228	1338	1208	1163	1126
Mean height	128.3	128.4	127.3	127.8	128.6	128.4
SD height	6.1	5.8	5.8	5.8	5.7	6.2
Twin correlation	0.95	0.61	0.94	0.63	0.58	0.68
Genotyped subsample						
N (individuals)	350	167	251	221	136	150
Mean	128.5	129.1	127.7	127.4	128.2	128.0
SD	6.0	5.8	5.7	6.1	5.6	5.7
Twin-correlation	0.97	0.71	0.95	0.68	0.57	0.69

that the correlation between maternal report and height measured in the laboratory was 0.96 in 5-year-old children in NTR. Mothers reported the age of their children at the moment of completing the survey and the date of the height measurement. In 5% of the children, the date at the time of height measurement was not available. Therefore, in this 5%, we took the age at the time of questionnaire completion. The correlation between age at questionnaire completion and age at height measurement is 0.95, and the mean difference in age is 0.01 years.

Region

At the time of reporting height, parents also reported the four digits of the postal code of their current address. In the Netherlands, postal codes map to geographical locations. The postal code consists of four digits and two letters, where the first two digits map to region and the second two digits and letters map to city, neighborhood within the city, and street. In our analyses, region is specified by the first two digits of the postal code, resulting in 90 regions which are displayed in Fig. 2. They cover on average 462 km² and have a mean population of around 192,000 (total area of the country is 41,543 km², including ~19% water bodies). Most regions encompass several municipalities. In the total sample, the number of children per postal code unit ranged from 10 to 194 ($M = 81.6$, $SD = 38.4$). In the genotyped sample, the

number of children per postal code unit ranged from 1 to 43 ($M = 15.6$, $SD = 8.6$).

Principal components

Genotype data in 1375 individuals were collected by the following genotype platforms: Affymetrix 6, Axiom and Perlegen, Illumina 1 M, 660 and GSA-NTR. The SNP data obtained on the 6 platforms were pruned in Plink to be independent, with additional filters to ensure Minor Allele Frequency (MAF) > 0.01, Hardy–Weinberg Equilibrium (HWE) $p > 0.0001$ and call rate over 95%. Subsequently, long range Linkage Disequilibrium (LD) regions were excluded as described in Abdellaoui et al. (2013), because elevated levels of LD result in overrepresentation of these loci in the PCs, disguising genome-wide patterns that reflect ancestry. For each platform, the NTR data were merged with the data of the individuals from the 1000 Genomes reference panel for the same SNPs, and Principal Components were calculated using SMARTPCA (Prince et al. 2006), where the 1000 genomes populations were projected onto the NTR participants (Privé et al. 2020). Population outliers were identified using pairwise PC plots. People who were identified as outliers from the central population on the basis of visual inspection of these pairwise PC plots, were excluded, rendering the final clustering homogeneous. The NTR platform genotype data of this cluster were aligned to the GoNL reference panel V4 (The Genome of the Netherlands Consortium

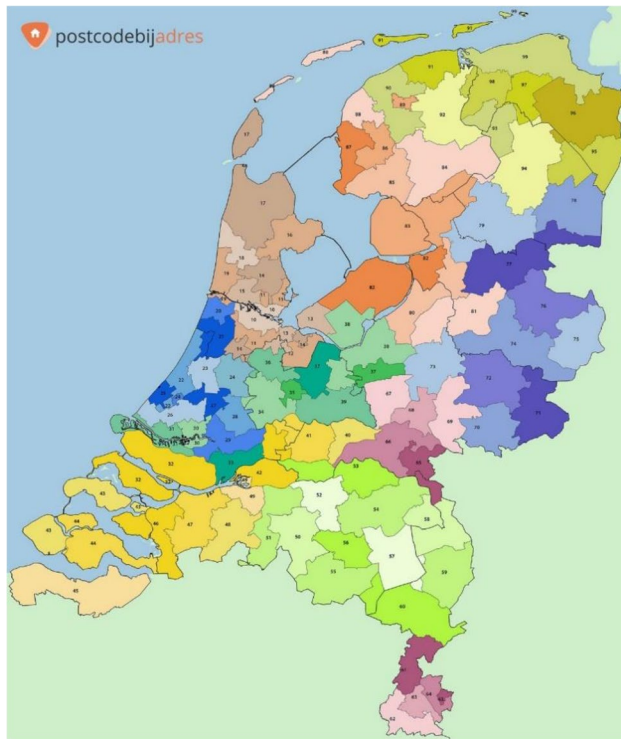


Fig. 2 Map of the 90 regions in the Netherlands based on first two digits of the postal code *Note*. This figure is reprinted from ‘Postcodekaart van Nederland’ by postcodebijadres (2020), retrieved July 29, 2020, from <https://postcodebijadres.nl/postcodes-nederland>

2014), merged into a single dataset, and then imputed in MaCH-Admix (Liu et al. 2013). From the imputed data, SNPs were selected that satisfied $R^2 \geq 0.90$, and that were genotyped on at least one platform. These SNPs were subsequently filtered on $MAF < 0.025$, $HWE p < 0.0001$, call rate $\geq 98\%$, and the absence of Mendelian errors. Again, the long-range LD regions were removed from these SNP data. With this selection of SNPs, 20 new PCs were calculated with SMARTPCA (Prince et al. 2006), to indicate the residual Dutch genetic stratification.

