

1 Genetic Nurture: Estimating the direct genetic effects of pediatric anthropometric traits

2

3 Samuel Ghatan<sup>1</sup>, Jard de Vries<sup>1</sup>, Jean-Baptiste Pingault<sup>2,3</sup>, Vincent W Jaddoe<sup>4,5</sup>, Charlotte

4 Cecil<sup>4,6</sup>, Janine F Felix<sup>4,5</sup>, Fernando Rivadeneira<sup>1,4</sup>, Carolina Medina-Gomez<sup>1,#</sup>

5

6 1. Department of Internal Medicine, Erasmus MC, University Medical Center

7 Rotterdam, Rotterdam, The Netherlands

8 2. Department of Clinical, Educational and Health Psychology, University College

9 London, United Kingdom

10 3. UCL Human Genomics, University College London, United Kingdom.

11 4. The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam,

12 Rotterdam, The Netherlands

13 5. Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam,

14 Rotterdam, The Netherlands

15 6. Department of Child and Adolescent Psychiatry and Psychology, Erasmus MC

16 University Medical Center Rotterdam, Rotterdam, the Netherlands

17

18

19 #Corresponding author

20

21 Corresponding author:

22 Carolina Medina-Gomez,

23 Department of Internal Medicine

24 Erasmus University Medical Center Rotterdam

25 Postbus 2040, Rotterdam, 3000 CA

26 The Netherlands

27 e-mail: [m.medinagomez@erasmusmc.nl](mailto:m.medinagomez@erasmusmc.nl)

28

29 Keywords: Genetic nurture, indirect genetic effects, polygenic risk score, children, growth,

30 body mass index

31

32 **Abstract (250 words max, one paragraph)**

33 Parental genetic variants can indirectly influence the traits of their child through the  
34 environment, a concept termed "genetic nurture", or indirect genetic effects (IGE). This study  
35 estimated the direct genetic effects (DGE), via direct allelic transmission, and IGE shaping  
36 height, body mass index (BMI), and bone mineral density (BMD) in a multi-ethnic Dutch  
37 pediatric cohort, examining children with repeated measurements at ages six, nine, and  
38 thirteen. We imputed missing parental alleles from the phased haplotypes of 1,931,478  
39 variants (MAF >1%), utilizing snipar (single nucleotide imputation of parents). We  
40 constructed polygenic risk scores (PRSs) and jointly regressed the proband's trait on their  
41 own PRS, while controlling for the proband's maternal and paternal PRSs. A total of 4,488  
42 probands, with genetic data, underwent at least one of the three specified measurements. We  
43 found statistically significant DGE estimates for the three traits across ages six, nine and  
44 thirteen. For instance, 71-77% of the BMI variance explained by the BMI-PRS can be  
45 attributed solely to the DGE. IGE estimates reached significance only for BMI measured at  
46 ages nine (Beta: 0.05, 95%CI: 0.01-0.09) and thirteen (Beta: 0.05, 95%CI: 0.01-0.09).  
47 Maternal and paternal IGE were of a similar magnitude in all our analyses. Our findings  
48 indicate that genetic nurture has limited influence on anthropometric traits during formative  
49 years. In addition, we do not observe differences between the maternal and paternal indirect  
50 contributions to these traits, opposite to the stronger maternal nurturing effect reported for  
51 other traits.

52

53

54

55 **Introduction**

56

57 Heritability, in the context of genetics, refers to the proportion of phenotypic variance  
58 determined by genetic factors [1]. Inherited DNA is passed from parent to offspring during  
59 reproduction. Inheritance is governed by Mendel's laws of genetics, such that each parent  
60 passes half of their autosomal DNA to their offspring through meiosis. Consequently, parents  
61 and offspring share a resemblance in heritable traits. Parents can also influence the traits of  
62 their offspring through the shaping of their offspring's environment, a phenomenon often  
63 referred to as nurture. If the offspring's trait is influenced by the environment and is heritable,  
64 then the genetic variants of the parents can indirectly influence the offspring's trait via the  
65 parent's own trait and therefore the environment of the child [2]. Thus, the variants of the  
66 parent can influence the trait of their offspring through two causal paths: directly by allele  
67 transmission and indirectly through the environment. The latter phenomenon was coined  
68 "genetic nurture". Kong et al and Bates et al first demonstrated this concept in a human  
69 population by examining the genetic nurture, also known as indirect genetic effects (IGE), for  
70 educational attainment [3, 4]. Kong et al showed that the non-transmitted alleles of the  
71 parents had an effect on the educational attainment of the proband that is 29.9% ( $P = 1.6 \times$   
72  $10^{-14}$ ) of that of the transmitted polygenic score. We can estimate the direct genetic effects  
73 (DGE), i.e. the direct allelic transmission, and IGE by jointly regressing the proband's  
74 phenotype with the polygenic risk score (PRS) of the proband (i.e., child), the mother and the  
75 father. A PRS consists of the sum of risk alleles associated with a particular trait, weighted by  
76 the effect sizes of those variants as identified in genome-wide association studies (GWAS). In  
77 the joint regression model, the DGE are estimated as the magnitude of the effect of the  
78 proband's PRS on the trait of interest, whereas, the IGE are estimated by taking the average  
79 effect estimate of the mother's and father's PRS within the model [5].

80 Nurturing conditions play an important role in child development. Growing evidence  
81 suggests that early childhood experiences, particularly family environment shape, among  
82 others, dietary and physical activity behaviors associated with growth velocity [6, 7]. For  
83 example, the adequate intake of basic nutrients through childhood and adolescence plays a  
84 crucial role in determining height, accumulation of muscle and fat mass, and accrual of peak  
85 bone mass [8-11]. These traits are partially influenced by genetics and are also shaped by the  
86 environmental conditions provided by parents. Large GWAS have identified thousands of  
87 genetic variants associated with height [12], body mass index (BMI) [13], and bone mineral  
88 density (BMD) [14], providing valuable information to estimate an individual's genetic  
89 liability for growth. However, classical GWAS can only estimate the combined effects of the  
90 direct and indirect effects [2, 15]. The influence of the family environment on phenotypic  
91 variation can change throughout an individual's life, as interactions among family members  
92 evolve. For example, the magnitude of parental nurturing may diminish as a child grows  
93 older [16]. It is then plausible that parental IGE, if present, will have the greatest impact on  
94 anthropometric traits during childhood and adolescence. To our knowledge, no study has  
95 examined the genetic nurturing effect on anthropometric traits in a pediatric population across  
96 multiple developmental stages. Therefore, we aimed to estimate the direct and indirect  
97 genetic effects in in developmental traits (i.e., height, BMI and BMD) at multiple timepoints  
98 through childhood development and test how these effects might vary. Our study surveyed  
99 longitudinal data of participants from a Dutch multi-ethnic cohort at ages six, nine and  
100 thirteen.

