When Less Conditioning Provides Better Estimates: Overcontrol and Collider Bias in Research on Intergenerational Mobility*

This version: April 4, 2019

Michael Grätz[a, b]

[a] Swedish Institute for Social Research (SOFI), Stockholm University, Sweden

[b] Institute for Futures Studies, Stockholm, Sweden

E-Mail: michael.gratz@sofi.su.se

Abstract

The counterfactual approach to causality has become the dominant approach to understand causality in contemporary social science research. Whilst most sociologists are aware that unobserved, confounding variables may bias estimates of causal effects, the issues of overcontrol and collider bias have received comparatively less attention. In particular, widely used practices in research on intergenerational mobility require conditioning on variables that are endogenous to the process of the intergenerational transmission of advantage. I review four of these practices from the viewpoint of the counterfactual approach to causality and show that overcontrol and collider biases arise when these practices are implemented. I use data from the German Socio-Economic Panel Study (SOEP) to demonstrate the practical consequences of these biases for conclusions about intergenerational mobility. Future research on

* Please quote as: Grätz, Michael. 2019. When Less Conditioning Provides Better Estimates: Overcontrol and Collider Bias in Research on Intergenerational Mobility. Paper presented at the Spring Meeting of the Research Committee 28 of the International Sociological Association in Frankfurt.

intergenerational mobility should reflect more upon the possibilities of bias introduced by conditioning on variables.

Keywords

causality, collider bias, directed acyclic graphs, intergenerational mobility, overcontrol bias

In the last decades, research practices in the social sciences have seen major changes due to the emergence of the counterfactual approach to causality as the primary way to understand causality in social science research (e.g. Angrist and Pischke, 2009 in economics; Gangl, 2010; Morgan and Winship, 2015 in sociology; Holland, 1986; Pearl, 2009; Rubin, 1974 in statistics; VanderWeele, 2015 in epidemiology; Acharya *et al.*, 2016; Imai *et al.*, 2011 in political science). An important aspect of the counterfactual approach to causality is the increased importance for researchers to spell out and to justify the assumptions needed to interpret statistical estimates as causal effects because of the possibility of omitted variable bias.

Whilst the issue of omitted variable bias due to unobserved variables is nowadays widely known among sociologists, the dangers of overcontrol and collider bias have received comparatively less attention. With a focus on collider bias, Elwert and Winship (2014) discussed some of the issues arising in this context but these insights have not yet fully translated into research practices in all research fields within the social sciences. A possible reason for this lack of impact of the counterfactual approach to causality on research practices in the social sciences is that the insights coming from this approach have not yet been applied with respect to each research field. The aim of this manuscript is to apply the insights arising from applying the counterfactual approach to causality to the study of the intergenerational transmission of advantage.

The main aim of research on intergenerational mobility is to provide descriptive estimates of the similarity in resources between parents and their offspring (Torche, 2015). These descriptive estimates provide important yardsticks about the degree of equality of opportunity in contemporary and historic societies. What is important to note, however, is that even though the focus of this research is on providing descriptive estimates, there are important insights that can be gained from applying the counterfactual approach to causality. These insights have, however, not yet been fully acknowledged by all researchers studying intergenerational

mobility. Torche (2015) touched on some of the issues that arise if we apply the counterfactual framework of causality to the study of intergenerational mobility, for instance the interpretation of underlying mechanisms based on path-analytical models. However, there are other widely used practices in research on intergenerational mobility that seem problematic from the perspective of the counterfactual approach to causality.

In this manuscript, I discuss important insights that can be obtained by applying the counterfactual approach to causality to the analysis of intergenerational mobility. I discuss four widely used practices in mobility research that lead to overcontrol and collider bias. I illustrate the practical consequences of these biases using data from the German Socio-Economic Panel Study (SOEP), a data set that is widely used to study intergenerational mobility in Germany (e.g. Grätz and Pollak, 2016; Hertel, 2017; Müller and Pollak, 2004). Although I illustrate these issues only with respect to one country, the issues I identify will arise in other countries as well. The aim of my treatment is not to criticize past research but to bring the issues of overcontrol and collider bias to the attention of the community of researchers who work on intergenerational mobility so that they can take these issues into account in their future work.

BACKGROUND AND THEORETICAL CONSIDERATIONS

Description and Causal Inference in Research on Intergenerational Mobility

The counterfactual approach to causality has changed the way causal analyses are conducted in the social sciences (Angrist and Pischke, 2009; Gangl, 2010; Holland, 1986; Morgan and Winship, 2015; Pearl, 2009; Rubin, 1974; VanderWeele, 2015). Many researchers in sociology reacted to the emergence of counterfactual thinking by emphasizing that they are not interested in providing causal but only aim at obtaining descriptive estimates of processes of interest. Descriptive estimates are important and a research field such as intergenerational mobility cannot proceed without these descriptive estimates as they provide important statistics about

equality of opportunity in contemporary and historic societies. However, there are important insights coming from the counterfactual approach to causality that also apply to how descriptive estimates are obtained. My impression is that these insights have so far often been overlooked, for instance by research on intergenerational mobility.

Any distinction between descriptive and causal research is artificial, as every identification of causal effects is based on assumptions. On the other hand, all descriptive research is motivated by an idea about the underlying causal processes of interest. Most sociologists acknowledge nowadays that bias through omitted, unobserved variables provides a threat to the identification of causal effects. Even if we are not able to control for all unobserved variables and therefore cannot identify causal effects, we would, however, like to make sure that we do not control for variables that make it even more unlikely that the descriptive estimates, which we report, represent underlying causal effects.

To the group of variables that we do not want to control for belong variables that lead to overcontrol bias and variables that lead to collider bias. Controlling for variables that lie on the causal pathway between the two variables of interest, i.e. the treatment and the outcome, leads to overcontrol bias (Elwert and Winship, 2014). Collider bias arises if we control for a variable (a mediator) that is not only affected by the cause (treatment) but also by another variable that also affects the outcome (Elwert and Winship, 2014). A particular problematic version of collider bias arises when the variables that confound the relationship between the mediator and the outcome are themselves affected by the treatment. Acharya *et al.* (2016) called this type of collider bias intermediate variable bias. In this situation, even controlling for these confounding variables introduces a bias, as the direct effect will then be estimated with overcontrol bias.

Controlling for a variable on the causal pathway can be intentional if researchers want to test mechanisms. In this respect, a field of causal mediation analysis is developing in particular in epidemiology and in political science (Acharya *et al.*, 2016; Imai *et al.*, 2011; Robins and

Greenland, 1992; Pearl, 2009; VanderWeele, 2015). These researchers developed traditional mediation analysis (Baron and Kenny, 1986) further trying to address some of the biases I identify below. Research on intergenerational mobility has not yet integrated these insights. There are, however, research practices in research on intergenerational mobility that require causal mediation analysis, for instance, the practice to control for children's educational attainment when estimating the association between social origin and children's labor market outcomes (research practice 3 below). The starting point of such an analysis is, as in any type of mediation analysis, the estimation of a total effect, which is then decomposed into an indirect and a direct component.

In this manuscript, I focus on four widely used practices in research on the intergenerational transmission of advantage that have in common that they lead to overcontrol and collider bias. In other words, these practices require researchers to control for variables that lie on the causal pathway connecting the two components of the intergenerational mobility process, social origin and social destination. In two instances, there is also a substantial danger of collider bias arising through these practices. In the following, I discuss these widely used research practices, the overcontrol and collider biases arising through them, and the implications of these biases for our conclusions about intergenerational mobility.

