EJEP-D-15-00242 · Final / Second Revised · 2 July 2015

The current deconstruction of paradoxes:
One sign of the ongoing methodological “revolution”.

Miquel Porta,[1,2,3] Paolo Vineis,[4,5] Francisco Bolúmar[3,6,7]

[1] Hospital del Mar Institute of Medical Research (IMIM), Universitat Autònoma de Barcelona, Catalonia, Spain

[2] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, USA

[3] CIBER Epidemiología y Salud Pública (CIBERESP), Spain

[4] MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College, London

[5]Human Genetics Foundation (HuGeF), Torino, Italy

[6] Faculty of Medicine, University of Alcalá, Alcalá de Henares, Madrid, Spain

[7] City University of New York School of Public Health at Hunter College, New York

Correspondence: Miquel Porta
Hospital del Mar Institute of Medical Research - IMIM
Universitat Autònoma de Barcelona
Carrer del Dr. Aiguader, 88
E-08003 Barcelona, Catalonia, Spain
Tel: +34 93 316 0700 and +34 93 316 0781
Fax: +34 93 316 0410
E-mail: mporta@imim.es

Abstract

The current deconstruction of paradoxes is one among several signs that a profound renewal of methods for clinical and epidemiological research is taking place; perhaps for some basic life sciences as well. The new methodological approaches have already deconstructed and explained long puzzling apparent paradoxes, including the (non-existent) benefits of obesity in diabetics, or of smoking in low birth weight. Achievements of the new methods also comprise the elucidation of the *causal structure* of long-disputed and highly complex questions, as Berkson's bias and Simpson's paradox, and clarifying reasons for deep controversies, as those on estrogens and endometrial cancer, or on adverse effects of hormone replacement therapy. These are signs that the new methods can go deeper and beyond the methods in current use. A major example of a highly relevant idea is: when we condition on a common effect of a pair of variables, then a spurious association between such pair is likely. The implications of these ideas are potentially vast. A substantial number of apparent paradoxes may simply be the result of *collider biases*, a source of selection bias that is common not just in epidemiologic research, but in many types of research in the health, life, and social sciences.

The new approaches develop a new framework of concepts and methods, as collider, instrumental variables, d-separation, backdoor path and, notably, Directed Acyclic Graphs (DAGs). The current theoretical and methodological renewal –or, perhaps, "revolution"– may be changing deeply how clinical and epidemiological research is conceived and performed, how we assess the validity and relevance of findings, and how causal inferences are made. Clinical and basic researchers, among others, should get acquainted with DAGs and related concepts.

Keywords: paradox; methods; paradigm; clinical research; obesity; diabetes; smoking; low birth-weight; causal inference; causality; causal structure; Berkson's bias; Simpson's paradox; estrogens; collider; instrumental variables; Directed Acyclic Graphs (DAGs).

Introduction

The past few years have been times of silent and intense paradox deconstruction. In some clinical areas, paradoxes have been deconstructed, explained and, sometimes, demolished. The process of deconstruction, which continues, may never have been so intense in clinical and epidemiological research, where apparent paradoxes have always been observed and questioned, as in other fields.[1] Of course, the causal interpretation of medical paradoxes was often, inherently, dubious. Yet, many mysteries lasted decades, while others remain bewildering. Clinical and epidemiological paradoxes are now regularly shown to have little or no causal significance. Or worse: to be nothing, to not exist. The pace of deconstruction is due to a formidable underlying process: the development and application of new, more formal and powerful methods to formulate hypotheses, analyse data (in particular, control biases), interpret results, and make causal inferences. The new methods already achieved several feats. They deserve the attention of clinicians interested in research methods.

One weight paradox

A remarkable example of the potential and actual achievements of the new methods is the deconstruction of some paradoxes involving body weight, such as the putative benefits of obesity in diabetics. While obesity is a well-established risk factor for type 2 diabetes (T2D), among people who already developed T2D the obese are often observed to fare better than T2D individuals with normal weight. Obese diabetics appear to survive longer and to have a milder clinical course than non-obese diabetics. But it is now being shown that the observation lacks causal significance. (Yes, indeed, an observation may be real and yet lack causal meaning).[2,3] As we will further see below, apparent paradoxes involving weight are diverse and complex, and some may reflect real causal processes. In some critical situations, being obese may confer advantages.[4]

The new methods are building methodological knowledge upon methods and knowledge previously generated by graph theory, computer science, artificial intelligence, statistics or epidemiology itself.[2,5-13] One way to explain why observations as the mentioned “obesity

paradox" lack causal significance is that "conditioning on a collider" (in our example, focusing only on individuals who developed T2D) creates a spurious association between obesity and survival.[2,3,5-15]

A collider is a variable directly affected by two or more other variables in a causal diagram[3] (Figure 1); in our example diabetes is a collider since it is affected by obesity and, independently, by an unknown variable U (e.g., some environmental chemicals or genetic variants).[16] Conditioning on a collider means attempting to control for the collider through restriction (in the design or in the analysis), stratification, or regression adjustment; such action will tend to induce a noncausal association (often referred to as *collider bias*) between the shared causes of the collider.[2,3,5-7,14,17,18]

The idea is highly relevant and rather broad: when in the design or in the analysis of a study we condition on a common effect of a pair of variables, then a spurious association between such pair is likely. The implications of this idea are potentially vast –not just for clinical and biomedical research, but for research in most health, life, and social sciences.

