VALIDATING MODELS IN PUBLIC HEALTH RESEARCH

Amber D. Elkins
Dennis M. Gorman
School of Public Health, Texas A&M Health Science Center,
TAMU 1266, College Station, TX 77845, USA

ABSTRACT

The application of systems method to the understanding of public health problems (e.g., alcohol and drug abuse, chronic disease, obesity, tobacco use, and violence) has grown considerably in the past decade. System methods are seen by many of their advocates within public health as complimenting traditional behavioral and epidemiological research methods, while others see them as a fundamentally different way of understanding and explaining public health problems. Those who see the methods as complimentary often use empirical data from studies employing traditional methods and statistical analysis to validate the output of simulation models. As in other fields of applied research in which modeling has become popular, this tendency to equate a model's correspondence to data with the model corresponding to reality is especially pronounced when the goal of the modeling is to inform public policy. The present paper discusses the problems that arise when using data from an empirical study to assess the validity of a simulation model. It illustrates these problems through an examination of a specific example from the public health literature. The example demonstrates that, rather than empirical data being superior to the model, each is better considered as simply capturing a different aspect of a real-world system. Alternative means of assessing model usefulness are also discussed.

Keywords: Model validation, system simulation, public health.

INTRODUCTION

The application of systems methods (notably system dynamics modeling, agent-based modeling, and social network analysis) to the understanding of a wide range of public health problems has grown considerably in the past decade (Elkins & Gorman, 2014; Galea, Riddle & Kaplan, 2010; Luke & Stamatakis, 2012). Much of the impetus for this has come from recognition of the complexity of many public health problems and a search for analytic methods better able to capture the underlying dynamic processes at work compared to traditional study designs and statistical approaches. The limits of traditional research designs (e.g., randomized trials and cohort studies) and the statistical analyses typically used to analyze data from such studies (e.g., regression analyses and descriptive statistics) have become especially noticeable in research on public health problems where multiple heterogeneous interacting elements produce emergent, population-level effects that involve feedback mechanisms and develop in a non-linear fashion (Diez Roux, 2011; Hammond, 2009; Luke & Stamatakis, 2012). Such problems include alcohol abuse, drug use, violence, obesity, tobacco-use, and chronic diseases (e.g., diabetes and heart disease), all of which are conditions particularly resistant to traditional individual-level interventions (McKinley & Marceau, 1999; Susser, 1995). Systems methods, it is argued, can be used not only to better understand the complexity of such problems but also to identify leverage points for interventions and to assess potential effectiveness of different types of policies and programs designed to influence population-level health (Hawe, Shiell, & Riley, 2009; Trickett et al., 2011). Thus, as in other fields of study, the attraction of systems methods in public health resides not only in their promise to provide better understanding of natural and social phenomena but to also to provide a means of ameliorating societal problems (see Oreskes, 1998). While such methods provide a means for studying and ameliorating societal problems, such benefits only come with proper application of those methods.

System methods are also viewed in different ways by public health researchers. Specifically, they are seen by many of their advocates within public health as complimenting traditional behavioral and epidemiological research methods, and as in no way an attempt to displace such methods (Kaplan, 2013). Others however see them as a fundamentally different way of understanding and explaining public health problems, and as presenting a "challenge" to traditional research methods (Luke & Stamatakis, 2012). Those who see the methods as complimentary often use empirical data from studies employing traditional methods and statistical analysis to validate the output of simulation models. Alfred Korzybski (1933) famously stated: "The map is not the territory," yet this predominant approach to model validation in public health research assumes traditional empirical methods and statistical techniques capture the "territory" with such accuracy that they can be used as a yardstick against which to judge the performance and adequacy of a model.

The current paper questions whether public health simulation models can, and should, be "validated" through comparison to empirical data. The next section briefly describes the underlying rationale for this approach to model valiation. The third secion of the paper examines some of the problems with this approach that have been raised in the broader modeling literature. This is followed by a detailed discussion of a specific example from the public health literature that illustrates these problems. The example is a system dynamics model of college drinking, developed by Scribner and colleagues, which is comprised of five compartments (abstainer through heavy episodic drinker) and three processes governing transitions (social norms, social interactions and individual risk) (Ackleh et al., 2009; Scribner et al., 2009). We specifically focus on the comparison of the model output with data from a survey of college drinking behavior, the Social Norms Marketing Research Project (SNMPR) (DeJong et al., 2006; Scribner et al., 2008). Following the examination of this specific example, the paper concludes with a disussion of some other approcahes to model evaluation that might be more useful in assessing public health systems models.