Understanding Practices in Intergenerational Mobility Research from the Viewpoint of the Counterfactual Approach to Causality

In the following, I discuss four research practices that are widely used in studies estimating intergenerational mobility but that seem problematic from the viewpoint of the counterfactual approach to causality. The focus of my discussion is on the causal models researchers need to entertain in order to justify the application of these practices. My impression is that many

researchers are not aware of the assumptions underlying these research practices and the overcontrol and collider biases that arise if they are employed.

Throughout the following discussion, I use Directed Acyclic Graphs (DAGs) to motivate the issues at stake. DAGs are a convenient and easy way to operationalize theoretical assumptions and are an important toolbox of counterfactual thinking (Elwert, 2013; Elwert and Winship, 2014; Pearl, 2009). DAGs make it possible to discuss the assumptions underlying empirical research in an intuitive way. For instance, Breen (2018) used DAGs to discuss collider bias arising in three-generational mobility studies.

DAGs operationalize hypothesized causal effects through arrows. I follow the convention in the literature to indicate variables that are conditioned on by a rectangular box (Elwert, 2013). In addition, I do indicate unobserved variables in the models by the term "Unobserved". Even though they are only indicated through one node, there usually are several possible unobserved variables in each graph so it is best to think about every node with the term "Unobserved" as indicating a vector of unobserved variables that can confound the relationship of interest. I perceive the influence of unobserved variables (omitted variable bias) as a general problem as there is always the danger that unobserved variables affect the estimates of intergenerational mobility.

Figure 1 represents graphically the relationship between parental resources (social origin) and child outcomes (social destination). These relationships are the focus of research on intergenerational mobility. The figure also indicates that there may be unobserved variables confounding the relationship between social origin and social destination. This is why estimates of intergenerational mobility usually have to be interpreted associationally and cannot be interpreted causally.

[FIGURE 1 ABOUT HERE]

According to Torche (2015), social mobility research should focus on estimating the relationships portrayed in Figure 1. Many studies on intergenerational mobility take additional steps to go beyond the bivariate relationship portrayed in this figure. These additional steps require additional assumptions. They also require researchers to have in mind a specific causal model (also called a data-generating process) of the relationships they are interested in studying. It is in this context that the four research practices, which I discuss in the following, lead to overcontrol and collider bias.

Research Practice 1: Controlling for Multiple Dimensions of Social Origin in the Same Model

The first research practice that introduces overcontrol bias is the practice to include several measures of social origin, for instance parental education, occupation, status, income, social capital, cultural capital, and wealth, in the same model as independent variables. One of the founders of empirical sociology Paul F. Lazarsfeld (1939) argued that different measures of social origin could be used interchangeably. In the following decades, however, the status attainment approach became the dominant approach to the study of the intergenerational transmission of advantage. The status attainment approach was based on path-analytical models in which the effects of various indicators of family background were distinguished by conditioning on the other indicators (Blau and Duncan, 1967; Featherman and Hauser, 1978; Sewell *et al.*, 1969; Sewell and Hauser, 1975).

The status attainment tradition influences contemporary practices in research on intergenerational mobility. Authors following this tradition have argued that different dimensions of family background have independent effects on respondent's educational and occupational attainment (Bukodi and Goldthorpe, 2013; Erola *et al.*, 2016; Hällsten and Pfeffer, 2017; Hällsten and Thaning, 2018; Jæger, 2007; Jæger and Holm, 2007; Mood, 2017;

Pfeffer, 2018; Wong, 1998). These claims were made using models in which different indicators of social origin were entered in the same model and in which the results showed that these different indicators had statistically significant and substantively large associations with the outcome (some measure of children's social destination) after conditioning on the other indicators of social origin.

In Figure 2, I illustrate the causal model, i.e. the data-generating process, that researchers need to entertain if they include measures of parental education, occupation, and income into the same model. The same logic applies if more indicators of social origin (parental wealth, cultural capital, social capital, and others) are added to the model but the three variables I choose are enough to illustrate my point.

[FIGURE 2 ABOUT HERE]

A first insight from the DAG in Figure 2 is that the estimate of the effect of parental education is biased if the same model controls for parental occupation and parental income. Clearly, parental income and parental occupation are variables lying on the causal pathway between parental education and social destination and they should therefore not be controlled for when researchers are interested in estimating the causal effect of parental education on children's outcomes.

Whilst probably most researchers are going to agree that parental education precedes parental occupation and parental income, the situation is less clear with respect to other measures of social origin. When entering parental occupation and parental income in the same model researchers have to assume that they are not affecting each other. But parental income and parental occupation are two characteristics directly related to a person's job. It is therefore likely that they affect each other. What is more, it is difficult to make an argument that parental

income precedes or proceeds parental occupation. Rather, both are jointly determined but the precise process is usually unknown to mobility researchers.

Sociologists seem to believe that occupation generally precedes income. For instance, Mood (2017: 266) argued that "earnings are a more "final" measure of success, affected by rather than affecting education and occupation". In addition, Erola *et al*. (2016) employed a model in which parental occupation preceded parental income. However, often parental occupation and parental income are determined at the same moment. If that is the case, it is very difficult to argue that one precedes the other. Because the relationship between parental income and parental occupation is theoretically ambiguous, it is best not to control for the other variable in the same model but to compare the estimates obtained using separate models. Alternatively, researchers should model the process by which parental income and parental occupation influence each other. I am not aware of any study that has done the latter.

To sum up the discussion of the research practice to include measures of parental education, income, and occupation in the same model predicting children's outcomes, the counterfactual approach to causality cautions against including several indicators of social origin in the same model as independent variables. This viewpoint is clearly at odds with the practices of stratification scholars following in the tradition of status attainment research (Blau and Duncan, 1967; Featherman and Hauser, 1978; Sewell *et al.*, 1969; Sewell and Hauser, 1975). For instance, Mood (2017: 282) argued that "from the child's viewpoint, parental education, occupation, and income are contemporaneous". There are certainly reasons for taking such a view but it cannot be maintained if we are interested in knowing what the causal effect of increasing parental education on children's outcomes is. However, this is the policy-relevant question as it is difficult to imagine a policy that raises parental education but that does neither affect parental occupation nor parental income. Controlling for occupation and income at the parental level when estimating the effects of parental education on children's outcomes

introduces overcontrol bias precisely because a high level of education allows parents to have a high income and a high level of occupation, which then may be beneficial for their offspring.

Research Practice 2: Controlling for Father's and Mother's Characteristics in the Same Model

The second research practice that I analyse, including separate characteristics of mothers and fathers in the same model, is directly related to the first one. This practice has received a lot of support in intergenerational mobility research, as it is theoretically appealing. Researchers have argued that not including maternal along paternal characteristics into the same model of intergenerational mobility provides a misrepresentation of social mobility in particular in contemporary societies (Beller, 2009; Bloome and Western, 2011; Buis, 2013; Jæger, 2007; Kalmijn, 1994; Korupp *et al.*, 2002; Marks, 2008). Models that include both maternal and paternal characteristics are usually interpreted as showing that both mothers and fathers influence children's outcomes.

This viewpoint is theoretically appealing. However, the methodological problems that arise when this approach is implemented have so far been overlooked. From the counterfactual perspective, it is problematic to control for characteristics of fathers and mothers in the same model for the same reason that makes it problematic to control for different indicators of social origin in the same model. Paternal and maternal education, occupation, and income influence each other and it is not clear which of these two (or, if combined with the first research practice that includes several measures of social origin in the same model, six) variables precedes the other. All what we know is that there is, even in modern societies, a strong tendency among men and women to form relationships with partners with similar educational and socioeconomic characteristics, a process called assortative mating (Blossfeld, 2009; Schwartz, 2013).

