Stockholm Research Reports in Demography | no 2021:4

Problems in the family

Controlling for age, period or cohort in sibling comparison designs

Ben Wilson, Maarten J. Bijlsma, Siddartha Aradhya and Alice Goisis

ISSN 2002-617X | Department of Sociology

Stockholm University
Demography Unit

Problems in the family

Controlling for age, period or cohort in sibling comparison designs

Ben Wilson[1,2]

Maarten J. Bijlsma[3,4]

Siddartha Aradhya[1]

Alice Goisis[3,5]

1: Department of Sociology, Stockholm University
2: Department of Methodology, London School of Economics
3: Max Planck Institute for Demographic Research
4: Groningen Research Institute of Pharmacy (GRIP), PTEE, University of Groningen
5: UCL Social Research Institute

Abstract

The sibling comparison design is increasingly used across the social sciences as a means of controlling for observed and unobserved confounding, but this design is not without its methodological challenges. One challenge is to control for age, period or cohort when studying exposures that are measured in units of time. In such cases, variation between siblings in the exposure is typically collinear with variation in age, period or cohort – which makes controlling for these factors highly problematic. We address this challenge by showing how it is possible to control for cohort (or age or period) and obtain unbiased estimates of the effect of an exposure that is measured in units of time, albeit given certain assumptions. Using a simulation study that compares a series of estimators, we show that bias can be minimised by including a group who are discordant on age, period or cohort but concordant on the exposure.

Keywords: Sibling comparison, Sibling fixed effects, Bias, Confounding, Age, Period, Cohort, APC

Stockholm Research Reports in Demography 2021:4
ISSN 2002-617X

© Ben Wilson, Maarten J. Bijlsma, Siddartha Aradhya and Alice Goisis

This work is licensed under a Creative Commons Attribution 4.0 International License.

Introduction

There is a long history of quantitative social science that uses a sibling comparison design. This design is typically used when trying to infer the causal effect of one variable on another, often referred to as the effect of an exposure on an outcome. One of the problems for causal inference is confounding, where the association between an exposure and an outcome is at least partly explained by another variable that causes both the exposure and outcome (for discussion and formal definitions, see [1,2]). The sibling comparison design is useful because it enables researchers to control for a range of confounding factors that are shared between siblings, including many factors relating to childhood, parents, and (some) shared genetics.[3,4] While far from new,[5–8] its use appears to be growing in sociology and across the social sciences, not only because of the increasing availability of high quality data in which siblings can be identified and linked, but also because of the increasing ease of estimating models by including what are often referred to as 'sibling fixed effects' (SFEs, which represent a fixed intercept for each sibling group, e.g. those with the same mother).[3,4] In a simple search of PubMed (1971-2019), we found more than 1,000 articles that appeared to mention the use of SFEs, sibling comparison or sibling control.[A] These methods are being applied in a wide range of disciplines, with many examples in sociology,[9–14] demography,[15–18] economics,[19,20] and epidemiology.[21,22] At the same time, there are a considerable number of methodological papers that discuss the relative merits of the sibling comparison design and the issues that arise when using this research method.[3,23–27]

Methodological research has raised a number of crucial issues for researchers who use sibling comparisons, such as the difficulties of interpreting null results,[26] the challenges of dealing with a lack of independence between siblings,[25] and the limits of interpretation with respect to causal inference.[24] Together, this research suggests the need to be cautious when interpreting the results of sibling comparison designs. Moreover, there is an additional challenge for studies that use a sibling comparison design to try to estimate the effect of exposures that are measured in units of time. The challenge is that such 'time exposure' effects are likely to be confounded by age, period or birth cohort (APC), but it is unclear how to control for these APC confounders in a sibling comparison design because they are collinear with the exposure when comparing siblings within the same family. We illustrate this problem in detail in the following section (see: *Understanding the issue*). It may arise, for example, when studying the impact of maternal age on children's birth weight. In this case it is desirable to control for the child's birth cohort, given that younger cohorts are more likely to be born at older maternal ages and less likely to have low birthweight (e.g. because of improvements to medical care). It is also desirable to control for SFEs (e.g. because of confounders shared between siblings such as mother's obstetric history or lifestyle that are often unobserved in the data). However, controls for birth cohort

[A] (((((("sibling control"[Title/Abstract]) OR "between siblings"[Title/Abstract])) OR "sibling design"[Title/Abstract])) OR "sibling comparison"[Title/Abstract]

are collinear (i.e. perfectly correlated) with the exposure – maternal age – after the inclusion of SFEs. This is because the numerical difference in both variables (maternal age and birth cohort) is the same for any pair of siblings. This issue has been highlighted by prior research, which has shown – with a focus on the study of maternal age effects –that it is not easily resolved.[16,28] Not only does the issue have the potential of invalidating the results of prior research, but it may also discount the use of sibling comparison designs in future, at least in cases where exposures are measured in units of time and researchers also wish to control for age, period or cohort. Given the increasing use of sibling comparison designs that meet these criteria, this is far from a niche concern.

Here, we respond to this concern by showing applied researchers how they can study an exposure that is measured in units of time while also controlling for age, period or cohort. One of our main contributions is that we compare a range of approaches, some of which have been used independently in applied research (i.e. in separate studies using different data). Our comparison enables a more reliable assessment of different modelling approaches, and their utility for research. Our results demonstrate how researchers might control for birth cohort (or age, or period) in a sibling comparison design when exposures of interest are measured in units of time but potentially confounded by APC variables. To meet our aims, we carry out a simulation study. Our research design is informed by guidance on the use of simulation studies to evaluate statistical methods.[29,30] In particular, we follow the ADEMP structure, which suggests five aspects of research design that should be stated explicitly and justified in such a study: (1) Aims, (2) Data-generating mechanisms, (3) Estimands, (4) Methods and (5) Performance measures.[29] In the following sections, we define these five criteria and their relevance for our overall objective. Before doing so we provide an elaboration of the problem, alongside several motivating examples.