MODEL VALIDATION

The rapidity of the adoption of systems methods within the field of public health has meant that some of the underlying philosophical issues regarding the use of such methods have not been explored and debated in much detail. One such issue that is particularly pronounced with those models intended to help solve social problems is the need to demonstrate that they resemble with some degree of accuracy the real-world systems to which they pertain (Oreskes, 1998). The closer the resemblance, so the reasoning goes, the more justified one is in conducting virtual experiments using the simulation model and the more confidence one can have that the results of such experiments can guide interventions and policies in the real-world system. The term most commonly used to describe the assessment of a simulation model in terms of how well it resembles the real-world system to which it pertains is "validation."

We acknowledge at the outset that the term "validation" is highly contested within the modeling community, and is frequently confused with related terms such as "verification", "accreditation", or "evaluation" (Balci, 1997; Grant & Swannack, 2008; Kleindorfer, O'Neill, & Ganeshan, 1998; Martis, 2006; Oreskes, Shrader-Frehette, & Belitz, 1994). There is also a wide range of activities that can be described under the general rubric of validation (Grant & Swannack, 2008; Rykiel, 1996). A review of the broader debate as to what constitutes "validation" and of the various activities that this term is used to describe is outside the scope of this paper. Instead, our focus is on the process of comparing model predictions with observations of the real-world system, a process that is often erroneously considered to be the only or primary validation criterion (Grant & Swannack, 2008).

Not surprisingly given its emphasis on solving societal problems, the demand that systems models be validated in terms of their correspondence to the real-world system is prevalent in the public health research literature. In addition, as in other areas in which models are validated using such a criterion, the standard approach to assessing the model's correspondence to reality is to compare it to the results of an empirical study. Thus, a common first step used in models that attempt to assess the effects of policies and prevention initiatives is to compare the model output to historical trends in the conditions that are the target of the intervention (e.g., Homer, Hirsch & Milstein, 2007; Jones et al., 2006). In assessing the validity of the model, its output is usually compared to the results obtained from empirical studies of the same phenomenon. So for example, from this perspective the expectation is that a valid model of the effects of low-level environmental exposure to lead in children should generate output that resembles empirical data pertaining to lead poisoning among children who have experienced low-level exposure (see Oreskes, 1998). The underlying assumption of such an approach is that: "Empirical data can help make model input assumptions as valid as possible *and can be used to test the output of models and their power to explain real-world phenomena of interest*" (Hammond, 2009, pages 5-6, emphasis added). The more the model can reproduce the historical data, the more confidence one can have in its ability to predict future trends under conditions of different policy options (Homer, 1996). This has long been a common practice within the field of modeling, and often involves a subjective assessment of the "see how well the simulated data matches the observed data test" (Rykiel, 1996, page 242).

PROBLEMS WITH VALIDATING OPEN SYSTEMS

As noted above, one of the underlying assumptions of the approach to model validation that focuses on comparing model output to data is that the latter captures with some accuracy the underlying dynamics of the real world system that it is measuring. At its extreme, this would look like Figure 1, with a perfect match occurring between the empirical data and the real-world system. In the overwhelming majority of research project, however, such perfection is unattainable due to problems such as selection bias, residual confounding and measurement error, and so an exact mapping of the data onto the real-world system is unlikely. However, one can assume that the match between the two is considered to be good by those who compare empirical data to model output as a means of validating the latter. For were there not some expectation that the data resemble the real-world system with some accuracy then there would be no point in comparing the output of the simulation model to the data as a means of generating confidence in the model's ability to predict the future state of the system.

