Best Practices for Systems Science Research

Authors: Matt Kasman and L. Kurt Kreuger

Editors: Elizabeth Bruch and Ross A. Hammond

Advisory board: Georgiy Bobashev, Kristen Hassmiller Lich, Peter Hovmand, Patty Mabry, Bobby Milstein, Mary E. Northridge, and Scott Weidman

Acknowledgement: the creation of this document was funded by NIH Grant R25HD082971

Introduction

Systems science research provides avenues to address questions that frequently bedevil traditional analytic techniques due to the presence of substantial interdependence over time. Given that public health is often concerned with problems involving organizations, populations, and environments that dynamically interact and affect one another, it is unsurprising that there is substantial interest in systems science within this field (Kasman et al., 2021). The application of systems science to topics related to public health policy is still a relatively recent development that is both rapidly expanding as more people are exposed to it and changing as new tools are introduced. As this research approach matures in this multidisciplinary space, there is an increasing need for a set of guidelines that can help guide effective and appropriate usage for researchers as well as those who fund and consume systems science research.

This document emerged out of a day-long workshop hosted at the National Institute of Health that took place on April 18th, 2016. The workshop gathered a number of experts in this area to discuss and come to some agreement regarding best practices in systems science research. During this discussion, topics for consensus statements of best practices were agreed upon, and an outline of these statements was drafted by the group. Participants included: William Riley (NIH), Ross Hammond (Brookings Institution), Michael Spittel (NIH), Patty Mabry (Indiana University), Elizabeth Bruch (University of Michigan), Hazhir Rahmandad (MIT), Nathaniel Osgood (University of Saskatchewan), Jonathan Caulkins (Carnegie Mellon University), Maria Mayorga (North Carolina State University), Georgy Bobashev (RTI), Stephen Eubank (VBI), Volker Grimm (Karlsruhe Institute), Deborah Marshall (University of Calgary), Ellis McKenzie (Fogarty International Center), Scott Weidman (NAS), David Mendez (University of Michigan), Lauren Ancel-Meyers (UT Austin), Ana Diez-Roux (Drexel), Mary Northridge (Columbia University), Kristen Hassmiller-Lich (UNC), Peter Hovmand (Washington University), John Sterman (MIT), Bobby Milstein (ReThink Health and MIT), and Matt Kasman (Brookings Institution). This outline was then expanded upon and refined by the lead authors, with guidance from the editors, as well as feedback and agreement from the advisory board.

The remainder of the paper is organized around the topics or statements that the workshop members agreed constituted a set of best practices for the field. This set of best practices is not meant to be exhaustive, but rather establishes a baseline set of evaluation criteria to assure that Systems Science produces insights that are useful, feasible, credible, and ethical. We describe best practices across eight categories. We describe each best practice category in detail, and then present a summary checklist of best practices that can be used when planning or evaluating particular studies in context.

References

Kasman, M., Breen, N. and Hammond, R.A. (2021). Complex Systems Science. In *The Science of Health Disparities Research* (eds I. Dankwa-Mullan, E.J. Pérez-Stable, K.L. Gardner, X. Zhang and A.M. Rosario).

1. Research Purpose

A research project should have clearly stated goals and scope. There should be a well-defined set of questions that the project will attempt to answer and a plan for assessing its success in doing so.

Systems science models can be used for a wide range of different purposes (Epstein, 2008; Hammond, 2015; Page, 1999). The specific goals of a project have important implications for model design choices, model construction, and engagement with data. Assessment of a model by readers or reviewers is best done in light of the intended purpose of the model, and is aided by a clear statement of this purpose.

We select three dimensions along which projects may differ in purpose: project aim, reference point, and empirical engagement. For each of these dimensions, we provide a clarifying question that can be used to situate a project along that dimension and a brief description.

1. *Project aim*: is a project intended to study the etiology of a system; explore plausible mechanisms underlying policy effects, interventions, or other perturbations; assess tools; forecast a range of outcomes; set appropriate goals and evaluate existing goals, or create a pedagogical device?

   A project can focus on explicating the etiology of a system (i.e. how components of a system dynamically interact with one another), or on providing insight into causal effects (typically for a system whose operation has been the subject of prior research or posited by theory). Alternatively, a project can explore the quality of available tools, including: study designs, statistical inference or classification methods, datasets, and measurement instruments. Finally, projects can aim to create pedagogical tools that illustrate specific mechanisms or processes.

2. *Reference point*: is the project retrospective or prospective in nature?

   Systems science research can attempt to explain patterns and outcomes that have already occurred (and can thus be compared to model behavior) or can provide insight into potential patterns and outcomes that might occur in a particular context. It is also possible for models to be functionally ahistorical, with no clear connection to existing or potential trends.

3. *Empirical engagement*: to what extent is the project intended to represent a specific setting or settings, and to what extent is the model calibrated to empirical data?

   Some models (typically those that are theoretical or pedagogical in nature) are largely or entirely “data free,” relying on stylized representations of model components and dynamics. Others are more empirical in nature, using data from observations or experiments to characterize model components and dynamics.

Table 1 provides illustrative examples of completed research, classified along each of the three dimensions of difference that we identify. These can be used to help guide the categorization of research projects or as exemplars for research projects in specific categories or combinations of categories.

Table 1: Research examples by dimensions of difference

Example	**Summary**	**Project Aim**	**Reference Point**	**Empirical Engagement**
Schelling (1971)	Highly stylized model of residential selection showing how preferences for same-race neighbors affects segregation	Etiological	Ahistorical	Low
Explanation:	The purpose of the model is to explore the influence of a dynamic mechanism (preference-driven mobility) on residential segregation. It doesn't engage with retrospective movement patters or prospectively predict changes. Model components and dynamics aren't intended to represent a specific setting, but end-state residential segregation levels can be generally compared with those in the real world to assess the model's explanatory power.			
Bruch and Mare (2006)	Extends the classic Schelling model of residential segregation to explore the role of neighborhood preferences in forming observed levels of racial segregation	Etiological	Retrospective	Strong
Explanation:	Like the Schelling model, the purpose is to explore the influence of preference-based mobility on residential segregation. However, model output is intended to be retrospectively compared to real-world patterns, and empirical data are used to characterize model components and dynamics.			
Reardon et al. (2014)	Understand causal pathways driving socioeconomic sorting in U.S. college enrollment	Etiological	Retrospective	Strong
Explanation:	The aim is to understand how a set of effect mechanisms affect patterns of socioeconomic sorting in college enrollment. The model attempts to retrospectively explain observed enrollment patterns, and model components and dynamics are based empirical data.			
Reardon et al. (2018)	Predict effect of "race-neutral" admissions policies relative to race-based affirmative action	Causal	Prospective	Strong

