Genetic influences on educational outcomes during and after the Soviet era: Revisiting evidence from Estonia

Ivan A. Kuznetsov[1,*], Margherita Malanchini[2,3], Oliver Pain[3,4], Jonathan Coleman[3,4], Philip S. Dale[5], Jean-Baptiste Pingault[6,3], Kathleen Rastle[7], Estonian Biobank Research Team[1], Peeter Hõrak[8], Robert Plomin[3], Andres Metspalu[1], Kelli Lehto[1], Vasili Pankratov[1,9,#], Kaili Rimfeld[3,7,#,*]

[#]*joint last authors*

[1] *Institute of Genomics, University of Tartu, Tartu 51010, Estonia*

[2] *School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK*

[3] *Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK*

[4] *NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, London, UK*

[5] *Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, NM, USA*

[6] *Department of Clinical, Educational and Health Psychology, University College London, London, UK*

[7] *Department of Psychology, Royal Holloway, University of London, Egham, UK*

[8] *Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia*

[9] *Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark*

*Corresponding authors: Ivan A. Kuznetsov, Kaili Rimfeld
Email: ivan.kuznetsov@ut.ee (IK), kaili.rimfeld@kcl.ac.uk (KR)

Estonian Biobank Research Team: Andres Metspalu, Lili Milani, Tõnu Esko, Reedik Mägi, Mari Nelis and Georgi Hudjashov

Key words: Gene-environment interaction; socioeconomic outcomes; changes in heritability; participation bias; societal changes; educational attainment

Abstract

The origins of individual differences in socioeconomic outcomes, including educational attainment and occupational status, reflect a combination of genetic and environmental factors whose relative contributions may shift across historical and societal contexts. Previous research suggested a doubling of genetic influence on these traits in independent Estonia compared to the Soviet era. Using the Estonian Biobank, now tenfold larger compared to the original study, we aimed to replicate and extend these findings. We found only limited evidence for the increase in genetic influence on social outcomes in independent Estonia with results differing between analytical setups. Most importantly, we show that genetic contribution to socioeconomic outcomes, as well as to height and body mass index, varies within each historical cohort depending on birth year and recruitment strategy. This highlights the limitations of generalising from the study cohort due to various participation biases and potential heterogeneity of social environment in both historical periods.

Introduction

Educational and occupational success, collectively known as socioeconomic status (SES), are linked to many significant life outcomes, including physical and mental health, life satisfaction, happiness, and longevity[1–4]. Individual differences in SES, considered as a composite or via its components (occupational status, educational attainment, and income), were long assumed to be shaped primarily by environmental factors; however, twin and family-based research has demonstrated that both genetic and environmental influences contribute significantly to individual variation[5–8]. Genetic influence on SES is probabilistic rather than deterministic, and the observed magnitude can vary with shifts in historical and social context[5,9–11]. In addition, the relative contribution of genetic factors, commonly expressed as heritability (the proportion of individual differences attributable to inherited DNA differences), varies across approaches to estimating it, yielding very different results.

Heritability estimates are not fixed and are context-dependent. Regardless of the method used to estimate heritability, they can vary across populations and environments[5,9–11]. Hence, heritability may vary between cohorts or groups, partly because these estimates can differ in different environments and may change with significant environmental shifts, including major social policy changes[5,9]. Heritability for educational outcomes tends to be higher when environmental influences are minimised. With less environmental variation, a larger proportion of individual differences is attributed to genetic factors[9,12,13].

We previously showed a massive shift in the aetiology of SES in Estonia following an abrupt social change following the collapse of the Soviet Union. Genetic differences explained up to twice as much variance in SES (occupational status, educational attainment) among individuals who were 15 or younger when Estonia regained independence, that is, those who spent most of their adult life in independent Estonia, compared to those who were older at the time and spent more of their adult life under Soviet rule[14]. These results suggested that regaining independence led to an increase in meritocratic selection of individuals in occupation and education, thereby reducing the impact of environmental differences and increasing the importance of genetic differences in explaining individual outcomes in SES compared to the Soviet era, when environmental variation had a larger effect. We also reported suggestive sex differences, with the shift much larger in females than in males, but we were cautious in our conclusions due to limited power in the sex-stratified analyses. Recent evidence from Germany examining cohorts around reunification reported that the genetic influence on educational attainment was comparable in East and West Germany before reunification, but increased in East Germany afterwards, providing further evidence for changes in heritability for SES outcomes with shifts in historical context[10].

There is converging evidence for the substantial heritability of SES. Approaches that rely on comparing the similarity of identical and non-identical twin pairs result in heritability estimates for socioeconomic outcomes as high as 50%[5,8,15]. Alternatively, heritability can be estimated by comparing the phenotypic, or trait, similarity between individuals and their genetic similarity, based on hundreds of thousands of common genetic variants, or SNPs (single-nucleotide polymorphisms). These SNP-based heritability estimates of SES outcomes average around 20%[7,16–18]. Thus, SNP heritability for educational and occupational outcomes is less than half that

of family-based heritability estimates[19]. This discrepancy likely arises because SNP heritability is estimated from genotyped or imputed common genetic variants and therefore captures only the additive effects of common SNPs represented on genotyping arrays[20]. Alternatively, it has been argued that twin heritability overestimates true heritability[21], which explains (or rather, explains away) the missing heritability (the gap between family and SNP-based heritability) due to unaccounted confounding factors such as environmental confounding, population stratification, or assortative mating[21]. Many of the same biases, including population stratification and assortative mating, also affect SNP heritability estimates.

A third method for estimating the genetic influence on complex traits is to utilise genome-wide polygenic scores (PGS). PGS utilise summary statistics from genome-wide association studies (GWAS) aimed at identifying specific genetic markers associated with a phenotype of interest by aggregating thousands of SNP associations across the genome to predict the trait of interest[22]. These PGS can serve as predictors of socioeconomic variables[12]. For example, the latest PGS for educational attainment have been shown to explain 12-16% of the trait variance in a sample of European genetic ancestry[23]. Estimates based on PGS predictive accuracy can vary substantially depending on the details of the GWAS and on genetic and environmental similarities between the GWAS cohort and the cohort used to evaluate the PGS[24–28]. As with heritability estimates, the predictive accuracy of PGS may not only reflect direct genetic effects but also factors such as population stratification, assortative mating, and genetic nurture or dynastic effects[29–31].

To determine whether cohort differences in estimated genetic influence on SES reflect genuine historical change or artefacts of measurement and sampling, this study re-examines the questions raised in the previous report. The Estonian Biobank sample has increased from 20,000 genotyped individuals, who comprised about 2% of the adult population in Estonia, to 210,000 participants, or 20% of the adult population[32,33]. This growth provides three advantages. First, the earlier study had limited power, particularly for sex-stratified analyses, which is no longer the case. Second, the recruitment procedure changed after the earlier study, enabling direct tests of participation bias and within-population generalisability. Third, methods and discovery resources have improved; there are now more effective methods available to construct PGS and GWAS discovery samples for relevant traits that are larger and more powerful, thereby increasing the precision of estimates.

In this preregistered study (https://osf.io/ug5jp/), we aimed to analyse SNP-heritability and polygenic-score prediction for educational attainment, occupational status, height, and BMI across Soviet and post-Soviet birth cohorts; assess whether educational-attainment results hold for both quantitative and binary outcomes; consider the possibility of gradual environmental change; and implement sensitivity analyses that address outcome distributions, participation bias, and recruitment-wave differences. This study aimed to verify whether the previously reported shift in genetic influence on socioeconomic outcomes following Estonia's independence could be reproduced by re-analysing the same cohort using updated methods and the same research question. We also examined whether these findings replicate — that is, whether they are observed in a larger sample, remain consistent across traits, are robust to differences in outcome definitions, and are unaffected by sampling and analytical approaches.

Results

Data and analysis overview

The Estonian Biobank (EstBB) cohort comprises a volunteer-based sample of the Estonian adult population (aged 18 years or older)[32,33]. The EstBB participants were recruited between 2000 and 2021 through two major data collection waves (Fig. 1a). During the first wave (2000-2016), approximately 52,000 individuals were recruited. A subset of these, genotyped by 2017, formed the sample used in the study by Rimfeld et al.[14]. Approximately 158,000 participants were recruited after 2016. These two periods differ substantially in their recruitment strategies. During the first wave, general practitioners informed their patients about the opportunity to join the EstBB, typically during routine health visits. Additionally, during the pilot period of wave 1, volunteers among university staff and students were actively recruited, likely leading to an overrepresentation of individuals with a university degree in the sample used by Rimfeld et al.[14] (Supplementary Table 1). In the second wave, participants were primarily recruited through a broad media campaign, including online and television advertisements**.** Neither wave is entirely representative with respect to educational attainment (EA), with the second wave in particular showing a notable overrepresentation of individuals with tertiary education (Supplementary Table 1).

Participants were divided into two birth cohorts, designated as the "Soviet era" and "post-Soviet era," to maintain consistency with the previous study. We included individuals born before 1976 (older than 15 years old in 1991, when the Soviet Union collapsed and Estonia regained independence) as the "Soviet era", while younger individuals were included in the "post-Soviet era". We repeated our analyses using an alternative grouping method, with 1981 as the birth-year cutoff (i.e., older than 10 years at the time of the collapse of the Soviet Union), again following the Rimfeld et al. approach[14]. The sample size of unrelated participants in each group and descriptive statistics for age, sex, and SES-related variables are presented in Supplementary Tables 2 and 3, respectively.

In this study, we mainly focus on educational attainment (EA) and occupational status (OS). To put the results for these phenotypes into perspective, we performed the same analyses for two anthropometric traits, height and body mass index (BMI) (Fig. 1b). The sample sizes for the traits range from 87,458 to 115,218 individuals, all unrelated within the corresponding era and participation wave.

We perform the following three sets of analyses (Fig. 1c) utilising all available samples. First, we analyse SNP heritability (h^2_{SNP}) in subcohorts divided by era or era and wave, and by decade of birth. Second, we compare the predictive accuracy (R^2) of a polygenic score (PGS) across subcohorts by era and wave, and by half-decade of birth, and analyse the effects of phenotype distribution and participation bias on PGS performance. Third, we directly test for PGS×Era statistical interactions in a set of models with varying parameters.

In addition, we perform a reproduction of the original study by Rimfeld et al.[14] using the same sample of individuals (Supplementary Note 1). In general, we see the same trend as in the original study: point estimates of both h^2_{SNP} and PGS R^2 for EA are always higher in the post-Soviet era compared to the Soviet one across all variations of the analysis, including different phenotype specification, different birth year cutoff between Soviet and post-Soviet and different methods used to estimate heritability. However, these differences are statistically significant (without multiple testing correction) only in a single specific setup (PGS R^2 for years of education with 1981 as the cutoff birth year).

