HHS Public Access

Author manuscript

Epidemiology. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:

Epidemiology. 2017 January ; 28(1): 1–5. doi:10.1097/EDE.0000000000000578.

The limits of risk factors revisited: is it time for a causal architecture approach?

Katherine Keyes and **Sandro Galea**

Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY

Dean's Office, Boston University School of Public Health, Boston, MA

Keywords

causal inference; causation; interaction; representative sampling

New challenges for the 21st century

Epidemiology today stands at a juncture. Our methods are increasingly sophisticated, and ever-greater pools of data are being collected and could be available for epidemiologic inquiry. Yet the core approaches inherent to, and the utility of, epidemiology are being questioned perhaps as never before by a growing number of stakeholders.[1–3] Big data, precision medicine, various -omics initiatives, technology-informed intervention, and other new research horizons leave open questions about where epidemiology does (and does not) fit in.

The methodologic advances in our field in the past quarter century have led to the establishment of an armamentarium that allows epidemiology to take its place alongside other sciences with their own well-established, methodologic conventions. They also have led to a particular orientation we take in our work. Our textbooks and advanced epidemiology coursework predominantly instruct our students on how to estimate risk factors, and many of our methodological tools for confounder control provide us with increasingly sophisticated ways to estimate these risk factors more precisely in the presence of complex data structures and potential sources of confounding.

There is little doubt that traditional risk factor epidemiologic methods have contributed in critical ways to our understanding of the effects of smoking, environmental toxins, pre- and peri-natal exposures, and viral agents of infectious and chronic disease, among others. However, it is also clear that many other public health successes of the past century owe little to our methodological developments as a discipline, and that there is growing concern that an effort to increasingly identify small effects marginalizes us as a field.[3–5]

Corresponding author: Katherine M. Keyes, Assistant Professor of Epidemiology, Columbia University Department of Epidemiology, Mailman School of Public Health, 722 West 168th Street, Suite 503, New York, NY 10032, Tel: (212) 304-5652, kmk2104@columbia.edu.

Conflicts of interest: None.

Why we cannot (seriously this time) continue risk-factor approaches

The calls for a move away from risk factors have been compelling and numerous over the past 20 years. It is time to heed the growing drumbeat of concern about these risk factor approaches and propose an alternate approach to how we may conceptualize our hunt for causes of disease, because, in part: (1) these approaches lead us to focus on even more precise estimates of risk factors that we already know are important, or alternatively, already know are not very important; (2) these approaches encourage us to continue to pursue a biomedical model of disease; and (3) the world is too complicated to distill into a set of correlates. These are not new ideas, but perhaps worth restating, with citations from some seminal papers that articulated some of these key ideas over the past decades.

First, the traditional risk-factor approach keeps us bound to what Beverly Rockhill termed a decade ago as “the increasingly reductionist hunt for causes.”[4] Because the field of risk factors is relatively well ploughed and we have a set of “usual suspects” (poverty, smoking, diet, toxins), we are often left estimating risk ratios of relatively small magnitude and trying to make the case for their role in public health. One needs only to look towards the field of genetic epidemiology to see this reductionist hunt in action. Further, we attempt to refine and reframe the extent to which previously identified associations are causal with increasingly restrictive samples with precise analyses. As just one common example, we apply propensity score matches to the data and only analyze those pairs or sets for which there is a match. This is entirely reasonable if the goal is to isolate the causal effect of an exposure on an outcome; only those pairs with a match on the propensity to be exposed contribute to our ability to assess the causal impact of the exposure. But if our overall goal is to improve population health, propensity score matching is perhaps much less useful if those pairs who match on propensity to be exposed are not representative of the population in which we would like to intervene, and if the factors that create extremes in the data and for which no match is possible are also factors that interact with exposures of interest.

Second, we often, and perhaps increastingly, remain ‘prisoners of the proximate’,[6] as Anthony McMichael termed, and we are almost 20 years ago away from when Carl Shy provocatively implored us to shift from biomedical models of disease to models that incorporate the complex social world through which disease generates.[7] We are further still away in time from other scholars who have encouraged us to consider the data-generating mechanisms through which exposures become embedded if we want any hope in changing population health.[8–10] Have we effectively heeded the calls for a movement away from proximate risk factors? In some ways, yes. Rapid progress in areas such as social epidemiology and lifecourse epidemiology have challenged the dominant risk factor paradigm as potentially inadequate to the questions that are of central interest to current public health problems. In the ensuing years, theory and methods to more rigorously assess fundamental causes have proliferated. Yet more broadly, beyond epidemiology, examination of the ‘social’ unless linked to biological and genetic substrates is, in our experience, increasingly evaluated as low priority in medical journals and study sections, suggesting that our survival in the field is questionable if we pursue a social and environmental agenda without an -omic hook.