Explanation:	Building upon Reardon et al. (2014), this model explores how specific policies might impact college enrollment patterns. Model components and dynamics are based on an expanded set of empirical data sources.			
Luke et al. (2017)	Provide insight into the potential effects of different combinations of tobacco retail policies across communities	Causal	Prospective	Strong
Explanation:	This model shows how different policies and policy combinations might affect tobacco purchase patterns across community contexts. Model components and dynamics are based on empirical data sources.			
Teyhouee et al. (2017)	Study efficacy of mobile phone sentinel system designed to detect food-borne illness outbreaks	Tool Assessment	Prospective	Moderate
Explanation:	This model explores how different detection tools perform at the task of outbreak detection. It is predictive in nature, and is based on a combination of observed and synthetic data sources.			
Wilensky and Rand (2015)	"Heroes and Cowards" game showing emergence of complex behavior from simple rules	Pedagogical	Ahistorical	None
Explanation:	This model is a useful demonstration of how varying simple, individual-level rules and starting conditions can influence system-level outcome patterns. It is pedagogical in nature, not intended to represent existing or potential processes, and does not represent any real-world setting.			
Homer et.al. (2016)	Assess the likely effects and affordability of combined regional investments to transform health system performance	Causal	Prospective	Strong
Explanation:	This empirically-grounded model estimates the extent to which summary measures of population health, healthcare cost, social inequity, and workforce productivity could change over time if leaders were able to (a) deliver higher value care; (b) reinvest savings and expand global payment; (c) enable healthier behaviors; and (d) expand socio-economic opportunities. To assess affordability, the model also calculates total program spending, as well as net financial benefit.			
Homer et.al. (2007)	Show how a tug-of-war between health care payers and providers creates dysfunction, and explore how to defuse it.	Etiological	Retrospective and Prospective	Moderate
Explanation:	This generic model shows how progress in managing chronic illness in the U.S. has slowed since 1980, largely due to competition between payers and providers, resulting in price inflation and an unstable climate for health investments. It also demonstrates the likely impact of efforts to manage biological risks that directly affect illness onset, as well as broader efforts to enable healthier behaviors and to enhance living conditions.			

Levy, et al.[a],	Simulation model that forecasts tobacco prevalence by elucidating policy impact at each point in the cessation mechanism: quit attempts, choice of quit method, and effectiveness of quit method.	Mechanistic/Forecasting	Prospective	Strong
Explanation:	This model has real world implications for policy decision making. By showing the impact of tobacco control policies on each of three determinants of the cessation aspect of tobacco prevalence, the model relies on empirical data to estimate the impacts of a number of evidence-based policy combinations on tobacco prevalence. This model illustrates how simulators can be constructed to help decision makers forecast the intended and unintended consequences of their decisions.			
Levy et al.[b], 2010	Tobacco policy simulation to understand policy requirements needed to achieve Healthy People 2010 goals.	Goal Setting/Evaluation	Prospective	Strong
Explanation:	This model starts with a specified desired outcome. The model is used to discover what it would take, in terms of resources and policy effects, to achieve the desired goal of reducing tobacco prevalence to 12% by 2010 (Healthy People 2010). The model shows what combination of existing policies with demonstrated effects would be needed to achieve the goal. Even with unlimited resources and utilizing all of the most effective evidence strategies the goal was shown to unobtainable in the proposed timeframe. Modeling can be used to both evaluate existing goals for resource requirements and to set appropriate goals that coincide with available resources and evidence.			
Milstein et.al. (2007)	Explore plausible trajectories for diabetes prevalence to assess the achievability and compatibility of future national objectives	Pedagogical	Prospective	Strong
Explanation:	This empirically-grounded model demonstrates that the 2010 national objective for reducing diabetes prevalence is unattainable, given historical dynamics of incidence, diagnosis, and mortality. It explains why mental models based largely on the dynamics of infectious disease are flawed when used to set goals for preventing chronic diseases.			

References

Bruch, E. E., & Mare, R. D. (2006). Neighborhood choice and neighborhood change. American Journal of sociology, 112(3), 667-709.

Epstein, J. M. (2008). Why model?. Journal of Artificial Societies and Social Simulation, 11(4), 12.

Hammond RA (2015). Considerations and best practices in agent-based modeling to inform policy. Washington, D. C.

Hammond, R. A., & Ornstein, J. T. (2014). A model of social influence on body mass index. Annals of the New York Academy of Sciences, 1331(1), 34-42.

Homer J, Hirsch G, Milstein B. Chronic Illness in a Complex Health Economy: The Perils and Promises of Downstream and Upstream Reforms. System Dynamics Review, 23(2/3):313-343, 2007.

Homer J, Milstein B, Hirsch G, Fisher E. Combined Regional Investments Could Substantially Enhance Health System Performance And Be Financially Affordable. Health Affairs. 2016;35(8):1435-1443.

Levy, D.T., Mabry, P.L., Graham, A.L., Orleans, C.T. & Abrams, D.B. (2010) Exploring Scenarios to Dramatically Reduce Smoking Prevalence: A Simulation Model of the Three-Part Cessation Process. American Journal of Public Health, 100(7), 1253-1259. http://ajph.aphapublications.org/cgi/content/abstract/100/7/1253

Levy, D.T., Mabry, P.L., Graham, A.L., Orleans, C.T. & Abrams, D.B. (2010). Reaching Healthy People 2010 by 2013: A SimSmoke simulation. American Journal of Preventive Medicine, 38(3Suppl), S373-381. http://www.sciencedirect.com/science/article/pii/S0749379709008599

Luke, D. A., Hammond, R. A., Combs, T., Sorg, A., Kasman, M., Mack-Crane, A., Ribisl, K.M. & Henriksen, L. (2017). Tobacco town: computational modeling of policy options to reduce tobacco retailer density. American journal of public health, 107(5), 740-746.

Milstein B, Jones A, Homer J, Murphy D, Essien J, Seville D. Charting plausible futures for diabetes prevalence: a role for system dynamics simulation modeling. Prev Chronic Dis. 2007;4(3):1-8.

Page, S. E. (1999). Computational models from A to Z. Complexity, 5(1), 35-41.

Reardon, S. F., Kasman, M., Klasik, D., & Baker, R. (2014). Agent-based simulation models of the college sorting process.

Reardon, S. F., Baker, R. , Kasman, M. , Klasik, D. and Townsend, J. B. (2018), What Levels of Racial Diversity Can Be Achieved with Socioeconomic-Based Affirmative Action? Evidence from a Simulation Model. J. Pol. Anal. Manage., 37: 630-657. doi:10.1002/pam.22056

Schelling, T. C. (1971). Dynamic models of segregation. Journal of mathematical sociology, 1(2), 143-186.

Teyhouee, A., McPhee-Knowles, S., Waldner, C., & Osgood, N. (2017). Prospective Detection of Foodborne Illness Outbreaks Using Machine Learning Approaches. In D. Lee, Y.-R. Lin, N.

