European Sociological Review, 2025, **41**, 808–821
https://doi.org/10.1093/esr/jcae037
Advance access publication 21 October 2024
Original Article

OXFORD

A caution on sibling comparisons in studying effects of the rearing environment

Per Engzell[1,2,3,*] and Martin Hällsten[3,*]

[1]UCL Social Research Institute, University College London, London WC1H 0AA, UK
[2]Nuffield College, University of Oxford, Oxford OX1 1NF, UK
[3]Swedish Institute for Social Research, Stockholm University, 106 91 Stockholm, Sweden
[*]Corresponding author. Email: p.engzell@ucl.ac.uk; martin.hallsten@sofi.su.se

Recent studies use sibling fixed effects to estimate the influence of the family environment on children, a practice we call the 'discordant family design'. These studies suffer from a disconnect between the use of within-family variation, on the one hand, and relevant theories which mostly refer to variation between families on the other. In addition, reverse causality, within-family confounding, selection into identification, and measurement error complicate their interpretation further. We discuss three applied examples—the effects of parenting, family income, and neighbourhood context—and provide some general guidance. To avoid misinterpretation, researchers should have a strong grasp of the variance that enters into estimation, and not just the potential confounders a given strategy is designed to deal with.

Introduction

The credibility revolution has shifted research focus from description to causal inference, but this presents a challenge for sociology because of its focus on static personal characteristics like social background, gender, race, and immigration status. Fixed-effects models are one of the few tools for causal inference that are easy to implement with the observational data that sociologists often use. One common application in this context is the sibling comparison design, also known as the family fixed-effects or within-family design. This design discards all variation between families, and is often described as adjusting for all confounders that siblings share. In its original use, the focus is on effects of siblings' characteristics on outcomes, for example, to study economic returns to education, controlling for familial confounding (Gorseline, 1932; Hauser and Sewell, 1986). More recently, scholars have begun using this model for analysing the effects of the environment one grows up in, that is, of parents and neighbourhoods (Duncan et al., 1998; Blau, 1999; Ermisch and Francesconi, 2001; Ermisch et al., 2004; Tamm, 2008; D'Onofrio et al., 2009; Jæger, 2011; Chia, 2013; Sariaslan et al., 2013, 2014, 2021; Elstad and Bakken, 2015; Lehti et al., 2019; Breinholt and Holm, 2020; Achard, 2022; Grätz et al., 2022; Markussen and Røed, 2022; Jensen et al., 2023).[1]

We examine this new strand of sibling comparison designs where the treatment of interest refers to characteristics of the rearing environment. We refer to this class of models as the 'discordant family design' (DFD). Figure 1 illustrates the idea. We have depicted a family with two siblings, where E denotes an exposure and Y denotes an outcome. In a typical sibling comparison, exposures and outcomes both occur at the level of siblings, independent of parents (Figure 1, left). In DFD, the source of exposure *is* the parents (Figure 1, right). This design, then, relies on parents, or families, treating siblings differently, as non-discordant family units do not contribute to the effect estimate. Alternatively, with treatments such as family income that do not differ for siblings at any given time, it relies on imperfect overlap in sibling life courses. We consider studies of neighbourhood effects as an example of the same design, as they present many of the same problems.

Studies using this design often report null findings, which are then used to reject the notion that the rearing environment has any influence on child outcomes. Some examples are that parenting styles are claimed not to influence the development of children's cognitive

Received: March 2023; **revised:** May 2024; **accepted:** September 2024

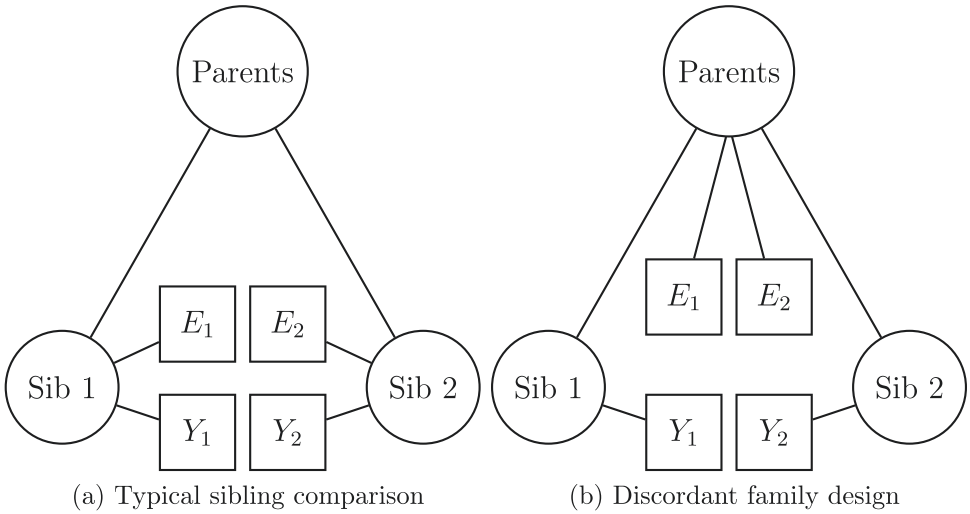

Figure 1 Two types of family fixed-effects models, E_1,E_2: exposure, Y_1,Y_2: outcome. Note: Examples of the typical sibling comparison design (left) include effects of schooling on earnings, childbearing on earnings, or schooling on health and mortality. Examples of DFD (right) include the effects of parenting activities, parental income, or neighbourhood deprivation on early skill accumulation or later life outcomes

skills (Grätz et al., 2022), that family income does not influence mental well-being (Sariaslan et al., 2021), or that neighbourhoods do not influence the risk of entering criminal careers (Sariaslan et al., 2013). Should we trust these conclusions?

One crucial assumption of DFD is that treatment effects are equal between and within families. We argue that this is unlikely to hold. Substantive theories in the area generally refer to variation between families, not within them. Between-family variation captures stable differences in the lived environment, while within-family variation reflects idiosyncratic fluctuation. For example, if the treatment is parenting styles, stable differences refer to durable dispositions of the kind that sociologists describe as class or habitus (Bourdieu, 1990; Lareau, 2011). If the treatment is income, stable differences refer to permanent income, which economists assume is what guides consumption and investment (Friedman, 1957; Haider and Solon, 2006).

Relevant theory, then, usually implies that a shared environment is the relevant treatment. By contrast, the case that within-family variation should have an influence is weaker. At the same time, there are a host of other reasons that could lead parenting and child outcomes to correlate within families. There is ample evidence of bidirectional parent–child effects, where parental inputs arise in response to the child's interests and abilities. Parents can either attempt to encourage a child's talents or to compensate for individual difficulties, giving rise to endogeneity of an unknown sign (Grätz and Torche, 2016; Dierker and Diewald, 2024; García-Sierra, 2024). Together with additional challenges such as within-family confounding, selection into identification, and measurement error, the resulting bias is difficult to assess.

Of course, if authors are genuinely interested in temporary fluctuations in the family environment, the thrust of our critique does not apply. Then researchers would have to contend with the usual challenges to identification such as reverse causality, confounding, selection, and measurement error, but they would not be making the category error that we point out. Some contributions offer thoughtful discussions of these issues and some even attempt—rather ambitiously—to estimate causal effects of within- and between-family variation separately (Tominey, 2010; Chevalier et al., 2013). Most of the papers we have reviewed, however, do not make this distinction. Our note is therefore an invitation for researchers to think more carefully about the distinction between- and within-family variation and which one it is that the relevant theory entails.

The discordant family design

Figure 2 presents a causal diagram (Pearl, 1995) of the typical sibling comparison design. U_j denotes unobserved confounders shared among siblings in family j, E_{ij} denotes an exposure that is specific to each sibling, and Y_{ij} denotes the outcome of interest. If we were to observe U_j, we could estimate:

$$Y_{ij} = \beta_0 + \beta_1 E_{ij} + \beta_2 U_j + \varepsilon_{ij}, \tag{1}$$

and β_1 would give us an unbiased estimate of the effect of E_{ij} on Y_{ij}. When U_j is not observed, the sibling comparison design offers an elegant solution. It estimates β_1 by transforming individual values to their deviation from a within-family mean:

$$(Y_{ij} - \bar{Y}_j) = \beta_1(E_{ij} - \bar{E}_j) + (\varepsilon_{ij} - \bar{\varepsilon}_j). \qquad (2)$$

Because U_j does not vary within families, it drops out of the equation and β_1 can be estimated without bias. Examples of this design include studies of the effect of schooling on earnings, childbearing on earnings, or schooling on health and mortality.

