

Editorial on the Special Issue “The Identification of Causal Mechanisms in Demographic Research”

Johannes Huinink, Josef Brüderl

1 Introduction

Explaining demographic behaviour and population change means identifying the causal mechanisms which “drive” them over time. Based on theoretical modelling and guided by empirical findings in prior studies, demographic and social research pursues the improvement of knowledge about those mechanisms and the relationships between the involved factors.

In demography, as in the social sciences in general, theoretical and methodological advancements over the past 50 years have greatly contributed to accomplishing this goal. Methods of longitudinal data collection as well as individual- and multi-level longitudinal data analysis have gained relevance. This trend was paralleled by the development of the life course perspective in the social sciences and conceptual refinements in cohort analysis in demographic research.

Meanwhile, collecting and analysing longitudinal data is a standard procedure in individual- and multi-level demographic research. Many studies using this methodological inventory have been conducted, enriching our knowledge on individual decision-making and behaviour considerably. Compared to cross-sectional data, longitudinal data significantly improve the conditions for identifying the “true” effects of underlying causal mechanisms. While retrospective information is already of great use, prospective panel designs enable a more appropriate and manifold collection of relevant information, as well as more refined statistical modelling of the interdependence between individual behaviour, its dispositional and motivational drivers, its situational conditions, and its outcomes over time. Panel data are also useful for another prominent class of methods, i. e. techniques of event history analysis (*Blossfeld/Rohwer 2002; Kreyenfeld 2021*).

In this Special Issue of Comparative Population Studies, we review the degree to which methodological innovations in panel studies have been useful in properly identifying causal mechanisms in the study of demographic behaviour, and ultimately population change. In the first contribution, methodological issues of panel data analysis are discussed and illustrated by the example of estimating the effect of motherhood on life satisfaction. The next four articles address the core question of the Special Issue with regard to major fields of demographic research:

nuptiality and partnership formation, fertility, spatial mobility and migration, and morbidity and mortality. The final two articles present illustrative empirical studies, both using data from the German Family Panel Study (pairfam). In doing so, they demonstrate the opportunities as well as the apparent limitations of longitudinal data analysis in identifying causal effects in social research.

2 Causal Mechanisms

What are the key causal mechanisms of demographic behaviour? There is an extensive literature on causal mechanisms across all scientific disciplines in general, and in the social sciences in particular. In the social sciences, the term “social mechanisms” is also used at times. We will not review this broad literature on causal and social mechanisms in detail here. On the social sciences in general, see for example *Hedström* and *Ylikoshi* (2010) or *Mayntz* (2020). Recently, *Wunsch* and *Gourbin* have also addressed the issue in their article on causal assessment in demography (*Wunsch/Gourbin* 2020). This literature proposed various definitions of the concept of causal mechanisms. In the following, we introduce and elaborate upon our own definition, which shall be simple but instructive for our introduction into the topic of this Special Issue.

A causal mechanism M (in its simple, mono-causal version) is a process by which a change in an attribute of an entity A , denoted by variable X (causing factor), at time point t_1 or during an interval $[t_0, t_1]$ induces a change in another attribute of A and/or an attribute of another entity B , denoted by variable Y (outcome), at time point t_2 or during a time interval $[t_2, t_3]$, with $t_3 > t_2 > t_1 > t_0$.

According to this definition, the change in X , say $\Delta x_A(t_1)$ or $\Delta x_A([t_0, t_1])$, precedes the change in Y in time: $\Delta y_A(t_2) = f(\Delta x_A(t_1))$ or $\Delta y_A([t_2, t_3]) = f(\Delta x_A([t_0, t_1]))$. $t_2 - t_1 > 0$ is the length of the time gap the mechanism “needs” to take effect. x_A and y_A denote realisations of the variables X and Y for an entity A . If the outcome occurs for an entity B , the realisation would be y_B respectively.¹ The size of the causal effect, i.e. the size of the change in Y (outcome), principally varies with the size of the change in X as well as with the time elapsed since the change in X according to the function f , which usually can be expressed mathematically.

A causal mechanism as defined here usually cannot be observed directly. One only realizes observable manifestations of the mechanism and follows the process empirically more or less closely over time. Based on theoretical modelling, which is always preliminary, as well as depending on the quality of applied methods of observation and the data analysis, we can “identify” the causal mechanism, i.e. the causal effect of X on Y , only in a more or less valid way. For this reason, causal

¹ Even though X and Y may change continually over time, it is assumed that the condition that a cause temporally precedes an effect always holds, although by a potentially extremely small period of time. For instance, it can be the case that the size of a change in X during $[t_0, t_1]$ must surpass a certain threshold at t_1 before a causal effect on Y is induced at t_2 or during $[t_2, t_3]$.

mechanisms are often conceptualized probabilistically. Because we are not able to identify all relevant circumstances under which the mechanism is at work, we must deal with uncertainty. X and Y are treated as random variables, and usually only the effects of a change in X on the (mean of a) probability distribution of a change in Y can be estimated (*Wunsch/Gourbin 2020: 3*).

In principle, according to this understanding of a causal mechanism, the collection of longitudinal data is a pre-emptory condition for tracing the “real” processes of the causal mechanism as it evolves over time. This means observing changes in properly measured variables X and Y , and the extent to which a change in X effects a change in Y over a certain period of time. Ideally, the data collection also should cover the time span across which the causal mechanism exerts its effect as precisely as possible. Therefore, the time schedule and intervals of data collection should be fine-grained enough to appropriately capture changes in X and Y . This also helps avoiding collecting synchronous information.