Osgood, & R. Thomson (Eds.), Social, Cultural, and Behavioral Modeling (pp. 302–308). Cham: Springer International Publishing.

Wilensky, U & Rand, W. (2015). Introduction to Agent-Based Modeling: Modeling Natural, Social and Engineered Complex Systems with NetLogo. Cambridge, MA. MIT Press.

2. Research Approach

A specific systems science approach has been identified, and that approach is appropriate given features of the modeling problem and research goals.

In this section, we provide information that can help researchers decide when systems science approaches are well suited for their research questions, and how to identify specific approaches that can answer those questions. We first discuss when using a systems science approach is appropriate; second, introduce a series of dimensions along which research questions differ that can help guide the selection of a specific approach; and finally, present examples of three different systems science approaches in the public health realm.

When to use a systems science research approach:

With increases in computational power and a growing recognition of the limitations of traditional research approaches, systems science approaches have gained favor among investigators in a variety of fields. When used appropriately and effectively, these research methods confer many advantages to investigators (Epstein, 2008; Hammond, 2009; Hammond et al., 2017; Sterman, 2000; Mabry et al., 2010). In particular, they can:

1. capture the dynamic and interdependent relationships between components within a system, which may occur at or between multiple levels of scale, and how these relationships may adaptively change over time;
2. help explain how system-level behavior can arise from the interactions between individual system components in nonlinear ways;
3. act as “policy laboratories” or “flight simulators,” allowing researchers to test interventions that cannot be feasibly explored in the real world due to time, cost, or ethical constraints;
4. guide the collection of future data;
5. build theory that informs future research and policy;
6. identify research gaps, and prioritize their importance for pursuing a given outcome;
7. be used to evaluate an existing policy or policy combination, in order to determine whether a specified goal will be met within a given timeline;
8. be used to inform goal setting, helping to identify realistic goals given resource constraints and timeframes; and
9. systematically synthesize available data, especially from different fields of study, and at different levels of analysis of the system.

However, systems science approaches are not suitable for all research questions. Real-world experimentation and traditional empirical analysis may be more appropriate when the focus of investigation is *whether* an association exists between variables, rather than *how* these associations can be characterized in terms of simultaneous causal pathways and feedback loops. This is particularly true if investigators desire high internal validity and characterization of the relationships between variables in precise, quantitative terms (Miller & Page, 2007; Homer, 2014). Whereas the passage of time is an important dimension in system science and allows for the observation of behavioral adaptation, researchers may favor other approaches if the focus is on associations between variables at static points in time. In addition, traditional approaches are better suited when the relationships between variables are direct and linear, such that changes in inputs correspond to proportional changes in outputs (Luke & Stamatakis, 2012).

Dimensions of difference between research questions:

If systems science approaches are applicable to the research question, the next step is choosing the best specific approach. Since no single approach is best suited to all situations, we have identified eight important dimensions of difference between research questions to help guide selection. While this list is not exhaustive, evaluating the research question through the lens of these eight dimensions will help ensure the systems science approach aligns with the research goal (see Table 1).

1. **The perspective from which the research question is analyzing the system: top-down, bottom-up, or process-oriented.** A research question with a *top-down perspective* is interested in how system behavior is a result of the way system components are structured and connected—via causal pathways and feedback loops—rather than individual behaviors and outcomes; although all systems science research incorporates causal mechanisms and feedback, top-down models inherently explicate these, which can have important implications for how models are developed with and communicated to stakeholders (Marshall et al., 2015; Minyard et al., 2018). For example, a research question focused on drug addiction might observe how changes in recidivism rates result from the way that system components—employers, law enforcement, or housing availability—interact. A research question with a *bottom-up perspective* is concerned with how individual entities within the system behave and interact to shape system-wide outcomes. Researchers exploring migration patterns might look at how the properties of individuals in the system influence the way they move and interact with one another, thereby shaping population-level patterns in migration. A system composed of resource-constrained processes would benefit from a *process-oriented perspective*. Here, rather than system-level feedbacks or individual agent interactions, complexity emerges from interactions between different processes and the sharing of limited processing resources (Zhang, 2018). This perspective is commonly used in health care and disease screening where the agency of patients and system-level feedbacks play a less dominant role. For example, an investigator may examine the impact that different employee scheduling protocols may have on emergency department wait times or cost-effectiveness.

2. **The nature of active entities.** Researchers must consider the nature of the active entities that are most important for the system under study. The active entities are those whose behavior, statuses, and outcomes are central to the dynamics of the system. It could be

individual people, companies or other organizations, aggregated sub-populations, or services provided within a community. For example, a hospital could be considered as a collection of individual patients and doctors, making the active entities the patients and doctors. If the focus is on interactions between multiple hospitals, or between hospitals and law enforcement agencies, the patients and doctors might usefully be aggregated, making the hospital the active entity. Active entities are not limited to tangible objects. They can be something more abstract, such as level of satisfaction, amount of stress, or fear of legal sanctions, which could be described within an individual person or at the population level. For example, if the focus is on the performance of the hospital in offering health services, the active entity might be the processes within the hospital (and the associated process resources).

3. **The importance of heterogeneity among system entities.** If a researcher is interested in how the dynamics of a system change across different populations or how a policy intervention might have different effects in a variety of discrete locations, heterogeneity is a significant consideration. For example, if researchers want to create a model that examines the effects of a soda tax on soft drink consumption, then heterogeneity across factors like race, income, and urbanicity may be important, particularly if interventions are being considered across a spectrum of distinct geographic regions. Conversely, entities may be characterized by average patterns of behavior, without distinguishing between particular types of entities. This may be sufficient for showing how general trends emerge from the dynamics of a system, particularly in cases in which entities are homogeneous across key dimensions, when more granular data is unavailable, or when broad interventions will be enacted that cannot be tailored by heterogeneous characteristics. When heterogeneity is low, often models can aggregate similar entities (e.g. individuals).

4. **The importance of geographic and social space.** Objects, ideas, and other phenomena—both tangible and intangible—may spread between entities (e.g. individuals or organizations) through their interactions with each other in a way shaped by social structure. Entities themselves may also move through geographic space. For example, investigators interested in the spread of epidemics may require an approach that can simulate geographic distribution and transmission of disease across a city, country, or even the world. A research question aimed at analyzing the spread of body image norms through a community might require an approach that can simulate social network ties between individuals and account for how they change over time.

5. **The handling of updates.** Investigators must consider whether updates to the system occur gradually (continuous), such as the gradual change in population size over time, or at specific times (events), such as a discrete change from a susceptible to infected status. Change events might update after the passage of a fixed amount of time, such as every simulated day, due to the action of other entities, such as a patient exiting a hospital once a

doctor completes the treatment, or based on a hazard rate, such as a person having a certain daily probability of making a purchase decision.