What about DFD? Technically, it is just an instance of the generic fixed-effects design in Equation 2. The family mean term $\bar{E}_j$ captures durable features of the rearing environment, while the within-deviation represents idiosyncratic variation. We illustrate this idea in Figure 3, where we label parenting inputs E, and distinguish between a shared component E_j and one E_{ij} that is unique to each sibling. As before, U_j denotes unobserved confounders shared among siblings and Y_{ij} is the outcome of interest.

If the assumptions encoded in Figure 3 are correct, fixed-effects estimation works as intended, and DFD is valid. The problem is that Figure 3 is a poor representation of most theories of parental influence. Specifically, it assumes that the causal effect of parenting runs from E_{ij} to Y_{ij}, and not from E_j to Y_{ij}, or some combination thereof. That is a substantive assumption that needs to be defended in light of theory and subject matter knowledge. In our view, it is more reasonable to assume that shared and unique parenting have distinct effects:

$$Y_{ij} = \gamma_0 + \gamma_1 E_j + \gamma_2 E_{ij} + \gamma_3 U_j + \epsilon_{ij}. \qquad (3)$$

Moreover, we believe that in many contexts, the influence of shared parenting is likely to be larger than that of idiosyncratic variation in parenting between siblings, so that $\gamma_1 > \gamma_2$. In the extreme case where γ_2 is zero, we end up with the causal diagram in Figure 4. Family membership j perfectly determines both U_j and E_j, which drop out of the equation. Attempting to estimate the influence of parenting E_j on child outcomes with DFD would then yield a null effect. It is, of course, unlikely that the marginal effect of E_{ij} on Y_{ij} is precisely zero, but only that its effect differs from that of E_j is enough to bias our estimate of the effect of family inputs.

Further, as we illustrate in Figure 5, it is plausible that E_{ij} is associated with Y_{ij} through other processes than direct causation. Candidates include individual confounding, reverse causation, and selection into identification. Individual confounding, indicated by the path through U_{ij}, would occur if variation in parenting and child outcomes within families share a cause. One example is birth order, which shapes the expectations and thereby behaviour of both parents and children. Reverse causation, indicated by the path from Y_{ij} to E_{ij}, occurs whenever parenting responds to a child's outcome, as is likely for things like scholastic achievement or behaviour. Finally, selection into identification (Miller et al., 2023), indicated by the node D_j, emerges because sibships need to be discordant to enter into estimation. Discordance is a function of both E_{ij} and Y_{ij}, and conditioning on it opens a collider path through D_j.[2] All these processes might alter the correlation between E_{ij} and Y_{ij} conditional on U_j, independent of any causal effect.

The problems illustrated by the dashed lines in Figure 5 have been described in earlier literature on limitations of family fixed-effects models (Bound and Solon, 1999; Sjölander et al., 2021). Another insight

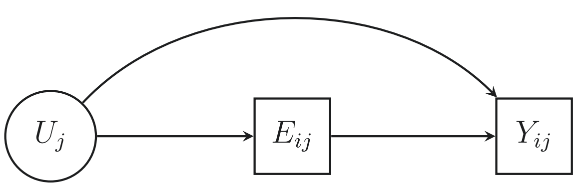

Figure 2 Typical sibling comparison design. U_j: shared confounders, E_{ij}: individual exposure, Y_{ij}: individual outcome

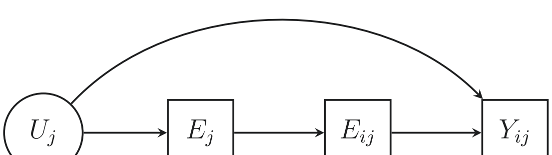

Figure 3 DFD. U_j: shared confounders, E_j: shared environment, E_{ij}: discordant environment, Y_{ij}: individual outcome

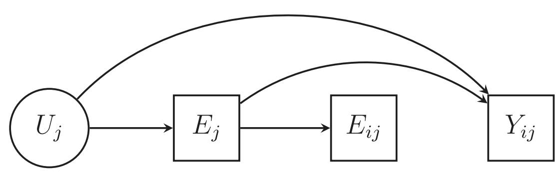

Figure 4 DFD when shared rearing is the relevant treatment. By conditioning on family membership j, the effect of E_j on Y_{ij} is lost

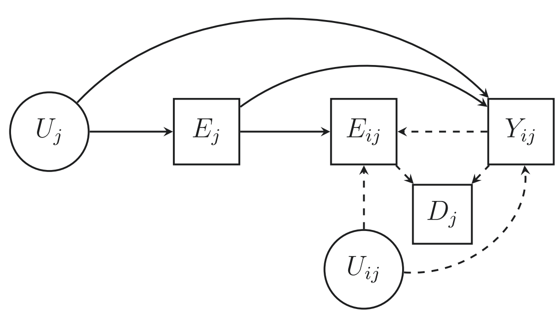

Figure 5 DFD with individual confounding ($E_{ij} \leftarrow U_{ij} \rightarrow Y_{ij}$), reverse causality ($Y_{ij} \rightarrow E_{ij}$), and selection into identification ($E_{ij} \rightarrow D_j \leftarrow Y_{ij}$)

from this literature is that measurement error is generally heightened in fixed-effects models, which risks biasing effects toward zero (Griliches, 1979). While all of these issues are important, in this note, we focus on the discrepancy between the theoretically relevant variation and that which DFD allows us to analyse.

Three applied examples

Is shared or discordant environment the causally relevant treatment? We now consider three applied cases where DFD has been used. We discuss what theory has to say about this question and then we dig deeper into specific methodological implications. The first case is the effects of parenting practices on child outcomes. We then consider the effects of family income and neighbourhood context.

Parenting practices

One application of DFD has been to investigate the impact of parenting on outcomes such as academic achievement or behaviour. For example, Grätz et al. (2022) estimate the effects of parenting styles and practices on early adolescents' noncognitive skills using data from the German Twin Family Panel (TwinLife) with twin fixed-effects models combined with longitudinal information. Parenting here is a set of self-reported items about how the parents exercise emotional support and control, and how often they engage their child in activities such as singing, reading, museum visits, or club activities. They find a null effect and conclude that there is 'no support to the notion that parenting styles, parental activities, and extracurricular activities in early adolescence affect the development of children's noncognitive skills' (Grätz et al., 2022, p. 398).

Is it reasonable to expect within-family differences in parenting to have an effect on behaviour? That depends on the theory. Although the theoretical basis of work in this vein is often vague, one common reference is that of Lareau (2011), whose influential work introduced the concept of 'concerted cultivation' as a distinct form of middle-class parenting. In her words:

> [T]he differences among families seem to cluster together in meaningful patterns.... [M]iddle-class parents tend to adopt a cultural logic of child-rearing that stresses the concerted cultivation of children. Working-class and poor parents, by contrast, tend to undertake the accomplishment of natural growth.... [T]he biggest gaps were not within social classes but, as I show, between them. It is these class differences and how they are enacted in family life and child rearing that shape the way children view themselves in relation to the rest of the world. (Lareau, 2011, pp. 3–4)

It seems clear that this work describes differences between families and not within them. Another frequently invoked term is that of 'parenting styles', first introduced by Baumrind (1966). In her original formulation, Baumrind describes these styles as *types of parent*, not as behaviours, noting how the permissive, authoritarian, and authoritative parent will tend to treat their child. For example, '[t]he authoritarian parent attempts to shape, control, and evaluate the behavior and attitudes of the child in accordance with a set standard of conduct, … [while t]he authoritative parent attempts to direct the child's activities in a rational, issue-oriented manner' (Baumrind, 1966, pp. 890–891). Baumrind (1966, p. 905) concludes that these differences need to be 'tested empirically with a variety of subgroups'.