Models

The classical twin model

In the classical twin model, the phenotypic variance can be decomposed into three components: Additive genetic (A), Common environmental (C) and unique Environment (E) component, which includes measurement error. As in most earlier publications, we will not consider genetic dominance variance for height (but see Joshi et al. 2015).

Assuming A, C, and E are mutually independent, we have the following decomposition of phenotypic variance:

$$\text{var}(y) = \sigma_A^2 + \sigma_C^2 + \sigma_E^2$$

The variance component model can be written as a path model in which A, C and E are standardized to have unit variance (see Fig. 3):

$$y_1 = \mu + a*A_1 + c*C_1 + e*E_1,$$

$$y_2 = \mu + a*A_2 + c*C_2 + e*E_2,$$

where y_1 represents the phenotype of first twin and y_2 of the second twin in a twin pair. A, C and E represent individual factor scores for twin 1 and twin 2, and a, c, e represent population specific factor loadings or path coefficients.

If A, C and E have unit variance, the variance decomposition is:

$$\text{var}(y) = a^2 + c^2 + e^2.$$

In terms of the path coefficient model, the covariance between the twins equals $\sigma_{mz} = a^2 + c^2$, in MZ twins, and $\sigma_{dz} = \frac{1}{2}a^2 + c^2$ in DZ twins.

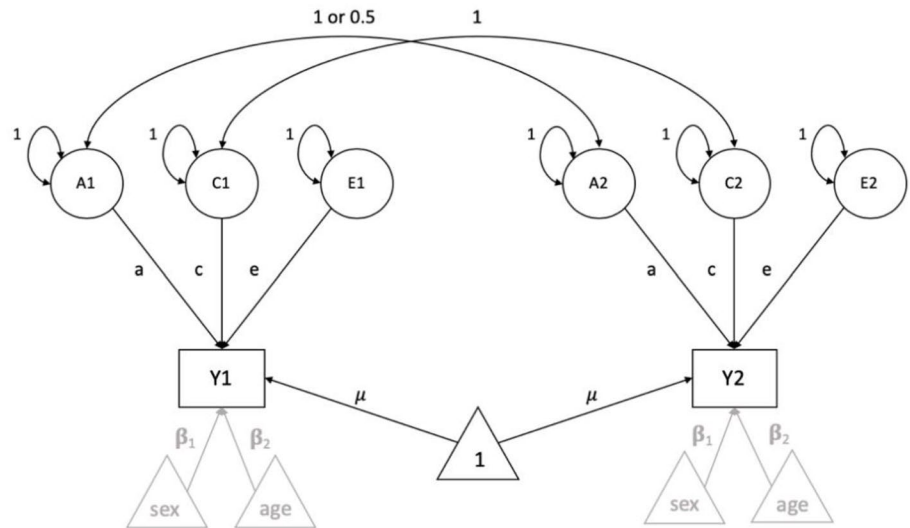
Multilevel twin model

When specifying a CTM as an MLM, the variance components of the CTM are parametrized as within and between family components. The additive genetic variance is separated into two parts: a part that is shared by the members of a twin pair on the second level, A_C , and a part that is unique to each individual on the first level, A_U . The path coefficients associated with the A_C and A_U are equal. The variance of the common genetic factor (r) and the unique genetic factor ($1-r$) depend on the zygosity of the twin pair: for MZ $r = 1.00$, while for DZ $r = 0.50$. The common environmental factor, representing between family influences, is a level two component. Unique environmental factors E represent within family, level one, influences. The means (intercepts) μ are specified on the first level and are assumed to be equal for first- and second-born twins and zygosity. The ACE model in multilevel parametrization is illustrated in Fig. 4. Here, we included age at the individual level, because it represents the age at reported height measure and thus could differ between twins.

Multilevel twin model with third level clustering variable and individual level covariates

Other clustering variables can be added to this model, as displayed in Fig. 5. A higher-order clustering variable can be added to the third level of this model in two steps. On the third level, the higher-order clustering variable is added

Fig. 3 The Classical Twin Model including three latent factors per person, representing Additive genetic, Common and unique Environmental influences. Two additional covariates, age and sex, are presented in a schematic way in grey



with a variance of 1 and a path loading of 1 to a latent variable on the second level, which has a variance of 0 and a freely estimated path loading from Region (reg) to the observed phenotype. Although the Region latent variable could directly affect the child-level phenotype and does not need to pass through the family level, we draw it here to indicate the nesting that region-level effects pass through the family level before impacting the child level. The same 3-level model which also includes PC1 as a fixed covariate is displayed in Fig. 5.