6. **The transparency of model structure.** There is often a need to communicate the structural assumptions of models to experts from diverse backgrounds, either internal to external to the modeling team. This is important for a number of reasons, including supporting expert participation, communicating learning to stakeholders, or persuading policy makers. Each modeling approach is associated with a general level of inherent transparency, influencing the amount of work needed to effectively communicate model structure (e.g. model components, relationships between them, feedback loops).

7. **The complexity of computation.** Systems models are simulations, which require computing hardware. There can be differences of orders of magnitude between different models, which can translate into requiring days or longer of simulation time on multi-node computing clusters instead of minutes or hours on a single laptop. This is magnified when a model has stochasticity, thereby requiring repeated realizations of identical model parameterizations.

Table 2: Comparisons of three complex systems approaches

Axis of difference	ABM	DES	SD
Perspective	Bottom-up	Process-oriented	Top-down
Focal entity	Individual	Process	Aggregate
Heterogeneity	High	High	Low to moderate
Geographic/social granularity	High	High	Low to moderate
Handling of updates	Event	Event	Continuous
Transparency	Moderate	High	High
Complexity of computation	Low to high	Low to high	Low to moderate

Examples of the application of different systems science approaches:

Table 2 shows how three commonly used systems science approaches can be roughly categorized in each of our defining dimensions. This is meant as a general guide, however, as each of these approaches is flexible enough to be able to transcend some or all of the classifications we give here. Agent-based modeling (ABM) simulates individual entities (e.g. individuals, organizations) interacting with one another and their environment over time.

Discrete Event Simulation (DES) describes resource-constrained processes acting on relatively passive entities (e.g. an emergency department with finite resources offering care to patients). System Dynamics (SD) models focus on explicitly representing relationships between system components and feedback mechanisms. Below, we discuss examples of how these approaches' attributes were leveraged to engage in research.

The Model of Infectious Disease Agents Study (MIDAS) is a network of scientists who have used ABM to study the spread of communicable diseases and explore measures to limit transmission. Among the many contributions of MIDAS, an ABM was developed that simulated the global transmission of an H1N1 outbreak among billions of distinct agents. These insights were used to, among other things, refine emergency response policy during pandemics (Epstein, 2009; Parker & Epstein, 2011). Smaller-scale ABMs, including those by Burke et al. (2006), provided evidence about how targeted contact tracing and vaccination can contain smallpox outbreaks (Burke et al., 2006).

ABM allowed MIDAS researchers to take a *bottom-up perspective*, observing how individual entities behave and adapt to the spread of disease (e.g., vaccinating or self-quarantining) to influence macroscopic outcomes (e.g., spread of disease over geographic regions) (Hammond, 2015). Researchers represented huge numbers of simultaneous entities (agents)—individuals—who are both the *active agents* and *heterogeneous* with regard to immunological and behavioral response. Additionally, and importantly in the study of infectious disease spread, ABM allowed researchers to capture how individuals move about *geographic space* over *discrete time* intervals—between their home, workplace, hospitals, and elsewhere—coming in physical contact with others and facilitating the spread of disease.

DES has been used widely in research on improving the efficiency and organization of hospitals and healthcare systems. For example, researchers at the Mayo Clinic used DES to analyze the effect of different policies designed to minimize the number of beds needed for cardiovascular surgery patients while maintaining timely patient service. Limiting the day-to-day variability of surgery schedules, adding weekday surgery, and moving long-stay patients out of the ICU were all identified by the modelers as potential strategies (Marmor et al., 2013). DES was chosen because the research question analyzed the system from a *process-oriented perspective*: the optimization of queues (e.g., patients moving from waitlist, to surgery, to recovery) and the use of limited resources—such as doctors and beds—that are allocated according to needs. Through a DES model, patients in the system can be characterized as *passive* entities who change state according to the flow of queues and cannot actively affect their queue position through behavior, and who are *heterogeneous* across certain variables (e.g. recovery time).

SD modeling has been used in a variety of health contexts (Homer and Hirsch, 2006), including research on physiological phenomena like energy homeostasis. In an SD model of weight gain, obesity, and fertility in women, Sabounchi et al. (2014) explore competing tradeoffs between the time required to lose weight and age-related fertility declines. Such insights can help guide preconception obesity intervention policy and planning. SD modeling was used to create a *top-*

down representation of fetal size and maternal weight as a set of stock variables that are influenced and acted upon through feedback mechanisms by other variables. This approach was successful because there were well-defined relationships between system components, and no need for *heterogeneity* within these system components.

Finally, researchers have the option of combining approaches (including those described above) in *hybrid* models that can take advantage of the strengths of multiple approaches (Swinerd and McNaught, 2012). Kreuger (2018) develops a model of tobacco smoking and purchasing behavior that consists of an ABM with multiple agents, each of which contains an SD module that depicts internal addiction dynamics. Hybrid modeling was chosen due to ABMs strengths in capturing social network and agent heterogeneity, while SD excels at describing the feedback-rich addiction theories used.

References:

Burke, D. S., Epstein, J. M., Cummings, D. A., Parker, J. I., Cline, K. C., Singa, R. M., & Chakravarty, S. (2006). Individual-based computational modeling of smallpox epidemic control strategies. *Academic Emergency Medicine*, *13*(11), 1142-1149.

Epstein, J. M. (2008). Why model?. *Journal of Artificial Societies and Social Simulation, 11*(4), 12.

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature*, *460*(7256), 687.

Hammond, R. A. (2009). Complex systems modeling for obesity research. *Preventing chronic disease, 6*(3).

Hammond RA (2015). Considerations and best practices in agent-based modeling to inform policy. Washington, D. C.

Hammond RA, Osgood N, Wolfson M. (2017). Using Complex Systems Simulation Modeling to Understand Health Inequality. In: Kaplan AD, Roux C, Simon S, Galea S, editors. Growing Inequality: Bridging Complex Systems, Population Health and Health Disparities. Washington, D. C: Westphalia Press.

Homer J. Levels of Evidence in System Dynamics Modeling. System Dynamics Review, 30:75-80, 2014.

Homer, J. B., & Hirsch, G. B. (2006). System dynamics modeling for public health: Background and opportunities. American Journal of Public Health, 96(3), 452–458. https://doi.org/10.2105/AJPH.2005.062059

Luke, D. A., & Stamatakis, K. A. (2012). Systems science methods in public health: dynamics, networks, and agents. *Annual review of public health*, *33*, 357-376.

Mabry, Patricia L., Stephen E. Marcus, Pamela I. Clark, Scott J. Leischow, and David Méndez. "Systems science: a revolution in public health policy research." (2010): 1161-1163.

Marmor, Y. N., Rohleder, T. R., Cook, D. J., Huschka, T. R., & Thompson, J. E. (2013). Recovery bed planning in cardiovascular surgery: a simulation case study. *Health care management science*, *16*(4), 314-327.

