

Effects of Genetic Propensity for Education on Labor Market and Health Trajectories across the Working Life*

Stefano Lombardi^{† 1,2,3}, Nurfatima Jandarova^{† 4}, Kristina Zguro^{† 5}, Jarkko Harju⁴, Aldo Rustichini⁶, and Andrea Ganna^{5,7,8}

¹VATT Institute for Economic Research, Helsinki, Finland, 00100

²IFAU and Uppsala Center for Labor Studies, Uppsala, Sweden, 75105

³IZA, Bonn, Germany, 53113, and Rockwool Foundation, Berlin, Germany, 10119

⁴Economics department at Tampere University, and FIT, Tampere, Finland, 33100

⁵Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Finland, 00290

⁶Economics department at University of Minnesota, MN, USA, 55455

⁷Broad Institute of MIT and Harvard, MA, USA, 02142

⁸Analytic and Translational Genetics Unit, Massachusetts General Hospital, MA, USA, 02114

Abstract

Education is a major source of inequality in income and health. Polygenic indices for educational attainment (EA-PGI) capture both direct and indirect genetic influences on education, but their effects on income and health remain unclear. Using Finnish registry data on 51,056 graduates followed annually since graduation for up to 25 years, we report three findings. First, higher EA-PGI strongly predicts income growth, but only among higher-educated people: tertiary-educated graduates at the 90th percentile earn €45,392 (13.1 %) higher discounted lifetime income than those at the 10th percentile. This effect is not mediated by overall health and is entirely absent for the secondary (high school)-educated workers, who do not benefit from higher EA-PGI levels. Second, EA-PGI does not predict income differences at labor market entry or the quality of the first employer, but rather a higher job-to-job mobility toward higher-quality firms that drives the long-run income divergence. Third, controlling for parental EA-PGI in 12,871 parent-offspring trios reduces the discounted lifetime income gap by 71 %, and the effect of paternal (but not maternal) EA-PGI on offspring income exceeds that of the offspring's own EA-PGI. These findings suggest that genetic factors associated with educational attainment predict income trajectories primarily through faster and more frequent changes to higher-paying employers. However, much of this association reflects indirect paternal genetic effects, consistent with enduring paternal patterns of intergenerational job and income transmission.

Classification: Social Sciences, Economics; Biological Sciences, Genetics

Keywords: income inequality; polygenic index; gene-environment interaction; firm pay premia; income and health trajectories

*Corresponding authors: Andrea Ganna (andrea.ganna@helsinki.fi), Stefano Lombardi (stefano.lombardi@vatt.fi).

[†]Co-first authors.

Significance Statement

Genetic differences are increasingly studied as contributors to socioeconomic inequality, yet little is known about how their influence unfolds over the working life. Using Finnish administrative records linked to molecular genetic data, we follow individuals' earnings trajectories from labor market entry over 25 years. Genetic predispositions are unrelated to income at entry, but substantial differences emerge later in the career. These gaps are concentrated among university graduates and arise primarily through differential sorting into higher-paying firms. Parental—especially paternal—genetic endowments also account for a meaningful share of these differences. Our findings provide new evidence on how genetic and social processes interact over the life course to shape income inequality.

Introduction

Understanding the origins and persistence of inequalities in health, education, and income—collectively referred to as socioeconomic status—is central to economic and social policy. While individual effort and choices contribute to these disparities, socioeconomic outcomes are also shaped by circumstances beyond individual control, including family background and genetic predispositions. The extent to which such circumstances influence life outcomes has important implications for equality of opportunity in society.¹

Social scientists and epidemiologists have long sought to identify the drivers of socioeconomic inequalities,^{2–5} recognizing that genetic and environmental factors interact over the life cycle to shape individual outcomes.^{6,7} The recent development of polygenic indices (PGIs), which summarize genetic predispositions for complex traits, has enabled researchers to directly examine how genetic endowments relate to education and other dimensions of socioeconomic status.^{8–11}

Recent advances in this literature have focused on distinguishing direct genetic effects—arising from genetic variants randomly transmitted at conception—from indirect genetic influences, such as environmentally mediated effects of parental genotypes (“genetic nurture”).^{12–14} Identifying direct genetic effects makes it possible to study how environmental processes amplify or mitigate genetic predispositions and to investigate the mechanisms through which genetic differences translate into socioeconomic outcomes such as income.^{15,16}

Despite the advances in this growing literature, evidence on how genetic predispositions shape income over the working life remains limited. The existing studies that use within-family designs provide static estimates that do not capture how income differences accumulate over the life cycle. They also do not examine potentially relevant mediating channels such as employer sorting, and often rely on coarse, self-reported measures of income and health. Complementary work has examined outcomes such as occupational prestige but without directly analyzing income.¹⁷

This paper examines when and through which mechanisms genetic predispositions translate into income differences over the working life. Using Finnish administrative registers on income, health, and education, we follow individuals’ labor income for 25 years after labor market entry. The data tracks workers in and out of employment and their employers over time, enabling us to analyze how individuals’ own and parental genetic endowments shape income trajectories and the roles of education, firm quality, and health in generating them. We find that genetic predispositions for education are unrelated to income at labor market entry, but substantial differences emerge and widen over the career. These gaps are concentrated among university graduates and arise through differential sorting into higher-paying firms.

The analysis combines polygenic indices for educational attainment (EA-PGIs) with matched employer–employee registers. Linking workers to their employers over their careers allows us to apply the Abowd–Kramarz–Margolis (AKM) framework to estimate worker productivity and firm wage premia.^{18,19} The worker productivity index captures worker-specific earnings capacity that is portable across employers, while firm pay premia measure systematic differences in firm wage policies after accounting for worker productivity. This setting allows us to shed light on the labor market processes generating unequal income trajectories across genetic groups: differences in worker productivity, mobility to higher-paying firms, and wage growth within and across firms.

Our empirical strategy exploits the random transmission of genetic variants at conception and implements a within-family design in which outcomes are regressed on the EA-PGI of the individual and those of both parents. This approach addresses population stratification and other forms of genetic confounding, allowing us to estimate direct genetic effects on income. Because the labor market returns to skills differ substantially by education level, the effects of genetic predispositions on income may vary across educational groups. We therefore present results separately for individuals with college and high school education.

Our analysis advances the literature in two ways. First, by studying income trajectories rather than static outcomes, we distinguish between two fundamentally different sources of long-run income inequality: differences present at labor market entry and differences arising from divergent income growth during the career. Similar income disparities later in life can emerge either from large initial gaps or from similar starting points followed by diverging trajectories. These two patterns correspond to very different mechanisms. Initial differences point to inequalities generated before labor market entry, for example through education or early skill formation. Divergent trajectories instead reflect processes operating during the career, such as job mobility, promotions, or health dynamics. Distinguishing between these mechanisms is therefore crucial for understanding how genetic predispositions contribute to the production and accumulation of income inequality over the life cycle.

Second, within this dynamic framework, we analyze a key channel through which genetics may shape income trajectories: job mobility and sorting into firms with different wage-setting quality. Firms play an important role in explaining wage inequality among workers with similar observable skills.^{20,21} Estimating firm pay premia—capturing how much firms pay relative to other firms after controlling for worker productivity—allows us to move beyond documenting income gaps and identify the labor market mechanisms linking genetic predispositions to income trajectories.

Results

EA-PGI predicts lifetime income trajectories, but only among individuals with tertiary education

The study includes 51,056 individuals with genome-wide genetic data linked to longitudinal health and socioeconomic information. The data covers employment histories (i.e., employee-employer links with job spells length), annual income from labor, and educational records (education level, field and school/university identifiers). This information spans thirty years (1987–2019) and all individuals are followed from graduation up to 25 years later. We estimate a dynamic model to analyze the relation between EA-PGI and individual income over time, adjusting for calendar year, birth year, gender, the first ten genetic principal components, and biobank indicator.

Individuals included in this study are generally better educated and with a higher share of women compared to the nationwide population of fresh graduates in the same calendar period (Supplementary Table A.2a). Reassuringly, when we reweight the sample to reflect the characteristics in the population, the results are virtually unaffected (Supplementary Table A.11).

We first confirm that EA-PGI is significantly associated with years of educational attainment, with results comparable to what previously reported in the literature ($R^2=7.1\%$, Supplementary Table A.12).²² Next, we provide novel estimates on income over the life cycle. Comparing earned income trajectories between the 10th and 90th percentile of the EA-PGI distribution over 25 years since graduation, we show that income levels are initially very similar across groups (Figure 1a). However, trajectories begin to diverge substantially over time. The gap in average annual income between 90th and 10th percentile of EA-PGI 10 years from graduation is €3,262 [CI €2,741 - €3,783] and widens further to €7,534 [CI €6,382 - €8,687] by 25 years from graduation. To put this in context, the average yearly earnings in the sample is €24,932 and €34,858 at 10 and 25 years after graduation, respectively. The cumulated income over the 25 years after graduation, discounted to obtain its present value at graduation, is €309,659 [CI €307,099 - €312,219] for individuals at the 10th percentile and €350,418 [CI €347,300 - €353,536] for those at the 90th percentile of the EA-PGI distribution (Table 1). The gap in cumulated income between the 10th and 90th percentile is €40,759, equivalent to 13.1% and corresponding to 14% of the median income over the same period.

Strikingly, this genetic gradient in income varies markedly by educational attainment. The income gap by genetics is driven entirely by individuals with tertiary education (Figures 1b and 1c), while no systematic differences are observed among the people with secondary education (63.4% of the graduates has tertiary degree). Among the tertiary-educated, differences in income trajectories widen up during the first 15 years after graduation and then stabilize, at which point individuals are, on average,

40 years old and approaching peak labor market attachment.

Such a differential result likely reflects two forces: individuals with relatively high EA-PGI experience higher economic returns to ability, and EA-PGI influences educational sorting, shaping the types and levels of education they pursue. Although we cannot fully separate sorting from the economic returns to higher EA-PGI, we next analyze how EA-PGI correlates with a measure of ability and examine the role of employers in explaining the observed income differences.

Figure 1: Average annual income by EA-PGI level, over time and by education

Panel (a) uses full analysis sample ($N = 51,056$), while Panels (b) and (c) use subset of workers based on their highest qualification being either secondary ($N = 18,692$) or tertiary degree ($N = 32,364$), respectively. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Average income estimated from a regression of annual income on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95 % CIs.

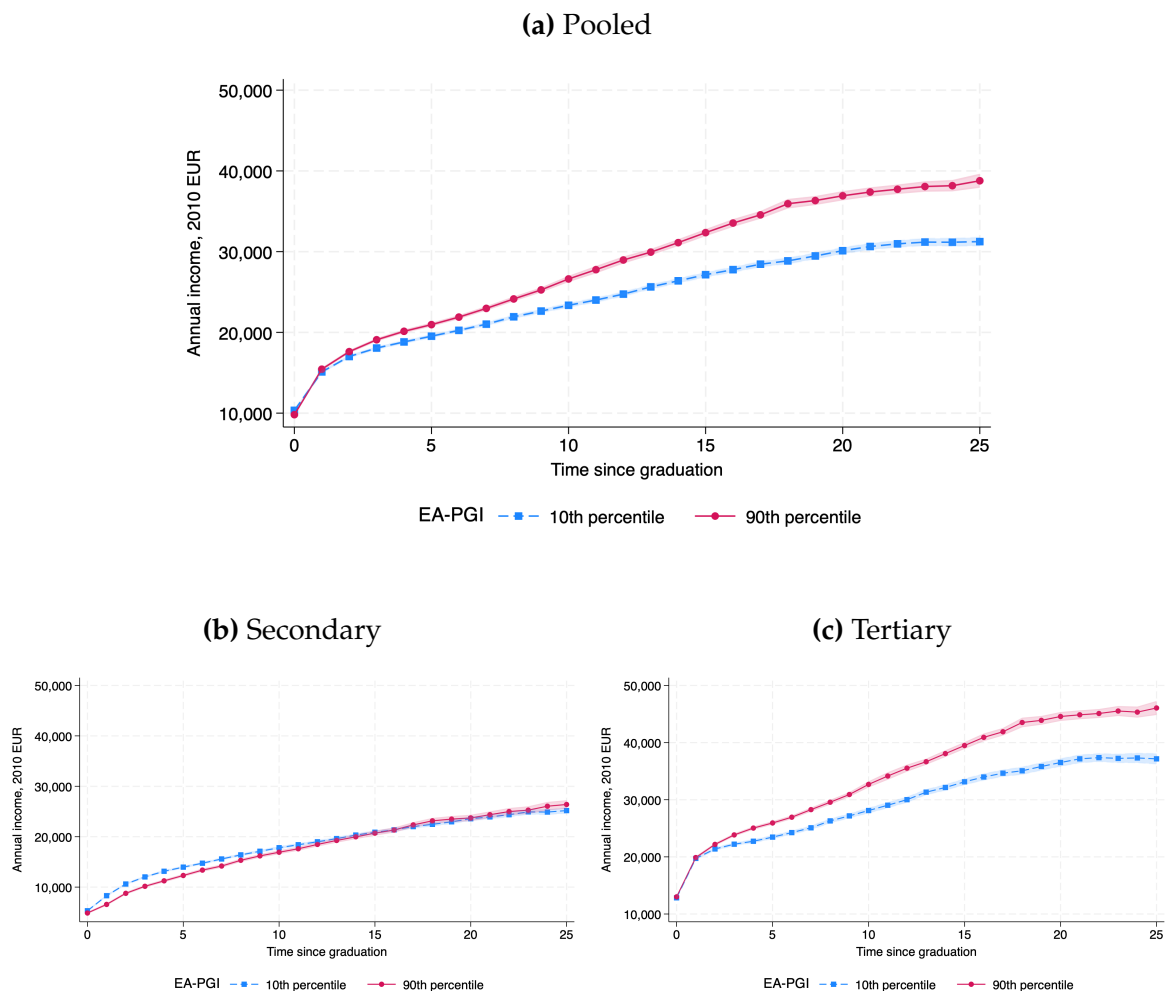


Table 1: Cumulated lifetime income by EA-PGI level

The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles. Column (1) uses full analysis sample ($N = 51,056$), while columns (2) and (3) use subset of workers based on their highest qualification being either secondary ($N = 18,692$) or tertiary degree ($N = 32,364$), respectively. Average lifetime income adjusted by regressing cumulated income on EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income		
	Pooled (1)	Secondary (2)	Tertiary (3)
EA-PGI percentiles			
10th	309 659 (1 306)	262 386 (1 429)	346 194 (1 944)
20th	316 476 (1 037)	260 040 (1 153)	353 786 (1 539)
30th	321 505 (898)	258 309 (1 046)	359 386 (1 304)
40th	325 804 (844)	256 829 (1 042)	364 174 (1 175)
50th	329 893 (857)	255 422 (1 116)	368 728 (1 137)
60th	333 908 (930)	254 041 (1 248)	373 198 (1 188)
70th	338 222 (1 061)	252 556 (1 439)	378 004 (1 328)
80th	343 389 (1 265)	250 777 (1 709)	383 758 (1 579)
90th	350 418 (1 591)	248 358 (2 120)	391 585 (2 006)
Obs.	51 056	18 692	32 364

EA-PGI is associated with the tertiary-educated workers' productivity

To understand the mechanisms through which the EA-PGI influences income trajectories, we first examine its relationship with both employee and employer productivity in the labor market. To quantify productivity, we use full population data within an Abowd–Kramarz–Margolis (AKM) framework, the workhorse model in economics for decomposing wage variation into worker and firm components.^{18,21}

The model exploits repeated measurements of worker's wage, relating it to individual- and firm-specific indicators. After model estimation, each worker is assigned a corresponding estimated intercept, which we refer to as *worker productivity index*, as it captures the persistent component of the worker's wage that is portable when switching jobs across firms, net of the effect of firm-quality and time-varying worker's characteristics.