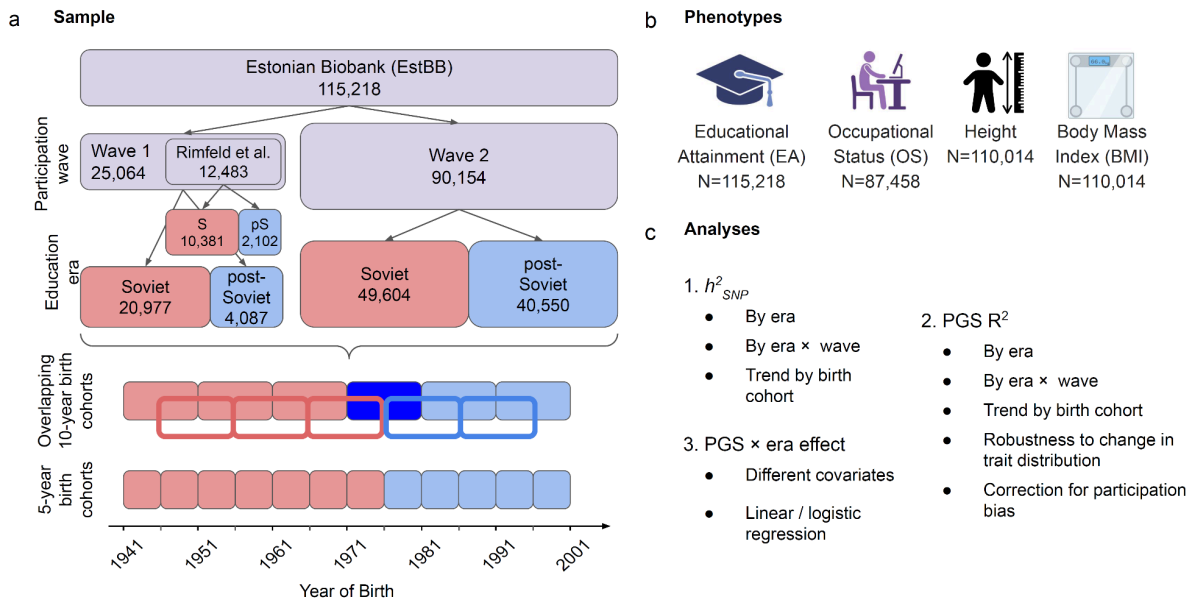

Fig. 1. Study overview. (**a**) Data overview showing the composition of the subcohorts used in the analyses. The sample size, after excluding 2nd-degree and closer relatives, is presented for the educational attainment (EA) phenotype in each cohort, defined by era and participation wave. The EstBB comprises wave 1 and wave 2, which differ in both recruitment period and strategy. The dataset analysed by Rimfeld et al.[14] represents a subset of wave 1. The Soviet (S, peach) and post-Soviet (pS, light blue) eras were defined using a 15-year cutoff at the time of the Soviet Union's collapse. In some analyses, 10- or 5-year birth cohorts were used; adjacent 10-year cohorts overlapped by 5 years. Colours correspond to the era encompassing a given birth cohort, whereas a cohort split evenly across eras is shown in dark blue. (**b**) The phenotypes examined in the study, along with the sample size for each, are reported for unrelated individuals (excluding 2nd-degree and closer relatives) within the corresponding era and participation wave. (**c**) Overview of the analyses performed in this study. Numbers indicate the order in which the analyses are presented in the text.

Analysis of heritability

The results of Rimfeld et al.[14] suggested that SNP heritability for EA and occupational status (OS) might be higher in the post-Soviet era than in the Soviet era, although these differences were not statistically significant. We estimated SNP heritability (h^2_{SNP}) separately for the Soviet and post-Soviet eras using the restricted maximum likelihood (REML) method implemented in LDAK[19]. Here and further in the main analysis that follows, EA was measured as years of education (i.e., a continuous measure corresponding to completed years of schooling), and OS on a discrete numeric scale of occupational status categories, with higher values indicating occupations of higher socioeconomic status. EA heritability in the post-Soviet era was significantly higher than in the Soviet era ($p = 2.7\times10^{-3}$), while there was no significant difference in OS heritability ($p = 0.26$) (Fig. 2a,b and Supplementary Table 4). The difference in EA h^2_{SNP} remained significant when h^2_{SNP} was estimated using an alternative model, implemented in GCTA-GREML ($p = 7.5\times10^{-3}$) (Supplementary Fig. 1a and Supplementary Table 5). The results remained qualitatively unchanged with the alternative era-based cutoff (Supplementary Fig. 2a,c and Supplementary Table 6).

EA is commonly analysed as a continuous trait[7,23], and we followed the same approach in the previous analysis. However, despite being transformed into years of education, it retains features characteristic of an ordinal trait. For example, it cannot exceed the value corresponding to a PhD degree. When the mean EA changes, the shape of its distribution also changes, affecting the scale on which h^2_{SNP} is estimated. Thus, h^2_{SNP} on the observed scale may differ between groups even when it does not differ on the liability scale. To address this problem, we repeated the analysis, treating EA as a binary trait corresponding to the presence or absence of a university degree. We estimated its heritability and transformed it to the liability scale (Fig. 2c, Supplementary Fig. 1e, and Supplementary Table 5). No significant difference in heritability between Soviet and post-Soviet eras was detected ($p = 0.19$, $p=0.22$ for 10-year cutoff). We note, however, that this analysis has lower statistical power due to information loss from collapsing all EA categories into two. The point estimate of h^2_{SNP} remains higher in the post-Soviet era than in the Soviet era. At the same time, the difference in EA heritability between the Soviet and post-Soviet eras decreased from 0.036 to 0.024 after the transition to the binary trait liability scale, despite an absolute increase in heritability.

Although the difference in EA heritability between the Soviet and post-Soviet eras might be a scale artefact, it could still reflect a true effect. Considering the collapse of the Soviet Union as an event that substantially affected the genetic influence on EA suggests pronounced changes in the temporal trend of EA heritability across the Soviet and post-Soviet eras. To examine whether this historical transition corresponds to a notable shift in the heritability of EA, we estimated EA h^2_{SNP} within overlapping decade-long birth subcohorts (Fig. 1a). Each bin spanned ten years to ensure sufficient sample sizes for reliable estimates, with a five-year step between bins to provide higher temporal resolution. The EA heritability fluctuates over time, demonstrating no clear or consistent trend change between the Soviet and post-Soviet eras (Fig. 2d). This pattern was reproduced with the heritability estimates from GCTA-GREML and for the binary EA on the liability scale (Supplementary Fig. 3). These findings indicate that the difference in EA h^2_{SNP} between the eras on the observed scale does not reflect a discrete shift associated with the collapse of the Soviet Union, but rather represents ongoing temporal fluctuations driven by unobserved factors.

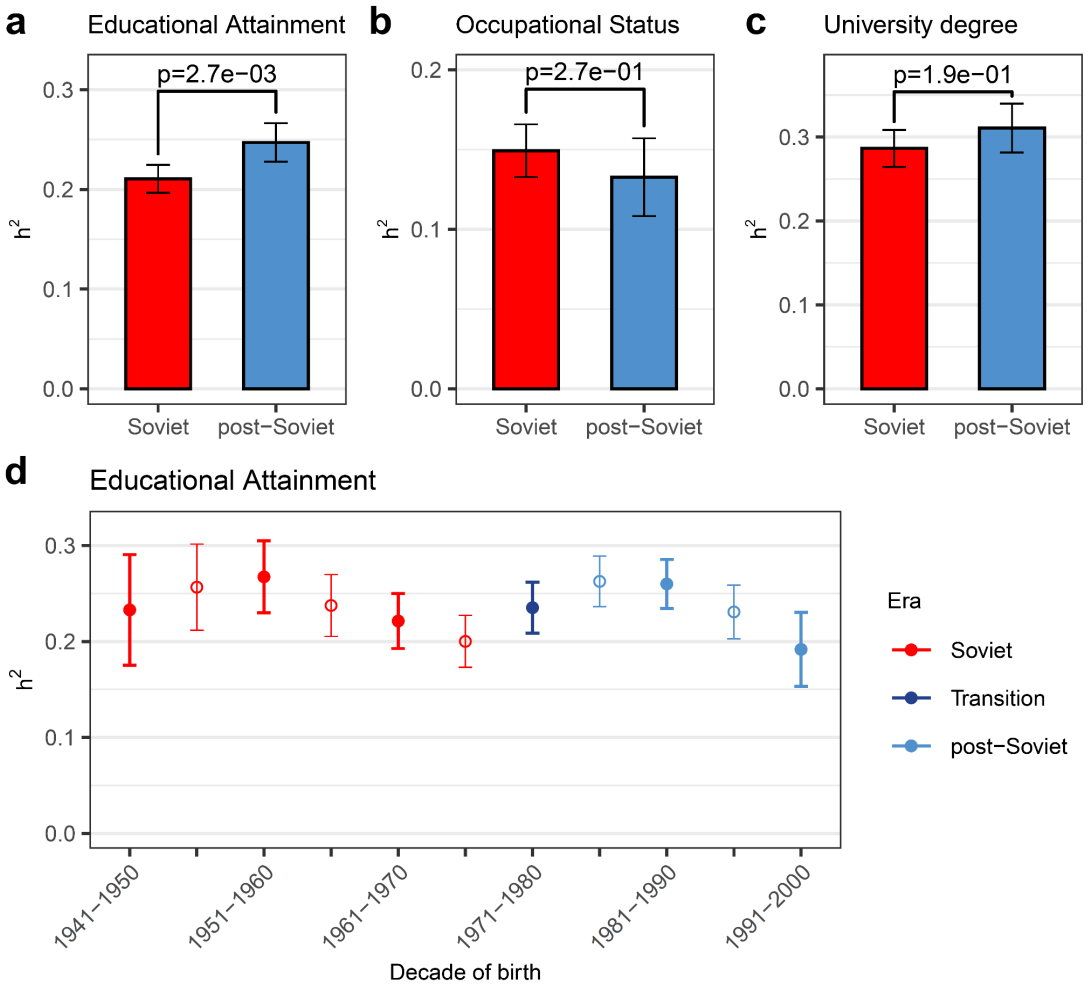

Fig. 2. Trait heritability (h^2_{SNP}) for educational attainment (EA) and occupational status (OS). (**a**) EA and (**b**) OS heritability is compared across the Soviet and post-Soviet eras on the continuous scale. (**c**) Heritability of EA as a binary trait (University Degree acquired or not), estimated on the liability scale. (**d**) Heritability of EA on the continuous scale in decade-long birth subcohorts with 5-year overlap between adjacent subcohorts. Thick error bars and filled circles or thin error bars and empty circles represent non-overlapping subcohorts. Those subcohorts belonging to the Soviet era are coloured in red, and the post-Soviet era in light blue. A transition group comprising individuals from both the Soviet and post-Soviet eras is shaded in dark blue. Error bars correspond to 95% CI.