As an example of this research practice, Figure 3 presents the relationships between father's occupation, mother's occupation, and child education. Researchers who enter father's and mother's occupation in the same model need to assume that they do not affect each other. But because of assortative mating this is certainly not the case. Therefore, controlling for the other parent's occupation leads to overcontrol bias.[1] For this reason, from a counterfactual perspective it is best not to control for the other parent's characteristics when estimating the relationship between father's and mother's characteristics and children's outcomes.

[FIGURE 3 ABOUT HERE]

It is important to point out that this critique is not a critique of the theoretical viewpoint that both fathers and mothers influence children's outcomes but of the methods that research has so far employed to provide empirical support for this claim. Previous research did not provide any support for the claim that mothers and fathers do have independent effects on children because the studies all employed models that controlled for both maternal and paternal characteristics in the same model (Beller, 2009; Bloome and Western, 2011; Buis, 2013; Jæger, 2007; Kalmijn, 1994; Korupp *et al.*, 2002; Marks, 2008). The estimates obtained controlling for the other's parents characteristics are, however, different from the causal effects of maternal or paternal characteristics on children's outcomes.

Ideally, researchers that want to test whether mothers and fathers have independent causal effects on children should analyse the effects of increases in maternal or paternal resources that the researchers can convincingly show to be unrelated to the other parent. For instance, researchers could exploit educational reforms that affected women but had no effects on men. It is only these estimates that can be used to identify the unique effects of fathers and

mothers on children's outcomes. I am not aware of any study that has so far implemented such an approach.

Research Practice 3: Controlling for Children's Educational Attainment when Estimating the Relationship between Social Origin and Children's Labor Market Outcomes

Especially in research on occupational mobility, the association between social origin and social destination is often estimated conditional on educational attainment (Bernardi and Ballarino, 2016; Blau and Duncan, 1967; Breen, 2004; Featherman and Hauser, 1978; Ishida *et al.*, 1995; Sewell and Hauser, 1975). This research tradition refers to the relationship between social origin, education, and social destination as the origin-education-destination (OED) triangle. Similarly, some studies investigated the role of education in income mobility (Bloome et al., 2018; Bloome and Western, 2011; Gregg *et al.*, 2017). A related topic is estimating how the effect of parental occupation on respondent's occupation is moderated by respondent's education. This has been done by including an interaction between parental occupation and children's education (Hout, 1988; Torche, 2011). Finally, a further variation of this approach conditions on education by running the analysis on a selected sample, e.g. focusing only on graduates from institutions of tertiary education (Jacob *et al.*, 2015) or only on those leaving school with the lowest educational degree (Holtmann *et al.*, 2017).

These three different research questions can be conceptualized through a common data-generating process. Figure 4 presents the general model between social origin (parental income, education, or occupation), child education, and social destination (child occupation or income) that studies following this research practice estimate.

[FIGURE 4 ABOUT HERE]

The figure shows that estimating these relationships requires conditioning on an endogenous variable (child education). Admittedly, this is the purpose of this research practice: to decompose the total effect of social origin on social destination in an indirect one mediated by educational attainment and a direct one net of educational attainment. In any case, there are two issues arising which affect the interpretation of these estimates.[2]

First, the effect of social origin is not estimated in an analysis that estimates only the net effect of social origin on social destination after conditioning on educational attainment. Certainly many researchers are aware of this issue, but it can run counter to the intuitive interpretation of research results (Acharya *et al.*, 2016). From this point of view, the highest danger to misinterpret the estimates have studies that use a selected sample (e.g. Holtmann *et al.*, 2017; Jacob *et al.*, 2015). These studies cannot estimate the effect of social origin on social destination because they do not have the necessary information. This limitation is not always reflected in the language researchers employ. For instance, Jacob *et al.* (2015) spoke throughout the paper of "the impact of social origin", even though their study did estimate only the direct effect of social origin conditional on completing a degree at an institution of tertiary education on children's occupation but not the total effect of social origin on children's occupation (and this is also only the case if we believe that there were no unobserved, confounding variables).

It is also difficult to estimate the direct effect in counterfactual terms. Any change in social origin is likely to affect the process of educational attainment and therefore selection into the sample used in these studies. Without knowing the total effect, it is impossible to determine the consequences of changing social origin.

Second, a further problematic issue in Figure 4 is that educational attainment is a collider variable as an arrow enters into educational attainment from both social origin and the node "Unobserved$_2$" (Breen, 2018; Elwert and Winship, 2014). "Unobserved$_2$" refers to unobserved

variables that affect both educational attainment and social destination, e.g. respondents' occupational outcomes. There are many possible unobserved variables of this kind, for instance, respondent's motivation, effort, and noncognitive skills. There may be an overlap between the variables included in "Unobserved$_1$" and "Unobserved$_2$" so that one may argue that researchers should in any case control for unobserved variables. However, there can be some unobserved variables that connect educational attainment and labour market outcomes but that are not influencing social origin. In any case, it is important to theoretically separate these two sources of bias.

If a researcher cannot control for the variables included in "Unobserved$_2$", the resulting estimates of the direct and indirect effects will suffer from collider bias. What makes the situation even more complicated is that social origin can affect the variables included in "Unobserved$_2$". In that case, controlling for the variables included in "Unobserved$_2$" will also lead to bias, to overcontrol bias in estimating the direct effect of social origin on social destination. Acharya *et al.* (2016) called this special type of collider bias intermediate variable bias.

The problems of collider and intermediate variable bias are usually not dealt with or even discussed in research analysing the OED triangle. These biases provide, however, a serious threat to the identification of the direct effect of social origin on social destination, i.e. the effect of social origin on social destination conditional on educational attainment.

Research Practice 4: Controlling for Children's Academic Performance when Estimating the Relationship between Social Origin and Children's Educational Attainment

A very similar triangle than the OED triangle can be drawn with respect to educational attainment as an outcome variable. This triangle reflects the, in particular among European sociologists, popular distinction between primary and secondary effects (Boudon, 1973;

Erikson *et al.*, 2005; Jackson, 2013; Jackson *et al.*, 2007). A related approach, leading to the same issues of overcontrol and collider bias, is estimating the association between social origin and social destination conditional on cognitive skills in the children's generation (Bukodi *et al.*, 2017; Bukodi *et al.*, 2014; Erikson, 2016).

The idea underlying the distinction between primary and secondary effects is that social origin influences offspring's educational attainment via two channels. First, social origin influences children's educational performance measured through test scores or school grades (so-called primary effects). Second, social origin affects children's educational choices (so-called secondary effects). Crucially, the effects of social origin on educational choices are identified in this research practice by conditioning on academic performance.

Therefore, as in the other examples above, this research practice requires the researcher to condition on a variable that is endogenous to the process of interest, in this case a measure of academic performance (Morgan, 2012). Figure 5 presents the relationships of interest graphically. This figure is very similar to the OED triangle and, unsurprisingly, the issues of overcontrol and collider bias are in both situations the same.

[FIGURE 5 ABOUT HERE]

There are two important insights coming from applying the counterfactual approach to causality to this research practice. First, it is important to point out that any analysis that conditions on educational performance, i.e. that estimates the relationship shown in Figure 5, does not estimate the effect of social origin on children's educational attainment. Of course, most researchers using this research practice are aware of this difference and argue that this is the aim of their analysis (Boudon, 1973; Erikson *et al.*, 2005; Jackson, 2013; Jackson *et al.*, 2007). Nevertheless, this research paradigm has affected research practices and, as in the case

of the direct effect of social origin, studies that only report the direct effect estimates, i.e. the so-called secondary effects, are difficult to interpret.

