Preprint Ver 2 - April 2020. This report has not been peer reviewed yet.

Evolution and Reproducibility of Simulation Modeling in Health Policy over Half a Century

Mohammad S. Jalali, PhD[1,2,*], Catherine DiGennaro[1], Abby Guitar[2], Karen Lew[2], and Hazhir Rahmandad, PhD[2]

[1] MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA, 02114

[2] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 02139

Corresponding author:

Mohammad S. Jalali, MSc, PhD
Assistant Professor, Harvard Medical School
Senior Scientist, Massachusetts General Hospital
Research Affiliate, MIT Sloan School of Management
msjalali@mgh.harvard.edu | (617) 724-3738

Word count:

3,280

Evolution of Simulation Modeling in Health Policy over Half a Century

Mohammad S. Jalali, PhD[1,2]*, Catherine DiGennaro[1], Abby Guitar[2], Karen Lew[2], and Hazhir Rahmandad, PhD[2]

[1] MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA, 02114
[2] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 02139

Abstract

Background: Simulation models are increasingly used to inform health policy. We provide an overview of applications of simulation models in health policy, analyze the use of best reporting practices, and assess the reproducibility of existing studies.

Method: Studies that used simulation modeling as the core method to address any health policy questions were included. Health policy domain distribution and changes in quality over time were well-characterized using MeSH terms and model characteristics, respectively. Reproducibility was assessed using predefined, categorical criteria.

Findings: 1,613 studies were analyzed. We found an exponential growth in the number of studies over the past half century, with the highest growth in dynamic modeling approaches. The largest subset of studies is focused on disease policy models (70%), within which pathological conditions, viral diseases, neoplasms, and cardiovascular diseases account for one-third of the articles. Nearly half of the studies do not report the details of their models. A subset of 100 articles (50 highly cited and 50 random) were selected to analyze in-depth criteria for reporting quality and reproducibility. Significant gaps between best modeling practices could be found in both the random and highly cited samples; only seven of 26 in-depth evaluation criteria were satisfied by more than 80% of samples. We found no evidence that the highly cited samples adhered better to the modeling best practices.

Interpretation: Our results suggest crucial areas for increased applications of simulation modeling, and opportunities to enhance the rigor and documentation in the conduct and reporting of simulation modeling in health policy.

Research in context

Evidence before this study

Simulation models are regularly used to assess the impacts of diverse policies and to minimize unintended consequences within a wide range of complex health issues. Simulation methods are impactful, but also rely on complex models that are not fully transparent to other researchers and decision-makers. Therefore, the rigor and transparency of modeling process, from design to implementation, validation, reporting, and dissemination, are essential. The status of recent efforts to increase the reproducibility and quality of simulation modeling remains unknown. We searched PubMed and Web of Science for research that used simulation modeling to address any health policy question, using the keywords "simulation", "policy", and their variants.

Added value of this study

Simulation models have become increasingly popular and advanced over the last half century in order to better encompass and represent the real-world complexity of health systems. Yet there is also significant variation in their application areas, with more than 70% of the studies focused on biomedical diseases. We also identified major discrepancies (more than half the sample) with best practices in reporting of essential information, such as model equations, required for reproduction. These discrepancies were not different between the top-cited studies and those with fewer citations.

Implications of all the available evidence

To realize its full potential, simulation modeling can be applied to a broader range of health policy topics, and requires enhanced application of documentation and modeling best practices.

Introduction

Increasingly complex health systems and advancing computational tools have promoted the application of simulation models to inform health policy. Simulation models help analyze and interpret the complexity of a health issue, enabling the design, evaluation, and improvement of policies while minimizing unintended consequences [1].

Simulation applications span a wide range of disciplines such as health care reform [2], health care delivery [3], cancer research [4], and infectious diseases [5], among others. These models are often complex, follow different methods and practices, and are integrated into major decisions with significant impact. Therefore, we need a more systematic view into which application areas are covered and how the methodological rigor and replicability of studies have evolved. However, these goals have not yet been systematically pursued [6]. To address this gap, three inter-related aims motivate the current study. First, we provide a broad comparative view of the state of simulation modeling research across various health domains and modeling approaches to inform current trends and identify potential gaps. Second, we provide a systematic assessment of the modeling practices in existing research, focusing on rigor and quality in reporting of design, implementation, validation, and dissemination, which influence the credibility and impact of simulation modeling research.