To contextualise these results, we compared the heritability of two anthropometric traits, height and body mass index (BMI), between the Soviet and post-Soviet eras (Fig. 3a,c). Both traits exhibited significantly higher heritability in the post-Soviet era, indicating that such a difference is not specific to educational attainment. Heritability estimated in decade-long birth subcohorts again showed no association with the era transition (Fig. 3b,d). Moreover, the temporal h^2_{SNP} patterns were distinct from one another and from that of educational attainment, suggesting that the factors underlying temporal fluctuations in heritability are likely trait-specific.

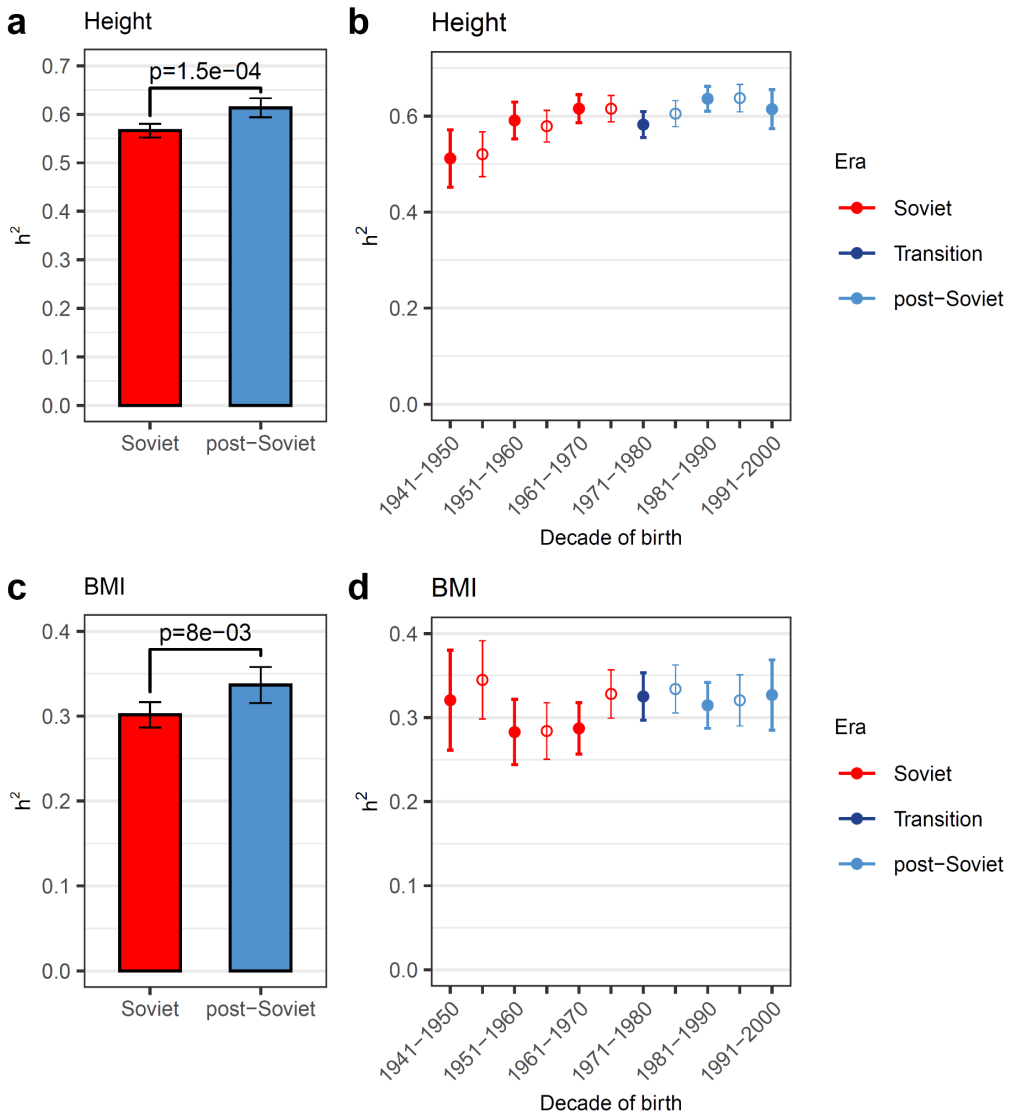

Fig. 3. Trait heritability (h^2_{SNP}) for height and Body Mass Index (BMI). (**a**) Height and (**c**) BMI heritability are compared across the Soviet and post-Soviet eras on the continuous scale. (**b**) Height and (**d**) BMI heritability in decade-long birth subcohorts with a 5-year overlap between adjacent subcohorts. Thick error bars and filled circles or thin error bars and empty circles represent non-overlapping subcohorts. Subcohorts from the Soviet era are coloured red, and those from the post-Soviet era are in light blue. A transition group comprising individuals from both the Soviet and post-Soviet eras is shaded in dark blue. Error bars correspond to 95% CI.

Differential participation bias can affect representativeness of subsamples in different ways, potentially leading to differences in h^2_{SNP} estimates between samples without actual differences between the underlying population cohorts[34]. For heritability comparisons across subcohorts to be valid, the results must generalise to the underlying population and not be driven by participation bias. Accordingly, two criteria should hold. First, comparisons within participation waves should not contradict patterns observed in the full sample. Second, no significant differences are expected between participation waves within the same era. In practice, significant cross-era differences in h^2_{SNP} for EA, height, and BMI were observed in wave 2, with higher heritability in the post-Soviet era, consistent with the main results (Fig. 4, Supplementary Table 4). The large standard errors in wave 1, due to its limited sample size, make it unlikely to detect significant h^2_{SNP} differences. The comparison of h^2_{SNP} estimates between waves within the same era yielded a single significant result: height heritability in the Soviet era differed significantly between wave 1 and wave 2 ($p=3.3\times10^{-5}$, $p_{Bonf}=2.6\times10^{-4}$). Thus, when it comes to EA, we do not see any differences in the behaviour of wave 1 and wave 2. Still, significant differences in h^2_{SNP} between wave 1 and wave 2 for height suggest that the two waves cannot be considered two random samples from the same underlying population, which impairs generalisability.

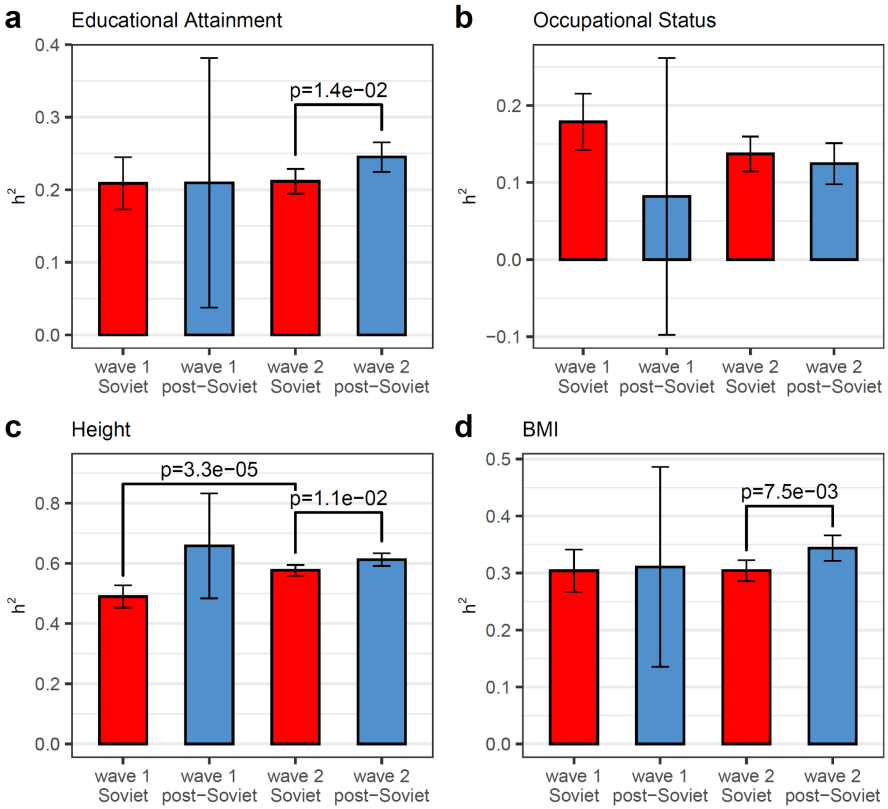

Fig. 4. Trait heritability in the Soviet and post-Soviet eras by participation wave. Trait heritability (h^2_{SNP}) is shown in groups divided by era of basic school education (left panels) and wave of biobank enrolment for (**a**) educational attainment (EA), (**b**) occupational status (OS), (**c**) height, and (**d**) body mass index (BMI). Pairwise comparisons of h^2_{SNP} were conducted either between waves within each era or between eras within each wave. Only p-values below the significance threshold of 0.05 (uncorrected for multiple testing) are shown. Error bars correspond to 95% CI.

Predictive accuracy of polygenic scores

Rimfeld et al.[14] reported that the polygenic score for educational attainment (PGS_{EA}) explains more variance in educational attainment and occupational status in the post-Soviet era than in the Soviet era. In this study, we evaluated and compared across eras the predictive power of a

polygenic score constructed from the most recent and well-powered EA GWAS summary statistics (EA4[23]), excluding Estonian Biobank and the 23andMe cohorts excluded from the meta-analysis. The SBayesR method[35] was used to re-estimate SNP effects while accounting for linkage disequilibrium (LD) structure.

Differences in the genetic architecture of a trait between the discovery population used for GWAS-based PGS construction and the target population in which the PGS is applied can reduce predictive accuracy. Therefore, we first verified that the genetic architecture of EA in the GWAS sample was compatible with that in the EstBB. The genetic correlation between EA in the EA4 meta-analysis sample (excluding the EstBB and 23andMe cohorts) and in the EstBB, estimated using LD score regression, was not significantly different from one (r_g=0.98, SE=0.03, p=0.53). This high genetic correlation indicates that EA in the EstBB shares a similar genetic architecture with the EA4 sample, supporting the validity of applying EA4-based polygenic scores in the Estonian population.