To give a concrete example, Dollmann (2016) estimated how a change in an educational policy in Cologne (Germany) affected educational choices. The policy changed whether parents could overrule the teacher's recommendation for which track a child should attend in secondary school. The effect of the reform was only estimated conditional on educational performance. This practice is in line with the primary and secondary effects framework but ignores that parents can respond to policy changes by affecting children's educational performance. Therefore, the study did not provide an unbiased estimate of the causal effect of the policy reform on the association between social origin and track attendance in the German education system.

There are other examples in published research. For instance, both Bukodi *et al.* (2017) and Hällsten and Thaning (2018) reported only estimates of the association between social origin and social destination controlling for academic performance. Therefore, these studies did not provide any estimates of the total effects of social origin on children's education. The language used in these study did, however, not reflect that they only provided estimates of the direct effect of social origin remaining after conditioning on children's educational performance.

Second, there is again a problem of potential collider bias arising from the fact that once we condition on educational performance, the analysis becomes sensitive to any unobserved variables that affect both educational performance and educational attainment. There are many likely unobserved variables of this kind, e.g. motivation, effort, and noncognitive skills. Variables that can lead to collider bias are indicated by the node "Unobserved$_2$" in Figure 5.

The omission of these variables provides a likely source of bias in the estimation of the direct effect of social origin on educational attainment net of educational performance. At the

same time, some of the variables included in "Unobserved$_2$" may be affected by social origin and controlling for them would lead to overcontrol bias, i.e. we have the situation which Acharya *et al*. (2016) called intermediate variable bias. For this reason, other, more direct ways to identify the effect of social origin on educational decision-making are needed to provide support to the model of primary and secondary effects (Morgan, 2012).

DATA AND METHODS

Data

In the empirical part of this manuscript, I demonstrate the consequences of the overcontrol and collider biases discussed above for our conclusions about intergenerational mobility. In order to do so, I employ data from the German Socio-Economic Panel Study (SOEP; Goebel *et al*. 2018). These data are widely used to study intergenerational mobility in Germany (e.g. Grätz and Pollak, 2016; Hertel, 2017; Müller and Pollak, 2004). The sample I use includes male and female respondents born between 1970 and 1998. They are the (adult) children's generation to whom parental characteristics are added.

Variables

I describe shortly the variables used in the empirical analysis. I use simple continuous or dummy variables of the concepts of interest. Measures that are more complex could be constructed but as the aim of the empirical analysis is to illustrate the consequences of overcontrol and collider biases, the variables I employ are sufficient and allow me a straightforward discussion of the important issues. Whilst I acknowledge that issues of measurement are important (and much attention in the intergenerational mobility literature is devoted to them), they are not the focus of this manuscript. The issues of overcontrol and

collider bias I discuss arise independently of the way in which social origin and social destination are measured.

Occupational Status. Occupation in the children's generation is the central outcome variable in research on occupational mobility. I employ a continuous variable of occupational status based on the International Socio-Economic Index of Occupational Status (ISEI; Ganzeboom *et al.*, 1992).

Educational Attainment. I measure educational attainment in the children's generation through years of education. This variable is the outcome variable in studies on educational mobility. In addition, many studies on occupational mobility use this variable as a control variable estimating the association between social origin and social destination conditional on educational attainment in the children's generation.

Social Origin. I employ four measures of social origin. These variables are based on parental characteristics of the respondents (children of these parents). I measure father's and mother's education and occupation. Father's and mother's education refer to whether the father or mother of the respondent holds an *Abitur* degree (a high school certificate and requirement to study at university). Father's and mother's occupation refer to the occupational status (measured again via ISEI) of the respective parent. These four measures of social origin allow me to discuss all issues raised in the theoretical considerations. Further extensions (e.g. to include measures of parental income and wealth) could easily be applied with similar implications.

Cognitive Skills. To discuss the consequences of including measures of academic performance in models predicting educational attainment, I employ a measure of cognitive skills that is based on a test that was conducted as part of the survey when the respondents were 16 to 17 years old.

Gender. In all models, I control for respondent's gender by including a dummy variable for male (adult) children.

Descriptive statistics on all variables used in the analysis are reported in Table 1.

[TABLE 1 ABOUT HERE]

Analytical Strategy

I rely on OLS regression analysis, the standard tool to estimate intergenerational mobility. Sometimes, researchers use other techniques to estimate intergenerational mobility, such as log-linear modelling. However, all the issues I discuss in this manuscript do occur in these models as well. Focusing on the linear case allows me to keep the technical aspects simple and to focus on the conceptual issues that I emphasize in this manuscript.

RESULTS

In the following, I report a set of regression models that illustrate the consequences of employing the four research practices discussed in this manuscript. The aim of this exercise is to demonstrate that the overcontrol and collider biases introduced by these practices are not just hypothetical scenarios but that these biases are substantive in size and affect our conclusions about intergenerational mobility.

As in the theoretical considerations, I start by discussing the issue of including several indicators of family background in the same model. In Table 2, I present models predicting both respondent's occupational status and their educational attainment, measured through years of education. The different models show what happens if different indicators of social origin are included in separate models (Models 1 and 2 as well as Models 4 and 5) as well as when these measures are entered in the same models (Models 3 and 6).

[TABLE 2 ABOUT HERE]

I focus in the interpretation of the findings on how the estimates of the associations between indicators of social origin and adult children's outcomes change when other indicators of social origin are added to the model. For instance, Model 2 shows that having a father with a high level of parental education (an *Abitur* degree) leads to a 0.63 standard deviations higher occupational status in the children's generation. The association is, however, only about one third of this size (0.21) once we control for father's occupational status (Model 3). This comparison shows that the overcontrol bias introduced by controlling for two measures of social origin in the same model drastically affects the amount of mobility observed in an analysis. As in this example, overcontrol bias due to controlling for several indicators of social origin leads usually to an underestimation of the association between social origin and children's occupational status.

Using educational attainment as an outcome variable (Models 4 to 6) instead of occupational status leads to very similar results. Overcontrol bias introduced by a control for father's occupational status (Model 6) leads us to underestimate the association between parental and child education (Model 5).

It is important to keep in mind that many studies only report the model that includes several indicators of social origin, i.e. Models 3 and 6 in my example. For that reason, these studies do not allow the reader to identify the size of the overcontrol bias.

The second research practice I discuss in this manuscript is the inclusion of measures of maternal and paternal characteristics in the same model. In Table 3, I show models that predict occupational status and years of education. The models compare results if father's and mother's

characteristics are entered into separate (Models 1 and 2 as well as Models 4 and 5) and into the same models (Models 3 and 6).

[TABLE 3 ABOUT HERE]

Again, I interpret the results from the perspective of the counterfactual approach to causality. Of course, there are many unobserved variables that are likely to confound the relationship between the measures of social origin and the measures of social destination so that it is impossible to argue that these estimates represent causal effects. However, my argument is that Model 2 presents a better estimate of the association between maternal and child occupational status (0.29) than the smaller estimate obtained after introducing overcontrol bias through conditioning on paternal occupation in Model 3 (0.18). Similarly, the association between paternal occupational status and child education is 1.11 (Model 4) but it is underestimated in Model 6 (0.88), which conditions on maternal occupational status. To conclude, as with respect to the first research practice, the practice to include measures of maternal and paternal characteristics in the same model leads to downwardly biased estimates of intergenerational persistence. These biases are substantively large in size and affect our conclusions about intergenerational mobility.