Understanding the issue

There are at least three broad types of family comparison design, based on a comparison between: (a) *twins*, (b) *siblings*, and (c) *cousins*. We focus on siblings here, but return to the other family comparisons in the discussion. In a sibling comparison design, the aim is to control for factors that are shared between siblings, which makes it a potentially powerful approach because it not only controls for many observed confounders, but also many unobserved (potential) confounders, including a range of factors relating to family background. Regardless of how they are estimated (e.g. using SFEs), the identification of such comparisons relies upon siblings being discordant on the exposure. This means that such a design cannot be used to examine the role of exposures that are shared (i.e. concordant) between siblings, such as mother's country of birth.

Despite this limitation, there are countless examples of exposures that are discordant for (at least some) siblings in most populations. In many cases, this discordance varies over time, for example by age, period and cohort. This is most obviously the case when the exposure is time-varying, for example when studying the effect of income, or body-mass index. It is also the case when the exposure is measured in units of time. For example, researchers have used sibling comparison designs to study the effects of: parental age,[16,22,31,32] age at parental divorce,[9,33] birth intervals,[15,34] and age at arrival for immigrants who arrive as children.[17,19,20,35] We use the latter as our motivating example here, while noting that what applies for this example will apply for many others (a point we elaborate upon in the discussion).

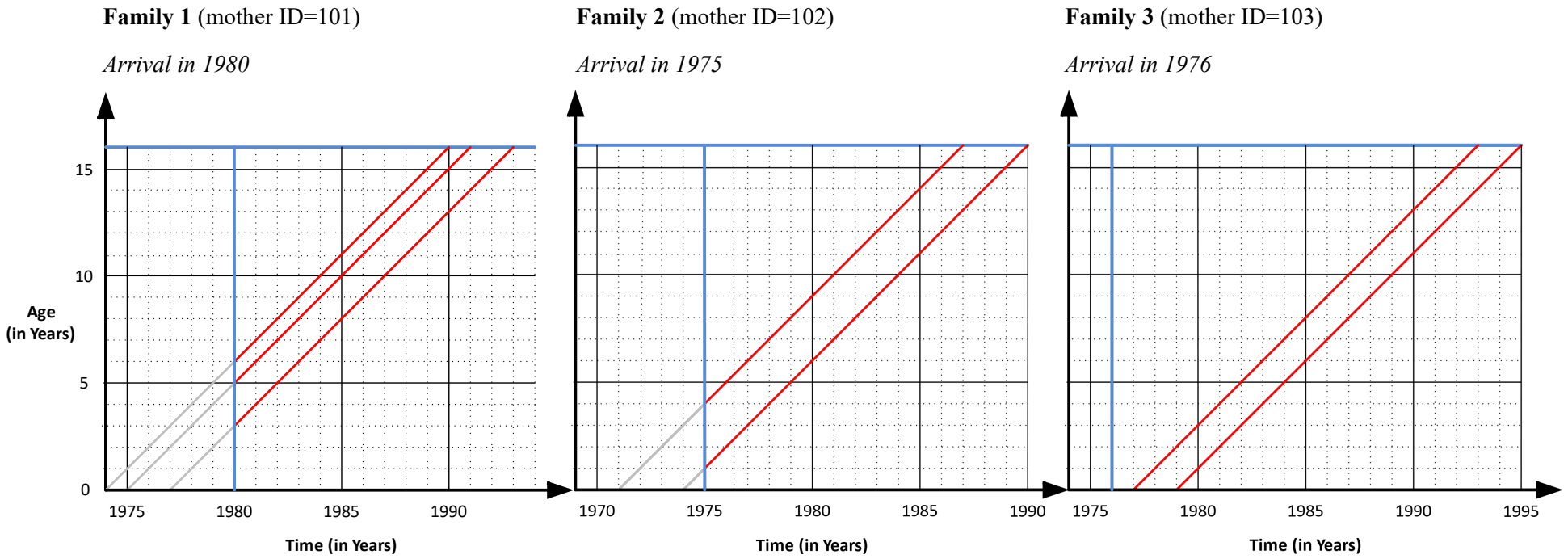
Researchers have tried to estimate the impact of age at arrival on various social outcomes of childhood immigrants, such as their education, income and health.[17,19,20,35] The use of SFEs is appealing in this case because there are many potential confounders that are shared between siblings, in particular family background and reasons for migration, which are rarely measured in observational data. In this example, the problem is that age at arrival is intrinsically linked with age, period and birth cohort. For example, a one year increase in birth cohort (e.g. from 2000 to 2001) implies a one-year decrease in age at arrival (e.g. from age 10 to age 9 for someone arriving in 2010). On its own, this is not a material issue because not all individuals with a specific age at arrival are from the same birth cohort. However, when sibling fixed effects are used, then effect estimates derive from a comparison between siblings from the same family (e.g. with the same mother). In this instance, age at arrival and birth cohort are almost always collinear. We demonstrate this in Table 1 and the accompanying Lexis diagrams in Figure 1.

Table 1: Within-family collinearity: An example of Swedish population data for siblings with the same mother

Family	ID: Mother	ID: Sibling	Country of birth: Mother	Country of birth: Sibling	Year of arrival: Mother	Birth cohort: Sibling	Age at arrival: Sibling	Birth cohort: deviation from family mean	Age at arrival: deviation from family mean	Within family collinearity
1	101	1	Iraq	Iraq	1980	1974	6	- 1.3	1.3	Yes
1	101	2	Iraq	Iraq	1980	1975	5	- 0.3	0.3	Yes
1	101	3	Iraq	Iraq	1980	1977	3	1.7	- 1.7	Yes
2	102	4	Iraq	Iraq	1975	1971	4	- 2.0	2.0	Yes
2	102	5	Iraq	Iraq	1975	1974	1	1.0	- 1.0	Yes
2	102	6	Iraq	Iraq	1975	1974	1	1.0	- 1.0	Yes
3	103	7	Iraq	Sweden	1976	1977	n/a	- 1.3	n/a	No
3	103	8	Iraq	Sweden	1976	1979	n/a	0.7	n/a	No
3	103	9	Iraq	Sweden	1976	1979	n/a	0.7	n/a	No