We used PGS_{EA} as a predictor of educational attainment and occupational status in two models. In the first model, we estimated the incremental R^2, defined as the proportion of variance in the trait explained by the PGS in addition to that explained by covariates (sex, age, sex×age, age^2, and 40 genetic principal components). In the second model, the traits were pre-adjusted for these covariates by regressing them out, and R^2 was then calculated from a univariate linear regression with the PGS as the sole predictor. We then compared the R^2 estimates obtained for the Soviet and post-Soviet eras, both across the full sample and separately for each participation wave.

For both EA and OS, R^2 values obtained from either model did not differ significantly between the Soviet and post-Soviet eras when the full sample was used (Fig. 5a,d, Supplementary Fig. 4a,d, and Supplementary Tables 7 and 8). When the comparison was performed separately for the two waves, no significant differences were observed in the incremental R^2 for EA (Fig. 5b and Supplementary Table 7). Differences in R^2 for the pre-adjusted EA were significant in both wave 1 (p = 0.032) and wave 2 (p = 0.04) before correction for multiple testing (Supplementary Fig. 4b and Supplementary Table 8). However, the directions of these differences were opposite across waves: in wave 2, the PGS explained more variance in the Soviet than in the post-Soviet eras, the opposite of what was reported in Rimfeld et al. For OS, both incremental R^2 and R^2 for

the pre-adjusted trait were significantly higher in the post-Soviet era than in the Soviet era in wave 1, but not in wave 2 (Fig. 5e and Supplementary Fig. 4d).

As with heritability analysis, a comparison of PGS R^2 is valid and generalisable only if there are no significant differences between participation waves within the same era. The PGS R^2 results did not meet this criterion. Incremental R^2 and R^2 for the pre-adjusted trait differed significantly between waves for EA in the Soviet era and for OS in the post-Soviet era (Fig. 5b,e and Supplementary Fig. 4b,d). Taken together with the opposite directions of R^2 change for EA, these results suggest that differences in predictive accuracy are primarily driven by sample-related characteristics rather than historical context or birth cohort. Consequently, comparisons of PGS predictive accuracy between the Soviet and post-Soviet eras should be interpreted with caution unless sample-related biases are explicitly controlled for.

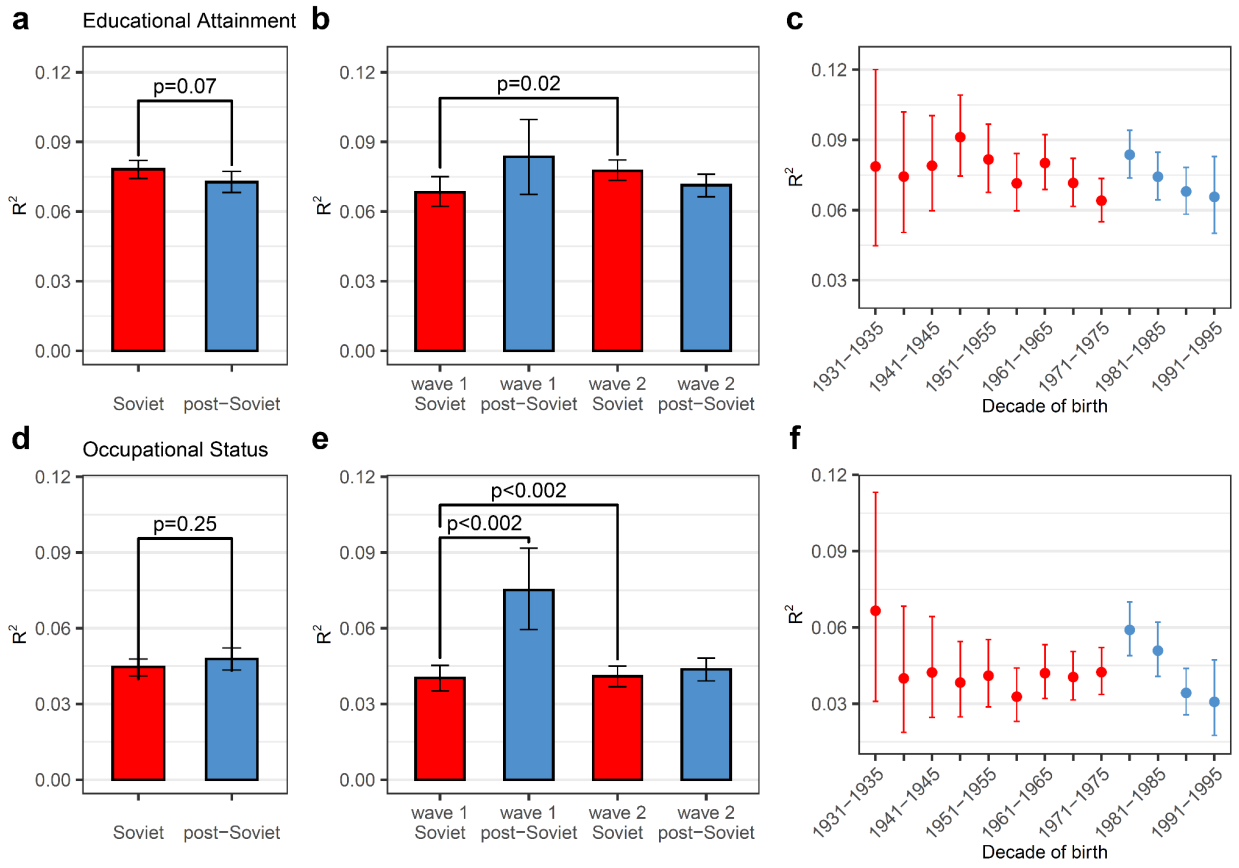

Fig. 5. Trait variance explained by PGS (incremental R^2) in the Soviet and post-Soviet eras. R^2 is shown in groups divided by era (in the left panels), by era and wave of biobank enrolment (in the middle panels), or by half-decade of birth (in the right panels) for educational attainment (**a, b** and **c**, respectively), and occupational status (**d, e** and **f**, respectively). Pairwise comparisons of R^2 were conducted between eras (**a**, **d**), and between waves within each era or between eras within each wave (**b**, **e**). Those cohorts from the Soviet era are coloured red, and those from the post-Soviet era are coloured light blue. Only p-values below the significance threshold of 0.05 (uncorrected for multiple testing) are shown for panels **b** and **e**. Error bars correspond to 95% CI.

To examine how the variance explained by PGS changes over time, we divided the sample into half-decade birth subcohorts and calculated incremental R^2 for each subcohort. For EA and OS, R^2 in the earliest post-Soviet birth cohort demonstrates an upward shift relative to the latest Soviet cohort (Fig. 5c,f). However, this increase was primarily driven by participants from wave 1 (Supplementary Fig. 6a,b) and was followed by a decline in R^2 in later cohorts. Results for height and BMI at both the era and year-of-birth levels are presented in Supplementary Fig. 4-7 and Supplementary Tables 7 and 8.

Stability of polygenic score predictive power

Despite the large sample size, we were unable to replicate the difference in EA R^2 between the Soviet and post-Soviet eras reported by Rimfeld et al. However, we observed that R^2 can vary considerably even between cohorts from the same era when recruitment procedures differ. One plausible explanation is that R^2 is affected by the distribution of the dependent variable[36]. Indeed, the cohorts, whether grouped by participation wave or by era, show distinct EA distributions (Supplementary Table 1).

To examine how such distributional differences influence R^2, we performed a resampling analysis. For each of the four groups defined by era (Soviet, post-Soviet) and participation wave (wave 1, wave 2), we repeatedly drew random subsamples with replacement from that group

("source group"). Individuals were sampled in such a way that the EA distribution in each subsample matched that of another group ("target group") (Fig. 6). This procedure was also repeated while matching jointly on EA and sex (Supplementary Fig. 8).

We then recalculated R^2 for each resampled dataset. In most cases, the resulting R^2 values shifted from the source group's original R^2 toward the target group's R^2, with the mode of the resampled R^2 distribution lying between them. The extent of the shift ranged from negligible to nearly complete. These results demonstrate that R^2 is sensitive to differences in the distribution of the trait, which may arise either from participation bias or from genuine population differences.