Analyses

All analyses were performed in R (R Core Team 2020) with the package OpenMx (Boker et al. 2011; Neale et al. 2016; Pritikin et al. 2017). Age at measurement was converted to z-scores. Due to scaling, the variance of PC scores is extremely low compared to the variance of the other variables in the model. Therefore, we multiplied these scores by 1000 to avoid ill-conditioning in the parameter covariance matrix, since ill-conditioning can cause optimization problems. First, in the full sample, a variance decomposition of the variance in height was obtained in the regular genetic covariance structure modeling. We included the z-scores of age at measurement and sex as covariates. Then, we repeated the analysis in the multilevel model to illustrate the equivalence of the two approaches. Following this, we added region as a third level in the multilevel parametrization. We repeated these steps in the genotyped group to investigate the representativeness of this subsample. Finally, in the genotyped subset, we added the PC scores as individual level covariates in the 3-level model.

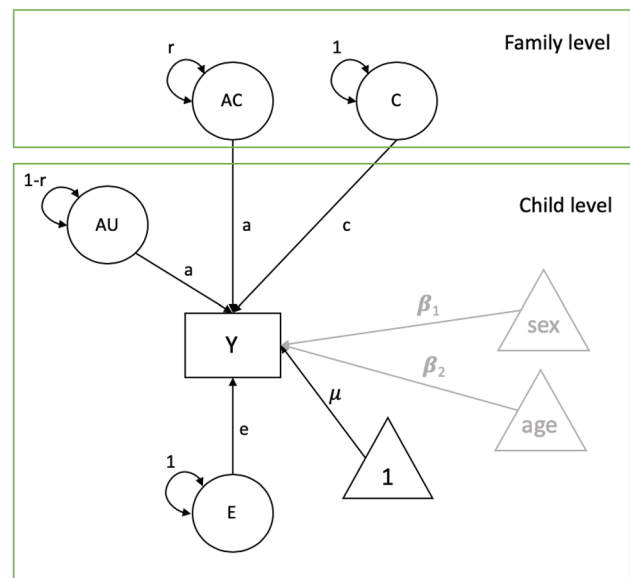
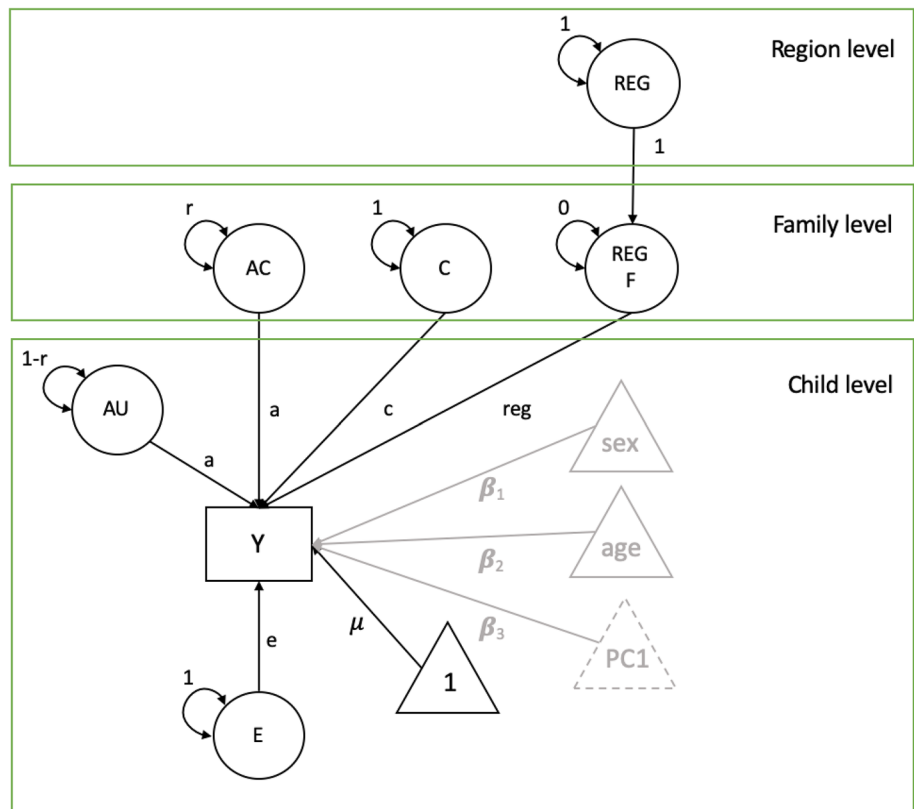


Fig. 4 Multilevel parametrization of the ACE model, where Y represents the phenotype, latent variables A_U represent the unique additive genetic influences and E unique environment. μ is the intercept, sex and age are covariates, presented in a schematic way. On the family level, C is common environment and A_C common genetic influences. The path coefficients, a, c, e, β_1 and β_2 represent regression coefficients. The r parameter represents variance (1 for MZ twins and 0.5 for DZ twins)

We tested the contribution of region to the variance of height by comparing the difference of fit in the 3-level model and the 2-level model without region with the log-likelihood ratio test. Under certain regularity conditions (Steiger et al. 1985), the difference in fit between these models is distributed as Chi squared with one degree of freedom. For all analysis we employed an alpha level of 0.01.