*Note: In **Family 1** (mother ID = 101), there is the same difference between each pair of siblings in terms of age at arrival as there is in terms of birth cohort (see the last three columns). The same is true for **Family 2** (mother ID = 102), which includes two twins who were born in the same year. The only families where this is not be the case are those with children born after arrival (e.g. **Family 3**, mother ID = 103) and those families where childhood migrants arrive in different years. We note that the latter are very rare, at least in the case of Sweden (for which we have access to whole population data). Age at arrival is the same (recorded as 'n/a') for those siblings who were born after their mother's migration. In some datasets this is recorded as zero, however we note that zero could also indicate foreign-born children who arrive before their first birthday (although they are again quite rare, at least in the case of Sweden). To make our simulations easier to interpret we ignore (i.e. exclude) these foreign-born infant arrivals here.*

Figure 1: Lexis diagrams to demonstrate age, period and cohort for siblings nested within families, including their age at arrival



*Note: Each of these three lexis diagrams represents a family (a group of siblings with the same mother). The vertical blue lines represent the year in which each mother arrived in Sweden. Each diagonal line represents the life course of a sibling (or a pair of siblings) born in a given year (when age=0). In **Family 1** (mother ID = 101), there is the same difference between each pair of siblings in terms of age at arrival (in years) as there is in terms of birth cohort (in years). The same is true for **Family 2** (mother ID = 102), which includes two twins who were born in the same year. The only families where this would not be the case are those with children born after arrival (**Family 3**, mother ID = 103) and those families where childhood migrants arrive in different years (who are very rare, at least in the case of Sweden). The uppermost blue line in each diagram represents age 16. For outcomes measured at this age, or any older age, the diagram shows that variation in age at arrival is collinear with variation in birth cohort when we compare foreign-born siblings within the same family.*

As highlighted by prior research,[16,28] studies that use sibling comparison designs often overlook the collinearity between their exposure and APC variables. Generally, in the presence of perfect collinearity – such as a model that includes age, period and cohort – statistical software will often return an error message or automatically omit one of the collinear variables (or part of one of the collinear variables) from the specified estimator. This happens when the software can identify that all parameters are not estimable. However, researchers cannot rely on the software to 'notice' this issue. In particular, it may be hidden by the operationalisation of APC variables. It is common practice for researchers to categorize the three APC components in unequal intervals, which circumvents but does not solve the identification issue in 'traditional' APC settings.[36] Similarly, we note that an unequal intervals approach may be used when controlling for APC variables in a sibling comparison design. For example, age at arrival may be measured in five-year categories, while birth cohort is measured in one-year intervals. Although this may allow researchers to estimate parameters for both age at arrival and birth cohort in a sibling comparison design, it does not solve the issue of identification, and it may well result in biased estimates and a misinterpretation of the (independent) effects of temporal exposures and confounders.[37] This issue is investigated below in our simulation of grouped factor controls.

As far as we are aware, there are only two situations where there is no linear dependency between age at arrival and birth cohort for siblings within the same family. The first is the situation where immigrant siblings have the same mother but arrived as children in different years. These siblings are discordant on both the exposure and birth cohort, but these two variables are not collinear. Considering the example in Table 1, we might imagine if sibling number 3 from Family 1 arrived separately from the rest of her family, say in 1990, rather than in 1980 (when her mother and siblings immigrated), thereby adjusting her arrival age to 13 (but not her birth cohort). We postpone a general appraisal of this type of sibling group (and related methodological issues) until the discussion, except to note that in this example (of childhood immigrants) they are very rare, at least in the case of Sweden (based on whole population data). For the sake of this motivating example, it is sufficient that they be ignored.

The second situation where age at arrival and birth cohort are not collinear is illustrated by Family 3 in Table 1 and Figure 1, where all siblings are born after their mother's migration. Scholars typically refer to this group as the second generation.[38–40] As such, they are concordant on age at arrival – i.e. they all have the same value (shown as 'n/a' in Table 1) – but at least some siblings in this family are discordant on birth cohort. As shown in Table 1 and Figure 1, the same is not true for Family 1 or Family 2, in which all siblings arrived in Sweden as children.

We note that the idea of including a group who are concordant on the exposure in sibling comparison designs has been proposed and implemented in prior empirical research.[33] However, we are not aware of any research that has tried to evaluate the utility of this approach more generally, including as compared with other means of estimating sibling comparison designs. There has been a recent debate about the relationship between the well-known identification issues in APC models and their relation to sibling comparison designs, including a discussion of whether it is even possible to mitigate the collinearity issues that we describe.[16,28] We build upon these studies by comparing the bias that arises when using a range of modelling approaches to control for APC variables in a sibling comparison design. Our study is more comprehensive than prior research because we compare a range of estimators – most of which have been used in isolation in prior studies – using a standardised (simulation) approach. In addition, we go beyond prior studies by evaluating an approach that enables birth cohort to be identified using sibling groups who are discordant on birth cohort but concordant on the exposure.

Research design

(1) Aims

Our broad objective is to understand how to control for birth cohort (or any APC variable) when exposures of interest are measured in units of time and potentially confounded by, or collinear with, birth cohort. Since we seek to provide recommendations for applied research, we might equally say that we wish to explore how best to specify models using SFEs, where 'best' is the approach that is least likely to lead to misleading research conclusions. Since researchers typically use SFEs as a means of obtaining unbiased (causal) estimates of the effects of an exposure on an outcome, we focus on bias as our principle measure of performance (see (5) below). More specifically, we have the following research questions:

A. *How should we control for birth cohort in sibling comparison designs when the exposure of interest is collinear with birth cohort:*
   1. *...and birth cohort is not a confounder?*
   2. *...and birth cohort is a confounder, but there is no age or period effect?*
   3. *...and birth cohort is a confounder and its effect varies by sex (as an example of the effect of birth cohort varying across groups)?*