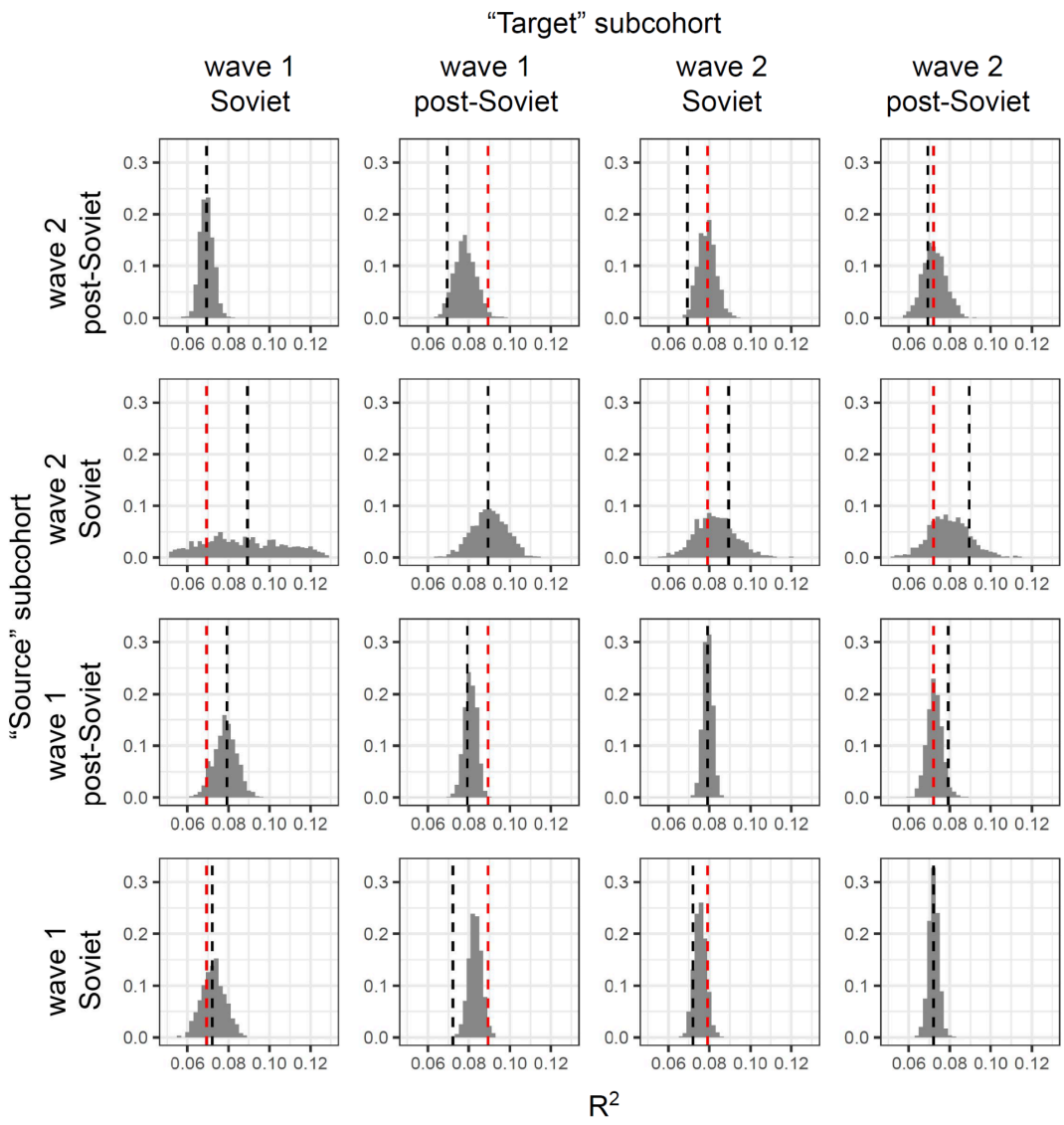

Fig. 6. Distribution of EA variance explained (R^2) by PGS_{EA} in subsamples generated to match the EA distribution of a target subcohort. Subsamples were drawn with replacement from a "source" subcohort defined by recruitment wave and era, either matching the EA distribution. Rows indicate the source subcohort from which individuals were sampled, columns indicate the target subcohort whose EA distribution was used for matching. For each source–target pair, 1,000 resampled subsamples were generated, and R^2 was calculated as the squared Pearson correlation between the pre-adjusted EA residuals (adjusted for sex, age, sex×age, age^2, and 40 PCs) and PGS_{EA}. Histograms show the distribution of R^2 across resamples. Black lines indicate the original R^2 in the full source subcohort, red lines indicate the R^2 in the target subcohort. Diagonal panels represent cases where source and target groups are identical.

As shown in Fig. 6, PGS R^2 differs between some cohorts from the same era but different recruitment waves. This may reflect, among other factors related to recruitment strategies, differences in the distribution of the phenotype itself. These distributional differences likely arise because individuals with different characteristics, particularly educational attainment, were more likely to participate in different waves. To test whether adjusting for participation bias could reduce these differences, we applied weights to the EstBB sample using census data from the period of data collection, accounting for age, sex, and EA. In the weighted subsamples, the distribution of EA across eras and waves was matched to the census distribution within each sex and 5-year age group. As an alternative approach, the joint distributions of EA, age and sex were matched to those in the census.

Comparing R^2 across subcohorts (Fig. 7 and Supplementary Table 9) revealed that weighting had only a modest effect. Although the confidence intervals widened due to the reduced effective sample size after weighting, waves 1 and 2 still showed a significant difference in R^2 for OS within the post-Soviet era. The difference between waves also became significant for BMI in the Soviet era. These findings indicate that participation bias related to age, sex, or educational attainment cannot fully explain the observed differences in PGS R^2. This implies that additional, unobserved factors contribute to the variability in R^2, limiting its utility for comparing genetic effects across time-related cohorts. However, other wave-specific factors (e.g. health problems since the wave 1 recruitments were mostly conducted by general practitioners) may also play a role, and our current adjustment for participation bias is not exhaustive.

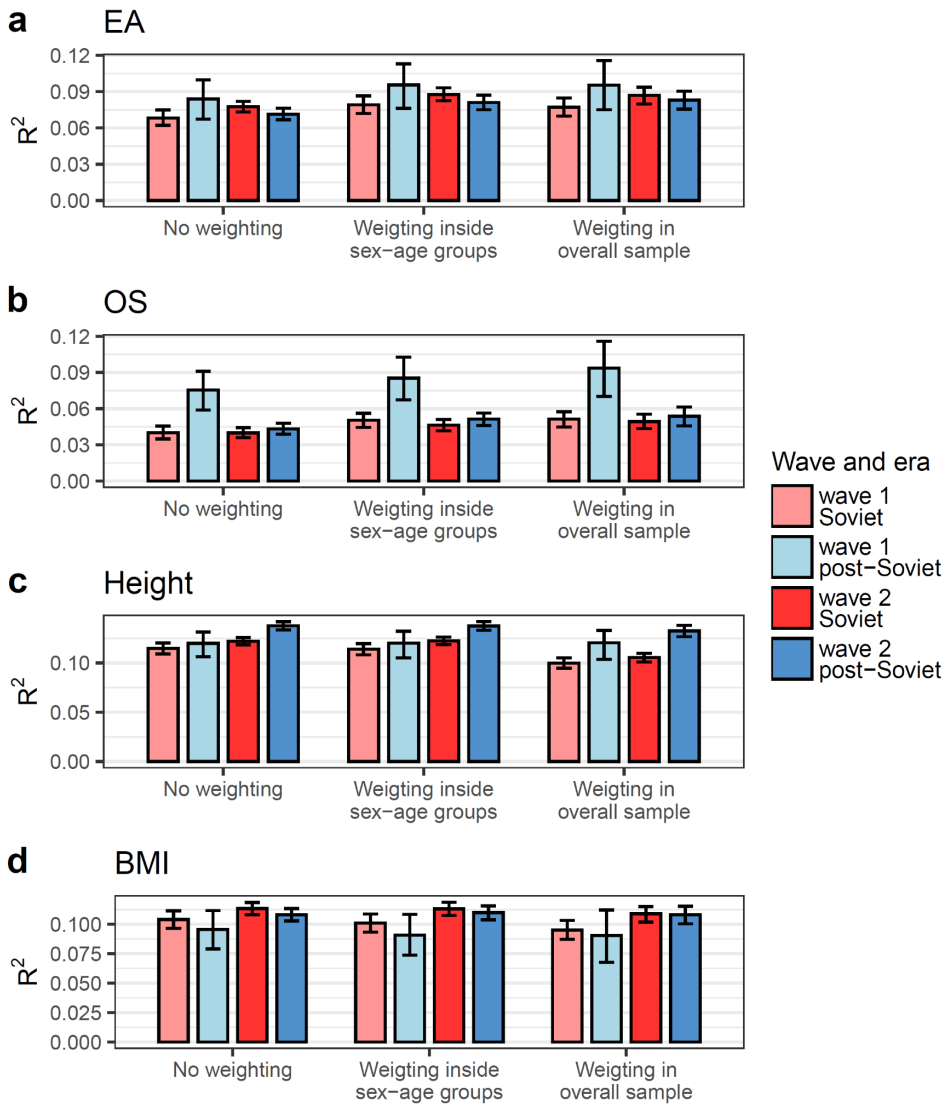

Fig. 7. Trait variance explained by PGS in weighted subsamples. Subsamples are based on era and wave of biobank participation. In all cases, incremental R^2 is shown. Results of three weighting procedures are shown: no weighting; based on the EA distribution in each sex-age group ("inside sex-age groups"); based on the joined distribution of sex, age, and EA ("in the overall sample"). Censuses of 2011 and 2021 were used as population reference for sex, age, and EA information for wave 1 and wave 2, respectively. Error bars correspond to 95% CIs from bootstrap analysis. PGS for EA was used for EA (**a**) and OS (**b**). PGSs for height and body mass index (BMI) were used for the corresponding traits (**c** and **d**, respectively).

Variation in polygenic score effects across eras

Another common approach to studying gene–environment (G×E) interactions is to use linear regression that includes an interaction term between the PGS and the environmental factor of interest. Like the other approaches, this method has several caveats. The resulting estimates are sensitive to differences in phenotypic variance (as in the case of PGS R^2), apparent interactions due to scaling or model misspecification, and issues of generalizability and external validity[37]. Nevertheless, it provides an alternative means to triangulate our findings.

We first applied linear regression models with EA (measured in years of education) and OS as dependent variables in a subsample of unrelated individuals (no 2nd-degree or closer relatives). The models included wave, era, year of birth (YoB), sex, and all pairwise interactions, including squared YoB, as covariates. Following standard recommendations[38], we included PGS_{EA} and interaction terms between PGS_{EA} and wave, YoB, and sex, in addition to the primary PGS×Era interaction. Analyses were also stratified by participation wave (Fig. 8a,c).

The PGS×Era interaction was significant in the full sample and in both wave-specific subsamples, indicating a stronger PGS_{EA} effect in the post-Soviet era. For OS, the interaction was significant in the full sample and in wave 1.

Era is directly derived from YoB, with the post-Soviet era coded as 1, corresponding to higher YoB values. Interestingly, the regression results showed that PGS×Era and PGS×YoB effects were in opposite directions, suggesting a non-linear relationship between PGS_{EA} effects and birth cohort. To examine this, we repeated the analysis excluding the PGS×YoB term (Fig. 8b,d). This led to an attenuation of PGS×Era effects in all models, with only the OS effect in the full sample remaining significant ($p = 0.046$). These results indicate that PGS_{EA} effects vary continuously across birth cohorts within each era, declining with increasing birth year. When this trend is controlled for (i.e., when PGS×YoB is included), significant PGS×Era interactions emerge for both EA and OS. However, when comparing the average PGS_{EA} effects across the Soviet and post-Soviet eras, rather than modelling birth-year-related trends, the difference is not statistically significant for EA, and only marginally significant for OS.

Because EA and OS are ordinal rather than continuous traits, treating them as continuous in linear regression can bias interaction estimates[37]. To assess robustness, we transformed EA and OS into binary outcomes and applied logistic regression. EA was coded as the presence or absence of a university degree, and OS distinguished managers, professionals, technicians and associate professionals from all other categories, yielding two approximately equal groups. In this framework, PGS×Era remained a significant predictor of EA even without the PGS×YoB term in wave 1 (Supplementary Fig. 9). For OS, PGS×Era was significant in all cases. However, when using the year of birth 1981 (10 years in 1991) as the cutoff between Soviet and post-Soviet eras, PGS×Era was no longer significant for either categorical or binary OS (Supplementary Fig. 10 and 11), emphasising that the results depend on specific cohort definitions and sample composition.