Fig. 5 Multilevel parametrization of the ACE model with Region as a third level, which loads on the region variable REG F on the family level, on which the observed variable Y is regressed with its coefficient estimating the effect of region. Sex and age are covariates. We compared the model where we included PC1 as a covariate with the model where we did not include PC1 as a covariate (we tested PC1, PC2, and PC3, but depict only PC1 to avoid clutter, and because our final model included only PC1). Note that the dummy REG F latent variable serves as a placeholder to stress the nesting of families in regions, but technically, it is not needed



Results

The plot of the average height by region revealed a north–south trend, with the children in the northern regions being taller than those in the southern regions of the Netherlands (of the 12 provinces in the Netherlands, the northern province Drenthe had the highest mean height ($M = 129.40$) and the southern province Noord-Brabant had the lowest mean height ($M = 127.01$)). Figure 6 displays the mean height of 7-year-olds per region. In the genotyped group, height correlated with PC1 (i.e., the PC showing a north–south gradient) ($r = 0.16$), but not with other PCs ($r = 0.01$ for PC2, $r = -0.01$ for PC3). Therefore, we incorporated PC1 into subsequent analyses and omitted PC2 and PC3.

The 2-level model fitted significantly worse than the 3-level model with region as level three clustering variable ($\Delta-2LL = 22.93$, $\Delta df = 1$, $p < 0.001$). So, region in the Netherlands accounts for a statistically significant proportion (1.8%) of the variance in height in 7-year-olds. Table 2 displays the parameter estimates and the standardized variance components of the models. Comparing the parameter estimates of the models shows that the variance attributable to region in the 3-level model was captured by the C-component in the 2-level model.

Results of analyses for genotyped sample

In the genotyped group, region explained 1.6% of the variance, which almost equals the percentage 1.8% reported above. The likelihood ratio test of this component was not significant: $\Delta-2LL = 0.85$, $\Delta df = 1$, $p = 0.36$. However, we ascribed this to a lack of power given the appreciably smaller sample size (in terms of individuals, $N = 7346$, vs. $N = 1375$). The parameter estimates and standardized variance components are displayed in Table 3.

Results of analyses for genotyped sample with PC1 as covariate

Table 4 displays the parameter estimates and standardized variance components of the 2- and 3-level model with PC1 included as a fixed covariate. When we included PC1 in the 3-level model, the variance explained by region went from 1.6% (before inclusion of PC1; see previous section) to $< 0.001\%$. This indicates that when PC1 is included as a covariate, region no longer explains any phenotypic variance in height. This was confirmed by the likelihood ratio test comparing the 2-level and 3-level model. As expected, with PC1 as a covariate the 2-level model fitted equally well as the 3-level model ($\Delta-2LL < 0.001$,

