Approaches and Methods for Causal Analysis of Panel Data in the Area of Morbidity and Mortality[*]

Rasmus Hoffmann, Gabriele Doblhammer

Abstract: We aim to give an overview of the state of the art of causal analysis of demographic issues related to morbidity and mortality. We will systematically introduce strategies to identify causal mechanisms, which are inherently linked to panel data from observational surveys and population registers. We will focus on health and mortality, and on the issues of unobserved heterogeneity and reverse causation between health and (1) retirement, (2) socio-economic status, and (3) characteristics of partnership and fertility history.

The boundaries between demographic research on mortality and morbidity and the neighbouring disciplines epidemiology, public health and economy are often blurred. We will highlight the specific contribution of demography by reviewing methods used in the demographic literature. We classify these methods according to important criteria, such as a design-based versus model-based approach and control for unobserved confounders. We present examples from the literature for each of the methods and discuss the assumptions and the advantages and disadvantages of the methods for the identification of causal effects in demographic morbidity and mortality research.

The differentiation between methods that control for unobserved confounders and those that do not reveals a fundamental difference between (1) methods that try to emulate a randomised experiment and have higher internal validity and (2) methods that attempt to achieve conditional independence by including all relevant factors in the model. The latter usually have higher external validity and require more assumptions and prior knowledge of relevant factors and their relationships. It is impossible to provide a general definition of the sort of validity that is more important, as there is always a trade-off between generalising the results to the population of interest and avoiding biases in the estimation of causal effects in the sample. We hope that our review will aid researchers in identifying strategies to answer their specific research question.

Keywords: Causality · Health · Mortality · Panel data · Methods

---

1   Introduction

Researchers in population studies have long dealt with descriptive analyses (*Smith* 2009) and ever more refined methods to use aggregate data for developing life tables, estimating vital rates and describing population dynamics. While causality has not been at the centre of demographic research (*Smith* 2009) for quite some time, the demographic community exploring mortality and morbidity has embraced a variety of causal study designs and methods. With the greater availability of panel studies and registers, demographers also began to ask causal questions (*Engelhardt et al.* 2009; *Moffitt* 2005) and took up questions and methods from other disciplines, such as economics or epidemiology, that have a long tradition in causal analysis. However, there is not one specific approach to causal analysis, neither in demography nor in the other disciplines. As *Moffitt* (2005: 92) pointed out: to conclude about causality "synthesis and reconciliation studies that are based on a variety of different approaches" are needed.

In this article we take up this argument and discuss demographic studies with causal approaches in the field of mortality and morbidity. First, we briefly describe the most common methods and designs focused on causal relationships in the social sciences. We use the general term "treatment", which may reflect a binary treatment, such as retirement or divorce, but may also be a continuous treatment, such as socio-economic status (SES) or number of children. Then we use the methods described as search terms in a structured literature review based on the leading demographic journal "Demography". The aim of such a review is to explore how prominent causal research features in a journal which is generally assumed to display the core research in the discipline of demography, and which methods and data were used to control for unobserved heterogeneity or selection. We have limited our search to journals that explicitly have "Demography" or "Population Study" in their title, but we must bear in mind that demographers also frequently publish in epidemiological journals, as the boundaries between demographic research on health and mortality and neighbouring disciplines are often blurred. While it is conceivable that epidemiologic journals are preferred because of higher impact factors and faster peer review, it may also be that demographers find their research appropriate for an epidemiologic rather than a demographic community when the issue is causality, and vice versa when methodological issues are involved. Finally, we give an overview of causal approaches in three important topics of research on morbidity and mortality by describing studies about the dual relationship between health outcomes and (1) retirement, (2) SES, and (3) characteristics of partnership and fertility history.

Reflecting on our choice of the three research areas in demography, it is worth mentioning that we preferred non-recursive causal research questions that deal with the direction of causality, i.e. the dual relation between health/mortality and some other factor where health selection plays a major role in addition to unobserved heterogeneity. We consider these to be very clear and very difficult causal problems, ideal to illustrate the use of different causal methods. At the same time, there are many demographic questions of recursive causality that are of no less importance,

e.g. the causal effect of a policy on health, where reverse causality is not an issue. In principle, the same fundamental problems apply to both types of causality, but in recursive causal questions the alternative to direct causality is rather indirect causality or no causality, while in non-recursive causality one tends to focus on the direction of causality, which does not mean that the alternatives "indirect causality" and "no causality" do not exist here either. We close with a discussion of our findings.

2   Causal methods in the social sciences

2.1   Fixed effects (FE) and random effects models (RE)

Both models assume individual specific effects and try to control for omitted variable bias due to unobserved heterogeneity. In this definition, "individual" can be (1) an individual or (2) a population segment, such as social group, it may be (3) kinships, mothers, grandmothers, twins, etc., or an aggregate group measure, e.g. in an (4) area or cohort analysis (*Moffitt* 2005, 2009).

The random effects assumption is that the individual-specific unobserved effects are uncorrelated with the independent variables. The fixed effects model does not need the random effects assumption. Therefore, fixed effects estimates are consistent even if the random effects assumption does not hold. Random effects estimates will be biased in this situation. However, if the random effects assumption holds, both models will provide consistent estimates (if the "strict exogeneity assumption" holds), but the random effects estimator is more efficient (*Brüderl/Ludwig* 2015; *Wooldridge* 2002).

The first difference removes the time-invariant heterogeneity in the individual-level FE model; in the case of area- or group-level FE models, this is done by subtracting the group average over time. The estimator is also called the within-variance estimator and provides the average effect in the subgroup of the population that received the treatment. The latter often raises the criticism of extrapolation from the treatment group to the total population (*Cameron/Trivedi* 2005; *Nerlove* 2005).

It is common to combine FE and RE in one model specification with prior testing of whether the assumptions are fulfilled (*Baltagi* 2008; *Hsiao* 2014). There is also the "between-within estimator" (also known as the Mundlak estimator), which combines the advantages of random and fixed effect estimators (*Schunck* 2013). As far as we know, it has not yet been used in demography.

2.2   Instrumental variables

The instrumental variable approach addresses the problem of unobserved confounding due to selection; it is less concerned with the bias that individual unobservables introduce into the estimation process, but rather with the problem that the treatment effect might be biased by non-random selection into the treatment group. An instrumental variable needs to be mean-independent of the

unobservables directly influencing the outcome. In addition, it must be relevant, and thus be correlated with the probability of receiving the treatment (*Moffitt* 2009). These relationships may be specified in two equations: first the equation estimating the relationship between the instrumental variable and the treatment, and the second, estimating the relationship between the predicted treatment (from the first equation) and the outcome. Often these equations are combined into one, integrating the second equation into to first, which is then called the reduced form. There are several types of instrumental variables and *Moffitt* (2005, 2009) distinguishes between

1. cross-sectional ecological variables, such as differences in policies, laws and social structure, which are independent of an individual's own choice

2. population-segment fixed effects instruments, where the segments are defined by social or demographic groups and the instruments pertain to groups, such as welfare reforms for certain income or demographic groups (e.g. single mothers)

3. siblings and related instruments, where the instrument is the deviation of each individuals' treatment from the average group-specific treatment, e.g. one sibling has children, the other does not.

4. natural experiments, which are often defined as a residual category of instruments which appear to be random, such as the month of birth, the birth of (naturally conceived) twins, etc.

Combining the instrumental variable approach with FE panel analysis is considered the superior approach to simultaneously control for selection and for estimation bias by individual unobserved heterogeneity (*Moffitt* 2009). Another prominent type of instrumental variable approach is Mendelian Randomisation, where the instrument is genes (*Mills et al.* 2020).

2.3   Regression discontinuity

Regression discontinuity can be used when individuals or other units are assigned to a programme or policy depending on a cut-off point of a continuous measure, e.g. eligibility according to age (*Ludwig/Miller* 2007). While other sources of external variation are (unexpected) historical events, in regression discontinuity designs, it is a known (deterministic or probabilistic, i.e. "fuzzy") institutional rule to assign or withhold treatment that creates the relevant variation (*Gangl* 2010); see *Imbens* and *Lemieux* (2008) for a review. The basic idea is that –depending on the relationship between the assignment variable and the outcome – the exposure at the cut-off point is as good as random. Thus, a comparison of the outcome of those just below and just above the cut-off point provides an estimate of the effect of the programme or policy (*Hu et al.* 2017).