We could cite other examples, but the point should be evident: theories of parenting usually describe stable differences between families, and it is not self-evident how transferable they are to variation within families. This does not mean that there is no variation within groups, within families, or within individuals over time. However, it does suggest that the causes and consequences of such variation may differ from those described by theories about the effects of parenting on child outcomes.

When interpreting empirical results from this design, one must ask: where do within-family differences in parenting come from? Consider some of the treatments that this literature has looked at: visits to museums and libraries, conversations about politics or schoolwork, encouragement of reading, music, hobbies, and so on (Jæger and Møllegaard, 2017; Grätz et al., 2022). We have to imagine a very unusual family for these differences to be both randomly distributed and sustained enough to matter. The thought experiment that DFD asks us to perform is essentially this: a parent tosses a coin and decides to exclude one child but not the other from these enriching activities. Few parents would hazard to try.[3]

A more likely scenario is that the stimuli that siblings receive are tailored to their personal makeup. Children are co-creators of their environment with a substantial power to shape how their parents act and react towards them (Scarr and McCartney, 1983; Avinun and Knafo, 2014; Breinholt and Conley, 2023). For example, a child who displays interest in music may trigger parents' investment in music education, going to concerts, and so on. The problem that treatment depends on earlier realizations of the outcome, or on unobserved third variables such as ability, has been thoroughly discussed in relation to twin-based estimates of the returns to schooling (Bound and Solon, 1999) and is no less relevant in the case of parenting.

If one can sign the resulting bias, may DFD nevertheless help bound the true effect? This seems unlikely, as parents' behaviour can relate positively or negatively to child endowments. Let us exemplify. With returns to schooling, it is common to assume that unobserved ability has a positive effect on both schooling and wages, which would lead to upward inconsistency. By contrast, parental behaviour may either reinforce or compensate initial differences (Grätz and Torche, 2016). For example, a child who enjoys reading could be more likely to receive books and educational games (Jæger, 2011; Engzell, 2021). But importantly, such gifts or activities could also be motivated by parents' desire to compensate for a lack of reading interest. Without further data, these two scenarios are difficult to separate.

What about monozygotic twins? Some studies use this population as a way to safeguard against endogeneity (Jæger and Møllegaard, 2017; Grätz et al., 2022). On the face of it, this may seem like a powerful strategy. The logic is that with genetically identical twins, differences in personal constitution are less likely to explain parental behaviour. After all, the genetic lottery is one of the major sources of difference between regular siblings (Visscher et al., 2006; Conley et al., 2015).[4] However, by removing birth spacing, twin studies make it less likely that discordant environment reflects fluctuation in parents' circumstances, and more likely that it is guided by some conscious choice—raising further doubts about exogeneity.

If there are genuine differences in parenting between siblings, discordant families that treat their children differently are likely to differ from those that treat their children more similarly. With DFD, only the former will contribute to the estimate, and this is an example of selection into identification (Miller et al., 2023). Feinberg and Hetherington (2001) report that differential parenting is most salient when parenting is cold and harsh. Moreover, given evidence that parent–child relations are bidirectional (Pardini, 2008), cold and harsh parenting can potentially involve, or be a response to, parent–child conflict. Studying twins will arguably heighten this problem, as the lack of birth spacing means that many activities would be shared by default.[5]

Another possibility that must be taken seriously is that much within-family variation is simply misreporting and other errors of observation. The consequence of random errors is to bias estimated effects toward the null (Bohrnstedt and Carter, 1971). This problem is amplified in fixed-effects models because de-meaning of variables reduces much of the signal without reducing the noise (Griliches, 1979). Data reported by Grätz et al. (2022, pp. 405–406) suggest that measurement error is a concern. They find that within-family differences in parenting are larger with child reports than with parent reports, in line with the knowledge that children tend to be less reliable as respondents (Looker, 1989). They also find that parents' and children's reports are poorly correlated (Grätz et al., 2022, pp. 415–416). At the same time, the assumption of random measurement errors is itself a strong one, and one could envision various scenarios that lead to other expectations.

In sum, discordant environment is likely to reflect parental behaviour induced by the child, selection into identification, and measurement error. While standard selection bias and random errors will attenuate associations (Engzell and Jonsson, 2015), other processes may produce inconsistency in the opposite direction and it is difficult to sign the resulting bias. Particularly hard to assess is the nature of responsive parenting: depending on whether parents try to reinforce or compensate differences, the inconsistency may be upward or downward.

Family income

Another application of DFD has been the effect of family income on children's outcomes. An early example was Blau (1999) who analysed the effect of parental income on children's cognitive, social, and emotional development. In addition to the discordant family model, he also used more traditional models comparing children of sibling mothers (grandparent or cousin fixed effects). A more recent example is Sariaslan et al. (2014) who analysed the effect of family income on adolescent violent criminality and substance misuse in Sweden. Family income was measured as an average at the child's ages 1 through 15. While they found an income gradient in a simple regression, with DFD, the gradient was null. In a similar study, Sariaslan et al. (2021) analysed the effect of childhood family income and subsequent psychiatric disorders, substance misuse, and violent crime arrests in Finland. Family income was measured in a single year, at age 15. In a DFD model, there was no effect of family income on the outcome.

Should we expect year-on-year fluctuations in income to affect children's well-being years later? The dominant theory of how income shapes investments in children is that of Becker and Tomes (1979, 1986). This theory stipulates that parents' economic behaviour depends not on concurrent income as much as expectations about permanent income, that is, the stable economic situation of a family over the foreseeable future. As Torche (2015, p. 45) explains:

> The analysis of intergenerational mobility has its conceptual basis in the notion of 'permanent income' (Friedman, 1957), which states that it is the permanent expectation of income that determines consumption and ultimate economic welfare. So

> the relationship of interest is between parents' and children's permanent standing.... From a permanent income perspective, transitory fluctuation and error from one year to the next is a form of measurement error.

The theory is clear: behaviour is governed by permanent income. This guides economic decisions and consumption, not short-term or transitory income.[6] Annual income is a poor indicator of permanent income because of income fluctuations, and researchers therefore typically attempt to measure permanent income (Solon, 1989; Brady et al., 2018). Blau (1999, p. 266) noted that 'permanent income does not vary across the observations on a given mother, so the mother and child fixed-effects methods cannot identify the effects of permanent income', but this insight appears to have been lost.

When analysing family income empirically, a number of problems arise. Since family income is a more distal treatment, the problem of reverse causation that is evident with parenting practices should be less acute. However, given that siblings live in the same household, the only source of variation is over time. Hence, there must be a time gap between siblings to have some variation in parents' income. This time gap has important consequences that are not always taken into account.

Consider a model for income with permanent and transitory components: $I_j = P_j + T_{jt}$. The sibling difference of measurement at different time points (e.g. 1 and 2) is then $\Delta_{S1,S2} = (P_j + T_{j2}) - (P_j + T_{j1}) = T_{j2} - T_{j1}$. Since the permanent component is by definition invariant, it will be netted out as a family-constant factor when comparing siblings or twins. This means that the remaining variation used to identify the effects of family income is the transitory component that we often regard as an idiosyncratic measurement error. By design, the family fixed-effects model with a time gap between siblings (measurement of family income) will therefore produce a null effect even if P_j has a causal relation to the outcome.

The transitory component of earnings may still have a theoretical implication, in gauging how income elastic the family is. A theoretically informed hypothesis could be that (some) families are severely credit-constrained and would spend every extra penny on improving their children's outcomes (e.g. education, and mental well-being). T_{jt} is a relevant measure to test this hypothesis, and a null effect from a fixed-effects model would thus reject that hypothesis.