Third, the natural consequence of a risk-factor approach focused on individual-level associations, especially those proximal to disease onset, is the current era of predictive and personalized preventive medicine. If we have a set of twenty risk factors, each explaining some small amount of variance, it is seductive to create predictive equations attempting to predict disease in individuals based on data collected in populations. In fact, predictive medicine is now being recommended as a potential use of epidemiologic data across several disciplines.[11,12] While predictive medicine for disease treatment has been fruitful in many areas of clinical medicine, the promise of such an approach for disease prevention has been unrealized. The basic mathematical limitations of predictive equations for understanding much less intervening on population health have been well documented, and the utility of such models for public health is questionable.[5,15]

In sum, our risk-factor-based value system can, based on the inputs that we give and the questions that we ask, push us into a corner. Charged with conducting comparative work that can guide population health impact (e.g. “What can we do that would improve the health of the population the most?”), we often find ourselves focusing on better estimating the causal effects of single exposures (e.g., “Did X cause Y?”). While “Did X cause Y?” is fundamental to science, we question here whether we are asking about the right X’s, whether it matters that X caused Y without knowing about the context and co-occurring causes around X, and if trial after trial of increasingly precise estimation of X’s is of real relevance.

Why change?

Informed by the history that precedes us, and an appreciation of the challenges facing the field, we suggest that there are two reasons that compel us to consider change. First, if our goals as epidemiologists and scientists are truly to uncover and reveal truth and generalized knowledge about the natural history of disease, then we have little choice but to confront the network of co-occurring causes that produce disease. Further, if our goals are to identify those causes that result in the greatest harm for the greatest number in order to improve population health, then a focus on prevalence and co-occurrence is also paramount. We know that the number of cases that are caused by a certain factor depend entirely on the prevalence of the co-occurring causes that interact with that factor, therefore the advance of an epidemiology of consequences requires a focused effort on understanding the complex architectures and networks of causes that underlie disease rather than the effects of single exposures in isolation.

What are the solutions?

The solutions to the problems we raise in this essay are unlikely to be methodologic, because we already have the methodologic solutions. The past two decades have witnessed a rapid array of methodological developments in epidemiology that are ripe for pursuing questions that take us beyond the risk factor paradigm, including systems science methods,[16,17] quantitative approaches to understanding neighborhoods and places, articulations of interaction and mediation,[18] and methods to examine population interventions that allow us to control potential confounding with a minimum of

assumptions.[19] It is possible that the reason this wealth of methods has not yet fueled a paradigm shift is that application of these methods requires a relatively high level of statistical and mathematical background at a level outside the scope of many even doctoral-level researchers. Yet through continued training of epidemiologists and researchers in aligned sciences, these methods will gain foothold. In the meantime, there are more basic tools that we all know, and have known for decades, which can put us on the path to a better understanding of how we should better understand population health.

Rather than continuing to refine our estimates of effects, we suggest that we need to move away from efforts to estimate series of associations towards understanding a causal architecture. A causal architecture approach capitalizes on our methods and our insights to ask a new generation of questions. Operationally, it means that our focus should shift from risk factor identification to (1) modeling prevalence of causes within and across populations, including their interactions and (2) representative sampling to understand effect distributions in populations and allow for between-population comparisons (rather than within population risk factor examinations).

First, modeling prevalence of causes within and across populations requires no new methods but rather requires that we use the methods that we already know for new and different purposes. For example, we know that two populations can have the same prevalence of exposure, but different causal effects, if the distribution of component causes differs across these populations. Further, two populations can have the same causal effect, but different distributions of the prevalence of causes, which has direct implications for how we intervene on those causes. It is well understood in epidemiology that the magnitude of the effect of an exposure on disease is dependent on the prevalence of the causal factors that interact with that exposure. While we accept this as a fundamental epidemiologic principle, we often fail to consider its implications seriously. For some exposure-outcome relations, the prevalence of interacting factors may be relatively constant across populations. We have perhaps been lulled into false confidence in this area by our triumph in identifying the association, for example, between smoking and lung cancer. This association is relatively consistent (or at least, always really strong[20]), indicating that the factors that interact with smoking to produce lung cancer have a relatively homogeneous prevalence across many populations. Yet even with smoking and lung cancer there have been numerous factors identified that interact with smoking, varying the magnitude of the association.[21]