We find a statistically significant correlation between the worker productivity index estimated via the AKM model and EA-PGI (both standardized to have mean 0 and standard deviation 1), but primarily for the tertiary-educated workers (Figure 2). For this group, a one standard deviation increase in EA-PGI is associated with a 0.115 [CI 0.101 - 0.129] standard deviation higher worker productivity, compared to a 0.017 [CI 0.003 - 0.031] increase among the secondary-educated individuals. Hence, individuals with high EA-PGI tend to be highly productive on the labor market (and therefore are on average paid higher wages), as long as they obtain a tertiary education degree.

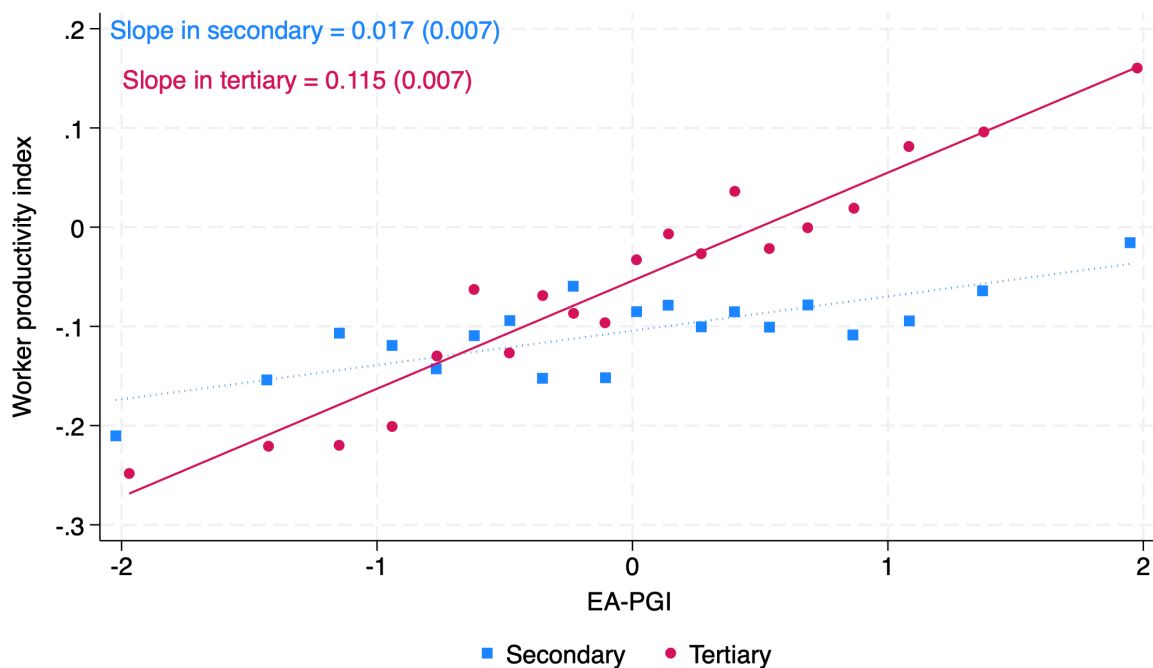
We further examine whether the correlation between EA-PGI and worker productivity persists across education groups after adjusting for fine-grained education-related characteristics (education field, and academic institution). Although the association between EA-PGI and worker productivity attenuates after adjustment (Supplementary Table A.6), it remains statistically significant among tertiary-educated workers and is significantly stronger than among those with secondary education.

High EA-PGI individuals transition more rapidly and more frequently to higher-quality firms

We next investigate the role of firms in shaping income differences across levels of EA-PGI among tertiary-educated individuals. Workers with higher EA-PGI change employers slightly more frequently (4.3 jobs vs 4.5 jobs over 25 years, P-value=0.007 for the 10th and 90th percentile EA-PGI, respectively; Figure 3a). To assess the quality dimension of these moves, we use the AKM regression to estimate a *firm quality index* which captures the firm-specific component of paid wage that is common to all workers at that firm, after accounting for worker ability and observed characteristics. We use the firm quality index to rank all firms in which workers in the sample are employed, and use this as an outcome in our trajectories model.

Figure 2: Worker productivity index and EA-PGI level by education

The vertical axis reports a worker productivity index estimated via an AKM regression (details reported in Section 1.2); the horizontal axis reports ventiles of the EA-PGI distribution. The binscatter plot shows the relation between these two quantities standardized to have mean 0 and standard deviation 1 and after having residualized them with respect to gender, year of birth indicators, and first 10 genetic principal components (PCs). The data used to obtain the plot is a cross-section of 31,866 individuals in the analysis sample with non-missing worker productivity index. The lines correspond to sub-samples based on highest education achieved. The figure also reports the estimated slopes (and standard errors in parentheses) from the corresponding linear regression of worker productivity on EA-PGI controlling for gender, year of birth, calendar year and biobank indicators, and first ten genetic PCs.



We show that, on average, higher EA-PGI individuals transition to significantly higher-quality firms (Figure 3b). Interestingly, the quality of their first employer is not statistically different across the EA-PGI distribution, but the firm quality gap widens as early as three years after graduation. This pattern suggests that individuals do not initially self-select into higher- or lower-quality firms based on their EA-PGI, but sorting by EA-PGI appears to begin early in the career, with higher EA-PGI individuals progressively moving toward better-quality firms.

In contrast, among individuals with secondary education, firm quality trajectories appear similar across EA-PGI levels (Supplementary Figure A.4). These findings suggest that the higher income trajectories observed among tertiary-educated individuals with high EA-PGI are largely driven by greater access to higher-quality, higher-paying firms over time, rather than by difference in educational attainment per se, initial labor market entry, or the frequency of employment transitions alone.

Income growth disparities between high and low EA-PGI are attributable to differences in mobility between rather than within jobs

Having established that tertiary-educated individuals with higher EA-PGI tend to transition more rapidly to higher-quality firms, we next examine how much of the income differences between EA-PGI groups can be attributed to mobility within versus between firms. To do so, we decompose changes in labor income between consecutive years since graduation into within-firm wage growth (stayers), employer-to-employer mobility, and transitions into and out of non-employment.²³

Overall, within-firm wage growth contributes most to cumulative income gains over time and does so similarly for individuals at the 10th and 90th percentiles of EA-PGI (Figure 4a). However, the contribution of between-firm mobility to earnings growth becomes increasingly important over time for those at the 90th percentile compared to those at the 10th percentile (Figure 4b). After 25 years, the cumulative job-to-job contribution to earnings growth differs by approximately 0.1 log-points (about 10%) between the two groups, suggesting that a substantial portion of the income divergence is driven by firm-to-firm mobility. Transitions into or out of non-employment account for only a negligible share of these disparities (Supplementary Figure A.5).

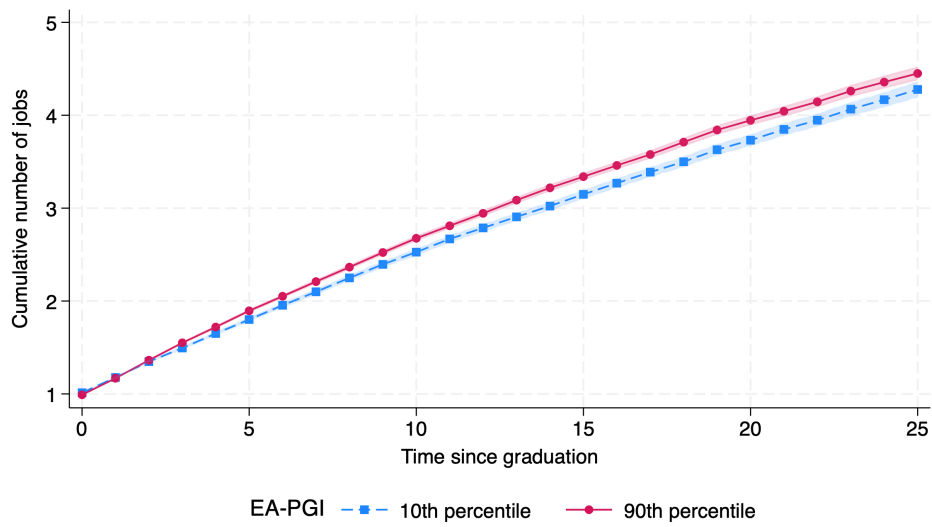
Overall health contributes similarly to the EA-PGI–income association for secondary- and tertiary-educated individuals

Health is a well-established determinant of educational and income disparities,^{24,25} making it a potential intermediate channel between EA-PGI and income. To examine whether the correlation between EA-PGI and income is mediated by disease burden, we calculate the Charlson Comorbidity Index (CCI), which captures the first occur-

Figure 3: Employer mobility and quality of tertiary-educated individuals by EA-PGI level, over time

Panel (a) plots the average number of employment spells, and Panel (b) - the average firm quality index over time since graduation. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Both panels are estimated from a regression of respective outcome on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The estimation sample is restricted to tertiary-educated workers with non-missing firm identifiers ($N = 31,015$) in panel (a) and non-missing firm quality index ($N = 22,733$) in panel (b). The shaded areas correspond to 95% CIs.

(a) Cumulative number of jobs since graduation



(b) Firm quality since graduation

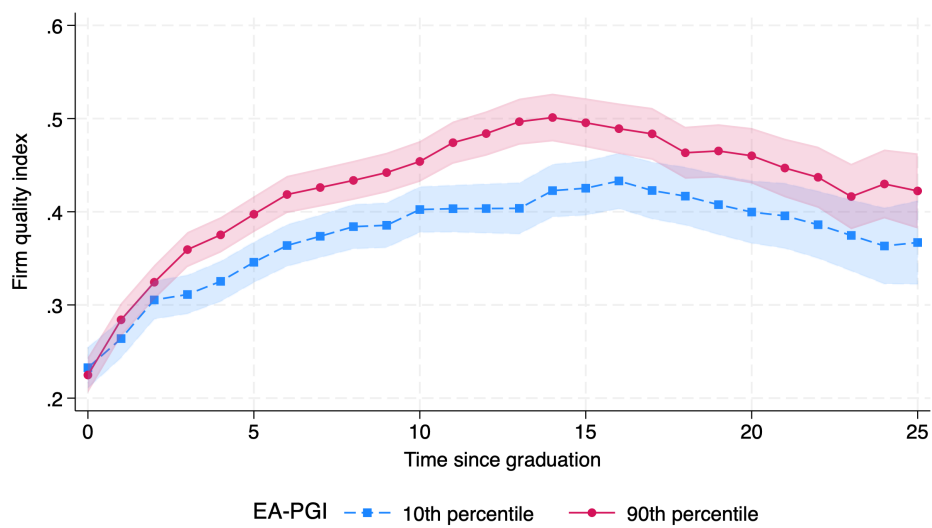
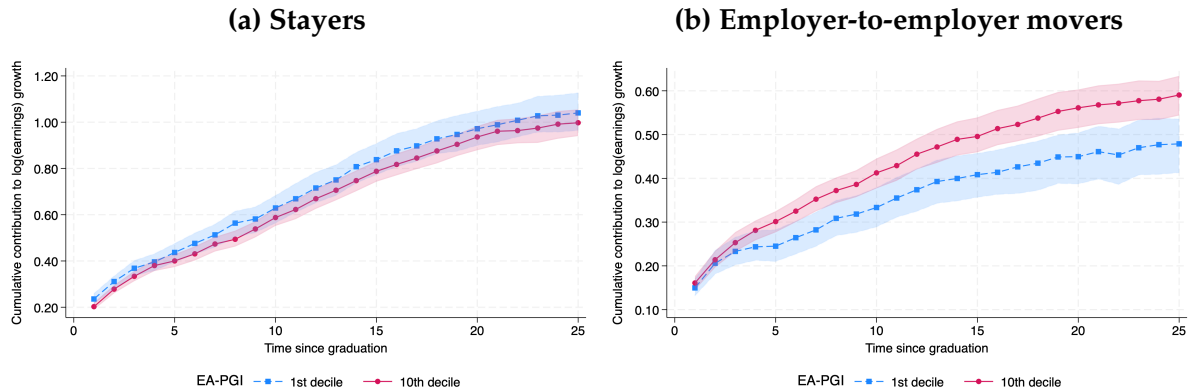


Figure 4: Decomposition of cumulative growth in log annual income of tertiary-educated individuals by EA-PGI level, over time

Panel (a) reports cumulative contribution of within-firm earnings growth and Panel (b) - of employer-to-employer mobility to overall cumulative log earnings growth over time since graduation. The blue line corresponds to 1st and the red - to 10th decile of EA-PGI distribution. The decomposition follows.²³ The sample for the decomposition is restricted to tertiary-educated workers ($N = 32,364$). The shaded areas correspond to 95% CIs obtained via 500 block bootstrap iterations, where at each iteration we sample with re-immission 32,364 whole income histories from the pool of tertiary graduates.



rence of 17 major chronic conditions over the life course^{26,27} and use it as an outcome in our trajectories regression.