Marshall, D. A., Burgos-Liz, L., IJzerman, M. J., Crown, W., Padula, W. V., Wong, P. K., ... & Osgood, N. D. (2015). Selecting a dynamic simulation modeling method for health care delivery research—Part 2: Report of the ISPOR Dynamic Simulation Modeling Emerging Good Practices Task Force. Value in health, 18(2), 147-160.

Miller, J. H., & Page, S. E. (2009). Complex adaptive systems: An introduction to computational models of social life (Vol. 17). Princeton university press.

Minyard K, Smith T, Turner R, Milstein B, Solomon L. Community and Programmatic Factors Influencing Effective Use of System Dynamic Models System Dynamics Review. 2018.

Parker, J., & Epstein, J. M. (2011). A distributed platform for global-scale agent-based models of disease transmission. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, *22*(1), 2.

Sabounchi, N. S., Hovmand, P. S., Osgood, N. D., Dyck, R. F., & Jungheim, E. S. (2014). A novel system dynamics model of female obesity and fertility. *American journal of public health*, *104*(7), 1240-1246.

Sterman, J. D. (2000). Business dynamics: systems thinking and modeling for a complex world (No. HD30. 2 S7835 2000).

Swinerd, C., & McNaught, K. R. (2012). Design classes for hybrid simulations involving agent-based and system dynamics models. *Simulation Modelling Practice and Theory*, *25*, 118–133. https://doi.org/10.1016/j.simpat.2011.09.002

Kreuger, L. K. (2018). Data and Design: Advancing Theory for Complex Adaptive Systems. PhD dissertation. University of Saskatchewan, Saskatoon, Saskatchewan.

Zhang, X. (2018). Application of discrete event simulation in health care: a systematic review. *BMC Health Services Research*, 18:687. https://doi.org/10.1186/s12913-018-3456-4

3. Research team/preparation

A research team should benefit from expertise in multiple, specific domains.

This section is meant to judge the potential value or success of proposed work rather than completed work. It is meant to help set realistic expectations of what good modeling takes, rather than as a tool to judge work already completed. As such, this is meant simply to highlight relevant features that have proven effective in the past. Importantly, however, not all skills are essential for all modeling projects and not all may be given equal weight.

There are 4 main skill areas that should be considered when building a team to conduct systems modeling research:

1. **Content familiarity** includes knowledge of the theory, empirical evidence, and policy within a given domain. For example, when using a systems model to evaluate health interventions within a clinical environment, it is helpful to have expertise on the team from clinical (physicians or nurses) and administrative (clinic managers) perspectives to help inform the process. This is true not only in interpreting results, but in the construction of the model itself since, as will be discussed further in the sections on design, assumptions, and development, drawing on existing domain theory, knowledge of the available empirical evidence, and prevailing policy considerations will strengthen both model design and the value of modeling results.

2. **Software design experience** refers to software engineering best practices and knowledge of the programming language for the specific modeling package chosen (e.g. NetLogo language or Java). This includes expertise in file versioning and documentation, software testing methods, software design best practices, database management, and computational complexity scaling.

3. **Data analytical experience** refers to knowledge of various analytical and statistical methods important to a type of complex problem, as well as methods to integrate multiple data sources into the same system model. Examples of analytical methods could include social network analysis, GIS, and survey analysis methods such as Discrete Choice Experiments, as well as general skills in statistics and data science.

4. **Systems modeling experience** includes knowledge of modeling best practices, such as discussed in this document. A modeling expert would also have experience translating domain knowledge into model algorithms. While it may be possible for non-experts to build models that appear convincing, there is often considerable nuance and many implicit modeling assumptions required (Hoffer, 2013; Sterman, 2001).

While it’s important for the research team to encompass the four general skill sets above, the conduct of reliable systems science research requires rather nuanced understanding within these general categories. The sections that follow explain in greater detail the sort of experience that needs to be brought to bear in order to produce high-value results.

References:

Hoffer, L. D. (2013). Unreal models of real behavior: The agent-based modeling experience. *Practicing Anthropology*, *35*(1), 19–23.

Sterman, J. D. (2001). System Dynamics Modeling: Tools for learning in a complex world. California Management Review, 43(4).

4. Model assumptions

Assumptions are clearly explicated and are appropriate given research purpose and approach.

Every model incorporates simplifying assumptions. The assumptions made by statistical models are well-known and frequently discussed when considering their application and interpretation (e.g. independence of observations and distributions of underlying data). Systems science allows for highly flexible models that are not hindered by the same assumptions; however, they require a different set of assumptions. Moreover, the assumptions made by systems science models are neither uniform across models (even models that may appear superficially similar) nor obvious to the casual observer. Therefore, it is incumbent upon systems science researchers to carefully consider the assumptions that are inherent in the models that they design and deploy, and to explicate these assumptions.

We identify four ways in which assumptions are used in systems science modeling:

1) *Process timeframe.* Assumptions that determine the time horizon within which the phenomena that a model attempts to represent must and can be meaningfully captured.
2) *To determine what is included in -- and excluded from -- a model (i.e. setting model boundaries).* Everything that is excluded is assumed to not affect model behavior within a given context and timeframe. These assumptions also define the populations being modeled, and the relevant environments in which they operate.
3) *Defining endogenous model components*. These assumptions determine what model components are allowed to dynamically change during the course of a model run and how they do so.
4) *Defining exogenous model components.* These assumptions determine static components of models. These include initial distributions of variables and relationships between them, variables that do not change during the course of model runs, and mechanisms that remain fixed.

We identify six broad sources for making assumptions in systems science models. When possible, best practices are to utilize the strategies in priority order moving roughly from #1 to #5 listed below, with #6 describing a categorically distinct source of model assumptions. It is possible to inform assumptions with multiple, mutually reinforcing sources.

1) *The assumption is obtained directly from data*. This is the preferred method for deriving assumptions because it can be precisely tailored for specific model contexts and applications. For example, an ABM of college enrollment includes a specified correlation between student caliber (a single construct representing observable student achievement and potential) and family resources that is taken from an analysis of relevant variables available in the ELS dataset (Lauff & Ingels, 2015; Reardon et al., 2014; Reardon et al., 2018). The appropriateness of these assumptions largely rests on data

quality (including the compatibility of data sources if more than one are being employed), analytic rigor, and the match between analytic results and model constructs.