Thus, if we continue to focus on estimating series of risk factors, we document associations whose magnitude is bound in time and place. These associations may no longer apply to the present, if the population is dynamic enough, and leads to the seemingly contradictory findings that so befuddle the public. This suggests that our focus on estimating a laundry list of effect measures to characterize the association between various exposures and outcomes is misplaced. An alternate approach would be to focus our effort on theorizing and examining the prevalence of co-occurring factors will allow us to more richly characterize the conditions and architecture that underlie specific risk factor associations. This includes modeling interaction, which is an approach well known to epidemiologists. Such modeling is scale dependent, of course, and one can almost always find interaction on at least one scale as long as two exposures are associated with an outcome. Careful consideration of the

appropriate mathematical scale for each particular question, context, and set of exposures is necessary. More broadly than including interaction terms in our models, though, we suggest that focusing on how such interactions may vary effect estimates across time and place in ways that illuminate underlying data generating processes is a shift in focus for many epidemiologists.

Second, we call for more representative sampling. We realize that this call may seem controversial, especially given a recent series of debates regarding the importance, or not, of representative sampling in epidemiologic studies. In a recent point–counterpoint, Rothman and colleagues argued that when the goal of an epidemiologic research question is generalizing about disease processes, representative sampling can detract from our ability to identify causal effects and explain how they influence health.[22] As an analogy, articulated both by Rothman and others in the field,[23] neuroscientists study the behavior of mice not to improve the health of the world's population of mice, but to infer the mechanisms of our biological make-up. We do not collect a representative sample of all mice in order to conduct our experiments and build towards scientific inference.

This "mice" argument rests on the premise that the goal of epidemiology in human populations is to simulate, to the extent we can, the conditions of the experimental lab. This idea has had appeal in the field for decades, perhaps because our field's foundations were first and ably proposed by clinicians trained in research through the randomized clinical trial, rather than by social scientists who were more comfortable with other approaches to causal theory such as triangulation, or by clinicians steeped in the tenets of preventative medicine.

These roots have led us to embrace one dominant conceptual paradigm—the counterfactual and potential outcomes approaches —in which the effects of risk factors on outcomes are isolated by careful control of confounding to achieve comparability between the exposure groups. After a causal effect has been identified under a no confounding assumption, it can then be elaborated to understand mechanisms and co-occurring causes. The rigor of the experimental paradigm, where we can vary one or more causes while holding constant the conditions that led up the causal exposures, is held then as the gold standard for estimating causal effects of individual exposures even in non-experimental conditions. From that worldview, representative sampling is worthless.

However, seen from a different perspective, the debate over the usefulness of representative samples underlies the deeper challenge in epidemiology that directly confronts our values as population health scientists. Our quest to isolate the effects of exposures increasingly requires restrictive and purposive samples to achieve comparability, yet that this may actually limit our ability to understand disease processes in populations. This concern is well documented in epidemiologic and causal inference literature, often grouped under the rubric of factors that affect 'transportability'.[24–26] We have, however, largely not heeded these concerns. If we approach our research questions by first asking in whom we want to improve health, rather than what caused poor health, then a focus first on populations—including defining that population and achieving broad representation within it—is critical.

Interestingly, this view has proliferated in other fields such as neuroscience,[27] where it is recognized that non-generalizeable samples may yield data that is scientifically interesting but of little help in improving population health. Of course, we must have valid answers to any questions in order to generalize them beyond our study samples. If representative sampling leads to invalid answers, then it is not worth pursuing. But critically assessing and theorizing about what populations we want to infer to, designing studies that allow us to directly measure population impact, and elevating concerns about external validity to the same level of focus in the field as are our concerns about internal validity,[28] are increasingly necessary. As big data becomes bigger, the resources will be available to achieve such goals, given that we align our values with such perspectives.

One example of such a causal architecture approach could be applied to the study of obesity, an outcome with clear public health relevance and a plethora of identified risk factors. There has been a great deal of effort in quantifying the effect of gene variants such as the fat mass and obesity-associated (FTO) gene on risk for obesity as well as explaining population variance.[29–31] Given that the social, political, and topographic environment is critical to obesity risk, interaction is clearly present. In fact, emerging evidence indicates cohort of birth modifies the effect of the FTO gene,[32] suggesting that the social structures in which we are embedded and develop shape the way in which molecular markers influence our health. Further expansion on the way in which risks for obesity are distributed across populations, interact with each other in dynamic ways, and spread across networks of individuals underlie a causal architecture approach and force us to creatively develop data and designs that articulate the broader structure in which obesity and its related health outcomes are embedded. Examples of causal architecture approaches in epidemiology are also beginning to grow, for example, Westreich (2014) demonstrated how estimates of the effects of pregnancy prevention on virologic failure among HIV-positive women can vary considerably when considering varying characteristics of populations,[33] and clarified the role of population interventions in epidemiologic literature in ways that illustrate the potential role for causal architecture approaches becoming embedded in epidemiologic queries.

