

Integrating Complex Systems Thinking into Epidemiologic Research

Ashley I. Naimi

The less a science is advanced, the more its terminology rests on an uncritical assumption of mutual understanding. —Quine¹

Complex systems science is a new field that continues to garner prominence. Arguments for a more widespread adoption of systems science in epidemiology are increasing. These arguments are typically framed in specific ways. First, claims are made that well-established methods in epidemiology are limited, and that these limitations justify the adoption of complex systems methods. Second, complex systems methods have been contrasted with newer (so-called causal inference) approaches, and important differences have been claimed to exist between them. Third, the conduct of complex systems science is often presented as requiring little more than the adoption of a particular set of analytic tools.

This commentary will try to cast these claims under a critical light. Its purpose will be to show that many of them do not stand up to closer scrutiny. With the first two arguments, I will suggest that claims supporting the use of complex systems methods are unconvincing and subject to important fallacies. Assertions about the limitations of traditional epidemiologic methods are often vague, or based on straw-man arguments. Furthermore, claims about the role of certain assumptions required for complex systems methods have led to an informal fallacy known as an “irrelevant conclusion,” where the arguments fail to address the issue in question. Finally, I will propose that complex systems science has little to do with a particular set of methods, but consists of an approach to framing questions about population health that necessitates the convergence of several disparate scientific areas. The central thesis of this commentary is that integrating “systems thinking” into epidemiologic research is of central importance, and will require a more rigorous treatment of the relations between these two fields.

COMPLEX SYSTEMS SCIENCE & POPULATION HEALTH RESEARCH

Precisely defining what constitutes a complex system is itself a complex matter.² However, a complex system can be *described* as one whose properties are not fully explained by an understanding of its component parts.³ The degree of *complexity* in a system can loosely be defined as the amount of *information* needed to describe it, where information is characterized as a function of the number of possible states in which a given system may exist.^{4(P. 12)} *Complex systems science* can be defined as the study of the properties of and the relations between units of a complex system, and the study of the relations between units and their environments that give rise to collective or emergent features of that system.⁵ This definition consists of several key components that merit emphasis: complex systems have *units*;

From the Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA.
The author reports no conflicts of interest.

Correspondence: Ashley I. Naimi, Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, 130 DeSoto Street, 503 Parran Hall, Pittsburgh, PA 15261. E-mail: ashley.naimi@pitt.edu.

Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 1044-3983/16/2706-0843
DOI: 10.1097/EDE.0000000000000538

these units have *properties*; these units share *relations*; and these units exist in *environments*.

However, the operational definition of complex systems science is quite different in the public health and epidemiologic literature. This definition is often made by claiming that typical approaches to knowledge generation in epidemiology are based on illegitimate and restrictive assumptions, and that complex systems approaches avoid these assumptions. While several examples exist,^{6–9} the Table, reproduced from Luke and Stamatakis,¹⁰ is an instructive case in point. Using this table, the authors attempt to make the case for “a mismatch between the characteristics and assumptions of traditional data analysis approaches...and the characteristics of the data...from complex systems.”^{10(P. 360)}

DEFINING AND HANDLING NONLINEARITY IN EPIDEMIOLOGY

Upon closer inspection, several questions arise about the comparisons made in the Table. Take, for example, the claim that traditional approaches fail to account for nonlinearity. Despite the claim’s ubiquity,^{10–17} the term “nonlinear” is seldom defined. This term has been used to describe

R1: The presence of certain link functions (e.g., logistic, logarithmic).¹⁸

R2: The situation in which a model’s *parameters* are characterized by polynomial, trigonometric, or other functions.^{19(P. 5–10)}

R3: The situation in which a model’s covariates are characterized by spline, polynomial, or other transformations.²⁰

R4: The situation in which a given individual’s exposure affects the outcome of other individuals.¹⁴