Figures 1 and 2 present Directed Acyclic Graphs (DAG),[2,5-9] a causal diagram now commonly used. Figure 1 includes only the three arrows that are necessary for collider bias to occur: an arrow from obesity (the exposure) to diabetes (the variable that the analysis is being stratified on), an arrow from U to diabetes, and an arrow from U to mortality (the outcome).

For simplicity of presentation, Figures 1 and 2 omit some important arrows, including those from diabetes and low birth weight to mortality, as well as some common causes for the variables on the causal diagrams. While an arrow simply indicates the possibility of a causal connection, a missing arrow represents a claim of zero effect and a conditional independence; for example between U and obesity. In causal diagrams, causal assumptions are also encoded in the missing arrows. Like all model frameworks, causal diagrams are only as good as the assumptions on which they are based.

Now, imagine that obesity and the environmental chemicals are the only two independent causes of diabetes. In this situation, all diabetic persons would have been exposed to obesity or the

environmental chemicals (or both). Thus, all diabetics that are not obese would necessarily have accumulated the chemicals, a fact that (let's assume) is associated with a higher mortality than obesity (which is not implausible).[16] Therefore, there would be an inverse association between obesity and mortality among diabetic persons. This would be so even if, as assumed in Figure 1, obesity has no causal effect on mortality. Conditioning on a collider is a form of selection bias.[2,6,12,14,18] Nevertheless, where the collider is affected by a cause of the outcome (as in Figures 1 and 2), some authors consider it also a type of confounding.[2]

By the way, should we recommend persons with T2D to remain overweight? Clearly, no.[19] Until now we already felt the 'no' was right, but we couldn't figure out why as exactly as DAGs allow.

Nevertheless, whether conditioning on a collider exacerbates or reduces bias in estimating a specific causal effect depends on the causal relations among variables affecting the collider and on the nature of their joint effect (additive or multiplicative) in the risk scale. For example, in Figure 1 there is no arrow from U to obesity, in accordance with the assumption (which is not the only plausible one[16]) that obesity is independent of U in the population. Yet, if we added an arrow from U to obesity in the DAG in Figure 1 (i.e., because we think that the available evidence supports the hypothesis that U directly affects obesity[16]), then selection on the collider diabetes might reduce bias in estimating the effect of obesity on mortality.

The bias just mentioned becomes particularly common when we deal with an outcome that is in fact a combination of diseases –e.g., breast cancer–, each of which has different etiologies and prognoses.

In the methodological sense mentioned above, collider is a term almost unheard of in the biomedical literature until recently. It will soon be quite common. The new approaches not only explain paradoxes, they provide a new framework of concepts and methods that it is already inevitable to critically consider and, often, use.

Directed Acyclic Graphs (DAGs): the flagship of the new methods

Alluring, DAGs are just the most visible component of the new methods. But collider, common causes, causal diagrams, instrumental variable (and the related Mendelian randomisation), d-separation rules, backdoor path, identifiability, transportability, positivity, ignorability, collapsibility, exchangeability, g-estimation, potential outcome, negative controls, immortal time bias...[3] these are the new concepts –some, not so new in mathematics, physics or economics– that many clinical and epidemiological researchers –and some basic scientists as well– may now need to learn. No kidding. Which should be fun to many, though the learning and incorporation will not be as easy as it may seem: the ongoing methodological "revolution"[20,21] is partly founded on complex mathematics, concepts as "counterfactuals", techniques as inverse probability weighting (IPW), and, yes, rules to draw the attractive DAGs (Figure 1). Attractive, but less intuitive than they may look. Very important: causal diagrams are a way to *encode* the researchers subject-matter knowledge, and assumptions, about the qualitative *causal structure* of a problem. In clinical and epidemiological research, and in other sciences, diagrams have been used to encode knowledge about systems of variables for decades.[5,7] When such knowledge is limited (and why perform research if it is not?), DAGs help make explicit the ignorance, uncertainties, hypotheses, and plausible causal scenarios. Thus, only experts[8,9] –e.g., on medicine, obesity, genetics, and environmental chemicals– may suggest whether in Figure 1 an arrow should link U and obesity (some forms of obesity may be influenced by sets of genetic variants, some environmental chemicals may be obesogens).[16]