B. *For each of the above, does bias arise when we estimate the effect of our exposure:*
   1. *...in absence of controls?*
   2. *...in absence of controls when the outcome is standardised?*
   3. *...after controlling for birth cohort using linear control?*
   4. *...after controlling for birth cohort using full factor controls?*
   5. *...after controlling for birth cohort using grouped factor controls?*

C. *How do our answers to these questions change if we include a group of siblings in our study population who are discordant on birth cohort but concordant on the exposure?*

D. *How do the estimates produced in answer to question C vary according to the proportion of siblings who are discordant on birth cohort but concordant on the exposure?*

(2) Data-generating process

Here, we use simulation based on parametric models. These models include the necessary complexity in order to answer the above research questions, but otherwise we try to keep them simple enough in order to make the simulation clear and intelligible to the widest range of researchers as possible. Our data generating processes are pure simulations, i.e. not based on real data, so that the coefficients underlying the data generation are known and hence can be used as a benchmark against which to compare estimates. We note that our conclusions will be driven by the presence or absence of bias, and not by the magnitude of such bias. Our choice is furthermore motivated by the fact that real world data is limiting because (real-world) coefficients underlying their generation are unknown.

We choose to analyse three different scenarios, corresponding to those described in research questions A1-A3. Each of these is modelled using a separate data generating process (DGP), as follows:

$$outcome = \delta.exposure + \alpha.siblingFE + \beta.X + \varepsilon \quad \text{(DGP1)}$$

$$outcome = \delta.exposure + \alpha.siblingFE + \beta.X + \gamma.cohort + \varepsilon \quad \text{(DGP2)}$$

$$outcome = \delta.exposure + \alpha.siblingFE + \beta.X + \gamma.cohort + \mu.(cohort * X) + \varepsilon \quad \text{(DGP3)}$$

Where $outcome$ is a continuous variable (representing years of education) and the $exposure$ (representing age at arrival, which is measured in discrete ages 0-18) has a true effect on this $outcome$ of δ, which is equal to 1 throughout all simulations and DGPs. All DGPs also include $siblingFE$ (a sibling fixed effect) with true effect α, a covariate vector X (in all simulations this includes only one binary covariate, representing sex) with true effect β, and a randomly distributed error term ε. DGP2 and DGP3 include $cohort$ (birth cohort, measured in discrete years) with true effect γ, and DGP3 includes an interaction between birth cohort and the covariate (sex) with true effect μ. In all DGPs, we only simulate sibling pairs for samples of the same size (in each scenario, see the appendix for more information on how this is done). However, we note that (on their own) neither the number of siblings within families nor the sample size will impact our conclusions.

For each of these DGPs, we then estimate our target parameter using five different estimators, corresponding to the research questions B1-B5. We therefore produce estimates for 15 different scenarios (one at a time) for which we can compare various performance measures. In order to answer research question C, we repeat this process to produce estimates for a further 15 scenarios with the inclusion of a group that are discordant on birth cohort but concordant on the exposure.

To be more specific, we initially use each of these three DGPs to generate a large (donor) pool of siblings, nested within two types of family: (1) families where all siblings were born abroad, therefore meaning that they are discordant on both age at arrival and cohort (DIS-DIS), but that these two variables are collinear conditional on the SFEs, and (2) families where all siblings were born after their mother's arrival, therefore meaning that they are discordant on birth cohort but concordant on age at arrival (DIS-CON), which in turn means that age at arrival and cohort are not collinear conditional on the SFEs. After creating this donor pool, we are then able to draw observations (or more specifically, sibling groups) at random, as well as varying the proportion of families that are DIS-DIS and DIS-CON.

(3) Estimands

We follow our motivating example and imagine that we wish to estimate the effect of age at arrival for immigrants who arrive as children (*exposure*) on their subsequent level of education at age 30 (*outcome*). In this example, age is held constant, thereby avoiding any of the well-known 'traditional' issues of APC modelling. The other variables in our DGPs are birth cohort (*cohort*) measured in years, a family identifier that is shared between siblings (*sibling FE*) and sex of each sibling (*confounder*), which we include primarily as a means of demonstrating a situation (in DGP3) where the true effect of birth cohort varies across individuals between and within families. The final term (e) is a randomly distributed error term that follows a normal (Gaussian) distribution. We note that our example can easily be translated to other research contexts by changing the variables that we use in our motivating example, notably the exposure (e.g. to maternal age) and the outcome (e.g. to child health). Since age is fixed, controlling for cohort is equivalent to controlling for period, and we discuss this generalisation further after the results.

(4) Methods

We compare five different estimators (which match the numbering of questions B1-B5), each of which can be used to estimate the effect of age at arrival using a model that includes SFEs. These are described below and can be summarised as:

Estimator 1. *No control*

Estimator 2. *Standardised outcome*

Estimator 3. *Linear control (continuous)*

Estimator 4. *Full factor controls (discrete)*

Estimator 5. *Grouped factor controls (discrete)*

As noted, each of these estimators have been used at least once before in prior applications of sibling comparison designs. However, they have not been compared in any prior research, and presently we lack an understanding of their ability to generate unbiased effect estimates in the scenarios discussed above. Estimator 1 estimates a sibling model with no control for birth cohort. This is a common initial step in applied research studying exposures that are collinear with calendar time (prior to controlling for birth cohort).[31] Given that DGP1 does not include a cohort effect, we expect Estimator 1 to produce an unbiased estimate of the effect of the exposure for this DGP. Estimator 2 estimates the same model as estimator 1, except to first standardise the outcome across birth cohorts. We do this using a z-transformation (i.e. so that the outcome has a mean of zero and a standard deviation of 1 within each birth cohort). This estimator has been recently proposed as a potential means of avoiding controlling for birth cohort explicitly, while also estimating effects net of the influence of cohort.[41] Although the authors do not make any explicit claims about the validity of the approach, we nevertheless include it here as a potential means of avoiding issues relating to linear dependency.