2.4   Difference-in-differences

Difference-in-differences is similar to regression discontinuity, but improves it by adding a control group. This method compares the change in the outcome for an exposed group before and after an event (e.g. the implementation of a policy) to the change in the outcome over the same time period for a non-exposed group (*Athey/Imbens* 2006; *Jones/Rice* 2011). The two groups may have different levels of the outcome before the policy, thus confounding by unobserved time-constant factors that differ between the exposed and unexposed is taken into account (*Hu et al.* 2017). Under the assumption that both groups follow a common trend, the difference in the change in the outcome between the exposed and non-exposed groups can be interpreted as an effect of the event or the policy (*Harper et al*. 2014).

2.5   Interrupted time series

Where time series data are available and there is a clear-cut event at a specific point in time, interrupted time series analysis can be used to estimate the effect of this event. This event can be a shock to individuals or change in an institution, programme or policy. Repeated-measures logistic regression analysis can be used to detect any sudden change in the level of the outcome (in regression terms: a change of intercept) or a more sustained change in the trend of the outcome (a change of slope) around the time of the event. The analysis estimates the causal effect by comparing the outcomes before and after the event. Interrupted time series is different from a difference-in-differences analysis because it does not use a control group. When randomised controlled trials are not feasible, a time series design is an alternative to estimate the effect of events (*Fretheim et al.* 2013) because it controls for prior trends before the event and studies the dynamics of change afterwards (*Burdorf* 2012; *Taljaard et al.* 2014; *Wagner et al.* 2002). However, it is problematic in situations with lagged-effects or other events affecting the outcome at the same time.

2.6   Growth curve model

The main focus of a (latent) growth curve model (GC) is on changes or development over time. This requires the subjects to be followed over time with repeated measures of each variable of interest. The goal of this model is to make inferences about the features of growth trajectories, i.e. the initial levels of outcome measures and their rate of change. In GCs, the changes are represented by growth parameters or trajectories (which are specified as latent variables)*: the intercept*, the initial value of the outcome measure, and *the slope,* which indicates how much the curve grows or the rate of outcome changes over time (*Pakpahan et al.* 2015). In terms of causal analysis, the model can estimate the causal effect of the initial level on the rate of change. GCs assume that the subject's growth trajectories vary randomly around the overall mean of growth trajectories (*Bollen/Curran* 2006; Wang/Wang 2012).

2.7 Propensity scores

The propensity score was defined by *Rosenbaum* and *Rubin* (1983) to be the probability of treatment assignment conditional on observed baseline covariates. In randomised experiments, the true propensity score is known and is defined by the study design. In observational studies, the true propensity score is not, in general, known. However, it can be estimated using the study data. In practice, the propensity score is most often estimated using a logistic regression model in which treatment status is regressed on observed baseline characteristics. The estimated propensity score is the predicted probability of treatment derived from the fitted regression model and conditional on the propensity score; the distribution of observed baseline covariates will be similar between treated and untreated subjects. Propensity scores control for observable characteristics at baseline in the hope that they also control for correlated unobservable characteristics.

Four different propensity score methods are used to remove the effects of confounding when estimating the effects of treatment on outcomes

1. Propensity score matching forms matched sets of treated and untreated subjects who share a similar value of the propensity score (*Rosenbaum/ Rubin* 1983) and is used to estimate the average treatment effect among the treated (ATT) (*Imbens* 2004).

2. Stratification by the propensity score groups individuals into mutually exclusive subsets based on their estimated propensity score, often into five equal-size groups using the quintiles of the estimated propensity score. Within each stratum, the effect of treatment on outcomes can be estimated and, using meta-analysis, the mean treatment-effect for the study population can be estimated.

3. Inverse probability of treatment weighting (IPTW) using the propensity score uses weights based on the propensity score to create a synthetic sample in which the distribution of measured baseline covariates is independent of treatment assignment (*Morgan/Todd* 2008).

4. Covariate adjustment using the propensity score regresses the outcome variable on an indicator variable denoting treatment status and the estimated propensity score.

More and more often, propensity scores are being estimated using machine learning methods (*Brand et al.* 2019).

2.8 Structural equations

A structural equation model (SEM) is a multivariate regression model that extends standard regression by allowing multiple outcomes, known as "endogenous" variables and unobserved "latent" variables. For each endogenous variable there is a corresponding regression equation, which can depend on other endogenous variables, as well as on exogenous variables. Exogenous variables are the predictor variables that are not determined by any other variable in the model. A SEM

combines the approach of confirmatory factor analysis for the measurement model and path analysis for the structural model. The measurement model describes how well the observed indicator variables measure the underlying latent variable, whereas the structural model describes the causal relationship among the variables. This combination is the core advantage of SEM: together, they can simultaneously take into account random measurement errors, the multiple dependent variables of the model, and estimate direct, indirect and total effects (*Acock* 2013; *Bollen* 1989; *Wang/Wang* 2012).

2.9   G-computation

Similar to structural equations, G-computation is a new and flexible approach to causal analysis that is especially useful for mediation analysis. The parametric g-formula, originally developed by biostatisticians (*Robins/Hernán* 2009), has recently been used in demographic studies on interdependent life course processes with time-varying confounding (*Bijlsma/Wilson* 2020). The method first runs multivariable regression models to estimate interrelationships between variables, based on the assumptions of a specified Directed Acyclic Graph (DAG). Then it uses these fitted regression models to perform a series of micro-simulations and compare counterfactual scenarios to estimate causal effects. The g-formula approach assumes that all relevant variables are observed. But it is possible to explore the risk of unmeasured confounding by sensitivity analyses (*Carnegie et al.* 2016; *Lin et al.* 2017; *VanderWeele* 2015). The g-formula approach can be summarised as a series of three analytical phases (*Bijlsma/Wilson* 2020): 1. Causal diagram: A DAG is created to describe the causal interrelationships to be studied. It describes the assumed relationships between the observed variables, including time-varying effects, confounders, mediators and outcomes. 2. Estimation: Based on the DAG, a series of multivariable regression models are estimated using observed data. These models can take any parametric (functional) form, and the number of models will depend on the number of variables involved. 3. Simulation: With the estimated parameters from the models, a series of causal processes can be simulated.

3   Causal methods used in demographic research published in the journal "Demography"

We searched the journal "Demography" using the online search tool provided by the publisher Springer. Publications from 2010 until 13 May 2020 were included and the following search was applied (("Unobserved heterogeneity" OR "Random effects" OR "Regression Discontinuity" OR "Interrupted time series" OR "G-Computation" OR "Structural Equation" OR "Growth Curves" OR "Instrumental variables" OR "Propensity Score" OR "difference-in-differences" OR "fixed effects") AND ("health" OR "morbidity" OR "mortality" OR "death") AND ("panel" OR "longitudinal")).

After the screening of titles and method sections, this search produced 37 publications (Table 1). The most common causal method was fixed effects models

Tab. 1: Result of the structured literature review in the journal "Demography"

Author	Year	Title	Methods
Julie A. Phillips	2012	Factors Associated With Temporal and Spatial Patterns in Suicide Rates Across U.S. States	FE
Thomas P. Vartanian; Linda Houser	2012	The Effects of Childhood SNAP Use and Neighborhood Conditions on Adult Body Mass Index	FE
Ryan K. Masters	2012	Uncrossing the U.S. Black-White Mortality Crossover: The Role of Cohort Forces in Life Course Mortality Risk	FE
Todd E. Elder	2013	The Predictive Validity of Subjective Mortality Expectations: Evidence From the Health and Retirement Study	FE
Elizabeth Nauman et al.	2015	Rural-to-Urban Migration and Changes in Health Among Young Adults in Thailand	FE
Kieron Barclay; Martin Kolk	2015	Birth Order and Mortality: A Population-Based Cohort Study	FE
Charles L. Baum	2017	The Effects of College on Weight: Examining the 'Freshman 15' Myth and Other Effects of College Over the Life Cycle	FE
Osea Giuntella	2016	Assimilation and Health: Evidence From Linked Birth Records of Second- and Third-Generation Hispanics	FE
Kieron J. Barclay; Martin Kolk	2018	Birth Intervals and Health in Adulthood: A Comparison of Siblings Using Swedish Register Data	FE
Ross Macmillan et al.	2018	Gender and the Politics of Death: Female Representation, Political and Developmental Context, and Population Health in a Cross-National Panel	
Emma Zang; Cameron Campbell	2018	Males' Later-Life Mortality Consequences of Coresidence With Paternal Grandparents: Evidence From Northeast China, 1789-1909	FE
Thomas Laidley et al.	2019	New Evidence of Skin Color Bias and Health Outcomes Using Sibling Difference Models: A Research Note	FE
Petter Lundborg et al.	2016	The Effect of Schooling on Mortality: New Evidence From 50,000 Swedish Twins	FE
Max Brüning; Josselin Thuilliez	2019	Mortality and Macroeconomic Conditions: What Can We Learn From France?	FE
Jonas Helgertz; Tommy Bengtsson	2019	The Long-Lasting Influenza: The Impact of Fetal Stress During the 1918 Influenza Pandemic on Socioeconomic Attainment and Health in Sweden, 1968-2012	FE
Tabea Bucher-Koenen et al.	2020	Double Trouble: The Burden of Child-rearing and Working on Maternal Mortality	FE

