

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Helin, Jouni; Jokinen, Juho; Koerselman, Kristian; Nokkala, Terhi; Räikkönen, Eija

**Title:** It runs in the family? Using sibling similarities to uncover the hidden influence of family background in doctoral education and academic careers

**Year:** 2023

**Version:** Published version

**Copyright:** © The Author(s) 2022.

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Helin, J., Jokinen, J., Koerselman, K., Nokkala, T., & Räikkönen, E. (2023). It runs in the family? Using sibling similarities to uncover the hidden influence of family background in doctoral education and academic careers. *Higher education*, 86, 1-20. <https://doi.org/10.1007/s10734-022-00897-7>



# It runs in the family?

## Using sibling similarities to uncover the hidden influence of family background in doctoral education and academic careers

Jouni Helin<sup>1,2</sup> · Juho Jokinen<sup>3</sup> · Kristian Koerselman<sup>1,3</sup> · Terhi Nokkala<sup>1</sup> · Eija Räikkönen<sup>2</sup>

Accepted: 28 June 2022  
© The Author(s) 2022

### Abstract

Family background has been shown to be a strong determinant of educational attainment, yet relatively little is known about the role that family background plays in PhD attainment or in the selection into academic careers. In this study, we estimate sibling correlations from Finnish full population register data to comprehensively assess the importance of family background in selection into academia. Our results show that family background accounts for over a third of the overall variation in becoming a PhD and subsequently an academic — a share which is up to four times as large as implied by conventional comparisons by parental education. However, we did not find evidence that family background would be an exceptionally strong determinant of doctoral outcomes when compared to other educational outcomes. Our findings further suggest that sibling similarities in PhD attainment and academic careers may largely be attributed to sibling similarities in prior educational achievement rather than to other family background characteristics.

**Keywords** Inequality · Higher education · PhD attainment · Academic careers · Sibling correlations

### Introduction

Research on the selection of PhDs and of academic faculty is more relevant now than ever. PhDs and academics are key producers of scientific knowledge, knowledge which in turn has become increasingly important for our continuing social and economic development (Pinto & Teixeira, 2020; Välimaa et al., 2016). In addition to the benefits to society as a whole, a research degree or an academic career can also bring benefits to the individual

---

✉ Jouni Helin  
jouni.t.helin@jyu.fi

<sup>1</sup> Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland

<sup>2</sup> Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

<sup>3</sup> Jyväskylä University School of Business and Economics, University of Jyväskylä, Jyväskylä, Finland

pursuing it. Among other benefits, PhD holders for example usually enjoy relatively high earnings and relatively high employment security (Mertens & Röbbken, 2013; Britton et al., 2020). The degree of diversity among academics may also influence current and future student generations' perceptions of who can become an academic (e.g. Gurin et al., 2002). It is thus important to examine inequalities in selection into academia not only with an eye to social welfare, but also with an eye to fairness in terms of who is selected, and who is not.

It is generally accepted that higher education should be open to prospective students regardless of their family background, and this reasoning should reasonably extend also to PhDs and academic careers (cf. Merton, 1979; Roemer, 2004). Yet although there is a sizeable literature on the role which family background plays in selection into the undergraduate level of higher education (e.g. Haas & Hadjar, 2020; Marginson, 2016, 2018; Reimer & Jacob, 2011; Shavit et al., 2007), efforts to quantify the role of background factors in selection into doctoral education or into the academic profession are comparatively scarce.

What quantitative evidence on between-family inequalities in selection into academia we do have mostly relies on comparisons by parental socioeconomic status (e.g. Bachsleitner et al., 2018; Mastekaasa, 2006; Mullen et al., 2003; Triventi, 2013), implicitly working off the assumption that all families sharing a specific position in the social hierarchy are alike. This hides a considerable amount of variation in childhood circumstances, and thus causes the importance of family background to be underestimated (Björklund & Jäntti, 2020). To our knowledge, there is furthermore no research comparing between-family inequalities in PhD attainment and academic careers to those in other educational and professional trajectories, and as a consequence, our understanding of the magnitude of the former is limited. Additionally, without a grasp of what characteristics of family background it is that drive inequalities, it is hard to target policy interventions to reduce such inequalities, and to promote equal access into academia.

We address these research gaps by asking the following questions:

1. (a) What is the role of family background in determining the attainment of PhDs and academic careers, and (b) how does this compare to the role of family background in other educational and professional trajectories?
2. To what extent do specific family characteristics contribute to the overall importance of family background in PhD attainment and academic careers?

To answer these research questions, we estimate sibling correlations from register data covering all individuals born in Finland between 1965–1975. A sibling correlation is a measure of the role that family background plays in determining an outcome. Unlike comparisons based on parental socioeconomic status, a sibling correlation captures all parental and immediate environmental characteristics which children from the same family share — even such characteristics which are not directly observable. By estimating sibling correlations, we thus provide a broader and more holistic measure of between-family inequalities in PhD attainment and academic careers compared to earlier studies that rely on comparisons by socioeconomic background. Furthermore, by comparing sibling correlations in PhD attainment and academic careers to those in other educational qualifications, we can assess whether the role of family background is different in selection into academia than into other professional trajectories. Finally, we examine the extent to which specific family background characteristics contribute to the overall importance of family background in selection into academia. This not only allows us to quantify the share of family background effects in these outcomes that would remain hidden without the use of sibling correlations,

but also provides some guidance for future research on which family characteristics seem to matter the most in becoming an academic.

## Background

Family background is an important determinant of educational performance, of the educational choices individuals make, and ultimately, of educational attainment (Björklund & Salvanes, 2011; Erikson & Jonsson, 1996). Moreover, early life influences can be long lived. Since becoming an academic requires completing multiple educational transitions all the way to attaining a PhD, we have good reason to think that family background would be an important determinant of academic careers as well (Hermanowicz, 2012; Posselt & Grodsky, 2017). It is also important to acknowledge that there are multiple ways through which families may confer advantages as well as place barriers in the path to doctoral education and beyond (e.g. Bahack & Addi-Racah, 2022; Gardner & Holley, 2011; McCallum, 2016). By neglecting some of these factors, we are at risk of seeing only a part of role that family background plays in selection into academia.

The family is the primary unit of socialization, and parents equip their children with different levels of socio-cultural skills, knowledge and networks that promote success in education and in the labor market (Bourdieu, 1986; Gardner & Holley, 2011; Marginson, 2016; Perna, 2006). Similarly, parents shape the preferences and aspirations of children (Erikson & Jonsson, 1996; Perna, 2006), such as the degree to which they value doctoral education and a career in academia (Bahack & Addi-Racah, 2022; Rockinson-Szapkiw et al., 2018). Parents also invest financial and non-financial resources, for example their time, to promote the skills formation and opportunities of their children, and these investments may further be constrained by family structure (e.g. Björklund et al., 2010; Perna, 2006). For example, the more children in a family, the less time may be able to be devoted to each child. Some traits and characteristics, such as those related to cognitive ability, non-cognitive ability, and health, are to a degree passed on from parents to children already before birth, and empirical evidence suggests that these factors may be an important source of inequality in educational outcomes (e.g. Björklund & Salvanes, 2011; Branigan et al., 2013). Furthermore, some of the family background influences manifest indirectly through the home surrounding community. Children growing up in different neighborhoods may be exposed not only to different neighborhood characteristics, such as school quality or availability of public services, but also to different peers and role models — all of which may shape educational outcomes (Akerlof, 1997; McCallum, 2016; Solon et al., 2000). The influence that family background may have in selection into academia thus arises from a broad combination of characteristics, from parental investments and endowments to childhood neighborhoods (Björklund & Salvanes, 2011; Posselt & Grodsky, 2017). The key challenge for quantitative research is how best to capture them all.