Our third objective is to assess the reproducibility of existing simulation modeling research. Reproducibility is the cornerstone of scientific research and the case for enhancing reproducibility has been made repeatedly and has given rise to various guidelines for authors to produce reproducible healthcare simulation models [7-13]. Reproducibility is key to building a cumulative science, revising problems in light of new data, and building confidence in the reliability of existing findings [14-16]. Documenting the state of the field informs the path forward and enables measurement of progress on this important issue.

We pursue these three objectives by conducting a systematic review of simulation modeling research in health policy. The current review does not summarize relevant substantive findings. Instead, our focus is on the trends in application areas, methods, modeling practices, and documentation and reproducibility.

Methods

Search strategy and selection criteria

PubMed was used as the primary search database where we searched for the terms simulation and policy (and any variation of those key terms) in the title and abstract. Our definition of "simulation" is not limited to any discipline and we included research from a wide range of fields. We complemented the results by searching 71 journals categorized in "Health Policy and Services" within Web of Science (WoS) for articles containing simulation in the title or abstract. We limited the search in both databases to English-language peer-reviewed articles. Our search included any articles indexed in these databases before March 2016. We reviewed the abstracts of the resulting sample to identify articles using simulation as the main method of research to answer a health policy question. The full text was inspected in cases where the abstract did not establish the inclusion criteria. Studies which only mentioned simulation modeling and/or policy but did not employ simulation modeling as the main method were excluded in addition to reviews and meta-analyses. Figure 1 summarizes the search and inclusion/exclusion process.

Data extraction

We extracted the title, abstract, publication year, journal, and author information directly from PubMed and WoS. To extract model type and more detailed reporting characteristics, we obtained and examined the full text of each article. To access the associated Medical Subject Headings (MeSH) terms from the articles selected for analysis, we developed a web scraping script in Python that was employed on PubMed in July 2018. We used a Python package (scholar.py) to scrape citation data from Google Scholar in October 2018.

Chronological, clustering, and trend analysis

We identified the categorizations of each MeSH term using the National Institute of Health's MeSH Browser and determined the distribution of the full sample of articles across second-level categorizations (e.g., within *Diseases [C]*, *Neoplasms [C04]*). In reporting the MeSH term frequencies below, we exclude the following categories because they were very general (and thus frequent) but not informative about policy areas of interest (we kept the articles containing them; each article may contain multiple second-level categorizations): *Eukaryota*, *Amino Acids, Peptides and Proteins*, and *Hormones, Hormone Substitutes and Hormone Antagonists*.

Using the same MeSH term frequency data, we present the distribution of all second-level MeSH terms across the articles. The four quadrants that are color coded to represent first-level

categorizations within the MeSH terms include: *Anatomy*, *Chemicals and Drugs*, *Diseases*, and *Organisms*.

Finally, adopting from Adams and Gurney [17], we aggregated collaboration and location data for authors of each of the papers to determine a multilateral collaboration score (MCS) and related this metric with the results of modeling rigor and reporting evaluation. See the Supplementary Text for more information.

Categorization based on modeling approach

The review of titles and abstracts across all articles began with an evaluation of four high-level properties of models and their reporting: (1) static or dynamic (time-dependency), (2) stochastic or deterministic, (3) event-driven or continuous, and (4) model documentation, i.e., whether model equations are included in the paper/appendix, referenced in another paper, or not available. The first three categorizations were used in lieu of reporting the simulation modeling approach, i.e., Markov decision modeling, microsimulation, compartmental modeling, system dynamics, agent-based, etc., because there is significant overlap among these categories and no uniform categorization exists. Categorical properties were selected because they are mutually exclusive and informative (see Table S1 for definitions of categorical criteria). We also did not attempt to replicate the actual simulations in the papers, only reporting whether the relevant equations were included. Two trained research assistants and the corresponding author conducted an initial independent evaluation of 100 articles each and compared their results with each other in order to establish consistent evaluation criteria. The research assistants then completed the coding.

