Genus

Causal assessment in demographic research

Guillaume Wunsch[1,2]* and Catherine Gourbin[1]

* Correspondence: g.wunsch@
uclouvain.be
[1]Centre for Research in
Demography, UCLouvain, Place
Montesquieu 1/L2.08.03, B-1348
Louvain-la-Neuve, Belgium
[2]Royal Academy of Belgium,
Brussels, Belgium

Abstract

Causation underlies both research and policy interventions. Causal inference in demography is however far from easy, and few causal claims are probably sustainable in this field. This paper targets the assessment of causality in demographic research. It aims to give an overview of the methodology of causal research, pointing out various problems that can occur in practice. The "Intervention studies" section critically examines the so-called *gold standard* in causality assessment in experimental studies, randomized controlled trials, and the use of quasi-experiments and interventions in observational studies. The "Multivariate statistical models" section deals with multivariate statistical models linking a mortality or fertility indicator to a series of possible causes and controls. Single and multiple equation models are considered. The "Mechanisms and structural causal modelling" section takes into account a more recent trend, i.e., mechanistic explanations in causal research, and develops a structural causal modelling framework stemming from the pioneering work of the Cowles Commission in econometrics and of Sewall Wright in population genetics. The "Assessing causality in demographic research" section examines how causal analysis could be further applied in demographic studies, and a series of proposals are discussed for this purpose. The paper ends with a conclusion pointing out, in particular, the relevance of structural equation models, of triangulation, and of systematic reviews for causal assessment.

Keywords: Causality, Structural modelling, Randomized trials, Counterfactuals, Multivariate statistical models, Causal graphs

Introduction

This paper targets the assessment of causality in demographic research. It proposes and discusses various recommendations for improving research aiming at causal inference. The text can be considered as a methodological support for demographers interested in causal analysis.

Causal studies are not only important for understanding and explaining a given phenomenon, such as the recent decrease in life expectancy in the USA (Woolf and Schoomaker, 2019), but also for adopting better policy actions, such as developing more efficient public health policies. A famous historical example, pointed out by Cameron and Jones (1983), is John Snow's study of cholera in London in 1854, where he linked the disease to the quality of the water supplies. Causation underlies both

research and policy interventions. Causal inference in demography is however far from easy and few causal claims are probably sustainable in these fields, especially outside the temporal and spatial context for which they are proposed. Maíre Ní Bhrolcháin has clearly shown all the problems that can occur, in her study of the effects of parental divorce on children (Ní Bhrolcháin, 2001). This restriction should not however hamper research on the ways of improving causal assessment in demography, the purpose of this paper. We start by painting the context.

Description

For most of its history, demography has chiefly dealt with descriptive studies focusing, for example, on fertility differences over time or area, or on projecting population characteristics in the future, such as death rates by age and sex. In many cases, some form of qualitative assessment of causality has been attempted by associating, for example, mortality differences by regions with the socio-economic characteristics of these regions or with selective migration. M. Barbieri (2013), for instance, pays lip service to these possible determinants of regional mortality in France, but does not actually attempt establishing the causes of the differentials observed. It is however clear that causal inference was not the purpose of her paper. To give another more recent example, Baptista and Queiroz (2019) have investigated the relation between CVD mortality and economic development, measured by gross domestic product per capita, in Brazilian micro-regions from 2001 to 2015. They end the paper by stating that their goal was not to investigate the causal relationship between CVD mortality and GDP per capita but to raise and examine some research questions regarding the association between socioeconomic factors, measured by GDP per capita, and CVD mortality. Numerous other examples could be given.

Aggregate data

In past times, in the absence of powerful computers and having recourse to aggregate data for the most time, the individual determinants of fertility, mortality, or migration could not be identified. Correlations could be established at the aggregate level but the extrapolation to individual behaviors, at the micro level, runs into the well-known problem of *ecological fallacies* (see, e.g., Lopez Rios and Wunsch, 1990). These do not however affect studies at the macro level of analysis, such as looking for the impact of the health care system on regional mortality differences. Both variables are, in this case, at the macro level of analysis; for an example, see Lopez Rios et al. (1992). Based on a series of examples in demography, Ní Bhrolcháin and Dyson (2007) make a strong case for looking at causation at the aggregate level of analysis, using a set of criteria supportive of causal inference.

Individual data

The situation drastically changed with the advent of computers and of *retrospective* surveys in the study of fertility and migration. One could now examine micro data and relate individual fertility or migration histories to the characteristics of the individuals obtained from the surveys. Of course, individual retrospective surveys are of no use in mortality studies: *dead men tell no tales*, as the saying goes. Indirect measures of infant

and child mortality can nevertheless be obtained from proxies. Retrospective surveys can however be used in morbidity and health studies. In the field of mortality and morbidity, various *prospective* longitudinal surveys were conducted in epidemiology, such as the *Framingham Heart Study* or the Doll and Hill prospective study among British doctors showing the link between tobacco consumption and lung cancer (Doll and Hill, 1964). Moreover, *record linkage* of censuses, surveys, and registers has more or less recently become available for longitudinal research at the micro level (Wunsch and Gourbin, 2018).

What is a cause?

Though these population studies usually mention searching for the *determinants* or *factors* of a phenomenon, they are in fact often looking for *cause/effect* relations. Is *cause* such a dirty word that demographers balk at using it? In this omission, demographers are however not alone. Pearl and Mackenzie (2018, p. 11) have indeed written that "Despite heroic efforts by the geneticist Sewall Wright (1889–1988), causal vocabulary was virtually prohibited for more than half a century." But what do we mean by cause, in particular in observational studies? Very succinctly, if a variation in X produces (i.e., increases the probability of) a variation in Y, *and* if one can explain *why or how* ΔX produces ΔY, then one can postulate that X is a (probabilistic) cause of the effect Y. In addition, actual causes should precede their effects in time. In other words, one must describe the intelligible mechanism and its parts, or sub-mechanisms, intervening between cause X and effect Y (Glennan, 2011). More specifically, according to a current definition, a mechanism consists of entities and activities organized in such a way that they are responsible for the phenomenon (Illari and Williamson, 2012). For causal assessment, we need both difference—making and mechanistic knowledge. More on this in F. Russo (2009, 2014), and in the following sections.

Outline

The outline of the paper is as follows. The aim is to examine to what extent various methods of causal assessment that are common in other sciences can possibly be used in demography. The "Intervention studies" section examines the so-called *gold standard* in causality assessment in experimental studies, i.e., randomized controlled trials, and the use of quasi-experiments in observational studies. For many scientists and philosophers, only interventions (or manipulations) can tell us if a variable is really a cause or not. The "Multivariate statistical models" section deals with multivariate statistical models linking, for example, a mortality or fertility indicator to a series of possible causes and controls. Single and multiple equation models are considered. The "Mechanisms and structural causal modelling" section takes into account a more recent trend, i.e., mechanistic explanations in causal research, and develops a structural causal modelling framework stemming from the pioneering work of the Cowles Commission in econometrics and of Sewall Wright in population genetics. The "Assessing causality in demographic research" section proposes a series of recommendations that could possibly lead to improving causal analysis in demographic studies and, more generally, in other social sciences. The paper ends with a conclusion pointing out, in particular, the relevance of structural equation models, of triangulation, and of systematic reviews for causal assessment.