This array of explanations makes it difficult to judge the merit of claims that traditional methods cannot handle “nonlinearity.” These difficulties are only compounded in the absence of context about study objectives. The claims may well be true when the goal is to visualize the degree of segregation that arises when individual agents are subject to a set of

behavioral rules.²¹ But when the goal is to predict or compare means, risks, rates, or odds, traditional methods can account for several types of nonlinearity with ease. In fact, Nelder and Wedderburn’s²² seminal contribution enabled fitting nonlinear models when the outcome belongs to the exponential family of distributions, and the conditional mean of the outcome can be linked to the covariates through some smooth and invertible linearizing function. These *generalized linear models*, perhaps the most standard method in our field, were expressly developed to handle nonlinearities described in scenarios R1 and R2, and easily accommodate “nonlinearities” encountered in scenario R3.²³ But what of R4?

Philippe and Mansi¹⁴ correctly describe the situation in which an individual’s exposure affects another’s outcome as “structural interaction” or dependent happenings. But they go on to suggest that this leads to “nonlinearity” that cannot be examined using “linear” methods of analysis. This situation was long ago recognized by Ross²⁴ and described by Cox²⁵ as “interference.” Recently, several authors have provided definitions of direct, indirect (or spillover), total, and overall effects in the presence of such interference.^{26–30} Though interference certainly complicates analysis, and methods such as agent-based modeling may usefully capture this complexity, such effects are not *a priori* incompatible with standard “linear” methods of analysis. For example, VanderWeele et al.³¹ note how contrasts defined under interference can be estimated using standard maximum likelihood or estimating equation techniques.

Under more careful scrutiny, the argument that “nonlinearity” necessitates the use of so-called complex systems methods (e.g., agent-based models, systems dynamic models) breaks down. At the very least, such arguments are subject to vagueness (due to the lack of a clear definition of “nonlinearity”) or equivocation (enabling the evasion of criticism by switching between the many definitions of “nonlinearity”). More consequentially, the alleged need for “complex systems methods” is based on a straw-man argument in which the properties of standard epidemiologic methods are misrepresented to justify the need for alternative approaches.

This critique applies specifically to claims about nonlinearity. However, similar arguments can be made about other entries in Table. Extensive literature exists, for example, on how to deal with dynamic feedback loops (often referred to as time-varying confounding),³² multiple levels of analysis,³³ and the violation of parametric assumptions, such as normally distributed error terms.³⁴

NATURE AND SCOPE OF ASSUMPTIONS IN COMPLEX SYSTEMS SCIENCE: EXCHANGEABILITY AND VALIDITY

Another claim has arisen regarding assumptions required when using agent-based models. For example, El-Sayed et al.^{12(P. 6)} argue that agent-based models “bypass Holland’s ‘Fundamental Problem of Causal Inference’...” which entails the need for several assumptions to infer causation. Similarly, Marshall and Galea^{11(P. 96)} argue that to estimate average causal

TABLE. Comparison of Traditional and Complex System Analytic Assumptions, Reproduced from Luke and Stamatakis, 2012¹⁰

Domain	Traditional Analytic Techniques Assumptions	Complex Systems Assumptions
Functional form	Linearity	Nonlinearity
Common distributions	Normality	Non-normality
Characteristics of actors	Homogeneity	Heterogeneity
Level of analysis	Single level	Multiple levels
Temporality	Static or discretely longitudinal	Dynamic, with feedback
Fundamental relationships	Among variables	Interaction of actors
Perspective	Reductionist	Holistic

effects, “several standard causal inference assumptions are not relevant in agent-based modeling (e.g., exchangeability is assured by design).”

The claim that exchangeability is not required with agent-based models is true in a very specific and critically important sense. In an agent-based simulation model, both “experimental” and “control” interventions are applied to the same population of agents. Thus, “[a]ssuming no changes in initial conditions and update rules, each simulation is identical within the bounds of stochasticity,”^{12(P. 6)} and the condition (e.g., all other things being equal) underlying assumptions such as exchangeability are met by definition.