In the social sciences and demography, we usually do not study simple mono-causal versions of a causal mechanism, but rather much more complex social processes – so-called “compound mechanisms”. Compound mechanisms are composed of more than one mono-causal mechanism. Changes in Y are induced by changes in causal factors $X_1, \dots, X_n, n > 1$. We extend the definition presented above and specify compound mechanisms with the formal expression: $\Delta y = f(\Delta x_1, \dots, \Delta x_n)$, given that n causal factors are involved which also may be attributes of more than just one entity A .²

The included mechanisms might work in parallel or in sequence. In the first case (mechanisms working in parallel), changes in causal factors X_i and X_j ($i, j = 1, \dots, n, i \neq j$) independently exert a “direct” effect on Y . The second case (mechanisms working in sequence) covers two sub-cases. In the first, a change in one causal factor X_i induces a change in another causal factor X_j , which then induces a change in Y : $\Delta x_j = f(\Delta x_i)$ and $\Delta y = g(\Delta x_j)$. In this case, the causal effect of X_i on Y is “indirect”, i.e. exerts an effect on Y via X_j . In the second, a change in X_j simultaneously induces a change in X_i and Y : $\Delta x_i = f(\Delta x_j)$ and $\Delta y = g(\Delta x_j)$. Here, X_j is a so-called confounder producing a (spurious) correlation between changes in X_i and Y which is not the consequence of a causal mechanism connecting X_i and Y . Usually, combinations of these different constellations occur in empirical analyses.

Compound mechanisms might also be composed of mono-causal mechanisms working on different process levels. In such a case, X_1, \dots, X_n might stand for attributes of entities on different levels of aggregation. We will mainly distinguish between inner-individual levels, extra-individual levels, and the level of the acting individuals themselves.

Finally, we must consider that the outcomes of causal mechanisms depend on the situational conditions they are working in. These conditions are constant during the time period the mechanisms need to take effect and can be specified

² For the sake of simplicity, the equations in the following are written without time- and entity-related indices.

by variables Z_1, \dots, Z_m , $m \geq 1$. They can restrict or support the mechanisms in their functioning. One can specify: $\Delta y = f(\Delta x_1, \dots, \Delta x_n \mid z_1, \dots, z_m)$. The relatively stable conditions Z_1, \dots, Z_m may also be observed on different process levels.³

Some important conclusions can be drawn from these considerations with regard to the empirical analyses of complex social processes we deal with in demographic research.

First, we usually must consider more than one relevant causal factor, i.e. X_1, \dots, X_n , $n > 1$, which themselves might affect each other through known or unknown causal mechanisms. If we are interested in the direct causal effect of a single potential causal factor X_i on Y – i.e. its “true” direct causal contribution to the change in Y , we must isolate the effect of Δx_i on Δy . This means that we must properly account for mediating and for confounding effects of other causal factors. The total causal effect of X_i on Y is then estimated by the sum of its direct and indirect effects.

Second, reverse causality presents a severe complication we may encounter when studying compound mechanisms. There is a mechanism by which a change in factor X_i has an effect on Y , say $\Delta y(t_2) = f(\Delta x_i(t_1))$, but after that, by another mechanism, $\Delta y(t_2)$ affects a further change in X : $\Delta x_i(t_3) = g(\Delta y(t_2))$, and so forth. This phenomenon, which is common in demographic processes, requires particularly careful observation of the process under study. For a detailed discussion see an article written by *Brüderl and Ludwig (2015)*.

Third, some causal factors at play might be unobserved due to not being measurable or due to poor or incomplete theory. This unobserved heterogeneity with respect to confounding or mediating attributes might bias estimates of the direct causal effect of a factor X_i . In order to identify the “true” impact of X_i , this issue must be properly addressed.

Fourth, the causal mechanisms at play might not work in an additive way. In other words, the effects of causal factors X_{i1} and X_{i2} on the outcome variable Y may not be independent of one another: factor X_{i1} might modify the effect of factor X_{i2} on the outcome Y , i.e. $\Delta y = f(\Delta x_{i1} \mid x_{i2})$, and/or vice versa. This means that X_{i1} and X_{i2} interact with or moderate (condition) each other in their functioning according to the respective mechanism. In the empirical model of estimating the effects of causal factors, such interaction effects must be considered.

Fifth, in social research often processes on higher levels of aggregation, such as the change in the size and structure of a population are investigated. Mechanisms working on more than one process level are involved in determining such outcomes. A reductionist programme examines that process level which allows us to identify the causal mechanisms which in combination are responsible for those changes in the units at higher process levels we are interested in. For instance, the approach of methodological individualism in the social sciences demands analyses on the

³ In the end, the impact of Z_j also is exerted by a causal mechanism as defined above. The moderating force of Z_j is assumed to be driven by the fact that it is constant during the period of time in which the changing X_i evolves its impact. However, Z_j can also change. In such a case, the change in Z_j leads to a change in the effect of X_i and a different outcome Y .

level of individual actors, assuming that their behaviour drives observed social dynamics on the aggregate level of a society's population. However, it might even be necessary to study underlying inner-individual, e.g., physiological, psychological, or neurological processes which eventually cause the observed behaviour.⁴ Modelling multi-level regression is a way of including factors from different aggregation levels simultaneously and for allowing them to interact with each other.