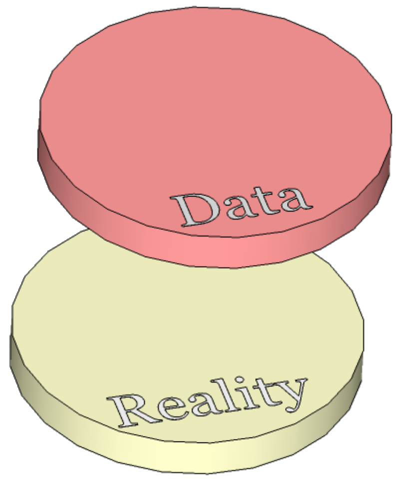

Figure 1: Idea Empirical Data Perfectly Captures the Real-World System

At a very basic level, judging the validity of simulation models in terms of results from empirical studies that use traditional research designs to collect data that is then analyzed using standard statistical methods is somewhat paradoxical. For, as noted earlier, one of the primary reasons for use of such models is that they provide an understanding of phenomena in terms of feedback, nonlinearity, and emergent properties that cannot easily be captured using traditional research designs and statistical methods. Thus, using traditional techniques to "validate" systems methods is at odds with the idea that the latter are, to use the term employed by Luke and Stamatakis (2011), a "challenge" to the former.

Beyond this, however, there are deeper philosophical issues with the assumption that the validity of a simulation model be judged in terms of how well it resembles or corresponds to data from an observational or analytic study. The philosophical roots of the critique of using data from empirical studies to validate simulation models emerged from the constructivist and anti-foundationalist schools of systems theory which challenge, to varying degrees, the idea that there exists a single reality that can be accurately measured and against which a model can be judged (for details see Kleindorfer, O'Neill, & Ganeshan, 1998). In recent years, Oreskes has presented a clearly articulated argument against the use of empirical research to validate models and has highlighted the marked tendency to use such an approach in applied areas of research (Oreskes, 1998; 2003; Oreskes, Shrader-Frechette & Belitz, 1994).

Oreskes' critique of the use of data from empirical studies to validate simulation models is founded upon the fundamental issue that the vast majority of such models of natural and social phenomena are open systems, that is they are inevitably incomplete or partial representations of the natural systems to which they pertain. More specifically, this openness falls within three general categories (Oreskes, 2003; see also Oreskes, 1998 in which this issue is discussed in terms of four similar categories called "flaws"). First, the way in which we conceptualize models is always incomplete, either because we deliberately choose to leave certain features out, or because we are unaware of all of the important features, or because we are mistaken or misguided about the nature of the problem. For example, in Scribner et al.'s (2009) conceptual model of college drinking (which is discussed in more detail below) the three "underlying processes" identified are social norms, social interactions, and individual risk. Other factors that might affect college drinking (e.g., price of alcohol, advertising, availability of other drugs, presence of prevention and treatment services) were deliberately excluded. Such a partial representation of the real system does not make for a "bad model", but it does make for an "open" model according to Oreskes (2003).

The second way in which models are open according to Oreskes (2003) is in terms of how well the numerical variables represent the core elements of the system. All models contain constructs whose qualities can only be partially discerned and distinguished, and assigning values to these qualities will often involve significant error (see also Rykiel, 1996). In the Scribner et al. (2009) model, "the essential features related to patterns of college drinking" are represented by five drinking compartments: abstainers, light drinkers, moderate drinkers, problem drinkers and heavy episodic drinkers (Scribner et al., 2009, page 806). Each compartment is assumed to contain individuals who are similar with regard to their drinking behavior. This is a reasonable simplifying assumption, but fewer or more compartments might have been used to represent the essential features of the model (e.g., Mubayi et al., 2011), or each drinker may have been assumed to be unique as would occur in an agent-based model (e.g., Fitzpatrick & Martinez, 2012).

Finally, what Oreskes (2003) describes as openness is also evident in how well the mathematical equations used in the model to capture the processes of interest. In the Scribner et al. (2009) model, for example, the social norm construct is operationalized as the rate at which individuals transition between drinking states (e.g., from abstainer to light drinker) and this is based on their perception that a certain level of drinking is typical among *all students on campus*. This is a perfectly reasonable assumption, but it might also be argued that such transition between drinking states is driven less by the drinking behavior of all student drinkers at a university or college than it is by the drinking behavior of one's immediate peers. Or it might be that the transition of drinkers within the same category is influenced by the context within which interactions take place (e.g., Mubayi et al., 2011). Thus, the mathematical equations used in the model could include a quite different transition rate.