Some scholars explicitly attempt to measure permanent income P_j to minimize the dependence on T_{jt}. Permanent income is usually operationalized by taking averages over time. This will increase the relative weight of P_j to T_{jt}, that is, the signal-to-noise ratio. However, this does not solve the problem. Consider a design using the sum of parents' income measured when each child was aged 3–10, where the mother's age at first birth is 20 and siblings born 3 years apart:

$$S_1 = \sum I_{j23}, I_{j24}, I_{j25}, I_{j26}, I_{j27}, I_{j28}, I_{j29}, I_{j30}, \quad (4)$$

$$S_2 = \sum I_{j26}, I_{j27}, I_{j28}, I_{j29}, I_{j30}, I_{j31}, I_{j32}, I_{j33}, \quad (5)$$

$$\Delta_{S1,S2} = \sum I_{j31}, I_{j32}, I_{j33} - \sum I_{j23}, I_{j24}, I_{j25}. \quad (6)$$

It is evident that most of the years will overlap, and most of the contributions to permanent income will be netted out. The within-family component of earnings instead contains the difference between parents' income at age 23–25 and income at age 31–33. This is a very specific income difference capturing growth in the parents' careers. It also reflects early life conditions for the elder sibling and school-age conditions for the younger sibling. How this difference relates to permanent income is unclear.

As the relationship of this income growth to theory is ambiguous, it may effectively capture a different process than intended. What could drive differences in earnings from age 23 to age 31? What if the parents at age 23 worked very little (due to prolonged education or childcare), but at age 31 work full time? This could, in some countries, be gender-specific, for example, apply to the mother and not the father. The difference could thus reflect a homemaking mother taking up work, or parents' finishing school and getting their first 'real' job, which carries broader implications than just increased income. An alternative approach is to minimize overlap, that is, to have two independent measures of permanent income. However, since birth spacing is small and year-to-year volatility large, this would result in very short intervals of income and would increase measurement error biases (i.e. the measure will capture transitory income).

The above problem illustrates a more general principle: one of the critical sources of variation in DFD models is siblings' imperfect overlap in lifespan. Hence, within-differences will derive from the period before a sibling is born or after leaving the nest. Such differences will be separated in time, which makes them collinear with family trajectories and other trends, raising issues resembling the classical age-period-cohort separation problem (Glenn, 2003; Fosse and Winship, 2019). Researchers may attempt to adjust for this by including terms for calendar time, birth order, age differences, and so on. But such efforts do not relieve the researcher from the fundamental duty of demonstrating the nature of the remaining variance, which tends to become more marginal and convoluted with each new control term added.

Families that display income stability, for example, because they were older at childbirth, and having completed education and labour market entry, will also be less discordant. The effective identifying sample will be families with a maximum difference between siblings, so the problem of selection into identification (Miller et al., 2023) must be taken into account in this case as well.

Neighbourhood effects

The literature on the effects of living or growing up in neighbourhoods of different affluence or social capital is large but reaches different conclusions depending on methods. In response to earlier work using regression to control for confounding, Plotnick and Hoffman (1999) used family fixed effects instead and found no evidence of neighbourhood effects. Many studies have since followed this path. For example, Sariaslan et al. (2013) analysed the effect of neighbourhood deprivation on adolescent violent criminality and substance misuse in Sweden. Neighbourhood deprivation was based on the neighbourhood at age 15 and measured via a principal component analysis of the share unemployed, share welfare recipients, share less-educated, share divorced, share immigrants, residential mobility, crime rate, and median disposable income. By comparing siblings, they found no effect of neighbourhood deprivation.

But here, too, there is reason to doubt that within-family variation captures the relevant treatment. The effects of neighbourhood exposure are likely to unfold over decades (Galster, 2012; Sharkey and Faber, 2014). This is highlighted by discussion around housing interventions such as the Moving to Opportunity experiment, which offered participants housing vouchers and followed their later outcomes. Evaluations have been mixed, sparking methodological controversy. For example, Wodtke et al. (2011) observe that 'theories of neighbourhood effects all specify mechanisms based on long-term exposure to disadvantaged neighbourhoods [but] most previous studies measure neighbourhood context only once or over just a short period' (cf. Sharkey and Elwert, 2011). The same point applies to DFD models where the neighbourhood is the treatment.

This becomes more salient when we focus on empirical applications. Consider a design using the explanatory variables average income and percent foreign-born in the neighbourhood. By averaging across neighbourhood residents, these measures will by design reduce measurement error, but measurement error is not eliminated. For example, several studies find substantial differences between one-point measures and averages over childhood in the association between neighbourhood exposure and later outcomes, reflecting attenuation bias (Sharkey and Elwert, 2011; Wodtke et al., 2011; Goldschmidt et al., 2017).

Consider an example neighbourhood NH with 10 neighbours N measured on their fixed characteristics (like being foreign-born) three years apart. In this period, one moves out, and another moves in. The overlap is large, and the within-difference of this measure will capture the difference in in- and outflow, that is, the marginal neighbour and not the representative neighbour:

$$NH_1 = \frac{1}{10}\sum N_1, N_2, N_3, N_4, N_5, N_5, N_6, N_7, N_8, N_9, N_{10}, \tag{7}$$

$$NH_4 = \frac{1}{10}\sum N_2, N_3, N_4, N_5, N_5, N_6, N_7, N_8, N_9, N_{10}, N_{11} \tag{8}$$

$$\Delta_{NH1,NH4} = \frac{N_{11} - N_1}{10}. \tag{9}$$

For characteristics that vary over time within individuals (like income, but not being foreign-born), the measure will also capture changes in incomes over time for all neighbours. Again, within-family differences focus on other things than permanent neighbourhood differences. This is theoretically appropriate if the relevant theory is one about neighbourhood change, but most theories in the area refer to, for example, repeated social interaction and exposure, and not isolated experiences.

The problem of selection into identification arises here too, because only discordant families will contribute to effect estimates. In these cases, the effective sample will be selected on families with an unstable residential experience, either because of repeated moves (which could contribute independently to child outcomes; Mollborn et al. 2018; Simsek et al. 2021), or living in neighbourhoods subject to accelerated social change.

A related example is the effect of attending immigrant-dense schools. In an important paper, Borgen (2024) notes that using school-fixed effects to eliminate selection means that the question is narrowed to be exclusively about peer effects, not school effects. A school-fixed effects model, where the variation in percent immigrant comes from different birth cohorts, is precisely an application of what we depict above. The model will estimate the effect of the marginal peer being an immigrant, not the effects of attending a school with, say, 5 percent versus 50 percent immigrants, a difference that will be netted out. Interestingly, Borgen (2024) finds no effect of attending an immigrant-dense school using school-by-program fixed effects and year-on-year variation in the proportion of immigrants. However, she does find an effect of attending an immigrant-dense school when she instead uses an admissions experiment, comparing students with the same GPA and school preferences who are admitted to different schools. This illustrates how a causal effect of between-school differences is not well identified using within-school variation.

Some further considerations

We now take up two counterclaims to our argument. First, we consider several ways to address the challenges to DFD we present, and show that insofar as they succeed, they do so by moving away from DFD and towards other, better-established designs. Second, we consider the role of the sibling correlation as an indicator of analysable variance, specifically, the argument that if sufficient variation between siblings remains there is little concern.

A major worry in DFD is reverse causality, as manifested in evidence of compensatory and reinforcing parental behaviour (Grätz and Torche, 2016; Dierker and Diewald, 2024). One way to establish the order of causality more credibly is through longitudinal data that enable the joint modelling of environmental stimuli and child outcomes, as well as their co-evolution over time (Zachrisson and Dearing, 2015; Dickerson and Popli, 2016). Such modelling alleviates but does not completely resolve the concern of reverse causality, and other obstacles remain. The key point, however, is that longitudinal data points towards a within-person rather than a between-sibling design, and the use of siblings in this setup is unclear.

Another alternative is to focus on children of siblings, that is, a cousin fixed-effects design. This strategy has been used to study the intergenerational transmission of education (Behrman and Rosenzweig, 2002), among other applications (Hällsten and Pfeffer, 2017; Mazrekaj et al., 2020). Some of the studies we have discussed above use this as an alternative strategy to DFD, adding credibility to their findings (Sariaslan et al., 2013, 2014, 2021). Such a design does away with the problems associated with treatments that occur in the same household, as the difference in exposure results from grown siblings who no longer share a nuclear family. In this sense, it is a compromise between within- and between-family designs, and shares the concerns of each—such as unobserved confounders. Crucially, however, insofar as it addresses the objections we have raised, it does so by moving away from DFD and closer to a typical sibling comparison design.