3 Challenges of causal analysis in demography

Over recent decades, empirical demographic research has changed considerably. It is increasingly acknowledged that a change in the size and structure of a population is ultimately the outcome of the demographically relevant activities and events (such as household formation, fertility, and mortality) which are conducted and experienced by its members and by individuals leaving or entering the population due to migration. This means that demographic behaviour, as we can call it, can be studied as part of the individual life course. In our terms, life courses are complex processes driven by compound causal mechanisms at various process levels (*Bernardi et al.* 2019). Following a life course perspective, one can identify typical issues in empirical demographic research, in accordance with the general considerations in the previous section.

First, life courses have a past and future which both matter for demographic behaviour. Regarding the past, the demographic behaviour of individuals is influenced by life experiences in various ways. They shape personal dispositions affecting individual decision-making (*Aizen/Klobas* 2014) as well as the capabilities and opportunities of future life conduct (*Dannefer* 1987; *DiPrete/Eirich* 2006). This leads to (self-) selection issues in the analysis of life course dynamics over time (*Brüderl/Ludwig* 2015). *Gattig* and *Minkus* deal with one such example in this Special Issue, analysing the interplay of life satisfaction and marriage among cohabitating couples (*Gattig/Minkus* 2021). Greater life satisfaction supports the intention to marry, and therefore married couples are the happier ones on average.

Regarding the future, the life course perspective considers individuals' expectations regarding the pros and cons of future activities and life plans. The anticipation of possible outcomes plays a major role in individual reasoning on costs and benefits of alternative activities and respective decision making. Studying the effects of anticipation is a particular challenge in causal analysis of demographic

⁴ Choosing the proper level of analysis is a question of theory. Emergent phenomena on a systems level (global attributes) as an outcome of a strong interaction between the elements of a system might be studied without going below the systems level. However, this can lead to issues in causal explanation. In physics for instance, it is relevant to know that the positive relationship between the temperature and pressure of a constant amount and volume of gas is caused by the fact that increasing temperature leads to increasing velocity of colliding gas-particles and the higher kinetic energy leads to the increase in pressure.

behaviour because it is hard to measure in a valid manner (*Kreyenfeld 2021; Hoem/Kreyenfeld 2006*).

A special case of self-selection into entire patterns of future life course trajectories is what *Huinink* and *Feldhaus* discussed as a longitudinal version of spurious correlation (*Huinink/Feldhaus 2009: 314*). Early life events or experiences might cause a synchronisation or harmonisation of trajectories in different dimensions of the further life course. This means that the correspondence between events in different life domains is not (only) driven by mechanisms working in the later life course. Instead, some outcomes may be predetermined by life scripts internalized by individuals early on. For example, being raised in a social environment favouring traditional gender norms may make individuals more likely to follow a comprehensive script prescribing how the work and family lives of men and women are to be combined in the future life course; e.g. with women devoting their time to the family and accordingly restricting other activities.

At the same time, deviance from life scripts due to unexpected events or unforeseen obstacles to life plans might initiate adaptation and coping mechanisms regarding the new situation (*Heckhausen/Buchmann 2019*). For instance, getting married or having a child can induce changes in individual traits and attitudes which otherwise would not have occurred. Attitudes regarding gender roles may traditionalise after the birth of a child based on the new household circumstances. Marked changes such as these may lead to shifts in individual values in order to reduce the uncomfortable consequences of cognitive dissonances which might otherwise occur (*Lesthaeghe 2002; Moors 2001*).

Second, the life course is multi-dimensional. Developments in different life domains or dimensions, such as education, work, family, and health, affect each other in various and usually complex ways. Among others, the question of reverse causality must be emphasised. In demographic research, the relationship between educational and occupational development and family formation processes is analysed particularly intensively in this respect. Another example in which reverse causality is very hard to disentangle is the interdependence between health status and family development, as well as social mobility, over the life course (*Hoffmann/Doblhammer 2021*).

Third, the life course is embedded in a multi-level structure of processes. Studying demographic change means analysing and revealing the consequences of individual demographic behaviour for the change of the size and structure of an overall population. Conversely, individual life courses are influenced by their natural and social environments. These include social dyads, intimate relationships, and wider social networks, as well as socio-economic structures or cultural and institutional environments. These environmental conditions limit the variety and affect the cost-benefit balance of options available to individuals. However, environmental conditions unfold their impact on the individual behaviour mainly through individuals' subjective perceptions and (re-) cognition. Overarching conditions take effect via the mentioned mental and emotional mechanisms, including the cognitive processes of evaluating the costs and benefits of an action.

This leads us to the relevance of intra-individual processes, such as the biological and social ontogenesis which is based on biological dispositions as well as on social experiences, particularly in early phases of the life course. These mental states, including values, attitudes, and convictions may stay quite stable and thereby condition individual intentions and behaviour (*Ajzen/Klobas* 2014). Studying these processes is not primarily the business of demographers, but the social sciences in general should devote more attention to these fields of the study of human nature.