Conclusions

We argue in this essay that we should concern ourselves more with a causal architecture approach rather than with our dominant approach. We suggest that this perspective is served by a focus on theorizing about the prevalence of causes and their co-occurrence across populations, using available innovative methods, and by the use of representative samples to allow us to understand the role of context in shaping the health of populations. Our traditional epidemiologic technique is to ask: Is exposure associated with disease? Is exposure associated with disease controlling for a set of confounders? Does exposure interact with any factors to produce disease? What are the mechanisms through which exposure works? We have a well-defined and increasingly sophisticated set of methods for addressing each of these questions, though sophistication and elegance of our modeling approaches alone are no match for the creativity and ingenuity of the researcher who is programming them.[34] A causal architecture approach, on the other hand, would ask: What is the structure of causes that underlie disease? Do these causes work together or separately? And most importantly, which causes are the most prevalent in the population. By ‘structure’,

in this context, we refer not only to directed acyclic graphs and other visual approaches to confounder control that are well documented in epidemiologic methods literature, but to establishing the prevalence, co-occurrence, and networks of causes that produce health within and across populations. It is an alternative framing and recalibration of our goals and our objectives as population health practitioners. We have the methods and the data; what is left is our will as scientists to push forward in these endeavors, and our skill as teachers to develop the next generation of scholars to push beyond the boundaries that we set for them.

Acknowledgments

An early version of this paper was presented at the 2014 Society for Epidemiologic Research meeting in Seattle, Washington.

Funding: Dr. Keyes is supported by a grant from NIAAA (K01AA02151).

References

1. Kuller LH. Point: is there a future for innovative epidemiology? Am J Epidemiol. 2013; 177(4):279–80. [PubMed: 23296359]
2. Ness RB, Andrews EB, Gaudino JA Jr, Newman AB, Soskolne CL, Sturmer T, Wartenberg DE, Weiss SH. The future of epidemiology. Acad Med. 2009; 84(11):1631–7. [PubMed: 19858828]
3. Galea S. An argument for a consequentialist epidemiology. Am J Epidemiol. 2013; 178(8):1185–91. [PubMed: 24022890]
4. Rockhill B. Theorizing about causes at the individual level while estimating effects at the population level: implications for prevention. Epidemiology. 2005; 16(1):124–9. [PubMed: 15613957]
5. Krieger, N. Does epidemiologic theory exist?. In: Krieger, N., editor. Epidemiology and The People's Health: Theory and Context. New York: Oxford University Press; 2011.
6. McMichael AJ. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. Am J Epidemiol. 1999; 149(10):887–97. [PubMed: 10342797]
7. Shy CM. The failure of academic epidemiology: witness for the prosecution. Am J Epidemiol. 1997; 145(6):479–84. discussion 485–7. [PubMed: 9063337]
8. Rose G. Sick individuals and sick populations. Int J Epidemiol. 1985; 14(1):32–8. [PubMed: 3872850]
9. Link BG, Phelan J. Social Conditions as Fundamental Causes of Diease. Journal of Health and Social Behavior. 1994; 35:80–94. [PubMed: 8014431]
10. Terris M. The Society for Epidemiologic Research and the future of epidemiology. J Public Health Policy. 1993; 14(2):137–48. [PubMed: 8408605]
11. Li K, Husing A, Fortner RT, Tjonneland A, Hansen L, Dossus L, Chang-Claude J, Bergmann M, Steffen A, Bamia C, Trichopoulos D, Trichopoulou A, Palli D, Mattiello A, Agnoli C, Tumino R, Onland-Moret NC, Peeters PH, Bueno-de-Mesquita HB, Gram IT, Weiderpass E, Sanchez-Cantalejo E, Chirlaque MD, Duell EJ, Ardanaz E, Idahl A, Lundin E, Khaw KT, Travis RC, Merritt MA, Gunter MJ, Riboli E, Ferrari P, Terry K, Cramer D, Kaaks R. An epidemiologic risk prediction model for ovarian cancer in Europe: the EPIC study. Br J Cancer. 2015; 112(7):1257–65. [PubMed: 25742479]
12. Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. N Biotechnol. 2012; 29(6):613–24. [PubMed: 22450380]
13. Keyes KM, Smith GD, Koenen KC, Galea S. The mathematical limits of genetic prediction for complex chronic disease. J Epidemiol Community Health. 2015; 69(6):574–9. [PubMed: 25648993]
14. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004; 159(9): 882–90. [PubMed: 15105181]