In much of the literature on selection into doctoral education and academia, the importance of family background has been measured by regressing a dependent variable, such as doctoral education enrollment, on predictors measuring socioeconomic background, such as parental education and/or occupation. Studies based on this approach suggest that undergraduates from higher socioeconomic background are more likely to transition to doctoral education (Bachsleitner et al., 2018; Helin et al., 2019; Mastekaasa, 2006; Mateos-González & Wakeling, 2022; Mullen et al., 2003; Zimdars, 2007), and to enter academia (Andersen, 2001; Helin et al., 2019; Möller, 2014; Oldfield & Conant, 2001),

even if Triventi (2013) finds mostly small and statistically insignificant differences in PhD program enrollment among undergraduates from 11 European countries.

These studies however offer an incomplete picture of the overall importance of family background in selection into academia. Although focusing on the transition from undergraduate to doctoral education has merits on its own, it will implicitly lead one to disregard between-family differences in educational trajectories leading up to that transition. More importantly, fixed parental background measures hide a considerable amount of both between-group and within-group heterogeneity in family circumstances. In this context, a regression of PhD attainment on parental education for example not only implicitly assumes that all parents with a specific level of education are the same, but also excludes all the influences of family background that do not covary with parental education. As a consequence, the importance of family background will thus necessarily be underestimated (e.g. Björklund & Salvanes, 2011).

An alternative approach is to use a sibling correlation as a measure of the role of family background. A sibling correlation represents the degree of similarity in outcomes of siblings born to the same family. The more the outcomes of siblings resemble each other, the higher the sibling correlation is. A perfect sibling correlation implies that within each family, all siblings have the same outcome, whereas an absence of a correlation implies that the outcome is randomly distributed across families. In other words, a sibling correlation measures the strength of the association between the outcomes of children in the same family. It captures the influence of all the characteristics that children in the same family share — even such characteristics that are not observed or are unobservable. Because children from the same family share more than their parents' socioeconomic status alone, a sibling correlation represents a broader measure of the importance of family background than the conventional comparisons based on parental background variables (Björklund & Salvanes, 2011; Björklund & Jäntti, 2020).

Even if there exists a rapidly growing body of sibling correlation research on educational and occupational attainment, only few studies report sibling correlations for specific levels of education or for specific occupations. Bredtmann and Smith (2018) found sibling correlations of 0.24 for upper secondary and 0.32 for tertiary degree attainment in Denmark. That is, family background accounted for 24% and 32% of the variation in these outcomes respectively, suggesting that sibling similarities are higher in tertiary than in upper secondary degree attainment. In terms of occupational attainment, Vladasel et al. (2021) found sibling correlations in different measures of entrepreneurship to range between 0.2–0.4, suggesting that family background accounted for 20% to 40% of the variation in becoming an entrepreneur.

In this paper, we estimate sibling correlations in the attainment of PhDs and of academic careers. To our knowledge, we are the first to do so, and we believe that this in and of itself is an important contribution to the literature. We furthermore estimate sibling correlations for other educational outcomes to serve as a frame of reference for interpreting our results.

One advantage of the sibling correlation approach is that a sibling correlation can be decomposed into a sum of contributions of specific family background characteristics, in addition to an unexplained part. With respect to other educational and occupational outcomes, previous studies have used such decompositions to show that parental characteristics explain a much larger share of sibling similarities than childhood neighborhood (Bredtmann and Smith, 2018; Lindahl, 2011; Raam et al., 2006; Vladasel et al., 2021).

Previous studies also suggest that the share of family background effects that can be captured by socioeconomic information on the parents is limited. Bredtmann and Smith

(2018) for example found that parents' education, occupation and income explain at most 39% of the sibling similarities in tertiary education attainment. Further, the influence of other family characteristics, such as family size, mother's age at first childbirth, or social problems, was largely captured by socioeconomic background. Thus, while socioeconomic background was an important predictor of sibling similarities in tertiary education attainment, a majority of the sibling similarities nevertheless remained unexplained by socioeconomic background (Bredtmann & Smith, 2018).

We contribute to the literature on the role of family background in doctoral education and academic careers by examining the extent to which sibling similarities in these outcomes can be attributed to parental education, family size, mother's age at first childbirth, the siblings' prior educational achievement, and their childhood neighborhoods. Doing so allows us both to quantify the share of family background effects in these outcomes that would remain unmeasured had we not used sibling correlations, and to quantify the contributions of these respective factors to the overall estimate.

A sibling correlation is a relatively directly observable empirical quantity, which can arguably be of general interest on its own. A sibling correlation is essentially constructed by counting how often the individuals who share a certain outcome are siblings, not entirely unlike how one might for example construct a gender ratio by counting how many of the individuals sharing a certain outcome are of the same gender. If one for example were to find that "all professors are men", this may be of general interest even to readers who disagree on the theoretical, normative, or political implications of such a finding. In the same way, the empirical finding that academic outcomes are clustered by family to a specific degree can be of general interest too, even to readers who disagree on its implications.

Even if a sibling correlation can stand on its own as an empirical finding, there also exists a relationship between the sibling correlation and equality of opportunity as a theory of social justice. The premise of equality of opportunity holds that differences in a social outcome are defensible as far as they are not determined by factors beyond individuals' own control, and that the determinants of an outcome can thus be decomposed into just and unjust determinants accordingly (Roemer, 1998, 2004). This serves as a normative benchmark for interpreting the sibling correlation decomposition of family effects. As one cannot choose the family one is born into, an extreme interpretation of equality of opportunity would posit any sibling correlation apart from zero as fundamentally unjust. Such an extreme interpretation is however contested by disagreement for example on the extent that differences in inherent ability should be rewarded, or on the extent that the society can legitimately intervene in parental involvement for example in their children's homework or formation of educational aspirations (Roemer, 2004; Swift, 2004). We thus follow in the tradition of the sibling correlation literature in holding that it is up to the researcher to produce the decomposition, but up to the reader to judge its normative significance, and we refer to Roemer (2004, 1998) and the references therein for a further discussion on the topic.

## Institutional context

Like other Nordic countries, Finland is characterized by a strong emphasis on equalizing opportunities in access to education at all levels (cf. Välimaa & Muhonen, 2018). Education in Finland is free of charge from pre-primary education through postgraduate

education, and the student financial aid system is comprehensive and relatively generous. These characteristics are likely to reduce financial barriers impeding participation in higher education.

Finnish students attend comprehensive school until age 16. After completing comprehensive school, an individual may choose to continue into upper secondary education, either in a vocational school or in an academically oriented Gymnasium, though it should be noted that a sizable share of the cohorts in our study does not complete either. Gymnasium concludes with a national matriculation examination, which provides general qualification to apply for tertiary education. Vocational school graduates rarely continue into tertiary education, even if they are eligible to do so.