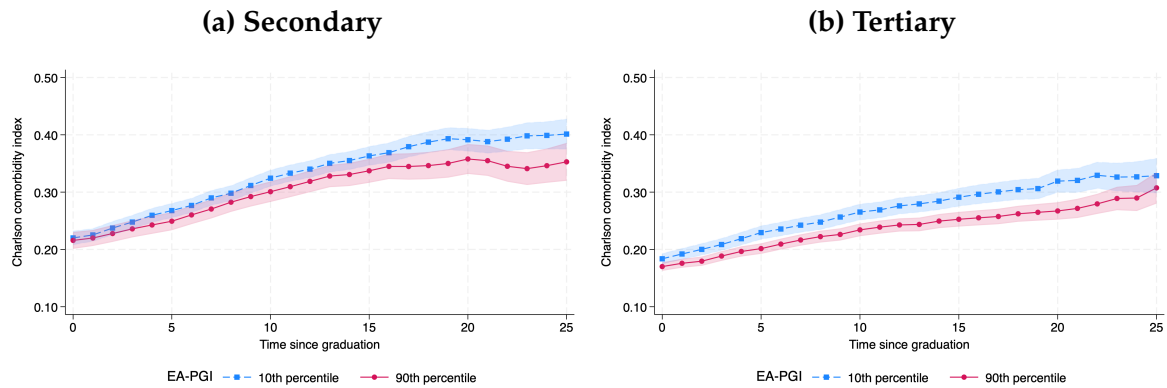
CCI is, on average, lower among individuals with tertiary compared to secondary education (Figure 5), reflecting a well-established lower incidence of major chronic diseases among high-educated individuals.^{28–30} Consistent with this pattern, higher EA-PGI is significantly associated with a lower cumulative disease burden as measured by the CCI. The magnitude of this association is highly similar across education groups, suggesting that the stronger association between EA-PGI and income observed among tertiary-educated individuals is not primarily explained by differences in disease burden. This result is confirmed when controlling for parental EA-PGI (Supplementary Figure A.6).

Parental EA-PGI, in particular that of fathers, predicts the income trajectories among tertiary-educated people

EA-PGI captures both direct genetic effects and indirect effects, such as environmentally mediated influences from parental genetics (i.e., genetic nurture), as well as population stratification and assortative mating.¹² Indeed, the people at the bottom decile of EA-PGI show remarkably lower socioeconomic status (parental education) than those at the top decile (Supplementary Table A.5). To better isolate direct genetic effects, we leverage parental genetic data. Specifically, we calculate maternal and paternal EA-PGI using SNIPAR for 12,871 parent–offspring trios.³¹ Of these, 4,586 were directly genotyped; the rest were imputed based on 4,482 duos and 3,803 sibling pairs.

Figure 5: Average health index by EA-PGI level, over time and by education

Panels (a) and (b) report the results for secondary- ($N = 18,669$) and tertiary-educated workers ($N = 32,294$) with non-missing Charlson Comorbidity Index, respectively. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Average health index estimated from a regression of Charlson Comorbidity Index on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95 % CIs.



We first confirm that the association between offspring EA-PGI and years of education is attenuated by approximately 25 % after accounting for both maternal and paternal EA-PGIs,³¹ and that both parental EA-PGIs are significantly associated with offspring education (Supplementary Table A.7), with effects of similar magnitude.

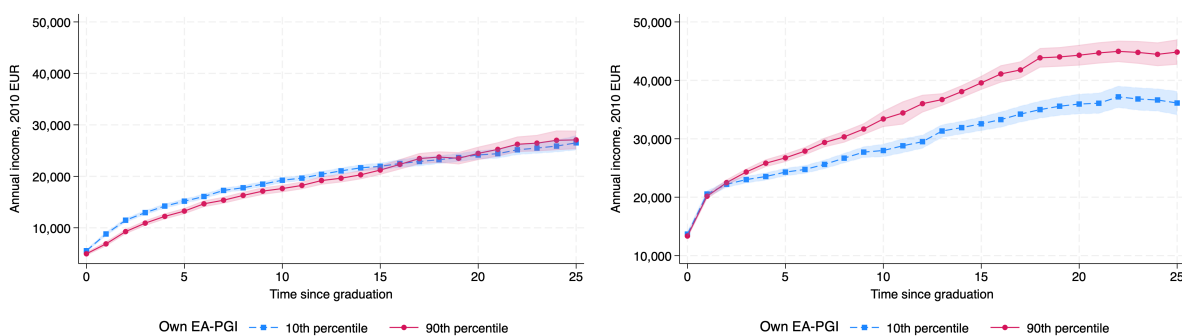
Next, we estimate the impact of offspring EA-PGI on income trajectories while adjusting for both maternal and paternal EA-PGIs. Controlling for parental EA-PGI accounts for income variation attributable to parental genetics as well as family background characteristics unevenly distributed across EA-PGI groups.¹⁵ Among secondary-educated individuals, the income gap across EA-PGI percentiles remains negligible, consistent with our earlier findings (Figure 6a). In contrast, among tertiary-educated individuals, adjusting for parental EA-PGI substantially reduces the income gap between the 90th and 10th percentiles of offspring EA-PGI by approximately 71 % (Figure 6b vs 6d; Table 2; Supplementary Tables A.8 and A.10).

However, unlike what was observed for years of education, only paternal (and not maternal) EA-PGI is significantly associated with offspring income trajectories (Supplementary Table A.8 and Supplementary Figure A.7). Strikingly, the effect of paternal EA-PGI on offspring income is larger than that of the offspring's own EA-PGI (€12,113 [CI €5,509 - €18,717] vs €5,048 [CI €-2,466 - €12,561] per one standard deviation in respective EA-PGI). To contextualize, regardless of offspring EA-PGI, fathers in the 10th percentile of EA-PGI are associated with a cumulative offspring income 25 years after graduation of €330,699 [CI €321,958 - €339,440], compared with €360,946 [CI €350,897 - €370,994] among fathers in the 90th percentile. These effects are consistent across both directly genotyped and imputed trios (Supplementary Table A.9).

Figure 6: Average annual income by EA-PGI level, over time and by education, unconditional and conditional on parental EA-PGI among parent-offspring trios

Panels (a) and (b) plot baseline trajectories among secondary- and tertiary-educated workers, respectively, without controlling for parental EA-PGI. Panels (c) and (d) plot the trajectories among secondary- and tertiary-educated workers after controlling for parental EA-PGI fully interacted with time since graduation. The estimation sample in Panels (a) and (c) is restricted to secondary- ($N = 5,063$) and in Panels (b) and (d) - to tertiary-educated ($N = 7,808$) workers with non-missing parental EA-PGI. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Average income estimated from a regression of annual income on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95% CIs.

(a) Baseline without parental PGI (secondary) (b) Baseline without parental PGI (tertiary)



(c) Controlling for parental PGI (secondary) (d) Controlling for parental PGI (tertiary)

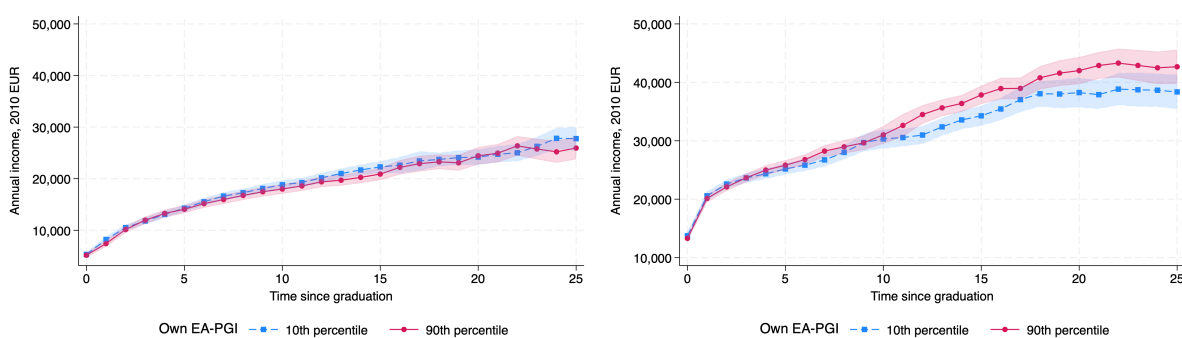


Table 2: Cumulated lifetime income of tertiary-educated individuals by EA-PGI level, conditional on parental EA-PGI among parent-offspring trios

The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles. The estimation sample is restricted to parent-offspring trios with tertiary-educated offspring ($N = 7,808$). Column (1) reports the estimates in the baseline specification without controlling for parental EA-PGI. Columns (2)-(4) report average cumulated lifetime income by own, maternal and paternal EA-PGI percentiles, respectively, conditional on parental EA-PGI. Average lifetime income adjusted by regressing cumulated income on EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income			
	Baseline	Controlling for parental EA-PGI		
	Own EA-PGI (1)	Own EA-PGI (2)	Mother EA-PGI (3)	Father EA-PGI (4)
<i>EA-PGI percentiles</i>				
10th	325 796 (4 015)	339 181 (5 690)	339 879 (4 311)	330 699 (4 459)
20th	332 743 (3 160)	341 428 (4 186)	341 893 (3 286)	335 844 (3 313)
30th	337 737 (2 677)	343 044 (3 232)	343 306 (2 713)	339 552 (2 664)
40th	341 952 (2 413)	344 408 (2 611)	344 571 (2 388)	342 722 (2 338)
50th	345 756 (2 328)	345 639 (2 320)	345 677 (2 313)	345 561 (2 309)
60th	349 747 (2 413)	346 930 (2 399)	346 834 (2 466)	348 447 (2 543)
70th	354 053 (2 685)	348 323 (2 884)	348 007 (2 823)	351 537 (3 019)
80th	359 280 (3 197)	350 014 (3 799)	349 535 (3 487)	355 397 (3 808)
90th	365 948 (4 022)	352 172 (5 198)	351 466 (4 500)	360 946 (5 126)
Obs.	7 808	7 808	7 808	7 808

Discussion

In this study, we examine the mechanisms through which known genetic factors associated with educational attainment affect socioeconomic inequalities. To our knowledge, this is the first large study to investigate direct and indirect (parental) EA-PGI effects on longitudinal income trajectories; prior work has used small samples or cross-sectional designs, which limits our understanding of how inequality accumulates over the life cycle and is transmitted. It is important to recognize that these studies have contributed to an aversion, anger and even fear of studying genetics in social stratification research.^{32,33} We include a document with Frequently Asked Questions (FAQs) to offer an accessible explanation of what our study does and, importantly, does not, find and how the results should be interpreted.

Labor income is the focus of our analysis because it provides a well-defined, downstream indicator of socioeconomic status, reflecting differences in education, health, and skills. Together with government transfers, labor income constitutes the main source of monetary inflows accumulated over the life course and is therefore a fundamental determinant of wealth.³⁴ Moreover, income inequality is a central concern in both economics²¹ and epidemiology,^{35–37} which allows us to incorporate genetics into well-established analytical frameworks.

We show that workers earn similar income upon labor market entry, irrespective of their EA-PGI level. After entry, however—and only among tertiary-educated individuals—income differences between those at the 90th and 10th percentiles of EA-PGI widen steadily over time, culminating in a 14% cumulative income gap 25 years after graduation. The fact that EA-PGI predicts income trajectories only among tertiary-educated individuals reflects two forces. First, individuals with relatively high EA-PGI experience higher economic returns to ability. Second, EA-PGI influences educational sorting itself, shaping the types and levels of education pursued. While we cannot quantify exactly the relative importance of the two channels (sorting and economic returns to higher EA-PGI), our analysis based on the AKM model estimation¹⁸ offers additional insights on the role of workers' ability and of employers in explaining the observed income differences.

We show that a measure of labor market productivity (or ability) is indeed positively associated with EA-PGI, but again only for tertiary-educated workers. This relationship persists even after controlling for detailed education-related characteristics (education field and type of academic institution). The finding that EA-PGI relates to labor market productivity only among highly educated individuals raises the question of whether some high-ability individuals with only secondary education might benefit from additional educational support or incentives to pursue tertiary studies. While we cannot provide a definitive answer to this, we note that any such discus-

sion must consider the role of family socioeconomic background³⁸ and return to the importance of family background later.

Our trajectory framework also allows us to document a new mechanism through which genetic endowments affect income: employer sorting. We show that there is no evidence of sorting into the first employer on the basis of EA-PGI. This is consistent with labor market entrants having limited information about firms and match quality, and with firms similarly knowing little about the productivity of new hires. Despite the absence of initial sorting, we show that EA-PGI subsequently matters for job mobility among university graduates, who are more likely to move to higher-paying employers. This provides novel evidence that employer learning about workers' productivity and job mobility³⁹ are important channels underlying the genetic gradient in income, but only for jobs accessed through higher-education degrees. This result complements a literature in economics that shows that employer-specific characteristics and pay policies play a crucial role in explaining wage inequality beyond occupation alone (Supplementary Figure A.1).^{20,21}

Because health is a well-documented determinant of educational and income disparities, we also examine whether it mediates the EA-PGI–income relationship. Previous studies have shown that EA-PGI is associated with multiple diseases.¹³ We find that the association between EA-PGI and cumulative disease burden is similar across education groups. This suggests that differences in health do not explain the stronger EA-PGI–income association among the tertiary-educated.

By Mendel's laws, parents of children with higher EA-PGI must have higher EA-PGI themselves, and higher EA-PGIs result in higher parental education. This pattern is confirmed in our data: individuals in the 10th percentile of EA-PGI come from substantially lower socioeconomic backgrounds than those in the 90th percentile (Supplementary Table A.5). We therefore examine the role of parental EA-PGIs in accounting for the observed income gap across offspring's EA-PGI levels. Adjusting for parental EA-PGI reduced the offspring income gap by approximately 71 %, indicating that indirect effects play an important role in shaping income development.

Notably, most of this attenuation is driven by the father's EA-PGI rather than the mother's, a pattern not observed for educational attainment itself. This may reflect differential mechanisms: maternal resources, such as time and cognitive engagement, are particularly relevant for early-life development and educational choices, whereas paternal resources may become more influential during labor-market entry, through direct income support, occupational networks, or knowledge transmission about high-quality firms.⁴⁰ These results are especially striking given Finland's relatively high level of gender equality in employment opportunities.⁴¹

This study advances the literature in several ways. First, we analyze a substantially larger genotyped sample, including 12,871 parent–offspring trios, than previous

studies, linked to decades of high-quality administrative data. This allows us to produce precise estimates of the genetic drivers of socioeconomic disparities over the life cycle. Second, by exploiting repeated measurements of workers linked to employers, we apply panel-data methods that account for worker and firm heterogeneity, which in turn allows us to provide novel evidence on the role of genetics in explaining the production of inequality in the labor markets. This allows us to decompose income variation into worker- and firm-level components, bridging genetics and modern labor economics.^{19,42,43} Lastly, we leverage continuous, rich information from individual tax records, avoiding self-report bias and measurement error that may affect survey-based measures of income.