It is worth noting that these three aspects of what Oreskes (2003) calls "openness" are also evident in the empirical or observational studies with which model output is compared in the validation process. Like simulations, empirical studies are almost always based on partial theories or conceptual models, and the concepts that comprise these are frequently abstract and vague in nature (e.g., social norms, peer group) (Babbie, 1995, pages 75-76). Likewise, the data collected are frequently based on inference-laden operational measures (e.g., "peers" are those with whom an individual attends school) and are often incomplete or inaccurate (e.g., due to non-response, attrition and faulty recall of subjects). Finally, the statistical analyses employed also have built in assumptions about the nature of the data and the relations between variables (e.g., that the data are normally distributed and the relationships are linear). Thus, comparing output from a simulation model of *phenomenon X* with the results of a statistical analysis of data from an empirical study of *phenomenon X* is a comparison of two partially and imperfectly captured systems of *phenomenon X*. They are, as Rykiel observes, "two moving targets that we try to overlay one upon the other" (Rykiel, 1996, page 235).

Figure 2 illustrates the idea that empirical data and the model output capture different aspects of the real-world system, wherein the model not matching the data might be a function of each capturing different aspects of a real-world system instead of the data being a superior representation against which the adequacy of the model output is to be judged. Accordingly, the data may not constitute the best test of the model (Rykiel, 1996). One of the implications of such a view of empirical data and simulation output is that it opens the door to the possibility that the latter may actually be a better representation of the real-world system of interest than the former for some purposes (Eck & Liu, 2008; Rykiel, 1996).

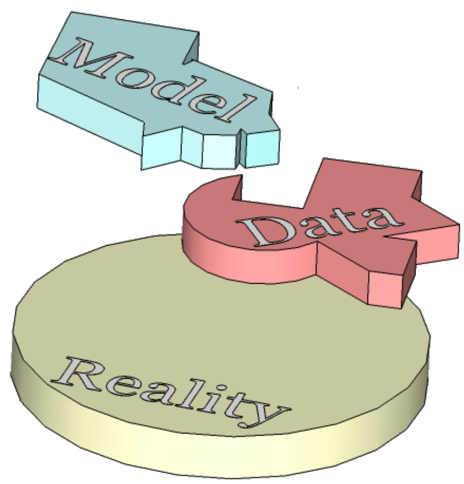

Figure 2. Model Output and Empirical Data Capture different Aspects of the Real-World System

EXAMPLE FROM PUBLIC HEALTH RESEARCH

We will explore these implications in more detail through an examination of Scribner et al.'s (2009) system dynamics model of college drinking, and specifically the use of data from the SNMRP to validate the predictions of the model" (Scribner et al., 2009, page 811). This is a good example to use to illustrate some of the issues raised by Oreskes (2003) concerning model validation as Scribner and colleagues explicitly state that they use survey data "to validate the predictions of the model" (Scribner et al., 2009: 811). In addition, in an earlier paper they state that the "obvious value" of comparing the model output to data "is that once the model has been validated with data, it can be used to make predictions" (Ackleh et al., 2009, page 497). Thus, there is explicit acceptance of the idea that empirical data can be used as a standard against which to assess the validity of simulation models.

We should make it clear however that we are not presenting a general critique of the model presented by Scribner et al. (2009). Indeed, we consider it an eloquent model that has yielded valuable insights into the nature of college drinking and allowed assessment of the possible effects of different policies targeted at this problem (see Fitzpatrick et al., 2012; Rasul et al., 2011). Rather, we are simply questioning whether there is much to gain from comparing the output of the model to the results obtained from an empirical study, and more specifically whether the data say very much about the usefulness and heuristic value of the model.

The model validation presented by Scribner et al (2009) involved two comparisons of the model output and the SNMRP data, one focused on the model's ability to predict the proportion of drinkers in each of the five drinking categories and one focused on the model's ability to predict the alcohol outlet density of campuses. With regard to the former, the analysis presented by Scribner et al. (2009) focused on four of the 32 campuses included in the SNMRP, representing a range from relatively low alcohol outlet density to relatively high alcohol outlet density (defined as the number of bars per undergraduate student within three miles of the campus; Scribner et al., 2008). Specifically, the analysis presented involved a comparison of the predictions of the model over a four-year period for each of the four campuses with the proportion in each compartment found in the SNMRP data. In nearly all of the 80 comparisons presented (5 drinking categories x 4 campuses x 4 years), the empirical data fell within the standard proportion estimator error bars generated by the model (see Figure 3 of Scribner et al., 2009).