A number of studies focus on events that occur when siblings are of different ages. These could be parental divorce, parental death (Kailaheimo-Lönnqvist and Erola, 2020), but also the timing of births (Duncan et al., 2018). These studies could be seen as more credible, as they do not rely on differences in treatment between siblings, only in the timing of these events. But a subtle point here is that by deriving their variance from differences in sibling age, these studies again move closer to a typical sibling comparison where the source of variation in treatment is characteristics of siblings and not parents. Meanwhile, they remain sensitive to concerns of collinearity with cohort, birth order, or broader time trends.[7]

Many authors are aware that DFD may remove 'too much' variation in the explanatory variable and leave no analysable variance at all in the extreme case. The sibling correlation is sometimes used as a diagnostic device. As the argument goes, if the sibling correlation is low enough, there is 'enough' remaining variance to identify the effect. For example, Sariaslan et al. (2013, p. 1061) state that the 'sibling correlation for the neighborhood-deprivation measure amounted to 0.46... implying that there was a fair amount of variability within families with which to continue our analyses'.[8] Similarly, Sariaslan et al. (2021, p. 1634) argue that the 'null findings within families could not be attributed to insufficient income variability, as the sibling correlations for the family income exposures increased from 0.71 to 0.80 across the four measurement points, thus indicating that between 20 per cent and 30 per cent of the observed income differences were unique to siblings within families'.

One can easily simulate data where, say, the family's permanent income has a causal effect on something and where sibling-specific variation is added to make up a measure of observed income, for example, if income is measured some years apart for the siblings. Running a DFD on the simulated data produces a zero effect. The degree of year-on-year variation can be tuned to meet realistic conditions. With a correlation between permanent and observed income of 0.65, a realistic value (see note 8), we get a sibling correlation of 0.43. This may sound like there is a lot of 'remaining variance' (1 – 0.43 = 0.57), but all of this is just noise. Similar examples apply to parenting and neighbourhoods. The sibling correlation in the parental trait does not tell us whether within-family variation in the parental trait is causally relevant. Therefore, it is uninformative as a diagnostic tool for whether DFD is a viable design.

Directions for future research

Should DFD be discarded altogether? We believe that unless the focus is on differences within families, the scope for DFD is limited. For within-family processes, the design is most likely to work when there is a discrete shock, the source of which is exogenous and external to the family. Examples may include job loss, migration, health shocks, and so on. Even in the absence of a shock, it helps to be clear about the question, 'where does your variation come from?' By laying bare the exact nature of variation that enters into estimation, as well as any threats to the identification proposed, it becomes possible to judge the results in a fair light and arrive at a more balanced interpretation.

As one example, Chetty and Hendren (2018) proposed to identify neighbourhood effects by using family moves that occur at different ages. This is an improvement over the neighbourhood studies we have discussed above, because there is a discontinuity in the form of a move, which allows results to be presented and discussed in more transparent terms. There is also a clear assumption underpinning this strategy: that families' decision to move does not depend on a child's age. Again, it is possible to raise objections. In fact, Heckman and Landersø (2022) present compelling evidence that this assumption most likely does not hold. Instead, parents' residential decisions are a dynamic process, where many move into 'better' neighbourhoods in the period immediately preceding a first birth. Subsequent moves are rare and decline especially rapidly as the firstborn approaches school age. Further, these patterns interact with social background in ways that threaten identification. Still, this type of critical exchange is made possible by clearly stating the identifying assumptions from the outset.

No design is free of flaws, and DFD has a role to play in the toolbox among other research designs, insofar as its assumptions can plausibly be defended and it is clear what form potential violations will take. However, our review of the issues above suggests that recent contributions may have been too optimistic about its potential. We suggest that authors should at a minimum grapple with the below seven questions before embarking on this design:

1. *Is within-family variation theoretically motivated to study?* Researchers should consider the relevant theory and whether it refers to variation between or within families. We believe the case is most clear-cut with family income, thanks to a well-developed theoretical apparatus that distinguishes between permanent and transitory components. In other cases, the distinction may be less developed but nevertheless matter, as we have argued for parenting and neighbourhood context. Many of the studies we have reviewed do not make a distinction between permanent and transitory rearing environment, which makes them liable to misinterpretation. Being clear about the distinction is therefore a first step in the right direction.
2. *Is the source of within-family variation known and transparent?* As we have argued, DFD is unlikely to be convincing without a clear grasp of what causes the remaining variation on which estimation is done. Unless there is an unambiguous source of that variation, endogeneity concerns are often as acute as in a between-family design. A DFD is therefore often most convincing with a source variation that is external to the family. In the case of parental income, this may include mass layoffs, industry-specific shocks, tax or benefit reforms, or lottery wins. On the other hand, with an exogenous treatment, the added value of DFD is doubtful and it may add concerns such as selection into treatment that do not apply with a more representative sample.
3. *Is it possible to rule out reverse causality within families?* The use of DFD is often motivated with respect to between-family confounding. Reverse causality is less often discussed but, paradoxically, using within-family variation may heighten concerns about the direction of causality. The clearest example is parenting, which generally arises in direct response to the type of outcomes that these studies examine, including children's skills, behaviour problems, or health. Concerns about reverse causality within families are not absent for the other treatments we consider either. For instance, parents may reduce their income to care for a sick sibling, or a family with an academically gifted sibling may move to a neighbourhood with good schools.
4. *Is it possible to rule out unobserved confounding within families?* Given that unobserved confounding is the usual motivation to adopt DFD, it may seem like a successful strategy to deal with this concern. However, studies that use the design rarely specify the confounders they have in mind, and instead, make blanket statements like 'all unmeasured confounders shared by siblings'. Without some idea of what the relevant confounders might be, it becomes difficult to judge whether confounding is more or less likely within families than between them. For example, if a younger sibling grows up in a higher-earning family compared to their older sibling, they might also have spent less time with their parents and more time in childcare (Cooper and Stewart, 2021).
5. *Is the design free from collinearity with birth order, cohort, age, and calendar time?* One main source of independent variation in DFD is siblings' imperfect overlap in lifespan. These differences are intertwined with family trajectories and other trends, which raises issues resembling the classical age-period-cohort problem. Researchers may adjust for this by including terms for calendar time, birth order, and age differences, but interpreting the remaining variance becomes increasingly complex. Some studies focus on events occurring when siblings are of different ages, such as parental divorce or death, aiming to identify differences in treatment based on age-related fragility. However,

age differences may also reflect cohort differences or other trends, posing a risk of confounding.

6. *Are we confident that there is no selection into identification?* DFD biases the sample towards larger families and families with greater discordance in treatment. As we have argued, discordance may be systematically related to background characteristics, as it entails more erratic or adaptive parenting, higher income fluctuation, or greater residential instability. Unlike some other issues, this one can be addressed by inspecting differences in observable characteristics between the full dataset and estimation sample. But the fundamental question of how average treatment effects may differ in the estimation sample and the population remains unknown.
7. *Are we confident that within-family variation is not just measurement error?* Researchers must ensure that reported variation in parental involvement and other treatments is not solely due to measurement error or inconsistent reporting by family members. In fixed-effects models, de-meaning reduces signal without reducing noise, potentially amplifying classical attenuation bias. This amplification relies on the assumption that measurement errors are uncorrelated with each other and the outcome. The correlation of errors depends on the data source: errors are more likely to correlate if reported by the same respondent, such as parents or siblings reporting on each other, whereas separate reports or external sources may lead to uncorrelated errors.

It is unlikely that all of these questions will ever be answered in the affirmative. This does not itself rule out the design, but whenever an answer is 'no', the burden of proof is on the authors. Further argument or evidence will be needed to substantiate the design's usefulness and guide the reader in interpreting its results.