101

## 102 **Methods**

103

104 Study characteristics

105 The Generation R Study is a multi-ethnic population-based prospective cohort study from  
106 fetal life onwards based in Rotterdam, The Netherlands. The study recruited pregnant women  
107 who were expected to deliver between 2002 and 2006, and encompasses 9,778 mothers and  
108 their children [17]. During these years, a vast array of data ranging from detailed  
109 questionnaires and regular medical examinations to advanced imaging techniques and  
110 extensive biosampling were gathered, and children and parents are in ongoing follow-up. For  
111 this study, height, BMI and BMD were measured in the probands during three visits to the  
112 research center at a mean age of 6.0, 9.7, and 13.5 years.

113

114 Acquisition of the phenotype data

115

116 In the Generation R Study, at each wave of the study, participants removed shoes, heavy  
117 clothing, and metal items before measurements. Height was measured to the nearest  
118 millimeter by a stadiometer (Holtain Limited, Dyfeld, UK). Weight was measured to the  
119 nearest gram using an electronic scale (SECA, Almere, The Netherlands). BMI was  
120 calculated [weight (kg)/height (m<sup>2</sup>)]. Height and BMI were then transformed into age- and  
121 sex-adjusted standard deviation scores (SDS) using LMSGrowth [18]. Total body BMD was  
122 assessed via Dual-energy- X-ray absorptiometry (DXA) scans using a GE-Lunar iDXA  
123 device operated by a skilled research assistant with daily quality checks, following standard  
124 manufacturer protocols [19]. Total body less head BMD was used for analysis as  
125 recommended by the International Society for Clinical Densitometry guidelines [20].

126

127 Genotyping and imputation of the Study population

128

129 Proband and parents were genotyped in three batches using three different arrays over a  
130 period of 10 years. Due to the limited overlap in overall content across genotyping arrays, the  
131 genotype data was first subjected to imputation to increase the number of overlapping  
132 variants. Proband and parents were genotyped in two batches. The first round of genotyping was carried  
133 out with either the Illumina HumanHap 610 or 660 Quad chips and is referred to here as  
134 Generation R3 (GENR3). It comprised data from 5,722 probands after quality control (QC).  
135 A detailed description of the genotyping can be found elsewhere [21]. Genotypes were then  
136 imputed to the 1000 genomes project phase3v5 (1KGP) reference panel using minimac3. The  
137 subset of genome-wide-variants with an  $r$ -squared  $> 0.95$  and a minor allele frequency (MAF)  
138  $> 1\%$ , after imputation, comprised 4,789,340 SNPs.

139 The second round of proband genotyping, referred to as Generation R4 (GENR4), was  
140 conducted together with a subset of participants' mothers in the study. In total, 3,424 samples  
141 (i.e., including both children and mothers) were genotyped in the HuGE facility at Erasmus  
142 Medical center. Genome Studio 2.0 was used to initially QC genetic variants and genotype  
143 clusters after genotyping. Data was further processed in PLINK 1.9 for the following steps.  
144 Exclusion of samples and variants with call rates below 97.5%, variants with an excess of  
145 heterozygosity ( $> 4$  standard deviations from the mean), Hardy-Weinberg equilibrium (HWE)  
146 deviation ( $P < 1 \times 10^{-4}$ ), or samples deemed to be sex mismatches and/or genetic duplicates. In  
147 addition, we detected and excluded potential sample swaps by comparing reported familiar  
148 relationships based on questionnaires across all Generation R samples (i.e., both children and  
149 mothers) and IBD-based relationships. Lastly, zCall was used on the previously uncalled  
150 genotypes aiming to improve the call rate of less-frequent to rare variants [22]. More  
151 stringent filtering of both samples and variants after zCall filtering was applied including  
152 missingness ( $> 99\%$ ) and HWE ( $P < 1 \times 10^{-7}$ ). In total, 1,809 children and 708,478 variants are  
153 included in the final genotyped dataset distributed to Generation R researchers. Next,

154 imputation was carried out. GENR4 variants were surveyed in the 1KG panel, and these not  
155 present or with different reported alleles, excluded. The remaining variants were phased using  
156 ShapeIT v2 [23], followed by imputation to the 1KG using minimac4 [24] with an in-house  
157 pipeline. The subset of variants with an  $r$ -squared  $> 0.95$  and MAF  $> 1\%$ , after imputation,  
158 comprised 2,921,550 SNPs.

159

160 Parents of probands were genotyped in two rounds, using the GSA-MD v2.0 ( $n = 1,530$   
161 (13%), only mothers), and GSA-MD v3.0 ( $n = 10,213$  (87%), both mothers and fathers)  
162 genotyping arrays. The QC of the first batch is described in detail in the previous section and  
163 supplementary figure 1 while the detailed QC of the second batch of GENR Parents is  
164 described in figure 2. For this second batch, Genome Studio 2.0 was employed to perform a  
165 technical QC of genotyped variants and clusters, followed by a strict PLINK 2 QC.

166 Exclusions were applied to samples and variants with call rates below 97.5%, variants with  
167 an excess of heterozygosity ( $> 4$  standard deviations from the mean), HWE deviation  
168 ( $P < 1 \times 10^{-5}$ ), or samples deemed to be sex mismatches and/or genetic duplicates. In addition,  
169 we detected and excluded potential sample swaps by a comparison of reported familiar  
170 relationships based on questionnaires across all Generation R samples (i.e., both children and  
171 parents) and KING reported relationships [25]. zCall was used on the previously uncalled  
172 genotypes aiming to improve the call rate of less-frequent to rare variants [22]. More  
173 stringent filtering of both samples and variants after zCall filtering was applied including  
174 missingness ( $> 99\%$ ) and HWE ( $P < 1 \times 10^{-7}$ ). Both sets of parents genotyped data were merged  
175 using PLINK 2, and variants not present in both array platforms or reporting different alleles  
176 were excluded. In total, the GENR Parents combined information of 660,868 variants present  
177 in 11,742 individuals. The combined GENR Parents dataset was phased using ShapeIT v2  
178 [26], followed by imputation to the 1KG reference panel using minimac4 [24] using our in-



179 house pipeline. Imputation resulted in a subset of 3,000,903 SNPs with an  $r$ -squared  $> 0.95$   
180 and  $MAF > 1\%$ . Lastly, retrieved informed consents and family cross-checks based in the  
181 four genotyped batches described in the sections above resulted in extra exclusions, noted in  
182 an inclusion file share among Generation R researchers. The working genome-wide files for  
183 Generation R comprise GENR3 (5,700 participants), GENR4 (1,802 participants) and GENR  
184 Parents (11,676, parents).