With regard to the density comparison, the model is said to have done a "reasonable job" in predicting outlet density for each of the four campuses (Scribner et al., 2009:814): specifically, as the empirical measure of density increased (from 5.25 to 32.81), so did the index generated by the model (from .01 to .77). However, there was almost no difference between the middle two campuses on the index (.23 and .24), whereas the bar-density of the two as measured by the survey was quite different (10.75 and 16.23). Extending this analysis to the entire SNMRP sample of campuses, Ackleh and colleagues found "…that all the 32 fits were quite satisfactory, with the model output within two standard deviations of the data" (Ackleh et al., 2009, p. 491). They also compared the model's alcohol density index for all 32 campuses with the empirical measure of the physical availability of alcohol and obtained an R^2 of 0.2293, which increased to 0.3112 when only the 20 residential campuses were included in the analysis (Ackleh et al., 2009).

Thus, in the case of the proportion of individuals in each drinking compartment, and to a lesser extent the bar density of the four campuses, Scribner et al.'s (2009) system dynamics model is able to predict with some accuracy the empirical data pertaining to each campus that was collected in the SNMRP. But does this degree of correspondence validate the model? Or is the SNMRP survey simply capturing a part of the real-world college drinking system which may or may not be a good representation of this, and therefore may or may not tell one much about the value of the system dynamics model? Is it, as Oreskes (2003) would argue, an open system that cannot be used to validate the model, which is a different open system?

Concern that the SNMRP data pertaining to drinking categories might only provide a partial view of the real-world system it is intended to capture (and hence be a very limited yardstick against which to validate the model) seems reasonable when one examines these data in a little detail. This shows the response rate across the four years of the study was just 53% (n=19,838), and that the final analysis sample was further reduced to those with available data for all variables, decreasing it from 19,838 to 17,051 students (Scribner et al., 2008, page 113). Additionally, the survey questions used are open to varying interpretation by respondents. For example, one question asked, "During the past 30 days, on how many occasions did you use alcohol (beer, wine, liquor)?" and responses choices included, "never," "1-2 times," "3-5 times," "6-9 times," "10-19 times," "20-39 times," and "40 or more times" (Scribner et al., 2008, page 114). The survey question did not provide a definition of "occasions," allowing room for varying definitions (e.g., an occasion might be a party lasting a few hours or a three-day vacation). In addition, the response categories are fairly broad: one could drink 20 times during the past 30 days or 35 times during the past 30 days, but each would receive the same score. While there is nothing wrong with this *per se*, it is likely that these issues pertaining to response rate and measurement uncertainty will produce a dataset that bears more of a resemblance to the one depicted in Figure 2 than the one depicted in Figure 1. Thus, the extent to which these data validate the model and increase confidence in its ability to predict changes in the real-world system is questionable.

The alcohol outlet data used to validate the model are much less subject to selection and reporting bias as these were obtained from the alcohol control boards in the states in which the 32 campuses were located, and which license alcohol outlets such as bars and package goods stores. Only one state was unable to provide such license data (and in this case, project staff visually recorded the outlets close to the campus) and 96% of the outlets were successfully geocoded to a street address. However, the type of uncertainty in how well numerical variables represent the core elements of a system that Oreskes (2003) observes makes for an open system was certainly present in turning these data into a measure of outlet density. In the validation exercise, this was operationalized in terms of the number of bars per undergraduate student within three miles of the campus (Scribner et al., 2009, page 814). Outlets other than bars might have been included. Indeed, off-sale outlets were included in the SNMRP dataset, but were found to be much less densely concentrated around campuses (see Table 1 of Scribner et al., 2008). Alcohol outlet density could also have been calculated by outlets-per-roadway-mile, rather than per 1,000 students. Buffers other than 3-miles could have been used, as indeed they were in Scribner et al. (2008). Again, there is nothing inherently wrong with the decisions made about how to represent alcohol outlet density in the statistical model of the data, but these decisions are likely to create a set of results that look more those in Figure 2 than those in Figure 1.