Our study also has limitations. First, our sample consists solely of Finnish individuals who tend to be positively selected compared to the general population. This limits the generalizability to other ancestries or institutional settings. To address the fact that genotyped individuals are not fully representative of the Finnish population, we applied inverse-probability weighting to make the sample representative of all fresh graduates in the same calendar period.⁴⁴ Weighted analyses produced results highly consistent with our main estimates (Supplementary Table A.11). Second, EA-PGI should not be interpreted as a pure “genetic endowment for educational attainment,” as it may capture pathways correlated with educational attainment, such as lifestyle or social factors, that are modifiable and context-dependent. As such, changes in these underlying environments or social conditions could alter the observed relationship between EA-PGI and income. Lastly, as noted, the large attenuation we observe when controlling for parental EA-PGI indicates that indirect effects—particularly those operating through fathers—play an important role in shaping income differences. However, our setting does not allow us to disentangle the (parental) genetic and alternative environmental components underlying this result, which we leave to future work.

Taken together, our results indicate that EA-PGI is a robust predictor of income across the working life course, but only among tertiary-educated individuals. These effects operate primarily through individuals with higher EA-PGI experiencing faster and more frequent labor-market transitions to higher-quality employers, which in turn offer higher wages and thereby further amplify income inequality. Moreover, a substantial share of the EA-PGI–income association arises through indirect paternal genetic effects, suggesting that in the relatively recent generations studied (average graduation year 2000), fathers continued to play an important role in shaping offspring labor-market trajectories through untransmitted genetic factors.

Materials and Methods

1.1 Data Sources

Genetic data. The genotyped sample was obtained from Finnish biobanks and consists of individuals who provided consent for research use of their blood samples. Participants in our dataset were drawn from several population-based epidemiological cohorts: 27,135 from the THL and 23,921 from the Blood Donor study.¹

To quantify the genetic contribution to educational attainment, we constructed a polygenic index for years of education (EA-PGI) based on the largest genome-wide association study of educational attainment to date,¹³ excluding Finnish cohorts to avoid overfitting. In our analysis sample, the PGI explained 7.1 %² of the variance in years of education (Supplementary Table A.12)—slightly lower than estimates from previous studies,⁴⁵ possibly due to differences in cohort composition or educational classification.

Our sample tends to be positively selected (younger, better educated, with a higher share of women) compared to the population of fresh graduates. To analyze whether sample composition is a likely driver of our results, we apply an inverse probability weighting approach to make the sample representative of the graduates population.⁴⁴ Supplementary Table A.2 provides details on the re-weighting procedure and shows that it is effective in making the sample representative of the general population along the central dimensions that are initially unbalanced (including entry income, which we do not re-weight in our routine). Using these weights when estimating the income trajectories yields results that are qualitatively very similar to those in our main analysis, both when using the full genotyped sample and when using the family trios (Supplementary Table A.11).

Register data. We link the genotyped data to administrative registers from Statistics Finland (FOLK databases), covering the years 1987–2019. In addition, we utilize these register data independently, as they include the entire population of individuals permanently residing in Finland at the end of each year. The registers provide detailed information on employment histories, which allows us to identify the main employer at the end of each calendar year. They also contain income by source, from which we identify yearly labor income. We include people with zero income in the income analysis, thereby avoiding conditioning on employment. Given the long time series used in the analysis, all monetary values are deflated to 2010 EUR to account for inflation

¹THL biobank website: <http://www.thl.fi/biobank>. Blood service biobank website: <https://www.veripalvelu.fi/en/biobank/>.

²Computed as incremental R^2 , following,¹³ and controlling for gender fully interacted with year of birth indicators, biobank indicator and first 10 genetic principal components

and ensure comparability across time. The choice of the reference year is arbitrary; we choose 2010 because it lies midway between 2000 – the average graduation year in the sample – and the last follow-up year, 2019.

The registers further include demographic variables (gender, year of birth) and detailed educational information (highest degree, 4-digit field, school or institution ID, vocational vs. academic track), occupation, and industry codes.

Health registers. Health outcomes are obtained from two nationwide registers maintained by the Finnish Institute for Health and Welfare: the Care Register for Health Care (Hilmo) and the Register of Primary Health Care Visits (Avohilmo). In this study, Hilmo covers inpatient visits, operations, and specialized outpatient visits for the period to 1987–2024, when diagnoses follow ICD-9 and ICD-10 coding. Avohilmo, which uses ICD-10, covers primary care outpatient visits since 2011. For individuals absent from Hilmo, Avohilmo is used to complement the coverage. Both registers contain patient identifiers, care episode details, and one or more discharge diagnoses.

1.2 Methods

Polygenic Indices

We construct polygenic indices (PGIs) by aggregating single nucleotide polymorphisms (SNPs), common genetic variants identified in Genome-Wide Association Studies (GWAS) as predictive of years of education and health-related outcomes (see e.g.⁷). SNPs are linearly combined using GWAS-derived effect sizes as weights, producing out-of-sample PGIs predictive of each trait of interest.

Our primary measure is the PGI for educational attainment (EA-PGI), standardized to mean zero and unit variance. Its distribution across education groups is shown in Supplementary Figure A.2. To control for ancestry and population stratification, we compute the first ten principal components of the genetic data and include them as covariates in all analyses.

Worker Ability and Firm Quality Measurement

To operationalize worker ability (or productivity) on the labor market θ_i , and firm quality ψ_J we estimate an (AKM) regression:^{18,21}

$$y_{it} = \mathbf{X}_{it}\beta + \psi_{J(i,t)} + \theta_i + \varepsilon_{it} \quad (1)$$

where y_{it} denotes log monthly labor income³ for individual i in year t ; \mathbf{X}_{it} includes

³The outcome in (1) is monthly income, defined as annual earnings divided by number of months worked. Results are robust to using hourly wages derived from the Structure of Earnings Register (SES),

education fully interacted with calendar year and cubic age polynomial; $\psi_{J(i,t)}$ represents firm fixed effects; and θ_i are worker fixed effects. Estimated $\hat{\theta}_i$ provides a measure of worker productivity (unobserved heterogeneity) and $\hat{\psi}_{J(i,t)}$ offers firm-specific wage premia, interpreted as firm quality. Workers with higher fixed effects earn more across firms relative to a reference worker, holding observables constant. Similarly, firms with higher AKM fixed effects pay higher average wages, consistent with higher productivity serving as a proxy firm quality, as they capture persistent wage differences across firms after accounting for worker characteristics.

Except for the additive separability of firm and worker fixed effects, no functional form assumptions are made on either of the fixed effects, which are estimated non-parametrically. The estimation sample includes all full-time employees aged 20–60. The main employer is defined as the highest-paying ongoing job at year-end. To ensure labor market attachment, we restrict to workers earning at least 50% of the national median monthly income. The model is estimated by using the sample of firms linked by workers job-to-job transitions, also called connected set.¹⁹ Firms that do not experience hires or separations during their observation period are not part of the connected set and do not have an associated firm fixed effect.

For the estimation of the model, employment spells in very small firms (<5 employees) or shorter than four months are excluded. The resulting panel comprises 3.7 million workers, 31.6 million person-year observations, and 177.0 thousand firms. The AKM estimation is performed separately for two periods (1987–2003 and 2004–2019) due to computational constraints. Correlations of worker and firm fixed effects across the two periods are reasonably high (Supplementary Figure A.8).

We standardize worker and firm fixed effects to have zero mean and unit variance, and link them to the genotyped sample. Supplementary Tables A.3 and A.4 provide summary statistics for the sample used in estimating AKM, and standard statistics and income variance decomposition following AKM estimations.

Income and Health Trajectory Model

To study how genetic predispositions affect income over the life cycle, we estimate:

$$y_{icmt} = \alpha + \tau_c + \tau_m + \beta_i PGI_i + \gamma X_i + \varepsilon_{icmt} \quad (2)$$

where y_{icmt} is labor income of individual i or the Charlson Comorbidity Index (see below), in birth cohort c , calendar year m , and number of years since graduation t . PGI_i is standardised EA-PGI; τ_c and τ_m are cohort and year effects; and X_i includes ten genetic principal components, gender, and biobank indicator (THL or Blood Donors).

which covers the whole public sector and a sample of about half of the private sector. To maximize sample size and coverage, we use monthly earnings as our baseline income measure.

The coefficients of interest are β_t , which capture the income differential for EA-PGI over time since graduation. Since PGIs are randomly assigned at conception, β_t has a causal interpretation, conditional on adequate control for population stratification via X_i . Any residual stratification would lead the coefficients to capture a compound effect of own genetics and other environmental factors. While assigning sign to the bias is not immediate, we believe that it is reasonable to consider our estimates as conservative lower bounds of true causal effects of own EA-PGI.⁴

Our model income trajectories specifications include continue EA-PGI evaluated at 10th and 90th percentiles of EA-PGI distribution. The results are similar when specifying EA-PGI via deciles (Supplementary Figure A.3). Finally, when computing cumulated lifetime income we discount income by a 3% interest rate, computing its present value upon graduation. In the computation, we sum over all discounted income rows (including zeros) between graduation year and up to 25 years later.

Income decomposition by EA-PGI group and over time since graduation

We implement the approach by²³ to decompose the log-income growth separately by EA-PGI group (1st and 10th deciles).

Each year since graduation t , workers are partitioned into one of four groups: *stayers* (workers who stay with the same employer); *employer-to-employer transitions* (workers who change firm); *entrants from non-employment* (hires from nonemployment); *exitters to non-employment* (incumbent workers separating to nonemployment). The average income growth between $t - 1$ and t is decomposed into four weighted contributions based on the four worker types (weighted by the share of workers in each worker type). Since entrants and exitters move between employment and non-employment, their contribution is obtained by comparing their average income to that of the workers who are continuously employed in the time period.

In line with,²³ and confirmed by our analysis, job-to-job movers' transitions are associated with large earnings gains for individuals. Moreover, the entrants from nonemployment earn substantially less than the continuously-employed workers to which their salary is compared to. Hence, their entry into employment lessens (subtracts from) the average earnings and their contribution to the average earnings growth is negative. The opposite occurs for exitters to nonemployment: they also tend to earn less than the continuously-employed workers, but because these low-paying jobs dissolve, this contributes positively to the earnings growth.

We present results by cumulating the income growth components over $t = 1, \dots, 25$.

⁴The PGI captures only the contribution of common genetic variants identified in external GWAS, not the full genetic architecture of education.

Charlson Comorbidity Index

The Charlson Comorbidity Index (CCI)^{26,27} assigns fixed weights to comorbid conditions associated with higher mortality risk. The index is the weighted sum of an individual’s comorbidities, with weights derived from Cox regression models.

We compute CCI scores using the ICCI R package,⁴⁶ which implements the ICD-9 and ICD-10 coding⁴⁷ via R’s comorbidity package.⁴⁸⁵ For each individual, we compute cumulative CCI scores by age, recalculating the index at successive cutoffs (0–19, 0–20, . . . , up to 0–50 years).

Definitions and Sample Restrictions. For the trajectory-based analyses, we construct a panel of all genotyped adults with either secondary or tertiary qualification, followed from year of graduation onward between 1987 and 2019. To ensure that individuals in our sample have completed their education phase, we remove those with secondary degree that have not been observed past age 30; people that obtain tertiary degree before age 30 are retained in the sample. The final sample includes 51,056 individuals and 963,715 person-year observations (see Supplementary Table A.1).

Acknowledgments

We would like to thank Adrian Adermon, Mikko Arvas, Pietro Biroli, Rosa Cheesman, Lena Hensvik, Paul Hufe, Andrea Ichino, Francis Kramarz, Martin Nybom, Gerard van den Berg, and participants to the European Social Science Genetics Network Conference 2023 (Bologna), Nordic Summer Institute 2025 (Uppsala), Workshop on Inequality and education 2025 (Modena), Workshop in Wages, Employment, and Inequality 2025 (Helsinki), for useful comments and discussions.

We thank all study participants for their generous participation in biobank research, and we are grateful to Blood Service Biobank and to THL Biobank for issuing the data permits to utilize the genetic data (BSB permit no. 005-2020 and THLBB2023_47). The project fall under IRB ethical approval HUS/1311/2020, Statistics Finland data permit TK/125/07.03.00/2020, and Findata permit THL/4014/14.02.00/2022. Lombardi gratefully acknowledges funding from the Research Council of Finland (decision no. 350399) and Yrjö Jahnsson Foundation (decision no. 20247803).

⁵The package accommodates multiple ICD versions. Source code: [ICCI GitHub repository](#).