Hence, DAGs are not a new "black box": they are as transparent as our knowledge, assumptions, and hypotheses. Yet, their functioning and reasoning are often difficult to follow, beyond intuition, and beyond classic reasoning on confounding. To follow a DAG concentrating and thinking hard is necessary, but not sufficient: you must also know the subject, and the rules to formulate and interpret them.[2,5-9,13,14,18]

Indeed, DAGs must be drawn and interpreted following stricter rules than the informal, heuristic graphs that we all use intuitively, and which have a long tradition in health research.[7,22] As currently proposed DAGs are important, and fun, because they force us to make explicit our knowledge, hypotheses, and assumptions. Regrettably, out of ignorance, laziness, and other

priorities many of us will stay away from the mathematical foundations of some of the new methods, as IPW, pseudopopulations, randomisation in observational studies, and the like; consequently, it is yet unclear how often some of the new methods will be used, and misused.[18,23] Nevertheless, since they require more of us, they may well spare us from "quick and dirty" approaches to complex clinical problems. The challenge for clinical researchers is thus to understand the basics of the new language and concepts, and to work with epidemiologists and statisticians with deep knowledge of the methods. In turn, the challenge for methodologists is to work with clinicians in integrating the complex biological, pathophysiological, social and environmental mechanisms of human diseases.[3,22,24-31]

Of course, it is not the aim of this paper to provide a tutorial on DAGs and related methods, nor to cover in depth their possible uses, and limitations, which they have. As structural models of bias, DAGS cannot accommodate important biases such as random confounding or multivariate measurement error. DAGs are also limited or unable to handle the time sequence of events, interactions among variables, or the magnitude and functional forms of their relations, which are all important for establishing causality.[2,5,6,13,14,17,18,32]

As all scientific tools, the new methods have indications, contraindications, and limitations (needless to say?). Valid causal inference, in particular, most often requires a plurality of approaches.[2,3,5,6,15,26-33] Nowhere in the most representative works of the leaders of the current theoretical and methodological renewal do we find claims that there is only one way to perform causal inference or other related scientific endeavours.[2,5,6-15,17-23,26,33-40]

The low birth weight paradox

Another prominent example involving weight, the 'low birth-weight paradox' stems from the following apparently paradoxical observation: low birth weight infants in groups exposed to a prenatal factor that increases the risk of low birth weight typically have a lower mortality rate than unexposed low birth weight infants. For instance, low birth-weight children born to smoking mothers have a lower infant mortality rate than low birth-weight children of non-smoking mothers. Should mothers at risk of delivering a low birth-weight child be

recommended to smoke? No, because: a) among children born with normal weight, those born to smoking mothers have a higher mortality rate than those born to non-smokers; and b) the spurious association between survival and smoking is due to collider bias. In captivating studies, Wilcox et al. have assessed the birth weight paradox by using empirical curves of the birth weight distribution and birth weight specific mortality.[41] And more recently, Hernández-Díaz et al. used DAGs as a conceptual framework to discuss empirical evidence on the birth weight paradox.[10,11] They have hence showed that the birth weight paradox can be explained by the presence of other unmeasured factors that reduce birth weight and increase mortality. In terms of DAGs, the apparent paradox could be the result of selection bias due to stratification on the collider low birth weight, which is affected by the exposure of interest (smoking) and shares common causes with the outcome (infant mortality) (Figures 2a and 2b). While causal DAGs facilitate discussion among subject-matter experts, they do not provide the final answer.[6,10,11,18]

Of course, in "pre-DAG" times there was path analysis, and structural nested models.[7] And scholars who thought in-depth and developed fine conceptual models. Brian MacMahon or Allen Wilcox are among those who worked elegantly to unravel the birth weight paradox. However, tellingly, none came up with the concept of collider and its derivatives. For example, a paper from 1997 underlines the importance of conceptual models (as DAGs do),[42] and recommends not to adjust for mediators, which today remains a common mistake in the literature. While the authors were not brilliant enough to discover collider bias,[42] they did propose an explanation for low birth-weight paradox, which remains valid: low birth-weight babies from smoking mothers may be more likely to have intra-uterine growth retardation, whereas low birth-weight babies from non-smokers may be more likely to be preterm, and therefore the latter are more likely to die.[42] The authors did not recommend smoking for high-risk mothers.

Conditioning on antecedent causes may also distort causal estimates. For instance, smoking starts a carcinogenic process that leads to lung cancer: the carcinogenic process is the cause of the protein pattern that the investigators seek to discover with proteomics.[43] Thus, matching (i.e., conditioning) on smoking is likely to increase the prevalence of the protein pattern among the controls and attenuate or mask the association between relevant protein patterns and cancer.