185

186 Construction of the combined Generation R dataset

187

188 The subsets of imputed variant files described above (i.e.,  $MAF > 1\%$  and  $r$ -squared  $> 0.95$ ),  
189 that were present across all three files, were then merged. Palindromic variants (coded as A-T  
190 or G-C) with ambiguous effect allele frequencies (40-60%) were removed. Variants that  
191 exhibited batch effects, as determined by examining PCA plots and by conducting a GWAS  
192 using batch as the outcome, were subsequently removed. Overall, the merged set comprised  
193 1,931,478 variants and 19,178 samples after QC. This merged set of variants was then phased  
194 with SHAPEIT2 [26] with the duohmm function, -W 5 parameter (window size) and the  
195 1KGP reference panel (All population) as a genetic map [27]. The duohmm function utilizes  
196 pedigree information to improve phasing.

197

198 Determination of the Genetic ancestry of the participants

199

200 Individuals of European genetic ancestry were identified using genetic data from the 1KGP  
201 as a reference. Our dataset and the European set of the 1KGP (e.g., CEU, TSI, GBR, FIN,  
202 IBS) were harmonized to ensure overlapping variants and consistent allele coding.

203 Palindromic variants and variants with  $MAF < 1\%$  were removed. In the merged dataset,

204 variants were then pruned ( $r^2 > 0.2$  in a 50kb window). Principal component analysis (PCA)  
205 was conducted using PLINK. Using the first two principal components, the center of the  
206 1KGP European samples was calculated. The maximum Euclidean distance of the 1KGP  
207 European samples was then calculated. All GENR samples whose maximum Euclidean  
208 distance fell beyond this radius were defined as non-European.

209

210 Imputation of missing parental alleles

211

212 DGE and IGE can only be estimated in complete trios. In duos, the missing/ungenotyped  
213 parental alleles have to be imputed. We utilized snipar (single nucleotide imputation of  
214 parents) to impute missing parental alleles in probands of European ancestry [5]. Briefly,  
215 snipar imputes ungenotyped parental alleles based on Mendelian laws of inheritance. Feasible  
216 nonlinear imputations of parental alleles can be grouped into three categories: (a) genotyped  
217 sibling pairs, (b) genotyped parent–offspring pairs, and (c) genotyped sibling pairs with one  
218 genotyped parent. By using phased genotype data, uncertainty at heterozygous positions can  
219 be resolved. Missing genotypes that cannot be inferred through family structure were imputed  
220 using the in-sample population allele frequencies. As a result, only parents of European  
221 ancestry were included in the imputation. Phased VCF files were converted to bgen files  
222 using QCtools. Identity by descent (IBD) segments shared between siblings were inferred  
223 using the implementation within snipar, which employs a Hidden Markov Model [5]. KING  
224 software [25] was used to obtain kinship coefficients and pedigree information. A kinship  
225 coefficient of 0.177 and above was used to define first degree relatives. KING identified  
226 10,229 parent offspring relations. Families consisted of 3,635 trios and 2,897 duos and 267  
227 proband-sibling pairs. Kinship files, as well as files denoting the age and sex of individuals  
228 were used to infer IBD segments and pedigree structure. Missing parental genotypes (Table

229 1) were imputed by snipar for 1,931,478 variants using phased haplotypes. After imputation  
230 and QC a total of 5,241 European and non-European probands had complete parental data  
231 available (Table 1).

232

233 Polygenic risk scores

234 To construct the PRS for each trait, variant effect estimates were obtained from GWAS  
235 summary statistic data for height [12], BMI [13], and estimated bone mineral density  
236 (eBMD) [14]. These GWAS were chosen due to their representation of the largest sample  
237 sizes available for each respective trait. To increase the variance explained by each PRS,  
238 LDpred2 [28] was used to construct genome-wide PRSs across autosomal chromosomes.  
239 LDpred2 uses a Bayesian approach that adjusts the effect estimates of GWAS summary  
240 statistics using linkage disequilibrium (LD) information to estimate a posterior mean effect  
241 size for each SNP, resulting in a more accurate PRS prediction. The LDpred automatic model  
242 was used, in which hyper-parameters are directly inferred from the data without the need for  
243 a validation set. Variants in the GWAS summary statistics were intersected with the  
244 1,931,478 imputed variants of our sample and the 1,444,196 “HapMap3+” variants [29].  
245 Palindromic variants were removed. In sample LD matrices were generated from the imputed  
246 variants of the probands. Once variant weights had been computed using LDpred, snipar was  
247 used to generate the PRSs for the probands, mothers and fathers. The same set of variants was  
248 used to construct the proband and parental PRSs. To obtain estimates of direct and indirect  
249 genetic effects, we jointly regressed the proband’s trait on their own PRS, while controlling  
250 for the proband’s maternal and paternal PRS. IGE are quantified as the average of the  
251 maternal and paternal PRS effects within equation 1.

252

253 (Equation 1) 
$$Y_{proband} = \delta PRS_{proband} + \beta PRS_{mother} + \beta PRS_{father} + \epsilon$$

254

255 Where,  $Y$  = proband phenotype,  $\delta$  = direct genetic effect,  $\beta(PRS_{mother} + PRS_{father})/2 =$   
256 indirect genetic effect. Linear regression models were constructed using R software and were  
257 adjusted by age, age<sup>2</sup>, sex, the interaction between age and sex, and principal components 1  
258 to 10. Principal components were constructed from the genetic data of the individuals being  
259 tested. Traits were standardized to have a mean of zero and a standard deviation of one. The  
260 PRS for eBMD and BMI was tested in the European subset of the sample, as the GWAS for  
261 these traits was derived from a European population and was poorly associated within our  
262 non-European sample. The height GWAS from which summary statistics were extracted was  
263 performed in a multiethnic population and as such, we used the complete GENR dataset to  
264 infer DGE and IGE. A two-sided Z-test was used to calculate the difference in effects  
265 between the maternal and paternal PRSs, as well as, the difference in effects between  
266 different age measurements. The ratio of IGE and DGE can be determined using the  
267 following method: DGE/IGE+DGE. To determine the fraction of the phenotype variance  
268 which is explained by the DGE we square this ratio.

269

270 Analysis in independent probands

271

272 Closely related individuals can cause inflation in the estimates of the association of the PRS  
273 and the trait tested [30]. This phenomenon arises, in part, because individuals from the same  
274 family share genetic information but also more environmental factors than unrelated  
275 individuals. For instance, individuals in nuclear families often share the same household. If  
276 these shared factors are associated with the trait of interest, it can lead to confounding in PRS  
277 analyses [30]. While GWAS methods attempt to control for broad population stratification,

278 they may not adequately control for subtle stratification caused by family relationships [31].  
279 To evaluate whether the presence of related participants of the Generation R Study led to  
280 inflated association effect sizes, we performed a sensitivity analysis in which children related  
281 up to the third degree (i.e., first cousins) were removed from the regression analysis.  
282 Kingship coefficients were derived from the analysis performed in KING.