For further reading, references are given throughout the text to some of our previous papers in the field of causal analysis[1], based on a structural causal modelling framework. The present paper thus reflects, to some extent, our past work in this domain. Of course, other relevant approaches to causal inference, such as qualitative research methods, can be found in the literature.

Intervention studies

Randomized controlled trials

The best-known example of an intervention study in experimental research is the randomized controlled trial (RCT). For example, in order to test a new drug against a disease, a sample of patients—rigorously selected according to well-defined criteria—is randomly divided into two groups, a treatment group and a control group. The first group receives the new drug and the other either a placebo or the best alternative available treatment. The outcome (recovery, in this case) is then compared between the two groups to see if the outcome is (statistically) significantly better in the treatment group than in the control group, i.e., if the new drug is better than a placebo[2] or the best treatment currently available. A rather sophisticated example of an RCT, the *Sure Outcome of Random Effects* (SORE) model—taking into account both observed pharmacological and residual effects, and for each effect two latent factors—is given in Mouchart et al. (2019). RCTs have been used in population research to test for instance effective contraceptive use (Melnick et al., 2016). Numerous RCTs have also been conducted in Africa to examine the effect of different types of interventions on HIV prevention. These studies have been critically examined by David Gisselquist (2013) who shows that, for various reasons stated in the title of his paper, these RCTs have been insufficient to guide HIV prevention.

As patients are randomly allocated in an RCT between the two groups, in a very large sample the only difference between the groups would be treatment versus no treatment. In other words, RCTs lead to closely matched groups. The method therefore controls for possible latent confounders. Though RCTs are considered the gold standard for testing cause/outcome relations, not only in epidemiology but also in statistics or in econometrics, the method is not without problems (Stock and Watson, 2003; Deaton and Cartwright, 2018). In particular, the method is *firstly* not feasible or ethical in many circumstances. *Secondly*, samples are usually small and the two groups can differ from one another by chance due to factors other than treatment. *Thirdly*, results in the real world can be quite different from those obtained in laboratory experiments. For example, oral contraceptives tested by RCTs would yield a contraceptive effectiveness close to 100%. This is not the case in the real world due to possible poor compliance. *Fourthly*, results are only valid for the sample and can vary across subgroups. Moreover, in another population, outcomes might be different. This is of course an issue not only for RCTs but relates to the external validity—beyond the population of reference—of all demographic studies, which are always context-dependent. *Lastly*,

---

[1]In particular to the works of the first author in collaboration with Federica Russo (Philosophy of Science) and Michel Mouchart (Statistics).

[2]Remember that a placebo effect is not a pure psychological artifact but can bring about a biological response (see the recent case file in Mérat, 2019).

RCTs do not give us the mechanism leading from the treatment or cause to the outcome. The link between cause and outcome actually remains a black box[3].

Natural or quasi-experiments

Most studies in demography in particular cannot rely on experiments, but in some cases, one can have resort to natural or quasi-experiments. For example, in a public health perspective, Chattopadhyay and Duflo (2004) have examined if the election of a woman leader leads to a better provision of water, by taking advantage of an Indian government reservation policy stipulating that in one third of the Village Councils in India, randomly selected, the leader must be a woman. Causality can run both ways however (Basu, 2014): women can have a greater sense of social responsibility, compared to males, and once elected ensure that villagers get good water, or villages with better public goods may tend to elect more women as leaders. Thanks to the natural experiment, Chattopadhyay and Duflo were able to rule out the hypothesis that causality ran from good provision of water to women being elected leaders. As Village Councils were randomly selected to be reserved for women, differences in investment decisions can be attributed to the reserved status of those Village Councils.

Natural experiments following an intervention have been used in demography, in the field of morbidity and mortality research among others (see a partial overview in MRC, 2012). For example, Herttua et al. (2008), using register data, have studied changes in alcohol-related mortality, in the Finnish population aged 15 and over, after a large reduction in alcohol prices. In Sri Lanka, mortality from suicide by self-poisoning with pesticide has been examined after legal restriction on pesticide imports (Gunnell et al., 2007). In a similar vein, Blum and Monnier (1989) have shown the impact on Russian mortality of the drastic measures taken by President Gorbachev to curb alcohol consumption, alcohol-related deaths in the USSR being an important feature of the high death rates from accidents and violence. With continuous variables, a *regression discontinuity* quasi-experimental design can sometimes be used by taking advantage of a cutoff or threshold in the putative cause. For example, Ludwig and Miller (2007) have exploited a discontinuity in the "Head Start" program funding across US counties, by virtue of the grant-writing assistance given to just the poorest 300 counties, for examining mortality rates for children from causes that could be affected by the health services offered by the program.

Though natural experiments should be used if possible in observational studies, they are not perfect substitutes for RCTs. First, as the MRC report (2012) states: "Only a small proportion of the 'multitude of promising initiatives' are likely to yield good natural experimental studies." Furthermore, contrary to RCTs, they do not perfectly get rid of possible latent confounding, as the assignment of individuals before and after the intervention is not truly carried out at random. Selective exposure to the intervention remains a problem.

Counterfactuals and manipulation

The randomized trial in experimental studies has led Donald Rubin (1974) to extend the design to observational studies, with his counterfactual approach to causation based

---

[3]Thus, the inclusion of latent sub-mechanisms in the *SORE* model referred to above.

on potential outcomes. A typical question would be: "If I had taken an aspirin an hour ago, would my headache (after reading this paper) be gone?" As I cannot both take an aspirin (counterfact) and not take an aspirin (fact), the correct answer cannot be given. But if I find someone very similar to me who has taken an aspirin, I can compare the outcomes of the two situations. The reasoning can be extended to a multiple number of treatments and to a population of individuals. For this purpose, Rubin has developed the technique of *propensity scores* in order to match controls to cases, as best as one can (Rosenbaum and Rubin, 1983).

The counterfactual approach is widely accepted at present, though it is not without problems; see for instance Russo et al. (2011). One of the problems is that Rubin requires that all subjects be potentially exposable to the various k treatments, i.e., causes, including no treatment. In this approach, "causes are only those things that could, in principle, be treatments in experiments" (Holland, 1986). An attribute (such as ethnicity) cannot therefore be a cause, because potential exposability cannot apply to it. In other words, we cannot change ethnicity in a subject in order to see if this change has an impact on an outcome. One could, of course, examine in each ethnic group if the outcome has the same causes but, in this case, one would actually be conditioning on and controlling for ethnicity.

The manipulation/intervention criterion is however not satisfactory. We know, for example, that in a population mortality and health differ by sex or ethnicity, which are to some extent the causes of these differentials, though we cannot change them at the individual level. The manipulation or intervention approach to causality, proposed by Woodward (2003) among others, cannot therefore be a sound basis for causal research in demography (see also Ní Bhrolcháin and Dyson, 2007). Moreover, following Robert E. Lucas' criticism (Lucas, 1976), the manipulation of one variable in a system can lead to changes in the other variables and in the system itself, in particular when the intervention is operated under a change of policy active on the global mechanism. This criticism can also be addressed in principle to Pearl's *do-operator*[4] (Pearl, 2000) in his *directed acyclic graphs* approach to causality. For this causal criterion to hold, the intervention must indeed meet a series of requirements (see Woodward, 2016). Finally, as Vandenbroucke et al. (2016) have stressed, the counterfactual/manipulation approach does not take into account the need to integrate diverse types of evidence to assess causality.