The third research practice, which I illustrate here, is the estimation of the effects of family background on occupational outcomes conditional on education. In Table 4, the so-called origin–education–destination (OED) triangle is estimated. Model 1 estimates the association between father's and adult children's occupational status. Model 2 estimates this relationship conditional on respondent's educational attainment.

[TABLE 4 ABOUT HERE]

The association between father's and child occupational status is 0.35 (Model 1). It is reduced once we condition on educational attainment to 0.11 (Model 2). It is very difficult to interpret this conditional estimate. In any case, Model 2 alone does not allow us to say anything about the effect of parental on child occupation. Studies that only present the estimate of Model 2 should therefore be interpreted with caution.

As discussed in the theoretical considerations, it is also difficult to argue that 0.11 is the "direct effect" (Bernardi and Ballarino, 2016) of father's on child occupation as conditioning on education open ups the possibility of collider bias. There are many possible candidate variables, such as ability and motivation, which may affect both educational attainment and occupational status and therefore confound the estimation of the "direct effect" of social origin on occupational status. Contrary to the overcontrol bias identified in research practices 1 and 2, it is impossible to say in which direction collider bias influences the estimates in this research practice and how large this bias is (Acharya *et al.*, 2016; Breen, 2018).

Finally, the last research practice that I discuss is to condition on academic performance when estimating the association between family background and child education. In Table 5, I present a simple form of estimates of the so-called "primary" and "secondary effects" (Boudon, 1973; Erikson *et al.*, 2005; Jackson, 2013; Jackson *et al.*, 2007). Model 1 shows the (gross) association between father's occupational status and respondent's education. Model 2 shows this relationship after conditioning on children's cognitive skills, a measure of academic performance, providing an estimate of the secondary effects.

[TABLE 5 ABOUT HERE]

As expected, the association between father's occupation and child education is smaller (0.52; Model 2) once we control for cognitive skills than without controlling for this endogenous variable (0.76; Model 1). Again, the question is how this conditional estimate can be interpreted. Certainly it should not be confused with the causal effect of social origin on educational attainment, which is estimated (under the strong assumption of no unobserved, confounding variables) by Model 1 and not by Model 2.

In addition, even this estimate of the direct effect obtained in Model 2 can be confounded by the collider bias introduced through conditioning on children's cognitive skills. This bias can invalidate any claim that the conditional estimate provides a representation of a "direct effect" of social origin on children's educational attainment after conditioning on children's academic performance. In addition, the direction and the size of this bias cannot be foreseen (Acharya *et al.*, 2016; Breen, 2018). For this reason, Morgan (2012) argued that other ways were needed to identify the influence of socioeconomic differences in educational decision-making on children's educational attainment.

DISCUSSION AND CONCLUSION

The emergence of the counterfactual approach to causality has affected research practices in the social sciences. However, there are still some widely used research practices that do not take into account the lessons that we can learn from this new understanding of causality. This manuscript discusses four of these practices, which are taken from the field of research on the intergenerational transmission of advantage.

The arguments presented in this manuscript underline the importance of gross estimates of the associations between social origin and social destination for research on intergenerational mobility (Torche, 2015). These bivariate associations are the main contribution intergenerational mobility research makes to our understanding of contemporary

and historic societies. Researchers have to be aware that adding control variables to their analyses often complicates identification and makes it harder to understand what models are actually estimating. What is more, conditioning on endogenous or collider variables can bias the estimates of intergenerational mobility. Apart from the four research practices that I discuss in this manuscript, there are other research practices than can lead to overcontrol and collider bias. For instance, Torche (2015) discussed the problems in interpreting mechanisms underlying the intergenerational transmission of advantage.

The issues identified in this manuscript have general significance for the way research on intergenerational mobility is conducted. The empirical analyses reported here, whilst being very simple and straight forward, have shown that the biases discussed in this manuscript do indeed affect our conclusions about the intergenerational transmission of advantage in contemporary and historic societies. In many practical applications, the consequences of overcontrol and collider bias are likely to be much larger than in the empirical applications discussed in this manuscript.

For instance, mobility research is often interested in describing variation in the association between social origin and social destination over time. In this context, the overcontrol and collider biases introduced by the four research practices discussed in this manuscript are very consequential. If researchers condition in the same model on different indicators of social origin and/ or on maternal and paternal characteristics (e.g. Bloome and Western, 2011; Duncan *et al.*, 2017; Mare, 1981; Shavit and Blossfeld, 1993), they can find indications of change in intergenerational mobility even if the gross association between every indicator of social origin and the measure of social destination did not change. This can simply happen because the associations between the different indicators of social origin and/ or between paternal and maternal characteristics can change over time. However, that is certainly not what we mean if we talk about changing effects of social origin on social destination over

time or across cohorts. It is therefore recommendable if research estimating changes in intergenerational mobility across cohorts focuses on the gross estimates of the associations between one indicator of social origin and a measure of social destination without conditioning on other measures of social origin (e.g. Breen *et al.*, 2009).

Similar arguments apply to comparisons across countries. If researchers condition on serval indicators of social origin in these comparisons, they can observe differences in the associations between one indicator of social origin and children's outcomes even if the effects of this indicator on children do not vary across countries. This can happen if the associations between different indicators of social origin vary across countries. Again, my main recommendation is to focus on the gross estimates and compare those across countries.

If researchers insist on employing the four research practices discussed in this manuscript, it is a good practice if they report the gross estimates without conditioning in addition to the conditional, net estimates (as, for instance, done by Duncan *et al.*, 2017; Mood, 2017; and Pfeffer, 2018). By these means, readers can at least see the initial, gross estimates for themselves and can interpret the results taking them into account. Some research reports only the conditional estimates making it impossible for the reader to assess the impact of overcontrol and collider bias.

Generally, I hope the arguments advanced and the empirical results presented here convince researchers working on the intergenerational transmission of advantage to take into account and to reflect upon the possibilities of overcontrol and collider bias in their research. There may be good reasons to employ any of the four research practices discussed critically in this manuscript but at least I believe that researchers should be aware of the pitfalls connected to them. Finally, as mentioned above, researchers in epidemiology and in political science have developed more systematic approaches to causal mediation analysis (Acharya *et al.*, 2016; Imai *et al.*, 2011; Robins and Greenland, 1992; Pearl, 2009; VanderWeele, 2015). Researchers who

work on intergenerational mobility may profit from integrating these methods into their toolbox.

Notes

1. A possible way to take into account the characteristics of both parents and to avoid overcontrol bias is to construct a single measure combining information from both parents. For instance, Hout (2018) constructed occupational scores using information on both fathers' and mothers' occupations and weighting these two components.
2. There are also more fundamental critics of the idea to split a causal effect into a direct and an indirect component. For instance, Rubin (2004) argued that "the concepts of direct and indirect causal effects are generally ill-defined and often more deceptive than helpful to clear statistical thinking in real, as op-posed to artificial, problems." (Rubin 2004: 162)

Acknowledgments

The data used in this publication were made available to me by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin. For comments on earlier versions of this manuscript, I would like to thank Richard Breen and Fabian Kratz.

References

Acharya, A., Blackwell, M. and Sen, M. (2016). Explaining causal findings without bias: detecting and assessing direct effects. *American Political Science Review*, **110**, 512–529.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Baron, R. M. and Kenny, D. A. (1986). Moderator-mediator variables distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, **51**, 1173–1182.

Beller, E. (2009). Bringing intergenerational social mobility research into the twenty-first century: why mothers matter. *American Sociological Review*, **74**, 507–528.