283

284 Re-imputation of observed parental genotypes

285

286 To assess the quality of the imputation we proceeded to re-impute measured parental alleles  
287 and then compared the re-imputed alleles to the observed alleles. To perform the analysis, we  
288 started with complete trios families of European ancestry (n = 8055), of which 2,987 were  
289 probands. 5,470 were parents. Next, the allele information of one parent from 844 probands  
290 (30%) was randomly removed. Missing parents genotypes were imputed using snipar and  
291 then the correlations between imputed genotypes and measured genotypes were tested for  
292 each individual across 1,931,478 variants. The mean correlation across all individuals and  
293 variants was as our measure of the imputation accuracy.

294

## 295 **Results**

296 Description of the data acquisition

297 Proband's measurements were collected during three visits to the research center, roughly  
298 four years apart. Measurements of probands took place at the median age of 6.0 (IQR: 5.9-  
299 6.2), 9.73 (IQR: 9.6-9.9), and 13.5 (IQR: 13.4-13.7). Of the 4,177 probands that had one of  
300 the three measurements and genotype data available, 2,587 (62%) had measurements across  
301 all three visits. Between 83-84% of the participants measured at each visit, were of European

302 ancestry. A full description of the participants taking part in each measurement wave can be  
303 found in Table 1.

304

| Data type  | N probands | Missing data imputed |
|--|------------|----------------------|
| Proband and both parents (European)              | 2,736      | No                   |
| Proband and one parent (European)                | 1,576      | Yes                  |
| Proband and sibling(s) but no parents (European) | 57         | Yes                  |
| Proband and both parents (Non-European)          | 872        | No                   |
| Total  | 5,241      |                      |

305

306 Table 1 – Overview of the familial data types within the Generation R Study.

307

308

309

310

311 Imputation of missing parental genotypes

312

313 In total 1,690 missing parental genotypes of European ancestry were imputed using snipar

314 (Table 2). The structure of the missing family data consisted of 1,319 missing father

315 genotypes with the mother and proband genotypes available, 257 missing mother genotypes

316 with the father and proband genotypes available, and 57 probands with no parental genotype

317 data but with a sibling genotyped. To assess the quality of the imputation we re-imputed

318 measured parental alleles and calculated the correlation between the re-imputed alleles and

319 observed alleles. When assessing the quality of the imputation, the mean correlation between

320 the observed alleles and the re-imputed alleles was 0.82, indicating a small bias in the  
321 imputed data.  
322

| Characteristic                    | Measurement age 6  | Measurement age 9  | Measurement age 13  |
|-----------------------------------|--------------------|--------------------|---------------------|
| Probands with family genetic data | 5241               | 5241               | 5241                |
| Attended assessment center        | 4178               | 3831               | 3232                |
| Age (years)                       | 6.01 (5.86-6.23)   | 9.73 (9.61-9.87)   | 13.54 (13.42-13.71) |
| Female (%)                        | 49.42              | 49.64              | 49.22               |
| European (%)                      | 84.38              | 84.4               | 83.2                |
| BMI (sds)                         | 0.17 (-0.34-0.75)  | 0.18 (-0.48-0.87)  | 0.14 (-0.56-0.90)   |
| NAs                               | 1                  | 74                 | 6                   |
| Height (sds)                      | -0.18 (-0.84-0.47) | -0.09 (-0.70-0.57) | 0.06 (-0.63-0.74)   |
| NAs                               | 1                  | 74                 | 6                   |
| TBLH-BMD (g/cm <sup>2</sup> )     | 0.55 (0.52-0.58)   | 0.67 (0.64-0.72)   | 0.85 (0.78-0.92)    |
| NAs                               | 108                | 117                | 176                 |

323

324 Table 2 – Descriptive statistics of the individuals contributing data for this study at ages, six,  
325 nine and thirteen.