Evaluation of modeling rigor and reporting

After the broad overview and categorization of the full sample of articles, we selected 100 articles on which to perform an in-depth evaluation. These articles were selected by first identifying 50 highly-cited articles, and then picking 50 additional articles at random from those that remained. In identifying the 50 highly-cited articles, we controlled for the recency of the articles using a year fixed effect in a regression to predict expected publication; we then classified highly cited articles as those with the largest fractional deviation from expected. The 50 highly-cited and 50 randomly selected articles are presented in Table S3 and Table S4, respectively.

Building on best modeling practice guidelines in the literature [18-20], we developed a set of 26 concrete criteria for evaluating the quality and rigor of simulation-based modeling articles. These criteria spanned reporting of the model's context, conceptualization, and formalization, along with analysis of the results and any external influences (see the criteria and their definitions in Table S2 and further detail on criteria assessment in the Supplementary Text). Coding criteria across two assistants were calibrated over an initial sample of 10 articles with the help of the first author. Uncalibrated agreement rate on this initial sample across two coders was 75%. Once consistency in criteria was established the two assistants completed the coding, showing a 90% agreement level. The remaining 10% of split coded items were discussed and resolved; reported findings are based on the resulting consensus.

Results

Study Identification and Selection (inclusion and exclusion criteria)

Figure 1 presents the summary of study selection. We identified 5,092 articles from an initial search in the PubMed and WoS databases, of which 300 were duplicates found in both. We then reviewed the abstracts and titles of the resulting 4,792 articles. Of these, 1,855 articles met the initial criteria for full-text review. Next 272 papers were excluded following the full-text inspection, leaving 1,613 studies that contained simulation modeling as a core method in addition to a policy analysis (Figure 1). This is the full sample for the study.