Tab. 1:   Continuation

Author	Year	Title	Methods
Ryan D. Edwards	2018	If My Blood Pressure Is High, Do I Take It To Heart? Behavioral Impacts of Biomarker Collection in the Health and Retirement Study	FE
Jere R. Behrman; Susan W. Parker	2013	Is Health of the Aging Improved by Conditional Cash Transfer Programs? Evidence From Mexico	FE, D-D, PS
Jennifer Caputo et al.	2020	Midlife Work and Women's Long-Term Health and Mortality	FE, GC
David Frisvold; Ezra Golberstein	2013	The Effect of School Quality on Black-White Health Differences: Evidence From Segregated Southern Schools	FE, IV
Kenneth A. Couch et al.	2015	The Long-Term Health Implications of Marital Disruption: Divorce, Work Limits, and Social Security Disability Benefits Among Men	FE, PS
Ana Lucia Abeliansky; Holger Strulik	2018	How We Fall Apart: Similarities of Human Aging in 10 European Countries	FE, RE
Iliana V. Kohler et al.	2017	The Demography of Mental Health Among Mature Adults in a Low-Income, High-HIV-Prevalence Context	FE, RE, GC
Hui Zheng et al.	2011	Heterogeneity in the Strehler-Mildvan General Theory of Mortality and Aging	RE
Ross M. Stolzenberg	2011	Do Not Go Gentle Into That Good Night: The Effect of Retirement on Subsequent Mortality of U.S. Supreme Court Justices	IV
Ross M. Stolzenberg; James Lindgren	2010	Retirement and Death in Office of U.S. Supreme Court Justices	IV
Volha Lazuka	2019	Early-Life Assets in Oldest-Old Age: Evidence From Primary Care Reform in Early Twentieth Century Sweden	D-D
Susan W. Parker et al.	2018	Health Insurance and the Aging: Evidence From the Seguro Popular Program in Mexico	D-D
Nicole Guertzgen; Karsten Hank	2018	Maternity Leave and Mothers' Long-Term Sickness Absence: Evidence From West Germany	D-D
Michal Engelman; Heide Jackson	2019	Gradual Change, Homeostasis, and Punctuated Equilibrium, Reconsidering Patterns of Health in Later Life	GC
Martin O'Flaherty et al.	2016	The Family Life Course and Health: Partnership, Fertility Histories, and Later-Life Physical Health Trajectories in Australia	GC
Feinian Chen et al.	2015	For Better or Worse: The Health Implications of Marriage Separation Due to Migration in Rural China	GC
Nicholas J. Bishop et al.	2016	Estimating the Co-Development of Cognitive Decline and Physical Mobility Limitations in Older U.S. Adults	GC

Tab. 1: Continuation

Author	Year	Title	Methods
Steven A. Haas et al.	2017	The Life Course, Cohort Dynamics, and International Differences in Aging Trajectories	GC
Liliya Leopold	2019	Health Measurement and Health Inequality Over the Life Course: A Comparison of Self-rated Health, SF-12, and Grip Strength	GC
Randall Kuhn et al.	2011	The Effects of Children's Migration on Elderly Kin's Health: A Counterfactual Approach	PS
Jeremy Pais	2014	Cumulative Structural Disadvantage and Racial Health Disparities: The Pathways of Childhood Socioeconomic Influence	LCA

FE: fixed effect models, RE: random effect models, IV: instrumental variable, D-D: difference-in-difference, GC: growth curve models, PS: propensity score. We consider LCA (latent class analysis) to be similar to GC.

(23) with individual FE, but also kinship FE (mothers, grandmothers, siblings, twins) and region/country/cohort FE. The latter were most often used with aggregate data and thus did not use individual panel data. FE models were sometimes combined with other design features, such as random effects (RE), the difference-in-differences approach, propensity score matching/regression, or growth curve models. Both register data and panel data were used, and panel and registers were often linked to accommodate the various designs. The second most common method was (latent) growth curve models (9), followed by difference-in-differences designs (4), instrumental variables (3), propensity scores (3), and random effects models (3). A second search in the journal "Population Studies" confirmed the overall picture with a strong reliance on FE models with sibling designs based on register data (not shown).

4   Topic 1: The dual relationship between retirement and health

The transition to retirement is an important step in the life course and an important study subject in demography. This is because retirement policies are a widely used instrument to address demographic change and the harmonisation of labour market demands, individual preferences, health development and public finances is a core problem of public policy. The empirical analysis of the effects between health and retirement is complex because the retirement process depends on health, and retirement has an effect on subsequent health (*Oksanen/Virtanen* 2012; *Radó/Boissonneault* 2018). It is relatively straightforward to hypothesise and to find empirical evidence that poor health leads to earlier retirement (*van Rijn et al.* 2014). Thus, we concentrate on the question as to how retirement affects health, which is the more complicated question, because health is a confounder that influences both the retirement transition and health after retirement. Available theory suggests two possible effects: On the one hand, leaving an active role on the labour market can lead to a form of crisis (*Atchley* 1975); on the other hand, retirement can be a relief from work-related burdens and stress (*Westerlund et al.* 2009). Given these theoretical assumptions it is not surprising that the related empirical findings are inconsistent: Some studies show a positive influence of retirement on health (*Eibich* 2015; *Insler* 2014; *Jokela et al.* 2010; *Westerlund et al.* 2009), while others find a negative effect (e.g. *Behncke* 2012; *Stenholm et al.* 2014). A systematic review (S*him et al.* 2013) shows a positive correlation between retirement and mortality, but points out that the problem of selection (i.e. poor health not only being a consequence but also a cause of retirement) is only insufficiently dealt with. Two other systematic reviews suggest that the effect of health can be different between occupational status groups (effect heterogeneity) (*Schaap et al.* 2018; *van der Heide et al.* 2013).

   In the following we describe the four main causal methods that have been applied to longitudinal data to explore the effect of retirement on health.

   a) *Fixed effects models (FE)*: as mentioned above, FE models use multiple measurements per individual to take account of unobserved constant

confounders. However, the main confounder in our setting is prior health influencing both retirement and health after retirement, and this confounder might not be constant and not be accounted for by FE. Using data from the Health and Retirement Study (HRS) *Calvo et al.* (2013) combine fixed and random effects with an instrumental variable approach and find that early retirements decrease health.

b) Several studies used the statutory retirement age as an *instrumental variable* (*IV*) that influences retirement without affecting health. For example, *Hessel* (2016) uses data from the European Union Statistics of Income and Living Conditions (EU-SILC) and finds that retirement improves health. *Hanemann* (2017) uses data from HRS, the English Longitudinal Study of Ageing (ELSA) and the Survey of Health, Ageing and Retirement in Europe (SHARE), and finds that retirement improves physical health and deteriorates cognitive health. A potential problem with this method is that it identifies the local average treatment effect (LATE), i.e. the effect of retirement in the population for which the instrument applies. In this case, it is the population that retires because the statutory retirement age is reached. This effect can differ from the overall effect of retirement for all people who retire. There are indications that studies using IV tend to find positive health effects of retirement, while studies that control for confounding factors find negative effects (*Behncke* 2012).

c) The *regression discontinuity* design is similar to the IV approach in that it also uses the statutory retirement age, e.g. different age cut-offs in different countries. *Coe* and *Zamarro* (2011) use SHARE and *Eibich* (2015) uses data from the German Socio-Economic Panel (SOEP). Both find that retirement improves health. *Giesecke* (2019) uses administrative data from the German federal pension insurance and finds substantial effect heterogeneity between people with different pension types and different life time earnings; poorer people experience a mortality decrease while richer people a mortality increase. Also with this method, it is possible that only local causal effects are measured (among those who are just below and above the age threshold) that may not be generalisable to the whole population.

d) Finally, it is possible to study the effects of retirement on health with an *interrupted time series* analysis that compares the health trend before and after retirement. *Schuring et al.* (2015) study health trends before and after transitions into early retirement with data from the European Community Household Panel (ECHP). They conclude that lower educated people are more likely selected into early retirement because of poor health, while higher educated people who retire early experience a health decline after retirement.