326

327 Estimating direct and indirect genetic effects for childhood anthropometric traits

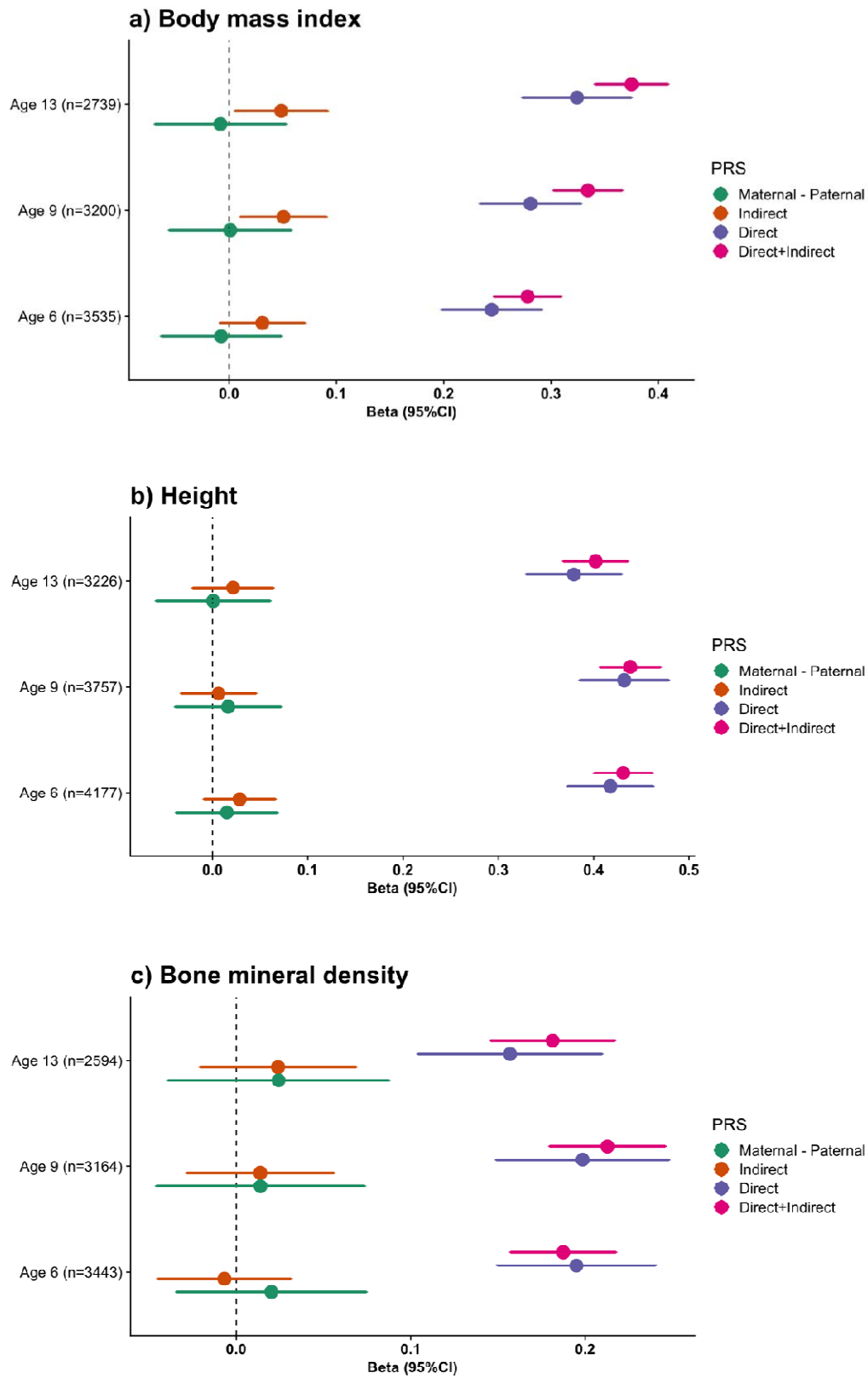
328

329 The PRS for BMI was constructed using 195,154 genetic variants across the genome and  
330 explained between 7.7-14.0% of the BMI phenotypic variance (supplementary table 1) at

331 each measurement (European subset). We observed a correlation coefficient ( $\rho$ ) of 0.55  
332 (95%CI: 0.53-0.57) between proband and mother PRS, a  $\rho$  of 0.57 (95%CI: 0.55-0.59)  
333 between proband and father PRS, and a  $\rho$  of 0.06 (95%CI: 0.02-0.10) between the parent's  
334 PRSs. We obtained statistically significant DGE estimates for BMI across all ages (Figure 1a,  
335 supplementary table 1). DGE explained between 6.0-10.5% of the BMI phenotypic variance  
336 (supplementary table 1) at the different ages the assessment was performed. Conversely, IGE  
337 estimates on BMI reached statistical significance for the measurements at age 9 (Beta: 0.05,  
338 95%CI: 0.01-0.09) and at age 13 (Beta: 0.05, 95%CI: 0.01-0.09). The IGE estimate at age 6  
339 was borderline significant (Beta: 0.03, 95%CI: -0.01-0.07). The ratio of DGE/IGE+DGE  
340 ranged between 0.84-0.88 across the different ages. Between 71-77% of the total variance  
341 explained by the BMI PRS is due to the direct effects alone. There were no significant  
342 differences between maternal and paternal effects or between age groups at any of the ages  
343 tested.



It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).



345 Figure 1 – Forest plots depicting the direct, indirect and maternal minus paternal genetic  
346 effects for: a) Body mass index b) Height c) and Bone mineral density measured at different  
347 ages in participants of the Generation R Study.

348

349 The height PRS was constructed using 210,390 genetic variants across the genome and  
350 explained between 16.2-19.2% of the phenotypic variance of height (supplementary table 1)  
351 at each measurement wave. The correlation between proband and mother PRSs was 0.60  
352 (95%CI: 0.59-0.61), proband and father PRSs, 0.62 (95%CI: 0.61-0.63) and between parent's  
353 PRSs, 0.17 (95%CI: 0.16-0.18). We obtained statistically significant unbiased DGE estimates  
354 for height across all ages (Figure 1b, supplementary table 1). DGE explained between 14.4-  
355 18.7% of height phenotypic variance (supplementary table 1). IGE Estimates of did not reach  
356 significance for any of the ages tested (p-value > 0.05). Direct effects explained between 94-  
357 99% of the height variance explained by its correspondent PRS. There were no significant  
358 differences between maternal and paternal effects or between age groups at any of the ages  
359 tested.

360

361 The PRS for BMD was constructed using 251,272 variants across the genome and explained  
362 between 3.3-4.5% of the total body less head BMD phenotypic variance (supplementary table  
363 1) at the different measurement waves (European subset). The correlation between proband  
364 and mother PRSs was 0.53 (95%CI: 0.51-0.55), proband and father PRSs were 0.56 (95%CI:  
365 0.54-0.58) and between parent's PRSs was 0.04 (95%CI: -0.01;0.06). We obtained  
366 statistically significant unbiased DGE estimates for BMD across all age groups (Figure 1c,  
367 supplementary table 1). DGE explained between 2.5-3.9% of the BMD phenotypic variance  
368 (supplementary table 1). IGE estimates did not reach statistical significance for any of the

369 ages tested. There were no significant differences between maternal and paternal effects or  
370 between age groups at any of the ages tested.

371

372 Sensitivity analysis

373

374 To assess the extent to which our PRS coefficients were inflated due to relatedness within our  
375 sample population, we conducted a sensitivity analysis including only independent  
376 individuals (supplementary table 2). Overall, DGE and IGE coefficients remained consistent  
377 between both analyses indicating minimal bias due to relatedness in our analysis  
378 (supplementary table 2).

379

## 380 **Discussion**

381

382 We examined genetic nurture effects that may shape childhood anthropometric traits across  
383 ages six, nine and thirteen in a multi-ethnic Dutch cohort study. We imputed missing parental  
384 genotypes based on Mendelian inheritance. Polygenic risk scores (PRSs) were constructed for  
385 mothers, fathers and probands, for three childhood anthropometric traits: height, BMI and  
386 BMD (total body less head), using hundreds of thousands of genetic variants. We obtained  
387 statistically significant estimates of direct genetic effects (DGE) for all traits. We also  
388 observed significant estimates of indirect genetic effects (IGE) for BMI but not height or  
389 BMD. In general, we observed modest estimates of genetic nurture effects of anthropometric  
390 traits throughout developmental ages, with no significant differences in IGE for any traits  
391 between different age ranges in children. Additionally, no differences were observed in the  
392 contribution of maternal and paternal genetic nurture effects. These results indicate that the