Finnish tertiary education is a dual system, which has historically consisted of universities and vocational colleges, with the latter being made into universities of applied sciences in the mid-1990s. Applying to doctoral education requires a master's degree or an equivalent qualification, and only universities<sup>1</sup> grant master's and PhD degrees. It should also be noted that a master's degree is the usual university-graduating degree in Finland, and as such, corresponds essentially to an (under)graduate degree instead of to a postgraduate degree. For the majority of our sample members, the path to a PhD and subsequently into academia has thus been through Gymnasium and university. Moreover, differences in institutional prestige between Finnish universities may be considered low if not non-existent (Välilä & Muhonen, 2018). To pursue a career in academia, it is thus important to complete a university degree, whereas the choice of university is of lesser importance. Universities are also the single most important employers of PhDs, with an employment share of almost 40%, followed by private companies and municipalities, both with an employment share of around 20% (Holopainen, 2017).

## Data and methods

### Data

We base our study on a Finnish full-population panel containing information drawn from multiple administrative registers held at Statistics Finland. These include: the Longitudinal Employer-Employee Data, which contain information on individuals' family relationships and other basic demographics as well as on individuals' main employment relationship observed during the last week of each year; the Register of Completed Education and Degrees, which contains information on individuals' post-compulsory educational qualifications and their completion dates; and the Matriculation Examination Board Register, which covers information on individuals' test scores in the Gymnasium matriculation examination. Observations in different registers are linked to each other by unique personal identifiers, and the matching between the registers is thus exact.

Our sample consists of all individuals born in Finland between 1965–1975 and residing in Finland at any point between 1988–2015. Siblings were defined as individuals who share the same mother. Because the information on mothers is missing for many foreign-born Finnish residents, we excluded foreign born residents from the sample (ca. 1.5%).

---

<sup>1</sup> More recently, Finnish universities of applied sciences have started to grant postgraduate degrees equivalent to university master's degrees.

Mothers were then identified for over 99% of the individuals, resulting in a main sample of 722,611 individuals clustered in 488,597 families.

## Measures

### Dependent variables

We measure *PhD attainment* as a binary variable which takes a value of 1 if the individual was observed to attain a PhD degree at the latest by the end of the year in which they turn 40, and 0 otherwise. Our measure of *academic career* takes a value of 1 if the individual holds a research degree<sup>2</sup> and has been employed at least for six years<sup>3</sup> in higher education by the end of the year they turn 40, and 0 otherwise.

Educational attainment at lower levels of the Finnish education system was examined using the following binary variables (all variables coded 1 for attainment and 0 otherwise). The variable *upper secondary education* represents attaining any upper secondary qualification (ISCED<sup>4</sup> 3) at the latest by the end of the year in which the person turns 25. The variable *tertiary education* represents attaining any tertiary degree (ISCED 5–7) at the latest by the end of the year in which the person turns 30. Since the path to a PhD most likely leads through Gymnasium and subsequent university education, we wanted to measure these tracks also separately. The variable *Gymnasium* represents attainment in the academic upper secondary education, and variable *university* represents attaining any master's degree. To compare academic careers to other professional careers, we additionally included variables denoting master's degree attainment in the following fields of study (based on ISCED-F 2013): Education, Arts, Humanities, Social sciences, Business and administration (B&A), Law, Sciences, Information and communication technologies (ICT), Engineering, Agriculture, and Health and welfare sciences.

### Independent variables

We selected explanatory variables representing specific family characteristics based on the literature review in [Background](#). These variables are listed in [Table 1](#). Concerning family structure, *family size* denotes the total number of siblings in the family, and *maternal age* denotes the age at which a given mother bore her first child. *Mother's education* and *father's education* were measured as the highest level of education of each parent, and spans eight levels, from compulsory education to PhD. Educational attainment is a strong predictor of income in Finland (Koerselman & Uusitalo, 2014; Suhonen & Jokinen, 2018), which is why parental education may be considered as a good overall proxy for families' socioeconomic status.

To examine the contribution of sibling similarities in prior educational achievement, we used information on the Matriculation Examination, a national standardized examination

---

<sup>2</sup> We condition the outcome on PhD and/or Licentiate's degree (an optional intermediate degree before a PhD) attainment because higher education institutions employ also a considerable number of non-academic staff.

<sup>3</sup> This represents the median among those with at least one year of employment in academia. Our results are however not sensitive to the choice of cutting threshold.

<sup>4</sup> For more information on the International Standard Classification of Education (ISCED), see UNESCO Institute for Statistics (2012, 2014).



**Table 1** Main sample descriptive statistics

|   | Mean     | SD    | N      |
|---|----------|-------|--------|
| <b>A. Dependent variables (main)</b>        |          |       |        |
| PhD attainment                              | 0.011    | 0.104 |        |
| Academic career                             | 0.006    | 0.078 |        |
| <b>B. Dependent variables (comparison)</b>  |          |       |        |
| Any upper secondary education               | 0.776    | 0.417 |        |
| Gymnasium                                   | 0.428    | 0.495 |        |
| Any tertiary education                      | 0.395    | 0.489 |        |
| University (master's degree)                | 0.119    | 0.324 |        |
| Master in Education                         | 0.015    | 0.121 |        |
| Master in Arts                              | 0.004    | 0.061 |        |
| Master in Humanities                        | 0.014    | 0.119 |        |
| Master in Social sciences                   | 0.013    | 0.115 |        |
| Master in B&A                               | 0.014    | 0.119 |        |
| Master in Law                               | 0.005    | 0.072 |        |
| Master in Sciences                          | 0.012    | 0.109 |        |
| Master in ICT                               | 0.006    | 0.079 |        |
| Master in Engineering                       | 0.021    | 0.144 |        |
| Master in Agriculture                       | 0.002    | 0.050 |        |
| Master in Health sciences                   | 0.009    | 0.096 |        |
| <b>C. Control variables</b>                 |          |       |        |
| Year of birth                               | 1969.759 | 3.207 |        |
| Male  | 0.511    | 0.500 |        |
| <b>D. Explanatory variables<sup>a</sup></b> |          |       |        |
| Family size                                 | 2.572    | 1.418 |        |
| Maternal age                                | 23.559   | 4.197 |        |
| Mother's education <sup>b</sup>             | 11.129   | 2.479 |        |
| Father's education <sup>b</sup>             | 11.238   | 2.858 |        |
| Matriculation examination (ME)              | 0.453    | 0.465 |        |
| ME grade in mother tongue                   | 2.171    | 2.416 |        |
| ME grade average (excl. mother tongue)      | 1.941    | 2.192 |        |
| <b>E. Observations</b>                      |          |       |        |
| Individuals                                 |          |       | 722611 |
| Families                                    |          |       | 488597 |
| Neighborhoods                               |          |       | 3058   |

<sup>a</sup> Calculated at the family level to prevent overweighting larger families

<sup>b</sup> The values shown represent parental nominal years of schooling, but indicator variables representing eight different levels of education were used in the analysis itself

taken at the end of Gymnasium. We use a total of three measures of siblings' matriculation examination attainment. First, we derived an indicator variable denoting whether an individual had participated in the examination at all. Second, we derived two measures of exam performance: grade in the mother tongue test, which is compulsory for all participants, and mean grade of the following four tests: second national language, a foreign language,

mathematics, and a combined test in humanities and natural sciences; tests which the participants had to choose at least three from. In each test subject, the standardized grades range from 0 (failed) to 7 (excellent). If an individual had not participated in the examination at all, grade in mother tongue and mean grade were set to 0. Because the analysis is fundamentally on the family-level, we use family-level means of all these variables.