2) *The assumption is taken from available literature, in particular empirical findings reported in peer-reviewed literature*. This is the second most preferred method for deriving assumptions. This is similar to the first category except that the analyses were undertaken by other researchers. Therefore, some additional caution must be used here in cases where multiple literature sources are being combined with one another or with direct data analyses (e.g. making sure that they are comparable) or where details of how particular estimates were obtained are not completely transparent. In general, the stronger the level of evidence (scientific rigor) of the cited literature and the greater the similarity of the population and conditions described in the literature is to those represented in the model being developed, the greater the confidence and suitability the literature is for making assumptions.
3) *The assumption is derived from theory*. While less desirable than strategies #1 or #2, a modeler often encounters situations where appropriate data from these sources is not available. For example, many models make an explicit or implicit assumption that consumers make rational decisions (Luke et al., 2017); this is a central assumption in classical economic theory. The appropriateness of these assumptions rests on how well-validated a theory is for a particular application.
4) *The assumption is derived from expert opinion*. Model assumptions derived, wherever possible through rigorous qualitative analytical methodology, from the intuition and experience of content area experts.
5) *Assumptions are derived through sensitivity analysis.* Model assumptions arrived at during the course of model testing.
6) *The assumption is a focus of model inquiry*. Etiological or retrospective models are often deployed to test the explanatory power or implications of specific hypothesized mechanisms. For example, Schelling's model (1971) of the emergence of residential segregation tests how a hypothesized behavior (residents attempting to move only when a relatively high proportion of their neighbors were of a different race) would affect housing patterns. Because they are intended to be explicitly tested, the appropriateness of these assumptions is predicated solely on how well they are operationalized within a model.

When describing a model, researchers should identify the assumptions they make; we believe that going through the three categories of types of assumptions—model inclusion, endogenous components, and exogenous components—may help researchers to generate an exhaustive list. For each assumption that researchers identify, they should be able to provide a rationale for its inclusion from among the four categories that we provide, and should also be able to argue for its appropriateness given their research purpose and approach. By doing so, their audience should have a clear sense of how a model comports with (or departs from) existing literature as well as how to interpret model findings.

References:

Lauff, E., & Ingels, S. (2015). *Education Longitudinal Study of 2002 (ELS: 2002)*. (National Center for Education Statistics (NCES); No. NCES 2015034). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Luke, D. A., Hammond, R. A., Combs, T., Sorg, A., Kasman, M., Mack-Crane, A., Ribisl, K.M. & Henriksen, L. (2017). Tobacco town: computational modeling of policy options to reduce tobacco retailer density. American journal of public health, 107(5), 740-746.

Reardon, S. F., Kasman, M., Klasik, D., & Baker, R. (2014). Agent-based simulation models of the college sorting process.

Reardon, S. F., Baker, R. , Kasman, M. , Klasik, D. and Townsend, J. B. (2018), What Levels of Racial Diversity Can Be Achieved with Socioeconomic-Based Affirmative Action? Evidence from a Simulation Model. J. Pol. Anal. Manage., 37: 630-657. doi:10.1002/pam.22056

Schelling, T. C. (1971). Dynamic models of segregation. Journal of mathematical sociology, 1(2), 143-186.

5. Model design

An effective model design conceptualizes the currently understood or proposed mechanisms, ensuring model components (such as inputs, dynamic processes, and outputs) are appropriate for the research purpose.

The process of model design can be thought of as one of conceptualization. Drawing on the inter-disciplinary knowledge of the research team, the problem needs to be described or mapped in such a way as to highlight the perceived causal relationships relevant to the problem at hand. The design will draw heavily on model assumptions, since they define what should be included in the model. Furthermore, they also define whether something is endogenous or exogenous, the former requiring some internal dynamic causal structure, the latter requiring some knowledge of how certain variables will be parameterized or estimated. Not only is this a valuable deliverable in its own right, but it also provides a strong foundation for the operationalization of the model during model development, discussed in the next section.

However, because the systems being studied are often complex and require experts from multiple domains, designs themselves can easily become difficult to understand. In this case, it is useful to remember that the modeling process should be iterative. The goal of the design is to represent the collective knowledge of the team at a given point of time. As the model is developed and validated (as described in the next section), learning will require that the model design be updated. For example, it can be often difficult to see a logical or empirical inconsistency in the design until it starts to become operationalized in a running model. Therefore, the team need not focus on ensuring that the model design is complete before moving on to the process of operationalizing it.

On the other hand, the model design must have sufficient complexity to guide the model development process. As discussed in the section on research approach, systems modeling is useful when a system possesses certain types of complexity. Therefore, the design of the model must include some features suitable for a systems lens. For example, if the research approach seeks to understand the interdependent relationships between actors in a system, the design of an ABM should include a set of assumptions and causal descriptions of those interactions.

One way to reinforce this balance between simplicity and sufficiency is to use a design protocol that outlines what types of features or model components should be described. There are multiple such protocols useful for some, or all, of the main modeling approaches. Causal loop diagrams are often used, especially though not exclusively with SD models, to map the underlying causal structure in a principled way (Sterman, 2000; Kim, 1992; Homer, 2019). The PARTE framework (Hammond, Osgood and Wolfson, 2017) is a useful approach for ABMs specifically. It is named after its constituent themes: properties, actions, and rules (which define the agents), and time and environment (which define the agent context). The ODD framework (Grimm et al., 2006; Grimm et al., 2010), also useful for ABM conceptualization. It encompasses a wide array of factors to specify in the design, such as model stochasticity, whether agents interact or learn, and where emergent behavior is expected. ODD+D (Müller et al., 2013) is an extension to specifically incorporate human decision-making in ABMs.

References:

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., … others. (2006). A Standard Protocol For Describing Individual-based And Agent-based Models. Ecological Modelling, 198(1), 115–126.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. Ecological Modelling, 221(23), 2760–2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., … Schwarz, N. (2013). Describing Human Decisions In Agent-based Models--ODD+ D, An Extension Of The ODD Protocol. Environmental Modelling & Software, 48, 37–48.

Kim, D. H. (1992). Guidelines for drawing causal loop diagrams. The Systems Thinker, 3(1), 5–6. Retrieved from https://thesystemsthinker.com/guidelines-for-drawing-causal-loop-diagrams-2/

Hammond, R. A., Osgood, N. D., & Wolfson, M. C. (2017). Using Complex Systems Simulation Modeling to Understand Health Inequality. In G. A. Kaplan, A. V. Diez Roux, C. P. Simon, & S. Galea (Eds.), Growing Inequality: Bridging Complex Systems, Health Disparities, and Population Health (pp. 11–24). Westphalia Press.

Homer J. Best practices in system dynamics modeling, revisited: a practitioner's view. System Dynamics Review 35:177-181, 2019.

Sterman, J. D. (2000). Business dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill.

6. Model development and testing

The model is accurately and adequately operationalized in software, subject to verification and validation, and following an incremental development process.

For confidence to be built in any model's output, researchers need to consider how to properly verify and validate their model (Sterman, 2000). *Verification* is the process of ensuring that the model design matches the model built. In other words, it is an attempt to answer the question "are we building the model right?" *Validation* is the process of ensuring that the model design matches the problem being modeled. In other words, addressing the question "are we building the right model?"