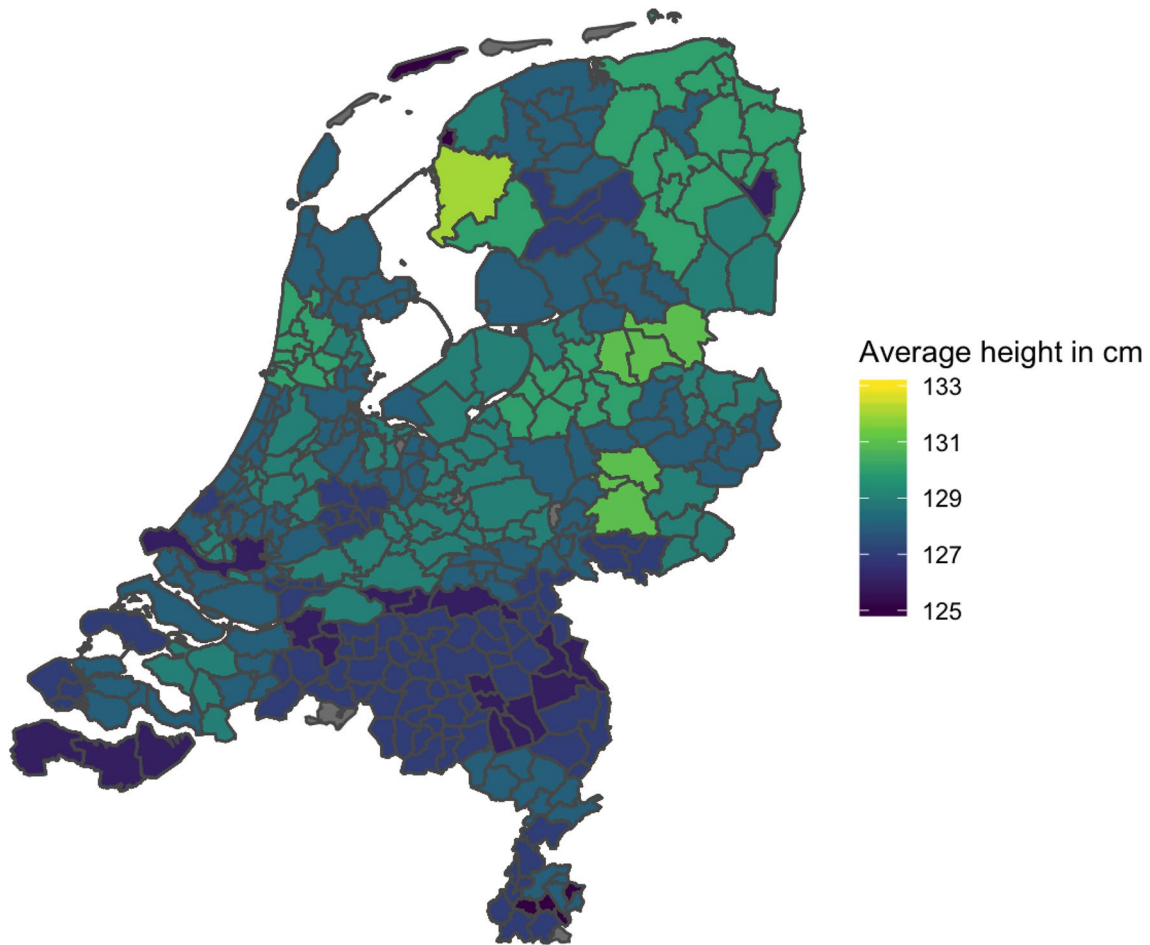


Fig. 6 Mean height (in centimeters) of 7-year-old children in centimeters by region in the Netherlands

Table 2 Results of CTM and 2-level and 3-level MLM analyses for the full sample: path coefficient estimates with standard errors (SE) and standardized variance components of the 2-level and the 3-level models (with age and sex as covariates)

Parameter	CTM	2-Level model	3-Level model
Intercept			
Intercept (SE)	128.4 (0.11)	128.4 (0.11)	128.5 (0.14)
Covariates			
β_{sex} (SE)	-0.62 (0.12)	-0.62 (0.12)	-0.62 (0.12)
β_{age} (SE)	1.43 (0.08)	1.42 (0.08)	1.42 (0.08)
a, c, e, Region path loadings			
a (SE)	4.70 (0.08)	4.70 (0.08)	4.70 (0.08)
c (SE)	2.90 (0.16)	2.90 (0.16)	2.80 (0.16)
e (SE)	1.39 (0.03)	1.39 (0.03)	1.39 (0.03)
Region (SE)			92.33 (2.13)
Total variance ($a^2 + c^2 + e^2$ (+region ²))	32.46	32.46	32.47
A (standardized)	0.681	0.681	0.681
C (standardized)	0.259	0.259	0.241
E (standardized)	0.060	0.060	0.060
REGION (standardized)			0.018

Key differences between the 2-Level and the 3-Level model are printed in bold

N=7346 twins in 3724 families

$\Delta df = 1$), suggesting no effects of region after inclusion of PC1 in the model.

Discussion

In this paper we specified a multilevel twin model in OpenMx and fitted it to data on children’s height. We added a higher-level variable, region in the Netherlands, in which the twin pairs were nested. Adding a third level variable enabled us to determine whether part of the variance in children’s height can be explained by differences in geographical region.

We found that 68% of the variance in 7-year-old children’s height is attributable to additive genetic factors. Common environmental factors accounted for 26%, and unshared environmental factors (including measurement error) for 6% of the variance. We found that regional differences accounted for a significant 1.8% of the phenotypic variance in the complete sample (1.6% in the genotype subsample). In a standard multilevel ACE-twin model, ignoring regional clustering, this variance was captured by the C-component. This is expected, because the common environmental component captures between-family variance, regardless of its source. At age 7, cohabiting MZ and DZ twins necessarily share region, so that the effect of region will contribute to C variance.

Table 3 Results of 2-level and 3-level MLM analyses in the genotyped sample (N = 1375 twins in 714 families)

Parameter	2-Level model	3-Level model
Intercept		
Intercept (SE)	128.4 (0.26)	128.5 (0.27)
Covariates		
β_{sex} (SE)	-0.62 (0.31)	-0.63 (0.31)
β_{age} (SE)	1.16 (0.18)	1.15 (0.18)
a, c, e, Region path loadings		
a (SE)	4.53 (0.20)	4.53 (0.20)
c (SE)	3.21 (0.34)	3.13 (0.16)
e (SE)	1.13 (0.04)	1.13 (0.04)
Region (SE)		0.71 (2.13)
Total variance ($a^2 + c^2 + e^2$ + region ²)	32.11	32.11
A (standardized)	0.640	0.640
C (standardized)	0.320	0.305
E (standardized)	0.040	0.040
REGION (standardized)		0.016

Key differences between the 2-Level and the 3-Level model are printed in bold

Path coefficient estimates with standard errors (SE) and standardized variance components of the 2- and 3-level model (with age and sex as covariates)

Table 4 Results of 2-level and 3-level MLM analyses for the genotyped sample with PC covariate (N = 1375 twins in 714 families)

Parameter	2-Level model	3-Level model
Intercept		
Intercept (SE)	128.5 (0.26)	128.5 (0.26)
Covariates		
β_{sex} (SE)	-0.74 (0.31)	-0.74 (0.31)
β_{age} (SE)	1.16 (0.18)	1.17 (0.18)
β_{PC1} (SE)	0.11 (0.02)	0.11 (0.02)
a, c, e, Region path loadings		
a (SE)	4.55 (0.20)	4.55 (0.20)
c (SE)	3.03 (0.36)	3.03 (0.36)
e (SE)	1.13 (0.04)	1.13 (0.04)
Region (SE)		8.97×10^{-6} (0.76)
Total variance ($a^2 + c^2 + e^2$ + region ²)	31.19	31.19
A (standardized)	0.664	0.664
C (standardized)	0.295	0.295
E (standardized)	0.041	0.041
REGION (standardized)		2.58×10^{-12}