Our three other estimators include an explicit control for birth cohort. Estimator 3 includes a linear control (a single continuous variable) for birth cohort. Given that the true cohort effect in DGP2 is linear, we might expect estimator 3 to produce an unbiased estimate of the effect of the exposure for this DGP if there were no problems relating to collinearity. Estimators 4 and 5 use factor controls. In the most similar paper to ours (in terms of research aims), Kravdal compares the results of a series of simulations using sibling models with different specifications to identify the effect of maternal age on infant mortality.[16] He includes specifications with birth cohort operationalised as a set of dummy (factor) variables, either for all individual years (except one) or for years grouped together. The latter is equivalent to our estimator 4, controlling for birth cohort using grouped factor controls, while the former

is equivalent to our estimator 5, which controls for birth cohort using full factor controls (dummies for individual years).

As noted above, we not only apply these five estimators to each of the three DGPs (giving estimates for 15 different scenarios), but we also do this twice (resulting in 30 scenarios overall): once for families where all siblings are discordant on both age at arrival and cohort (DIS-DIS), and once where all siblings are discordant on birth cohort but concordant on age at arrival (DIS-CON). For each of these scenarios we run a simulation with 999 iterations (generating a unique set of simulated data for each iteration).

(5) Performance measures

Since researchers typically use SFEs as a means of obtaining unbiased (causal) estimates of the effects of an exposure on an outcome, we focus on bias as our principle measure of performance. We run our simulation across a fixed number of iterations (1,000 - unless otherwise specified) and store all estimates. We then take the average of the estimated effect of the exposure across all iterations and report this for each combination of DGP (x3), estimator (x5), and approach (x2).

Results

Our first two research questions [A and B] focus on fifteen scenarios (five estimators, each applied to three DGPs) where all sibling groups are discordant on the exposure and birth cohort. Results for these scenarios are shown in Table 2. Estimators 1, 3, 4 and 5 all estimate an effect that is very close to the true effect of 1 (that we chose for our simulations). For DGP1 we find no obvious evidence of bias, irrespective of the way that we control for birth cohort, and this is as expected because there is no confounding by cohort in DGP1. The results for DGP1 not only serve to suggest that the simulation is operating as expected, but they can also be used as a benchmark (of the true estimate), which is especially useful in the case of Estimator 2, which standardises the outcome. For this estimator, the true effect is no longer 1, but is best approximated by the estimate for DGP1 (which is 0.05 in Table 2).

As shown in Table 2, we note that when birth cohort is a confounder of the effect of age at arrival (as in DGP2 and DGP3), all of the estimators produce biased estimates for a sibling comparison of an exclusively DIS-DIS population. If anything, the bias is slightly larger, on average, for DGP3, suggesting that the amount of bias may be linked to an increasing complexity of the underlying confounding that relates to birth cohort.

Table 2: Study population all discordant on the exposure and birth cohort (DIS-DIS)

	Estimate		
Estimator	DGP1	DGP2	DGP3
Estimator 1: No control	1.00	2.00	2.06
Estimator 2: Standardised outcome	0.05	0.21	0.00
Estimator 3: Linear control (continuous)	1.00	2.00	2.50
Estimator 4: Full factor controls (discrete)	0.98	1.94	2.32
Estimator 5: Grouped factor controls (discrete)	1.00	2.00	2.50

*Note: The true effect is equal to 1 throughout, except in the case of **Estimator 2** (the standardised outcome) where the true effect is not equal to 1. This is because the outcome has been transformed to a z-score (conditional on cohort), and for this reason we take the best estimate of the true effect to be the estimate obtained in DGP1 (where there is no cohort effect).*

Table 3: Study population includes siblings discordant on birth cohort but concordant on the exposure (DIS-CON)

	Estimate		
Estimator	DGP1	DGP2	DGP3
Estimator 1: No control	1.00	2.05	2.47
Estimator 2: Standardised outcome	0.06	0.25	0.00
Estimator 3: Linear control (continuous)	1.00	1.00	0.99
Estimator 4: Full factor controls (discrete)	1.00	1.00	1.00
Estimator 5: Grouped factor controls (discrete)	1.00	1.95	2.43

*Note: The true effect is equal to 1 throughout, except in the case of **Estimator 2** (the standardised outcome) where the true effect is not equal to 1. This is because the outcome has been transformed to a z-score (conditional on cohort), and for this reason we take the best estimate of the true effect to be the estimate obtained in DGP1 (where there is no cohort effect).*

Table 4: Varying the proportion of siblings discordant on birth cohort but concordant on the exposure (DIS-CON)

	Proportion DIS-CON [1]				
Estimator	**0.05**	**0.15**	**0.5**	**0.85**	**0.95**
Estimator 1: No control	2.04	2.04	2.04	2.04	2.04
Estimator 2: Standardised outcome	0.06	0.07	0.07	0.06	0.05
Estimator 3: Linear control (continuous)	1.00	1.00	1.00	1.00	1.00
Estimator 4: Full factor controls (discrete)	1.00	1.00	1.00	1.00	1.00
Estimator 5: Grouped factor controls (discrete)	1.99	1.98	1.96	1.93	1.88

1 - The proportion of siblings discordant on birth cohort but concordant on the exposure

*Note: The remaining sibling groups are all discordant on the exposure and birth cohort (DIS-DIS). Results are for 999 iterations using the same specification as DGP2. The true effect is equal to 1 throughout, except in the case of **Estimator 2** (the standardised outcome) where our best estimate of the true effect is the estimate obtained in DGP1 in Table 3 (where there is no cohort effect).*

To answer question C, we repeat the same 15 scenarios, but this time including some sibling groups who are discordant on birth cohort but concordant on the exposure (DIS-CON). We do this in an attempt to control for the true confounding effect of birth cohort. The results from these scenarios are biased in some cases but not in others (see Table 3). As in Table 2, estimates are unbiased when birth cohort is not a confounder (in DGP1). However, when birth cohort is a confounder (in DGP2 and DGP3) then evidence of bias is highly dependent upon the estimator that is used. Our main finding is that there is limited evidence of bias when using linear or full factor controls (in Estimators 3 and 4).