To accomplish these interrelated tasks, we suggest an *iterative* or *incremental development process* (Homer, 1996). This is an idea supported by the software engineering paradigm of *agile* development (Fowler & Highsmith, 2001) and entails regular review, from very early on and by some or all team members, of model features. There are different types of reviews that can be conducted, ranging from formal to informal, engaging the whole team or even specific pairs of members, and focusing on such things as model structure, model output, or model parameters (e.g. Freedman, 2000; Wiegers, 2002). But in all cases, the goal is to allow the team to maintain an updated and dynamic understanding of the behavior of the model.

This process supports confidence building in at least four ways.

1. **Finding model coding errors:** It helps the team discover programming errors early. Early detection of code errors is vital for maximizing project efficiency. Conducting regular reviews of a software project as complex as a systems model can often find software errors more efficiently than relying exclusively on testing practices.
2. **Avoiding the kitchen-sink model:** It helps the team control the level of complexity and detail in dynamic models. A practice to avoid is the adding of parameters or mechanisms that have no clear real-world meaning but are only intended to obtain the expected results from the model. This often leads to significant struggles in verification and validation. This not only subverts the natural learning process of model construction but also reduces the validity of model behavior. In this respect, dynamic models have more in common with software engineering projects than statistical models.
3. **Accelerating learning:** It allows the team to effectively understand the behavior of the model and assess the influence of key assumptions through various types of analysis, discussed more fully in the section below. Indeed, one of the primary sources of value of a modeling project is the clarity it brings to our individual and collective mental models, which allows the team to critique and study the assumptions they are making as individuals and learn from those made by other team members. This also facilitates regular comparison with available data, thereby improving model validity.
4. **Adapting to changing project requirements:** As a modeling project progresses, discoveries may bring about a shift in the project's assumptions, purpose, or approach.

For example, a study exploring leverage points in reducing patient wait times in an emergency department may discover that patient heterogeneity is more important than initially thought. This might require a change from an SD approach to an ABM or hybrid, the addition of behaviors for which empirical or theoretical knowledge must be gathered, or even a change in the purpose of the model. By following an iterative development process and regularly comparing mental models with model behavior, important shifts in project elements can be discovered early.

The unique characteristics of the modeling project will determine specifics such as how often model results are examined, how regularly the team meets, or how much of the team is present at each meeting. Even projects that involve simpler models should incorporate some iterative development, at the very least to investigate and understand potentially surprising model behavior. There are codified approaches to incremental development that researchers may find useful (e.g. Hovmand, 2014).

In addition to the model review process discussed above, model testing is critical in building confidence in model output, especially when model output generates surprising results. Thorough and timely model testing prevents the introduction of errors during iterative development and ensures the replicability of model results (as discussed in Section 8, below). There are many sources for testing best practices (e.g. Ammann, 2016; Craig, 2002; Kaner, 1999), and the modeling and software experts on the team should identify, explicate, and engage in those most appropriate for a given project.

References:

Ammann, P. and Offutt, J., 2016. Introduction to software testing. Cambridge University Press.

Craig, R.D. and Jaskiel, S.P., 2002. Systematic software testing. Artech house.

Fowler, M., & Highsmith, J. (2001). The agile manifesto. *Software Development*, *9*(August), 28–35. https://doi.org/10.1177/004057368303900411

Freedman, D.P. and Weinberg, G.M., 2000. Handbook of walkthroughs, inspections, and technical reviews: evaluating programs, projects, and products. Dorset House Publishing Co., Inc.

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

Hovmand, P. S. (2014). Group model building and community-based system dynamics process. In *Community Based System Dynamics* (pp. 17-30). Springer, New York, NY.

Kaner, C., Falk, J. and Nguyen, H.Q., 1999. Testing computer software. John Wiley & Sons.

Sterman JD. Truth and beauty: validation and model testing. In: *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston, MA: Irwin McGraw-Hill; 2000:845-890.

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, *1*(2), 143–186. https://doi.org/10.1080/0022250X.1971.9989794

Wiegers, K.E., 2002. Peer reviews in software: A practical guide (p. 256). Boston: Addison-Wesley.

7. Model Analysis

The behavior of the model is fully explicable through model components and causal dynamics and matches the research questions. Sensitivity analyses quantify how uncertainty in model inputs are associated with model outputs.

Model analysis requires some type of strategy to explore model output under various parametric and structural assumptions. The strategy will clearly depend on model purpose. For example, a model built to identify data gaps or to guide data collection might require a simpler analysis strategy than one seeking to predict multiple counter-factual scenarios. Selecting the appropriate experimental design or strategy is important. It might require as much time to properly conduct model analysis as spent in model development.

There are many types of experimental designs suitable for simulations studies. Sanchez and Wan (Sanchez & Wan, 2015) outline several. The simplest is some form of factorial design, where all allowed values of the different model parameters are considered. This allows identification of interesting parameter regimes and notable parameter interactions. For simple models, this might be sufficient, but more complicated models with many parameters that must be varied will require other designs, such as Latin hypercube sampling. Some systems models will also possess stochastic variation (due to the use of pseudo-random numbers) which must be accounted for by the chosen experimental method.

Whichever method, or combination of methods, is chosen, the goal is to efficiently explore model output while varying all pertinent assumptions and model parameters across wide ranges. This requires conscientiously defining ranges for parameters based on what is empirically plausible, feasible, or interesting, including extreme values. Reference should be made to empirical experiments, previous simulation studies, or expert opinion. Additionally, some parameters can be estimated using optimization algorithms, which repeatedly run the model while varying parameters to minimize some objective function, usually some difference measure between model output and empirical observations. These parameter estimates can be used to refine parameter ranges.

These sensitivity analyses are important for determining the generalizability and robustness of model outputs. Parameters or assumptions most impactful on key conclusions should be identified. The robustness of conclusions to the feasible ranges of assumptions should be discussed. Sources of uncertainty should be acknowledged. The goal should be to determine how sensitive the model's results are to numerical, behavioral, or structural uncertainties.

It is important to explain the main results in terms of the underlying structural mechanisms in the model. For example, an ABM that studies social influence and selection will likely include mechanisms for a social network between agents. A proper analysis will describe, for example, how and why network clustering controls influence and selection dynamics. Or, an SD model incorporating role modeling in the spread of behavioral changes could explain the conditions where role modeling's impact is increased or decreased, and why. Explanations and conclusions should be compared and contrasted with any established and relevant literature. And, as discussed more fully in Section 8, attention should be paid to the reproducibility of experimental procedures and analysis results.

This process can be very computationally expensive. Without adequately following an iterative model development process as outlined above, a model can become sufficiently complex to possess many more parameters with wider ranges than might be required otherwise. Furthermore, grounding results in knowledge of the model's structure becomes much more difficult.