Demographers are rarely in a situation where they can conduct an experiment, such as an RCT, take advantage of a natural one, or manipulate the cause (except virtually). The following sections are therefore dedicated to the assessment of cause-effect relations in observational studies without interventions.

Multivariate statistical models
Single equation models
Much research in demography has recourse to single multivariate equation models where an outcome (or "dependent") variable Y is related to a set of explanatory (or "independent") variables X_i through some functional form of relation f:

---

[4]Meaning *'given that we do…'*, or similarly *'set variable X to x'*.

$$Y = f(X_i) + \varepsilon$$

The so-called error term ε stands for the variables influencing Y that are not included in the model, in other words for the fact that a model is never a perfect representation of the data. Some of these latter variables, say Z_i, may however be associated, in addition to Y, with some of the X_i and have to be controlled for in order to avoid loss of exogeneity in the explanatory variables X_i. One also says that the Z_i confound in this case the relation between the X_i and the outcome Y. Controlling means here conditioning on the Z_i, and the final model becomes

$$Y = f(X_i, Z_i) + \varepsilon$$

For example, Green and Hamilton (2019) investigate, using single logistic regression models for infant, neonatal, and postneonatal mortality, whether maternal education-infant mortality gradients vary by race/ethnicity among infants from US-born and foreign-born mothers. In addition, they include controls for maternal characteristics, such as maternal age and marital status.

Though widely used[5], in often intricate frameworks, single equation models can be criticized from a causal point of view. First, they do not spell out the structure of relationships among the variables, though interaction effects are often considered. In single equation models, it is as if different structures of association among the variables had no impact on the generation of the outcome variable, a doubtful hypothesis indeed. For example, even a very simple model such as X is a cause of Z and of Y, and Z is a cause of Y, cannot be represented adequately by a sole equation. Secondly, as the structure among the variables is not specified, if one is not careful, single equation models can lead to incorrect controlling, such as conditioning on mediators[6], on colliders[7], or on the other components of a conjunctive cause[8], three errors to avoid. Choosing the appropriate confounding variables one should control for requires specifying the order of relations among the variables; on this issue, see Mouchart et al. (2016). Lastly, single equation models cannot deal with simultaneity issues, such as the "causal circle" (or feedback effect) X causes Y causes X, also called a directed cycle in graph theory. On these grounds, in the study of causes and effects, demographers should consider abandoning the use of single equation models in favor of more complex designs (such as in Bijwaard et al., 2019).

Multiple equation models

For the reasons given above, multiple equations models should be preferred to single equation ones for assessing cause-effect relations in demography. Econometricians have developed *simultaneous equations models* (SEMs) to deal with simultaneity issues, such as supply and demand in a market, or, more generally, in equilibrium systems with simultaneous feedback (Wooldridge, 2013, chapter 16). Simultaneity often results from a lack of information on the ordering of the variables, i.e., from insufficient or inadequate

---

[5]Perusing, as of end November 2019, the papers published during that year in the journals Demographic Research, Population Studies, Population, Genus, European Journal of Population, and Demography.
[6]Such as X_k in X_j causes X_k causes Y.
[7]Such as X_h in X_k and X_j cause X_h.
[8]Such as Z in $X \wedge Z$ causes Y.

data. If the data are aggregated by yearly periods, for instance, events occurring during these periods appear to be simultaneous though, in fact, they are not.

The multiple equations models discussed here are of another type. In this case, the equations represent the structure of the system of variables in a causal perspective (see Bollen and Pearl, 2013). They take into account the asymmetric relations between the variables, each variable in the system being a cause or an outcome of another one. These models are therefore called *structural equations models*[9]. A structural equations model can be represented by a *directed acyclic graph*, or DAG, while a simultaneous equations model usually cannot. Following Pearl (2000), a DAG visually represents the recursive decomposition of a joint distribution, specifically representing the causal structure among variables. A DAG cannot always represent, however, all the characteristics of the structural model. SEMs, on the contrary, usually describe non-causal, and therefore non-recursive, systems. For more on this subject, see the interesting discussion by Strotz and Wold (1960) on recursive versus non-recursive systems.

An excellent overview of structural equations modelling is given in Tarka (2018). Though in favor of this approach, Tarka nevertheless points out some issues concerning, in particular, the understanding of the role of the null hypothesis, the specification of such models and the possible omission of important variables or the inclusion of redundant variables, the testing of the fit of the model, and more generally the need for a strong theoretical background. The following section on mechanisms and sub-mechanisms will develop a more general framework for structural equations modelling, but first two examples of the latter are given below as an illustration of the methodology[10].

Two examples

Lopez Rios et al. (1992) have proposed a structural equations model, using multiple indicators per concept, for examining at the macro level the impact of the health care system on regional adult mortality differences in Spain, for all medical causes of death and by large groups of causes. Six concepts (or latent variables) and 31 indicators (or manifest variables) are taken into account. In this model, mortality depends upon the use of the health care system and the level of social development. The use of the health care system depends upon the population age structure of the regions, the available health infrastructure, and the level of social development. Finally, the latter two variables depend upon the level of regional economic development. This structural model has been estimated using the *LISREL* (for LInear Structural RELations) software developed by Karl Jöreskog and Dag Sörbom.

The second example is drawn from a publication by Gaumé and Wunsch (2010). Using individual data from the Norbalt surveys held in 1994 and 1999 in the three Baltic countries, the authors examine the determinants of self-rated health in the three countries and for the two periods, by way of structural equations modelling and directed acyclic graphs. The model includes as possible determinants of self-rated health: alcohol consumption, physical health, psychological distress, education, locus of control, and social support. The model takes into account the structure of relations among the variables, in particular the direct and indirect paths leading from the various possible determinants (or causes) to the effects (or outcomes). The authors have used

---

[9]In the literature, they are sometimes confusingly also called SEMs.
[10]The reader is referred to the papers for the full results of the models.

Bayesian inference to estimate the parameters of the model. The posterior distributions (posterior probabilities) have been obtained from the data and priors iteratively, using a *Markov Chain Monte Carlo* (MCMC) procedure. This linear recursive structural model has been fitted, for each of the three countries, using the *AMOS* (for Analysis of MOment Structures) software for causal modelling.

Mechanisms and structural causal modelling

A general framework

The two examples given above raise a series of questions shared by many researchers, the following among others. How are the concepts chosen? Can these be translated into measurable indicators? On what basis is the causal network of relations among variables specified? What is the external validity of the model, outside the population of reference? Michel Mouchart (statistics), Federica Russo (philosophy of science), and the first author of the present paper have developed over the past years a general framework for structural causal modelling (SCM) in the social sciences. The objective is to make explicit the conditions under which multiple equation models can allow causal inference, including the availability of relevant data. An overview of this approach is given in Russo et al. (2019), on which the present sub-section is based.