Notably, the amount of bias when using grouped controls is the same as when not controlling at all for birth cohort. The standardised outcome approach also appears to generate biased estimates when cohort is a confounder. In terms of the amount of bias, comparisons are broadly similar for DGP2 – where confounding due to birth cohort is constant across groups – and DGP3 – where confounding due to birth cohort is different for women and men.

To answer our final research question [D], we examine variation in bias according to the proportion of sibling groups that are discordant on birth cohort but concordant on the exposure (DIS-CON). With the exception of the proportion DIS-CON, the DGP is the same as DGP2 in Table 3. Our results (in Table 4) shows that the proportion DIS-CON does not appear to make a substantial difference to the magnitude of bias. There is some variation in the case of grouped factor controls, although the estimates from this approach remain highly biased throughout.

Discussion

One of the main goals of sibling comparison designs is to obtain estimates of effects that are less biased than those that would be obtained from designs that do not control for confounding that is shared between siblings. However, as we have shown in detail, this goal can be undermined by controlling for birth cohort, at least when the exposure of interest is measured in units of time. Our objective was not only to provide applied researchers with a comprehensive understanding of this issue, and its implications for research, but also to demonstrate how researchers might control for birth cohort when exposures of interest are measured in units of time. We have compared a series of alternatives and found that all of these may produce biased results, unless a comparison group – who are discordant on birth cohort but concordant on the exposure – is included in the analysis.

In line with this finding, we therefore recommend that researchers who wish to control for birth cohort in sibling comparison designs consider the merits of this discordant-concordant approach. Indeed, we

would argue that this appears to be the only approach that has the potential of producing unbiased estimates when birth cohort is a confounder and it is collinear with the exposure within families (i.e. conditional on a sibling fixed effect). There are many studies where this is likely to be the case, not only the age at arrival example considered here, but also in other cases where the exposure is measured in units of time, such as age at parental divorce. In some cases, we believe that a discordant-concordant group is relatively easy to include – for example by including those sibling groups who do not experience a parental divorce and are therefore concordant on age at parental divorce.

However, for other exposures, it may not be possible to find a discordant-concordant group. In the case of maternal age, for example, the only sibling group that is concordant on the exposure would appear to be siblings born at the same time (e.g. twins). Given that twins are also concordant on birth cohort, we do not believe that they will help to control for birth cohort when using a sibling comparison design. The discordant-concordant approach is therefore not without limitations. In addition to the fact that it may be impossible to apply in some settings, it also requires additional assumptions, including that the birth cohort effect is the same for members of the discordant-concordant group (i.e. those who are used to identify the confounding effect of birth cohort) as it is for other sibling groups who are discordant on both birth cohort and the exposure. In some settings, this may be a strong assumption to make, but in others it may be quite reasonable. We note that this is not the same as general concerns about situations in which the cohort effect varies across groups, including when it is non-linear. Indeed, we found no strong evidence of additional bias when the true cohort effect varies across groups (i.e. via the interaction between cohort and our covariate – sex – in DGP3).

Importantly for the generalisability of our findings, we have focussed on birth cohort, but note that a similar set of issues would have occurred if we instead focussed on calendar period (often called 'period effects'). Indeed, the reason to control for birth cohort is often in order to control for differences due to the period (for example the year) in which a given (sub)population experience events (such as a recession). Throughout this study, we therefore note that controlling for birth cohort might equally well be generalised to controlling for calendar period, at least within the confines of the research designs that we have evaluated in our simulation. We found that if birth cohort is allowed to be a confounder (DGP2) and there is a discordant-concordant group in the data, it is important to adjust for birth cohort to avoid bias. Both linear controls and full factor controls (discrete) were sufficient to adjust for this confounding, while standardized outcomes and grouped factor controls were not. We note that the former two approaches treat the data as-is, whereas the latter two approaches transform the data thereby imposing potentially biasing constraints on the data before modelling. Hence, when the linear dependency is broken in the data by the presence of a discordant-concordant group, the lack of potentially biasing

constraints on the data is a preferred condition. We therefore suggest that other approaches to the APC linear dependency problem that similarly place their constraints on the model, rather than by transforming the data, are likewise more likely to provide unbiased results in this setting – however, we stress that this would need to be investigated first, for example through a simulation study such as ours. We note that while many new advances have been made in recent years to address the age-period-cohort linear dependency problem[37,42–45], our study does not seek to provide new estimators to 'solve' this issue and limits itself to the specific issue of linear dependency in sibling comparison designs.

Our study design focuses on the most prominent research contexts in which researchers may wish to use a sibling comparison design and control for birth cohort. That said, future methodological research may generate additional insights by examining more complex designs, including those that seek to study outcomes at more than one point in time. In our example, we focussed on education at age 30, but we might imagine a situation where researchers wish to focus on outcomes at different ages – e.g. educational trajectories from ages 18-30, which would increase the potential for collinearity issues, as discussed in traditional APC models.[43,46–48] In this case, we expect that controlling for birth cohort would be an issue when researchers also wish to control for either age or period – or more generally, that the birth cohort collinearity we describe here for sibling comparison designs would arise when seeking to control for any two of the dimensions of APC. We note that controlling for two APC variables would not normally lead to collinearity issues in a setting without a sibling comparison.[43,46–48]

Future research may also consider how to control for birth cohort in sibling comparison designs that use random effects. Such an approach has been proposed by Kravdal,[16] and might also be extended to include a discordant-concordant group, perhaps with enabling some of the benefits of random effects (as compared with fixed effects) such as the ability to investigate the role of factors that are shared between siblings. The use of between-within models may also be worthwhile investigating for similar reasons. At the same time, we note that our approach here has been to evaluate statistical methods using simulation. It may be beneficial for future research to try to prove our recommendations algebraically, thereby providing a more generalisable statistical foundation for our results. However, this may not be a straightforward task, given the inclusion of sibling fixed effects. This explains why we chose our simulation approach, which also has the benefit of being more intuitive and accessible, in particular for applied researchers. As such, we conclude by hoping that our results provide a wider audience with a greater understanding of an important methodological issue when using sibling comparison designs, namely how to control for age, period or cohort when studying exposures that are measured in units of time.