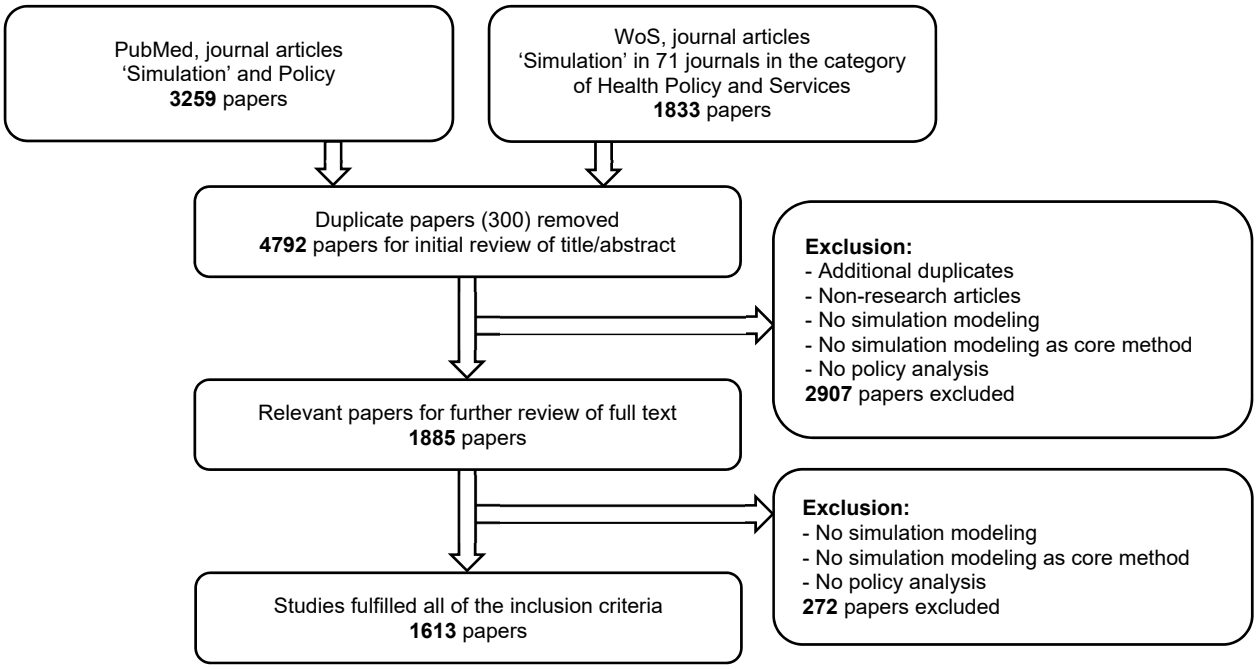

Figure 1: Study Selection Flow

Chronological, clustering, and trend analysis

Figure 2 presents the breakdown of research areas. The broad majority of all research areas across articles fall under *Diseases* (37·5%), followed by *Chemicals and Drugs* (33·96%), *Organisms* (27·55%), and *Anatomy* (1·25%) (Figure 2). Each category is broken down into the most common subcategories. In *Diseases*, the most commonly occurring categories are *Pathological Conditions*, *Signs and Symptoms*, and *Virus Diseases*; in *Chemicals and Drugs*, *Organic Chemicals* is the most common; in *Organisms*, *Viruses* constitutes the vast majority of terms; in *Anatomy*, *Body Regions* (11 articles) and *Cells* (9 articles) are most common. See Figure S1 for the number of articles in the top 20 research areas. It should be noted that 5% (77 out of 1613) of the articles evaluated did not contain MeSH terms and were not included in Figure 2. It is important to note that one article has the potential to contain MeSH terms in more than one segment of the figure.