What other strategies are available? Cooper and Stewart (2021) provide a helpful summary of studies attempting to identify the causal effect of parental earnings on child outcomes. Apart from fixed effects or other similar techniques, these include various quasi-experiments such as policy reforms or, in rare cases, randomized controlled trials (RCTs). Quasi-experiments include industry-specific shocks to the local economy, windfall profits from casino revenue and lottery wins, or expansions of programs such as the Earned Income Tax Credit (Akee et al., 2010; Dahl and Lochner, 2012; Løken et al., 2012; Cesarini et al., 2016). RCTs involve the random implementation of cash-transfer programs, often with a variety of treatment arms combining different inputs (Gennetian and Miller, 2002; Cancian et al., 2013). The interpretation of these studies highlights their relative strengths and limitations. For example, a lottery win is clearly exogenous conditional on participating but may not alter long-run income, leading to the same concern about permanent and transitory income that we have highlighted.

There is no shortage of experiments, whether actual or natural, for the other areas we consider: parenting and neighbourhood inputs. Reforms around parental leave can induce exogenous change in the time spent with a child (Carneiro et al., 2015). Intervention programs like the Perry Preschool or Abecedarian projects provide comprehensive support to improve parenting (García and Heckman, 2023). Some of these are omnibus programs that focus on a wide range of inputs, while others target educational activities more narrowly (Mayer et al., 2019; Kalil et al., 2023). A 'genetic nurture' design can leverage the influence of parents' non-transmitted genetic variants to study household environment (Wang et al., 2021). Housing mobility has been studied with voucher experiments such as the Moving to Opportunity study (Clampet-Lundquist and Massey, 2008; Ludwig et al., 2008), while catastrophic events such as natural disasters have been shown to induce moves and alter life-course trajectories (Sacerdote, 2012; Nakamura et al., 2022).

In general, DFD is likely to be most useful as one tool among many, where the differences in estimation results can be understood in light of the various biases that are thought to affect one estimation strategy or the other. Instead of shoehorning results into a binary framework of hypothesis confirmation or refutation, we would encourage researchers to use a wide variety of designs and refine theory in the light of disparate findings. This model of inference is a powerful tool for knowledge generation (Goldthorpe, 2016; Engzell and Mood, 2023) and has much in common with what Lieberson and Horwich (2008) call 'implication analysis'. One example we would consider successful in this regard is Holmlund et al. (2011) who assess a variety of strategies in estimating the causal effect of parents' schooling on children's schooling, including comparing twin mothers, parents of adoptive children, and education reforms. By juxtaposing methods subject to different presumptive biases, a coherent picture can emerge, but this requires openness about the limitations of each method.

Conclusion

If there is a wider lesson to be learned from the discordant family design, it is perhaps this. Any empirical strategy that claims to net out confounding factors needs a convincing story about what generates the remaining variation that is used to identify and estimate treatment effects. While the notion of adjusting for unobserved factors seems appealing, it lacks practical value without a comprehensive understanding of the remaining variation. The first question for any identification strategy should instead be, 'where does your variation come from?'

Opening up the black box of DFD as we have done helps us ask generative questions and tighten the link between estimation and theory. In doing so, we add to the literature that has pursued the same aim in other fixed-effects applications (Halaby, 2004; Plümper et al., 2005; Rohrer and Murayama, 2023). Crucially, in many cases, the relevant mechanisms and theories do not speak to inequality within but across families. The estimated parameter, therefore, does not equal the theoretically relevant target estimand (Lundberg et al., 2021). It is necessary to relate design and measurement to substantive theories, specifically asking how they relate to within-family differences.

We have discussed the limitations of DFD, both concerning theory and method. For several applied examples in the literature, we showed that the variance used in estimation is at best loosely connected to theories about child-rearing. It is hard to say anything about the sign and magnitude of the resulting bias, but there are more reasons to expect a bias toward the null than its opposite. Ultimately, researchers should proceed with caution before embarking on this design. Our survey recommends DFD not as a panacea to rid estimates of confounding but rather one of many complementary strategies that may or may not be useful depending on the substantive question at hand.

Notes

1 The origin of the approach is elusive. While it made a brief appearance in economics through a few papers around the millennial shift, it did not gain widespread traction in that field. Ermisch and Francesconi (2001) note that 'most empirical studies linking parents' behaviour and children's attainments have not addressed the problem of unmeasured heterogeneity with sibling estimators'. They attribute the introduction of the design to yet unpublished papers from 1996 and 1997. The earliest example we have been able to identify is Rosenzweig and Wolpin (1995); however, their focus is on intra-familial processes. In sociology and other disciplines, the adoption of the design gained momentum later, in the mid-2010s.

2 Like many biases, there is more than one way to formalize selection into identification. We follow Elwert and Winship (2014) in characterizing endogenous selection bias as a problem of conditioning on a collider variable. Because sibling pair discordance is a function of variation in the treatment (E_{ij}) and the outcome (Y_{ij}), selection into identification opens a collider path. An alternative way of understanding this problem is that the average treatment effect (ATE) in the discordant sample is non-representative of the population ATE (Miller et al., 2023).

3 While the empirical literature shows that not all parenting is uniform across siblings, uniform parenting still strongly dominates differential parenting (Holden and Miller, 1999; Feinberg and Hetherington, 2001).

4 Meanwhile, even 'identical' twins can differ considerably from birth due to a complex mix of genetic and epigenetic processes, and competition for resources in utero (Hall, 2003; Silva et al., 2011). In fact, around a quarter of monozygotic twins live under the impression that they are dizygotic (Conley et al., 2013).

5 Of course, some parents may treat siblings or twins differently simply as a way of encouraging their individuality, which should not be taken as a sign of harshness. At the same time, differential treatment can create feelings of resentment and competition, so most parents will probably aim for some degree of balance.

6 Not all researchers agree on the permanent income hypothesis. Mayer (1997, pp. 72–78) suggests that families with the same permanent income but losing or gaining income over time should differ in consumption behaviour. Still, her descriptive findings analysing children's test scores, behaviour, and educational attainment suggest that permanent income is much more important than trends (i.e. incomes measured at different time points in childhood).

7 A specific risk for these types of studies is that the age differences pick up generic trends in the outcome, such as educational expansion if one studies education. Kalil et al. (2016) offer a solution to the collinearity problem by using non-exposed siblings to estimate the cohort difference, and explicitly control away any trends. However, our experience is that this technique is extremely demanding for statistical power (Barclay and Hällsten, 2022).

8 For example, in Swedish register data of individuals born in 1973 observed between 1991 and 2018 (age 18–45) representing the parental generation, the correlation between permanent and observed income is between 0.6 and 0.7 at age 35–45. The correlation between incomes spaced one year apart is around 0.8 and 0.4 for incomes 10 years apart (i.e. the sibling correlation in parental income measured at these intervals would be 0.8 and 0.4).

Acknowledgements

Earlier versions of this work were presented at the 2022 European Consortium for Sociological Research (ECSR) Conference in Amsterdam and the Swedish Institute for Social Research, Stockholm University. We thank participants at these occasions, especially Adam Altmejd, Kirsten van Houdt, Stephanie Plenty, and Are Skeie Hermansen, as well as the Editor-in-Chief, editorial team, and three anonymous reviewers at *European Sociological Review*. This paper was submitted and conditionally accepted prior to Engzell's tenure as an Associate Editor for the journal.

Author contributions

Per Engzell (Conceptualization [equal], Investigation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Martin Hallsten (Conceptualization [equal], Investigation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal])

Funding

This work was supported by the Swedish Research Council, VR (grants no. 2019-02552 and 2022-02314).

References

Achard, P. (2022). *Ethnic Enclaves and Cultural Assimilation*, CentER Discussion Paper 2022-011. Tilburg: Tilburg University.

Akee, R. K. *et al.* (2010). Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits. *American Economic Journal: Applied Economics*, **2**, 86–115.

Avinun, R. and Knafo, A. (2014). Parenting as a reaction evoked by children's genotype: a meta-analysis of children-as-twins studies. *Personality and Social Psychology Review*, **18**, 87–102.

Barclay, K. and Hällsten, M. (2022). Does the impact of parental death vary by parental socioeconomic status? A study of children's educational and occupational attainment. *Journal of Marriage and Family*, **84**, 141–164.