References

- [1] J. E. Roemer and A. Trannoy. “Equality of Opportunity: Theory and Measurement”. *Journal of Economic Literature* 54.4 (2016), 1288–1332. ISSN: 0022-0515. DOI: [10.1257/jel.20151206](https://doi.org/10.1257/jel.20151206). URL: <https://pubs.aeaweb.org/doi/10.1257/jel.20151206> (visited on 12/12/2024).
- [2] O. Ashenfelter and A. B. Krueger. “Estimates of the Economic Return to Schooling from a New Sample of Twins”. *The American Economic Review* 84.5 (1994), 1157–1173.
- [3] R. Chetty, N. Hendren, P. Kline, and E. Saez. “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”. *Quarterly Journal of Economics* 129.4 (2014), 1553–1623.
- [4] R. Chetty, M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler. “The association between income and life expectancy in the United States, 2001–2014”. *Jama* 315.16 (2016), 1750–1766.
- [5] J. D. Bundy, K. T. Mills, H. He, T. A. LaVeist, K. C. Ferdinand, J. Chen, and J. He. “Social determinants of health and premature death among adults in the USA from 1999 to 2018: a national cohort study”. *The Lancet Public Health* 8.6 (2023), e422–e431.
- [6] F. Cunha and J. Heckman. “The Technology of Skill Formation”. *American Economic Review* 97.2 (2007), 31–47.
- [7] P. Biroli, T. Galama, S. von Hinke, H. Van Kippersluis, C. A. Rietveld, and K. Thom. “The Economics and econometrics of gene–environment interplay”. *Review of Economic Studies* (2025), rdaf034.
- [8] K. P. Harden and P. D. Koellinger. “Using genetics for social science”. *Nature Human Behaviour* 4.6 (2020), 567–576. ISSN: 2397-3374. DOI: [10.1038/s41562-020-0862-5](https://doi.org/10.1038/s41562-020-0862-5). URL: <https://www.nature.com/articles/s41562-020-0862-5> (visited on 12/13/2024).
- [9] A. Abdellaoui, L. Yengo, K. J. Verweij, and P. M. Visscher. “15 years of GWAS discovery: realizing the promise”. *The American Journal of Human Genetics* 110.2 (2023), 179–194.
- [10] D. Barth, N. W. Papageorge, and K. Thom. “Genetic Endowments and Wealth Inequality.pdf”. *Journal of Political Economy* 128.4 (2020), 1474–1522.
- [11] N. W. Papageorge and K. Thom. “Genes, education, and labor market outcomes: evidence from the health and retirement study”. *Journal of the European Economic Association* 18.3 (2020), 1351–1399.
- [12] A. Kong, G. Thorleifsson, M. L. Frigge, B. J. Vilhjalmsdottir, A. I. Young, T. E. Thorgeirsson, S. Benonisdottir, A. Oddsson, B. V. Halldorsson, G. Masson, et al. “The nature of nurture: Effects of parental genotypes”. *Science* 359.6374 (2018), 424–428.
- [13] A. Okbay et al. “Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals”. *Nature Genetics* 54.4 (2022), 437–449. DOI: [10.1038/s41588-022-01016-z](https://doi.org/10.1038/s41588-022-01016-z). URL: <http://dx.doi.org/10.1038/s41588-022-01016-z>.
- [14] T. Tan, H. Jayashankar, J. Guan, S. M. Nehzati, M. Mir, M. Bennett, E. Agerbo, R. Ahlskog, V. P. de Andrade Anapaz, B. O. Åsvold, S. Benonisdottir, L. Bhatta, D. I. Boomsma, B. Brumpton, A. Campbell, C. F. Chabris, R. Cheesman, Z. Chen, E. de Geus, E. A. Ehli, A. G. Elnahas, E. B. R. Team, F. Authors, A. Ganna, A. Giannelis, L. Hakaste, A. F. Hansen, A. Havdahl, C. Hayward, J.-J. Hottenga, M. A. Houmark, K. Hveem, J. Kaprio, A. Langhammer, A. Latvala, J. J. Lee, M.

- Lehtovirta, L. Li, K. Lin, R. K. Linner, S. Lombardi, N. G. Martin, M. McGue, S. E. Medland, A. Metspalu, B. L. Mitchell, G. Ni, I. M. Nolte, M. T. Oetjens, S. Oskarsson, T. Palviainen, R. B. Prasad, A. Reigo, K. Reis, J. Sidorenko, K. Silventoinen, H. Snieder, T. Tuomi, B. J. Vilhjalmsson, R. G. Walters, E. A. Willoughby, B. S. Winsvold, E. Ystrom, J. Flint, L. Yengo, P. M. Visscher, A. Kong, E. M. Tucker-Drob, R. Border, D. Cesarini, P. Turley, A. Okbay, D. J. Benjamin, and A. S. Young. “Family-GWAS reveals effects of environment and mating on genetic associations”. *medRxiv* (2024). preprint. DOI: [10.1101/2024.10.01.24314703](https://doi.org/10.1101/2024.10.01.24314703).
- [15] A. Rustichini, W. G. Iacono, J. J. Lee, and M. McGue. “Educational attainment and intergenerational mobility: A polygenic score analysis”. *Journal of Political Economy* 131.10 (2023), 2724–2779.
- [16] L. S. Carvalho. “Genetics and Socioeconomic Status: Some Preliminary Evidence on Mechanisms”. *Journal of political economy microeconomics* 3.3 (2025), 429–476.
- [17] E. T. Akimova, T. Wolfram, X. Ding, F. C. Tropf, and M. C. Mills. “Polygenic prediction of occupational status GWAS elucidates genetic and environmental interplay in intergenerational transmission, careers and health in UK Biobank”. *Nature Human Behaviour* 9.2 (2025), 391–405.
- [18] J. M. Abowd, F. Kramarz, and D. N. Margolis. “High wage workers and high wage firms”. *Econometrica* 67.2 (1999), 251–333. URL: <http://onlinelibrary.wiley.com/doi/10.1111/1468-0262.00020/abstract> (visited on 09/19/2016).
- [19] P. M. Kline. “Firm wage effects”. NBER Working Paper 33084. 2025.
- [20] D. Card, J. Heining, and P. Kline. “Workplace heterogeneity and the rise of West German wage inequality”. *The Quarterly journal of economics* 128.3 (2013), 967–1015.
- [21] P. Kline. “Firm wage effects”. *Handbook of Labor Economics* 5 (2024), 115–181.
- [22] B. Wang, J. R. Baldwin, T. Schoeler, R. Cheesman, W. Barkhuizen, F. Dudbridge, D. Bann, T. T. Morris, and J.-B. Pingault. “Robust genetic nurture effects on education: A systematic review and meta-analysis based on 38,654 families across 8 cohorts”. *The American Journal of Human Genetics* 108.9 (2021), 1780–1791.
- [23] J. Hahn, H. Hyatt, and H. Janicki. “Job ladders and growth in earnings, hours, and wages”. *European Economic Review* 133 (2021), 103654. DOI: [10.1016/j.euroecorev.2021.103654](https://doi.org/10.1016/j.euroecorev.2021.103654). URL: <http://dx.doi.org/10.1016/j.euroecorev.2021.103654>.
- [24] A. V. J. Pallesen, J. O. Mierau, F. K. Christensen, and L. H. Mortensen. “Educational and income inequalities across diseases in Denmark: a register-based cohort study”. *The Lancet Public Health* 9.11 (2024), e916–e924. DOI: [10.1016/S2468-2667\(24\)00128-2](https://doi.org/10.1016/S2468-2667(24)00128-2). URL: [https://doi.org/10.1016/S2468-2667\(24\)00128-2](https://doi.org/10.1016/S2468-2667(24)00128-2).
- [25] D. B. Newman, A. M. Gordon, and W. B. Mendes. “Income and education show distinct links to health and happiness in daily life”. *Nature Human Behaviour* (2025). Published online 8 Aug 2025. DOI: [10.1038/s41562-025-02264-9](https://doi.org/10.1038/s41562-025-02264-9). URL: <https://doi.org/10.1038/s41562-025-02264-9>.
- [26] M. E. Charlson, P. Pompei, K. L. Ales, and C. R. MacKenzie. “A new method of classifying prognostic comorbidity in longitudinal studies: development and validation”. *Journal of Chronic Diseases* 40.5 (1987), 373–383. DOI: [10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8).
- [27] R. A. Deyo, D. C. Cherkin, and M. A. Ciol. “Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases”. *Journal of Clinical Epidemiology* 45.6 (1992), 613–619. DOI: [10.1016/0895-4356\(92\)90133-8](https://doi.org/10.1016/0895-4356(92)90133-8).

- [28] E. Agardh, P. Allebeck, J. Hallqvist, T. Moradi, and A. Sidorchuk. “Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis”. *International Journal of Epidemiology* 40.3 (2011), 804–818. DOI: [10.1093/ije/dyr029](https://doi.org/10.1093/ije/dyr029). URL: <https://doi.org/10.1093/ije/dyr029>.
- [29] T. Tillmann, J. Vaucher, A. Okbay, H. Pikhart, A. Peasey, R. Kubinova, A. Pajak, A. Tamosiunas, S. Malyutina, F. P. Hartwig, K. Fischer, G. Veronesi, T. Palmer, J. Bowden, G. Davey Smith, M. Bobak, and M. V. Holmes. “Education and coronary heart disease: mendelian randomisation study”. *BMJ (Clinical research ed.)* 358 (2017), j3542. DOI: [10.1136/bmj.j3542](https://doi.org/10.1136/bmj.j3542). URL: <https://doi.org/10.1136/bmj.j3542>.
- [30] S. Vaccarella, D. Georges, F. Bray, O. Ginsburg, H. Charvat, P. Martikainen, H. Brønnum-Hansen, P. Deboosere, M. Bopp, M. Leinsalu, B. Artnik, V. Lorenzoni, E. de Vries, M. Marmot, P. Vineis, J. Mackenbach, and W. Nusselder. “Socioeconomic inequalities in cancer mortality between and within countries in Europe: a population-based study”. *The Lancet Regional Health - Europe* 25 (2023), 100551. DOI: [10.1016/j.lanepe.2022.100551](https://doi.org/10.1016/j.lanepe.2022.100551). URL: <https://doi.org/10.1016/j.lanepe.2022.100551>.
- [31] A. I. Young, S. M. Nehzati, S. Benonisdottir, et al. “Mendelian imputation of parental genotypes improves estimates of direct genetic effects”. *Nature Genetics* 54.6 (2022), 897–905. DOI: [10.1038/s41588-022-01085-0](https://doi.org/10.1038/s41588-022-01085-0). URL: <https://doi.org/10.1038/s41588-022-01085-0>.
- [32] D. Martschenko, S. Trejo, and B. W. Domingue. “Genetics and education: Recent developments in the context of an ugly history and an uncertain future”. *AERA Open* 5.1 (2019), 2332858418810516.
- [33] M. C. Mills. “Sociogenomics: theoretical and empirical challenges of integrating molecular genetics into sociological thinking”. *Handbook of sociological science*. Edward Elgar Publishing, 2022, 250–270.
- [34] S. E. Black, P. J. Devereux, F. Landaud, and K. G. Salvanes. “The (un) importance of inheritance”. *Journal of the European Economic Association* 23.3 (2025), 1060–1094.
- [35] J. W. Lynch, G. Davey Smith, G. A. Kaplan, and J. S. House. “Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions”. *BMJ* 320.7243 (2000), 1200–1204. DOI: [10.1136/bmj.320.7243.1200](https://doi.org/10.1136/bmj.320.7243.1200).
- [36] K. E. Pickett and R. G. Wilkinson. “Income inequality and health: A causal review”. *Social Science & Medicine* 128 (2015), 316–326. DOI: [10.1016/j.socscimed.2014.12.031](https://doi.org/10.1016/j.socscimed.2014.12.031).
- [37] H. Schwandt, J. Currie, M. Bär, J. Banks, P. Bertoli, A. Bütikofer, S. Cattan, B. Z. Y. Chao, C. Costa, L. González, V. Grembi, K. Huttunen, R. Karadakic, L. Kraftman, S. Krutikova, S. Lombardi, P. Redler, C. Riumallo-Herl, A. Rodríguez-González, K. G. Salvanes, P. Santana, J. Thuilliez, E. van Doorslaer, T. van Ourti, J. K. Winter, B. Wouterse, and A. Wuppermann. “Inequality in mortality between Black and White Americans by age, place, and cause, and in comparison to Europe, 1990–2018”. *Proceedings of the National Academy of Sciences of the United States of America* 118.40 (2021). Published online 28 Sept 2021, e2104684118. DOI: [10.1073/pnas.2104684118](https://doi.org/10.1073/pnas.2104684118).
- [38] A. Ichino, A. Rustichini, and G. Zanella. “College, cognitive ability, and socioeconomic disadvantage: policy lessons from the UK in 1960-2004”. *Review of Economic Studies* (forthcoming) (2024).
- [39] H. S. Farber and R. Gibbons. “Learning and wage dynamics”. *The Quarterly Journal of Economics* 111.4 (1996), 1007–1047.

- [40] D. Del Boca, C. Flinn, and M. Wiswall. “Household choices and child development”. *Review of Economic Studies* 81.1 (2014), 137–185.
- [41] W. World Economic Forum. “Global Gender Gap Report 2024”. Tech. rep. Accessed: 2025-11-12. World Economic Forum, 2024. URL: <https://www.weforum.org/publications/global-gender-gap-report-2024/>.
- [42] D. Card, A. R. Cardoso, J. Heining, and P. Kline. “Firms and Labor Market Inequality: Evidence and Some Theory”. *Journal of Labor Economics* 36 (S1 2018), S13–S70.
- [43] J. Song, D. J. Price, F. Guvenen, and N. Bloom. “Firming up inequality”. *The Quarterly Journal of Economics* 134.1 (2018), 1–50.
- [44] N. M. Davies, M. Dickson, G. Davey Smith, G. J. Van Den Berg, and F. Windmeijer. “The causal effects of education on health outcomes in the UK Biobank”. *Nature Human Behaviour* 2.2 (2018), 117–125. ISSN: 2397-3374. DOI: [10.1038/s41562-017-0279-y](https://doi.org/10.1038/s41562-017-0279-y). URL: <https://www.nature.com/articles/s41562-017-0279-y> (visited on 03/17/2025).
- [45] J. J. Lee, R. Wedow, A. Okbay, et al. “Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals”. *Nature Genetics* 50.8 (2018). Published online 23 Jul 2018, 1112–1121. DOI: [10.1038/s41588-018-0147-3](https://doi.org/10.1038/s41588-018-0147-3). URL: <https://doi.org/10.1038/s41588-018-0147-3>.
- [46] K. E. Detroit. *ICCI: Calculates the Charlson Comorbidity Index*. R package version 2.3.1. 2024. URL: <https://CRAN.R-project.org/package=ICCI>.
- [47] H. Quan, V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.-C. Luthi, L. D. Saunders, C. A. Beck, L. Feasby, and W. A. Ghali. “Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data”. *Medical Care* 43.11 (2005), 1130–1139. DOI: [10.1097/01.mlr.0000182534.19832.83](https://doi.org/10.1097/01.mlr.0000182534.19832.83). URL: <https://doi.org/10.1097/01.mlr.0000182534.19832.83>.
- [48] A. Gasparini. “comorbidity: An R package for computing comorbidity scores”. *Journal of Open Source Software* 3.23 (2018), 648. DOI: [10.21105/joss.00648](https://doi.org/10.21105/joss.00648). URL: <https://doi.org/10.21105/joss.00648>.

A Supplementary Figures and Tables

A.1 Supplementary Figures

Figure A.1: Standard deviation in wages from alternative wage models

Measures of dispersion in actual and residualized wages for full-time workers, using the sample described in “Worker Ability and Firm Quality Measurement” section. Residual wages are obtained by sequentially adjusting worker-level wages for the covariates listed in the legend using linear models. The Mincer specification adjusts for education indicators and a cubic age profile. All models are estimated separately by calendar year. 95% confidence intervals.