4 Longitudinal analysis of causal mechanisms and the advantages of panel data

Studying demographic behaviour as a source of demographic change means investigating compound mechanisms. As outlined in the previous sections, the identification of such causal mechanisms through empirical analysis is quite a challenge. For comprehensive treatments of the issues of causal inference in the social sciences, see *Morgan and Winship* (2015) or *Gangl* (2010). A recent overview on methods dealing with these issues in the context of demographic research has been published by *Wunsch and Gourbin* (2020). For the case of dichotomous treatments, such as the impact of life events on the attributes of individuals, see *Ludwig and Brüderl* (2021) in this Special Issue. We do not go into greater detail on these topics here, but rather present some insights on how panel data analysis is useful in the pursuit of identifying causal effects in compound mechanisms.

Randomised experiments are considered the “gold standard” for causal analysis. In principle, experimental data provide outcome data from before and after the treatment. A group of subjects is (randomly) drawn from a population of interest. Members of a randomly selected subgroup (treatment group TG) are exposed to a change in one single factor X (treatment) which theoretically should have a causal effect on an outcome variable Y. Y is measured before and after the treatment and $\Delta y(\text{TG})$ is calculated. This is compared to $\Delta y(\text{CG})$, estimated for the members of a control group (CG) which was not exposed to the change in X. An unbiased estimate of the average treatment effect is thus gained, under the assumptions that randomisation worked and that no confounding occurred. In this case, the unknown counterfactuals in the treatment group can be substituted by the observed outcomes in the control group.

The ability to use an experimental design with a purposefully implemented treatment is limited in demographic research due to practical and ethical considerations. Therefore, observational studies remain most common. With observational data, we must almost always deal with selective (i.e., non-random) treatment assignment and confounding. Observed confounders can be conditioned on, for instance by introducing them as controls in a regression model. However, there are very often plausible arguments that important unobserved confounders may exist. In such cases, the conditioning strategy will provide biased estimates.

As outlined above, longitudinal data are helpful when identifying causal mechanisms. They can be collected retrospectively, by panel surveys, or by

different ways of process tracing. Retrospective data are collected in a one-shot survey by collecting information on past processes or, in the case of individual life courses, past experiences, which the respondents report. If we collect (or register) information on attributes of individuals (or other units of analysis) at a sequence of time points using the same measurement instruments, we produce panel data. This approach allows the researcher to follow life course trajectories more closely and reliably. Typically, retrospective data are sufficiently valid only in the case of structural-objective information such as dates of life course events and social status indicators. Meanwhile, panel surveys aim to gather subjective information such as life satisfaction and attitudes on behavioural issues and their change over time. We also more reliably observe the temporal order of events and activities over the life course properly. Therefore, panel data can provide valid information on individual decision-making and behaviour over time. There are also disadvantages of panel surveys compared to retrospective data collection: Retrospective surveys are less expensive and less time-consuming than panel surveys and the problem of sample attrition is less serious.

Panel data enables significant improvements in causal analysis. In some cases, one can use so-called natural or quasi-experimental designs. In such cases, data collected by process tracing or (quasi-) panel designs allow the researcher to estimate effects of certain events, e.g. political interventions, on the outcomes of individuals who experienced the event or realised the interventions (e.g. *Gangl/Ziefle 2015; Hofmann et al. 2017; Kreyenfeld 2021*). However, non-random selection of cases into the treatment group must be examined in close detail and with in-depth substantive knowledge of the treatment assignment mechanism. As a complementary approach, some authors propose simulating counterfactuals by propensity score matching. By this procedure it shall be assured that the treatment effect of X can be estimated as an average treatment effect if one accounts for the dependence of the probability of experiencing the treatment given relevant other attributes of the units of analysis (*Morgan/Winship 2015; Gangl 2010*).

Panel data, as longitudinal data in general, can also be analysed with event history analysis methods. The outcome variable Y must be discrete in this case. Potential causal factors can be traced over time (time-dependent covariates) and effects on incident rates of a (later) shift in Y can be estimated. To study causal mechanisms inducing (or affecting the likelihood of) the occurrence of a certain life event, a rich toolbox of methods of event history analysis has been developed (*Blossfeld/Rohwer 2002; Kleiding 2014*).

A major advantage of panel data is that the influence of non-measured time-invariant attributes (unobserved heterogeneity), i.e. variables Z_j as defined above, can be accounted for (“differenced out”) in fixed-effects (FE) models. Meanwhile, more advanced models have been developed which have relaxed the assumption of time independence of the unobserved heterogeneity (*Gattig/Minkus 2021*). Nonetheless, FE models still present many issues or potential pitfalls which must be considered, as *Ludwig* and *Brüderl* discuss in their contribution in this Special Issue (*Ludwig/Brüderl 2021*).

Panel data also allow researchers to analyse how the effects of causal mechanisms evolve over time. As mentioned, the impact of a certain treatment at a certain point in time t_1 , i.e., $\Delta x(t_1)$, on an outcome, i.e. $\Delta y(t_2)$, $t_2 > t_1$, in many cases will be a function of the elapsed time since t_1 (Ludwig/Brüderl 2021). Here, one can also refer to growth curve modelling, which can be understood as a type of panel analysis (Brüderl et al. 2019).