*Childhood neighborhood* was measured as the individual's postcode of residence at age 12.<sup>5</sup> If an individual's residence was unknown at age 12, we looked for information at age 13, and so on until the age of 15. As a result, postcodes were identified for 706,998 individuals (98% of the main sample). Finally, we included control variables denoting individuals' *gender* and *year of birth*. Descriptive statistics for all the derived variables are presented in Table 1.

## Empirical analysis

To assess the role of family background in selection into academia, we follow Vladasel et al. (2021) and estimate sibling correlations via latent models. We use latent models because both PhD attainment and academic careers are rare in the population, and sibling similarities in the observed binary outcomes—the manifest sibling similarities—are dependent on these attainment rates. For example, if country A has a lower PhD attainment rate than country B, then the manifest sibling similarities will also be lower in country A even if siblings in both countries would share the same underlying propensity to pursue a PhD. A latent model adjusts for this and produces more comparable estimates not only between different countries but also between different levels of education.

To calculate a sibling correlation, we first estimate the following latent response model:

$$y_{ij}^* = \mathbf{X}_{ij}\beta + u_i + e_{ij} \quad (1)$$

where the observed binary outcome  $y_{ij}$  for sibling  $j$  in family  $i$  is related to the continuous latent outcome  $y_{ij}^*$  via a threshold model where  $y_{ij} = 1$  if and only if  $y_{ij}^* > \theta$ . The vector  $\mathbf{X}_{ij}$  denotes an optional set of control and explanatory variables. Further,  $u_i$  is a family-level random effect (which siblings have in common), and  $e_{ij}$  the individual-level residual (indicating within-family sibling differences). Importantly, model (1) decomposes the variance of  $y_{ij}^*$  that is unexplained by  $\mathbf{X}_{ij}$  into two components: the family-level variance  $\sigma_u^2$  representing variation due to differences *between* families, and the individual-level variance  $\sigma_e^2$  representing variation due to differences *within* families. The sibling correlation  $\rho$  is then calculated by dividing the family-level variance by the sum of family and individual-level variances:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}. \quad (2)$$

That is, the sibling correlation  $\rho$  represents the fraction of the variance in an outcome which can be attributed to shared family background.  $\rho$  ranges between 0 and 1; the higher the correlation is, the stronger the role of family background (Björklund & Salvanes, 2011; Björklund & Jäntti, 2020).

<sup>5</sup> The residential data are available from 1977 onwards. We thus observe postcodes for the 1965 cohort at earliest the age of 12.

To answer our first research question on the importance of family background in selection into academia, we first calculate unadjusted sibling correlations in PhD attainment and academic careers by estimating model (1) for both outcomes without any control or explanatory variables. These unadjusted sibling correlations thus represent the share of the *overall* variation in PhD attainment and academic careers that can be attributed to family background. Furthermore, we estimate unadjusted sibling correlations in the attainment of lower levels of education, and use these as comparison outcomes to place the role of family background in PhD attainment and academic careers into a relative perspective.

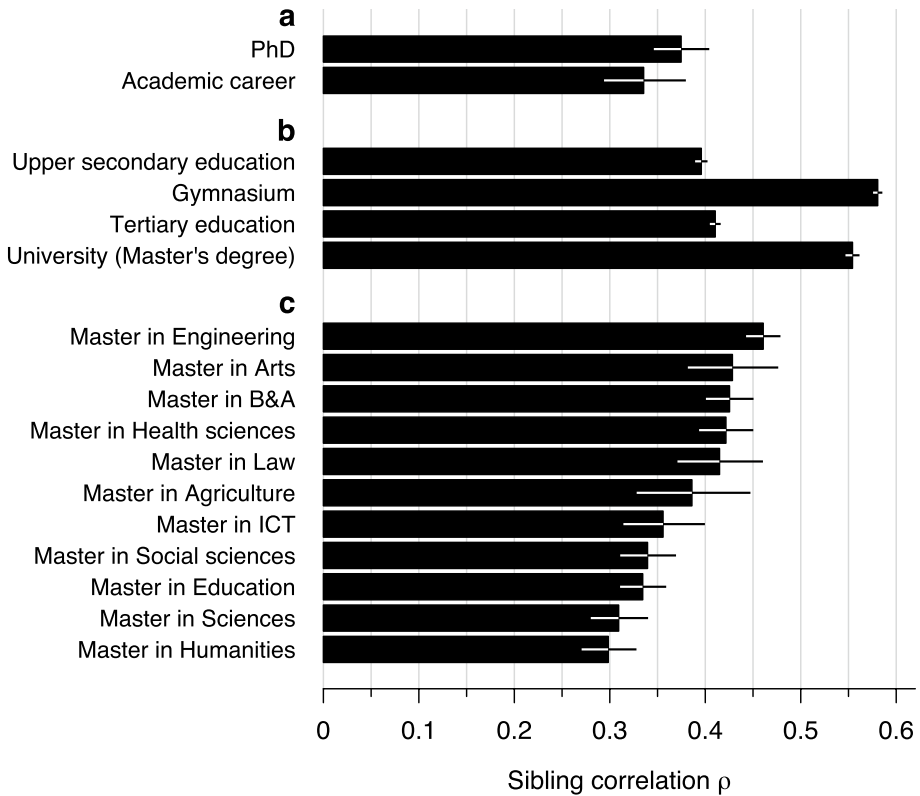
The analysis regarding our second research question is carried out in two parts. In the first part, we examine the extent to which parental education, prior educational achievement, family size and mother's age at first childbirth contribute to the between-family differences in PhD attainment and academic careers. We do this by first calculating baseline sibling correlations  $\rho^*$  while adjusting only for gender and year of birth. We then calculate residual sibling correlations  $\rho^*$  by including the aforementioned family characteristics as explanatory variables in the model both one at a time and simultaneously. The inclusion of a family characteristic will reduce the residual family-level variance in the model. The larger the share of the family-level variance that can be attributed to a particular family characteristic, the more the residual sibling correlation  $\rho^*$  will thus be reduced from its baseline value  $\rho$ . The relative difference between baseline  $\rho$  and residual  $\rho^*$  can be interpreted as the upper-bound contribution of the specific family characteristic in explaining the observed sibling similarities (e.g. Bredtmann & Smith, 2018; Vladasel et al., 2021). Similarly, by including multiple family characteristics simultaneously in the model, we may derive their joint contribution to the observed sibling similarities.