Acknowledgements

BW's contribution to this research was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant agreement No. 948727 (REFU-GEN), in addition to funding received from the Swedish Research Council (Vetenskapsrådet) via project 2017-01021, financial support from the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond, RJ) grant registration number M18-0214:1, and financial support from the Swedish Research Council for Health, Working life and Welfare (FORTE) grant numbers: 2016-07105, 2016-07115, 2018-00310 and 2019-00603.

SA acknowledges funding from CRITEVENTS. The project CRITEVENTS is financially supported by the NORFACE Joint Research Programme on Dynamics of Inequality Across the Life-course, which is co-funded by the European Commission through Horizon 2020 under grant agreement No 724363.

AG's contribution to this research was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant agreement No. 803959 (MARTE).

References

1. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986;15(3):413-419.

2. Kaufman JS. Commentary: Causal Inference for Social Exposures. *Annu Rev Public Health*. 2019;40(1):7-21. doi:10.1146/annurev-publhealth-040218-043735

3. Donovan SJ, Susser E. Commentary: Advent of sibling designs. *Int J Epidemiol*. 2011;40(2):345-349. doi:10.1093/ije/dyr057

4. Susser E, Eide MG, Begg M. Invited Commentary: The Use of Sibship Studies to Detect Familial Confounding. *Am J Epidemiol*. 2010;172(5):537-539. doi:10.1093/aje/kwq196

5. Record RG, McKEOWN T, Edwards JH. The relation of measured intelligence to birth weight and duration of gestation*. *Annals of Human Genetics*. 1969;33(1):71-79. doi:10.1111/j.1469-1809.1969.tb01631.x

6. Record RG, McKEOWN T, Edwards JH. An investigation of the difference in measured intelligence between twins and single births*. *Annals of Human Genetics*. 1970;34(1):11-20. doi:10.1111/j.1469-1809.1970.tb00215.x

7. Gorseline DE. *The Effect of Schooling upon Income*. Indiana University; 1932.

8. Zvi Griliches. Sibling Models and Data in Economics: Beginnings of a Survey. *Journal of Political Economy*. 1979;87(5):S37.

9. Sigle-Rushton W, Lyngstad TH, Andersen PL, Kravdal Ø. Proceed With Caution? Parents' Union Dissolution and Children's Educational Achievement. *Journal of Marriage and Family*. 2014;76(1):161-174. doi:10.1111/jomf.12075

10. Grätz M. Competition in the Family: Inequality between Siblings and the Intergenerational Transmission of Educational Advantage. *Sociological Science*. 2018;5:246-269. doi:10.15195/v5.a11

11. Grätz M. When Growing Up Without a Parent Does Not Hurt: Parental Separation and the Compensatory Effect of Social Origin. *European Sociological Review*. 2015;(5):546.

12. Härkönen J. Birth Order Effects on Educational Attainment and Educational Transitions in West Germany. *European Sociological Review*. 2014;30(2):166-179.

13. Hällsten M. Is Education a Risky Investment? The Scarring Effect of University Dropout in Sweden. *European Sociological Review*. 2017;33(2):169-181. doi:10.1093/esr/jcw053

14. Cordero-Coma J, Esping-Andersen G. Parental time dedication and children's education. An analysis of West Germany. *Research in Social Stratification and Mobility*. 2018;55:1-12. doi:10.1016/j.rssm.2018.03.006

15. Barclay KJ, Kolk M. The Long-Term Cognitive and Socioeconomic Consequences of Birth Intervals: A Within-Family Sibling Comparison Using Swedish Register Data. *Demography*. Published online February 13, 2017. doi:10.1007/s13524-017-0550-x

16. Kravdal Ø. Taking birth year into account when analysing effects of maternal age on child health and other outcomes: The value of a multilevel-multiprocess model compared to a sibling model. *DemRes*. 2019;40:1249-1290. doi:10.4054/DemRes.2019.40.43

17. Hermansen AS. Age at Arrival and Life Chances Among Childhood Immigrants. *Demography*. 2017;54(1):201-229. doi:10.1007/s13524-016-0535-1

18. Laidley T, Domingue B, Sinsub P, Harris KM, Conley D. New Evidence of Skin Color Bias and Health Outcomes Using Sibling Difference Models: A Research Note. *Demography*. 2019;56(2):753-762. doi:10.1007/s13524-018-0756-6

19. Böhlmark A. Age at immigration and school performance: A siblings analysis using swedish register data. *Labour Economics*. 2008;15(6):1366-1387. doi:10.1016/j.labeco.2007.12.004

20. Åslund O, Böhlmark A, Skans ON. Childhood and family experiences and the social integration of young migrants. *Labour Economics*. 2015;35:135-144. doi:10.1016/j.labeco.2015.05.004

21. Richardson EA, Shortt NK, Mitchell R, Pearce J. A sibling study of whether maternal exposure to different types of natural space is related to birthweight. *Int J Epidemiol*. 2018;47(1):146-155. doi:10.1093/ije/dyx258

22. Goisis A, Remes H, Barclay K, Martikainen P, Myrskylä M. Advanced Maternal Age and the Risk of Low Birth Weight and Preterm Delivery: a Within-Family Analysis Using Finnish Population Registers. *American Journal of Epidemiology*. 2017;186(11):1219-1226. doi:10.1093/aje/kwx177

23. Frisell T, Öberg S, Kuja-Halkola R, Sjölander A. Sibling Comparison Designs: Bias From Non-Shared Confounders and Measurement Error. *Epidemiology*. 2012;23(5):713-720. doi:10.1097/EDE.0b013e31825fa230

24. Sjölander A, Zetterqvist J. Confounders, Mediators, or Colliders: What Types of Shared Covariates Does a Sibling Comparison Design Control For? *Epidemiology*. 2017;28(4):540-547. doi:10.1097/EDE.0000000000000649

25. Sjölander A, Frisell T, Kuja-Halkola R, Öberg S, Zetterqvist J. Carryover Effects in Sibling Comparison Designs: *Epidemiology*. 2016;27(6):852-858. doi:10.1097/EDE.0000000000000541