The framework stems from the work on structural modelling by the Cowles Commission in econometrics in the early 1950s, and from Sewall Wright's path analysis in population genetics dating from the 1920s. According to Pearl and Mackenzie (2018, p.63), Wright's approach "is a landmark for the history of causality." It was later developed by Judea Pearl with his directed acyclic graphs approach to causation (Pearl, 2000). A close neighbor is *mediation analysis*, for instance in life course studies, that also examines the causal pathways among multiple variables (Daniel and De Stavola, 2019). SCM does not imply an explicit statistical model. In particular, the form of the relations between variables needs to be specified according to the problem and data at hand.

The purpose of SCM is to represent and explain a data generating process (DGP). The main features of the framework are the following.

Causal and structural

Focusing on causal analysis, the SCM approach depends upon reliable background information and evidence for proposing:

- The putative causes of outcomes and effects of causes, including the direct and indirect paths from a cause to an effect,
- The ordering of the variables and their role-function in the mechanism and sub-mechanisms producing the data, as developed in Wunsch et al. (2014),
- And more generally for specifying the intelligible organized structure of relations among variables.

Background knowledge typically involves existing theories concerning the domain of analysis, and theoretical reasoning, but also embraces previous results, preliminary analysis of data (including exploratory data analysis), and the advice of experts. It is on this basis that a preliminary hypothesis is formulated, in a hypothetico-deductive (H-D)

perspective. SCM is thus far from exclusively relying on the associations observed among variables, as would a purely data-driven approach. For a devastating critique of the latter, in particular of automated causal discovery based on the correlations among variables, see Freedman and Humphreys (1999). Causality cannot be assessed solely from associations observed in observational data. A similar critique has been put forward by Dawid (2009), who rejects causal discovery algorithms and DAGs aiming at the extraction of causal conclusions from observationally inferred conditional independencies.

Recursive decomposition and DAG

"Explaining" usually implies decomposing a complex phenomenon in terms of a set of simpler parts. In demography, for instance, this is the purpose of demographic analysis; see, e.g., the *Introduction* to Louis Henry's well-known book on the subject (Henry, 1972). In SCM, the causal explanation is based on a recursive decomposition of the joint distribution of the variables, representing the mechanism generating the data. The joint distribution is expressed as a product of conditional distributions where the conditioning variables form an increasing sequence and where each factor of this product represents a plausible sub-mechanism composed of entities and activities. For this reason, *directed acyclic graphs* (DAGs) provide a privileged tool of representation, though a DAG cannot always fully represent the characteristics of a joint distribution (such as moderator effects).

More formally, if one considers a vector of variables, the joint distribution can be written as:

$$P(X_1, ... X_p) = P(X_1)P(X_2|X_1)...P(X_p|X_1...X_{p-1})$$

Usually, one obtains a condensed or simplified recursive decomposition after retaining only the relevant conditioning variables. Often, however, one cannot achieve a complete decomposition in terms of single variables but in terms of "blocks" of variables. The paper by Wunsch et al. (2018) categorizes the distinct types of block-recursivity and examines the implications of block-recursivity for causal attribution.

Exogeneity and causation

Under a suitable exogeneity condition of non-confounding, one can view the conditioning variables as causes in the sub-mechanism where they appear (Wunsch et al., 2014). This requires in particular that the relevant confounders be controlled for, i.e., conditioned on.

Focusing on distributions

The basic objects of analysis are the set of empirical distributions. Equations are related at best to conditional expectations, although effects of causes may take other ways. To give a trite example, one can obtain the same mean length of life e_0 with different distributions of life-table deaths by age and examine why this is so. The culprit could be different distributions of medical causes of death.

Explanation and parametrization

In SCM, the explanation is based on a recursive decomposition. Representing a DGP by a probability distribution implies that this representation leaves unexplained some part of the DGP, namely the stochastic component of the model. Therefore, the

statistical explanation concerns the characteristics, or parameters, of the probability distributions.

Stability or invariance

Considering as structural a mechanism underlying the workings of a DGP requires that the model enjoy suitable properties of invariance under a class of "reasonable"[11] interventions or modifications of the environment. The point here is to look for a proper separation between the incidental and the structural aspects of the DGP. The issue is also that of properly defining the population of reference. A reason for this is that no model in demography, and more generally the social sciences, can pretend to be universal in time and in space. At variance with Kincaid (2004)[12], there are no universal and necessary laws in the social sciences per se, though in demography biological "laws" in the fields of fertility and mortality can be embedded in social processes.

An example

The present example, in the field of reproductive health, is taken form Gourbin et al. (2017). Having recourse to an analysis of *Demographic and Health Survey* (DHS) data, this study examines the causes of contraceptive use in the capital cities of four African countries. The methodology is based on recursive structural causal models represented by directed acyclic graphs. After a comprehensive search of the literature on the topic, discussions with experts, and a thorough description of the sample data, a *conceptual model* (Fig. 1) has first been put forward that reflects the organized network of relations among theoretical concepts.

Based once again on background knowledge, Fig. 2 presents the *operational model* taking this time the available data into account. Figure 2 actually represents a directed acyclic graph (DAG) where each variable or node in the graph depends upon the variables upstream, i.e., upon their "ancestors", in the absence of retroactive or feedback effects (Pearl 2000). Each arrow or link represents a putative causal effect and each endogenous variable (i.e., one that is determined by other variables in the model) is conditioned on its immediate causes or "parents" in the sub-mechanism, i.e., on only those variables that have a direct effect on this endogenous variable. This strategy controls for known confounders and takes possible interaction effects into account (Mouchart et al., 2016). For example, in Fig. 2, one conditions the outcome "contraceptive use" on its immediate or direct causes:

Contraceptive use | man's level of education, approval of family planning, woman's level of education, paid employment in the past 12 months, desire to have children

where the symbol "|" means "conditioned on." One does the same for each outcome in the graph. There are as many equations as there are outcomes. Woman's age at the time of the survey and her socialization environment are regarded as exogenous variables, i.e., they do not depend upon other variables in the model. The variables being in categorical format in this example, logistic regressions are used throughout for parameter estimation in each of the distinct sub-mechanisms.

---

[11]In Haavelmo's words (1944, p. 29).
[12]"Laws are universal in that they do not refer to particular entities" (Kincaid, 2004, p.171). This is never the case in the social sciences, the context or entity always playing a major role.