Bernardi, F. and Ballarino, G. (2016). *Education, Occupation and Social Origin: A Comparative Analysis of the Transmission of Socio-Economic Inequalities*. Cheltenham, UK: Edward Elgar Publishing.

Blau, P. M. and Duncan, O. D. (1967). *The American Occupational Structure*. New York, NY: The Free Press.

Bloome, D., Dyer, S. and Zhou, X. (2018). Educational inequality, educational expansion, and intergenerational income persistence in the United States. *American Sociological Review*, **83**, 1215–1253.

Bloome, D., and Western, B. (2011). Cohort change and racial differences in educational and income mobility. *Social Forces*, **90**, 375–395.

Blossfeld, H.-P. (2009). Educational assortative marriage in comparative perspective. *Annual Review of Sociology*, **35**, 513–530.

Boudon, R. (1973). *L'Inégalité des Chances*. Paris: Armand Colin.

Breen, R. (Ed.). (2004). *Social Mobility in Europe*. Oxford: Oxford University Press.

Breen, R. (2018). Some methodological problems in the study of multigenerational mobility. *European Sociological Review*, **34**, 603–611.

Breen, R., Luijkx, R., Müller, W. and Pollak, R. (2009). Nonpersistent inequality in educational attainment: evidence from eight European countries. *American Journal of Sociology*, **114**, 1475–1521.

Buis, M. L. (2013). The composition of family background: the influence of the economic and cultural resources of both parents on the offspring's educational attainment in the Netherlands between 1939 and 1991. *European Sociological Review*, **29**, 593–602.

Bukodi, E., Bourne, M. and Betthäuser, B. (2017). Wastage of talent? Social origins, cognitive ability, and educational attainment in Britain. *Advances in Life Course Research*, **34**, 34–42.

Bukodi, E., Erikson, R. and Goldthorpe, J. H. (2014). The effects of social origins and cognitive ability on educational attainment: evidence from Britain and Sweden. *Acta Sociologica*, **57**, 293–310.

Bukodi, E. and Goldthorpe, J. H. (2013). Decomposing social origins: the effects of parents' class, status and education on the educational attainment of their children. *European Sociological Review*, **29**, 1024–1039.

Dollmann, J. (2016). Less choice, less inequality? A natural experiment on social and ethnic differences in educational decision-making. *European Sociological Review*, **32**, 203–215.

Duncan, G., Kalil, A. and Ziol-Guest, K. M. (2017). Increasing inequality in parent incomes and children's schooling. *Demography*, **54**, 1603–1623.

Elwert, F. (2013). Graphical causal models. In Morgan, S. L. (Ed.), *Handbook of Causal Analysis for Social Research*. Dordrecht: Springer.

Elwert, F. and Winship, C. (2014). Endogenous selection bias: the problem of conditioning on a collider variable. *Annual Review of Sociology*, **40**, 31–53.

Erikson, R. (2016). Is it enough to be bright? Parental background, cognitive ability, and educational attainment. *European Societies*, **18**, 117–135.

Erikson, R., Goldthorpe, J. H., Jackson, M., Yaish, M. and Cox, D. R. (2005). On class differentials in educational attainment. *Proceedings of the National Academy of Sciences*, **102**, 9730–9733.

Erola, J., Jalonen, S. and Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility*, **44**, 33–43.

Featherman, D. L. and Hauser, R. M. (1978). *Opportunity and Change*. New York: Academic Press.

Gangl, M. (2010). Causal inference in sociological research. *Annual Review of Sociology*, **36**, 21–47.

Ganzeboom, H., de Graaf, P. and Treiman, D. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, **21**, 1–56.

Goebel, J., Grabka, M. M., Liebig, S. Kroh, M., Richter, D., Schröder, C. and Schupp, J. (2018). The German socio-economic panel study (SOEP). *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)*, DOI: 10.1515/jbnst-2018-0022.

Grätz, M. and Pollak, R. (2016). Legacies of the past: social origin, educational attainment, and labor market outcomes in Germany. In Bernardi, F. and Ballarino, G. (Eds.), *Education, Occupation, and Social Origin: A Comparative Analysis of the Transmission of Socio-Economic Inequalities*. Cheltenham: Edward Elgar.

Gregg, P., Jonsson, J. O., Macmillan, L. and Mood, C. (2017). The role of education for intergenerational income mobility: a comparison of the United States, Great Britain, and Sweden. *Social Forces*, **96**, 121–151.

Hällsten, M. and Pfeffer, F. T. (2017). Grand advantage: family wealth and grandchildren's educational achievement in Sweden. *American Sociological Review*, **82**, 328–360.

Hällsten, M. and Thaning, M. (2018). Multiple dimensions of social background and horizontal educational attainment in Sweden. *Research in Social Stratification and Mobility*, **56**, 40–52.

Hertel, F. R. (2017). *Social Mobility in the 20th Century: Class Mobility and Occupational Change in the United States and Germany*. Wiesbaden: Springer VS.

Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association*, **81**, 945–960.

Holtmann, A. C., Menze, L. and Solga, H. (2017). Persistent disadvantages or new opportunities? The role of agency and structural constraints for low-achieving adolescents' school-to-work transitions. *Journal of Youth and Adolescence*, **46**, 2091–2113.

Hout, M. (1988). More universalism, less structural mobility: the American occupational structure in the 1980s. *American Journal of Sociology*, **93**, 1358–1400.

Hout, M. (2018). Americans' occupational status reflects the status of both of their parents. *Proceedings of the National Academy of Sciences*, **115**, 9527–9532.

Imai, K., Keele, L., Tingley, D. and Yamamoto, T. (2011). Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, **105**, 765–789.

Ishida, H., Müller, W., and Ridge, J. M. (1995). Class origin, class destination, and education: a cross-national study of industrial nations. *American Journal of Sociology*, **101**, 145–193.

Jackson, M. (Ed.). (2013). *Determined to Succeed? Performance versus Choice in Educational Attainment*. Stanford, CA: Stanford University Press.

Jackson, M., Erikson, R., Goldthorpe, J. H. and Yaish, M. (2007). Primary and secondary effects in class differentials in educational attainment: the transition to A-level courses in England and Wales. *Acta Sociologica*, **50**, 211–229.

Jacob, M., Klein, M. and Ianelli, C. (2015). The impact of social origin on graduates' early occupational destinations: an Anglo-German comparison. *European Sociological Review*, **31**, 460–476.

Jæger, M. M. (2007). Educational mobility across three generations: the changing impact of parental social class, economic, cultural and social capital. *European Societies*, **9**, 527–550.

Jæger, M. M. and Holm, A. (2007). Does parents' economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime? *Social Science Research*, **36**, 719–744.

Kalmijn, M. (1994). Mother's occupational status and children's schooling. *American Sociological Review*, **59**, 257–275.

Korupp, S. E., Ganzeboom, H. B. G. and van der Lippe, T. (2002). Do mothers matter? A comparison of models of the influence of mothers' and fathers' educational and occupational status on children's educational attainment. *Quality and Quantity*, **36**, 17–42.

Lazarsfeld, P. F. (1939). Interchangeability of indices in the measurement of economic influences. *Journal of Applied Psychology*, **23**, 33–45.

Mare, R. D. (1981). Change and stability in educational stratification. *American Sociological Review*, **46**, 72–87.

Marks, G. N. (2008). Are father's or mother's socioeconomic characteristics more important influences on student performance? Recent international evidence. *Social Indicators Research*, **85**, 293–309.