In the second part of the analysis of our research question two, we examine the extent to which childhood neighborhoods contribute to the between-family differences in PhD attainment and academic careers. To do this, we follow earlier literature (e.g. Bredtmann & Smith, 2018; Vladasel et al., 2021), and first estimate neighborhood correlations by substituting the family random effects in model (1) with neighborhood random effects. Neighborhood correlations  $\rho_{zip}$  represent the extent to which the outcomes of individuals growing up in the same neighborhood resemble each other. We then derive the residual sibling correlation  $\rho^*$  by subtracting the neighborhood correlation from the baseline sibling correlation ( $\rho - \rho_{zip}$ ). Similar to the above, the relative difference between baseline  $\rho$  and residual  $\rho^*$  may be used as a measure of the contribution of neighborhood influences in explaining the sibling similarities in PhD attainment and academic careers. Finally, to rule out some of the selection of families into neighborhoods, we also adjust the neighborhood correlations for parental education (at the neighborhood level) and then repeated the  $\rho - \rho_{zip}$  procedure. We adjust all neighborhood correlations for individuals' gender and year of birth.

All models are estimated in Stata using maximum likelihood estimation via the *xtprobit* command, under the assumption that  $u_i \sim \mathcal{N}(0, \sigma_u^2)$  and  $e_{ij} \sim \mathcal{N}(0, 1)$ . Since we use population data, there is no need of sample-to-population inference. We however report standard errors or 95% confidence intervals as an approximate indicator of the measurement uncertainty inherent in each correlation (e.g. Gelman & Hill, 2006, pp. 17–18).

## Results

We first assess the importance of family background in determining the attainment of PhDs and academic careers by estimating unadjusted sibling correlations. An unadjusted sibling correlation represents the fraction of the overall variation in an outcome that can be



**Fig. 1** Sibling correlations in the attainment of (a) PhDs and academic careers, (b) lower levels of education, and (c) master's degrees by fields of study.  $N = 722611$ . ML estimates derived from random-effects probit models with 95% confidence intervals

attributed to family background, and captures the influence of all the observed and unobserved family characteristics that children in the same family share. We find sibling correlations of 0.37 for PhD attainment and 0.34 for academic careers. This suggests that over a third of the overall variation in selection into Finnish academia can be attributed to family background.

Though a sibling correlation can be interpreted on its own as a decomposition into sibling-shared and other factors, we can also compare the sibling correlation estimates for PhDs and academic careers to sibling correlations for educational outcomes at lower levels of education. We illustrate our estimates of these in in Fig. 1, with the exact point estimates also reported in Appendix Table 3. Though one might be led to believe that because the attainment of a lower level degree is a less selective and less prestigious outcome than the attainment of a PhD or of an academic career, family background should play a smaller role in producing it, this is not a logical necessity. Indeed, a comparison of the correlation estimates in Fig. 1a and b shows that sibling similarities in PhD attainment and academic careers are close to sibling similarities in attaining any upper secondary or tertiary degree, while the sibling similarities in Gymnasium and University degrees are higher than those for PhD attainment or for academic careers.

It is possible to argue that a PhD is a university degree which prepares the student for a specific type of profession, not unlike a graduate degree in medicine or engineering, and that it would thus also be relevant to compare sibling correlations in PhDs to sibling correlations in degrees in specific fields. We show estimates for the latter in Fig. 1c. Sibling correlations for PhD degrees are smaller than for graduate-level degrees in engineering, medicine, or arts, but larger than for degrees in humanities, education, or natural sciences. Taken together, with the estimates of sibling correlations in PhD degrees and academic careers thus falling in-between the estimates for other educational outcomes, and well below some of them, there is little evidence to suggest that family background would play a larger role in either the attainment of PhD degrees or of academic careers than in many other educational outcomes.

### The contribution of specific family characteristics

We next assess the contribution of specific family characteristics in explaining the between-family differences in selection into academia. That is, we aim to shed light on what it is in the family that make siblings' attainment of PhDs and academic careers resemble each other. The results are summarized in Table 2. Concerning parental education, column (2) shows that including mother's and father's education as explanatory variables leads to a sizeable decrease in the correlations in comparison to the baseline estimates shown in column (1). Parents' education account for 33% and 24% of the baseline sibling similarities in PhD attainment and academic careers respectively, with the remainder of the sibling similarities unexplained by parental education.

Turning to the contribution of between-family differences in siblings' prior educational achievement, column (3) of Table 2 shows that including the family-level participation rate in the matriculation examination as an explanatory variable instead results in residual sibling correlations of 0.22 in PhD attainment and 0.20 in academic careers. This suggests that between-family differences in attending the matriculation examination explain almost 40% of the baseline sibling correlations. The additional inclusion of family-level exam performance leads to further decreases in the residual sibling correlations. Column (4) shows that matriculation examination participation and grades jointly explain 72% and 65% of the baseline sibling similarities in PhD attainment and academic careers respectively. Moreover, column (7) shows that including parents' education adds little explanatory value in addition to the variance already captured by the matriculation examination.

By contrast, columns (5) to (6) in Table 2 show that family size and maternal age contribute only between 4–8% to the baseline sibling correlations. Furthermore, column (8) shows that by including all the previously presented family characteristics simultaneously as explanatory variables in the models, we are able to explain 77% of the sibling similarities in PhD attainment, and 66% of the sibling similarities in academic careers. A comparison of columns (7) and (8) however also shows that the contribution of family size and maternal age is practically zero once parental education as well as the matriculation examination are taken into account.

In terms of the role of childhood neighborhood, column (9) in Table 2 shows that between 6–7% of the sibling similarities in PhD attainment and academic careers may be attributed to childhood neighborhoods. This may be considered small given that the

**Table 2** The contribution of specific family characteristics in the attainment of PhDs and academic careers

| Outcome  | Sibling correlations |                  |                  |                  |                  |                   |                  |                  |                  |                  |
|--|----------------------|------------------|------------------|------------------|------------------|-------------------|------------------|------------------|------------------|------------------|
|  | Baseline $\rho$      |                  |                  |                  |                  | Residual $\rho^*$ |                  |                  |                  |                  |
|  | (1)                  | (2)              | (3)              | (4)              | (5)              | (6)               | (7)              | (8)              | (9)              | (10)             |
| PhD attainment   | 0.364<br>(0.015)     | 0.243<br>(0.017) | 0.220<br>(0.016) | 0.103<br>(0.017) | 0.352<br>(0.015) | 0.337<br>(0.015)  | 0.086<br>(0.018) | 0.085<br>(0.018) | 0.338<br>(0.006) | 0.360<br>(0.006) |
| Academic career  | 0.325<br>(0.022)     | 0.247<br>(0.024) | 0.201<br>(0.023) | 0.116<br>(0.025) | 0.313<br>(0.022) | 0.304<br>(0.022)  | 0.111<br>(0.025) | 0.111<br>(0.025) | 0.306<br>(0.012) | 0.320<br>(0.010) |
| Family characteristics   |                      |                  |                  |                  |                  |                   |                  |                  |                  |                  |
| Mother's education   |                      | YES              |                  |                  |                  |                   | YES              | YES              |                  |                  |
| Father's education   |                      | YES              |                  |                  |                  |                   | YES              | YES              |                  |                  |
| Matriculation examination (ME)   |                      |                  | YES              | YES              |                  |                   | YES              | YES              |                  |                  |
| ME grade in mother tongue  |                      |                  | YES              | YES              |                  |                   | YES              | YES              |                  |                  |
| ME grade avg. (excl. mother tongue)  |                      |                  | YES              | YES              |                  |                   | YES              | YES              |                  |                  |
| Family size  |                      |                  |                  |                  | YES              |                   | YES              | YES              |                  |                  |
| Maternal age   |                      |                  |                  |                  |                  | YES               | YES              | YES              |                  |                  |
| Neighborhood   |                      |                  |                  |                  |                  |                   |                  |                  | YES              |                  |
| Neighborhood (adj.) <sup>a</sup>   |                      |                  |                  |                  |                  |                   |                  |                  |                  | YES              |
| % of baseline estimate explained by included family characteristics <sup>b</sup> |                      |                  |                  |                  |                  |                   |                  |                  |                  |                  |
| PhD attainment   | –                    | 33.3%            | 39.6%            | 71.6%            | 3.5%             | 7.5%              | 76.5%            | 76.8%            | 7.3%             | 1.1%             |
| Academic career  | –                    | 24.1%            | 38.1%            | 64.5%            | 3.7%             | 6.6%              | 65.9%            | 65.8%            | 5.9%             | 1.6%             |
| Observations   | 722611               | 722611           | 722611           | 722611           | 722611           | 722611            | 722611           | 722611           | 706998           | 706998           |