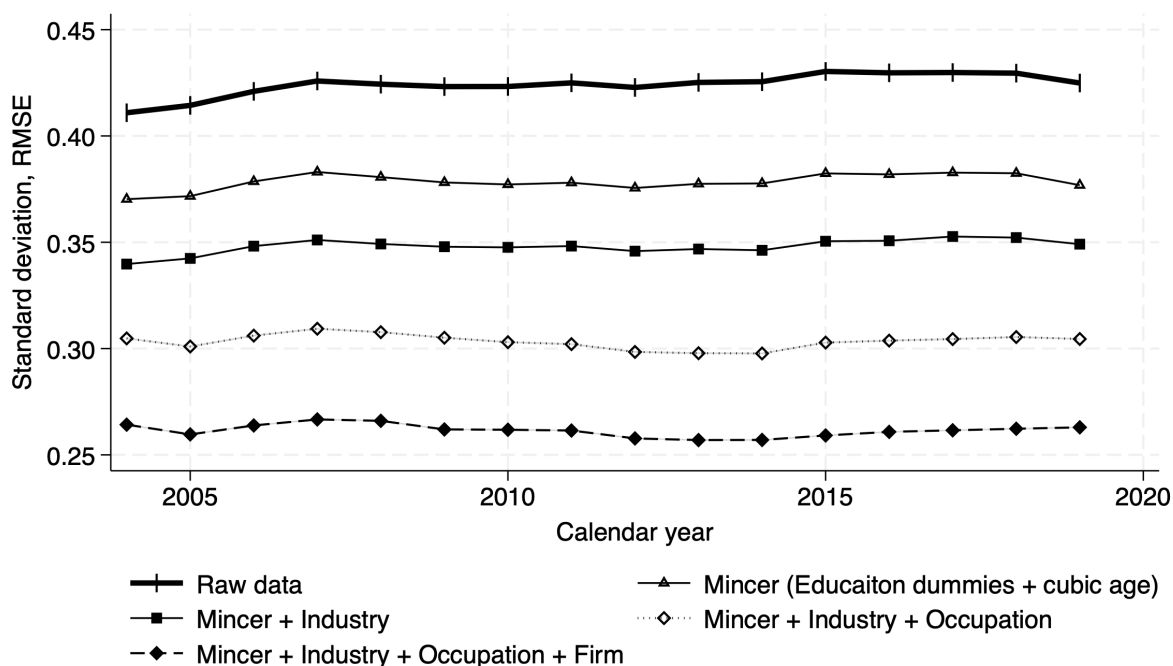


Figure A.2: Density of EA-PGI by highest education level and track

The density plot uses full analysis sample with non-missing education track information ($N = 50,940$). The blue lines correspond to secondary, and the red - to tertiary degree level. Solid lines correspond to academic, and the dashed lines - to vocational education tracks.

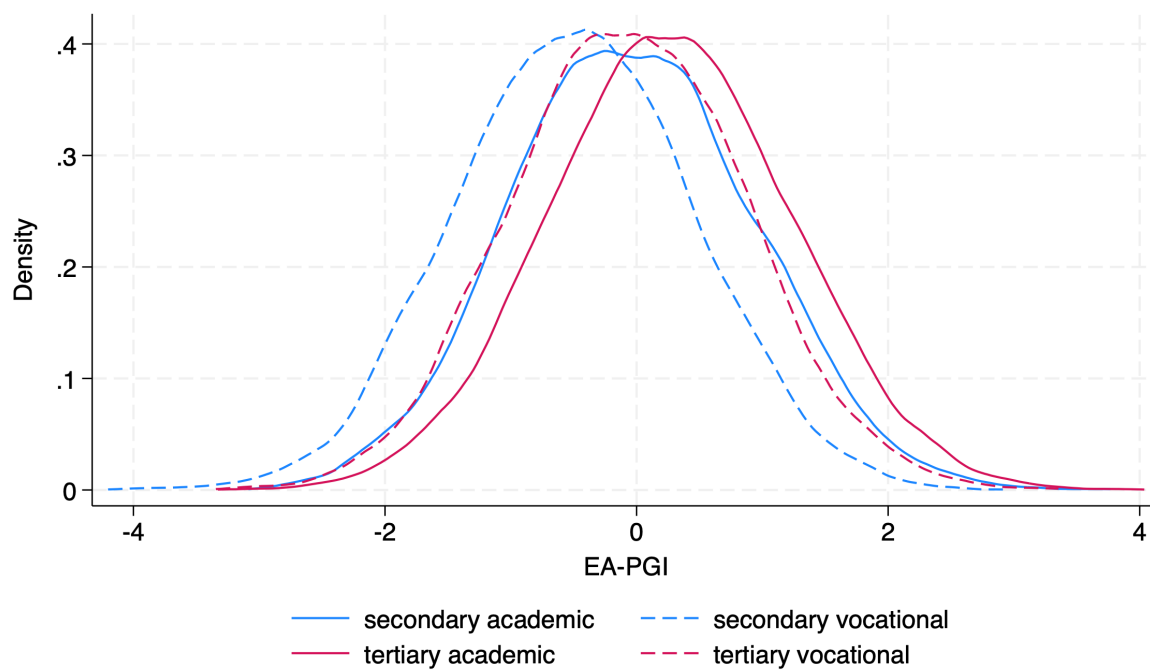


Figure A.3: Average annual income of tertiary-educated individuals by EA-PGI level, over time

The figure uses a subset of tertiary-educated individuals ($N = 32,364$). The lines correspond to ten deciles of the EA-PGI distribution. Average income estimated from a regression of annual income on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95% CIs.

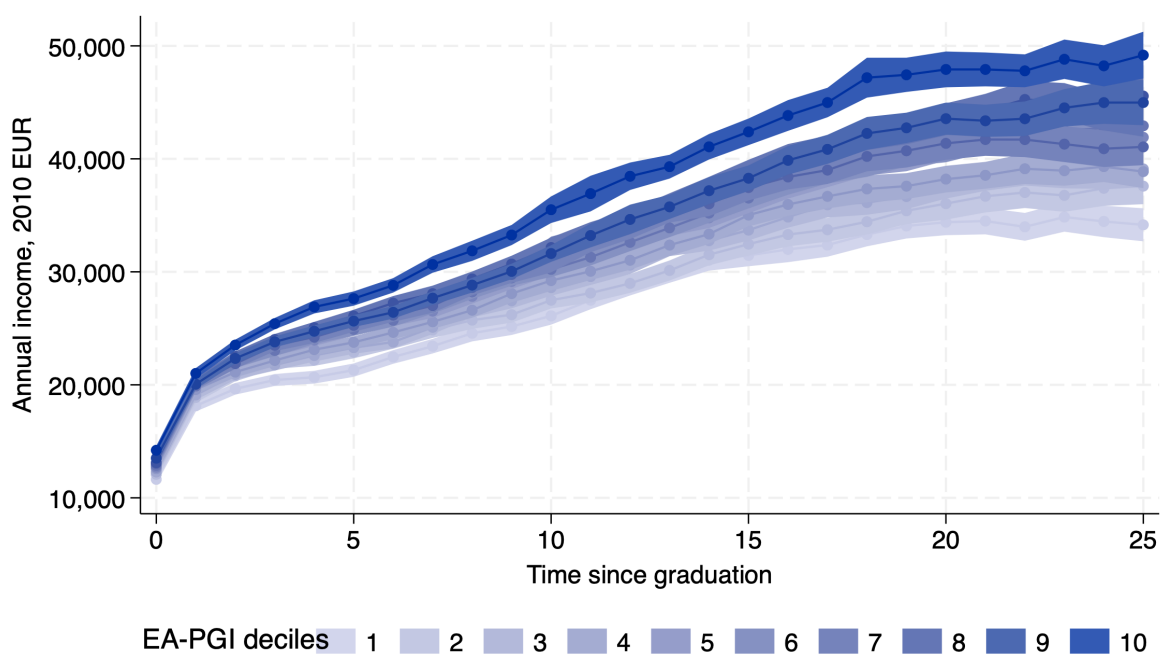
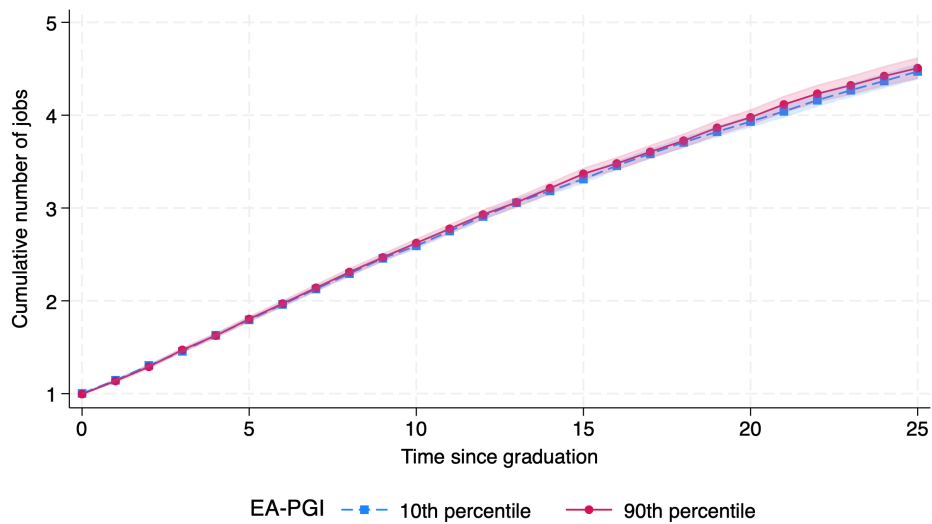


Figure A.4: Employer mobility and quality of secondary-educated individuals by EA-PGI level, over time

Panel (a) plots the average number of employment spells, and Panel (b) - the average firm quality index over time since graduation. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Both panels are estimated from a regression of respective outcome on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The estimation sample is restricted to secondary-educated workers with non-missing firm identifiers ($N = 17,432$) in panel (a) and non-missing firm quality index ($N = 15,579$) in panel (b). The shaded areas correspond to 95% CIs.

(a) Number of jobs since labor market entry



(b) Firm quality since labor market entry

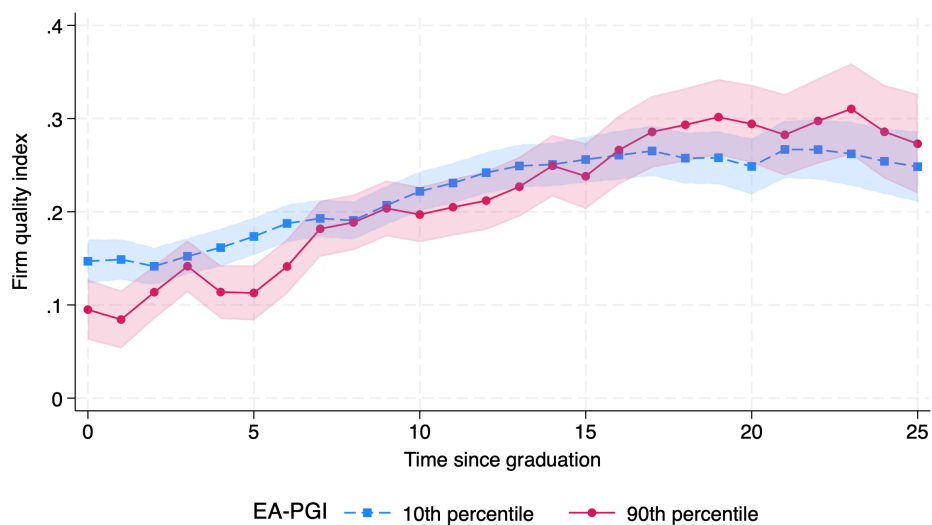
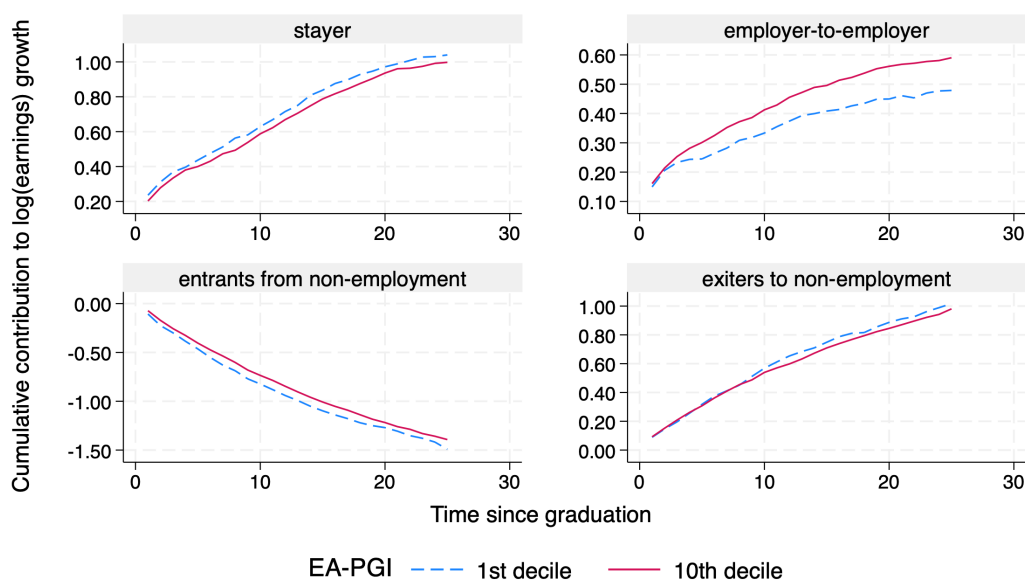


Figure A.5: Decomposition of cumulative growth in log annual income of tertiary-educated individuals by EA-PGI level, over time, including mobility into and out of non-employment

Panel (a) reports cumulative contributions of four mobility types to overall cumulative log earnings growth over time since graduation. Panel (b) reports average employment shares of four mobility types over time since graduation. The blue line corresponds to 1st and the red - to 10th decile of EA-PGI distribution. The decomposition follows the algorithm in.²³ The sample for the decomposition is restricted to tertiary-educated workers ($N = 32,364$).