Fig. 1 A conceptual model of the determinants of contraceptive use
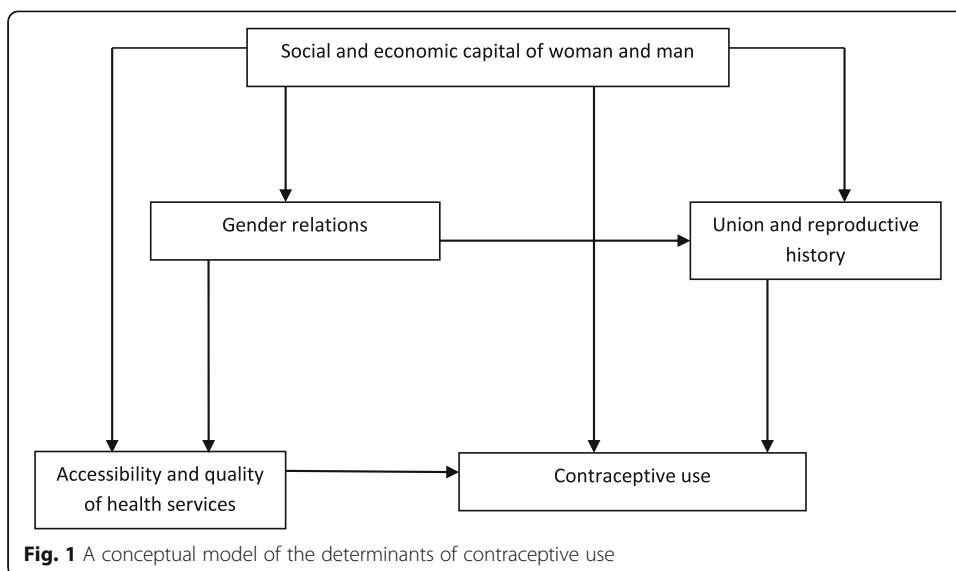
The empirical analysis has confirmed the importance of variables such as education, the desire for children, and partner agreement on family planning in explaining contraceptive use. It has also highlighted a structural union-reproductive indirect path (in bold on the right in the graph) linking female education to contraceptive use. This path was remarkably stable between countries and between the two large age groups considered. The directions of the relations remained the same and were always statistically significant. On the contrary, the analysis led to a tentative rejection of a socio-cultural indirect path (in bold on the left), as the latter was not confirmed by the data available. Possible reasons for this are discussed in the article, in particular the lack of some appropriate indicators in the available data concerning especially the concept "Accessibility and quality of health services" (see Fig. 1).

Fig. 2 An operational model of the determinants of contraceptive use
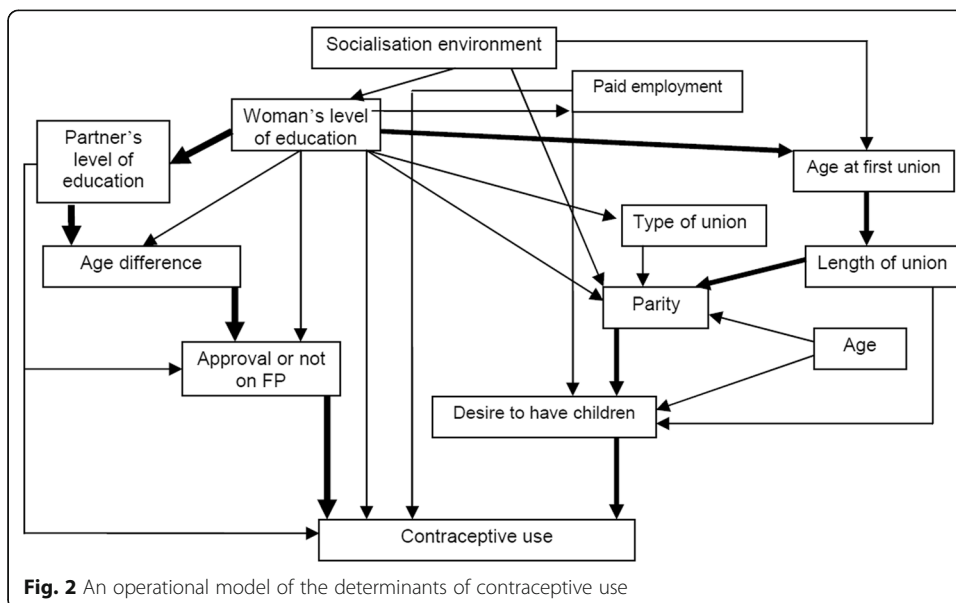
Assessing causality in demographic research

As pointed out in the "Intervention studies" section, Ní Bhrolcháin and Dyson (2007) have put forward and discussed a set of criteria supportive of causal inference. Several of these criteria are comparable to those proposed by Bradford Hill and thoroughly evaluated by Rothman and Greenland (1998)—see Appendix. In this section, we complement this approach by recalling the various steps in a research and by discussing a series of recommendations relative to their implementation in population research.

Research question

Gérard (2006), in particular, has underlined the need for clearly defining the question at the origin of the research, as it is a first step in the formulation of the underlying theory. In her study on divorce effects and causality, Ní Bhrolcháin (2001) has shown that a same question at issue can be understood in more than one way. Among others, is the question raised at the aggregate or at the individual level? The formulation of the question at the origin of the research is dependent upon what one already knows, i.e., upon one's background knowledge (see the "Mechanisms and structural causal modelling" section). The question will be at the basis of the conceptual framework to be developed.

Conceptual framework

The need for developing a conceptual framework has been stressed some decades ago by Hubert Blalock (1968) and more recently by Hubert Gérard (1989, 2006). The purpose of the conceptualization procedure is to organize the information provided by background knowledge, including the critical review of the literature and exploratory data analysis. The main theory should identify, to the best of one's knowledge, the relevant concepts for the problem at hand, the interrelations among these concepts, and specify the direction of these relations (see, e.g., Fig. 1). As concepts are theoretical constructs, they need to be clearly defined in all their dimensions. As a banal example, if one studies inequalities in mortality according to social class, one should define what is meant by the latter, as social class covers multiple dimensions, such as economic capital, cultural capital, social prestige, and social network. The structural relations between the variables should distinguish the putative causes of the outcome considered, from the variables that can confound the possible causal relations. For this, one should spell out the global mechanism and sub-mechanisms responsible for the data generating process.

Operational framework

As stated above, the main theory or conceptual framework is expressed in terms of concepts and relations between concepts. To test this theory, one needs to translate the conceptual framework into an operational framework or auxiliary theory (see, e.g., Fig. 2) where the concepts are represented by observable and potentially measurable variables or indicators (ideally, at least one indicator per dimension of the concept). Once again, this translation should be based on background knowledge, and it depends of course on the availability of relevant data. In some cases, as pointed out at the end of the "Mechanisms and structural causal modelling" section, no suitable indicators are available, and this issue must be discussed in the conclusions of the research as it can hamper the validation of

the theory. In other cases, the dimensions of a same concept may be weakly associated, and the question then is to break up or not the concept into its various more or less independent dimensions.

Structural modelling and DAG

In the "Multivariate statistical models" section, for assessing causality, a preference was given to multiple equations models compared to single equation models. The operational framework will most often show that variables are interrelated according to an organized network. A single equation model cannot represent the latter's complexity. The operational framework should ideally correspond to a recursive decomposition relating to the postulated mechanism and its sub-mechanisms and be represented by a directed acyclic graph (see, e.g., Fig. 2). Of course, this framework requires strong background information on the mechanism and sub-mechanisms involved. As David Freedman (2004, p. 274) has written: "You cannot infer a causal relationship from a data set by running regressions—unless there is substantial prior knowledge about the mechanisms that generated the data." And one needs in addition relevant high-quality data. If this is not achievable, because some of the relations between variables are unknown, recursivity between blocks of variables is nevertheless often possible and a partial causal assessment is feasible, as discussed in Wunsch et al. (2018). In all cases, it is recommended to translate one's theory into a causal graph, even if the latter remains incomplete. This will, inter alia, make clearer the network of relations among the variables, and the possible presence of confounders, of mediators, and of colliders.