Panel data, at least in principle, enable the analysis of processes of recursive causation in models with lagged dependent variables or multi-equation event history analysis (Allison et al. 2017; Leszczensky/Wolbring 2019; Mund/Nestler 2019; Lillard 1993). We mentioned the interdependence between attitudes and behaviour over the life course. However, these methods are still controversially discussed (Brüderl/Ludwig 2015).

Finally, multi-level panel data improve the opportunity of differentiating between cohort and age effects on individual life courses and models of APC analysis (age-period-cohort analysis), which is prominently used in demographic research (Ludwig/Brüderl 2021; Kratz/Brüderl 2021).

5 Introducing the contributions

This Special Issue is divided into three parts. The first part contains a methodological contribution on the analysis of panel data. In the second part, a series of four articles from central fields of demographic research evaluate the progress in identifying causal mechanisms of demographic behaviour through longitudinal data analysis in general and panel analysis in particular. The third part contains two examples of studies using panel data to analyse complex research questions in family demography. These seven contributions will be briefly introduced in the following.

In their methodological contribution in the first part of the Special Issue, Volker Ludwig and Josef Brüderl (2021) discuss the case of estimating the impact of a dichotomous treatment, such as a certain life event, on outcomes in the further life course using panel data. In doing so, the authors detail the “the methodological subtleties of the approach”. Even though the standard methods of panel data analysis improve the quality of estimating impact functions, violations of the required assumptions for achieving unbiased estimates are still possible in many ways. This is shown with the instructive example of estimating the effect of motherhood on life satisfaction. The authors also provide guidance for tackling these obstacles.

In the first article of the second part, Michael Feldhaus and Richard Preetz (2021) give an overview of the state of the art in the longitudinal analysis of the dynamics of intimate relationships. Starting with a brief introduction of theoretical approaches spanning from broader macro-analytical concepts to the micro-analytical model of more proximate mechanisms of partnership-related behaviour, they discuss issues of causal analyses from a methodological point of view. In the main part of the article, the authors present recent findings from the rich literature using longitudinal data analysis. They focus on the most relevant biographic transitions between partnership statuses, i.e., union formation and living apart together; cohabitation

and marriage; and divorce, re-partnering and remarriage. For each of these, they select research questions which have been prominently addressed in the literature. According to their overview, following the first significant step of introducing event history methods, the increased use of longitudinal panel data represents a second considerable improvement in identifying causal mechanisms in demographic research. The recent studies on the effect of marriage on wellbeing are mentioned as one example in which neglecting processes of self-selection would lead to incorrect results, but which can be corrected by adequate modelling using panel data. However, the authors discuss the limited quality of collected data as a major obstacle to achieving more accurate results, which is connected with the fact that the underlying mechanisms behind an identified correlation of demographic factors are often not considered directly enough.

Next, *Michaela Kreyenfeld* (2021) evaluates the study of causal factors for reproductive behaviour. She focusses on three important substantive topics under investigation in this field of research: the important “female employment and fertility nexus”, the impact of family policy on reproductive behaviour, and the explanation of fertility intentions and their behavioural link. She proceeds by discussing the merits, advantages, and limitations of event history analysis, panel analysis (fixed-effects models), and natural experiments in this field of demographic research. Here, the thoughtful considerations regarding issues of anticipation of principally uncertain consequences of fertility (“narrative of the future”) are particularly interesting. *Kreyenfeld* then turns to the causal analysis of policy effects and presents an illustrative empirical analysis of the effects of a parental leave reform on the female employment and fertility nexus. As a major conclusion from her overview on the state of the art as well as her own empirical study, she acknowledges the progress in the causal analysis of fertility achieved through different methods of causal modelling with longitudinal data. However, *Kreyenfeld* also recommends that these advancements cannot substitute the improved specification and measurement of causal factors. She particularly and correctly suggests that anticipation must be measured in greater detail and precision.

In the third article, *Sergi Vidal* and *Philipp Lersch* (2021) synthesise the advancements made in migration and spatial mobility research using longitudinal and panel data. After clarifying some terminological issues, the authors present different approaches to longitudinal data collection and modelling causal mechanisms in spatial mobility research and refer to prominent panel studies which conduct refined analyses of mobility behaviour. In a broad overview of recent empirical studies of spatial mobility, they show how well-suited data can significantly ease disentangling the highly interdependent processes in which the internal and external migration of individuals is embedded. These studies often use a life course approach. As migration decisions are commonly made by households, a “linked lives” perspective plays a prominent role and multi-actor designs are employed. Finally, another section in the article is devoted to the contextual conditions of spatial mobility and their change due to different kinds of migration. The authors show that panel designs more adequately allow for the observation, modelling,

and testing of respective causal explanations relating to “the diverse and complex patterns of mobility and immobility”.

In their article, *Rasmus Hoffmann* and *Gabriele Doblhammer* (2021) address the fourth major field of demographic research; mortality and morbidity. They follow an original strategy in evaluating the advancement of causal analysis in a field of demographic research which is particularly challenging in this regard. After a differentiated overview on several approaches of causal modelling in the social sciences in general, they present an analysis of relevant research articles published in the leading journal “Demography” since 2010.⁵ The authors scanned the journal for articles on morbidity and mortality in which the considered approaches of causal analysis are applied. In the articles they found fixed effects and growth curve models (applied to panel and register data) stand out among the methods of causal modelling. Furthermore, they focus on studies of the impact of retirement, socio-economic status, and partnership and fertility on health outcomes, and present the findings of these articles. In their conclusions regarding the causal analysis of health issues in demographic research, they point out that there is a particular need for interdisciplinary cooperation in studying “the multi-faceted influence factors”. They also recommend to better combine findings on different levels of aggregation including “results from more holistic concepts”.