CONCLUSION

The above discussion and specific example which we examined in detail suggest that approaching model validation in terms of a comparison of model output with empirical data is an exercise fraught with difficulties. The example illustrates the issues raised by Oreskes (2003) concerning the comparison of model output to empirical data as a means of validating the former and justifying its use to make predictions about the future state of the real world system to which it pertains. If such a comparison of model output to empirical data is of limited usefulness, how then might one go about assessing the value of a simulation model? Oreskes (1998) argues that we should move away from the use of the term validation entirely and instead focus on model evaluation. The former term, she contends, implies only an affirmative result, with the model nearly always resembling the data. Evaluation, on the other hand, implies an assessment in which the criteria for model success are clearly articulated and in which a negative appraisal is as likely as a positive one. These criteria for success would involve evaluating the model in ways other than the correspondence of its output to empirical data.

These alternative ways of evaluating a model include comparing it to other models, sensitivity analysis, and extreme condition tests, and in deciding upon which of these to employ it is important to consider relations between the amount of data available and level of understanding of the system influences in the particular problem one is addressing (Grant & Swannack, 2008; Rykiel, 1996). Where the level of understanding and amount of data available are low then conceptual evaluation is most relevant, for example whether the model can reproduce the relationships between model components and their dynamic behavior (Rykiel, 1996). Quantitative evaluation is most appropriate where both the level of understanding and availability of data are high. The tendency in much public health research has been to move to quantitative evaluation of models irrespective of the level of understanding of the system influences that affect the problem of interest and the amount of data available.

Ultimately, the value of any model resides in its ability to further our understanding of the real-world system we are studying. While one should not expect a model to be able to predict the future behavioral of a real-world system with absolute certainty, one should expect simulations results to provide new knowledge to help reduce (in some useful way) uncertainty with which to view the future of the real-world system of interest (Grant & Swannack, 2008). The relative amount of knowledge gained depends largely upon the current state of the knowledge about the system of interest. This roots in the assumptions that use of the systems approach to solve the problem implies one is dealing with a system for which there are relatively few data and likely little understanding, and the less one understands about a system, the more there is to learn about it. So, for example, the model developed by Scribner et al. (2009) was used to estimate the potential effects of lowering the legal drinking age on alcohol consumption on colleges and university campuses (Rasul et al., 2011). The results of the simulation show that lowering the legal drinking age would only be effective in the unlikely event of a combination of very high alcohol availability and very low enforcement of polices. This demonstrates the useful of the model in understanding the system influences that drive college drinking and its ability to help us understand the potential effects of various policy options. These seem better criteria by which to evaluate the model than whether it can generate output that look like empirical data pertaining to college drinking and the availability of alcohol on college campuses.

REFERENCES

Ackleh, A.S., Fitzpatrick, B.G., Scribner, R., Simonsen, N., and Thibodeaux, J.J. (2009). Ecosystem modeling of college drinking: Parameter estimation and comparing models to data. *Mathematical and Computer Modelling*, 50:481-497.

Babbie, E. (1995). *The Practice of Social Research* (7th edition), Wadsworth, Belmont, CA.

Diex Roux, A.V. (2011). Complex systems thinking and current impasses in health disparities research. *American Journal of Public Health*, 101:1627-1634.

DeJong, W., Schneider, S.K., Towvim, L.G., Murphy, M.J., Doerr, E.E., Simonsen, N.R., Mason, K., and Scribner, R.A. (2006). A multisite randomized trial of social norms marketing campaigns to reduce college student drinking. *Journal of Studies on Alcohol*, 67:868-879.

Eck, J.E., and Lui, L. (2008). Contrasting simulated and empirical experiments in crime prevention. *Journal of Experimental Criminology*, 4:195-213.

Elkins, A.D., and Gorman, D.M. (2014). Systems theory in public health, in *Oxford Bibliographies in Public Health* (D. McQueen, ed.), Oxford University Press, New York.

Fitzpatrick, B.G., & Martinez, J. (2012). Agent-based modeling of ecological niche theory and assertive drinking. *Journal of Artificial Societies and Social Simulation*,15(2):4.