393 genetic effects of anthropometric traits in children predominantly occur via genetic  
394 transmission, although some genetic nurture effects are present for BMI.  
395 Our study is not the first to examine the effects of genetic nurture on child BMI trajectories.  
396 Tubbs et al. examined the influence of maternal genetic nurture on child BMI trajectories in  
397 approximately 2,900 children and found that its effect increased with age, becoming stronger  
398 throughout development [32]. Notably, Tubbs et al. focused solely on maternal genetic  
399 nurture, without accounting for potential paternal influences. This finding contrasts with  
400 studies on twins, which suggest that the impact of the shared environment, including genetic  
401 nurture, diminishes over childhood and adolescence [33]. Further complicating the picture, a  
402 GWAS of BMI at different ages found that distinct genetic factors influence BMI during  
403 infancy (2 weeks to 18 months) compared to later childhood (18 months to 13 years) [34].  
404 Similarly, a study based on pooled twin data reported that genetic correlations for BMI  
405 decrease with age, indicating that new genetic factors emerge at different stages of childhood  
406 growth [35]. In the Generation R study, a polygenic risk score (PRS) for adult BMI explained  
407 varying amounts of BMI variance across age groups, with the highest variance at age 13,  
408 suggesting that different genetic variants influence BMI across childhood, or that adult BMI  
409 variants have weaker effects at younger ages. Additionally, a study tracking BMI PRSs from  
410 childhood to adulthood found that differences in weight between individuals in the highest  
411 and lowest deciles of the PRS were evident early in childhood and grew to 12 kilograms by  
412 age 18 [36]. Together, these findings highlight the dynamic role of genetic factors, with  
413 varying influences across different developmental stages.  
414  
415 In a prior study that assessed genetic nurturing effects in 21,637 adult probands from an  
416 Icelandic cohort study, a DGE/IGE+DGE ratio of 0.94 for height was reported [3] aligning  
417 with the estimated ratios from our study. Despite the large sample size of the Icelandic study,

418 the authors did not find a significant IGE for BMI [3]. In contrast, we report a significant IGE  
419 for BMI. As the Icelandic study was carried out on adults, it is plausible that the influence of  
420 the IGE on BMI is more accentuated during childhood. This aligns with the logic that parents  
421 have more control over their child's environment during childhood, but less control over their  
422 environment in adulthood. Still, we observed no significant disparities in IGE on BMI across  
423 the developmental ages of our examined cohort, suggesting that a potential divergence of IGE  
424 may occur later in adolescence or adulthood. Another study of roughly 56,500 adult probands  
425 from the UK biobank, Generation Scotland and the Swedish Twin registry produced a  
426 DGE/IGE+DGE ratio of 0.910 (s.e. = 0.009) for height, and 0.962 (s.e. = 0.017) for BMI  
427 [37]. Again, the IGE values for height in this study are similar to those obtained from the  
428 Generation R Study, while the IGE values for BMI in this study are lower. However, the  
429 Generation R Study displays larger standard errors for BMI IGE and the BMI GWAS  
430 summary statistics used in our study were different. No prior studies have investigated the  
431 IGE or DGE of BMD in either pediatric or adult populations, precluding a direct comparison  
432 with our research. While IGE for BMD did not reach statistical significance, we did observe  
433 increasing effect estimates with age. Correlations between parental scores for height, and to a  
434 lesser extent BMI, were inflated, indicating bias, possibly driven by assortative mating or  
435 population stratification, which are also captured by IGE [39, 40]. Assortative mating refers  
436 to the phenomenon in which individuals are more likely to mate with those who share a  
437 similar trait than is expected under random mating, and is a well-documented phenomenon of  
438 the genetics of height [5, 39, 40]. Previous studies have shown that much of the IGE on  
439 height is attributable to assortative mating, whereas, for BMI, assortative mating has been  
440 shown to have a negligible effect [37, 40]. BMD on the other hand displayed no evidence of  
441 assortative mating in our study, in accordance with previous estimations [39]. We did not  
442 observe any statistically significant difference between maternal and paternal IGE on the

443 outcome traits across the age groups tested. We are underpowered to draw reliable  
444 conclusions regarding parental differences in genetic nurturing effects. Yet, our results do not  
445 support the hypothesis that mothers have a stronger nurturing effect on children's  
446 anthropometric traits during their formative years, or the hypothesis that IGE for  
447 anthropometric phenotypes are mainly driven by interactions between the mother and fetus in  
448 the womb for the PRS that we studied. However, the nurturing effects of the mother may still  
449 influence the child via alternative environmental pathways.

450 The use of PRS derived from GWAS, which are inherently influenced by indirect genetic  
451 effects, can have clinical significance, especially when using PRS for disease screening [41,  
452 42]. Screening children using a PRS comprised of DGE would offer a more accurate  
453 classification of children suffering from a possible Mendelian disease, or accumulation of  
454 common risk alleles, than using a PRS consisting of both DGE and IGE [43]. Further, one  
455 potentially promising use of PRS is for personalized medicine, in which individuals with a  
456 high genetic risk of obesity would be prescribed medication for earlier prevention and  
457 treatment, whereas individuals with a low genetic risk of obesity would be prescribed  
458 lifestyle interventions [36]. In such a scenario, it is of importance to construct a PRS without  
459 IGE to distinguish those with a high genetic burden of disease from those with a high  
460 environmental burden. Additionally, IGE can lead to biases within Mendelian randomization,  
461 a type of causal association analysis, leading to spurious associations [44]. IGE can also have  
462 impacts on many downstream analyses of GWAS, including annotation, heritability estimates  
463 and genetic correlations [5].

464

465 Our study presents several limitations. The limited sample size of our cohort warrants  
466 replication of our results in additional (larger) cohorts to improve power and generalizability  
467 of the effect estimates. Although our study encompassed multiple ethnicities, we could not

468 conduct multi-ethnic analyses for BMI and BMD due to the absence of sufficiently large  
469 multi-ethnic GWAS summary statistics available. Additionally, current BMD GWAS have a  
470 limited sample size affecting power, therefore we constructed a BMD PRS derived from an  
471 eBMD GWAS. Despite also being a measure of bone mineralization, this measurement is  
472 only moderately correlated with the BMD obtained from dual-energy x-ray absorptiometry,  
473 which was measured in our cohort [14]. Lastly, our estimates of IGE may be inflated due to  
474 assortative mating, particularly for height. Here, we estimated DGE and IGE using PRSs  
475 derived from GWAS of unrelated individuals [12-14]. Therefore, the IGE could reflect  
476 biases, such as population stratification and assortative mating, present within the original  
477 GWAS. An improved approach would be to use variant effect estimates derived from family-  
478 based GWAS design, in which parental genotypes are controlled for when testing the  
479 association between proband genotypes and traits, which do not suffer from these biases [2,  
480 15, 45]. However, such family-based GWAS designs would require large samples of  
481 genotyped nuclear families, currently not available.