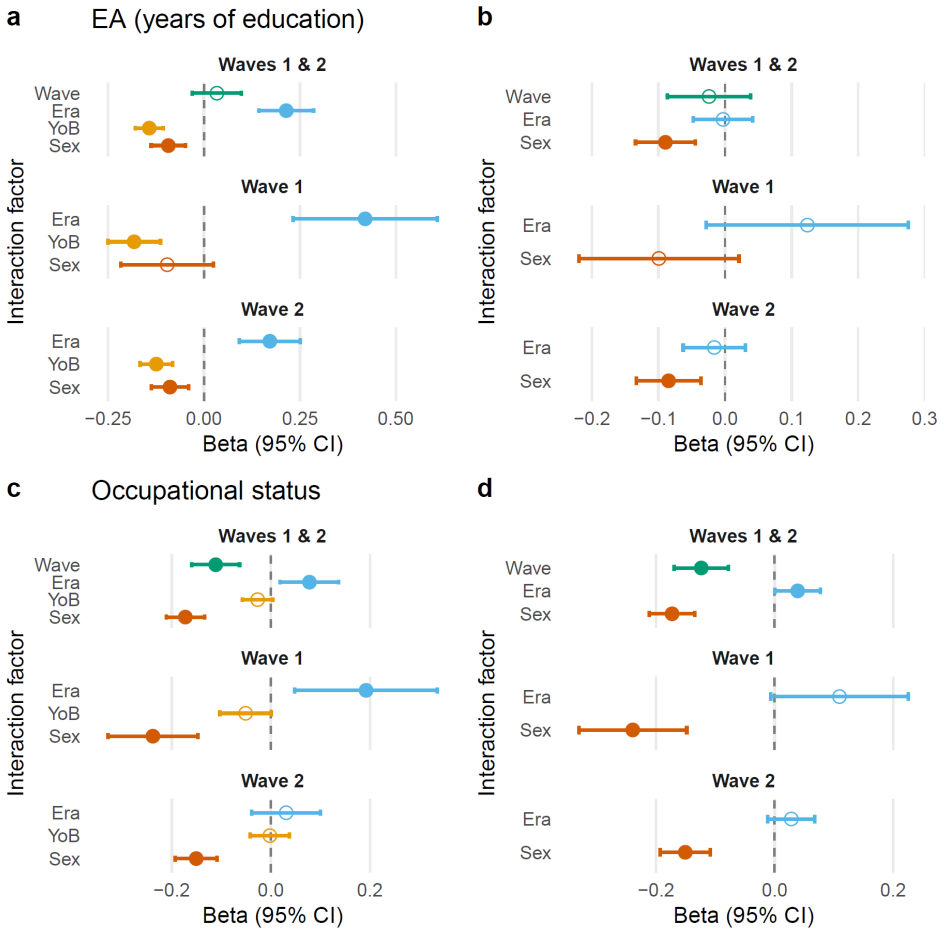

Fig. 8. PGS interaction effects with era and demographic factors across outcomes. Linear regression models for years of education (a–b) and occupational status (c–d) show estimated effect sizes (β, 95% CI) for interactions of the PGS_{EA} with era, wave, sex, and year of birth (YoB). Models in the left column include the PGS×YoB interaction; models in the right column exclude it. Points show estimated interaction effects (β) with 95% confidence intervals, indicating how the association between PGS_{EA} and the phenotype differs by wave, era, YoB, or sex. Filled circles denote statistically significant effects ($p < 0.05$). Each panel is stratified by participation wave (Waves 1 & 2, Wave 1, Wave 2).

Discussion

The cornerstone of science is reproducibility and replicability. Strong claims require strong evidence. However, many published social science results are false and attempts to replicate them often fail[39]. Furthermore, reproducibility alone is not enough; broader replicability is needed to ensure the robustness of scientific findings[40,41]. When an initial study is biased, and subsequent replications use the same methodology, they will systematically reproduce the original mistake. Therefore, it is vital to question biases and sample representativeness, and to triangulate by employing different approaches and datasets to address the same question[40,42]. Here, we aimed to revisit and extend our previous findings[14] using a larger sample, updated methods, and multiple analytical strategies.

Across analyses, we find limited evidence that genetic influence on educational attainment changed after the collapse of the Soviet Union. Estimates of educational attainment vary with analytical choices and, in some cases, suggest a somewhat higher genetic contribution in the post-Soviet cohort. However, these differences do not form a robust pattern, and we do not interpret them as evidence of a shift attributable to the Soviet–post-Soviet transition. Instead, heritability varies across birth cohorts without a clear breakpoint, and similar era differences are observed for height and BMI, traits that are not directly linked to educational policy or economic restructuring. This suggests that observed differences reflect sample structure, phenotype distributions, and cohort dynamics rather than a change in underlying genetic architecture.

This pattern of non-replication is not unique to our findings. The challenges of identifying robust gene–environment interactions (G×E) may be even greater than those of detecting the main effects of genetic or environmental influences on complex traits. Many reported G×E findings are unlikely to be true, in part due to small sample sizes and winner's curse, the complexity of modelling interactions and the difficulty of reliably measuring environmental exposures[43,44]. Social and psychological environments are complex and inconsistently defined, making replication especially challenging. Our results contribute to the literature by showing that initial patterns often weaken as sample size increases or when alternative specifications are used.

One specific reason that may explain the differences in results between the original and extended samples is participation bias. It is well documented that participants in the study samples are systematically different from the wider population they are intended to represent[45–48]. UK Biobank participants, for example, tend to be more educated, healthier, and have healthier lifestyles than the general UK population[45]. More importantly for our case, the bias may depend on gender, age, or recruitment strategy[34,49]. This would result in different cohorts of the dataset being biased in different ways relative to the underlying demographic groups in the populations. One feature of the Estonian Biobank with respect to participation bias is that there were two recruitment waves: wave 1 recruited before 2016, and wave 2 recruited after that. These two differed in their predominant recruitment strategy, with wave 2 showing a more substantial "healthy volunteer bias" than wave 1[33]. When we compare estimates across recruitment waves and historical eras, and then zoom in to 10- or 5-year birth cohorts, we observe changes in both heritability and PGS prediction; however, these changes do not reflect the transition from the Soviet to the post-Soviet era. This pattern suggests that at least some of the apparent Soviet–post-Soviet differences may reflect who enrolled in each wave, rather than genuine population-level change. Using weighting strategies to adjust for non-representativeness in age, sex, and educational attainment reduces some of these discrepancies but does not eliminate them, suggesting that additional, unmeasured aspects of participation likely influence our estimates.

Differences in how the phenotype is distributed across cohorts are also likely to contribute to our results. In our data, later-born cohorts tend to have more years of education, and the distribution of educational attainment also evolves over time. Heritability estimates and PGS R^2 depend on this scale; therefore, shifts in the distribution can lead to variation in estimated genetic influence

even if the underlying genetic architecture remains constant. Our analysis of binary EA heritability on a liability scale, which is not sensitive to the scale effects, did not reveal a significant difference between eras. In our distribution-matching analysis, aligning the EA distributions across cohorts shifted R^2 closer to the value in the target cohort. This suggests that part of the variation in predictive accuracy results from distributional differences rather than genuine changes in genetic effects.

This sensitivity is not unique to historical comparisons. Prediction accuracy varies even within ancestry groups, declining with genetic distance from the discovery sample[28,50] and performing less well in regional or socioeconomically distinct subgroups[27,51,52]. For traits such as educational attainment and cognitive ability, genetic prediction is further complicated by social stratification, assortative mating, participation bias, and broader societal changes that accumulate gradually over time[53]. Our findings align with this evidence and underscore the importance of caution when interpreting differences in polygenic score performance as evidence of changes in genetic influence. Our findings are better explained by gradual environmental changes, age-related effects, and survivor bias[50] than by a single historical step-change in genetic influence.

Our study has several strengths and limitations. We utilised a large sample and employed multiple analytical approaches (heritability, PGS performance, era comparisons, wave comparisons, and birth cohort binning), allowing us to test the robustness of our findings. Results varied somewhat depending on modelling choices and phenotype definitions, but the overall conclusion remained consistent: we find no robust evidence of a shift in genetic influence on educational attainment following the collapse of the Soviet Union. We also examined other phenotypes and a binary definition of university attendance, which is less sensitive to scale and distributional changes, and observed consistent patterns. However, like most volunteer-based biobanks, the Estonian Biobank is not population-representative, and sample composition differs across recruitment waves. Our adjustment to address non-representativeness through weighting reduced, but did not eliminate, these differences, as between-era differences did not change substantially. This suggests that unmeasured sources of participation bias remained in the sample. The sample is also ethnically homogeneous and geographically specific, and no external replication cohort was available, limiting generalisability beyond Estonia. Finally, although we focus on a clear historical divide, economic and social change unfolded gradually, and some

measures may not fully capture this complexity. Moreover, we did not address potential indirect genetic effects, as there were no clear differences between the eras. However, accounting for them is crucial for interpreting results in studies that report apparent differences in genetic effects, even though it increases the complexity of the analysis.

Overall, our findings provide limited evidence that genetic influence on educational attainment changed after the collapse of the Soviet Union. Differences in polygenic score performance across cohorts are more likely due to gradual environmental changes and changes in phenotype distributions than to changes in underlying genetic architecture. Participation bias in both our sample and the original GWAS complicates interpretation, as it can distort estimates of SNP heritability and prediction accuracy. These results highlight the importance of replication, methodological triangulation, and careful consideration of representativeness in genetically informed research.

Methods

Preregistered and actual study

The analysis plan was preregistered in the Open Science Framework prior to accessing the data (https://osf.io/ug5jp/).

The first step in the analysis plan was to replicate the findings of Rimfeld et al.[14] using a similar methodological approach (comparison of SNP heritability, PGS predictive power). According to the plan, bootstrap was used for the comparison of the proportion of variance explained by PGS in the traits. We also estimated genetic correlations between EA in the EstBB and the GWAS sample used for PGS calculation. Height and body mass index (BMI) were used along with EA and OS.

As the heritability analysis did not reveal clear differences between the eras that could be directly linked to the collapse of the Soviet Union and Estonia's regaining of independence for either trait, and because apparent differences can arise for unaccounted-for reasons, we limited subgroup analyses to those that remained theoretically and empirically justified. Specifically, we conducted sex-stratified analyses, as sex represents a key demographic dimension with potential

relevance to both participation and outcome patterns. However, we omitted urban/rural-stratified analyses, as well as those involving transmitted and non-transmitted alleles or cognitive and non-cognitive PGS components. These were all part of the preregistered plan, but our results show that both participation bias and differences in the underlying phenotype distributions across cohorts introduce systematic distortions. These biases violate core assumptions of these models, particularly the comparability of variance structures and the interpretation of between-group differences. We therefore report sex-stratified heritability estimates (Supplementary Fig. 12) in the Supplementary material, but other PGS comparisons were deemed insufficiently robust and potentially misleading, as they risk conflating historical changes in participation and measurement with genuine differences in aetiology.