Path coefficient estimates with standard errors (SE) and standardized variance components of the 2- and 3-level ACER model, including PC1

In a subsample of children with genetic PC scores, i.e., the genotyped subsample, we found a statistically significant correlation ($r = 0.16$) between height and the first genetic PC, representing the geographical north–south gradient in the Netherlands. This correlation is similar to previous results for height in a Dutch sample of adults and in line with the findings in European samples, where northern populations are on average taller than southern populations (Abdel-laoui et al. 2013). The correlations between the second and third PC and children’s height were negligible. After the inclusion of the first PC in the multilevel model, region no longer explained any variance.

This last result indicates that the variance in children’s height that is explained by region is attributable to differences in genetic ancestry. That is, although unmodeled regional clustering manifests as C, it does not mean that the inflation of the common environmental variance is due to genuine shared environmental factors like region. When we included the first PC, which reflects differences in allele frequencies between regions, no variance was explained by geographical region above and beyond what was already explained by the PC. Because offspring from the same family share their ancestry, a proportion of the variance that is captured in the C-component of the CTM is actually of a genetic nature. This does not, however, entirely exclude the presence of environmental effects that are explanatory of regional clustering in height. The PC representing the north–south gradient could be correlated with environmental

factors that might contribute to the relationship between PC1, height and regional clustering.

We note the following limitations of our study. First, we did not explicitly model qualitative differences in genetic architecture between boys and girls. There is some evidence that the additive genetic correlation in opposite-sex twins is lower than 0.50, suggesting that partly different genes operate in 7-year-old boys and girls (Silventoinen et al. 2007). However, the twin correlations in our sample did not suggest the presence of qualitative sex differences (we observed correlations of 0.61 and 0.63 in the DZ male and DZ female, versus 0.58 and 0.68 in the DZ opposite sex male–female and DZ opposite sex female–male twins, respectively).

Secondly, we surmise that the power to detect the region effect in the genotyped sample was low, given the sample size ($N = 1375$ in the genotyped sample). However, the effect sizes in both samples were very similar (1.8% vs 1.6%), and in the full sample ($N = 7346$) effect was statistically significance. Therefore, we trust that the regional effect is real.

A final limitation to note is that the current approach assumes that lower levels are fully nested in the higher-level. That is, members of a twin pair cannot differ on the clustering variable. It is therefore not possible to define a third-level clustering variable, when the variable of interest differs within a twin pair (e.g., adult twins who do not live in the same region). It is possible, however, to include variables in which both twins are not nested as a lower-level variance component. When the clustering variable is not specified as a higher-level (i.e., nesting) variable, the effect of clustering can also be manifested as any of the other variance components (i.e., A/C/D/E) when unmodeled. Furthermore, missing data for higher-level clustering variable (here: region) is not allowed. The higher-level variable needs to have a sufficient number of units for the model to have enough power to detect the effect of the higher-level variable (e.g., postal codes in our region example; Goldstein, 2011).

The current study showed that when data are nested in a higher-level variable, adding this higher-level variable to a multilevel model for twin data provides opportunities to further decompose the phenotypic variance. Clustering can be due to unwanted confounding, for example, batch effects. Applying a multilevel model to identify the nuisance variance that is explained by higher-level clustering would in this case serve as a correction. However, as is shown within this paper, the MLM can also be used to empirically study clustering.

Acknowledgements We warmly thank all twin families who take part in the Netherlands Twin Register studies.

Funding This project is part of the Consortium on Individual Development (CID). CID is funded through the Gravitation Program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO: 024-001-003).

The Netherlands Twin Registry (NTR) is funded by ‘Netherlands Twin Registry Repository: researching the interplay between genome and environment’ (NWO: 480-15-001/674) and BBMRI-NL (NWO-184.021.007 and 184.033.111). EvB acknowledges ‘Decoding the gene-environment interplay of reading ability’ (NWO: 451-15-017). MCN, CVD and DIB acknowledge NIH grants DA-49867 and DA-018673