482

483 In conclusion, we observed moderate estimates of genetic nurturing effects on height, BMI  
484 and BMD throughout developmental ages, with no significant differences in indirect genetic  
485 effects for any of the traits across different ages in children. Additionally, no differences were  
486 observed in the contribution of maternal and paternal genetic nurturing effects, providing no  
487 indication of non-genetic transmission of the mother's anthropometric traits on the child's in  
488 pregnancy. These results do not support the hypothesis that mothers have a stronger nurturing  
489 effect on children during their formative years, or the hypothesis that IGE for anthropometric  
490 phenotypes are mainly driven by interactions between the mother and fetus in the womb.

491

492 References

493

- 494 1. Visscher, P.M., W.G. Hill, and N.R. Wray, *Heritability in the genomics era — concepts*  
495 *and misconceptions*. Nature Reviews Genetics, 2008. **9**(4): p. 255-266.
- 496 2. Young, A.I., et al., *Deconstructing the sources of genotype-phenotype associations in*  
497 *humans*. Science, 2019. **365**(6460): p. 1396-1400.
- 498 3. Kong, A., et al., *The nature of nurture: Effects of parental genotypes*. Science, 2018.  
499 **359**(6374): p. 424-428.
- 500 4. Bates, T.C., et al., *The Nature of Nurture: Using a Virtual-Parent Design to Test*  
501 *Parenting Effects on Children's Educational Attainment in Genotyped Families*. Twin  
502 Res Hum Genet, 2018. **21**(2): p. 73-83.
- 503 5. Young, A.I., et al., *Mendelian imputation of parental genotypes improves estimates*  
504 *of direct genetic effects*. Nature Genetics, 2022. **54**(6): p. 897-905.
- 505 6. Kininmonth, A.R., et al., *Associations between the home environment and childhood*  
506 *weight change: a cross-lagged panel analysis*. International Journal of Obesity, 2022.  
507 **46**(9): p. 1678-1685.
- 508 7. Jelenkovic, A., et al., *Genetic and environmental influences on height from infancy to*  
509 *early adulthood: An individual-based pooled analysis of 45 twin cohorts*. Scientific  
510 Reports, 2016. **6**(1): p. 28496.
- 511 8. Perkins, J.M., et al., *Adult height, nutrition, and population health*. Nutr Rev, 2016.  
512 **74**(3): p. 149-65.
- 513 9. Norris, S.A., et al., *Nutrition in adolescent growth and development*. The Lancet,  
514 2022. **399**(10320): p. 172-184.
- 515 10. Rizzoli, R., et al., *Maximizing bone mineral mass gain during growth for the*  
516 *prevention of fractures in the adolescents and the elderly*. Bone, 2010. **46**(2): p. 294-  
517 305.
- 518 11. Heaney, R.P., et al., *Peak Bone Mass*. Osteoporosis International, 2000. **11**(12): p.  
519 985-1009.
- 520 12. Yengo, L., et al., *A Saturated Map of Common Genetic Variants Associated with*  
521 *Human Height from 5.4 Million Individuals of Diverse Ancestries*. bioRxiv, 2022: p.  
522 2022.01.07.475305.
- 523 13. Yengo, L., et al., *Meta-analysis of genome-wide association studies for height and*  
524 *body mass index in ~700000 individuals of European ancestry*. Human Molecular  
525 Genetics, 2018. **27**(20): p. 3641-3649.
- 526 14. Morris, J.A., et al., *An atlas of genetic influences on osteoporosis in humans and mice*.  
527 Nature Genetics, 2019. **51**(2): p. 258-266.
- 528 15. Howe, L.J., et al., *Within-sibship genome-wide association analyses decrease bias in*  
529 *estimates of direct genetic effects*. Nature Genetics, 2022. **54**(5): p. 581-592.
- 530 16. Baud, A., et al., *Indirect Genetic Effects: A Cross-disciplinary Perspective on Empirical*  
531 *Studies*. Journal of Heredity, 2021. **113**(1): p. 1-15.
- 532 17. Kooijman, M.N., et al., *The Generation R Study: design and cohort update 2017*.  
533 European Journal of Epidemiology, 2016. **31**(12): p. 1243-1264.
- 534 18. Cole, T.J., *The LMS method for constructing normalized growth standards*. Eur J Clin  
535 Nutr, 1990. **44**(1): p. 45-60.
- 536 19. Heppe, D.H.M., et al., *Maternal first-trimester diet and childhood bone mass: the*  
537 *Generation R Study1234*. The American Journal of Clinical Nutrition, 2013. **98**(1): p.  
538 224-232.



- 539 20. Lewiecki, E.M., et al., *Special report on the 2007 adult and pediatric Position*  
540 *Development Conferences of the International Society for Clinical Densitometry.*  
541 *Osteoporos Int*, 2008. **19**(10): p. 1369-78.
- 542 21. Medina-Gomez, C., et al., *Challenges in conducting genome-wide association studies*  
543 *in highly admixed multi-ethnic populations: the Generation R Study.* *European*  
544 *Journal of Epidemiology*, 2015. **30**(4): p. 317-330.
- 545 22. Goldstein, J.I., et al., *zCall: a rare variant caller for array-based genotyping: genetics*  
546 *and population analysis.* *Bioinformatics*, 2012. **28**(19): p. 2543-5.
- 547 23. Delaneau, O. and J. Marchini, *Integrating sequence and array data to create an*  
548 *improved 1000 Genomes Project haplotype reference panel.* *Nat Commun*, 2014. **5**:  
549 p. 3934.
- 550 24. Das, S., et al., *Next-generation genotype imputation service and methods.* *Nat Genet*,  
551 2016. **48**(10): p. 1284-1287.
- 552 25. Manichaikul, A., et al., *Robust relationship inference in genome-wide association*  
553 *studies.* *Bioinformatics*, 2010. **26**(22): p. 2867-2873.
- 554 26. Delaneau, O., J.-F. Zagury, and J. Marchini, *Improved whole-chromosome phasing for*  
555 *disease and population genetic studies.* *Nature Methods*, 2013. **10**(1): p. 5-6.
- 556 27. Auton, A., et al., *A global reference for human genetic variation.* *Nature*, 2015.  
557 **526**(7571): p. 68-74.
- 558 28. Privé, F., J. Arbel, and B.J. Vilhjálmsson, *LDpred2: better, faster, stronger.*  
559 *Bioinformatics*, 2021. **36**(22-23): p. 5424-5431.
- 560 29. Florian, P., et al., *Inferring disease architecture and predictive ability with LDpred2-*  
561 *auto.* *bioRxiv*, 2023: p. 2022.10.10.511629.
- 562 30. Choi, S.W., T.S.-H. Mak, and P.F. O'Reilly, *Tutorial: a guide to performing polygenic*  
563 *risk score analyses.* *Nature Protocols*, 2020. **15**(9): p. 2759-2772.
- 564 31. Selzam, S., et al., *Comparing Within- and Between-Family Polygenic Score Prediction.*  
565 *Am J Hum Genet*, 2019. **105**(2): p. 351-363.
- 566 32. Tubbs, J.D., et al., *The Genes We Inherit and Those We Don't: Maternal Genetic*  
567 *Nurture and Child BMI Trajectories.* *Behavior Genetics*, 2020. **50**(5): p. 310-319.
- 568 33. Nan, C., et al., *Heritability of body mass index in pre-adolescence, young adulthood*  
569 *and late adulthood.* *European Journal of Epidemiology*, 2012. **27**(4): p. 247-253.
- 570 34. Couto Alves, A., et al., *GWAS on longitudinal growth traits reveals different genetic*  
571 *factors influencing infant, child, and adult BMI.* *Science Advances*, 2019. **5**(9): p.  
572 eaaw3095.
- 573 35. Silventoinen, K., et al., *Changing genetic architecture of body mass index from*  
574 *infancy to early adulthood: an individual based pooled analysis of 25 twin cohorts.*  
575 *International Journal of Obesity*, 2022. **46**(10): p. 1901-1909.
- 576 36. Khera, A.V., et al., *Polygenic Prediction of Weight and Obesity Trajectories from Birth*  
577 *to Adulthood.* *Cell*, 2019. **177**(3): p. 587-596.e9.
- 578 37. Okbay, A., et al., *Polygenic prediction of educational attainment within and between*  
579 *families from genome-wide association analyses in 3 million individuals.* *Nature*  
580 *Genetics*, 2022. **54**(4): p. 437-449.
- 581 38. Weaver, C.M., et al., *The National Osteoporosis Foundation's position statement on*  
582 *peak bone mass development and lifestyle factors: a systematic review and*  
583 *implementation recommendations.* *Osteoporos Int*, 2016. **27**(4): p. 1281-1386.
- 584 39. Yengo, L., et al., *Imprint of assortative mating on the human genome.* *Nature Human*  
585 *Behaviour*, 2018. **2**(12): p. 948-954.