Mood, C. (2017). More than money: social class, income, and the intergenerational persistence of advantage. *Sociological Science*, **4**, 263–287.

Morgan, S. L. (2012). Models of college entry in the United States and the challenges of estimating primary and secondary effects. *Sociological Methods and Research*, **41**, 17–56.

Morgan, S. L. and Winship, C. (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research. Second Edition*. Cambridge: Cambridge University Press.

Müller, W. and Pollak, R. (2004). Social mobility in West Germany: the long arms of history discovered? In Breen, R. (Ed.), *Social Mobility in Europe*. Oxford: Oxford University Press.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference. Second Edition.* Cambridge: Cambridge University Press.

Pfeffer, F. T. (2018). Growing wealth gaps in education. *Demography*, **55**, 1033–1068.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, **3**, 143–155.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Education Psychology*, **66**, 688–701.

Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, **31**, 161–170.

Schwartz, C. R. (2013). Trends and variation in assortative mating: causes and consequences. *Annual Review of Sociology*, **39**, 451–470.

Sewell, W. H., Haller, A. O. and Portes, A. (1969). The educational and early occupational attainment process. *American Sociological Review*, **34**, 82–92.

Sewell, W. H. and Hauser, R. M. (1975). *Education, Occupation and Earnings*. New York: Academic Press.

Shavit, Y. and Blossfeld, H.-P. (Eds.). (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview Press.

Torche, F. (2011). Is a college degree the great equalizer? Intergenerational mobility across levels of schooling in the United States. *American Journal of Sociology*, **117**, 763–807.

Torche, F. (2015). Intergenerational mobility and equality of opportunity. *European Journal of Sociology*, **56**, 343–370.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford: Oxford University Press.

Wong, R. S.-K. (1998). Multidimensional influences of family environment in education: the case of socialist Czechoslovakia. *Sociology of Education*, **71**, 1–22.

TABLES

Table 1. Descriptive statistics

Variable	Mean	SD	Min	Max	*N*
Male	0.44	0.50	0	1	4,691
Occupational status	0.00	1.00	–1.85	2.60	4,691
Years of education	13.16	2.72	7	18	4,691
Cognitive skills	0.00	1.00	–3.19	2.34	449
Father's occupational status	0.00	1.00	–1.61	2.73	4,691
Mother's occupational status	0.00	1.00	–1.66	2.98	4,691
Father *Abitur*[1]	0.20	0.40	0	1	4,691
Mother *Abitur*[1]	0.17	0.37	0	1	4,691

Notes: An *Abitur* (the highest German secondary school leaving certificate, comparable to A-levels in the United Kingdom) degree indicates a high level of education.
Source: German Socio-Economic Panel Study (SOEP), v33.1.

Table 2. OLS regression models predicting occupational status and educational attainment

	Occupational status (ISEI)			Educational attainment (years of education)		
	(1)	(2)	(3)	(4)	(5)	(6)
Father's occupational status	0.35**		0.30**	1.11**		0.86**
	[0.32, 0.37]		[0.27, 0.33]	[1.04, 1.18]		[0.77, 0.94]
Male	–0.01	0.01	–0.00	–0.36**	–0.30**	–0.34**
	[–0.06, 0.05]	[–0.04, 0.07]	[–0.06, 0.05]	[–0.50, –0.21]	[–0.44, –0.15]	[–0.48, –0.20]
Father *Abitur*		0.63**	0.21**		2.31**	1.12**
		[0.56, 0.70]	[0.13, 0.29]		[2.13, 2.49]	[0.91, 1.33]
N	4,691	4,691	4,691	4,691	4,691	4,691

Note: 95% confidence intervals in brackets.
Source: SOEP v33.1 (DOI: 10.5684/soep.v33.1).
* $p < 0.05$, ** $p < 0.01$

Table 3. OLS regression models predicting occupational status and educational attainment

	Occupational status (ISEI)			Educational attainment (Years of education)		
	(1)	(2)	(3)	(4)	(5)	(6)
Father's occupational status	0.35**		0.27**	1.11**		0.88**
	[0.32, 0.37]		[0.24, 0.30]	[1.04, 1.18]		[0.80, 0.95]
Male	–0.01	0.00	–0.01	–0.36**	–0.33**	–0.36**
	[–0.06, 0.05]	[–0.05, 0.06]	[–0.06, 0.05]	[–0.50, –0.21]	[–0.48, –0.18]	[–0.50, –0.22]
Mother's occupational status		0.29**	0.18**		0.93**	0.56**
		[0.26, 0.32]	[0.15, 0.20]		[0.85, 1.00]	[0.48, 0.64]
N	4,691	4,691	4,691	4,691	4,691	4,691

Note: 95% confidence intervals in brackets
Source: SOEP v33.1 (DOI: 10.5684/soep.v33.1).
* $p < 0.05$, ** $p < 0.01$

Table 4. OLS regression models predicting occupational status

	Occupational status (ISEI)	
	(1)	(2)
Father's occupational status	0.35**	0.11**
	[0.32, 0.37]	[0.09, 0.14]
Male	–0.01	0.07**
	[–0.06, 0.05]	[0.03, 0.12]
Years of education		0.21**
		[0.20, 0.22]
N	4,691	4,691

Note: 95% confidence intervals in brackets
Source: SOEP v33.1 (DOI: 10.5684/soep.v33.1).
* $p < 0.05$, ** $p < 0.01$

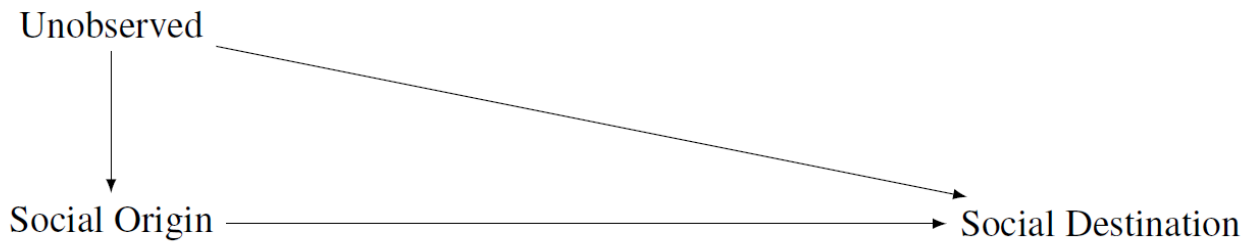
Table 5. OLS regression models predicting educational attainment

	Educational attainment (Years of education)	
	(1)	(2)
Father's occupational status	0.76**	0.52**
	[0.56, 0.96]	[0.33, 0.71]
Male	–0.31	–0.52**
	[–0.73, 0.11]	[–0.91, –0.13]
Cognitive skills		0.89**
		[0.69, 1.09]
N	449	449

Note: 95% confidence intervals in brackets
Source: SOEP v33.1 (DOI: 10.5684/soep.v33.1).
* $p < 0.05$, ** $p < 0.01$