Compliance with ethical standards

Conflict of interest ZT declares that she has no conflict of interest. STK declares that she has no conflict of interest. JJH declares that he has no conflict of interest. MDH declares that he has no conflict of interest. ELdZ declares that she has no conflict of interest. MCN declares that he has no conflict of interest. CEMvB declares that she has no conflict of interest. CVD declares that he has no conflict of interest. EvB declares that she has no conflict of interest. DIB declares that she has no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Parental informed consent was obtained for participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdellaoui A, Hottenga JJ, De Knijff P, Nivard MG, Xiao X, Scheet P et al (2013) Population structure, migration, and diversifying selection in the Netherlands. *Eur J Hum Genet* 21(11):1277–1285. <https://doi.org/10.1038/ejhg.2013.48>
- Abdellaoui A, Hugh-Jones D, Yengo L, Kemper KE, Nivard MG, Veul L et al (2019) Genetic correlates of social stratification in Great Britain. *Nat Hum Behav* 3(12):1332–1342. <https://doi.org/10.1038/s41562-019-0757-5>
- Barton NH, Etheridge AM, Véber A (2017) The infinitesimal model: definition, derivation, and implications. *Theor Popul Biol* 118:50–73. <https://doi.org/10.1016/j.tpb.2017.06.001>
- Baten J, Blum M (2014) Human height since 1820. In: van Zanden JL (ed) *How was life?: Global well-being since 1820*. OECD Publishing, Paris, pp 117–137
- Baten J, Komlos J (1998) Height and the standard of living. *J Econ Hist* 58(3):866–870. <https://doi.org/10.1017/S0022050700021239>
- Bentler PM, Stein JA (1992) Structural equation models in medical research. *Stat Methods Med Res* 1(2):159–181. <https://doi.org/10.1177/096228029200100203>

- Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T (2011) OpenMx: an open source extended structural equation modeling framework. *Psychometrika* 76(2):306–317. <https://doi.org/10.1007/s11336-010-9200-6>
- Boomsma DI, Molenaar PC (1986) Using LISREL to analyze genetic and environmental covariance structure. *Behav Genet* 16(2):237–250. <https://doi.org/10.1007/BF01070799>
- Boomsma DI, Orlebeke JF, van Baal GC (1992) The Dutch Twin Register: growth data on weight and height. *Behav Genet* 22(2):247–251. <https://doi.org/10.1007/BF01067004>
- Boomsma DI, Wijmenga C, Slagboom EP, Swertz MA, Karssen LC, Abdellaoui A et al (2014) The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 22(2):221–227. <https://doi.org/10.1038/ejhg.2013.118>
- Chen G, Saad ZS, Nath AR, Beauchamp MS, Cox RW et al (2012) fMRI group analysis combining effect estimates and their variances. *Neuroimage* 60(1):747–765. <https://doi.org/10.1016/j.neuroimage.2011.12.060>
- Colodro-Conde L, Couvy-Duchesne B, Whitfield JB, Streit F, Gordon S, Kemper KE et al (2018) Association between population density and genetic risk for schizophrenia. *JAMA Psychiatry* 75(9):901–910. <https://doi.org/10.1001/jamapsychiatry.2018.1581>
- Eaves LJ, Last KA, Young PA, Martin NG (1978) Model-fitting approaches to the analysis of human behaviour. *Heredity* 41(3):249–320. <https://doi.org/10.1038/hdy.1978.101>
- Estourgie-van Burk GF, Bartels M, Van Beijsterveldt TC, Delemarre-van de Waal HA, Boomsma DI et al (2006) Body size in five-year-old twins: heritability and comparison to singleton standards. *Twin Res Hum Genet* 9(5):646–655. <https://doi.org/10.1375/twin.9.5.646>
- Falconer DS, McKay TFC (1996) Introduction to quantitative genetics. Burnt Mill Engl. <https://doi.org/10.2307/2529912>
- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinb.* 53:399–433. <https://doi.org/10.1017/S0080456800012163>
- Gelman, A (2005) Analysis of variance—why it is more important than ever. *The annals of statistics* 33(1): 1–53.
- Goldstein H (2011) Multilevel statistical models. Wiley, Chichester
- Guo G, Wang J (2002) The mixed or multilevel model for behavior genetic analysis. *Behav Genet* 32(1):37–49. <https://doi.org/10.1023/A:1014455812027>
- Heath AC, Neale MC, Hewitt JK, Eaves LJ, Fulker DW et al (1989) Testing structural equation models for twin data using LISREL. *Behav Genet* 19(1):9–35. <https://doi.org/10.1007/BF01065881>
- Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *J Educ Psychol* 24(6):417–441. <https://doi.org/10.1037/h0071325>
- Hunter, MD (2020). Multilevel modeling in classical twin and modern molecular behavior genetics. *Behavior Genetics. This Issue*
- Jelenkovic A, Sund R, Hur YM, Yokoyama Y, Hjelmborg JVB, Möller S et al (2016) Genetic and environmental influences on height from infancy to early adulthood: an individual-based pooled analysis of 45 twin cohorts. *Sci Rep* 6(1):1–13. <https://doi.org/10.1038/srep28496>
- Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T et al (2015) Directional dominance on stature and cognition in diverse human populations. *Nature* 523(7561):459–462. <https://doi.org/10.1038/nature14618>
- Karp R, Martin R, Sewell T, Manni J, Heller A (1992) Growth and academic achievement in inner-city kindergarten children: the relationship of height, weight, cognitive ability, and neurodevelopmental level. *Clin Pediatr* 31(6):336–340. <https://doi.org/10.1177/000992289203100604>
- Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38(4):963–974. <https://doi.org/10.2307/2529876>
- Ligthart L, van Beijsterveldt CE, Kevenaar ST, de Zeeuw E, van Bergen E, Bruins S et al (2019) The Netherlands twin register: longitudinal research based on twin and twin-family designs. *Twin Res Hum Genet* 22(6):623–636. <https://doi.org/10.1017/thg.2019.93>
- Liu EY, Li M, Wang W, Li Y (2013) MaCH-Admix: genotype imputation for admixed populations. *Genet Epidemiol* 37(1):25–37. <https://doi.org/10.1002/gepi.21690>
- Longford NT (1993) Regression analysis of multilevel data with measurement error. *Br J Math Stat Psychol* 46(2):301–311. <https://doi.org/10.1111/j.2044-8317.1993.tb01018.x>
- Martin NG, Eaves LJ (1977) The genetical analysis of covariance structure. *Heredity* 38(1):79–95. <https://doi.org/10.1038/hdy.1977.9>
- McArdle JJ, Prescott CA (2005) Mixed-effects variance components models for biometric family analyses. *Behav Genet* 35(5):631–652. <https://doi.org/10.1007/s10519-005-2868-1>
- Neale MC, Cardon LR (1992) Methodology for genetic studies of twins and families. NATO ASI Series. Kluwer Academic Press, Dordrecht. <https://doi.org/10.1007/978-94-015-8018-2>
- Neale MC, Hunter MD, Pritikin JN, Zahery M, Brick TR, Kirkpatrick RM et al (2016) OpenMx 2.0: extended structural equation and statistical modeling. *Psychometrika* 81(2):535–549. <https://doi.org/10.1007/s11336-014-9435-8>
- Paterson L, Goldstein H (1991) New statistical methods for analysing social structures: an introduction to multilevel models. *Br Edu Res J* 17(4):387–393. <https://doi.org/10.1080/0141192910170408>
- Postcodebijaadres (2020) Postcodekaart van Nederland. Retrieved from <https://postcodebijaadres.nl/postcodes-nederland>. Accessed July 29 2020
- Prince AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909. <https://doi.org/10.1038/ng1847>
- Pritikin JN, Hunter MD, von Oertzen T, Brick TR, Boker SM (2017) Many-level multilevel structural equation modeling: an efficient evaluation strategy. *Struct Eq Model* 24(5):684–698. <https://doi.org/10.1080/10705511.2017.1293542>
- Privé F, Luu K, Blum MG, McGrath JJ, Vilhjálmsson BJ (2020) Efficient toolkit implementing best practices for principal component analysis of population genetic data. *Bioinformatics* 36(16):4449–4457. <https://doi.org/10.1093/bioinformatics/btaa520>
- R Development Core Team (2020) R: a language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org> (ISBN 3-900051-07-0)
- Rabe-Hesketh S, Skrondal A, Gjessing HK (2008) Biometrical modeling of twin and family data using standard mixed model software. *Biometrics* 64(1):280–288. <https://doi.org/10.1111/j.1541-0420.2007.00803.x>
- Reich D, Price AL, Patterson N (2008) Principal component analysis of genetic data. *Nat Genet* 40(5):491–492. <https://doi.org/10.1038/ng0508-491>
- Rijsdijk FV, Sham PC (2002) Analytic approaches to twin data using structural equation models. *Brief Bioinform* 3(2):119–133. <https://doi.org/10.1093/bib/3.2.119>
- Schalekamp JC (2009) Bataven en Buitenlanders: 20 Eeuwen Immigratie in Nederland. Wind Publishers, Huizen, pp 15–40
- Scharpf RB, Ruczinski I, Carvalho B, Doan B, Chakravarti A, Irizarry RA (2011) A multilevel model to address batch effects in copy number estimation using SNP arrays. *Biostatistics* 12(1):33–50. <https://doi.org/10.1093/biostatistics/kxq043>
- Sellström E, Bremberg S (2006) Is there a “school effect” on pupil outcomes? A review of multilevel studies. *J Epidemiol Commun Health* 60(2):149. <https://doi.org/10.1136/jech.2005.036707>
- Silventoinen K, Krueger RF, Bouchard TJ, Kaprio J, McGue M (2004) Heritability of body height and educational attainment in an