ML estimates derived from random-effects probit models. All models adjusted for gender and year of birth. Standard errors in parentheses. Residual sibling correlations in columns (9) and (10) are derived by subtracting neighborhood correlations from their respective baseline sibling correlations in column (1). Standard errors for columns (9) and (10) are derived by bootstrapping

<sup>a</sup> Adjusted for average parental education at the neighborhood level

<sup>b</sup>  $(\rho - \rho^*)/\rho \cdot 100\%$

selection of families into neighborhoods is not taken into account in this figure. To rule out some potential selection, we adjusted the estimates for average parental education levels within neighborhoods. Column (10) shows that, net of such selection, neighborhood characteristics contribute less than 2% to the sibling similarities in PhD attainment and academic careers.

## Discussion and conclusions

Family background is known to be an important determinant of educational attainment at lower levels of education (e.g. Björklund & Salvanes, 2011; Bredtmann & Smith, 2018). In this article, we have shown that this is also the case for the attainment of PhDs, and of subsequent academic careers. We based this conclusion on an analysis of sibling correlations estimated off decades-long series of Finnish full population microdata. A sibling correlation captures all sibling-shared variation, not just factors varying by discrete levels of a specific parental background variable, and it should thus be seen as a more holistic measure of family background than the measures that have been used in quantitative assessments of selection into doctoral education and academia so far.

In the cohorts we studied, we found a sibling correlation of 0.37 in PhD attainment, and of 0.34 in the attainment of an academic career. In other words, slightly over a third of the variation in these outcomes can be attributed to family background. This indicates that entry into Finnish academia runs in the family. At the same time, our results also show that a majority of the determinants, i.e. the other two-thirds, lie outside of the family. Furthermore, when we compared the role of family background in PhDs and academic careers to its role in various types of lower-level degrees, the former seems unexceptional in magnitude. A possible explanation for our findings is that, even if the influence of family background is long lived, the relative importance of other factors may increase en route into academia. One can think of factors such as peers, mentors, institutional support, and starting an own family, but also chance and luck (Denson & Szelényi, 2020; Lindholm, 2004; Martinsuo & Turkulainen, 2011; McCulloch, 2022).

We further decomposed our two sibling correlations by estimating the respective contributions of different family-level characteristics to the overall correlations. It is known that analyses which group individuals by the values of a specific parental background variable are likely to grossly underestimate the role of family background in producing the outcome (cf. Björklund & Jäntti, 2020; Bredtmann & Smith, 2018). We confirmed that this is the case for estimates of the role of family background in the attainment of PhDs and academic careers as well. In our data, parental education can for example only explain between a quarter and a third of sibling similarities in PhD attainment and academic careers, numbers similar to those found by Bredtmann and Smith (2018), and implying that a comparison by parental education would underestimate the role of family background by a factor of three to four. This suggests that researchers who compare academic outcomes by the levels of a specific parental background variable are likely to miss the large majority of quantitatively important aspects of family background that will almost necessarily remain uncaptured by their categorization.

The contributions of different family-level characteristics to the overall sibling correlation is also of interest in its own right. In line with previous findings on educational attainment (e.g. Bredtmann & Smith, 2018; Raaum et al., 2006), we found family structure and childhood neighborhood to explain only a small fraction of between-family differences in selection into academia. Notably, after adjusting for some of the selection of families into neighborhoods, we found the neighborhood influences in PhD attainment and academic careers to be close to non-existent. By contrast, we found that the individual's own prior educational achievement explained a majority of the between-family differences in selection into academia. Sibling similarities in Gymnasium matriculation examination attainment accounted for 72% of the sibling similarities in PhD attainment, and 65% of the sibling similarities in academic careers. Even if it is not unexpected that prior educational achievement should predict PhD attainment (cf. e.g. Bachsleitner et al., 2018; Mullen et al., 2003), the explanatory power of the matriculation examination seems high, and suggests that a large share of family differences in the selection into academia is observable already at the end of childhood, at least in Finland.

To our knowledge, there are no other studies which examine sibling similarities in doctoral or post-doctoral outcomes, and in this sense it is an open question whether our findings can be generalized to other countries. There are however important reasons to believe that their validity and relevance exceeds that of the Finnish context. It is for example known that sibling correlations in years of schooling are fairly similar across countries, even if they tend to be somewhat lower in Finland and other Nordic countries than for example in Germany or the US (Björklund & Jäntti, 2020; Grätz et al., 2021). Our findings are also consistent with previous studies which have shown that parental characteristics explain sibling similarities better than do childhood neighborhoods (e.g. Bredtmann & Smith, 2018; Raaum et al., 2006; Vladasel et al., 2021), even if on an absolute level, neighborhoods are a non-trivial source of inequality at least in the US (Solon et al., 2000).

We have shown that the attainment of PhDs and academic careers is partially determined by the family to which one happens to be born. Under some interpretations of equality of opportunity (Roemer, 1998, 2004), this finding can be seen as indicative of a deficiency of social justice. Even if one feels that it is not, selection into academia not only affects the selected, but also those who are taught and advised by them, and our results can thus also be seen as indicative of a problem of representation among those who will shape the educational experiences of the next generation.

To the researchers studying the obstacles which individuals from disadvantaged backgrounds face in their academic careers, we believe that our study can be a point of reference which complements their work. It does this by showing that some factors determining academic success have their origins in family background, but that quantitatively speaking, most do not; by showing that some of these family background factors can be attributed to a specifically measurable parental characteristic, but that quantitatively, most can not; by showing that quantitatively and on average, academia is not exceptional in this respect compared to other educational and occupational outcomes; and by showing that quantitatively, many of the between-family differences in PhD attainment and in the attainment of academic careers are visible also already at younger ages. The last finding also provides an important avenue for those who seek to change the status quo: early educational



performance and selection into academic profession go hand in hand. A deeper understanding on how families foster educational performance is necessary not only in discovering the roots of this association but also in determining whether altering this association should be considered a socially just action.