The third part of the Special Issue presents two new empirical studies which in different ways show how the use of panel data, in both cases stemming from the German Family Panel (pairfam), can be used to obtain estimates of causal effects with respect to important research questions in family demography.

Alexander Gattig and *Lara Minkus* (2021) investigate the potential effect of marriage on cohabiting partners’ life individual satisfaction. As mentioned, self-selection is a major issue here. This might not only be true for time-independent traits of the respondents, but also for unobserved heterogeneity changing over time. The authors consider the latter case in their FEIS-model, showing evidence for a brief and small increase in life satisfaction due to marriage, which can be interpreted causally based on the mechanisms presented by the authors.

Kristin Hajek (2021) investigates the complex relationship between having an abortion, relationship satisfaction, and union dissolution. She examines the question of whether the effect of experiencing an abortion on union dissolution is mediated by relationship satisfaction, or whether a potential correlation between these events is spurious because relationship satisfaction works as a confounder. Panel data enable her to track changes in relationship satisfaction for a certain period of time and cover if and when one of the mentioned events occurs. The results of a sequence of models show that even though – as is well-known – relationship satisfaction supports union dissolution, it neither has a mediating effect, nor is it a confounder. Among others, low relationship satisfaction probably does not fuel the likelihood of an abortion. On the contrary, having an abortion might depress relationship satisfaction temporarily.

⁵ They also conducted such an analysis of the journal “Population Studies”, which is not presented in the article.

6 Conclusion

The contributions in this Special Issue shed light on the advancements made in causal analysis in demography using longitudinal research designs. They contribute considerably to the knowledge about in the contemporary state of demographic research with regard to explaining core dimensions of demographic behaviour. The Special Issue also shows which caveats deserve further attention. We very much hope that it finds a broad audience in demography and the social sciences and will stimulate the discussion on methodological and substantive challenges of causal analysis.

In summary, what needs to be achieved for the field to advance further? We will briefly emphasize three aspects.

First, to obtain unbiased estimates of the causal effects of individual factors, a few deficiencies and pitfalls in current modelling must be considered. A particular challenge is the design of proper models of self-referential processes and path-dependency. This is connected to the problem of estimating the effects of lagged dependent variables. Following our definitions, one case can be that $\Delta y([t_2, t_3]) = f(y([t_0, t_1]))$, i.e., that $\Delta y([t_2, t_3])$ is triggered by previous changes in Y itself. One example is the Mathew effect studied in life course research (*Dannefer 1987*). Another case can be distinguished, namely that $\Delta y([t_2, t_3]) = f(\Delta x([t_2, t_3] \mid y(t_2)))$, i.e., that Y itself modifies the effects of a causal factor X. Put simply, this means that where one stands at a certain point in time matters for where one can go next. An additional case is studied in non-linear systems dynamics, where the aggregated outcome of individuals' behaviour influences the likelihood of others adopting it (diffusion) or the likelihood of avoiding it (dampening). These processes can be studied with multi-level models.

Second, we also need progress in the measurement and collection of data relating to simple or compound mechanisms. For instance, panel studies usually follow a certain time schedule, e.g., annual waves. However, this might not be appropriate for the mechanisms under study. Often, we miss the moment of decisive change, such as an individual change in an attitude or a life event. Therefore, we might observe the changing factor too late, perhaps even only after the outcome one is interested in. In psychology, methods of experience sampling have been introduced to tackle this problem. In the social sciences, such approaches are waiting to be developed in a useful and applicable way. Innovative methods such as "smart surveys" – are another promising instrument for improving the collection of time-dependent information (*Toepoel et al. 2020*).

Third, with respect to substantive aspects of demographic research, we see a need for an intensification of interdisciplinary research. This is a must if one acknowledges the multi-dimensionality and multi-level structure of the compound mechanisms this research deals with. Whatever advancements will be made methodically, without the study of these processes and the explanation of demographic behaviour as the driver of demographic change, demographic research will always remain incomplete. This is particularly true when we investigate processes of reverse

causality. The research on morbidity as discussed in the contribution of *Hoffmann* and *Doblhammer* (2021) is an impressive example.

For instance, interdisciplinary research on fertility has found that genetic issues, which are rarely studied in demography, can help to explain heterogeneity in the reproductive behaviour of individuals (*Kohler et al.* 2004; *Mills/Tropf* 2020). For the case of migration research, geography and climate play an important role and will become even more relevant as the climate change proceeds. Scanning the demographic literature and research activities in prominent institutions of demographic research, we realise that demographic research is well on its way in this regard – maybe even more so than other social sciences.

Acknowledgements

Finally, we would like to warmly thank our colleagues who contributed to this Special Issue for their professional work and great engagement. We also thank the reviewers for their support, thoughtful comments, and recommendations. And last but not least, we thank the CPoS editorial team at the BiB for the professional support and pleasant cooperation.