5 Topic 2: The dual relationship between socio-economic status (SES) and health

The well-known association between SES and health can be explained by two opposed causal mechanisms: *social causation*, by which SES influences health, and *health selection*, by which health influences SES (*Goldman* 2001). For a complete account of causal models explaining health inequalities, one needs to add *indirect selection*, which means that a common underlying factor, such as genes, cognitive

ability, family factors, personality, or general life-style orientations, influences both SES and health, so that there is no causal effect between SES and health.

It is important to note that SES in itself is a highly debated concept. We use it here as a simplifying label for all studies that look at the relation between any aspect of SES (the most important being education, occupation and income) and health. But it is an open question whether there is a latent variable, such as the "socio-economic status" of a person, which can be measured and approximated by education, occupation or income as *indicators,* or whether SES is too imprecise and meaningful (causal) relationships can only be established between specific resources, such as education, income, or those related to occupational status or occupational class. On the one hand, specific variables (e.g. income) more convincingly fulfil the requirement that any causal interpretation should be based on theoretical knowledge of the pathways and mechanisms. On the other hand, such an analytical approach to social variables can become unrealistic because social variables do not work in isolation. In this regard, it is possible to study the effect of lottery wins on health (*Lindahl* 2005): while this isolated and randomised effect of income is closer to the causal "rules" of an experiment, the assumption is that it is far different from the mechanisms that create the health gradient between income groups in the general population.

While the association between SES and health is generally accepted, and probably also the assumption that all three models are real to some extent (*Smith* 1999), which of them actually contributes most to the social gradient in health is highly debated (*Kröger et al.* 2015). The highest level of agreement can probably be reached for the effect of education on health, partly because it has been studied with a variety of approaches, such as natural experiments on school reforms or compulsory schooling law (see below) and twin studies (*Madsen et al.* 2010). However, for the association between income and health there is still disagreement on the direction of causation between different studies and disciplines. While countless studies by social epidemiologists and medical sociologists implicitly or explicitly assume that income has an effect on health, e.g. *Galama* and *van Kippersluis* (2018) believe that the losses of income and wealth as a consequence of poor health is the most dominant causal relation between health and dimensions of SES. In the following we present a selection of approaches, studies and the respective results which, we believe, have contributed to this ongoing discussion. They also illustrate the methods that can be used for this and similar questions exploiting panel data for causal analysis.

   a) Two examples shall be presented with *fixed effects models*: first, *Foverskov* and *Holm* (2016) is one of the few studies that address all three causal models in the same study. They analyse the British Household Panel Survey and find no support for social causation, and limited support for health selection. Their conclusion is that indirect selection may be the most important mechanism. This conclusion is questionable because persons aged 30 to 60 are observed for 5 years and everything before age 30 is defined as a common background factor. Second, a quite different application of a fixed effects model to a natural experiment was done by *Frijters et al.* (2005), who only look at a causal effect

of changes in income on health after German reunification using the German Socio-Economic Panel. They find a statistically significant but small effect.

b) In the same vein as the study above, and again reflecting the great interest of economists in this question, a large number of studies analyse the effects between material wealth and health with *instrumental variables*. It is interesting to see how in the following three studies sources of external variation have been exploited to create causal evidence. *Michaud* and *Soest* (2008) use inheritances and find no evidence that wealth affects health, but strong evidence of effects from both spouses' health on household wealth in the Health and Retirement Study. *Lindahl* (2005) uses the Swedish Level of Living Survey and lottery prizes as an exogenous source of variation in income and finds causal effects from income to health. Finally, in several publications based on the Health and Retirement Study and the Panel Study of Income Dynamics, *Smith* (e.g. 2004) uses stock market changes as an instrument for changes in income and concludes that income as such has no effect on health, but the socio-economic status does, namely through education. As conceded by Smith himself, it is questionable whether the instrument used in his study or by the other authors above represent the causal effects of material wealth in the general population.

c) With regard to this effect of education on health, many studies have used school reforms as natural experiments and *regression discontinuity analysis*. *Lleras-Muney* (2005) finds a strong positive effect of education with US census data, and also *Kippersluis* (2010) confirms such an effect with linked survey and register data from the Netherlands. Conversely, *Albouy* and *Lequien* (2009) do not find such effect in French longitudinal data (Echantillon Démographique Permanent). In a systematic review and meta-analysis of this approach, *Gathmann et al.* (2015) find that more education yields small mortality reductions, but only for men. While this method may reveal the causal effect in a specific historical setting, this effect may not be the same as the one behind the social gradients in health in contemporary societies.

d) *Growth curve models* have also been used to study the interplay between the process of health development and the process of development of occupation and income during the life course. For example, *Halleröd* and *Gustafsson* (2011) analysed data from the Swedish Survey of Living Conditions and find support for both social causation and health selection. It can be problematic to use the same variable across a longer period in the life course (which is necessary for this method), both for pragmatic reasons of availability and comparability and for theoretical reasons of the meaning of a variable (e.g. occupational status) in different ages.

e) *Structural equation models* have been applied in several ways to study the relationship between SES and health, often over long periods in the life course. We mention here, first, studies that use this method to test how different aspects of SES and health in different ages directly and indirectly influence a health outcome later in life (e.g. *Chandola et al.* 2006 based on the UK National Child Development Study; *Pakpahan et al.* 2017 based on the Survey of Health, Ageing and Retirement in Europe). In these models, aspects of all three causal mechanisms described above can be found, but are not directly compared. Second, there are studies explicitly aimed at a quantitative comparison of social causation versus health selection as opposed causal mechanisms over longer parts of the life course. Examples of such studies are *Warren* (2009) and *Hoffmann et al.* (2018) who both find stronger evidence for social causation than for health selection; Warren with the Wisconsin Longitudinal

Study and Hoffmann with the Survey of Health, Ageing and Retirement in Europe. In a related paper *Hoffmann et al.* (2019) use this framework to compare the effects between health and education, occupation, and income (see discussion about SES above). A crucial and untestable assumption of this approach is that all relevant confounders are taken into account so that treatment assignment and outcome are conditionally independent. There are more methods within the wide definition of structural equation models that have been used to study effects between SES and health, e.g. *path analysis* (*Palloni et al*. 2009) and *G-computation* (*Bijlsma et al.* 2017).

Finally, it is worth mentioning that also other methods from our overview above have been used for causal analysis between SES and health, e.g. *propensity score matching*, often combined with *difference-in-differences*, but these are more focused on the effect of policies on an SES- or health-related outcome.

6 Topic 3: The dual relationship between partnership, fertility history, and health/mortality

The relationship between characteristics of family formation and health has been a long-standing issue in demography. Following the recent review by *Hank* and *Steinbach* (2018) we differentiate between the aspects of partnership and fertility history; both are confronted with the underlying question as to whether certain characteristics of these biographies have protective or detrimental effects on later-life health, or whether any empirical relations are caused by selection forces into or out of partnership, and parenthood. While these two aspects are often explored separately, more frequently there are also attempts to jointly model their effects on health (see below, e.g. *O'Flaherty et al.* 2016).