FIGURES

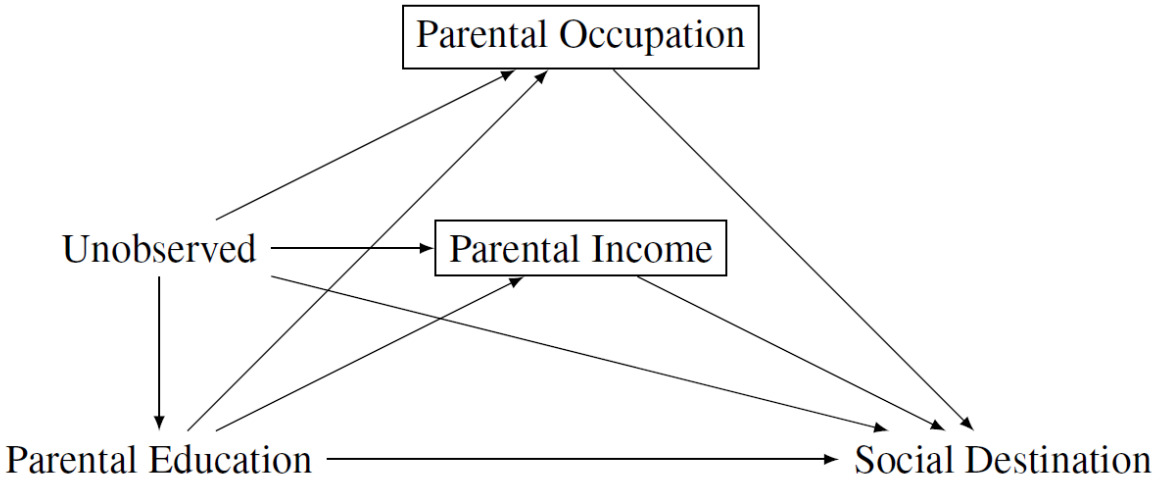
Figure 1. The general model underlying the analysis of intergenerational mobility



Social Origin: Measured via parental education, income, or occupation.

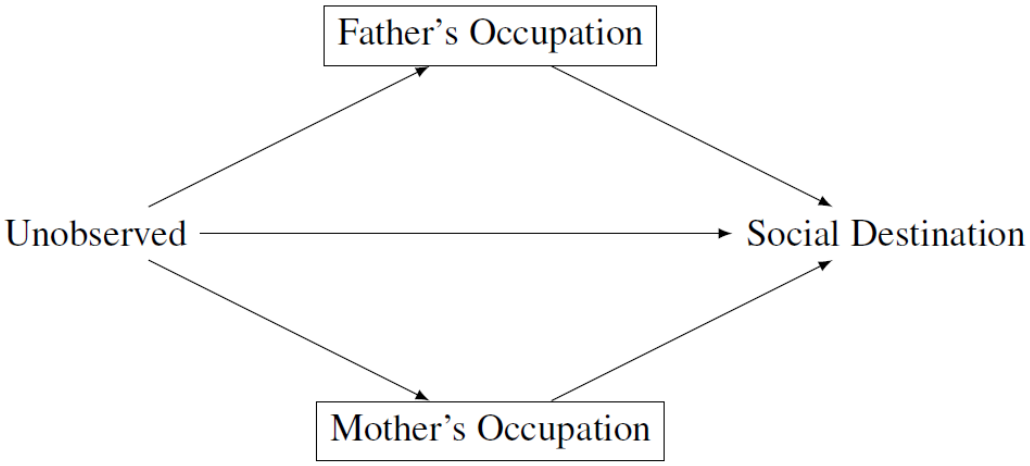
Social Destination: Measured via child education, income, or occupation.

Figure 2. Controlling for multiple dimensions of social origin in the same model



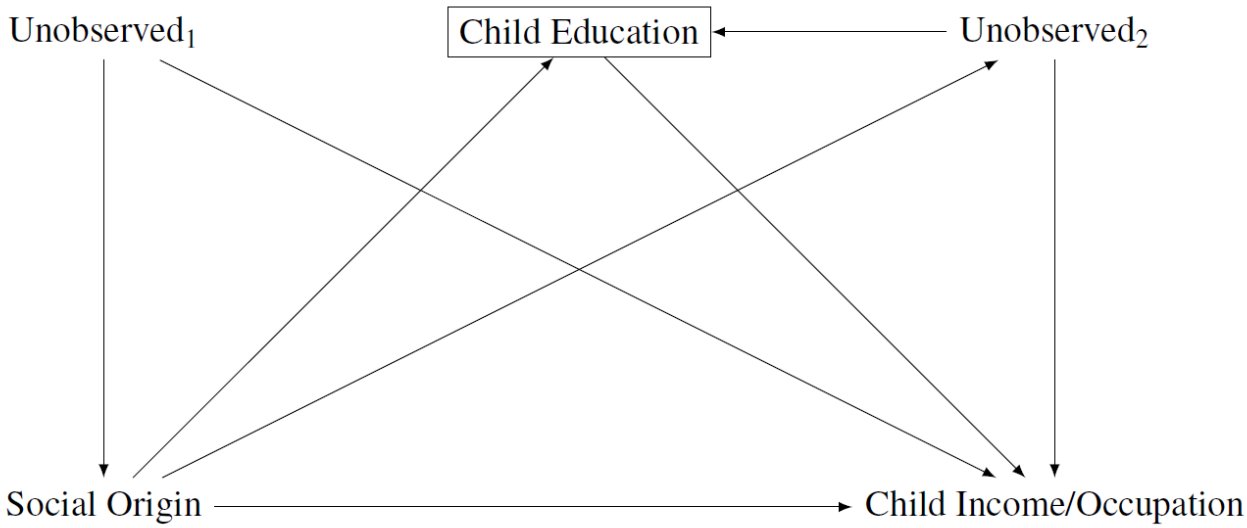
Social Destination: Measured via child education, income, or occupation.

Figure 3. Controlling for father's and mother's characteristics in the same model



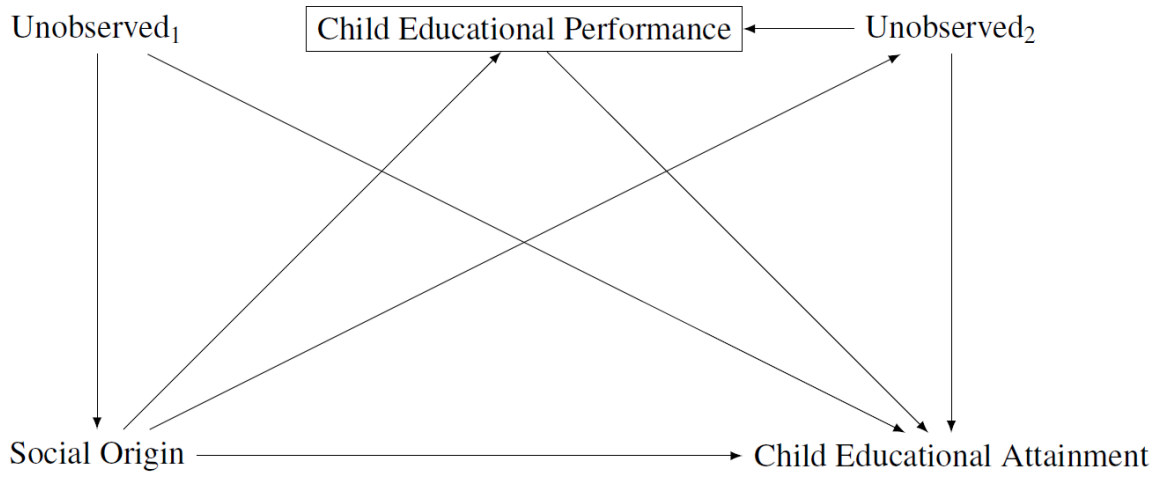
Social Destination: Measured via child education, income, or occupation.

Figure 4. Controlling for children's educational attainment when estimating the relationship between social origin and children's labor market outcomes



Social Origin: Measured via parental education, income, or occupation.

Figure 5. Controlling for children's academic performance when estimating the relationship between social origin and children's educational attainment



Social Origin: Measured via parental education, income, or occupation.