This is another example of the contributions of causal diagrams in providing simple visual explanations for difficult methodological issues.[44]

Paradoxes deconstructed and clarified so far −which often involved obesity− include the following: a survival advantage of being obese in populations diagnosed with a medical condition (e.g., the higher mortality among normal-weight people with dysglycemia is not causal but a product of the closer inverse association between obesity and smoking in this subpopulation);[45] the lower mortality among post-diagnosis overweight patients with colorectal cancer compared with normal weight patients;[46] positive effects of obesity on mortality among individuals with cardiovascular disease;[47] telling patients with chronic disease to gain weight;[19] better survival of patients with dementia and a higher BMI than patients with dementia and low BMI.[48] The range of disorders in which false obesity paradoxes (again: observed for real but lacking causal meaning) may mislead clinicians, scientists and citizens is quite wide: it includes chronic kidney disease, chronic heart failure, chronic obstructive pulmonary disease, and acute coronary syndrome.[49] Others are likely to ensue: obesity will not be the only variable at fault.[50] And, surely, DAGS will not explain all paradoxes: many have reasons in the limitations of our subject-matter knowledge. Yet, let's just consider again the possibilities of reviewing the assumptions and methods of studies that show this kind of "paradox":

> Individuals with a disorder exposed to a factor that increases the risk of the disorder have a better clinical course than individuals with the disorder unexposed to the risk factor.

Did they by any chance condition on a collider...?

Two additional examples of paradoxes that the new approaches have helped deconstruct are 'survivor bias' and a selection bias built in hazard ratios (HRs).[2,51] An example of the survivor bias paradox is the black-white mortality crossover: mortality is greater for blacks and other disadvantaged groups relative to whites at younger ages, but the pattern reverses at oldest ages. The effect may occur if there is confounding by an exposure or by unobserved factors that conferred a survival advantage.[2]

An example of the selection bias built in HRs can be given in the context of a randomised controlled trial.[51] In such context the apparent paradox is that the HR for a harmful exposure can go from greater than 1.0 to less than 1.0 during follow-up. Subjects more susceptible to the outcome of interest are preferentially excluded from the treatment arm as they develop the outcome over time –precisely because they received a therapy with harmful effects to which they were susceptible (all other factors to which they were susceptible are equally distributed between the 2 arms). With time, the proportion of susceptible subjects progressively increases in the placebo arm relative to the treatment arm. The bias due to the differential permanence of less susceptible subjects over time, because of differential depletion of susceptibles, is the built-in selection bias of period-specific HRs. This bias may explain that after some years of follow-up the HR is less than 1.0 even if treatment has no truly preventive effect in any individual at any time.[51] This built-in selection bias of the HR has also been deconstructed using causal diagrams.[52]

Real and potential achievements of the new approaches

Amazingly, but not surprisingly, the new approaches provide insights that are beyond many methods in current use. In particular, the new methods can take us deeper and beyond the methods of "modern epidemiology", the methodological and conceptual current that eclosed in the 1980s.[53] Methods and ideas promoted by modern epidemiology, for instance, contributed substantially to increase the validity and relevance of clinical research, and favoured the emergence of clinical epidemiology, first, and then evidence-based clinical care. Now a no less profound change is taking place. But as Miguel Hernán sometimes warns, senior scholars may not have noticed it, since they seldom have sufficient quality time to study new, complex methods; in the meantime, the younger generation is naturally learning and using them. DAGs are becoming a common feature in the landscape of medical articles. Again, just a sign.

Actually, we recently saw a tweet −yes, in Twitter− celebrating "A DAG in The Lancet!". A deserved surprise: it appeared to be the first DAG in a major clinical journal –albeit in a humble letter to the editor, perhaps symbolically.[54] But it was not such 'first': another DAG was published in 2011 raising exactly the same issue.[55] The aim of the letter was to improve the

estimation made by a previous paper of the proportion of cardiovascular disease due to adiposity that can be explained by metabolic mediators (high blood pressure, cholesterol, and glucose). The excess risk mediated can be estimated with and without adjustment for the mediators. Confounders of the association between adiposity and cardiovascular disease were considered, but confounders of the mediator-outcome relation were not. Family history of cardiovascular disease, diet and physical activity could be some of such confounders.[54]

We think it will not be long before the ongoing methodological innovations become mainstream in clinical research, as they are becoming in epidemiologic research. And why should they not affect some areas of basic research as well, even if the rules of inference are different? [3,27] For instance, the idea that when one has conditioned (implicitly or explicitly) on a common effect of a pair of variables, or on a descendant of such collider, then a spurious association between such pair is likely seems also relevant for mechanistic microbiological research. Why or under what circumstances would it not be? Some biochemical and genetic relationships currently considered causal –and not even perceived as paradoxical− could be due to problems as collider-conditioning bias.