6.1 Partnership

Over a long period, the difference in health between single, married, divorced and widowed individuals was at the centre of research. It has been repeatedly suggested that marriage has a protective effect on health due to economic and social benefits, as well as beneficial lifestyle choices particularly among men. As Hank and Steinbach, however, point out in their review, selection into marriage may be related to better health in the first place, by affecting an individual's chance of getting married. On the other hand, marital disruption and divorce appears to have detrimental effects both in the short and long-term, even after remarriage. And the effect of divorce differs according to marriage satisfaction and by gender. Women's health appears to suffer more after divorce than men's (*Monden/Uunk* 2013).

a) The following example demonstrates how a *fixed effects model*, to control for unobserved confounding, together with *propensity score matching*, to control for observed confounding, was used to explore the effect of divorce on men's health. Using the Survey of Income and Program Participation (SIPP) panel study matched to US social security administrative data, *Couch et al.* (2015)

followed continuously married men for 20 years and compared them with their counterparts experiencing divorce. In addition, they filtered the data for previous health issues to help control for health selection into divorce and find that among those not re-marrying, divorce increases men's long-term probability of both self-reported work limitations and federal disability benefit receipt. They attribute the negative health impact to the lack of marital resources.

b) The following two studies used *growth curve models* to explore marriage separation in the context of Chinese migration, and of marriage disruption in Australia. Based on the "China Health and Nutrition Survey", *Chen et al.* (2015) investigated health trajectories of left-behind rural individuals whose spouses migrated for work. Their linear growth curve models take into account that individuals start with different levels of self-rated physical health and that each individual could experience a different rate of change across age dependent on their marital status and the residence of the spouse. The time varying-covariates allow that each individual is taken as its own control to account for within- and between-individual unobserved heterogeneity. Their results point towards a clear health disadvantage of married individuals whose spouses are absent compared with those whose spouses are living in the same household. Longer absence led to worse health and health deficits were stronger for men than women.
The second article incorporates the whole family life cycle from age 18 to 50 exploring both fertility and partnership histories (*O'Flaherty et al.* 2016). Using the Household, Income and Labour Dynamics in Australia (HILDA) survey they used a twofold analysis strategy. First, applying sequence analysis, they grouped individuals with similar fertility and partnership histories and used this categorisation as a primary independent variable for the second stage, where they applied *growth curve models* to establish the relationship with later-life health. They found gendered results with a stronger link of long-term family trajectories and health among men than women. Early or no family formation, marital disruption or high fertility was particularly harmful for men, and high fertility levels with a disrupted marital biography for women.

6.2 Fertility history

Turning to the second aspect, *Hank* and *Steinbach* (2018) in their review provided an extensive overview of the different fertility characteristics, such as parity, childlessness, birth spacing/intervals etc., that have been explored. They discussed both biological and social factors which may cause the relationship between (late) life health and fertility characteristics among both genders, stressing the lack of knowledge about the relative importance of these factors. The biological factors include diseases such as breast cancer, as well as other cancers of the female reproductive system, which were shown to be associated with pregnancy, childbirth and lactation; among the social factors, differences in socio-economic status, social relationships, and health behaviours across the life-course have been suggested. At the same time, health selection into certain fertility characteristics may play an important role, producing biased estimates if not accounted for (*Doblhammer/ Oeppen* 2003).

a) *Fixed effects siblings design* approaches are common with register data due to the availability of family links in the data and large sample sizes. *Einiö et al.*

(2016) applied such a design, which is described in more detail below, when addressing the research by *Barclay* and *Kolk* (2015, 2018), and explored the relationship between the number of children and later-life mortality among Finns. They confirmed earlier findings, which did not use causal modelling techniques, that all-cause mortality relative to those with two children is highest among childless women followed by women with one child and could also extend these findings to men. They concluded that living conditions in adulthood contributed to the association between the number of children and mortality to a greater extent than childhood background, and chronic conditions contributed to the excess mortality of the childless, probably revealing health selection into childlessness. These designs suffer from the problem of extrapolation because they can only use sibships and thus exclude childless individuals and those with one child only. In addition, they need heterogeneity in the outcome, thus sibships with all individuals still alive are also excluded. The results of the study, however, were reproducible for the total population without a sibling design. When studying mortality by cause of death, sample size issues are common in sibling designs despite the large number of observations in register studies. Although the sibling comparison design with sibling fixed effects analysis is very common in demography, concerns about the causal interpretation of the findings have been raised. As far as we know, these concerns have not been widely acknowledged in demography: e.g. problems related to precision and bias of estimates (*Gilman/Loucks* 2014), the violation of the assumption that the exposure and outcome of an individual do not affect the exposure and outcome of his/her siblings ("sibling carryover") (*Sjölander et al.* 2016), or overcontrolling by the indiscriminate control for confounders, mediators and colliders (*Sjölander/ Zetterqvist* 2017).

b) Recent studies extended the topic to siblings and explored the effect of sibship size, birth order, and birth intervals between siblings on their later life health. Using the Swedish multigenerational population register data, *Barclay* and *Kolk* (2015) applied *fixed-effect discrete-time survival analysis* using a within-family comparison of siblings with the same biological mother–father pairing. The fixed effects were applied at the level of siblings to adjust for all factors that remained constant within the sibling group (e.g. sibling size) as well as factors difficult to observe and measure, such as shared socio-economic background and general parenting style. Exploiting within-family variation in mortality by cause of death, they showed that the relative effect of birth order was greater among sisters than among brothers, particularly for mortality attributable to cancers of the respiratory system and to external causes. Social pathways only mediated the relationship between birth order and mortality risk in adulthood to a limited degree. In a second article with the Swedish multi-generational register data linked to the Swedish military conscription register and applying the same modelling strategy, *Barclay* and *Kolk* (2018) concluded that birth intervals had little effect on long-term health outcomes of brothers.

c) To explore the relationship between the level and rate of change in cognitive functioning and associations with fertility history, *Read* and *Grundy* (2017) used ELSA and applied *growth curve models* with random effects to capture individual differences, and *fixed effects* to estimate the average growth of the entire sample. The results showed associations between the number and timing of births and cognitive functioning in older age, in particular adverse effects of high parity, early childbearing and low parity which appeared to reflect underlying socio-economic and health disparities.

d) *Structural equation models* have been applied to explore fertility trajectories and later-life depression in England (*Grundy et al.* 2020). Using the English Longitudinal Study of Ageing (ELSA), they applied path analysis within the structural equation framework to investigate direct and indirect effects. Early parenthood, experience of a short birth interval, and high parity (four or more children) appeared to be linked to depressive symptoms in later life. These factors mainly worked indirectly, mediated by other life course experiences related to SES, partnership and lifestyle characteristics.

7   Discussion

Our empirical account of causal methods applied to longitudinal observational data in the field of demography rests on two parts: first, we searched the causal methods used in the studies in the leading demographic journal "Demography" using specific keywords. *Fixed effects models* accounting for individual or family invariant characteristics in combination with other methods were the leading approach, followed by *growth curve models*. These methods were applied to both panel and register data, sometimes the two were linked but also FE area/cohort models with aggregate trend data were used. Second, we used our own knowledge and experience to analyse three important areas of research in demography and described the contributions of prominent methods to the evidence base in these fields and to the concrete research questions. The first approach is more systematic and superficial, while the second is admittedly more arbitrary, but allows a deeper insight into how a variety of methods have been used to answer core demographic questions. These methods can complement each other, but can also produce conflicting evidence. Overall, *fixed effects models* seem to be the most prominent causal approach in demographic analysis of health and mortality.

FE models eliminate stable unmeasured factors related to invariant characteristics. For individuals, this works well for assessing treatment effects of variant characteristics, but excludes the analysis of invariant characteristics, such as completed family size, completed parity, or birth order. In these cases, individual FE models are not possible because all invariant characteristics are cancelled out of the equation. Researchers therefore try to control for invariant unobserved characteristics, such as genes, upbringing, living circumstances, values and norms, by using FE for higher level entities, such as siblings, mothers, or grandmothers. Register data, in particular, lend themselves to within-family comparisons using siblings-designs with FE since observations of this nature are seldom available in panel studies. The FE approach with individual fixed effects is more naturally tailored for the study of SES, partnership status, or retirement processes on health since these characteristics are largely time variant. It does not work for the study of (long-term) effects of (completed) fertility characteristics and other characteristics related to early life because there is either little or no variation in these features over the life course. *Growth curve models* are an extension of FE models and are the second most common approach in the journal "Demography". These models are often combined with FE approaches. In contrast to FE models, they permit

estimates of the effect of observables on the outcome at baseline as well as on the growth rate.

In our literature review, a couple of studies used instrumental variables to identify recursive causality by exploring pension eligibility, political climate, differences in infant mortality between black and white, the decline in child labour, or the decline in family size and their relation to health. Apart from unobserved heterogeneity and reverse causation, the hierarchical structure of data presents an additional source of bias, and multi-level modelling is often applied to address this bias. Usually regional macro-data are combined with individual-level data to differentiate between contextual and individual-level determinants of mortality and morbidity. Two of the many examples for Germany are a study of individual area-level effects on mortality on the basis of data from the German pension fund (*Kibele* 2014), and individual and contextual determinants of health among ethnic German immigrants (*Kreft/Doblhammer* 2012).