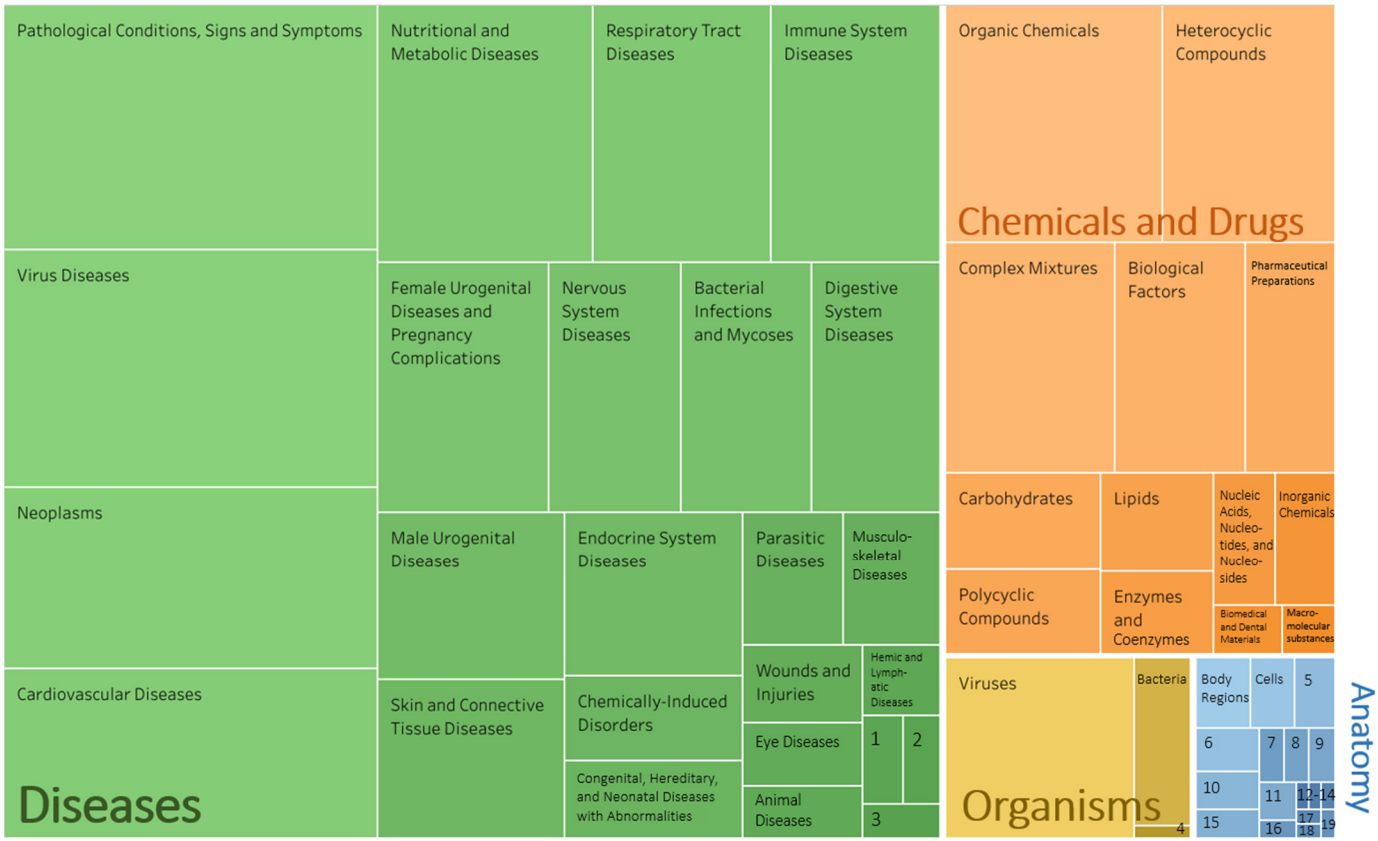

Figure 2: MeSH term categorization. Diseases (green), Chemicals and Drugs (orange), Organisms (yellow), and Anatomy (blue). For the boxes labeled by numbers, the key is as follows: 1. Occupational Diseases, 2. Otorhinolaryngologic Diseases, 3. Stomatognathic Diseases, 4. Organism Forms, 5. Fluids and Secretions, 6. Hemic and Immune Systems, 7. Musculoskeletal System, 8. Urogenital System, 9. Embryonic Structures, 10. Cardiovascular System, 11. Plant Structures, 12. Digestive System, 13. Respiratory System, 14. Nervous System, 15. Tissues, 16. Sense Organs, 17. Animal Structures, 18. Stomatognathic System, 19. Integumentary System.

Categorization based on modeling approach

Figure 3 presents the trend in reporting of models from 1967 to 2016. Until 2007, the proportion of articles that did not report a model at all was greater than the ones that did. Between 2007 and 2016, more studies reported model equations than those that did not report. The overall percentage of studies that cite an earlier study with their model equations (9%) increases from 1967 to 2016, with the largest increase seen between 2007 and 2011.

Figure 3 also presents a breakdown of model reporting across the four model type permutations. 79% of articles containing models that are both Static and Deterministic (n=135) did not report their model equations, while 15% did within the text of the paper/appendix and 6% cited another article referencing the model equations. 44% of articles containing models that are both Static and Stochastic (n=193) did not report their model equations, while 52% did within the text of the paper/appendix and 4% cited another article referencing the model equations. 62% of articles containing models that are both Dynamic and Deterministic (n=370) did not report their model equations, while 30% did within the text of the paper/appendix and 8% cited another article referencing the model equations. 34% of articles containing models that are both Dynamic and

Stochastic (n=868) did not report their model equations, while 55% did within the text of the paper/appendix and 11% cited another article referencing the model equations. Approximately half of the articles were event-driven, and half were continuous, as shown in Figure S2.