References:

Sanchez, S. M., & Wan, H. (2015). Work smarter, not harder: A tutorial on designing and conducting simulation experiments. In Winter Simulation Conference (WSC), 2015 (pp. 1795–1809). https://doi.org/10.1109/WSC.2015.7408296

8. Reporting model design and results

Both the design of a model as well as results of model runs are available for scientific scrutiny. The model is described in a way that allows for its core components and key findings to be reproduced.

Dissemination is a crucial, if often overlooked, aspect of research. The goal of dissemination is twofold: the audience should be able to understand what was done (and why it was done that way) as well as what researchers learned. In approaching this goal, researchers should keep in mind heterogeneity within their audience with regards to perspective and background.

When describing completed systems science research, the gold standard is *comprehension* and *reproducibility*. An interested and capable consumer should ideally have enough information available that they can understand a model's design, including the rationale and methodology used in its construction; how it was implemented; the ways in which researchers tested and deployed the model; and patterns of behavior exhibited by the model. The consumer should then be able to *reproduce the methodology used and obtain similar results*.

The information that researchers should provide includes:

- A clear description and documentation of model components and dynamics. At a minimum, variables included in the model, whether these are fixed or updated during model runs, frequency of updates, and source of updates should be reported.
- A full list of software and libraries used (including, where applicable, versions)
- Transparent reporting of the sources of data, model building process, stakeholders engaged, sources of funding, and potential conflicts of interest
- Enumeration of which model inputs were calibrated and description of the process
- A coherent and complete account of how the model was used
- Data underlying model parameterization and use across all runs
- Model output data (either complete data or meaningfully aggregated). For models with substantial stochasticity, it is best to conduct a large number of simulation runs reporting the resulting outcome distribution along with some measure of uncertainty quantification.
- Any third party audit the model has been subjected to

For a variety of reasons (e.g. use of proprietary data or software), all of this may not always be feasible. However, a good faith attempt should be made to provide as accurate and complete a depiction as possible. Although there is no single "correct" way to describe a model and its use, there are a number of frameworks and guides that have been proposed and employed (e.g. ODD, TRACE, PARTE); researchers might consider turning to one or more of these in order to ensure the completeness and accessibility of their reporting (Donoho et al., 2009; Grimm et al., 2010; Grimm et al., 2014; Hammond, 2015; Jackson et al., 1991; Waltemath et al., 2011; Rahmandad et al., 2012).

Finally, researchers should recognize and respond to the needs and interests of multiple audiences. For example, if their model application and results may suggest areas where future data collection or systems science research might be especially fruitful, this should be communicated to others in the field. Similarly, if a model produces results that have important implications for policy and practice, researchers should both describe their model and relevant results in a user-friendly manner. This might include diagrams, tables, and text summarizing a model's operation and sources of input data; dynamic depictions of model runs; or accessible output data visualizations. In such situations, technical details of the model can be provide in supplementary material.

References:

Donoho, D. L., Maleki, A., Rahman, I. U., Shahram, M., & Stodden, V. (2009). Reproducible research in computational harmonic analysis. *Computing in Science & Engineering*, *11*(1), 8-18.

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S., ... & Thorbek, P. (2014). Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. *Ecological modelling*, *280*, 129-139.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: a review and first update. *Ecological modelling*, *221*(23), 2760-2768.

Hammond RA (2015). Considerations and best practices in agent-based modeling to inform policy. Washington, D. C.

Jackson, R. H., Boggs, P. T., Nash, S. G., & Powell, S. (1990). Guidelines for reporting results of computational experiments. Report of the ad hoc committee. *Mathematical programming*, *49*(1), 413-425.

Rahmandad H and Sterman J. Reporting guidelines for simulation-based research in the social sciences. System Dynamics Review 28(4):396-411, 2012.Waltemath, D., Adams, R., Beard, D. A., Bergmann, F. T., Bhalla, U. S., Britten, R., ... & Garny, A. (2011). Minimum information about a simulation experiment (MIASE). *PLoS computational biology*, *7*(4), e1001122.

Best Practices Checklist

A set of criteria that researchers, funders, and reviewers can use to engage in and evaluate proposed, ongoing, or completed projects

1. Research Purpose: project goals and research questions

- ☐ Project has clearly stated goals (i.e. research questions, uses and users addressed)
- ☐ Scope of project is well-defined
- ☐ Plan for assessing whether research goals have been met

2. Research Approach: system science methods used

- ☐ Systems science can be used to address the research questions
- ☐ A specific systems approach is identified
- ☐ The systems science approach is appropriate for the research goals given its handling of:
  - ☐ Perspective
  - ☐ Active entities
  - ☐ Heterogeneity
  - ☐ Geographic or social space
  - ☐ Updating
  - ☐ Design transparency
  - ☐ Computational complexity

3. Research Team and Preparation: personnel skills and roles

- ☐ The research team has personnel with expertise sufficient to meet research goals, including:
  - ☐ Content familiarity
  - ☐ Software skill
  - ☐ Data analytic skills
  - ☐ Complex systems modelling

4. Research Assumptions: support from theory and data

- ☐ It is clear how assumptions define:
  - ☐ What is included in (and excluded from) the model
  - ☐ Endogenous (dynamic) model components
  - ☐ Exogenous (static) model components
- ☐ Researchers identify the basis for each assumption (data, literature, theory, direct inquiry)

5. Model design: conceptualization of causal relationships

- ☐ Incorporate all model assumptions
- ☐ Design should be simple enough for the team to understand and complex enough to describe system behavior
- ☐ Select one or more design protocols (e.g. causal loop diagram, PARTE, ODD, ODD+D)

6. Model development and testing: translating design into a functional model

- ☐ The model is accurately and adequately operationalized in software
- ☐ The model is developed in an incremental fashion with regular model review meetings by the whole team, especially intended users
- ☐ The model has been subjected to verification and validation tests

7. Model Analysis: using the model and its output

- ☐ The behavior of the model is fully explicable through model components and dynamics
- ☐ Model behavior addresses research questions
- ☐ Sensitivity analyses are used to quantify how varying assumptions affects model output

8. Dissemination Plan: communicating the model and findings

- ☐ Researchers provide information including:
  - ☐ A clear description and documentation of model components and dynamics
  - ☐ A full list of software and libraries used (including, where applicable, versions)
  - ☐ Sources of data, model building process, stakeholders engaged, sources of funding, and potential conflicts of interest
  - ☐ How model inputs were calibrated
  - ☐ Model parameterization
  - ☐ Model use
  - ☐ Model output, including quantification of uncertainty across repeated runs and results from sensitivity analyses
  - ☐ Any third party audits the model has been subjected to
- ☐ The information provided is sufficient to scrutinize the model and its output and replicate findings
- ☐ Researchers communicate model design and findings to relevant audiences using clear, understandable language