26. Gilman SE, Loucks EB. Another casualty of sibling fixed-effects analysis of education and health: An informative null, or null information? *Soc Sci Med*. 2014;118:191-193. doi:10.1016/j.socscimed.2014.06.029

27. Smith GD. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *Int J Epidemiol*. 2011;40(3):537-562. doi:10.1093/ije/dyr117

28. Keiding N, Andersen PK. Comment by Keiding and Andersen on Barclay and Myrskylä (2016): Linear effects of maternal age and period trends cannot be distinguished. Published online 2016. Accessed February 7, 2020. https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fj.1728-4457.2016.00105.x&file=padr105-sup-0003-SuppMat.pdf

29. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 2019;38(11):2074-2102. doi:10.1002/sim.8086

30. Skrondal A. Design and Analysis of Monte Carlo Experiments: Attacking the Conventional Wisdom. *Multivariate Behavioral Research*. 2000;35(2):137-167. doi:10.1207/S15327906MBR3502_1

31. Barclay K, Myrskylä M. Advanced Maternal Age and Offspring Outcomes: Reproductive Aging and Counterbalancing Period Trends. *Population and Development Review*. 2016;42(1):69-94. doi:10.1111/j.1728-4457.2016.00105.x

32. Goisis A, Remes H, Barclay K, Martikainen P, Myrskylä M. Paternal age and the risk of low birth weight and preterm delivery: a Finnish register-based study. *J Epidemiol Community Health*. 2018;72(12):1104-1109. doi:10.1136/jech-2017-210170

33. Kravdal Ø, Grundy E. Children's age at parental divorce and depression in early and mid-adulthood. *Population Studies*. 2019;73(1):37-56. doi:10.1080/00324728.2018.1549747

34. Miller R, Karra M. Birth Spacing and Child Health Trajectories. *Population and Development Review*. 2020;46(2):347-371. doi:https://doi.org/10.1111/padr.12335

35. Mehta NK, Martikainen P, Cederström A. Age at Immigration, Generational Status, and Death Among Children of Immigrant Mothers: A Longitudinal Analysis of Siblings. *American Journal of Epidemiology*. 2019;188(7):1237-1244. doi:10.1093/aje/kwz055

36. Holford TR. Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in Medicine*. 2006;25(6):977-993. doi:https://doi.org/10.1002/sim.2253

37. Luo L, Hodges JS. Block Constraints in Age–Period–Cohort Models with Unequal-width Intervals. *Sociological Methods & Research*. 2016;45(4):700-726. doi:10.1177/0049124115585359

38. Guillot M, Khlat M, Wallace M. Adult mortality among second-generation immigrants in France: Results from a nationally representative record linkage study. *Demographic Research*. 2019;40(54):1603-1644. doi:10.4054/DemRes.2019.40.54

39. Scott K, Stanfors M. The transition to parenthood among the second generation: Evidence from Sweden, 1990–2005. *Advances in Life Course Research*. 2011;16(4):190-204. doi:10.1016/j.alcr.2011.09.003

40. Drouhot LG, Nee V. Assimilation and the Second Generation in Europe and America: Blending and Segregating Social Dynamics Between Immigrants and Natives. *Annual Review of Sociology*. 2019;45(1):null. doi:10.1146/annurev-soc-073117-041335

41. Aradhya S, Härkönen J, Boertien D, Bernardi F. Heterogeneous effects of parental separation on the educational outcomes of children. In: *Giornate Di Studio Sulla Popolazione 2019*. ; 2019. https://www.sis-aisp.it/ocs-2.3.4/index.php/popdays2019/popdays2019/paper/view/941

42. Fosse E, Winship C. Bounding Analyses of Age-Period-Cohort Effects. *Demography*. 2019;56(5):1975. doi:10.1007/s13524-019-00801-6

43. Carstensen B. Age–period–cohort models for the Lexis diagram. *Statistics in Medicine*. 2007;26(15):3018-3045. doi:10.1002/sim.2764

44. O'Brien RM, Hudson K, Stockard J. A Mixed Model Estimation of Age, Period, and Cohort Effects. *Sociological Methods & Research*. 2008;36(3):402-428. doi:10.1177/0049124106290392

45. Winship C, Harding DJ. A Mechanism-Based Approach to the Identification of Age-Period-Cohort Models. *Sociological Methods & Research*. 2008;36(3):362-401. doi:10.1177/0049124107310635

46. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: age–period–cohort models. *Statistics in medicine*. 1987;6(4):469-481.

47. Goldstein H. Age, Period And Cohort Effects — A Confounded Confusion. *Journal of Applied Statistics*. 1979;6(1):19-24. doi:10.1080/02664767900000003

48. Hobcraft J, Menken J, Preston S. Age, period, and cohort effects in demography: a review. *Popul Index*. 1982;48(1):4-43.

Appendix: Further details on the simulations

There are several different approaches that could be used to generate samples of the same size for each of the above scenarios (in each iteration). Even though we specify the number of families, and the number of siblings within each family, as inputs that are the same for each DGP, there are several constraints (i.e. data deletions) that we need to apply to the DGPs, which means that the generated data in the initial donor pool varies in size. Some of these constraints are the same across scenarios, in particular the need to drop children arriving over 18-years-old. (Note that if we applied this constraint in the initial DGP then it has the potential to alter the effect of the exposure because of the link between late arrival and being born in an older cohort.) Other constraints vary across scenarios, such as the constraint to drop DIS-CON families from (the first 15) scenarios that are meant to exclude them. The combined impact of these constraints leads to variation in the number of families and siblings (in the initial donor pool) across different scenarios. To maximise comparability we therefore take a stratified sample of 10,000 siblings from each donor pool. To do this, we first check the proportion of families of each size in the data, and then we randomly select those families within sibling size strata but keep the percentage in each stratum stable. That ensures that the sample size is stable across iterations.

Stockholm Research Reports in Demography

Stockholm University,
106 91 Stockholm,
Sweden
www.su.se | info@su.se | ISSN 2002-617X

Stockholm University

Demography Unit