Baumrind, D. (1966). Effects of authoritative parental control on child behavior. *Child Development*, **37**, 887–907.

Becker, G. S. and Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy*, **87**, 1153–1189.

Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, **4**, S1–S39.

Behrman, J. R. and Rosenzweig, M. R. (2002). Does increasing women's schooling raise the schooling of the next generation? *American Economic Review*, **92**, 323–334.

Blau, D. M. (1999). The effect of income on child development. *Review of Economics and Statistics*, **81**, 261–276.

Bohrnstedt, G. W. and Carter, T. M. (1971). Robustness in regression analysis. *Sociological Methodology*, **3**, 118–146.

Borgen, S. T. (2024). It's not all about the peers: reintroducing school context to the school segregation literature. *Acta Sociologica*, **67**, 66–87.

Bound, J. and Solon, G. (1999). Double trouble: on the value of twins-based estimation of the return to schooling. *Economics of Education Review*, **18**, 169–182.

Bourdieu, P. (1990). *The Logic of Practice*. Stanford: Stanford University Press.

Brady, D. *et al.* (2018). How to measure and proxy permanent income: evidence from Germany and the US. *The Journal of Economic Inequality*, **16**, 321–345.

Breinholt, A. and Conley, D. (2023). Child-driven parenting: differential early childhood investment by offspring genotype. *Social Forces*, **102**, 310–329.

Breinholt, A. and Holm, A. (2020). Heterogeneous effects of less educated mothers' further education during early childhood on children's educational performance in adolescence. *Research in Social Stratification and Mobility*, **68**, 100506.

Cancian, M., Yang, M. -Y. and Slack, K. S. (2013). The effect of additional child support income on the risk of child maltreatment. *Social Service Review*, **87**, 417–437.

Carneiro, P., Løken, K. V. and Salvanes, K. G. (2015). A flying start? Maternity leave benefits and long-run outcomes of children. *Journal of Political Economy*, **123**, 365–412.

Cesarini, D. *et al.* (2016). Wealth, health, and child development: evidence from administrative data on Swedish lottery players. *Quarterly Journal of Economics*, **131**, 687–738.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: childhood exposure effects. *Quarterly Journal of Economics*, **133**, 1107–1162.

Chevalier, A. *et al.* (2013). The impact of parental income and education on the schooling of their children. *IZA Journal of Labor Economics*, **2**, 1–22.

Chia, Y. F. (2013). Dollars and pounds: the impact of family income on childhood weight. *Applied Economics*, **45**, 1931–1941.

Clampet-Lundquist, S. and Massey, D. S. (2008). Neighborhood effects on economic self-sufficiency: a reconsideration of the moving to opportunity experiment. *American Journal of Sociology*, **114**, 107–143.

Conley, D. *et al.* (2013). Heritability and the equal environments assumption: evidence from multiple samples of misclassified twins. *Behavior Genetics*, **43**, 415–426.

Conley, D. *et al.* (2015). Is the effect of parental education on offspring biased or moderated by genotype? *Sociological Science*, **2**, 82–105.

Cooper, K. and Stewart, K. (2021). Does household income affect children's outcomes? A systematic review of the evidence. *Child Indicators Research*, **14**, 981–1005.

D'Onofrio, B. M. *et al.* (2009). A quasi-experimental analysis of the association between family income and offspring conduct problems. *Journal of Abnormal Child Psychology*, **37**, 415–429.

Dahl, G. B. and Lochner, L. (2012). The impact of family income on child achievement: evidence from the earned income tax credit. *American Economic Review*, **102**, 1927–1956.

Dickerson, A. and Popli, G. K. (2016). Persistent poverty and children's cognitive development: evidence from the UK millennium cohort study. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **179**, 535–558.

Dierker, P. and Diewald, M. (2024). Compensation or accentuation? How parents from different social backgrounds decide to support their children. *European Sociological Review*, jcae010.

Duncan, G. J. *et al.* (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review*, **63**, 406–423.

Duncan, G. J. *et al.* (2018). Maternal age and child development. *Demography*, **55**, 2229–2255.

Elstad, J. I. and Bakken, A. (2015). The effects of parental income on Norwegian adolescents' school grades: a sibling analysis. *Acta Sociologica*, **58**, 265–282.

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

Engzell, P. (2021). What do books in the home proxy for? A cautionary tale. *Sociological Methods & Research*, **50**, 1487–1514.

Engzell, P. and Jonsson, J. O. (2015). Estimating social and ethnic inequality in school surveys: biases from child misreporting and parent nonresponse. *European Sociological Review*, **31**, 312–325.

Engzell, P. and Mood, C. (2023). Understanding patterns and trends in income mobility through multiverse analysis. *American Sociological Review*, **88**, 600–626.

Ermisch, J. F. and Francesconi, M. (2001). Family structure and children's achievements. *Journal of Population Economics*, **14**, 249–270.

Ermisch, J., Francesconi, M. and Pevalin, D. J. (2004). Parental partnership and joblessness in childhood and their influence on young people's outcomes. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **167**, 69–101.

Feinberg, M. and Hetherington, E. M. (2001). Differential parenting as a within-family variable. *Journal of Family Psychology*, **15**, 22–37.

Fosse, E. and Winship, C. (2019). Analyzing age-period-cohort data: a review and critique. *Annual Review of Sociology*, **45**, 467–492.

Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Galster, G. C. (2012). The mechanism(s) of neighbourhood effects: theory, evidence, and policy implications. In van Ham, M., Manley,D., Bailey, N., Simpson, L. and Maclennan, D. (Eds.), *Neighbourhood Effects Research: New perspectives*. Berlin: Springer, pp. 23–56.

García, J. L. and Heckman, J. J. (2023). Parenting promotes social mobility within and across generations. *Annual Review of Economics*, **15**, 349–388.

García-Sierra, A. (2024). Parental responses to children's early health disadvantages: evidence from a British twin study. *European Sociological Review*, jcae016.

Gennetian, L. A. and Miller, C. (2002). Children and welfare reform: a view from an experimental welfare program in Minnesota. *Child Development*, **73**, 601–620.

Glenn, N. D. (2003), Distinguishing age, period, and cohort effects. In Mortimer, J. T. and Shanahan, M. J. (Eds.), *Handbook of the Life Course*. Berlin: Springer, pp. 465–476.

Goldschmidt, T., Hällsten, M. and Rydgren, J. (2017), *Are they hunkering down? Revisiting the relationship between exposure to ethnic diversity, intergroup contact, and group trust*, Department of Sociology Working Paper Series No. 30, Stockholm University.

Goldthorpe, J. H. (2016). *Sociology as a Population Science*. Cambridge: Cambridge University Press.

Gorseline, D. E. (1932). *The Effect of Schooling upon Income*. Bloomington: Indiana University (PhD Dissertation).

Grätz, M. and Torche, F. (2016). Compensation or reinforcement? The stratification of parental responses to children's early ability. *Demography*, **53**, 1883–1904.

Grätz, M., Lang, V. and Diewald, M. (2022). The effects of parenting on early adolescents' noncognitive skills: evidence from a sample of twins in Germany. *Acta Sociologica (Copenhagen, Denmark)*, **65**, 398–419.

Griliches, Z. (1979). Sibling models and data in economics: beginnings of a survey. *Journal of Political Economy*, **87**, S37–S64.

Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, **96**, 1308–1320.

Halaby, C. N. (2004). Panel models in sociological research: theory into practice. *Annual Review of Sociology*, **30**, 507–544.

Hall, J. G. (2003). Twinning. *Lancet (London, England)*, **362**, 735–743.

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

Hauser, R. M. and Sewell, W. H. (1986). Family effects in simple models of education, occupational status, and earnings: findings from the Wisconsin and Kalamazoo studies. *Journal of Labor Economics*, **4**, S83–S115.

Heckman, J. and Landersø, R. (2022). Lessons for Americans from Denmark about inequality and social mobility. *Labour Economics*, **77**, 101999.