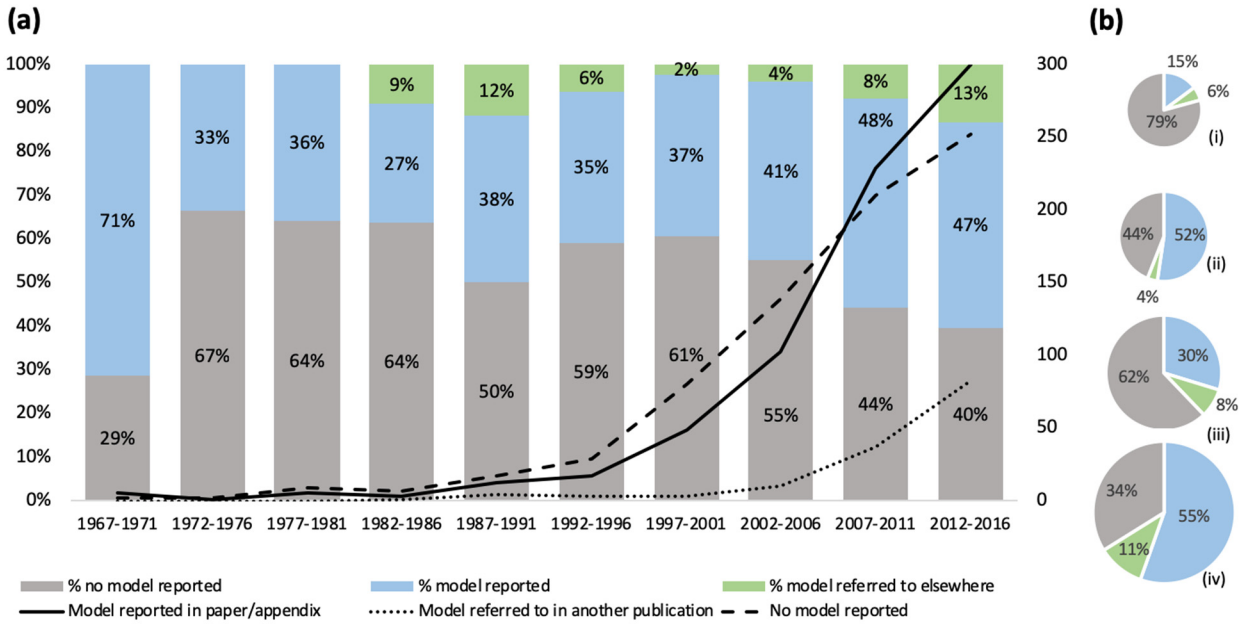

Figure 3: Model reporting trend from 1967 to 2016 and in each permutation of model categorization. (a) articles were divided into five-year publishing date increments and then assessed by whether they contained model equations or citations of model equations. (b) each article was categorized by type of model and sorted by whether the model equation used was reported in the paper/appendix, in another paper, or not at all. The models were divided into the following four descriptive categories: *static* and (i) deterministic or (ii) stochastic, or *dynamic* and (iii) deterministic or (iv) stochastic. The size of pie charts indicates sample size of each category.

Three logistic regressions were performed to analyze whether articles reported their models (either directly or by citation to original model) based on whether it was static or dynamic, event-driven or continuous, and stochastic or deterministic, and whether model reporting was significantly impacted by the age of the article. Table S5 summarizes these results. Overall, articles using static models are less likely to report those models than papers using dynamic models. In addition, event-driven and stochastic models are both more likely to be reported. Also, see Figure S2 for the trend of articles over time based on their modeling approaches.

Evaluation of modeling rigor and reporting

Figure 4 presents whether the 100 sampled articles satisfy the 26 evaluation criteria. Of the 26 criteria, four had significantly higher rates of satisfaction in the random selection of articles: *evaluation and testing* ($p = 0{\cdot}13$), *generalizability discussion* ($p=0{\cdot}09$), *high-level model visualization* ($p = 0{\cdot}07$), and *discussion about strategies/policies* ($p<0{\cdot}001$). Two had significantly higher rates of satisfaction in the most-cited articles: *quality of calibration fit* ($p=0{\cdot}08$) and *model calibration* ($p=0{\cdot}01$) (Figure 4). Modeling code availability and reproducibility discussion were the lowest in both groups with only 1-2% of the articles from either group satisfying these criteria. Conversely, limitations discussion, assumptions, scope,

objective, problem definition, parameter values and data sources, and the reporting of quantitative results were satisfied by the vast majority (>80%) of selected studies in both categories.