Thus, we live exciting days of profound methodological renewal. The deconstruction, explanation, or demolition of some paradoxes is just a symptom of it, or a product. A very important one, in any case. But there are other signs alerting us that something relevant is going on, as the elucidation of the causal structure of long-disputed and highly complex questions. Two such questions are Berkson's bias or fallacy,[12,18] and Simpson's paradox.[56] Another issue that DAGs have helped clarify is the controversy on estrogens and endometrial cancer that ravaged during the 1980s.[8] Or the results of observational and experimental studies on some adverse effects of hormone replacement therapy.[57]

In fact, the signs of the methodological renewal include the explanation of why very smart scientists did not understand each other and strongly disagreed. [8,12] Why was that? Because the mix of formal and intuitive tools they used could not go as deep as the current one goes in elucidating the causal structure of the problems at stake.[18] When new theories, concepts, methods, and techniques are able to clarify such controversies… whew, they deserve our attention. With power and elegance, DAGs are helping understand complex issues that we

addressed unsatisfactorily using words, probabilities, heuristic or intuitive diagrams, and complex epidemiological methods. Not least, DAGs have clarified the previously murky relationship between selection bias and confounding.[2,3,6,12,18]

Furthermore, the new methods are bridging the long annoying gap between experimental and observational designs. No less. Annoying because we always felt that there are many fundamental relationships between the two main types of designs. And because we sensed that sometimes the benefits of randomisation do not last long: all sorts of clinically meaningful protocol deviations –often, just minutes or days after randomisation– transformed more than one randomised clinical trial into a cohort study.[3] Now it is more clear why and how sometimes experiments must and can be analysed as observational cohort studies; and, perhaps (this is less evident) why and how some observational studies can be analysed as experiments −provided that a number of strong assumptions about the absence of residual confounding are made.[6,57]

The new methods are also strengthening the connections among clinical and epidemiological research and other branches of science. First, as mentioned before, because DAGs can obviously help address problems relevant in many fields of scientific inquiry. Second, for instance, through a formal treatment of *negative controls*,[3] which bench scientists employ in their experimental studies to rule out non causal interpretations of results. Methodologists have now defined the conditions under which the use of negative controls for exposure and outcome may detect confounding, selection bias, and information bias in observational studies in some basic, clinical, and epidemiological settings.[58]

Beyond paradoxes

In addition to the accomplishments mentioned above, recent developments in methodology correspond to important changes in the understanding and practice of causal inference in medicine. The 1940s and 1950s saw a burgeoning of the randomised clinical trial as a paradigm for causality, with emphasis on single causal factors (treatments or otherwise "actionable" agents). This was in line with the microbiological tradition; in fact, the first trials were on antibiotics. In Hernán and Robins' language (see below) the prevailing paradigm was the

"counterfactual" interpretation,[6] based on the notion that elimination of the cause would lead to the disappearance of the effect or several effects (if the patient has been treated, counterfactual is the absence of treatment).

Many years ago, guidelines for causal inference in observational contexts were inspired by Henle-Koch's rules, which were originally developed for microbiology. Austin Bradford Hill and others weakened such rules to adapt them to chronic multifactorial diseases.[3,27] Still, many historical examples of causes of chronic diseases involved single, powerful agents such as tobacco and asbestos. Though these exposures were not "necessary causes", their association with the effect was strong, and it was thought that the concomitant exposure to interacting agents or effect modifiers was able to complete the causal picture (a sort of "implicit, undrawn DAG"), and explain disease occurrence.[30]

In the last two decades, however, we have learnt that "weak" causal agents (e.g., some carcinogens and endocrine disruptors) actually contribute to cause chronic diseases; this is largely in contrast to the classical examples of "strong" agents as tobacco or asbestos.[59] One prominent example is air pollution, now recognised as a human carcinogen.[60] Etiology is currently better understood as a *process* involving the complex interplay of numerous agents that act along several mechanistic pathways, thus creating the opportunity for many colliders to bias the results. This is not extraneous to clinical practice: patients' response to therapies is also more complex than we thought 50 years ago, and it depends on complex networks that include co-morbidities, co-exposures, and individual susceptibility. Thus, both in etiologic research and in clinical medicine we acknowledge that an occurrence in a single patient can be due to the concomitant action of several component causes within the causal process; understanding and influencing such process often requires an appropriate mechanistic dissection of its pathways.[31,61] Not just epidemiology, clinical research has also become strongly "mechanistic".[3,24,62] This is an important substantive change that also underlies and fuels the methodological revolution. A change that has implications: *a)* weak causes have an impact on disease occurrence, and we need to face them (e.g., through preventive policies)[63,64]; *b)* causes may act through several pathways and, thus, the study of such processes (e.g., through biomarkers and omics) is relevant to identify intermediate steps where intervention may be effective; and *c)* distinguishing genuine from artefactual causes, and fundamental from ancillary

causes, requires complex reasoning that integrates biological, clinical, environmental, social and methodological knowledge (e.g., through DAGs and other frameworks).[2,5,6,15]