## Limitations

We acknowledge several limitations to our study. For reasons of data availability, we have excluded from our sample both foreign-born individuals and individuals without a known mother, and our results are thus not representative of these groups of Finnish residents. Both groups are however very small in size for the 1965–1975 birth cohorts used in our study. We have similarly excluded from our sample permanent emigrants from Finland, and our results should thus not be interpreted as informative of them.

Over 90% of PhDs employed by Finnish institutes of higher education are employed by a research university (Haila et al., 2016), with the remainder employed by a university of applied sciences. We were unable to distinguish between employment at either type of institution, and could thus also not estimate the role of family background in selection into academic careers separately for either type of institution. Because of the much higher prevalence of university employment among PhDs, our results should be interpreted as being primarily informative of academic careers at research universities.

Another limitation relates to our neighborhood measurement, which refers to residence at a specific age rather than the complete set of neighborhoods the individual has resided in, and this can downward bias estimates of neighborhood effects (Raaum et al., 2006). One could furthermore ask whether a postcode is not too large of a geographical area to accurately capture the influence of immediate peers and other role models. Postcodes are however the smallest geographical unit available in our data, and are similar in size to those used in many previous studies (e.g. Bredtmann & Smith, 2018; Lindahl, 2011; Vladasel et al., 2021).

One should also bear in mind that a sibling correlation measures the role of siblings' shared family background, but not all family background is fully shared between siblings. Siblings for example share only part of their genetic make-up, and different siblings almost necessarily also have different childhood experiences, with older siblings for example spending the first years of their life with younger parents and a smaller family than their younger siblings will. Like other common family background measures, sibling correlations do not pick up such within-family differences and may thus be thought to underestimate the total importance of family background as seen from the individual child's perspective.

Finally, our focus has been on investigating between-family inequality in PhD attainment and academic careers. There exist further dimensions of inequality which are not fully captured in these outcomes, such as those related to career progression, remuneration, and work tasks. While some studies have addressed these dimensions (e.g. Chiappa & Mejias, 2019; Passaretta et al., 2019), we welcome more research on family background influences with respect to these other outcomes as well.

## Appendix

**Table 3** Unadjusted and adjusted sibling correlations ( $N = 722611$ )

| Outcome                      | Sibling correlations |                  |
|------------------------------|----------------------|------------------|
|                              | unadjusted $\rho$    | adjusted $\rho$  |
|                              | (1)                  | (2)              |
| PhD                          | 0.375<br>(0.015)     | 0.364<br>(0.015) |
| Academic career              | 0.335<br>(0.022)     | 0.325<br>(0.022) |
| Upper secondary education    | 0.396<br>(0.003)     | 0.399<br>(0.003) |
| Gymnasium                    | 0.581<br>(0.002)     | 0.608<br>(0.002) |
| Tertiary education           | 0.410<br>(0.003)     | 0.423<br>(0.003) |
| University (Master's degree) | 0.554<br>(0.004)     | 0.548<br>(0.004) |
| Master in Engineering        | 0.461<br>(0.009)     | 0.475<br>(0.009) |
| Master in Arts               | 0.428<br>(0.024)     | 0.436<br>(0.025) |
| Master in B&A                | 0.425<br>(0.013)     | 0.433<br>(0.015) |
| Master in Health sciences    | 0.422<br>(0.014)     | 0.423<br>(0.013) |
| Master in Law                | 0.415<br>(0.023)     | 0.409<br>(0.023) |
| Master in Agriculture        | 0.386<br>(0.031)     | 0.385<br>(0.031) |
| Master in ICT                | 0.356<br>(0.022)     | 0.366<br>(0.023) |
| Master in Social sciences    | 0.340<br>(0.015)     | 0.351<br>(0.013) |
| Master in Education          | 0.334<br>(0.012)     | 0.345<br>(0.015) |
| Master in Sciences           | 0.309<br>(0.015)     | 0.310<br>(0.015) |
| Master in Humanities         | 0.298<br>(0.015)     | 0.300<br>(0.015) |

ML estimates derived from random-effects probit models. Standard errors in parentheses. Fig. 1 in the main text is based on the unadjusted estimates shown in column (1). Estimates in column (2) are adjusted for gender and year of birth (cf. baseline  $\rho$ 's in Table 2).

**Funding** Open Access funding provided by University of Jyväskylä (JYU). Jouni Helin gratefully acknowledges funding from the Finnish Cultural Foundation's Satakunta Regional Fund (personal grant number 75202315).

## Declarations

**Conflict of Interests** The authors declare that they have no competing interests.

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

## References

- Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica*, 65(5), 1005–1027.
- Andersen, H. (2001). The norm of universalism in sciences. Social origin and gender of researchers in Denmark. *Scientometrics*, 50(2), 255–272.
- Bachsleitner, A., Becker, M., Neumann, M., & Maaz, K. (2018). Social background effects in the transition to a doctoral degree—Empirical evidence from a German prospective study. *Research in Social Stratification and Mobility*, 57, 24–34.
- Bahack, H., & Addi-Raccach, A. (2022). PhD first-generation and continuing generation students' academic experience and strengths. *Higher Education*, 1–17.
- Björklund, A., & Jäntti, M. (2020). Intergenerational mobility, intergenerational effects, sibling correlations, and equality of opportunity: A comparison of four approaches. *Research in Social Stratification and Mobility*, 70, 100455.
- Björklund, A., Lindahl, L., & Lindquist, M. J. (2010). What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents. *The BE Journal of Economic Analysis & Policy*, 10.
- Björklund, A., & Salvanes, K. G. (2011). Education and family background: Mechanisms and policies. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education*, (Vol. 3 pp. 201–247). Elsevier.
- Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). New York: Greenwood Press.
- Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social Forces*, 92(1), 109–140.
- Bredtmann, J., & Smith, N. (2018). Inequalities in educational outcomes: How important is the family? *Oxford Bulletin of Economics and Statistics*, 80 (6), 1117–1144.
- Britton, J., Buscha, F., Dickson, M., van der Erve, L., Vignoles, A., Walker, I., Waltmann, B., & Zhu, Y. (2020). *The earnings returns to postgraduate degrees in the UK*. Institute for Fiscal Studies.
- Chiappa, R., & Mejias, P. P. (2019). Unfolding the direct and indirect effects of social class of origin on faculty income. *Higher Education*, 78(3), 529–555.
- Denson, N., & Szelényi, K. (2020). Faculty perceptions of work-life balance: The role of marital/relationship and family status. *Higher Education*, 1–18.
- Erikson, R., & Jonsson, J. O. (1996). Introduction. explaining class inequality in education: The Swedish test case. In R. Erikson & J. O. Jonsson (Eds.), *Can education be equalized? The Swedish case in comparative perspective* (pp. 1–63). Boulder: Westview.
- Gardner, S. K., & Holley, K. A. (2011). “Those invisible barriers are real”: The progression of first-generation students through doctoral education. *Equity & Excellence in Education*, 44(1), 77–92.
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