Yet this reasoning reveals precisely how agent-based models bypass key causal inference assumptions. With this approach, one seeks to ascertain the behavior of the simulated environment under different “computational interventions.” The object of inference is no longer the population system under study, but the computational model itself. However, if interest lies in estimating the effect of changing a line of programming code on the resulting simulated data, the same line of reasoning can be used to claim that exchangeability is not relevant for *any* modeling strategy (from simple logistic regression models, to more involved approaches such as the parametric g-formula). In effect, any method can be used to “bypass” the fundamental problem of causal inference by expressing interest in the simulated environment rather than the (human) population of interest.

To illustrate this point, consider a recent agent-based network model that sought to assess whether hypothetical interventions specifically targeting highly networked individuals would yield larger reductions in the prevalence of obesity than more generally targeted interventions.³⁵ In addition to several system features, the authors parameterized the effect of one person’s (ego) obesity status on another’s (alter) obesity state using a logistic regression model with an odds ratio of 1.16, taken directly from the work of Christakis and Fowler.^{36(Table S2)} Given this assumed effect, they examined the behavior of their agent-based model under different hypothetical interventions, and found that targeting highly networked simulated agents did not outperform an “at random” targeting strategy.

However, several authors have questioned the validity of Christakis and Fowler’s results.^{37–39} Indeed, the most commonly cited concern is whether their estimated associations could be explained by confounding due to latent homophily,⁴⁰ in which two individuals become connections in a network because of shared latent characteristics. Using directed acyclic graphs, it has been shown that homophily results in a *d*-connected path between the ego’s prior obesity status (the exposure) and the alter’s subsequent obesity state (the outcome).³⁹ In other words, the population of alters connected to obese egos may not be exchangeable with the population of alters connected to nonobese egos.⁴¹ Thus, while this agent-based model provides information on the properties of the

simulation under different computational interventions, one must assume exchangeability (among other causal inference assumptions) to infer that targeting highly networked *individuals* will yield the same obesity distribution as targeting highly networked *agents* in the simulated environment.

THE SUBSTANCE OF SYSTEMS SCIENCE

A final consideration worth noting is the distinction between methodology and complex systems science. Implied in much of the research using so-called complex systems methods is the notion that employing such techniques amounts to complex systems science. Some authors have even stated as much.^{42(P. 1157),8(P. 1298)} Such a conflation has occurred in other areas of study as well. For example, in the field of causal inference it would be conceivable, but incorrect, to assume that the use of a particular method (e.g., marginal structural models or targeted minimum loss-based estimation) is what ultimately confers a causal interpretation.^{43,44}

In contrast to this methods-based characterization is a more general portrayal. Boulding,^{45(P. 208)} for example, long ago noted that systems science “provides a framework or structure on which to hang particular disciplines in an orderly and coherent corpus of knowledge.” More contemporary authors have also made a clear distinction between systems thinking and modeling.⁴⁶ This distinction has important implications for epidemiologists. If, for example, interest lies in interventions that will alter the distribution of obesity in the population, complex systems science prompts us to consider a host of inter-related attributes, structures, and organizations potentially affecting the distribution of obesity, including individual physiology, the social environment, agricultural systems, the built environment, and political and economic systems.

CONSTRUCTIVELY QUESTIONING COMPLEX SYSTEMS SCIENCE IN PUBLIC HEALTH RESEARCH

The current state of the literature on complex systems science in epidemiology is growing. Yet critical terminology remains abstruse and crucial questions unarticulated. This is not surprising, considering the recency of the field. Indeed, it is no irony that the opening quote in this commentary is cited in a landmark text on complex systems science with agent-based models.^{47(P. 1)} To that end, one might pose several questions related to complex systems thinking in epidemiologic research. These include

1. Do systems methods (e.g., agent-based models, systems dynamic models) really account for nonlinearity in ways that standard statistical approaches don’t? If so, what kind(s) of nonlinearity, and how specifically do they account for it?
2. How do the theoretical, computational, and empirical properties of systems methods compare to their “standard” counterparts. Some of this work has begun in other fields.⁴⁸

3. How is “feedback” different from time-dependent confounding? How, specifically, do systems methods account for this feedback?
4. How are the causal inference assumptions (e.g., exchangeability, positivity, consistency, correct model specification) relevant when agent-based models or other systems tools are employed?
5. What are the benefits and tradeoffs of adopting “systems methods” versus engaging in “systems science,” more broadly defined?