Fitzpatrick, B.G., Scribner, R., Ackleh, A.S., Rasul, J., Jacquez, G., Simonsen, N., and Rommel, R. (2012). Forecasting the effect of the Amethyst Initiative on college drinking. *Alcoholism: Clinical and Experimental Research*, 36(9):1608-1613.

Galea, S., Riddle, M., and Kaplan, G.A. (2010). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology* 39:97-106.

Hammond, R.A. (2009). Complex systems modeling for obesity research. *Preventing Chronic Disease*, 6(3). http://www.cdc.gov/pcd/issues/2009/jul/09_0017.htm.

Hawe, P., Shiell, A., and Riley, T. (2009). Theorising interventions as events in systems. *American Journal of Community Psychology*, 43:267-276.

Homer, J.B. (1996). Why we iterate: scientific modeling in theory and practice. *System Dynamics Review*, 12:1-19.

Homer, J., Hirsch, G., and Milstein, B. (2007). Chronic illness in a complex health economy: the perils and promises of downstream and upstream reforms. *System Dynamics Review*, 23:313-343.

Jones, A.P., Homer, J.B., Murphy, D.L., Essien, J.D.K., Milstein, B., and Seville, D.A. (2006). Understanding diabetes population dynamics through simulation modeling and experimentation. *American Journal of Public Health* 96:488-494.

Kaplan, G.A. (2013). Einstein vs Occam – aka: why we (might) need complex systems models to improve population health and eliminate health inequalities. Paper presented at the 13th Annual Scientific Meeting of the American Academy of Health Behavior, March 17-20, Santa Fe, NM.

Kleindorfer, G.B., O'Neill, L., and Ganeshan, R. (1998). Validation in simulation: Various positions in the Philosophy of Science. *Management Science, 44*(8):1087-1099.

Korzybski, A. (1933). *Science and Sanity: An Introduction to Non-Aristotelian System and General Semantics*, Institute of General Semantics, Englewood, NJ.

Luke, D.A., and Stamatakis, K.A. (2012). Systems science methods in public health: Dynamics, networks and agents. *Annual Review of Public Health*, 33:357-376.

Mubayi, A., Greenwood, P., Wang, X., Castillo-Chavez, C., Gorman, D.M., Salt, R.F, and Gruenewald, P.G. (2011). Types of drinkers and drinking settings: An application of a mathematical model. *Addiction*, 106:749-758.

Oreskes, N. (1998). Evaluation (not validation) of quantitative models. *Environmental Health Perspectives*, 106 (Suppl. 6):1453-1460.

Oreskes, N. (2003). The role of quantitative models in science, in *Models in Ecosystem Science*, (C. Draper, W. Canham, J. Cole, and W.K. Lauenroth, eds.), Princeton University Press, Princeton, NJ.

Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263:641-646.

Rasul, J.W., Rommel, R.G., Jacquez, G.M., Fitzpatrick, B.G., Ackleh, A.S., Simonsen, N., and Scribner, R.A. (2011). Heavy episodic drinking on college campuses: Does changing the legal drinking age make a difference. *Journal of Studies on Alcohol and Drugs*, 72:15-23.

Rykiel, Jr., E.J. (1996). Testing ecological models: the meaning of validation. *Ecological Modelling*, 90:229-244.

Scribner, R. Mason, K. Theall, K., Simonson, N., Schneider, S.K., Towvim, L.G., and DeJong, W. (2008). The contextual role of alcohol outlet density in college drinking. *Journal of Studies on Alcohol and Drugs,* 69(1):112-120.

Scribner, R., Ackleh, A.S., Fitzpatrick, B.G., Jacquez, G., Thibodeaux, J.J, Rommel, R., and Simonsen, N. (2009). A systems approach to college drinking: Development of a deterministic model for testing alcohol control policies. *Journal of Studies on Alcohol and Drugs,* 70(5):805-821.

Susser, M. (1995). Editorial: The tribulation of trials – intervention in communities. *American Journal of Public Health*, 85:156-158.

Trickett, E.J., Beehler, S., Deutsch, C., Green, L., Hawe, P., McLeroy, K., Miller, R., Rapkin, B., Schensul, J., Schulz, A., & Trimble, J. (2011). Advancing the science of community-level interventions. *American Journal of Public Health*, 101:1410-1419.