In the following, we want to discuss the categorisation of causal methods into design-based versus model-based approaches (*Koch/Gillings* 2006) and a similar categorisation into methods that address unobserved confounders versus methods that deal with observed confounders. These two categorisations are largely overlapping in the sense that the conditioning on (time-constant) unobservables is generally design-based (fixed effects, instrumental variables, regression discontinuity, and difference-in-differences), while model-based methods address observable confounders (interrupted time series, growth curve models, propensity scores, structural equation models, and G-computation). The first group originates in econometrics, while the second originates in biostatistics and epidemiology. However, this distinction can be fuzzy, e.g. the g-formula can use fixed effects intercepts and difference-in-differences approaches will sometimes use additional covariates in order to strengthen the parallel trend assumption. Twin studies, which is arguably a design, use twin-fixed effects, with fixed effects generally being a model with no need for a special design (in the sense that it exploits external variation) other than the longitudinal design. Finally, it is difficult to decide whether interrupted time series is a design-based approach (because it is a quasi-experimental method that needs a natural experiment situation) or a model-based approach (because specific modelling approaches have been invented to deal with them, e.g. ARIMA). It is easier to claim that interrupted time series cannot deal with unobserved confounders. We would argue that all studies have a model and a design, and they deal or do not deal with unobserved confounders, therefore we give preference to the categorisation into unobservable vs observable confounders.

Methods conditioning on unobservables generally have higher internal validity, while methods conditioning on observables have higher external validity. However, this external validity depends on the sample and whether it is a random sample of the population of interest. The trade-off between these two validities when assessing different causal methods has long been discussed (*Moffitt* 2005; *Smith* 2013). Disciplinary traditions and changes over time influence which of these two aspects of scientific inquiry is deemed more important. *Moffitt* (2005) mentions the "danger in maximising internal validity at the expense of external validity" and

Smith (2013) advocates a balance of randomisation, representation and realism as integrated aspects of a meaningful causal analysis. Since not all researchers across disciplines are open to all methods, this calls for an intensified mix and comparison of methods which can yield the highest synergies in interdisciplinary cooperation, e.g. of demographers, sociologists and (social) epidemiologists in research on health and mortality.

What is the special contribution of demographers to the field of causality in health research? One may say:

(1) While demographers, like other social scientists, use many different longitudinal surveys they also have a special preference for the use of register data, sometimes linked with panel data and the exploitation of designs that uniquely fit to register data, such as kinship designs.

(2) As demographers are used to work on the total population level, they tend to explore causal relationships in health on the total population level rather than for specific groups in specific situations. Thus, they often search for high external validity which may come with reduced internal validity and may weaken the claim of causality.

(3) Compared to epidemiologists, they use general health outcomes (in addition to mortality) and explore factors of health that may have a number of different causal pathways, such as the relationship between fertility history, SES, or marital status and health. Both the general health outcome and the multi-faceted influence factors certainly complicate the search for causality and often the relative importance of these factors cannot be determined. This also calls for the interdisciplinary cooperation mentioned above because results for more specific and isolated variables need to complement results from more holistic concepts.

(4) There is a long demographic tradition of conducting ecological studies with panel data which, although not suitable for exploring causal mechanisms, provide important information on associations between socio-economic factors and health and mortality. Complex decomposition methods have been developed to distinguish the effects of age structure from those of disease/mortality rates on regional or temporal differences in health/mortality. One could also argue that this focus on ecological study designs and decomposition has delayed the adoption of causal analysis once individual panel data became available.

We close this chapter with a citation of *Paul W. Holland* (2003: 9-10), who reminds us that it is important to establish methods and their rules for identifying causal effects, but that adherence to these rules alone does not justify the ranking of empirical research into better or worse: "Being able to assert that the association is based on a causal connection is, in many circumstances, merely a status symbol, one that confers importance to the finding without any consequence for improved public health […] it is the use of an association for important purposes that has enduring value and not its status as a causal variable."

Acknowledgements

References

Acock, Alan 2013: Discovering Structural Equation Modeling using Stata: Stata Press.

Albouy, Valerie; Lequien, Laurent 2009: Does compulsory education lower mortality? In: Journal of Health Economics 28,1: 155-168.
https://doi.org/10.1016/j.jhealeco.2008.09.003

Atchley, Robert C. 1975: Adjustment to Loss of Job at Retirement. In: The International Journal of Aging and Human Development 6,1: 17-27.
https://doi.org/10.2190/EHU3-VCRV-VCRJ-04NU

Athey, Susan; Imbens, Guido W. 2006: Identification and inference in nonlinear difference-in-differences models. In: Econometrica 74,2: 431-497.

Baltagi, Badi 2008: Econometric analysis of panel data: John Wiley & Sons.

Barclay, Kieron J.; Kolk, Martin 2018: Birth intervals and health in adulthood: a comparison of siblings using Swedish register data. In: Demography 55,3: 929-955.
https://doi.org/10.1007/s13524-018-0673-8

Barclay, Kieron; Kolk, Martin 2015: Birth order and mortality: a population-based cohort study. In: Demography 52,2: 613-639. https://doi.org/10.1007/s13524-015-0377-2

Behncke, Stefanie 2012: Does retirement trigger ill health? In: Health Economics 21,3: 282-300. https://doi.org/10.1002/hec.1712

Bijlsma, Maarten J. et al. 2017: Unemployment and subsequent depression: A mediation analysis using the parametric G-formula. In: Social Science & Medicine 194: 142-150.
https://doi.org/10.1016/j.socscimed.2017.10.011

Bijlsma, Maarten J.; Wilson, Ben 2020: Modelling the socio-economic determinants of fertility: a mediation analysis using the parametric g-formula. In: Journal of the Royal Statistical Society: Series A (Statistics in Society) 183,2: 493-513.
https://doi.org/10.1111/rssa.12520

Bollen, Kenneth A. 1989: Structural Equations with Latent Variables. New York, NY: Wiley-Interscience.

Bollen, Kenneth A.; Curran, Patrick J. 2006: Latent curve models: A structural equation perspective: John Wiley & Sons.

Brand, Jennie E. et al. 2019: Uncovering Sociological Effect Heterogeneity using Machine Learning. arXiv preprint.

Brüderl, Josef; Ludwig, Volker 2015: Fixed-effects panel regression. In: *Best, Henning; Wolf, Christof*: The Sage handbook of regression analysis and causal inference. SAGE Publications Ltd: 327-357.

Burdorf, Alex 2012: The need for novel strategies to analyze the dynamic pattern of worker's health over time and the consequences for sustained employability. In: Scandinavian Journal of Work, Environment & Health 38,6: 485-488.
https://doi.org/10.5271/sjweh.3325

Calvo, Esteban; Sarkisian, Natalia; Tamborini, Christopher R. 2013: Causal effects of retirement timing on subjective physical and emotional health. In: The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 68,1: 73-84. https://doi.org/10.1093/geronb/gbs097

Cameron, A. Colin; Trivedi, Pravin K. 2005: Microeconometrics: methods and applications: Cambridge university press.

Carnegie, Nicole Bohme; Harada, Masataka; Hill, Jennifer L. 2016: Assessing sensitivity to unmeasured confounding using a simulated potential confounder. In: Journal of Research on Educational Effectiveness 9,3: 395-420. https://doi.org/10.1080/19345747.2015.1078862

Chandola, Tarani et al. 2006: Pathways between education and health: a causal modelling approach. In: Journal of the Royal Statistical Society: Series A (Statistics in Society) 169,2: 337-359. https://doi.org/10.1111/j.1467-985X.2006.00411.x

Chen, Feinian et al. 2015: For better or worse: The health implications of marriage separation due to migration in rural China. In: Demography 52,4: 1321-1343. https://doi.org/10.1007/s13524-015-0399-9

Coe, Norma B.; Zamarro, Gema 2011: Retirement effects on health in Europe. In: Journal of Health Economics 30,1: 77-86. https://doi.org/10.1016/j.jhealeco.2010.11.002

Couch, Kenneth A.; Tamborini, Christopher R.; Reznik, Gayle L. 2015: The long-term health implications of marital disruption: Divorce, work limits, and social security disability benefits among men. In: Demography 52,5: 1487-1512. https://doi.org/10.1007/s13524-015-0424-z

Doblhammer, Gabriele; Oeppen, Jim 2003: Reproduction and longevity among the British peerage: the effect of frailty and health selection. In: Proceedings of the Royal Society of London. Series B: Biological Sciences. 270,1524: 1541-1547. https://doi.org/10.1098/rspb.2003.2400

Eibich, Peter 2015: Understanding the effect of retirement on health: Mechanisms and heterogeneity. In: Journal of Health Economics 43: 1-12. https://doi.org/10.1016/j.jhealeco.2015.05.001

Einiö, Elina; Nisén, Jessica; Martikainen, Pekka 2016: Number of children and later-life mortality among Finns born 1938-50. In: Population Studies 70,2: 217-238. https://doi.org/10.1080/00324728.2016.1195506