The analysis of fecundity is out of the scope of this study and will be conducted separately.

Participants

The participants in this study were drawn from the Estonian Biobank (EsBB), a volunteer-based cohort of the Estonian adult population[32,54]. It includes genetic and phenotypic data on 210,438 individuals (72,708 men and 137,730 women), representing approximately 20% (14% men and 24% women) of Estonia's contemporary adult population. Participants' year of birth ranges from 1905 to 2003 (mean, 1972; SD, 16.5). The participants were recruited over a two-decade period, from 2001 to 2021, across the country, covering all regions and a variety of different settings, thereby providing socio-economic and cultural heterogeneity. Besides genetic and demographic data, participants provided health information, blood samples, and lifestyle details. Participants born outside the country were excluded from the analysis.

Ethics statement

The activities of the EstBB are regulated by the Human Gene Research Act, which was adopted in 2000 specifically for the operations of the EstBB. Individual-level data analysis in the EstBB was carried out under ethical approval “1.1-12/3593” from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application “4-1.6/GI/79” from the Estonian Biobank.

Genotyping and quality control

Samples were genotyped on the Infinium Global Screening Array (GSA) of different versions (depending on the time of recruitment) with approximately 550,000 overlapping positions. Samples with <95% call rate or a mismatch between genetic and self-reported sex were excluded. Before the imputation step, all non-SNP polymorphisms and strand-ambiguous SNPs were filtered out. The final number of SNPs before the imputation step was 309,258. The genotypes were imputed with Beagle 5.4[55] using the Estonian Reference panel as a reference set[56]. Related individuals with a 2nd-degree and closer relationship were excluded from every analysed subsample. Relationships were inferred with KING[57].

Ancestry and PCA

Genetic ancestry grouping was estimated using imputed genotypes with bigsnpr, following the original workflow[58]. For ancestry inference, genotypes were imputed using 1000 Genomes Project phase 3 samples[59] for better representation of global haplotype diversity. Individuals from 'Europe (East)', 'Europe (North-West)' and 'Finland' inferred ancestry groups were kept for further analysis. Next, individuals with self-reported ethnicity other than 'Estonian' were excluded from the subsequent analyses, except for the reproduction of the results of Rimfeld et al.[14], where only the genetic outliers were filtered out. These steps were implemented to recruit a relatively homogeneous group of participants with respect to genetics and culture.

A principal component analysis (PCA) was conducted to capture genetic structure. Before the analysis, genotypes were filtered for minor allele frequency >0.01, Hardy–Weinberg equilibrium (HWE) p-value $>10^{-5}$ and missingness <0.05. Long-range linkage disequilibrium regions were removed[60]. Genotypes were pruned for linkage disequilibrium with PLINK2[61,62] with a window size of 50kb, step 5kb and r^2 threshold of 0.1. The first 100 principal components (PCs) were calculated on this SNP set using flashPCA version 2[63].

Phenotypes

Information on the highest level of education was obtained by taking the maximum reported value across the Estonian Population Register and the EstBB questionnaires. Continuous and binary traits corresponding to EA were considered. The continuous 'years of education' phenotype was derived according to the ISCED 2011[64] methodology. The link table for the reported level of education, ISCED 2011 and 'years of education' is presented in Supplementary

Table 10. Alternatively, attainment of a Bachelor's (University) degree or higher was used as a binary phenotype (0 - not having a Bachelor's degree; 1 - having a Bachelor's degree or higher).

Occupational status was assessed as in the study by Rimfeld et al. 2018[14]. The International Standard Classification of Occupations (ISCO)[65] categories were used. The data from the questionnaire about "current professional status" and "main professional status" were combined. Military professions were excluded from the records. First, the mean value was calculated for each of the fields in case of multiple answers. Next, the mean of these values was used as the trait value.

Height and weight were measured by medical personnel. First available measurements were used. The age of the measurement was used as a covariate in the corresponding analyses. Height and Body Mass Index (BMI) were transformed to the logarithm scale to avoid the scale effect[66,67]. Individuals with values deviating from the means on at least one of the traits by more than 4 SD were excluded from the analysis.

Heritability estimation and comparison

Trait heritability was estimated by LDAK version 5.1[19,68], where not otherwise stated. The imputed genotypes were filtered using PLINK2 to retain only biallelic single-nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) >0.01, a Hardy–Weinberg equilibrium (HWE) p-value $>10^{-5}$. Next, SNPs were thinned using a squared correlation (r2) threshold of 0.98 within a 100-kb window. Per-chromosome GRMs were calculated assuming the LDAK model, assuming equal weights and scaling parameter (--power) of -0.25. These per-chromosome GRMs were then merged into a single GRM. LDAK generalised REML (restricted maximum likelihood) solver was used to get heritability estimates. Sex, age, sex×age, age^2, and 40 genetic principal components (PCs) were used as covariates.

For GCTA-GREML analysis[69], the genotyped SNP data were used (following methodology from Rimfeld et al.[14]) after keeping SNPs with minor allele frequency >0.01, Hardy–Weinberg equilibrium (HWE) p-value $>10^{-5}$ and missingness <0.015. GRMs were calculated per chromosome and merged afterwards. GCTA-GREML was applied to get heritability estimates. Sex, age, sex×age, age^2, and 10 PCs were used as covariates.

The binary EA (University degree) heritability estimates were transformed from the observed to the liability scale based on census population prevalence[70,71] using the Robertson transformation[72]. Heritability values were compared with a paired Z-test.

Polygenic score calculation and performance comparison

EA PGS was calculated using summary statistics from the largest GWAS of educational attainment currently available (EA4), with Estonian individuals and the 23andMe cohort excluded from the meta-analysis[23]. The height and BMI PGSs were calculated using summary statistics from GWAS in the European ancestry cohort of the UK Biobank conducted by the Pan-UKBB project[73]. To create polygenic scores, we extracted a set of 1,075,599 autosomal HapMap 3 SNPs with a minor allele count >5 and info score >0.7 in the EstBB. Polygenic scores were calculated using SBayesR (gctb_2.02) with default parameters (--pi 0.95,0.02,0.02,0.01 --gamma 0.0,0.01,0.1,1 --chain-length 10000 --burn-in 2000 --out-freq 10), including an LD matrix built using data on 50,000 UK Biobank participants[35].

To assess the PGS performance, the incremental R^2 was calculated using sex, age, sex×age, age^2, and 40 PCs as covariates. Alternatively, traits were pre-adjusted for these covariates, and then R^2 was derived from a univariate linear regression with PGS as a predictor. 95% confidence intervals (95% CIs) were calculated using bootstrap with 1000 replicates. Two-tailed p-values were derived from the same bootstrap replicates. R^2 95% CIs for half-decade birth cohorts were calculated using Fisher z-transformation, assuming that incremental R^2 is approximately equal to the squared correlation coefficient.

Statistical significance

A p-value of 0.05 was used as the threshold for statistical significance in all tests. Although we analysed up to four traits and conducted multiple comparisons between estimates across subcohorts, we did not apply a correction for multiple testing. This decision was based on several considerations. First, the tests are not independent, as the traits examined are substantially genetically correlated[74]. Second, the subcohorts defined by recruitment wave largely overlap with the broader Soviet and post-Soviet eras (though not entirely, as relatives were filtered independently within each era). Finally, because this study challenges the findings reported by

Rimfeld et al.[14], applying a conservative correction would reduce the likelihood of detecting false negatives rather than controlling false positives, making our non-replication results more robust.

Genetic correlations

To estimate the genetic correlation between EstBB EA and EA4 meta-analysis EA, we first conducted a GWAS in the EstBB cohort. We used the SAIGE method[75] following the guidelines provided. Genetic correlation was estimated from the GWAS summary statistics using LD score regression with default LD scores derived from 1000 Genomes European ancestry individuals[74].

Matching the EA distribution

To evaluate how differences in the distribution of EA affect the predictive performance ofPGS_{EA}, we conducted a resampling analysis across four subcohorts defined by recruitment wave and era. First, the EA phenotype was pre-adjusted for sex, age, sex×age, age^2, and 40 genetic principal components by regressing these covariates on EA and using the residuals. For each subcohort ("source group"), individuals were randomly sampled with replacement to create new subsamples whose EA distribution matched that of a different subcohort ("target group"). Matching was performed either by EA alone or jointly by EA and sex. For each subsample, the variance explained by the PGS_{EA} was quantified as the squared Pearson correlation between the pre-adjusted EA residuals and the PGS, which is an equivalent of R2 in an univariate linear modek. This procedure was repeated 1,000 times for each source-target combination to generate distributions of R^2 values. These distributions were compared with the original R^2 values of the source and target subcohorts to assess how differences in EA (and sex) distributions influence PGS performance.

Correction for participation bias

No established procedure for correction for participation bias exists for the EstBB. Thus, we used three variables known to have a strong effect on participation in biobank initiatives, whose summary data are easily accessible for the corresponding population: age, sex and EA[45,49]. Increasing the number of variables used for the correction decreases the effective sample size in further analyses. To explore how this affects the results, two weighting schemes were implemented: (1) marginal adjustment, where the EA distribution within each sex and age

stratum was matched to that of the census; and (2) joint adjustment, where the full joint distribution of age, sex, and EA matched that in the census population.

We used the 2011 census data[70] as a population reference for wave 1 of participation and the 2021 census data[71] for wave 2. This is justified by the time between census data collection and biobank participation, as well as by the fact that participants kept in the study were older than 25 at the time of enrolment and had therefore mostly completed their highest level of education. The age groups were divided into 5-year bins. EA had three categories: primary, secondary, and tertiary education. The post-stratification weighting method was used for weight calculation, further used in weighted linear regression to calculate incremental R^2. Bootstrap procedure with 1000 replicates was used to calculate R^2 95% CIs. In each replicate, a subsample with replacement was selected, after which weighting was applied.

Variation in PGS effects across eras

To assess whether the predictive power of PGS_{EA} for EA and OS differed between the Soviet and post-Soviet eras, we conducted interaction analyses using linear and logistic regression models.

We modelled EA (years of education) and OS (treated as continuous) as dependent variables in unrelated individuals (excluding 2nd-degree and closer relatives). The primary linear regression model included main effects of PGS, era (coded 0 for the Soviet era and 1 for the post-Soviet era), and their interaction term (PGS×Era). The model additionally controlled for participation wave (0 = wave 1; 1 = wave 2), year of birth (YoB, z-standardised), sex (0 = male; 1 = female), and all pairwise interactions among these covariates, including a quadratic YoB term. Following recommendations for modelling G×E interactions[38], we also included interaction terms between PGS and wave, YoB, and sex to reduce confounding due to omitted interactions. Because era is a binary variable derived from YoB, we also estimated models that excluded the PGS×YoB term.