- 586 40. Young, A.S., *Estimation of indirect genetic effects and heritability under assortative*  
587 *mating*. bioRxiv, 2023: p. 2023.07.10.548458.
- 588 41. Manousaki, D., et al., *Increased Burden of Common Risk Alleles in Children With a*  
589 *Significant Fracture History*. Journal of Bone and Mineral Research, 2020. **35**(5): p.  
590 875-882.
- 591 42. Ghatan, S., et al., *The Polygenic and Monogenic Basis of Paediatric Fractures*. Current  
592 Osteoporosis Reports, 2021. **19**(5): p. 481-493.
- 593 43. Mosley, J.D., et al., *Clinical associations with a polygenic predisposition to benign*  
594 *lower white blood cell counts*. Nature Communications, 2024. **15**(1): p. 3384.
- 595 44. Brumpton, B., et al., *Avoiding dynastic, assortative mating, and population*  
596 *stratification biases in Mendelian randomization through within-family analyses*.  
597 Nature Communications, 2020. **11**(1): p. 3519.
- 598 45. Trejo, S. and B.W. Domingue, *Genetic nature or genetic nurture? Introducing social*  
599 *genetic parameters to quantify bias in polygenic score analyses*. Biodemography Soc  
600 Biol, 2018. **64**(3-4): p. 187-215.

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616 List of abbreviations:

617 Genome-wide association study (GWAS), linkage disequilibrium (LD), polygenic risk score  
618 (PRS), body mass index (BMI), Bone mineral density (BMD), Direct genetic effects (DGE),  
619 Indirect genetic effects (IGE).

620

621 Declarations

622 Not applicable.

623

624 Ethics approval and consent to participate

625 The Generation R Study was approved by their respective institutional ethics review

626 committees and all participants provided written informed consent.

627

628 Consent for publication

629 Not applicable.

630

631 Data and code availability

632 Height and BMI summary statistics can be obtained from GIANT consortium website

633 ([https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_fil](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files)

634 [es](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files)). Estimated bone mineral density summary statistic data can be obtained via the GEFOS

635 consortium website (<http://www.gefos.org/>). All data generated or analyzed during this study

636 are included in this published article and its supplementary information files.

637

638 Competing interests

639 Not applicable.

640

641 Funding

642 This project has received funding from the European Union’s Horizon 2020 research and  
643 innovation program under the MARIE SKŁODOWSKA-CURIE grant agreement no.  
644 860898. The general design of the Generation R Study is made possible by financial support  
645 from Erasmus MC, Erasmus University Rotterdam, the Netherlands Organization for Health  
646 Research and Development and the Ministry of Health, Welfare and Sport. This project  
647 received funding from the European Research Council (ERC) under the European Union’s  
648 Horizon 2020 research and innovation programme attributed to JBP (IRISK, grant agreement  
649 No. 863981). This project received funding from the European Union’s Horizon 2020  
650 research and innovation program (874739, LongITools).

651

652 Authors' contributions

653 S.Ghatan, C.Medina-Gomez, and F.Rivadeneira designed the study. S.Ghatan performed the  
654 analysis. S.Ghatan, F.Rivadeneira, and C.Medina-Gomez drafted the manuscript. All authors  
655 contributed to the interpretation of data and critical revision of the manuscript. All authors  
656 read and approved the final version of the manuscript.

657

658 Acknowledgments

659 We express our gratitude to the study participants, whose participation made this work  
660 possible, and to the numerous colleagues who contributed to the collection, phenotypic  
661 characterization of clinical samples, as well as genotyping and analysis of GWAS data. We  
662 gratefully acknowledge the contribution of children and parents, general practitioners,  
663 hospitals, midwives and pharmacies in Rotterdam. The Generation R Study is conducted by  
664 Erasmus MC, University Medical Center Rotterdam in close collaboration with the School of  
665 Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal  
666 Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam

667 and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC),  
668 Rotterdam. The study protocol was approved by the Medical Ethical Committee of the  
669 Erasmus Medical Centre, Rotterdam. Written informed consent was obtained for all  
670 participants. The generation and management of GWAS genotype data for the Generation R  
671 Study was done at the Human Genomics Facility, HuGe-F, housed within the Laboratory for  
672 Population Genomics of the Department of Internal Medicine at Erasmus MC. Genetic  
673 Laboratory of the Department of Internal Medicine, Erasmus MC, The Netherlands. We  
674 thank Pascal Arp, Karol Estrada, Mila Jhamai, Manoushka Ganesh, Gaby van Dijk, Marijn  
675 Verkerk, Lizbeth Herrera, Marjolein Peters, and Dr. Linda Broer for their help in creating,  
676 managing and QC the GWAS database.