Shannon,⁴⁹ the founder of information theory, long ago cautioned that the increasingly widespread application of concepts and tools of his field to other scientific domains, although exciting, carried “an element of danger.” Establishing interdisciplinary applications of information theory, he argued, “is not a trivial matter of translating words to a new domain,” but requires “a thorough understanding of [its] mathematical foundation” and “the slow and tedious process of hypothesis and experimental verification.” Given the logical and experimental foundations of our field, epidemiologists have much to contribute to the work of integrating complex systems science into population health research.

ABOUT THE AUTHOR

ASHLEY I. NAIMI is an Assistant Professor at the University of Pittsburgh. His research focus is on causal inference and epidemiologic methods in relation to the social determinants of adverse reproductive and perinatal outcomes.

ACKNOWLEDGMENTS

I thank Drs. Sander Greenland, Jay S. Kaufman, Stephen R. Cole, and Cosma Shalizi for helpful comments on a previous version of the manuscript. I thank Dr. Kamran Sedig for clarifying the definition of complex systems science.

REFERENCES

1. Quine WV. Truth by convention. In: Benacerraf P, Putnam H, eds. *Philosophy of Mathematics: Selected Readings*. Cambridge: Cambridge University Press; 1936: 329–354.
2. Ladyman J, Lambert J, Wiesner K. What is a complex system? *EJPS*. 2013; 3:33–67.
3. Gallagher R, Appenzeller T. Beyond reductionism. *Science*. 1999; 284:79.
4. Bar-Yam Y. *The Dynamics of Complex Systems*. Reading, MA: Perseus Books; 1997.
5. Bar-Yam Y. General features of complex systems. In: Keil D, ed. *Knowledge Management, Organizational Intelligence and Learning, and Complexity*, vol. 1. United Kingdom: EOLSS Publishers Co Ltd; 2002: 43–95.
6. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. *Am J Epidemiol*. 2008;168:1–8.
7. Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. *Int J Epidemiol*. 2010;39:97–106.
8. Ip EH, Rahmandad H, Shoham DA, et al. Reconciling statistical and systems science approaches to public health. *Health Educ Behav*. 2013;40(1 suppl):123S–131S.
9. Nianogo RA, Arah OA. Agent-based modeling of noncommunicable diseases: a systematic review. *Am J Public Health*. 2015;105:e20–e31.
10. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. *Annu Rev Public Health*. 2012;33:357–376.
11. Marshall BD, Galea S. Formalizing the role of agent-based modeling in causal inference and epidemiology. *Am J Epidemiol*. 2015;181: 92–99.
12. El-Sayed AM, Scarborough P, Seemann L, Galea S. Social network analysis and agent-based modeling in social epidemiology. *Epidemiol Perspect Innov*. 2012;9:1.
13. Pearce N, Merletti F. Complexity, simplicity, and epidemiology. *Int J Epidemiol*. 2006;35:515–519.
14. Philippe P, Mansi O. Nonlinearity in the epidemiology of complex health and disease processes. *Theor Med Bioeth*. 1998;19:591–607.
15. Jayasinghe S. Conceptualising population health: from mechanistic thinking to complexity science. *Emerg Themes Epidemiol*. 2011;8:2.
16. Resnicow K, Page SE. Embracing chaos and complexity: a quantum change for public health. *Am J Public Health*. 2008;98:1382–1389.
17. Carvalho MS, Coeli CM, Chor D, Pinheiro RS, Fonseca Mde J, Sá Carvalho LC. The challenge of cardiovascular diseases and diabetes to public health: a study based on qualitative systemic approach. *PLoS One*. 2015;10:e0132216.
18. Tchetgen Tchetgen EJ, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context. *Epidemiology*. 2015;26:402–410.
19. Seber GAF, Wild CJ. *Nonlinear Regression*. New York, NY: Wiley; 1989.
20. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009;20:18–26.
21. Schelling TC. Dynamic models of segregation. *J Math Sociol*. 1971; 1:143–186.
22. Nelder JA, Wedderburn RWM. Generalized linear models. *JRSS-A*. 1972; 135:370–384.
23. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*. 1995;6: 356–365.
24. Ross R. An application of the theory of probabilities to the study of a priori pathometry. Part I. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 1916; 92:204–230.
25. Cox DR. *Planning of Experiments*. New York, NY: John Wiley & Sons; 1958.
26. Struchiner CJ, Halloran ME, Robins JM, Spielman A. The behaviour of common measures of association used to assess a vaccination programme under complex disease transmission patterns—a computer simulation study of malaria vaccines. *Int J Epidemiol*. 1990;19: 187–196.
27. Halloran ME, Struchiner CJ. Study designs for dependent happenings. *Epidemiology*. 1991;2:331–338.
28. Halloran ME, Struchiner CJ. Causal inference in infectious diseases. *Epidemiology*. 1995;6:142–151.
29. Hudgens MG, Halloran ME. Toward causal inference with interference. *J Am Stat Assoc*. 2008;103:832–842.
30. Tchetgen Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. *Stat Methods Med Res*. 2012;21:55–75.
31. VanderWeele TJ, Tchetgen Tchetgen EJ, Halloran ME. Interference and sensitivity analysis. *Stat Sci*. 2014;29:687–706.
32. Robins J, Hernán M. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. *Advances in Longitudinal Data Analysis*. Boca Raton, FL:Chapman & Hall; 2009: 553–599.
33. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge; New York:Cambridge University Press; 2007.
34. Tsiatis AA. *Semiparametric Theory and Missing Data*. New York, NY:Springer; 2006.
35. El-Sayed AM, Seemann L, Scarborough P, Galea S. Are network-based interventions a useful antiobesity strategy? An application of simulation models for causal inference in epidemiology. *Am J Epidemiol*. 2013;178:287–295.
36. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med*. 2007;357:370–379.