Analyses were performed in the full sample and separately within each participation wave. The statistical significance of the PGS×Era term was used to evaluate whether PGS effects differed between eras.

To assess robustness to treating ordinal traits as continuous[37], we repeated the analyses using logistic regression with binary outcomes. EA was dichotomised as the presence vs. absence of a

university degree. OS was dichotomised by grouping managers, professionals, and technicians and associate professionals (ISCO major groups 1–3) versus all other occupational categories (excluding “unclassified” and “armed forces”), yielding approximately balanced classes.

All models included 40 genetic principal components to control for population structure.

Data availability

Access to the Estonian Biobank data (https://genomics.ut.ee/en/content/estonian-biobank) is restricted to approved researchers and can be requested.

Code availability

Custom R code used for statistical analyses is available on GitHub (https://github.com/ivkuz/HeritabilityChangeEst).

Acknowledgements

We acknowledge the participants of the Estonian Biobank. The research was conducted using the Estonian Center of Genomics/Roadmap II, funded by the Estonian Research Council (project number TT17). Data analysis was carried out in part in the High-Performance Computing Center of the University of Tartu. IK and AM were supported by the Centre of Excellence for Personalised Medicine, funded by grant TK214 from the Estonian Ministry of Education and Research. KR was supported by a UK Medical Research Council Grant UKRI1503. KL was supported by the Estonian Research Council (grant no. PSG615) and by the Estonian Centre of Excellence for Well-Being Sciences (EstWell), funded by grant TK218 from the Estonian Ministry of Education and Research. PH was funded by the Estonian Research Council grant PRG1137.

Conflicts of interest

The authors declare no competing interests.

References

1. Adler, N. E. *et al.* Socioeconomic status and health. The challenge of the gradient. *Am. Psychol.* **49**, 15–24 (1994).
2. Cutler, D. M., Lleras-Muney, A. & Vogl, T. *Socioeconomic Status and Health: Dimensions and Mechanisms*. (Oxford University Press, 2011). doi:10.1093/oxfordhb/9780199238828.013.0007.
3. Batty, G. D., Deary, I. J. & Gottfredson, L. S. Premorbid (early life) IQ and later mortality risk: systematic review. *Ann. Epidemiol.* **17**, 278–288 (2007).
4. von Stumm, S., Deary, I. J. & Hagger-Johnson, G. Life-course pathways to psychological distress: a cohort study. *BMJ Open* **3**, e002772 (2013).
5. Heath, A. C. *et al.* Education policy and the heritability of educational attainment. *Nature* **314**, 734–736 (1985).
6. Branigan, A. R., McCallum, K. J. & Freese, J. Variation in the heritability of educational attainment: An international meta-analysis. *Soc. Forces* **92**, 109–140 (2013).
7. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
8. Silventoinen, K. *et al.* Genetic and environmental variation in educational attainment: an individual-based analysis of 28 twin cohorts. *Sci. Rep.* **10**, 12681 (2020).
9. Colodro-Conde, L., Rijsdijk, F., Tornero-Gómez, M. J., Sánchez-Romera, J. F. & Ordoñana, J. R. Equality in educational policy and the heritability of educational attainment. *PLoS One* **10**, e0143796 (2015).
10. Fraemke, D. *et al.* Polygenic associations with educational attainment in East versus West Germany: Differences emerge after reunification. *Psychol. Sci.* **36**, 559–573 (2025).
11. Pingault, J.-B. *et al.* Research Review: How to interpret associations between polygenic scores, environmental risks, and phenotypes. *J. Child Psychol. Psychiatry* **63**, 1125–1139 (2022).
12. Plomin, R. & von Stumm, S. The new genetics of intelligence. *Nat. Rev. Genet.* **19**, 148–159 (2018).
13. Shakeshaft, N. G. *et al.* Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLoS One* **8**, e80341 (2013).

14. Rimfeld, K. *et al.* Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nat. Hum. Behav.* **2**, 269–275 (2018).

15. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).

16. Hill, W. D. *et al.* Molecular genetic contributions to social deprivation and household income in UK Biobank. *Curr. Biol.* **26**, 3083–3089 (2016).

17. Marioni, R. E. *et al.* Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* **44**, 26–32 (2014).

18. Hyytinen, A., Ilmakunnas, P., Johansson, E. & Toivanen, O. Heritability of lifetime income. *SSRN Electron. J.* (2013) doi:10.2139/ssrn.2253264.

19. Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).

20. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).

21. Young, A. I. Solving the missing heritability problem. *PLoS Genet.* **15**, e1008222 (2019).

22. Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).

23. Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, 437–449 (2022).

24. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).

25. Wang, Y. *et al.* Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat. Commun.* **11**, 3865 (2020).

26. Privé, F. *et al.* Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am. J. Hum. Genet.* **109**, 12–23 (2022).

27. Mostafavi, H. *et al.* Variable prediction accuracy of polygenic scores within an ancestry group. *Elife*

9, e48376 (2020).

28. Ding, Y. *et al.* Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature* **618**, 774–781 (2023).
29. Nivard, M. G. *et al.* More than nature and nurture, indirect genetic effects on children's academic achievement are consequences of dynastic social processes. *Nat. Hum. Behav.* **8**, 771–778 (2024).
30. Tan, T. *et al.* Family-GWAS reveals effects of environment and mating on genetic associations. *medRxiv* (2024) doi:10.1101/2024.10.01.24314703.
31. Guan, J. *et al.* Family-based genome-wide association study designs for increased power and robustness. *Nat. Genet.* **57**, 1044–1052 (2025).
32. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).
33. Milani, L. *et al.* The Estonian Biobank's journey from biobanking to personalized medicine. *Nat. Commun.* **16**, 3270 (2025).
34. Song, S., Benonisdottir, S., Liu, J. S. & Kong, A. Participation bias in the estimation of heritability and genetic correlation. *Proc. Natl. Acad. Sci. U. S. A.* **122**, e2425530122 (2025).
35. Lloyd-Jones, L. R. *et al.* Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, 5086 (2019).
36. Kutner, M., Nachtsheim, C., Neter, J. & Li, W. *Applied Linear Statistical Models with Student CD*. (McGraw-Hill/Irwin, 2004).
37. Domingue, B. W., Trejo, S., Armstrong-Carter, E. & Tucker-Drob, E. M. Interactions between polygenic scores and environments: Methodological and conceptual challenges. *Sociol. Sci.* **7**, 465–486 (2020).
38. Keller, M. C. Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol. Psychiatry* **75**, 18–24 (2014).
39. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
40. Munafò, M. R. & Davey Smith, G. Robust research needs many lines of evidence. *Nature* **553**,

399–401 (2018).

41. Munafò, M. R. *et al.* A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
42. Munafò, M. R., Higgins, J. P. T. & Smith, G. D. Triangulating evidence through the inclusion of genetically informed designs. *Cold Spring Harb. Perspect. Med.* **11**, a040659 (2021).
43. Virolainen, S. J., VonHandorf, A., Viel, K. C. M. F., Weirauch, M. T. & Kottyan, L. C. Gene-environment interactions and their impact on human health. *Genes Immun.* **24**, 1–11 (2023).
44. Munafò, M. R. & Flint, J. Replication and heterogeneity in gene x environment interaction studies. *Int. J. Neuropsychopharmacol.* **12**, 727–729 (2009).
45. Schoeler, T. *et al.* Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat. Hum. Behav.* **7**, 1216–1227 (2023).
46. van Alten, S., Domingue, B. W., Faul, J., Galama, T. & Marees, A. T. Correcting for volunteer bias in GWAS increases SNP effect sizes and heritability estimates. *Nat. Commun.* **16**, 3578 (2025).
47. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* **47**, 226–235 (2018).
48. Griffith, G. J. *et al.* Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat. Commun.* **11**, 5749 (2020).
49. Pirastu, N. *et al.* Genetic analyses identify widespread sex-differential participation bias. *Nat. Genet.* **53**, 663–671 (2021).
50. Zhang, L. & Conley, D. Within-Family GWAS does not Ameliorate the Decline in Prediction Accuracy across Populations. *bioRxiv* (2024) doi:10.1101/2024.12.12.628188.
51. Zaidi, A. A. & Mathieson, I. Demographic history mediates the effect of stratification on polygenic scores. *Elife* **9**, e61548 (2020).
52. Hou, K. *et al.* Calibrated prediction intervals for polygenic scores across diverse contexts. *Nat. Genet.* **56**, 1386–1396 (2024).
53. Gusev, S. How population stratification led to a decade of sensationally false genetic findings. *The*

Infinitesimal https://theinfinitesimal.substack.com/p/how-population-stratification-led (2025).

54. Milani, L. *et al.* From Biobanking to Personalized Medicine: the journey of the Estonian Biobank. *medRxiv* 2024.09.22.24313964 (2024) doi:10.1101/2024.09.22.24313964.
55. Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
56. Mitt, M. *et al.* Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869–876 (2017).
57. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
58. Privé, F. Using the UK Biobank as a global reference of worldwide populations: application to measuring ancestry diversity from GWAS summary statistics. *Bioinformatics* **38**, 3477–3480 (2022).
59. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
60. Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–5; author reply 135–9 (2008).
61. PLINK 2.0. http://www.cog-genomics.org/plink/2.0/.
62. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
63. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776–2778 (2017).
64. UNESCO Institute for Statistics. *International Standard Classification of Education (ISCED) 2011*. (UNESCO Institute for Statistics, 2012). doi:10.15220/978-92-9189-123-8-en.
65. The leading source of labour statistics. *ILOSTAT* http://www.ilo.org/public/english/bureau/stat/isco/ (2019).
66. Slavskii, S. A. *et al.* The limits of normal approximation for adult height. *Eur. J. Hum. Genet.* **29**,

1082–1091 (2021).

67. Silverman, M. P. & Lipscombe, T. C. Exact statistical distribution of the body mass index (BMI): Analysis and experimental confirmation. *Open J. Stat.* **12**, 324–356 (2022).

68. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).

69. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

70. 2011 Population and Housing Census. https://www.stat.ee/en/statistics-estonia/population-census-2021/2011-population-and-housing-census.

71. Population census 2021. https://www.stat.ee/en/statistics-estonia/population-census-2021/.

72. Dempster, E. R. & Lerner, I. M. Heritability of Threshold Characters. *Genetics* **35**, 212–236 (1950).

73. Karczewski, K. J. *et al.* Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. *bioRxiv* (2024) doi:10.1101/2024.03.13.24303864.

74. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

75. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).