Holden, G. W. and Miller, P. C. (1999). Enduring and different: a meta-analysis of the similarity in parents' child rearing. *Psychological Bulletin*, **125**, 223–254.

Holmlund, H., Lindahl, M. and Plug, E. (2011). The causal effect of parents' schooling on children's schooling: a comparison of estimation methods. *Journal of Economic Literature*, **49**, 615–651.

Jæger, M. M. (2011). Does cultural capital really affect academic achievement? New evidence from combined sibling and panel data. *Sociology of Education*, **84**, 281–298.

Jæger, M. M. and Møllegaard, S. (2017). Cultural capital, teacher bias, and educational success: new evidence from monozygotic twins. *Social Science Research*, **65**, 130–144.

Jensen, S. S., Lindemann, K. and Weiss, F. (2023). Parental job loss and the role of unemployment duration and income changes for children's education. *European Sociological Review*, jcad068.

Kailaheimo-Lönnqvist, S. and Erola, J. (2020). Child's age at parental death and university education. *European Societies*, **22**, 433–455.

Kalil, A. *et al*. (2016). Father presence and the intergenerational transmission of educational attainment. *Journal of Human Resources*, **51**, 869–899.

Kalil, A. *et al*. (2023). *Nudging or nagging? Conflicting effects of behavioral tools*. Becker Friedman Institute for Economics Working Paper 2023-02. University of Chicago.

Lareau, A. (2011), *Unequal Childhoods*. Berkeley: University of California Press.

Lehti, H., Erola, J. and Karhula, A. (2019). The heterogeneous effects of parental unemployment on siblings' educational outcomes. *Research in Social Stratification and Mobility*, **64**, 100439.

Lieberson, S. and Horwich, J. (2008). Implication analysis: a pragmatic proposal for linking theory and data in the social sciences. *Sociological Methodology*, **38**, 1–50.

Løken, K. V., Mogstad, M. and Wiswall, M. (2012). What linear estimators miss: the effects of family income on child outcomes. *American Economic Journal: Applied Economics*, **4**, 1–35.

Looker, E. D. (1989). Accuracy of proxy reports of parental status characteristics. *Sociology of Education*, **62**, 257–276.

Ludwig, J. *et al*. (2008). What can we learn about neighborhood effects from the moving to opportunity experiment? *American Journal of Sociology*, **114**, 144–188.

Lundberg, I., Johnson, R. and Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects

statistical evidence to theory. *American Sociological Review*, **86**, 532–565.
Markussen, S. and Røed, K. (2022). Are richer neighborhoods always better for the kids? *Journal of Economic Geography*, **23**, 629–651.
Mayer, S. E. (1997), *What Money Can't Buy: Family Income and Children's Life Chances*. Cambridge, MA: Harvard University Press.
Mayer, S. E. *et al.* (2019). Using behavioral insights to increase parental engagement: the parents and children together intervention. *Journal of Human Resources*, **54**, 900–925.
Mazrekaj, D., De Witte, K. and Cabus, S. (2020). School outcomes of children raised by same-sex parents: evidence from administrative panel data. *American Sociological Review*, **85**, 830–856.
Miller, D. L., Shenhav, N. and Grosz, M. (2023). Selection into identification in fixed effects models, with application to head start. *Journal of Human Resources*, **58**, 1523–1566.
Mollborn, S., Lawrence, E. and Root, E. D. (2018). Residential mobility across early childhood and children's kindergarten readiness. *Demography*, **55**, 485–510.
Nakamura, E., Sigurdsson, J. and Steinsson, J. (2022). The gift of moving: intergenerational consequences of a mobility shock. *Review of Economic Studies*, **89**, 1557–1592.
Pardini, D. A. (2008). Novel insights into longstanding theories of bidirectional parent–child influences: introduction to the special section. *Journal of Abnormal Child Psychology*, **36**, 627–631.
Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, **82**, 702–710.
Plotnick, R. D. and Hoffman, S. D. (1999). The effect of neighborhood characteristics on young adult outcomes: alternative estimates. *Social Science Quarterly*, **80**, 1–18.
Plümper, T., Troeger, V. E. and Manow, P. (2005). Panel data analysis in comparative politics: linking method to theory. *European Journal of Political Research*, **44**, 327–354.
Rohrer, J. M. and Murayama, K. (2023). These are not the effects you are looking for: causality and the within-/between-persons distinction in longitudinal data analysis. *Advances in Methods and Practices in Psychological Science*, **6**, 251524592211408–251524592211414.
Rosenzweig, M. R. and Wolpin, K. I. (1995). Sisters, siblings, and mothers: the effect of teen-age childbearing on birth outcomes in a dynamic family context. *Econometrica*, **63**, 303–326.
Sacerdote, B. (2012). When the saints go marching out: long-term outcomes for student evacuees from Hurricanes Katrina and Rita. *American Economic Journal: Applied Economics*, **4**, 109–135.
Sariaslan, A. *et al.* (2013). The impact of neighbourhood deprivation on adolescent violent criminality and substance misuse: a longitudinal, quasi-experimental study of the total Swedish population. *International Journal of Epidemiology*, **42**, 1057–1066.
Sariaslan, A. *et al.* (2014). Childhood family income, adolescent violent criminality and substance misuse: quasi-experimental total population study. *The British Journal of Psychiatry*, **205**, 286–290.
Sariaslan, A. *et al.* (2021). No causal associations between childhood family income and subsequent psychiatric disorders, substance misuse and violent crime arrests: a nationwide finnish study of >650 000 individuals and their siblings. *International Journal of Epidemiology*, **50**, 1628–1638.
Scarr, S. and McCartney, K. (1983). How people make their own environments: a theory of genotype → environment effects. *Child Development*, **54**, 424–435.
Sharkey, P. and Elwert, F. (2011). The legacy of disadvantage: multigenerational neighborhood effects on cognitive ability. *American Journal of Sociology*, **116**, 1934–1981.
Sharkey, P. and Faber, J. W. (2014). Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects. *Annual Review of Sociology*, **40**, 559–579.
Silva, S. *et al.* (2011). Why are monozygotic twins different? *Journal of Perinatal Medicine*, **39**, 195–202.
Simsek, M., Costa, R. and de Valk, H. A. (2021). Childhood residential mobility and health outcomes: a meta-analysis. *Health & Place*, **71**, 102650.
Sjölander, A., Frisell, T. and Öberg, S. (2021). Sibling comparison studies. *Annual Review of Statistics and Its Application*, **9**, 71–94.
Solon, G. (1989). Biases in the estimation of intergenerational earnings correlations. *Review of Economics and Statistics*, **71**, 172–174.
Tamm, M. (2008). Does money buy higher schooling?: evidence from secondary school track choice in Germany. *Economics of Education Review*, **27**, 536–545.
Tominey, E. (2010). *The timing of parental income and child outcomes: the role of permanent and transitory shocks'*. Discussion Papers 10/21, Department of Economics, University of York, York.
Torche, F. (2015). Analyses of intergenerational mobility: an interdisciplinary review. *Annals of the American Academy of Political and Social Science*, **657**, 37–62.
Visscher, P. M. *et al.* (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genetics*, **2**, e41.
Wang, B. *et al.* (2021). Robust genetic nurture effects on education: a systematic review and meta-analysis based on 38,654 families across 8 cohorts. *American Journal of Human Genetics*, **108**, 1780–1791.
Wodtke, G. T., Harding, D. J. and Elwert, F. (2011). Neighborhood effects in temporal perspective: the impact of long-term exposure to concentrated disadvantage on high school graduation. *American Sociological Review*, **76**, 713–736.
Zachrisson, H. D. and Dearing, E. (2015). Family income dynamics, early childhood education and care, and early child behaviour problems in Norway. *Child Development*, **86**, 425–440.

Per Engzell is Associate Professor of Sociology at UCL Social Research Institute, University College London, Associate Member of Nuffield College, University of Oxford, and Docent at the Swedish Institute for Social Research, Stockholm University. His research interests include intergenerational mobility and educational stratification.

Martin Hällsten is Professor of Sociology at the Swedish Institute for Social Research, Stockholm University. Current research interests include multigenerational inequality processes.