37. Lyons R. The spread of evidence-poor medicine via flawed social-network analysis. *Stat Politics Policy*. 2011; 2:Article 2.
38. Noel H, Nyhan B. The “unfriending” problem: the consequences of homophily in friendship retention for causal estimates of social influence. *Soc Networks*. 2011; 33:211–218.
39. Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. *Sociol Methods Res*. 2011;40:211–239.
40. VanderWeele TJ. Sensitivity analysis for contagion effects in social networks. *Sociol Methods Res*. 2011;40:240–255.
41. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci*. 1999; 14:29–46.
42. Mabry PL, Bures RM. Systems science for obesity-related research questions: an introduction to the theme issue. *Am J Public Health*. 2014;104:1157–1159.
43. Hogan JW. Causal models. *Epidemiol*. 2009; 20:931–932.
44. Levine B. Causal models. *Epidemiology*. 2009;20:931; author reply 931–931; author reply 932.
45. Boulding KE. General systems theory—the skeleton of science. *Manag Sci*. 1956; 2:197–208.
46. Trochim WM, Cabrera DA, Milstein B, Gallagher RS, Leischow SJ. Practical challenges of systems thinking and modeling in public health. *Am J Public Health*. 2006;96:538–546.
47. Epstein J. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, NJ: Princeton University Press; 2006.
48. Grazzini J, Richiardi M, Sella L. Small sample bias in MSM estimation of agent-based models. In: Teglio A, Alfarano S, Camacho-Cuena E, Ginés-Vilar M, eds. *Managing Market Complexity: The Approach of Artificial Economics*. Lecture Notes in Economics and Mathematical Systems 662, Springer-Verlag. 2012; 237–47.
49. Shannon CE. The Bandwagon. *IRE Transactions on Information Theory*. 1956; 2:3.