References

- Ajzen, Icek; Klobas, Jane* 2014: Fertility intentions: An approach based on the theory of planned behavior. In: *Demographic Research* 29: 203-232. <https://doi.org/10.4054/DemRes.2013.29.8>
- Allison, Paul D.; Williams, Richard; Moral-Benito, Enrique* 2017: Maximum Likelihood for Cross-lagged Panel Models with Fixed Effects. In: *Socius* 3: 1-17. <https://doi.org/10.1177/2378023117710578>
- Bernardi, Laura; Huinink, Johannes; Settersten, Richard A. Jr.* 2019: The life course cube: A tool for studying lives. In: *Advances in Life Course Research* 41. <https://doi.org/10.1016/j.alcr.2018.11.004>
- Blossfeld, Hans P.; Rohwer, Götz* 2002: *New approaches to causal analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Brüderl, Josef; Kratz, Fabian; Bauer, Gerrit* 2019: Life course research with panel data: An analysis of the reproduction of social inequality. In: *Advances in Life Course Research* 41. <https://doi.org/10.1016/j.alcr.2018.09.003>
- Brüderl, Josef; Ludwig, Volker* 2015: Fixed-Effects Panel Regression. In: *Best, Henning; Wolf, Christof* (Eds.): *The Sage handbook of regression analysis and causal inference*. London: SAGE: 327-359. <https://dx.doi.org/10.4135/9781446288146.n15>
- Dannefer, Dale* 1987: Aging as Intracohort Differentiation: Accentuation, the Matthew Effect, and the Life Course. In: *Sociological Forum* 2,2: 211-236. <https://doi.org/10.1007/BF01124164>
- DiPrete, Thomas A.; Eirich, Gregory M.* 2006: Cumulative Advantage as a Mechanism. For Inequality: A Review of Theoretical and Empirical Developments. In: *Annual Review of Sociology* 32: 271-297. <https://doi.org/10.1146/annurev.soc.32.061604.123127>

- Feldhaus, Michael; Preetz, Richard* 2021: Analysing transitions in intimate relationships with panel data. In: *Comparative Population Studies* 46: 331-362. <https://doi.org/10.12765/CPoS-2021-12>
- Gangl, Markus* 2010: Causal Inference in Sociological Research. In: *Annual Review of Sociology* 36: 21-47. <https://doi.org/10.1146/annurev.soc.012809.102702>
- Gangl, Markus; Ziefle, Andrea* 2015: The Making of a Good Woman: Extended Parental Leave Entitlements and Mothers' Work Commitment in Germany. In: *American Journal of Sociology* 121,2: 511-563. <https://doi.org/10.1086/682419>
- Gattig, Alexander; Minkus, Lara* 2021: Does marriage increase couples' life satisfaction? Evidence using panel data and fixed-effects individual slopes. In: *Comparative Population Studies* 46: 123-148. <https://doi.org/10.12765/CPoS-2021-05>
- Hajek, Kristin* 2021: The Interplay of Having an Abortion, Relationship Satisfaction, and Union Dissolution. In: *Comparative Population Studies* 46: 97-122. <https://doi.org/10.12765/CPoS-2021-04>
- Heckhausen Jutta; Buchmann, Marlis* 2019: A multi-disciplinary model of life-course canalization and agency. In: *Advances in Life Course Research* 41. <https://doi.org/10.1016/j.alcr.2018.09.002>
- Hedström, Peter; Ylikoski, Petri* 2010: Causal Mechanisms in the Social Sciences. In: *Annual Review of Sociology* 36: 49-67. <https://doi.org/10.1146/annurev.soc.012809.102632>
- Hoem, Jan; Kreyenfeld, Michaela* 2006: Anticipatory analysis and its alternatives in life-course research. Part 1: The role of education in the study of first childbearing. In: *Demographic Research* 15: 461-282. <https://doi.org/10.4054/DemRes.2006.15.16>
- Hofmann, Barbara; Kreyenfeld, Michaela; Uhlendorff, Arne* 2017: Job displacement and first birth over the business cycle. In: *Demography* 54,3: 933-959. <https://doi.org/10.1007/s13524-017-0580-4>
- Hoffmann, Rasmus; Doblhammer, Gabriele* 2021: Approaches and methods for causal analysis of panel data in the area of morbidity and mortality. In: *Comparative Population Studies* 46: 69-96. <https://doi.org/10.12765/CPoS-2021-03>
- Huinink, Johannes; Feldhaus, Michael* 2009: Family Research from a Life Course Perspective. In: *International Sociology* 24,3: 299-324. <https://doi.org/10.1177/0268580909102910>
- Kleiding, Niels* 2014: Event History Analysis. In: *Annual Review of Statistics and Its Applications*: 333-360. <https://doi.org/10.1146/annurev-statistics-022513-115558>
- Kohler, Hans P.; Rodgers, Joseph L.; Christensen, Kaare* 2004: Is Fertility Behavior in Our Genes? Findings from a Danish Twin Study. In: *Population and Development Review* 25,2: 253-288. <https://doi.org/10.1111/j.1728-4457.1999.00253.x>
- Kratz, Fabian; Brüderl, Josef* 2021: The Age Trajectory of Happiness. In: *PsyArXiv*. April 18. <https://doi.org/10.31234/osf.io/d8f2z>
- Kreyenfeld, Michaela* 2021: Causal Modelling in Fertility Research: A Review of the Literature and an Application to a Parental Leave Policy Reform. In: *Comparative Population Studies* 46: 269-302. <https://doi.org/10.12765/CPoS-2021-10>
- Lesthaeghe, Ron* 2002: *Meaning and Choice: Value Orientations and Life Course Decisions*. NIDI-CBGS Monograph 37. The Hague: Netherlands Interdisciplinary Demographic Institute.
- Leszczensky, Lars; Wolbring, Tobias* 2019: How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study. In: *Sociological Methods & Research* 48. <https://doi.org/10.1177/0049124119882473>