Temporal ordering

It is usually admitted that causes should precede their effects in time. It is hence necessary to integrate time into causal models and to choose the data accordingly. A main advantage of longitudinal data, either retrospective or prospective, is to give the time-ordering of events. Nevertheless, we recall that retrospective data are affected, i.a., by recall biases and prospective data by drop-outs. Record linkage is another approach that can be used to time-order events at the individual level. For example, Rychtaříková et al. (2013) have used linked data from three Czech registers to examine the impact of maternal and paternal age at childbearing on congenital anomalies. More and more sources of data are being linked together, and methods for analyzing *Big Data*, structured and unstructured, are becoming increasingly available. However, temporal ordering does not imply causal ordering. In other words, as it is well known, association is not causation, temporally or otherwise. Without a good knowledge of the mechanisms leading from the causes to the effects, it is impossible to infer causality from the simple ordering of events. On the other hand, if the mechanism is not well known, observing the regular succession of events may put one on the way of eventually finding a convincing explanation. Regular succession is indeed one of the causal criteria proposed by David Hume in the eighteenth century and is still valid for exploratory purposes.

Multiple levels

A distinction was made in the "Introduction" section between aggregate or macro-level analysis and individual or micro-level analysis. Actually, both levels of explanation are often required to truly understand a given phenomenon, and multi-level models are

recommended for this purpose. Their aim is to separate the effects resulting from micro characteristics, from those emanating from macro features and the environment. The Polish sociologist Stefan Nowak (1989), many years ago, has stressed the need for constructing multi-level theories in the social sciences, taking into account causes and effects from more than one level. In demography, Daniel Courgeau (2007) has thoroughly examined the change in the paradigms of demographic research, from the macro to the micro level, and then to multilevel analysis. Here as elsewhere a strong conceptual framework is required, in order to disentangle the multi-level network of relationships between variables. Courgeau has shown for example, in a study of Norwegian inter-regional migration according to the fact that the individual is a farmer or not, that the parameters estimated at the micro and macro levels were contradictory. These differences could be explained by simultaneously including in the micro model the fact of being a farmer and the percentage of farmers living in the region (Courgeau and Baccaïni, 1998). To give a recent example in the field of subjective health, Teixeira Vaz et al. (2019) have examined life satisfaction among older people in Belo Horizonte according to their individual characteristics and those of their neighborhoods. The paper shows, among others, a lower prevalence of life satisfaction among those who lived in neighborhoods with high physical disorder levels (such as presence of trash and graffiti), after adjusting for individual and other contextual characteristics. One could extend this approach to include the spatial patterns (clustering) of neighborhood deprivation and of life satisfaction (Okrasa and Rozkrut, 2019).

Agent-based simulation

Frans Willekens has been a pioneer in building a bridge between the micro and macro levels with his *microsimulation* MicMac models (Willekens, 2005). For bridging the micro-macro gap, some studies are resorting to *agent-based simulation* modelling (ABM). For example, Billari et al. (2007) have developed an ABM marriage model for a population of interacting agents, taking into account the chances of marrying and the willingness to marry. The typical macro age-pattern of marriage emerges from this micro simulation. A detailed presentation of ABM is given, among others, in Grow and Van Bavel (2017). Obviously, simulation models create a virtual world and, as Diez Roux (2015, p.101) has pointed out, in this artificial world we cannot determine whether X causes Y in the real world, because the virtual world is our own creation. We can only create scenarios and examine, for example, the implications of counterfactuals. Nevertheless, as the simulation exercise consists of acting and inter-acting individual agents, where an agent's behavior can be made dependent upon the behavior of others, ABM could be used in causal research not only for creating counterfactual scenarios or for taking heterogeneity and time into account, but possibly also for bridging the individual level and the macro levels. As put forward by Casini and Manzo (2016, p.18), ABM allows the co-habitation of several levels of analysis: "By iterating the objects' behavior, by making the objects communicate, and by collecting the local products of these behaviors over time, the simulation of an ABM is able to produce the macro level step-by-step." For empirical validation, macro level results can then be confronted with the real world. The micro level rules of behavior should be based, to the best of one's knowledge, on sound empirical insight rather than on hypothetical assumptions. For example, in the health field, Ajelli et al. (2010) have used agent-based

modelling for studying the spread of an infection among individuals through contacts with household members, school and workplace colleagues, and by random contacts with the general population. However, the question remains to what extent macro-level rules can be construed from micro level ones. ABM seems especially useful, in causal analysis, for "opening-up" the black box by theoretical explorations, when part of the mechanism between cause and effect is unknown. Nevertheless, these explorations rely on different scenarios, some of which may lead to the same observed effect; in that case, the black box will remain black.

The need for qualitative data

The methodology discussed up to now has been mainly quantitative. However, qualitative methods can bring extra knowledge concerning causal processes. It is pointed out below, in the paragraph on triangulation, that the example given in the "Mechanisms and structural causal modelling" section, on the causes of contraceptive use in the capitals of four African countries, should be backed up by other means. Actually, the quantitative study was complemented by a qualitative one based on semi-structured in-depth interviews with women in Accra, Dakar, Ouagadougou, and Rabat (Bajos et al., 2013). This investigation showed that social reproductive norms are strongly linked with fertility inside marriage and that fertility of the woman must be proven when she is married. This reinforces our finding concerning the union-reproductive indirect path.

Triangulation

In many cases, different theories can be applied to the same data set. There can be many ways of looking at the data, and no sole model can be deemed "true." If these theories are confirmed by the data, they may lead to different causal conclusions. How do we choose between competing theories? A single study can rarely lead to sustainable causal claims. Triangulation is suggested in social research, and also in epidemiology (Vandenbroucke et al. 2016), as a way to support one's causal conclusions (see the pros and cons in Flick, 2017). This requires that results converge when they are obtained from different independent studies, on the same population and in the same context, with different methods of data collection and analysis. For instance, our study presented in the "Mechanisms and structural causal modelling" section should ideally be backed up by other surveys on the same population, by other data sources such as medical registries, and by other methods such as qualitative research. If results do not converge, one should tone down one's causal claims and try understanding why results diverge. For example, for complex phenomena, different theories, data, and methods may shed light on different facets of the object of study. In any case, triangulation should improve our knowledge of the phenomenon beyond what is made possible by one sole approach (Flick, op. cit.).

More systematic reviews

Many studies are available in the literature dealing with a specific research question. It seems necessary, from time to time, to take stock of the findings by way of systematic reviews based on clearly defined protocols. This is also important for providing the background knowledge required for further studies. Systematic reviews are currently done in the medical field and a good example are the *Cochrane Reviews.* According to

the Cochrane website, "*A systematic review attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a specific research question*". A first step is to define the criteria for selecting the studies. For instance, do we solely keep the papers in English or French, or do we include other languages as well? What is the time period considered? What geographical context do we cover? Which sources of studies will be chosen? … The studies selected should then be analyzed on the basis of clearly defined criteria. For example, for quantitative studies, what significance level is required for selecting the study in the review? Finally, a "summary of findings" should be provided, with information concerning the quality of the studies included in the review process, the potential biases, the variables entered in the analyses, and the main results observed. Jenicek (1987, chapter 2) has outlined a series of questions that can be considered in the analysis of the studies.

It should be pointed out, however, that systematic reviews can be affected by various biases. In particular, studies with statistically significant results are more frequently published than those with null or negative results (Easterbrook and Berlin, 1991). They are also published earlier. Studies with significant effects are more likely to be written in English and cited by other authors (Sterne et al., 2001). Therefore, the probability is higher that they be included in systematic reviews. Low methodological quality of some studies may also be an important source of biases. This is especially the case of smaller studies (Sterne et al., op. cit.).