(a) Cumulative contribution to log(earnings) growth



(b) Employment share

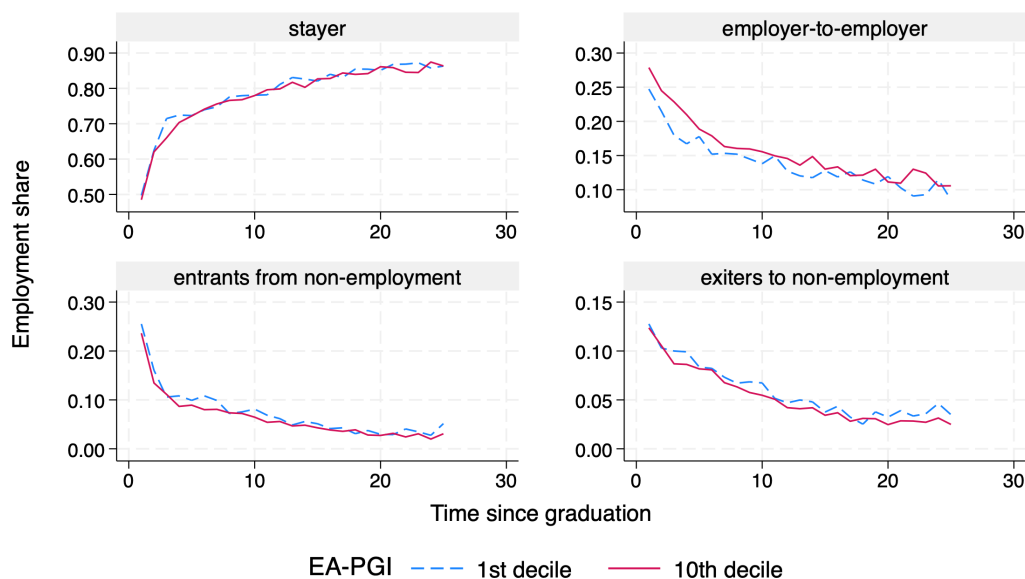


Figure A.6: Average health index by EA-PGI level, over time and by education, conditional on parental EA-PGI among parent-offspring trios

Panel (a) uses full sample of parent-offspring sample with non-missing Charlson Comorbidity Index ($N = 50,963$), while Panels (b) and (c) use subset of these workers based on their highest qualification being either secondary ($N = 18,669$) or tertiary degree ($N = 32,294$), respectively. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Average health index estimated from a regression of Charlson Comorbidity Index on the set of own, maternal and paternal EA-PGI, all fully interacted with indicators measuring years since graduation, and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95% CIs.

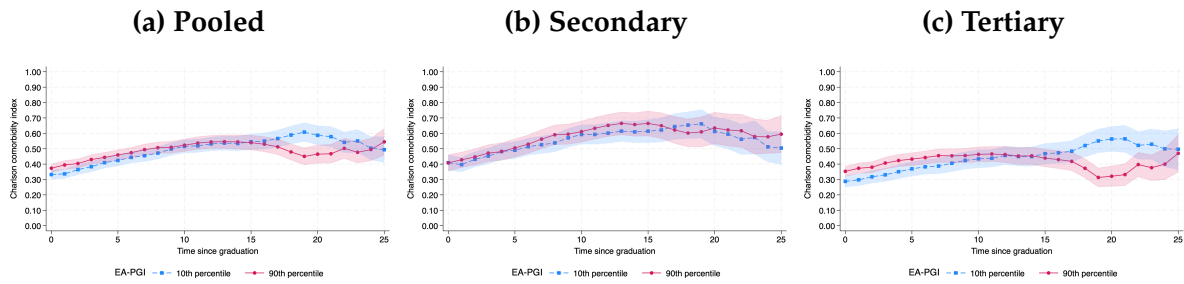


Figure A.7: Average annual income of tertiary-educated individuals by own or parental EA-PGI levels, over time

The figure uses a subset of parent-offspring trios with highest offspring qualification being tertiary degree ($N = 7,808$). The lines in Panel (a) correspond to 10th and 90th percentiles of offspring EA-PGI; in Panel (b) - 10th and 90th percentiles of paternal EA-PGI; and in Panel (c) - 10th and 90th percentiles of maternal EA-PGI. Average income estimated from a regression of annual income on the set of own, maternal and paternal EA-PGI, all fully interacted with indicators measuring years since graduation, and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95% CIs.

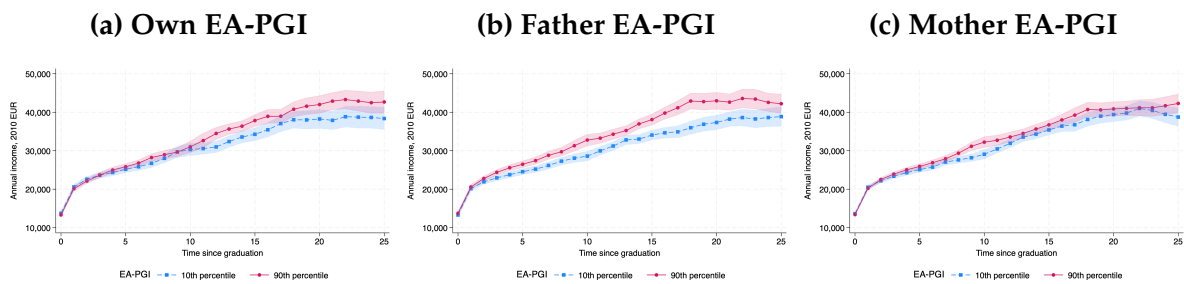
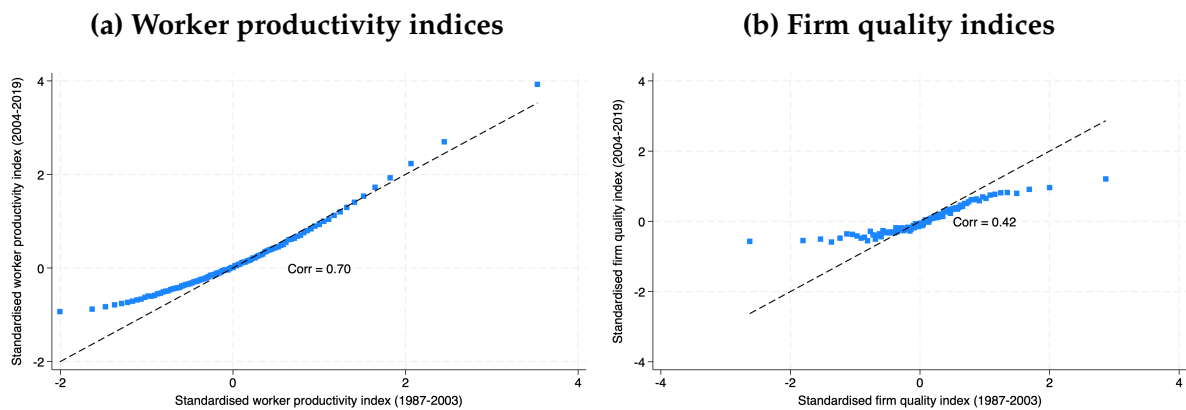


Figure A.8: AKM worker productivity and firm quality indices between two AKM estimation periods (1987-2003 and 2004-2019)

Panel (a) is a binscatter plot of worker productivity indices estimated using matched employer-employee data between 2004-2019 (on the y axis) and 1987-2003 (on the x axis). Panel (b) is a binscatter plot of firm quality indices estimated using matched employer-employee data between 2004-2019 (on the y axis) and 1987-2003 (on the x axis). The dashed black lines correspond to 45° line. The sample in Panel (a) are individuals with both worker productivity indices non-missing ($N = 13,715$); in Panel (b) - individuals with both firm quality indices non-missing ($N = 32,494$).



A.2 Supplementary Tables

Table A.1: Working sample in trajectory analysis

The table reports sample counts by working sample restrictions. Columns (1)-(3) report person-year observation counts in full sample and by biobanks, respectively. Columns (4)-(6) report unique individual counts in full sample and by biobanks.

	Person-year observations			Unique individuals		
	All (1)	THL (2)	BDB (3)	All (4)	THL (5)	BDB (6)
Start	5,374,521	3,963,254	1,411,267	176,523	132,171	44,352
Keep individuals with non-missing genetic PCs	5,335,564	3,963,254	1,372,310	175,050	132,171	42,879
Keep graduates only	3,215,453	1,957,308	1,258,145	98,810	59,416	39,394
Keep graduates with non-missing graduation year	3,215,453	1,957,308	1,258,145	98,810	59,416	39,394
Graduated between 1970 and 2020	3,139,889	1,882,118	1,257,771	96,186	56,803	39,383
Observed between 0 and 25 years since graduation	1,692,473	1,065,113	627,360	96,166	56,786	39,380
Followed from 0 years since graduation	1,000,872	576,894	423,978	57,956	29,205	28,751
Followed at least up to age 30 (if secondary)	963,715	567,236	396,479	51,056	27,135	23,921

Table A.2: Average population and sample characteristics of graduates

Panel (a) compares full analysis sample ($N = 51,056$) and Panel (b) compares subsample of parent-offspring trios ($N = 12,871$) to the population of fresh graduates. The population of fresh graduates is selected from the full Finnish population with same criteria described in Table A.1, except for conditioning on genotypes and genetic principal components. Column (1) reports average characteristics in the population of fresh graduates. Columns (2) and (3) report average characteristics in the corresponding sample, before and after applying inverse probability weights, respectively. Columns (4) and (5) report p -values for the equality of means between population and working sample, before and after applying inverse probability weights, respectively. All p -values are adjusted for multiple hypotheses testing using Holm correction. Columns (6) and (7) report population and sample counts, respectively. Inverse probability weights are estimated to balance the sample according to year of birth and graduation year (fully interacted with highest education level and gender), and rural area indicator fully interacted with gender.

(a) Genotyped sample

	Population (1)	Genotyped sample (2)	Rewighted sample (3)	$\Delta_{(2)}^{(1)}$ (4)	$\Delta_{(3)}^{(1)}$ (5)	$N^{(1)}$ (6)	$N^{(2)}$ (7)
Cohort: 1950-59	0.01	0.01	0.01	0.01	0.000	1.000	1,599,332
Cohort: 1960-69	0.17	0.22	0.16	0.000	1.000	1,599,332	51,056
Cohort: 1970-79	0.34	0.36	0.34	0.000	1.000	1,599,332	51,056
Cohort: 1980-89	0.36	0.29	0.37	0.000	1.000	1,599,332	51,056
Cohort: 1990-99	0.13	0.11	0.13	0.000	1.000	1,599,332	51,056
Graduation age: 16-20	0.36	0.31	0.36	0.000	1.000	1,599,332	51,056
Graduation age: 21-25	0.39	0.43	0.39	0.000	1.000	1,599,332	51,056
Graduation age: 26-30	0.25	0.25	0.24	0.001	1.000	1,599,332	51,056
Education: secondary	0.44	0.37	0.44	0.000	1.000	1,599,332	51,056
Education: tertiary	0.56	0.63	0.56	0.000	1.000	1,599,332	51,056
Male	0.48	0.39	0.48	0.000	1.000	1,599,332	51,056
Married	0.10	0.13	0.11	0.000	0.000	1,599,332	51,056
Rural	0.24	0.25	0.25	0.712	1.000	1,599,332	51,056
Cumulated income	279,534	302,077	277,373	0.000	0.692	1,599,332	51,056
Income at t=0	9,301	9,527	9,328	0.000	1.000	1,599,332	51,056

(b) Family trio sample

	Population (1)	Genotyped family trios (2)	Rewighted family trios (3)	$\Delta_{(2)}^{(1)}$ (4)	$\Delta_{(3)}^{(1)}$ (5)	$N^{(1)}$ (6)	$N^{(2)}$ (7)
Cohort: 1950-59	0.01	0.01	0.01	0.01	0.446	1.000	12,871
Cohort: 1960-69	0.17	0.15	0.17	0.001	1.000	1,599,332	12,871
Cohort: 1970-79	0.34	0.35	0.34	0.017	1.000	1,599,332	12,871
Cohort: 1980-89	0.36	0.38	0.36	0.000	1.000	1,599,332	12,871
Cohort: 1990-99	0.13	0.11	0.12	0.000	1.000	1,599,332	12,871
Graduation age: 16-20	0.36	0.34	0.36	0.000	1.000	1,599,332	12,871
Graduation age: 21-25	0.39	0.42	0.40	0.000	1.000	1,599,332	12,871
Graduation age: 26-30	0.25	0.24	0.24	0.086	1.000	1,599,332	12,871
Education: secondary	0.44	0.39	0.44	0.000	1.000	1,599,332	12,871
Education: tertiary	0.56	0.61	0.56	0.000	1.000	1,599,332	12,871
Male	0.48	0.40	0.48	0.000	1.000	1,599,332	12,871
Married	0.10	0.12	0.12	0.000	1.000	1,599,332	12,871
Rural	0.24	0.27	0.24	0.000	1.000	1,599,332	12,871
Cumulated income	279,534	282,107	283,464	0.446	1.000	1,599,332	12,871
Income at t=0	9,301	9,782	9,557	0.000	1.000	1,599,332	12,871

Table A.3: Descriptive statistics in AKM sample

The table reports descriptive statistics in full matched employee-employer panel used in AKM estimation. Column (1) reports sample mean, column (2) - standard deviation, and column (3) - person-year observation counts for the variables in rows.

	Mean (1)	SD (2)	N (3)
Cohort: 1946-1955	0.231	0.421	7,225,843
Cohort: 1956-1965	0.299	0.458	9,379,633
Cohort: 1966-1975	0.240	0.427	7,504,156
Cohort: 1976-1985	0.162	0.369	5,086,257
Cohort: 1986-1995	0.064	0.245	2,010,473
Cohort: 1996-2005	0.004	0.062	121,624
Age group: 20-29	0.200	0.400	6,875,154
Age group: 30-39	0.288	0.453	9,929,190
Age group: 40-49	0.288	0.453	9,926,283
Age group: 50-59	0.212	0.409	7,306,357
Age group: 60-69	0.011	0.106	391,987
Male	0.604	0.489	34,428,971
Education level: Compulsory	0.212	0.409	7,314,987
Education level: Secondary	0.447	0.497	15,406,896
Education level: Tertiary	0.340	0.474	11,707,088
Firm size	1,690	4,283	34,428,971
Annual total earning, 2015 EUR	27,319	30,515	34,428,936
Months worked in a year	11.680	1.241	34,428,971
Monthly total earning, 2015 EUR	2,330	2,585	34,428,936
Firm quality index	0.365	0.771	15,699,075
Worker productivity index	0.122	0.990	16,162,338

Table A.4: AKM summary statistics and variance decomposition

The table reports summary statistics and variance decomposition following AKM estimations. The dependent variable in AKM estimation is log monthly earnings calculated as the ratio of total annual earnings by number of months worked. The sample includes employees aged between 20 and 60, with monthly earnings above 50% of national median monthly income, working in firms with at least 5 employees and for at least four months in a calendar year. The estimations control for calendar year indicators, education level, cubic polynomial in age, as well as interactions of calendar year and age polynomial with education level. For further details, see Section 1.2.