- Lillard, Lee A.* 1993: Simultaneous equations for hazards: Marriage duration and fertility timing. In: *Journal of Econometrics* 56,1-2: 189-217. [https://doi.org/10.1016/0304-4076\(93\)90106-f](https://doi.org/10.1016/0304-4076(93)90106-f)
- Ludwig, Volker; Brüderl, Josef* 2021: What You Need to Know When Estimating Impact Functions with Panel Data for Demographic Research. In: *Comparative Population Studies* 46: 453-486. <https://doi.org/10.12765/CPoS-2021-16>
- Mayntz, Renate* 2020: Causal mechanism and explanation in social science. MPIfG Discussion Paper 20/7. Max Planck Institute for the Study of Societies: Cologne [<http://hdl.handle.net/21.11116/0000-0006-54B8-C, 07.10.2021>].
- Mills, Melinda C.; Tropf, Felix C.* 2020: Sociology, Genetics, and the Coming of Age of Sociogenomics. In: *Annual Review of Sociology* 46. 553-581. <https://doi.org/10.1146/annurev-soc-121919-054756>
- Moors, Guy* 2001: Values and living arrangements: a recursive relationship. In: *Waite, Linda J. et al.* (Eds.): *The Ties that Bind: Perspectives on Marriage and Cohabitation*. Aldine de Gruyter Publishers, Hawthorne: 212-226.
- Morgan, Stephen L.; Winship, Christopher* 2007: *Counterfactuals and Causal Inference*. Cambridge University Press.
- Mund, Markus; Nestler, Steffen* 2019: Beyond the Cross-Lagged Panel Model: Next-generation statistical tools for analyzing interdependencies across the life course. In: *Advances in Life Course Research* 41,100249. <https://doi.org/10.1016/j.alcr.2018.10.002>
- Toepoel, Vera; Lugtig, Peter; Schouten Barry* 2020. Active and passive measurement in mobile surveys. In: *The Survey Statistician* 82: 14-26.
- Vidal, Sergi; Lersch, Philipp M.* 2021: Panel Data in Research on Mobility and Migration: A Review of Recent Advances. In: *Comparative Population Studies* 46: 187-214 <https://doi.org/10.12765/CPoS-2021-07>
- Wunsch, Guillaume; Gournon, Catherine* 2020: Causal assessment in demographic research. In: *Genus* 76,18: 1-20. <https://doi.org/10.1186/s41118-020-00090-7>.

Prof. Dr. Johannes Huinink (✉). University of Bremen, SOCIUM – Research Center on Inequality and Social Policy. Bremen, Germany.
E-mail: huinink@uni-bremen.de
URL: <https://www.socium.uni-bremen.de/ueber-das-socium/mitglieder/johannes-huinink/>

Prof. Dr. Josef Brüderl. University of Munich, Department of Sociology. Munich, Germany. E-mail: bruederl@lmu.de
URL: https://www.ls3.sozioogie.uni-muenchen.de/personen/professor/bruederl_josef/index.html

Comparative Population Studies

www.comparativepopulationstudies.de

ISSN: 1869-8980 (Print) – 1869-8999 (Internet)

Published by

Prof. Dr. Norbert F. Schneider

Federal Institute for Population Research
D-65180 Wiesbaden / Germany



Managing Editor

Prof. Dr. Johannes Huinink
Dr. Katrin Schiefer

Editorial Assistant

Beatriz Feiler-Fuchs
Wiebke Hamann

Layout

Beatriz Feiler-Fuchs

E-mail: cpos@bib.bund.de

Scientific Advisory Board

Karsten Hank (Cologne)
Michaela Kreyenfeld (Berlin)
Marc Luy (Vienna)
Natalie Nitsche (Rostock)
Zsolt Spéder (Budapest)
Rainer Wehrhahn (Kiel)

Board of Reviewers

Bruno Arpino (Barcelona)
Kieron Barclay (Rostock)
Laura Bernardi (Lausanne)
Gabriele Doblhammer (Rostock)
Anette Eva Fasang (Berlin)
Michael Feldhaus (Oldenburg)
Tomas Frejka (Sanibel)
Alexia Fürnkranz-Prskawetz (Vienna)
Birgit Glorius (Chemnitz)
Fanny Janssen (Groningen)
Frank Kalter (Mannheim)
Stefanie Kley (Hamburg)
Bernhard Köppen (Koblenz)
Anne-Kristin Kuhnt (Duisburg)
Hill Kulu (St Andrews)
Nadja Milewski (Wiesbaden)
Roland Rau (Rostock)
Thorsten Schneider (Leipzig)
Tomas Sobotka (Vienna)
Jeroen J. A. Spijker (Barcelona)
Heike Trappe (Rostock)
Helga de Valk (The Hague)
Sergi Vidal (Barcelona)
Michael Wagner (Cologne)