In a more formal way, there are two basic ways to look at causality in this changed scenario.[6] One is called by Hernán and Robins[6] the "sufficient-component-cause framework", and considers sets of actions, events, or states of nature which together bring about the outcome under consideration. The model is able to account, and gives an account of the multiple causes (including weak causes) that in their combination lead to a particular effect. It means that the set or complex itself is sufficient to produce the effect, but none of the component elements, which can be weak and remote, is. For example, lung cancer in an individual may be explained by smoking, plus past exposure to asbestos, plus inherited susceptibility related to some SNPs in DNA repair genes; none of these factors is necessary and none is sufficient, but the complex of factors is sufficient to cause cancer in that single individual. In this context it is clear that weak risk factors can fit as contributing causes, that risk factors are causes (if they are true risk factors, unbiasedly estimated); a weak and probabilistic cause is no less a cause; in medicine, epidemiology, physics or economics a cause is no less a cause because it is not necessary nor sufficient.[3] And the study of pathways and mechanisms thus acquires relevance.

A second approach to causality is based on the "counterfactual model", which focuses on one particular cause or intervention, and gives an account of the various effects of that cause. In contrast to the sufficient-component-cause framework, the potential outcomes framework addresses the question, "What would have occurred if a particular factor were intervened upon and thus set to a different level than it in fact was?" The counterfactual framework does not require a detailed knowledge of the mechanisms by which the factor affects the outcome.[6] Thus, in a randomised trial typically we investigate a single cause, we are interested in its multiple effects, and we are not necessarily preoccupied with mechanisms.

Conclusions

We do not believe, imply or suggest that it is not possible to perform causal inference or other scientific tasks without DAGs or other components of the new methods. As all scientific tools,

the new methods have limitations; sometimes, alternatives; in some instances, they are simply not required (e.g., because they do not challenge or improve upon previously available methods). Valid causal inference, in particular, most often requires a plurality of approaches.

The current theoretical and methodological renewal –or perhaps "revolution"– seems to be changing deeply how clinical and epidemiological research is conceived and performed, how we assess the validity and relevance of findings, and how causal inference is understood and practised. Clinical researchers better get acquainted with DAGs and related concepts. Paradox deconstruction shows the scientific power and substantive relevance of the new approaches. Such power and relevance are also shown by the elucidation of the causal structure of long-disputed questions among leaders in science, and by the new sets of concepts and methods. Long accepted or puzzling paradoxes and other biased beliefs will continue to crumble −tomorrow and in the foreseeable future.

Acknowledgments

We thank Julia del Amo, Arnaud Chiolero, Cesar Victora and three anonymous reviewers for insightful comments to earlier versions of the manuscript. The work was supported in part by research grants from Instituto de Salud Carlos III – FEDER (FIS PI13/00020 and CIBER de Epidemiología y Salud Pública − CIBERESP), Government of Spain; Fundació La Marató de TV3 (20132910); and Government of Catalonia (2014 SGR 1012).

Figure legends

Figure 1. Directed Acyclic Graph (DAG) representing the causal structure between obesity, diabetes, mortality, and an unknown variable U (e.g., some genetic variants or environmental chemicals). Here, diabetes is a collider since it is affected by obesity and, independently, by U. Conditioning on the collider means attempting to control for the collider through restriction (in the design or analysis), stratification, or regression adjustment; such conditioning will tend to induce a noncausal association (collider bias) between obesity and U, the shared causes of diabetes.

Figure 2. Directed Acyclic Graph (DAG) representing two possible causal structures between smoking, low birth weight, mortality, and an unknown variable. The birth weight paradox can be the result of selection bias due to stratification on the collider (low birth weight), which is affected by the exposure (smoking) (Figure 2a), and shares common causes with the outcome (infant mortality) (Figure 2b). Figure 2a is the same as Figure 2b under the causal null hypothesis of no direct effect of the exposure on the outcome.

References

1 Al-Khalili J. Paradox: The Nine Greatest Enigmas in Physics. New York: Random House; 2012.

2 Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL, eds. Modern Epidemiology. 3rd. ed. Philadelphia: Lippincott-Raven; 2008:183–209.