- international context: comparison of adult twins in Minnesota and Finland. *A J Hum Biol* 16(5):544–555. <https://doi.org/10.1002/ajhb.20060>
- Silventoinen K, Bartels M, Posthuma D, Estourgie-van Burk GF, Willemsen G, van Beijsterveldt TC et al (2007) Genetic regulation of growth in height and weight from 3 to 12 years of age: a longitudinal study of Dutch twin children. *Twin Res Hum Genet* 10(2):354–363. <https://doi.org/10.1375/twin.10.2.354>
- Spears D (2012) Height and cognitive achievement among Indian children. *Econ Hum Biol* 10(2):210–219. <https://doi.org/10.1016/j.ehb.2011.08.005>
- Steiger JH, Shapiro A, Browne MW (1985) On the multivariate asymptotic distribution of sequential Chi square statistics. *Psychometrika* 50(3):253–263. <https://doi.org/10.1007/BF02294104>
- The Genome of the Netherlands Consortium, Francioli LC, Menelao A, Pulit SL, Van Dijk F, Palamara PF, Elbers CC et al (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 46(8):818. <https://doi.org/10.1038/ng.3021>
- Van den Oord EJCG (2001) Estimating effects of latent and measured genotypes in multilevel models. *Stat Methods Med Res* 10:393–407.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.