Conclusions

This paper started off by stating that demographic research has, for decades, been more concerned with descriptive results than with causal assessment, and the reasons for this have been proposed. This observation should however not be considered as disparaging. Indeed, a thorough description of the data is often a first and necessary step in explanation; thus, for instance, the continued importance of demographic analysis. At present, more and more studies go beyond description and attempt some form of explanation by searching for the factors, determinants, causes, … , of the phenomenon considered. Most of these studies are still based on single equation statistical models. We have suggested that there are good reasons to opt instead for multiple equations models describing the organized network of relations among variables. This is, in particular, important for choosing the correct variables to control for and to avoid excess controlling. Single equation models remain valuable as a tool for a preliminary analysis of data. Other methods of exploratory data analysis can also be used for this purpose, such as dimensionality reduction (principal components analysis, multidimensional scaling, etc.).

We have recalled and discussed several ways to improve causal inference in demographic research. In particular, it has been suggested that more systematic reviews of the literature should be conducted on pertinent research questions and that triangulation is required before asserting causal claims. In order to organize and visualize the network of relations between the variables identified in the literature, it was highly recommended to draw, to the best of one's knowledge, the corresponding directed graph. The research approach one takes depends of course on the quantity and quality of background knowledge, on the availability of relevant data, on intuition, and sometimes on serendipity. To conclude, few theories are probably tenable outside the context for which they have been developed. It should always be remembered that no causal model

Appendix

Table 1 Criteria of causal inference

Ní Bhrolcháin and Dyson (2007)	Time order	Rothman and Greenland (1998)	Strength
	Contiguity		Consistency
	Duration		Specificity
	Distinctiveness		Temporality
	Direction		Biologic gradient
	Proportionality		Plausibility
	Recurrence		Coherence
	No cause, no effect		Experimental evidence
	Mechanism		Analogy
	No alternative		

can be deemed "true" in demographic research, as a model is always a partial and simplified representation of the real world.

References
Ajelli, M., Gonçalves, B., Balcan, D., et al. (2010). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, *10*(190), 1–13.
Bajos, N., Teixeira, M., Adjamagbo, A., et al. (2013). Normative tensions and women's contraceptive attitudes and practices in four African countries. *Population*, *68*(1), 15–36.
Baptista E.A. and Queiroz B.L. (2019). The relation between cardiovascular mortality and development: A study of small areas in Brazil, 2001–2015, *Demographic Research*, Vol. *41*, Article 51, 1437-1452.
Barbieri, M. (2013). Mortality in France by *département*. *Population*, *68*(3), 375–418.
Basu, K. (2014). Randomisation, causality and the role of reasoned intuition. *Oxford Development Studies*, *42*(4), 455–472.
Bijwaard, G. E., Tynelius, P., & Myrskylä, M. (2019). Education, cognitive ability, and cause-specific mortality: A structural approach. *Population Studies*, *73*(2), 217–232. https://doi.org/10.1080/00324728.2018.1493135.
Billari, F. C., Prskawetz, A., Diaz, B., & Fent, T. (2007). The "wedding ring": An agent-based marriage model based on social interaction. *Demographic Research*, *17*(3), 59–82.
Blalock, H. M. (1968). The measurement problem: A gap between the languages of theory and research. Chap. 1. In H. M. Blalock, & A. B. Blalock (Eds.), *Methodology in social research*, (pp. 5–27). New York: McGraw-Hill.
Blum, A., & Monnier, A. (1989). Recent mortality trends in the U.S.S.R.: New evidence. *Population Studies*, *43*(2), 211–241.
Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. Chap. 15. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research*, (pp. 301–328). Springer.
Cameron, D., & Jones, I. G. (1983). John snow, the broad street pump and modern epidemiology. *International Journal of Epidemiology*, *12*(4), 393–396.
Casini, L., & Manzo, G. (2016). *Agent-based models and causality. A methodological appraisal*. In *IAS working paper series, 2016:7*, (80p). Linköping: University.
Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica*, *72*(5), 1409–1443.
Courgeau D. (2007). Multilevel synthesis. From the group to the individual, Springer.

Courgeau, D., & Baccaïni, B. (1998). Multilevel analysis in the social sciences. *Population, 10*(1), 39–71.

Daniel R. and De Stavola B.L. (2019). Mediation Analysis for life course studies. Chap. 1 in G.B. Ploubidis, B. Pongiglione, B. De Stavola, *et al.: Pathways to Health*, Springer, 1-40.

Dawid, A. P. (2009). Beware of the DAG! *JMLR: Workshop and Conference Proceedings, 6*, 59–86.

Deaton, D., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine, 210*, 2–21.

Diez Roux, A. V. (2015). Invited commentary: The virtual epidemiologist–promise and peril. *American Journal of Epidemiology, 181*(2), 100–102.

Doll, R., & Hill, A. B. (1964). Mortality in relation to smoking: Ten years' observations of British doctors. *British Medical Journal, 1*, 1399–1410.

Easterbrook, P. J., & Berlin, J. A. (1991). Publication bias in clinical research. *Lancet, 337*, 867–872.

Flick U. (2017). Triangulation. Chap. 19 in N.K. Denzin and Y.S. Lincoln (Eds.): *The Sage Handbook of Qualitative Research*, fifth edition, Sage, 444-461.

Freedman, D. (2004). Graphical models for causation, and the identification problem. *Evaluation Review, 28*(4), 267–293.

Freedman, D., & Humphreys, P. (1999). Are there algorithms that discover causal structure? *Synthese, 121*(1-2), 29–54.

Gaumé, C., & Wunsch, G. (2010). Self-rated health in the Baltic countries, 1994-1999. *European Journal of Population, 26*(4), 435–457.

Gérard, H. (1989). Théories et théorisation. In J. Duchêne, G. Wunsch, & E. Vilquin (Eds.), *Explanation in the social sciences. The search for causes in demography, Chaire Quetelet 1987*, (pp. 267–281). Louvain-la-Neuve: CIACO.

Gérard, H. (2006). Theory building in demography. Chap. 129. In G. Caselli, J. Vallin, & G. Wunsch (Eds.), *Demography analysis and synthesis. A treatise in population studies*, (vol. 4, pp. 647–659). San Diego: Elsevier Academic Press.

Gisselquist D. (2013). Randomized controlled trials for HIV/AIDS prevention among men in Africa: Untraced infections, unasked questions, and unreported data. Chap. 16 in G.C. Denniston, F.M. Hodges and M.F. Milos (Eds.): *Genital cutting: Protecting children from medical, cultural, and religious infringements*, Springer, 243-270.

Glennan, S. (2011). Singular and general causal relations: A mechanist perspective. In P. McKay Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences*, (pp. 789–817). Oxford: Oxford University Press.

Gourbin C., Wunsch G., Moreau L., and Guillaume A. (2017). Direct and indirect paths leading to contraceptive use in urban Africa. An application to Burkina Faso, Ghana, Morocco and Senegal, *Revue Quetelet / Quetelet Journal*, 5(1), 33-70.