15. Smith GD. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. Int J Epidemiol. 2011; 40(3):537–62. [PubMed: 21807641]
16. Marshall BD, Galea S. Formalizing the role of agent-based modeling in causal inference and epidemiology. Am J Epidemiol. 2015; 181(2):92–9. [PubMed: 25480821]
17. Hernan MA. Invited commentary: Agent-based models for causal inference-reweighting data and theory in epidemiology. Am J Epidemiol. 2015; 181(2):103–5. [PubMed: 25480820]
18. VanderWeele, TJ. Explanation in causal inference: methods for mediation and interaction. New York: Oxford University Press; 2015.
19. Westreich D, Cole SR, Young JG, Palella F, Tien PC, Kingsley L, Gange SJ, Hernan MA. The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. Stat Med. 2012; 31(18):2000–9. [PubMed: 22495733]
20. Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. BMC Cancer. 2012; 12:385. [PubMed: 22943444]
21. Saracci R. The interactions of tobacco smoking and other agents in cancer etiology. Epidemiol Rev. 1987; 9:175–93. [PubMed: 3315716]
22. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol. 2013; 42(4):1012–4. [PubMed: 24062287]
23. Miettinen, O. Theoretical Epidemiology. Albany, NY: Delmar Publishers, Inc; 1985.
24. Pearl, J.; Bareinboim, E. Transportability of Causal and Statistical Relations: A Formal Approach; Proceedings of the 25th AAAI Conference on Artificial Intelligence; August 7–11, 2011; Menlo Park, CA: AAAI Press; 2011. p. 247-254.
25. Hernan MA, VanderWeele TJ. Compound treatments and transportability of causal inference. Epidemiology. 2011; 22(3):368–77. [PubMed: 21399502]
26. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. Am J Epidemiol. 2010; 172(1):107–15. [PubMed: 20547574]
27. Falk EB, Hyde LW, Mitchell C, Faul J, Gonzalez R, Heitzeg MM, Keating DP, Langa KM, Martz ME, Maslowsky J, Morrison FJ, Noll DC, Patrick ME, Pfeffer FT, Reuter-Lorenz PA, Thomason ME, Davis-Kean P, Monk CS, Schulenberg J. What is a representative brain? Neuroscience meets population science. Proc Natl Acad Sci U S A. 2013; 110(44):17615–22. [PubMed: 24151336]
28. Keyes, K.; Galea, S. Chapter 12: Do the results matter beyond the study sample?. In: Keyes, K.; Galea, S., editors. Epidemiology Matters: A New Introduction to Methodological Foundations. New York: Oxford University Press; 2014.
29. Warrington NM, Howe LD, Paternoster L, Kaakinen M, Herrala S, Huikari V, Wu YY, Kemp JP, Timpson NJ, Pourcain BS, Davey Smith G, Tilling K, Jarvelin MR, Pennell CE, Evans DM, Lawlor DA, Briollais L, Palmer LJ. A genome-wide association study of body mass index across early life and childhood. Int J Epidemiol. 2015; 44(2):700–12. [PubMed: 25953783]
30. Park SL, Cheng I, Pendergrass SA, Kucharska-Newton AM, Lim U, Ambite JL, Caberto CP, Monroe KR, Schumacher F, Hindorff LA, Oetjens MT, Wilson S, Goodloe RJ, Love SA, Henderson BE, Kolonel LN, Haiman CA, Crawford DC, North KE, Heiss G, Ritchie MD, Wilkens LR, Le Marchand L. Association of the FTO obesity risk variant rs8050136 with percentage of energy intake from fat in multiple racial/ethnic populations: the PAGE study. Am J Epidemiol. 2013; 178(5):780–90. [PubMed: 23820787]
31. Peng S, Zhu Y, Xu F, Ren X, Li X, Lai M. FTO gene polymorphisms and obesity risk: a meta-analysis. BMC Med. 2011; 9:71. [PubMed: 21651756]
32. Rosenquist JN, Lehrer SF, O'Malley AJ, Zaslavsky AM, Smoller JW, Christakis NA. Cohort of birth modifies the association between FTO genotype and BMI. Proc Natl Acad Sci U S A. 2015; 112(2):354–9. [PubMed: 25548176]
33. Westreich D. From exposures to population interventions: pregnancy and response to HIV therapy. Am J Epidemiol. 2014; 179(7):797–806. [PubMed: 24573538]
34. Cartwright, N. The Dappled World: A Study of the Boundaries of Science. Cambridge: Cambridge University Press; 1999.