3 Porta M, ed. A Dictionary of Epidemiology. 6th. ed. New York: Oxford University Press; 2014.

4 Dixon JB, Egger GJ. A narrow view of optimal weight for health generates the obesity paradox. Am J Clin Nutr. 2014;99:969–70.

5 Pearl J. Causality: Models, Reasoning and Inference. 2nd. ed. Cambridge: Cambridge University Press; 2009.

6 Hernán MA, Robins JM. Causal Inference. New York: Chapman & Hall / CRC; 2015. http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book

7 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10:37–48.

8 Robins JM. Data, design, and background knowledge in etiologic inference. Epidemiology. 2001;12:313–20.

9 Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002;155:176–84.

10 Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight ‘‘paradox’’ uncovered? Am J Epidemiol. 2006;164:1115–20.

11 Hernández-Díaz S, Wilcox AJ, Schisterman EF, Hernán MA. From causal diagrams to birth weight-specific curves of infant mortality. Eur J Epidemiol. 2008;23:163–166.

12 Snoep JD, Morabia A, Hernández-Díaz S, Hernán MA, Vandenbroucke JP. A structural approach to Berkson’s fallacy. And a guide to a history of opinions about it. Int J Epidemiol. 2014;43:515–21.

13 Greenland S. Quantifying biases in causal models: classical confounding vs. collider-stratification bias. Epidemiology. 2003;14:300–6.

14 Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. Int J Epidemiol. 2010;39:417–20.

15 VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. New York: Oxford University Press; 2015.

16 Lee DH, Porta M, Jacobs DR, Vandenberg LN. Chlorinated persistent organic pollutants, obesity, and type 2 diabetes. Endocrine Reviews. 2014;35:557–601.

17 Richiardi L, Barone-Adesi F, Merletti F, Pearce N. Using directed acyclic graphs to consider adjustment for socioeconomic status in occupational cancer studies. J Epidemiol Community Health. 2008;62:e14.

18 Pearce N, Richiardi L. Three worlds collide: Berkson's bias, selection bias and collider bias. Int J Epidemiol. 2014;43:521–4.

19 Lajous M, Banack HR, Kaufman JS, Hernán MA. Should patients with chronic disease be told to gain weight? The obesity paradox and selection bias. Am J Med. 2015;128:334–6.

20 Porta M. The deconstruction of paradoxes, and a methodological revolution in clinical and epidemiological research. OUPblog (Oxford University Press Blog). 17 October. 2014. http://blog.oup.com/2014/10/deconstruction-paradoxes-sociology-epidemiology/

21 Pearl J. Are economists smarter than epidemiologists? Comments on Imbens's recent paper. Causal analysis in theory and practice [Blog]. October 27,. 2014. http://www.mii.ucla.edu/causality/?p=1241

22 Joffe M, Gambhir M, Chadeau-Hyam M, Vineis P. Causal diagrams in systems epidemiology. Emerg Themes Epidemiol. 2012;9(1):1.

23 Aalen O, Røysland K, Gran J, Kouyos R, Lange T. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. Stat Methods Med Res. 2014 (in press). http://smm.sagepub.com/content/early/2014/03/27/0962280213520436.full.pdf+html

24 Bolúmar F, Porta M. Epidemiologic methods: beyond clinical medicine, beyond epidemiology. Eur J Epidemiol. 2004;19:733–5.

25 Vineis P, Stringhini S, Porta M. The environmental roots of non-communicable diseases (NCDs) and the epigenetic impacts of globalization. Environ Res. 2014;133:424–30.

26 Hernán MA. Agent-based models for causal inference –reweighting data and theory in epidemiology. Am J Epidemiol. 2015;181:103–5.

27 Geneletti SG, Gallo V, Porta M, Khoury MJ, Vineis P. Assessing causal relationships in genomics: From Bradford-Hill criteria to complex gene-environment interactions and directed acyclic graphs. Emerg Themes Epidemiol. 2011;8:5. http://www.ete-online.com/content/8/1/5

28 Krieger N. Epidemiology and the People's Health. Theory and Context. New York: Oxford University Press; 2011.

29 Davey Smith G. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. Int J Epidemiol. 2011;40:537–62.

30 Richmond RC, Al-Amin A, Davey Smith G, Relton C. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. Early Human Devel. 2014;90:769–80.

31 Vineis P, Porta M. Causal thinking, biomarkers and mechanisms of carcinogenesis. J Clin Epidemiol. 1996;49:951–6.

32 Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. Eur J Epidemiol. 2015 Feb 17. DOI 10.1007/s10654-015-9995-7.

33 Glass TA, Goodman SN, Hernán MA, Samet JM. Causal inference in public health. Annu Rev Public Health. 2013;34:61–75.

34 VanderWeele TJ, Vansteelandt S, Robins JM. Marginal structural models for sufficient cause interactions. Am J Epidemiol. 2010;1717:506–14.