	Dependent variable: log monthly earnings	
	1987-2003	2004-2019
Standard deviation of outcome	0.5003	0.4614
N largest connected set	16 862 428	15 435 023
N singletons	275 680	374 028
N estimation sample	16 586 748	15 060 995
<i>Panel A: Summary of parameter estimates</i>		
N worker FE	1 881 715	1 842 564
N firm FE	126 605	50 430
Std. dev. of worker FE	0.2969	0.3208
Std. dev. of firm FE	0.1067	0.1027
Std. dev. of Xb	0.3416	0.2437
Std. dev. of residual	0.1587	0.1561
Corr(worker FE, firm FE)	0.1054	0.2496
RMSE	0.1693	0.1669
Adjusted R2	0.8846	0.8681
<i>Panel B: Share of outcome variance attributed to</i>		
Worker FE	0.3547	0.4868
Firm FE	0.0458	0.0499
Cov(worker FE, firm FE)	0.0269	0.0778
Xb and associated covariances	0.4712	0.2703
Residual	0.1014	0.1153

Table A.5: Descriptive statistics by EA-PGI level

The table reports descriptive statistics among tertiary-educated individuals ($N = 32,364$) by deciles of EA-PGI. Columns (1)-(3) report sample means in 1st, 2nd-9th and 10th decile of EA-PGI, respectively. Column (4) reports difference in sample means and standard error of the difference in parentheses for the 2nd-9th decile relative to 1st decile, respectively. Column (5) reports difference in sample means and standard error of the difference in parentheses for the 10th decile relative to 1st decile, respectively. Column (6) reports total sample count for each variable considered. All estimates control for first ten genetic principal components.

	Sample means			Diff.		(6)
	1st decile	2nd-9th deciles	10th decile	2nd-9th deciles	10th decile	
	(1)	(2)	(3)	(4)	(5)	
Male	0.267	0.341	0.402	0.074*** (0.011)	0.135*** (0.013)	32,364
Birth year	1977.2	1977.3	1977.6	0.170 (0.231)	0.455 (0.272)	32,364
Mother edu: compulsory	0.332	0.250	0.175	-0.082*** (0.010)	-0.157*** (0.012)	32,364
Mother edu: secondary	0.413	0.355	0.258	-0.059*** (0.011)	-0.155*** (0.013)	32,364
Mother edu: tertiary	0.247	0.386	0.554	0.139*** (0.011)	0.307*** (0.013)	32,364
Mother edu: missing	0.008	0.009	0.012	0.001 (0.002)	0.005 (0.003)	32,364
Father edu: compulsory	0.351	0.278	0.196	-0.073*** (0.010)	-0.155*** (0.012)	32,364
Father edu: secondary	0.388	0.314	0.218	-0.074*** (0.011)	-0.170*** (0.012)	32,364
Father edu: tertiary	0.220	0.366	0.551	0.147*** (0.011)	0.332*** (0.013)	32,364
Father edu: missing	0.042	0.042	0.035	0.001 (0.005)	-0.007 (0.005)	32,364
Age at graduation	24.506	24.854	25.245	0.349*** (0.054)	0.740*** (0.063)	32,364
Graduation: years since predicted graduation	2.256	2.211	2.126	-0.045 (0.049)	-0.131* (0.058)	32,364
Age at first job	26.332	26.448	26.739	0.116 (0.072)	0.407*** (0.085)	31,015
First job: years since predicted graduation	3.588	3.302	3.092	-0.286*** (0.071)	-0.496*** (0.084)	31,015
Annual average income at first job	11,729	12,811	14,208	1,082*** (242)	2,478*** (285)	32,364
AKM firm FE of first job	0.217	0.236	0.268	0.019 (0.022)	0.052* (0.026)	16,860
AKM firm FE at t=15	0.407	0.462	0.527	0.055 (0.035)	0.120** (0.043)	8,004

Table A.6: Worker productivity and EA-PGI

The table reports estimation results from a regression of worker productivity index on EA-PGI of interacted with education level. All regressions control for first ten genetic PCs, gender, year of birth, calendar year and biobank indicators. Column (1) reports the baseline results. Column (2) additionally controls for detailed field of education (3-digit) indicators. Column (3) adds interaction between level and detailed field of education. Column (4) adds indicators for academic institution identifiers. Standard errors reported in parentheses.

	Dependent variable: Worker productivity index			
	(1)	(2)	(3)	(4)
Secondary education \times EA-PGI	0.017* (0.007)	0.010 (0.007)	0.011 (0.007)	0.007 (0.007)
Tertiary education \times EA-PGI	0.115*** (0.007)	0.086*** (0.007)	0.082*** (0.007)	0.016* (0.006)
Level	Yes	Yes	Yes	Yes
Field	No	Yes	Yes	Yes
Level \times Field	No	No	Yes	No
Institution ID	No	No	No	Yes
Obs.	31,866	31,709	31,702	31,574
Avg. obs. per cell	15,933	352	145	23
Adj R2	0.198	0.251	0.256	0.330
RMSE	0.836	0.808	0.806	0.765

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.7: Years of education and EA-PGI, unconditional and conditional on parental EA-PGI among parent-offspring trios

The table reports estimation results from regression of predicted years of education (given highest qualification) on EA-PGI, unconditional and conditional on parental EA-PGI in parent-offspring trios. The top panel reports baseline estimations without controlling for parental EA-PGI. The bottom panel reports estimates conditional on parental EA-PGI. Column (1) uses full sample of parent-offspring trios ($N = 12,871$), including the cases where parental genotypes were imputed from parent-offspring duos and siblings. Column (2) uses subset of directly genotyped parent-offspring trios ($N = 4,586$). All estimations control for first ten genetic principal components, gender, year of birth, and biobank indicators. Standard errors reported in parentheses.

	Dependent var: predicted years of education	
	All family trios (1)	Directly genotyped trios (2)
Baseline without parental EA-PGI		
Own EA-PGI	0.553*** (0.016)	0.570*** (0.027)
Constant	14.691*** (0.491)	13.741*** (1.058)
Obs.	12 871	4 586
Controlling for parental EA-PGI		
Own EA-PGI	0.413*** (0.026)	0.441*** (0.040)
Mother EA-PGI	0.128*** (0.021)	0.110*** (0.030)
Father EA-PGI	0.093*** (0.021)	0.095*** (0.030)
Constant	14.641*** (0.482)	13.717*** (1.028)
Obs.	12 871	4 586

Table A.8: Cumulated lifetime income, own and parental EA-PGI, regression results

The table reports estimation results from regression of cumulated lifetime income on own and parental EA-PGI. Columns (1)-(3) report the baseline estimates without controlling for parental EA-PGI. Columns (4)-(6) report estimates conditional on parental EA-PGI. Columns (1) and (4) use full sample of parent-offspring trios ($N = 12,871$), columns (2) and (5) - subset of trios with secondary-educated offspring ($N = 5,063$) and columns (3) and (6) - subset of trios with tertiary-educated offspring ($N = 7,808$). All estimates control for first ten genetic principal components, gender, birth year, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income					
	Baseline			Controlling for parental EA-PGI		
	Pooled (1)	Secondary (2)	Tertiary (3)	Pooled (4)	Secondary (5)	Tertiary (6)
Own EA-PGI	13 724 (1 804)	-7 721 (2 132)	15 602 (2 546)	7 398 (2 746)	-3 613 (3 478)	5 048 (3 833)
Mother EA-PGI				1 769 (2 163)	-5 002 (2 680)	4 658 (3 016)
Father EA-PGI				8 211 (2 375)	-1 702 (2 795)	12 113 (3 370)
Constant	414 682 (61 295)	270 156 (58 545)	507 419 (84 159)	411 300 (61 302)	270 147 (57 824)	501 314 (84 234)
Obs.	12 871	5 063	7 808	12 871	5 063	7 808

Table A.9: Cumulated lifetime income by own and parental EA-PGI levels

The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles, conditional on parental EA-PGI. Columns (1)-(3) use full sample of parent-offspring trios with tertiary-educated offspring ($N = 7,808$). Columns (4)-(6) use subset of directly genotyped parent-offspring trios with tertiary-educated offspring ($N = 2,704$). Columns (1) and (4) report cumulated lifetime income at 10th and 90th percentiles of own EA-PGI, holding maternal and paternal EA-PGI constant. Columns (2) and (5) report cumulated lifetime income at 10th and 90th percentiles of maternal EA-PGI, holding own and paternal EA-PGI constant. Columns (3) and (6) report cumulated lifetime income at 10th and 90th percentiles of paternal EA-PGI, holding own and maternal EA-PGI constant. Average lifetime income adjusted by regressing cumulated income on the set of own, maternal and paternal EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income					
	All family trios			Directly genotyped trios		
	Own EA-PGI (1)	Mother EA-PGI (2)	Father EA-PGI (3)	Own EA-PGI (4)	Mother EA-PGI (5)	Father EA-PGI (6)
10th percentile	339 181 (5 690)	339 879 (4 311)	330 699 (4 459)	350 648 (9 782)	342 498 (6 842)	336 264 (6 792)
90th percentile	352 172 (5 198)	351 466 (4 500)	360 946 (5 126)	350 086 (8 547)	358 149 (6 833)	364 690 (7 444)
90-10 gap	12 991 (9 866)	11 587 (7 502)	30 247 (8 414)	-562 (16 426)	15 650 (10 933)	28 426 (11 639)
Obs.	7 808	7 808	7 808	2 704	2 704	2 704

Table A.10: Cumulated lifetime income of secondary-educated individuals by EA-PGI level, conditional on parental EA-PGI among parent-offspring trios

The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles. The estimation sample is restricted to parent-offspring trios with secondary-educated offspring ($N = 5,063$). Column (1) reports the estimates in the baseline specification without controlling for parental EA-PGI. Columns (2)-(4) report average cumulated lifetime income by own, maternal and paternal EA-PGI percentiles, respectively, conditional on parental EA-PGI. Average lifetime income adjusted by regressing cumulated income on EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income			
	Baseline	Controlling for parental EA-PGI		
	Own EA-PGI (1)	Own EA-PGI (2)	Mother EA-PGI (3)	Father EA-PGI (4)
<i>EA-PGI percentiles</i>				
10th	265 492 (2 782)	259 961 (4 621)	261 546 (3 938)	257 400 (3 954)
20th	262 054 (2 229)	258 352 (3 323)	259 384 (3 036)	256 677 (3 029)
30th	259 582 (2 022)	257 195 (2 568)	257 866 (2 532)	256 156 (2 503)
40th	257 496 (2 017)	256 219 (2 184)	256 508 (2 240)	255 710 (2 223)
50th	255 614 (2 150)	255 338 (2 154)	255 320 (2 161)	255 311 (2 159)
60th	253 639 (2 405)	254 413 (2 459)	254 078 (2 274)	254 906 (2 294)
70th	251 508 (2 775)	253 416 (3 057)	252 817 (2 566)	254 471 (2 624)
80th	248 921 (3 311)	252 206 (3 978)	251 177 (3 130)	253 929 (3 216)
90th	245 621 (4 074)	250 661 (5 294)	249 103 (4 008)	253 149 (4 253)
Obs.	5 063	5 063	5 063	5 063

Table A.11: Cumulated lifetime income by own EA-PGI level, education, genotype sample and weighting scheme

The table reports adjusted lifetime income (up to 25 years since graduation) by own EA-PGI percentiles. Panel A uses full analysis sample ($N = 51,056$ in pooled, $N = 18,692$ in secondary-educated and $N = 32,364$ in tertiary-educated subsamples), while Panel B uses sample of parent-offspring trios ($N =$ in pooled, $N = 5,063$ in secondary-educated and $N = 7,808$ in tertiary-educated subsamples). Columns (1), (3) and (5) report unweighted estimates. Columns (2), (4) and (6) report weighted estimates using inverse probability weights. Average lifetime income adjusted by regressing cumulated income on EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Panel B additionally controls for maternal and paternal EA-PGI. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Inverse probability weights are estimated to balance the sample according to year of birth and graduation year (fully interacted with highest education level and gender), and rural area indicator fully interacted with gender. Standard errors reported in parentheses.

	Dependent variable: Cumulated income					
	Pooled		Secondary		Tertiary	
	Unweighted (1)	Weighted (2)	Unweighted (3)	Weighted (4)	Unweighted (5)	Weighted (6)
Panel A: Genotyped sample						
10th percentile	309 659 (1 306)	291 728 (1 286)	262 386 (1 429)	257 996 (1 501)	346 194 (1 944)	331 362 (1 920)
50th percentile	329 893 (857)	308 756 (832)	255 422 (1 116)	249 549 (1 157)	368 728 (1 137)	350 947 (1 105)
90th percentile	350 418 (1 591)	325 930 (1 525)	248 358 (2 120)	241 029 (2 185)	391 585 (2 006)	370 700 (1 938)
Obs.	51 056	51 056	18 692	18 692	32 364	32 364
Panel B: Family trio sample (conditional on parental EA-PGI)						
10th percentile	303 725 (3 906)	304 546 (4 417)	259 961 (4 621)	263 787 (5 261)	339 181 (5 690)	345 745 (6 559)
50th percentile	313 190 (1 697)	313 886 (1 981)	255 338 (2 154)	258 422 (2 388)	345 639 (2 320)	351 212 (2 755)
90th percentile	322 764 (3 934)	323 347 (4 502)	250 661 (5 294)	252 986 (5 717)	352 172 (5 198)	356 750 (6 114)
Obs.	12 871	12 871	5 063	5 063	7 808	7 808

Table A.12: Predictive power of EA-PGI for years of education

The table reports estimation results from regression of predicted years of education (given highest qualification) on EA-PGI. Column (1) reports estimates without EA-PGI, and column (2) - with EA-PGI. All estimations controls for gender fully interacted with year of birth indicators, biobank indicator and first ten genetic principal components. Standard errors reported in parentheses.

	Dependent variable: predicted years of education	
	(1)	(2)
EA-PGI		0.530*** (0.008)
Constant	14.984*** (0.202)	14.979*** (0.194)
Obs.	51,056	51,056
R^2	0.086	0.158
Adj. R^2	0.085	0.157
Incremental R^2		0.071

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$