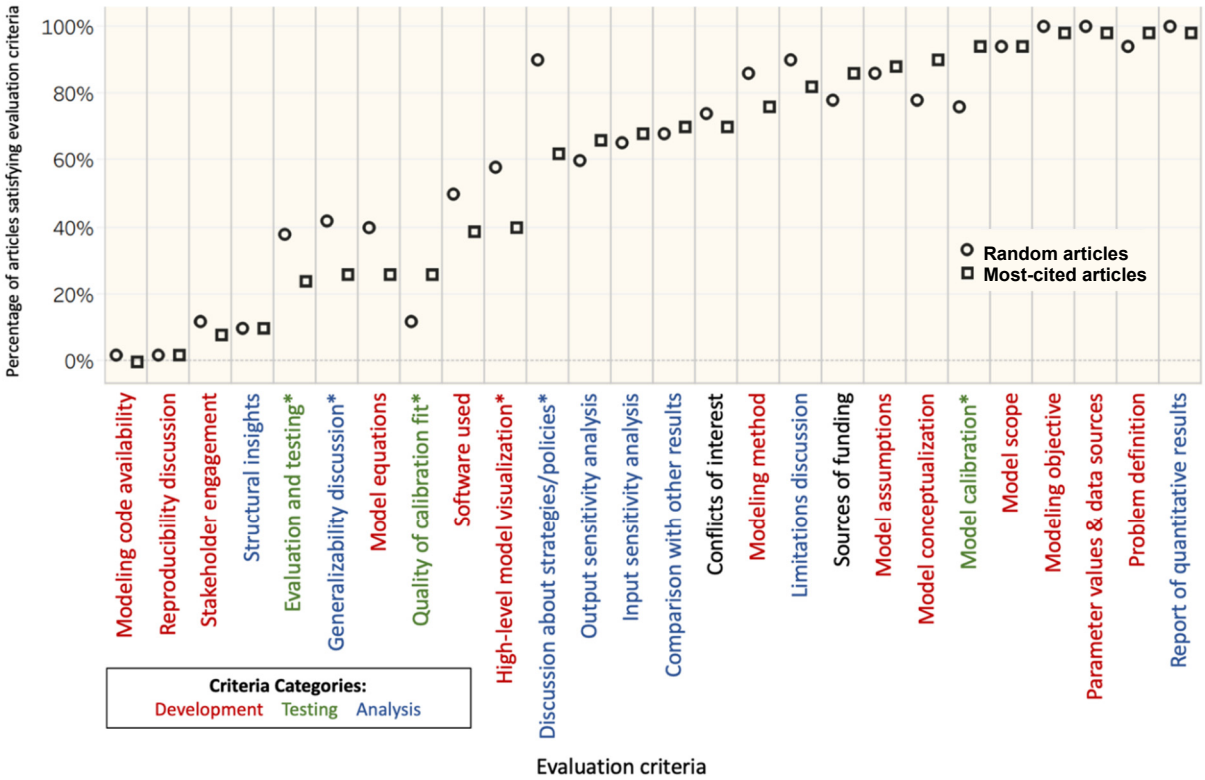

Figure 4: Percentage of articles satisfying 26 in-depth evaluation criteria in four areas. Each criterion was assessed for the 50 most-cited articles (square) and 50 random articles (circle). The percentage of articles from each group meeting the criteria are presented. Criteria marked with an * are significantly different between random and most-cited articles (p<0·1).

A principle component analysis suggests that the various criteria are not highly correlated and thus offer distinct measures for assessing the papers (See Table S6 for details). However, we found no strong predictor of aggregate quality measures. A linear regression was performed to analyze the overall evaluation score (maximum score=26 points) of the 100 sampled articles based on whether it was static or dynamic, event-driven or continuous, stochastic or deterministic, the journal impact factor, the number of citations, MCS, number of pages, time since publication, and the location of the authors' affiliation (U.S. or non-U.S.). An article's score increased only 0·07 points for each additional page (p=0·08). The other independent variables, including collaboration, number of citations received, and journal impact factor had no significant impact on the score (p>0·1). See Table S7 for the details of the regression analysis.