Engelhardt, Henriette; Kohler, Hans-Peter; Prskawetz, Alexia 2009: Causal analysis in population studies. In: *Engelhardt, Henriette; Kohler, Hans-Peter; Fürnkranz-Prskawetz, Alexia* (Eds.): Causal Analysis in Population Studies. Concepts, Methods, Applications. Dordrecht: Springer: 1-7. https://doi.org/10.1007/978-1-4020-9967-0_1

Foverskov, Else; Holm, Anders 2016: Socioeconomic inequality in health in the British household panel: Tests of the social causation, health selection and the indirect selection hypothesis using dynamic fixed effects panel models. In: Social Science & Medicine 150: 172-183. https://doi.org/10.1016/j.socscimed.2015.12.021

Fretheim, Atle et al. 2013: Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. In: Journal of Clinical Epidemiology 66,8: 883-887. https://doi.org/10.1016/j.jclinepi.2013.03.016

Frijters, Paul; Haisken-DeNew, John P.; Shields, Michael A. 2005: The causal effect of income on health: evidence from German reunification. In: Journal of Health Economics 24,5: 997-1017. https://doi.org/10.1016/j.jhealeco.2005.01.004

Galama, Titus J.; van Kippersluis, Hans 2018: A Theory of Socio-economic Disparities in Health over the Life Cycle. In: The Economic Journal 129,617: 338-374. https://doi.org/10.1111/ecoj.12577

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

Gathmann, Christina; Jürges, Hendrik; Reinhold, Steffen 2015: Compulsory schooling reforms, education and mortality in twentieth century Europe. In: Social Science & Medicine 127: 74-82. https://doi.org/10.1016/j.socscimed.2014.01.037

Giesecke, Matthias 2019: The retirement mortality puzzle: Evidence from a regression discontinuity design. In: Ruhr Economic Papers #800. https://doi.org/10.4419/86788928

Gilman, Stephen E.; Loucks, Eric B. 2014: Another casualty of sibling fixed-effects analysis of education and health: An informative null, or null information? In: Social Science & Medicine 118: 191-193. https://doi.org/10.1016/j.socscimed.2014.06.029

Goldman, Noreen 2001: Mortality Differentials: Selection and Causation. In: International Encyclopedia of Social and Behavioral Sciences (Second Edition): 851-856. https://doi.org/10.1016/B978-0-08-097086-8.31041-8

Grundy, Emily M. D.; Read, Sanna; Väisänen, Heini 2020: Fertility trajectories and later-life depression among parents in England. In: Population Studies 74,2: 219-240. https://doi.org/10.1080/00324728.2019.1649450

Halleröd, Björn; Gustafsson, Jan-Eric 2011: A longitudinal analysis of the relationship between changes in socio-economic status and changes in health. In: Social Science & Medicine 72,1: 116-123. https://doi.org/10.1016/j.socscimed.2010.09.036

Hanemann, Felizia 2017: Working Conditions, Retirement and Health – Longitudinal Evidence from Europe and the US. MEA Discussion Paper 07-2017. Munich: Munich Center for the Economics of Aging. http://dx.doi.org/10.2139/ssrn.3060413

Hank, Karsten; Steinbach, Anja 2018: Families and health: A review. In: *Doblhammer, Gabriele; Gumà, Jordi* (Eds.): A demographic perspective on gender, family and health in Europe. Cham: Springer: 23-39. https://doi.org/10.1007/978-3-319-72356-3_3

Harper, Sam et al. 2014: The effect of mandatory seat belt laws on seat belt use by socioeconomic position. In: Journal of Policy Analysis and Management 33,1: 141-161. https://doi.org/10.1002/pam.21735

Hessel, Philipp 2016: Does retirement (really) lead to worse health among European men and women across all educational levels? In: Social Science & Medicine 151: 19-26. https://doi.org/10.1016/j.socscimed.2015.12.018

Hoffmann, Rasmus; Kröger, Hannes; Geyer, Siegfried 2019: Social causation versus health selection in the life course: Does their relative importance differ by dimension of SES? In: Social Indicators Research 141,3: 1341-1367. https://doi.org/10.1007/s11205-018-1871-x

Hoffmann, Rasmus; Kröger, Hannes; Pakpahan, Eduwin 2018: Pathways between socioeconomic status and health: Does health selection or social causation dominate in Europe? In: Advances in Life Course Research 36: 23-36. https://doi.org/10.1016/j.alcr.2018.02.002

Holland, Paul W. 2003: Causation and Race. In: Educational Testing Service (ETS), Research Report Series, RR-03-03. Princeton NJ: Research & Development Division. https://doi.org/10.1002/j.2333-8504.2003.tb01895.x

Hsiao, Cheng 2014: Analysis of panel data: Cambridge University Press.

Hu, Yannan et al. 2017: Assessing the impact of natural policy experiments on socioeconomic inequalities in health: how to apply commonly used quantitative analytical methods? In: BMC Medical Research Methodology 17,1: 68. https://doi.org/10.1186/s12874-017-0317-5

Imbens, Guido W. 2004: Nonparametric estimation of average treatment effects under exogeneity: A review. In: Review of Economics and statistics 86,1: 4-29. https://doi.org/10.1162/003465304323023651

Imbens, Guido W; Lemieux, Thomas 2008: Regression discontinuity designs: A guide to practice. In: Journal of Econometrics 142,2: 615-635. https://doi.org/10.1016/j.jeconom.2007.05.001

Insler, Michael 2014: The health consequences of retirement. In: Journal of Human Resources 49,1: 195-233. https://doi.org/10.1353/jhr.2014.0000

Jokela, Markus et al. 2010: From midlife to early old age: health trajectories associated with retirement. In: Epidemiology 21,3: 284-290. https://doi.org/10.1097/EDE.0b013e3181d61f53

Jones, Andrew M.; Rice, Nigel 2011: Econometric evaluation of health policies. In: *Glied, Sherry; Smit, Peter C.*: The Oxford handbook of health economics. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199238828.013.0037

Kibele, Eva U. B. 2014: Individual-and area-level effects on mortality risk in Germany, both East and West, among male Germans aged 65+. In: International Journal of Public Health 59,3: 439-448. https://doi.org/10.1007/s00038-013-0480-3

Kippersluis, Hans van 2010: Understanding Socioeconomic Differences in Health. An Economic Approach: Rozenberg Publishers.

Koch, Gary G.; Gillings, Dennis B. 2006: Inference, Design-Based vs. Model-Based. Wiley Encyclopedia of Statistical Sciences: 1-5.

Kreft, Daniel; Doblhammer, Gabriele 2012: Contextual and individual determinants of health among Aussiedler and native Germans. In: Health & Place 18,5: 1046-1055. https://doi.org/10.1016/j.healthplace.2012.05.008

Kröger, Hannes; Pakpahan, Eduwin; Hoffmann, Rasmus 2015: What causes health inequality? A systematic review on the relative importance of social causation and health selection. In: European Journal of Public Health 25,6: 951-960. https://doi.org/10.1093/eurpub/ckv111

Lin, Sheng-Hsuan et al. 2017: Parametric mediational g-formula approach to mediation analysis with time-varying exposures, mediators, and confounders. In: Epidemiology 28,2: 266-274. https://doi.org/10.1097/EDE.0000000000000609

Lindahl, Mikael 2005: Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. In: Journal of Human Resources 40,1: 144-168.

Lleras-Muney, Adriana 2005: The Relationship Between Education and Adult Mortality in the United States. In: The Review of Economic Studies 72,1: 189-221. https://doi.org/10.1111/0034-6527.00329

Ludwig, Jens; Miller, Douglas L. 2007: Does Head Start improve children's life chances? Evidence from a regression discontinuity design. In: The Quarterly Journal of Economics 122,1: 159-208. https://doi.org/10.1162/qjec.122.1.159

Madsen, Mia et al. 2010: Does educational status impact adult mortality in Denmark? A twin approach. In: American Journal of Epidemiology 172,2: 225-234. https://doi.org/10.1093/aje/kwq072

Michaud, Pierre-Carl; van Soest, Arthur 2008: Health and wealth of elderly couples: causality tests using dynamic panel data models. In: Journal of Health Economics 27,5: 1312-1325. https://doi.org/10.1016/j.jhealeco.2008.04.002

Mills, Melinda C.; Barban, Nicola; Tropf, Felix C. 2020: An Introduction to Statistical Genetic Data Analysis. MIT Press.