35 Robins JM, Hernán MA, Brumbach B. Marginal structural models and causal inference in epidemiology. Epidemiol. 2000;11:550–60.

36 Van der Weele TJ, Hernán MA. From counterfactuals to sufficient component causes and vice versa. Eur J Epidemiol. 2006;21:855–8.

37 Hernán MA, Robins JM. Instruments for causal inference:an epidemiologist's dream? Epidemiology. 2006;17:360–72.

38 Tong S, Neale RE, Shen X, Olsen J. Challenges for epidemiologic research on the verge of a new era. Eur J Epidemiol. 2011;26:689–94.

39 Joffe M, Mindell J. Complex causal process diagrams for analyzing the health impacts of policy interventions. Am J Public Health. 2006;96:473–9.

40 Pearce N. Epidemiology in a changing world: variation, causation and ubiquitous risk factors. Int J Epidemiol. 2011;40:503–12.

41 Wilcox A. On the importance –and the unimportance– of birthweight. Int J Epidemiol. 2001;30:1233–41.

42 Victora CG, Huttly SR, Fuchs SC, Olinto MT. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. Int J Epidemiol. 1997;26:224–7.

43 Rundle A, Ahsan H, Vineis P. Better cancer biomarker discovery through better study design. Eur J Clin Invest. 2012 Dec;42(12):1350–9.

44 Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. Int J Epidemiol. 2013;42:860–9.

45 Preston SH, Stokes A. Obesity paradox: conditioning on disease enhances biases in estimating the mortality risks of obesity. Epidemiology. 2014;25:454–61.

46 Renehan AG. The 'obesity paradox' and survival after colorectal cancer: true or false? Cancer Causes Control. 2014;25:1419–22.

47 Banack HR, Kaufman JS. The obesity paradox: understanding the effect of obesity on mortality among individuals with cardiovascular disease. Prev Med. 2014;62:96–102.

48 Moga DC, Abner EL, Brouwer ES. Dementia and "obesity paradox": is this for real or are we missing something? An epidemiologist's perspective. J Am Med Dir Assoc. 2014 (in press).

49 Niedziela J, Hudzik B, Niedziela N, Gąsior M, Gierlotka M, Wasilewski J, et al. The obesity paradox in acute coronary syndrome: a meta-analysis. Eur J Epidemiol. 2014;29:801–12.

50 Röhrig N, Strobl R, Müller M, Perz S, Kääb S, Martens E, et al. Directed acyclic graphs helped to identify confounding in the association of disability and electrocardiographic findings: results from the KORA-Age study. J Clin Epidemiol. 2014;67:199–206.

51 Hernán MA. The hazards of hazard ratios. Epidemiology. 2010;21:13–5.

52 Hernán MA, Hernández-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15:615–25.

53 Porta M, Vandenbroucke JP, Ioannidis JPA, Sanz S, Fernandez E, Bhopal R, et al. Trends in citations to books on epidemiological and statistical methods in the biomedical literature. PLoS ONE. 2013;8(5):e61837.

54 Chiolero A, Kaufman JS. Metabolic mediators of body-mass index and cardiovascular risk. Lancet. 2014;383:2042.

55 Dahabreh IJ, Kitsios GD, Trikalinos TA, Kent DM. The complexity of ABO in coronary heart disease. Lancet. 2011;377:1493.

56 Hernán MA, Clayton D, Keiding N. The Simpson's paradox unraveled. Int J Epidemiol. 2011;40:780–5.

57 Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008;19:766–79.

58 Lipsitch M, Tchetgen E, Cohen T. Negative controls. A tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21:383–8.

59 Pearce N, Blair A, Vineis P, Ahrens W, Andersen A, Antó JM, et al. IARC Monographs: 40 years of evaluating carcinogenic hazards to humans. Environ Health Perspect. 2015;123:507–14.

60 Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. Environ Health Perspect. 2014;122:906–11.

61 VanderWeele TJ. Mediation and mechanism. Eur J Epidemiol. 2009;24:217–24.

62 Porta M, Álvarez-Dardet C. Epidemiology: bridges over (and across) roaring levels. J Epidemiol Community Health. 1998;52:605.

63 Espina C, Porta P, Schüz J, Hernández-Aguado I, Percival RV, Dora C, et al. Environmental and occupational interventions for primary prevention of cancer: A cross-sectorial policy framework. Environ Health Perspect. 2013;121:420–6.

64 Pearce N, Blair A, Vineis P, Ahrens W, Andersen A, Antó JM, et al. IARC Monographs: 40 years of evaluating carcinogenic hazards to humans. Environ Health Perspect. 2015;123:507–14.