Green T. and Hamilton T. (2019). Maternal educational attainment and infant mortality in the United States: Does the gradient vary by race/ethnicity and nativity? *Demographic Research*, Vol. 41, Article 25, 713–752.

Grow A. and Van Bavel J. (Eds.) (2017). *Agent-based modelling in population studies: Concepts, methods, and applications*, Springer International Publishing.

Gunnell, D., Fernando, R., Hewagama, M., et al. (2007). The impact of pesticide regulations on suicide in Sri Lanka. *International Journal of Epidemiology, 36*(6), 1235–1242.

Haavelmo T. (1944). The probability approach in econometrics, *Econometrica*, 12, Supplement, iii-vi + 1-115.

Henry, L. (1972). *Démographie - analyse et modèles*. Paris: Larousse.

Herttua, K., Mäkelä, P., & Martikainen, P. (2008). Changes in alcohol-related mortality and its socioeconomic differences after a large reduction in alcohol prices: A natural experiment based on register data. *American Journal of Epidemiology, 168*(10), 1110–1118.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960.

Illari, P., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*(1), 119–135.

Jenicek, M. (1987). *Méta-analyse en médecine*. Edisem, Quebec: Evaluation et synthèse de l'information clinique et épidémiologique.

Kincaid H. (2004). There are laws in the social sciences. Chap. 8 in H. Hitchcock (Ed.): *Contemporary debates in philosophy of science*, Blackwell, 168-185.

Lopez, R. O., Mompart, A., & Wunsch, G. (1992). Système de soins et mortalité régionale : Une analyse causale. *European Journal of Population, 8*(4), 363–379.

Lopez, R. O., & Wunsch, G. (1990). Méthodes spatio-temporelles pour l'analyse de la mortalité. *Espace, Population, Sociétés, 3*, 393–402.

Lucas, R. (1976). Econometric policy evaluation: A critique. In K. Bruner, & A. Metzler (Eds.), *The Phillips curve and labour markets, Carnegie-Rochester conference series on public policy, 1*, (pp. 19–46). New York: American Elsevier.

Ludwig, J., & Miller, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics, 122*(1), 159–208.

Melnick, A. L., Rdesinski, R. E., Marino, M., et al. (2016). Randomized controlled trial of home-based hormonal contraceptive dispensing for women at risk of unintended pregnancy. *Perspectives on Sexual and Reproductive Health, 48*(2), 93–99.

Mérat, M. C. (2019). Effet placebo. Il soigne vraiment ! *Science & Vie, 1225*, 64–80.

Mouchart, M., Bouckaert, A., & Wunsch, G. (2019). Pharmacological and residual effects in randomized placebo-controlled trials. *A structural causal modelling approach, Revue d'Epidémiologie et de Santé Publique, 67*(4), 267–274.

Mouchart, M., Wunsch, G., & Russo, F. (2016). Controlling variables in social systems - a structural modelling approach. *Bulletin of Sociological Methodology, 132*, 5–25.

MRC (2012). *Using natural experiments to evaluate population health interventions: Guidance for producers and users of evidence.* UK: Medical Research Council.

Ní, B. M. (2001). 'Divorce effects' and causality in the social sciences. *European Sociological Review, 17*(1), 33–57.

Ní, B. M., & Dyson, T. (2007). On causation in demography: Issues and illustrations. *Population and Development Review, 33*(1), 1–36.

Nowak S. (1989). Causality and determinism in the social science. In J. Duchêne, G. Wunsch and E. Vilquin (Eds.): *Explanation in the Social Sciences. The search for causes in demography*, Chaire Quetelet 1987, CIACO, Louvain-la-Neuve, 225-266.

Okrasa, W., & Rozkrut, D. (2019). Subjective and community well-being interaction in multilevel spatial modelling framework. *Statistics in Transition, 20*(4), 167–179.

Pearl, J. (2000). *Causality. Models, reasoning, and inference.* Cambridge University press, Cambridge, revised and enlarged in 2009.

Pearl, J., & Mackenzie, D. (2018). *The book of why*. New York: Basic Books.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effect. *Biometrika*, *70*(1), 41–55.

Rothman K.J. and Greenland S. (1998). *Modern epidemiology*. Lippincott – Raven Publishers, Philadelphia, 2nd edition.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non randomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Russo, F. (2009). *Causality and causal modelling in the social sciences. Measuring variations, Methodos series*. New York: Springer.

Russo, F. (2014). What invariance is and how to test for it. *International Studies in the Philosophy of Science*, *28*(2), 157–183.

Russo, F., Wunsch, G., & Mouchart, M. (2011). Inferring causality through counterfactuals in observational studies – Some epistemological issues. *Bulletin of Sociological Methodology*, *111*, 43–64.

Russo, F., Wunsch, G., & Mouchart, M. (2019). Causality in the social sciences: A structural modelling framework. *Quality & Quantity*, *53*(5), 2575–2588.

Rychtaríková, J., Gourbin, C., Wunsch, G., & Sipek, A. (2013). Should females and males avoid having their children late in life? Impact of parental ages at childbearing on congenital anomalies. *Demographic Research*, *28*(5), 137–176.

Sterne, J. A. C., Egger, M., & Davey, S. G. (2001). Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal*, *323*, 101–105.

Stock, J. H., & Watson, M. W. (2003). *Introduction to econometrics*. Boston: Addison Wesley.

Strotz, R. H., & Wold, H. O. (1960). Recursive vs. nonrecursive systems: An attempt at synthesis. *Econometrica*, *28*(2), 417–427.

Tarka, P. (2018). An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity*, *52*, 313–354.

Teixeira Vaz C., de Souza Andrade A.C., Proietti F.A., *et al.* (2019). A multilevel model of life satisfaction among old people: Individual characteristics and neighborhood physical disorder, *BMC Public Health*, 19, Article 861, 12 p.

Vandenbroucke, J. P., Broadbent, A., & Pearce, N. (2016). Causality and causal inference in epidemiology: The need for a pluralistic approach. *International Journal of Epidemiology*, *45*(6), 1776–1786.

Willekens, F. (2005). Biographic forecasting: Bridging the micromacro gap in population forecasting. *New Zealand Population Review*, *31*(1), 77–124.

Woodward, J. (2003). Making things happen. In *A theory of causal explanation*. Oxford: Oxford University Press.

Woodward J. (2016). Causation and manipulability, *Stanford Encyclopedia of Philosophy,* Stanford University.

Wooldridge J.M. (2013). *Introductory econometrics – A modern approach*, 5th edition, South-Western.

Woolf, S. H., & Schoomaker, H. (2019). Life expectancy and mortality rates in the United States, 1959-2017. *Journal of the American Medical Association*, *322*(20), 1996–2016.

Wunsch, G., & Gourbin, C. (2018). Mortality, morbidity, and health in developed societies: A review of data sources. *Genus*, *74*, 2. https://doi.org/10.1186/s41118-018-0027-9.

Wunsch, G., Mouchart, M., & Russo, F. (2014). Functions and mechanisms in structural-modelling explanations. *Journal for General Philosophy of Science*, *45*(1), 187–208.

Wunsch, G., Mouchart, M., & Russo, F. (2018). Causal attribution in block-recursive social systems: A structural modeling perspective. *Methodological Innovations*, *11*(1), 1–11.

Publisher's Note