Moffitt, Robert 2005: Remarks on the analysis of causal relationships in population research. In: Demography 42,1: 91-108. https://doi.org/10.1353/dem.2005.0006

Moffitt, Robert A. 2009: Issues in the estimation of causal effects in population research, with an application to the effects of teenage childbearing. In: *Engelhardt, Henriette; Kohler, Hans-Peter; Fürnkranz-Prskawetz, Alexia* (Eds.): Causal analysis in population studies. Concepts, Methods, Applications. Dordrecht: Springer: 9-29. https://doi.org/10.1007/978-1-4020-9967-0_2

Monden, Christiaan W. S.; Uunk, Wilfred J. G. 2013: For better and for worse: The relationship between union dissolution and self-assessed health in European panel data. In: European Journal of Population/Revue europeenne de demographie 29,1: 103-125. https://doi.org/10.1007/s10680-012-9269-y

Morgan, Stephen L.; Todd, Jennifer J. 2008: 6. A Diagnostic Routine for the Detection of Consequential Heterogeneity of Causal Effects. In: Sociological Methodology 38,1: 231-282. https://doi.org/10.1111/j.1467-9531.2008.00204.x

Nerlove, Marc 2005: Essays in panel data econometrics: Cambridge University Press.

O'Flaherty, Martin et al. 2016: The family life course and health: Partnership, fertility histories, and later-life physical health trajectories in Australia. In: Demography 53,3: 777-804. https://doi.org/10.1007/s13524-016-0478-6

Oksanen, Tuula; Virtanen, Marianna 2012: Health and retirement: a complex relationship. In: European Journal of Ageing 9,3: 221-225. https://doi.org/10.1007/s10433-012-0243-7

Pakpahan, Eduwin; Hoffmann, Rasmus; Kröger, Hannes 2017: The long arm of childhood circumstances on health in old age: Evidence from SHARELIFE. In: Advances in Life Course Research 31: 1-10. http://dx.doi.org/10.1016/j.alcr.2016.10.003

Pakpahan, Eduwin; Hoffmann, Rasmus; Kröger, Hannes 2015: Statistical Methods for Causal Analysis in Life Course Research: An illustration of a Cross Lagged Structural Equation Model, a Latent Growth Model, and an autoregressive latent trajectories model. In: International Journal of Social Research Methodology 20,1: 1-19. https://doi.org/10.1080/13645579.2015.1091641

Palloni, Alberto et al. 2009: Early childhood health, reproduction of economic inequalities and the persistence of health and mortality differentials. In: Social Science & Medicine 68,9: 1574-1582. https://doi.org/10.1016/j.socscimed.2009.02.009

Radó, Márta; Boissonneault, Michaël 2018: Short and long-term change in subjective well-being among voluntary and involuntary retirees. In:The Journal of the Economics of Ageing 17, 100178. https://doi.org/10.1016/j.jeoa.2018.11.003

Read, Sanna L.; Grundy, Emily M. D. 2017: Fertility history and cognition in later life. In: Journals of Gerontology Series B: Psychological Sciences and Social Sciences 72,6: 1021-1031. https://doi.org/10.1093/geronb/gbw013

Robins, James M.; Hernán, Miguel A. 2009: Estimation of the causal effects of time-varying exposures. In: *Fitzmaurice, Garrett et al.* (Eds.): Longitudinal data analysis: Chapman & Hall/CRC: 553-599

Rosenbaum, Paul R.; Rubin, Donald B. 1983: The central role of the propensity score in observational studies for causal effects. In: Biometrika 70,1: 41-55. https://doi.org/10.1093/biomet/70.1.41

Schaap, Rosanne et al. 2018: The effects of exit from work on health across different socioeconomic groups: A systematic literature review. In: Social Science & Medicine 198: 36-45. https://doi.org/10.1016/j.socscimed.2017.12.015

Schunck, Reinhard 2013: Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. In: The Stata Journal 13,1: 65-76. https://doi.org/10.1177/1536867X1301300105

Schuring, Merel et al. 2015: Educational differences in trajectories of self-rated health before, during, and after entering or leaving paid employment in the European workforce. In: Scandinavian Journal of Work, Environment & Health 41,5: 441-450. https://doi.org/10.5271/sjweh.3514

Shim, Matthew J. et al. C 2013: A Systematic Review of Retirement as a Risk Factor for Mortality. In: *Hoque, Nazrul; McGehee, Mary A.; Bradshaw, Benjamin S.* (Eds.): Applied Demography and Public Health. Dordrecht: Springer Netherlands: 277-309. https://doi.org/10.1007/978-94-007-6140-7_17

Sjölander, Arvid et al. 2016: Carryover Effects in Sibling Comparison Designs. In: Epidemiology 27,6: 852-858. https://doi.org/10.1097/EDE.0000000000000541

Sjölander, Arvid; Zetterqvist, Johan 2017: Confounders, Mediators, or Colliders: What Types of Shared Covariates Does a Sibling Comparison Design Control For? In: Epidemiology 28,4: 540-547. https://doi.org/10.1097/EDE.0000000000000649

Smith, Herbert L. 2009: Causation and its discontents. In: *Engelhardt, Henriette; Kohler, Hans-Peter; Fürnkranz-Prskawetz, Alexia* (Eds.): Causal analysis in population studies. Dordrecht: Springer: 233-242. https://doi.org/10.1007/978-1-4020-9967-0_10

Smith, Herbert L. 2013: Research Design: Toward a Realistic Role for Causal Analysis. In: *Morgan, Stephen L.* (Ed.): Handbook of Causal Analysis for Social Research. Dordrecht: Springer Netherlands: 45-73. https://doi.org/10.1007/978-94-007-6094-3_4

Smith, James P. 1999: Healthy bodies and thick wallets: the dual relation between health and economic status. In: Journal of Economic Perspectives 13,2: 145-166. https://doi.org/10.1257/jep.13.2.145

Smith, James P. 2004: Unraveling the SES-health connection. In: Population and Development Review 30, supplement, Aging, health, and public policy: demographic and economic perspectives: 108-132.

Stenholm, Sari et al. 2014: Age-related trajectories of physical functioning in work and retirement: the role of sociodemographic factors, lifestyle and disease. In: Journal of Epidemiology and Community Health 68,6: 503-509. https://doi.org/10.1136/jech-2013-203555

Taljaard, Monica et al. 2014: The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. In: Implementation Science 9,77. https://doi.org/10.1186/1748-5908-9-77

van der Heide, Iris et al. 2013: Is retirement good for your health? A systematic review of longitudinal studies. In: BMC Public Health 13,1180. https://doi.org/10.1186/1471-2458-13-1180

van Rijn, Rogier M. et al. 2014: Influence of poor health on exit from paid employment: a systematic review. In: Occupational and Environmental Medicine 71,4: 295-301. https://doi.org/10.1136/oemed-2013-101591

VanderWeele, Tyler 2015: Explanation in causal inference: methods for mediation and interaction: Oxford University Press.

Wagner, Anita K. et al. 2002: Segmented regression analysis of interrupted time series studies in medication use research. In: Journal of Clinical Pharmacy and Therapeutics 27,4: 299-309. https://doi.org/10.1046/j.1365-2710.2002.00430.x

Wang, Jichuan; Wang, Xiaoqian 2012: Structural Equation Modeling: Applications Using Mplus. Chichester, West Sussex England; Hoboken N.J: Wiley.

Warren, John Robert 2009: Socioeconomic Status and Health across the Life Course: A Test of the Social Causation and Health Selection Hypotheses. In: Social Forces 87,4: 2125-2153. https://doi.org/10.1353/sof.0.0219

Westerlund, Hugo et al. 2009: Self-rated health before and after retirement in France (GAZEL): a cohort study. In: The Lancet 374,9705: 1889-1896. https://doi.org/10.1016/S0140-6736(09)61570-1

Wooldridge, Jeffrey M. 2002: Econometric analysis cross section panel. MIT press Cambridge, MA.

Prof. Dr. Rasmus Hoffmann(✉). Universität Rostock. Rostock, Germany. University of Siegen. Siegen, Germany. E-mail: rasmus.hoffmann@eui.eu
URL: https://www.uni-siegen.de/lwf/professuren/mse/mitarbeiter/hoffmann.html?lang=de

Prof. Dr. Gabriele Doblhammer. Universität Rostock. Rostock, Germany. German Centre for Neurodegenerative Diseases (DZNE). Bonn, Germany
E-mail: gabriele.doblhammer@uni-rostock.de
URL: https://www.isd.uni-rostock.de/doblhammer/