- Grätz, M., Barclay, K. J., Wiborg, Ø.N., Lyngstad, T. H., Karhula, A., Erola, J., Präg, P., Laidley, T., & Conley, D. (2021). Sibling similarity in education across and within societies. *Demography*, 58(3), 1011–1037.
- Gurin, P., Dey, E., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review*, 72(3), 330–367.
- Haas, C., & Hadjar, A. (2020). Students' trajectories through higher education: A review of quantitative research. *Higher Education*, 79(6), 1099–1118.
- Haila, K., Karinen, R., Kaihovaara, A., Eronen, A., & Haapakorpi, A. (2016). Miten tohtorit työllistyvät [Employment situation of people with PhDs]. Opetus- ja kulttuuriministeriön julkaisuja 2016:3, Helsinki.
- Helin, J., Koerselman, K., Nokkala, T., Tohmo, T., & Viinikainen, J. (2019). Equal access to the top? Measuring selection into Finnish academia. *Social Inclusion*, 7(1), 90–100.
- Hermanowicz, J. C. (2012). The sociology of academic careers: Problems and prospects. In J. C. Smart & M. B. Paulsen (Eds.), *Higher education: Handbook of theory and research* (pp. 207–248). Dordrecht: Springer.
- Holopainen, H. (2017). Tohtoreiden monet urat: tohtoreiden osaamisen hyödyntäminen työelämässä yliopistojen ulkopuolella. Finnish Education Employers.
- Koerselman, K., & Uusitalo, R. (2014). The risk and return of human capital investments. *Labour Economics*, 30, 154–163.
- Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income—evidence from Sweden. *The Journal of Economic Inequality*, 9(2), 207–226.
- Lindholm, J. A. (2004). Pathways to the professoriate: The role of self, others, and environment in shaping academic career aspirations. *The Journal of Higher Education*, 75(6), 603–635.
- Marginson, S. (2016). The worldwide trend to high participation higher education: Dynamics of social stratification in inclusive systems. *Higher Education*, 72(4), 413–434.
- Marginson, S. (2018). Equity. In B. Cantwell, S. Marginson, & A. Smolentseva (Eds.), *High participation systems of higher education*. Oxford University Press.
- Martinsuo, M., & Turkulainen, V. (2011). Personal commitment, support and progress in doctoral studies. *Studies in Higher Education*, 36(1), 103–120.
- Mastekaasa, A. (2006). Educational transitions at graduate level: Social origins and enrolment in PhD programmes in Norway. *Acta Sociologica*, 49(4), 437–453.
- Mateos-González, J. L., & Wakeling, P. (2022). Exploring socioeconomic inequalities and access to elite postgraduate education among English graduates. *Higher Education*, 83(3), 673–694.
- McCallum, C. M. (2016). “Mom made me do it”: The role of family in African Americans' decisions to enroll in doctoral education. *Journal of Diversity in Higher Education*, 9(1), 50.
- McCulloch, A. (2022). Serendipity in doctoral education: the importance of chance and the prepared mind in the phd. *Journal of Further and Higher Education*, 46(2), 258–271.
- Mertens, A., & Røbken, H. (2013). Does a doctoral degree pay off? An empirical analysis of rates of return of German doctorate holders. *Higher Education*, 66(2), 217–231.
- Merton, R. K. (1979). The normative structure of science. In N. W. Storer (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 267–278). The University of Chicago Press.
- Möller, C. (2014). Wie offen ist die universitätsprofessur für soziale aufsteigerinnen und aufsteiger? *Soziale Welt*, 64(4), 341–360.
- Mullen, A. L., Goyette, K. A., & Soares, J. A. (2003). Who goes to graduate school? Social and academic correlates of educational continuation after college. *Sociology of Education*, 143–169.
- Oldfield, K., & Conant, R. F. (2001). Professors, social class, and affirmative action: A pilot study. *Journal of Public Affairs Education*, 7(3), 171–185.
- Passaretta, G., Trivellato, P., & Triventi, M. (2019). Between academia and labour market—the occupational outcomes of PhD graduates in a period of academic reforms and economic crisis. *Higher Education*, 77(3), 541–559.
- Perna, L. W. (2006). Studying college access and choice: A proposed conceptual model. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research*, (Vol. 21 pp. 99–157). Springer.
- Pinto, T., & Teixeira, A. A. C. (2020). The impact of research output on economic growth by fields of science: a dynamic panel data analysis, 1980–2016. *Scientometrics*, 123(2), 945–978.
- Posselt, J. R., & Grodsky, E. (2017). Graduate education and social stratification. *Annual Review of Sociology*, 43, 353–378.
- Raam, O., Salvanes, K. G., & Sørensen, E. (2006). The neighbourhood is not what it used to be. *The Economic Journal*, 116(508), 200–222.
- Reimer, D., & Jacob, M. (2011). Differentiation in higher education and its consequences for social inequality: Introduction to a special issue. *Higher Education*, 61(3), 223–227.

- Rockinson-Szapkiw, A. J., Sosin, L. S., & Spaulding, L. (2018). Does family matter? A phenomenological inquiry exploring the lived experiences of women persisting in distance education, professional doctoral programs. *International Journal of Doctoral Studies*, 13.
- Roemer, J. E. (1998). *Equality of opportunity*. Harvard University Press.
- Roemer, J. E. (2004). Equal opportunity and intergenerational mobility: Going beyond intergenerational income transition matrices. In M. Corak (Ed.), *Generational income mobility in North America and Europe* (pp. 48–57). Cambridge: Cambridge University Press.
- Shavit, Y., Arum, R., Gamoran, A., & Menahem, G. (2007). *Stratification in higher education: A comparative study*. Stanford University Press.
- Solon, G., Page, M. E., & Duncan, G. J. (2000). Correlations between neighboring children in their subsequent educational attainment. *Review of Economics and Statistics*, 82(3), 383–392.
- Suhonen, T., & Jokinen, J. (2018). Mikä on tutkintotodistuksesi tuotto? *Talous ja yhteiskunta*, 46(2), 30–37.
- Swift, A. (2004). Would perfect mobility be perfect? *European Sociological Review*, 20(1), 1–11.
- Triventi, M. (2013). Stratification in higher education and its relationship with social inequality: A comparative study of 11 European countries. *European Sociological Review*, 29(3), 489–502.
- UNESCO Institute for Statistics. (2012). International Standard Classification of Education (ISCED) 2011. UNESCO Institute for Statistics.
- UNESCO Institute for Statistics. (2014). ISCED Fields of Education and Training 2013 (ISCED-F 2013): Manual to accompany the International Standard Classification of Education. UNESCO Institute for Statistics.
- Välilmaa, J., & Muhonen, R. (2018). Reproducing social equality across the generations: The Nordic model of high participation higher education in Finland. In B. Cantwell, S. Marginson, & A. Smolentseva (Eds.), *High participation systems of higher education* (pp. 358–385). Oxford University Press.
- Välilmaa, J., Papatsiba, V., & Hoffman, D. M. (2016). Higher education in networked knowledge societies. In D. M. Hoffman & J. Välilmaa (Eds.), *Re-becoming universities? Higher education institutions in networked knowledge societies*. Springer.
- Vladasel, T., Lindquist, M. J., Sol, J., & van Praag, M. (2021). On the origins of entrepreneurship: Evidence from sibling correlations. *Journal of Business Venturing*, 36(5), 106017.
- Zimdars, A.K. (2007). Testing the spill-over hypothesis: Meritocracy in enrolment in postgraduate education. *Higher Education*, 54(1), 1–19.

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