Discussion

In this study, we conducted a broad review of published journal articles that used simulation modeling to inform health policy in order to identify trends, document application domains, and

assess modeling rigor and reproducibility. Between 1967 and 2016, this line of research has become more common and the type of models have shifted to better represent the complexity of real-world health issues, particularly within a dynamic and stochastic framework.

In our sample of simulation modeling articles, the fraction of disease-focused articles is the largest. Areas such as pathological conditions and viral diseases receive a great deal of modeling attention, but parasitic diseases, for example, receive less modeling attention; although they affect hundreds of millions of people each year and cases are concentrated in tropical regions where populations have fewer resources [21]. Simulation modeling is well-suited to aid in understanding of complex issues where resources are limited. Within the context of global health, a gap in the literature is detrimental when considering potential interventions to prevent the spread of parasitic diseases, especially given that the projected global rise in temperature promotes the spread of these diseases [22, 23]. There are other research areas with similar potential for applications of simulation models, for example chemically-induced disorders (under which opioid use disorder is studied) [24, 25] and congenital and hereditary diseases, impact the population in increasingly complex ways [26] but are not prominent in current applications.

Our analysis also highlights the significant room for improvement both in terms of reproducibility and rigor and quality in reporting of studies. The only significant predictor of reproducibility was the length of the article; longer articles tended to perform better. Nevertheless, given the wealth of online repositories for storing models, codes, and data, page limit should no longer be a consideration in documenting models and simulation modelers can increasingly benefit from these resources. On the other hand, factors such as collaboration, number of citations, and journal reputation showed no significant impact on reproducibility. This goes against some common intuitions about the quality of research [27] and indicates a gap between perceived quality and actual reproducibility.

Our review also shows that dynamic models tend to be better reported than static models, and the same is true for stochastic and event-driven vs. deterministic and continuous models. Beginning in 2002, publications have moved toward creating dynamic and stochastic models to better reflect the real world. This time period also witnessed improvements in reporting model equations. We hope this trend will continue to accelerate as simulation models become more frequently relied upon in health policy decisions.

The most-cited articles in our sample perform no better than random articles when assessed against a host of modeling process evaluation criteria. We also find that nearly half of the studies do not report their model equations. This is consistent with a recent study where only 7.3% of simulation modeling researchers responded when asked to post their codes to a research registry clearinghouse, and only 1.6% ultimately agreed to post these details [28]. We encourage researchers to open their work to others which provides many opportunities for learning from others and enhancing the work. Moreover, connecting a model's structural qualities to the purported insights is at the heart of developing the readers' intuition and thus having an impact; much more can be done on that front [13]. While some of these suggestions may appear burdensome, they are key to building confidence in models, gaining and maintaining the trust of decision-makers, and the cumulative improvement of modeling research in general. We presented these gaps in model reporting and rigor in hopes that future publications meet a higher standard of developing and reporting simulation models and facilitate their reproducibility.

An important limitation of this study is our search strategy that focused on articles using the term "simulation" (and all its variants) in the title or abstract. We expect a sizeable number of articles exist that use simulation modeling but do not use the term or its variants in the title and abstract, instead referring to various modeling approaches. Expanding the search strategy to include mathematical, computational, economic, or other modeling approaches would have enlarged the initial sample to tens of thousands of records, which would have been infeasible to review and analyze in depth. Considering this, we chose to review a large subset of the relevant literature that is explicit about using a simulation model, while acknowledging the under-count inherent to excluding non-explicit simulation modeling approaches. We hope this balancing act has not biased our results in any specific direction. Future research may include evaluating studies against various, field-specific definitions of "simulation modeling" to increase the coverage of the sample.

This review may also be limited by selecting only peer-reviewed studies published in English and indexed on PubMed and WoS by the end of 2016. We used PubMed's MeSH term categorization to identify the focus of each article; however, we are limited by potential overlap and errors in these categories. Our sampling frame stops at 2017, yet given the rapid growth of the field updates every few years to track new trends should add value. In addition, models in health policy may be developed privately, discussed in institutional reports, or remain unpublished. Thus, the journal articles may be missing some of the relevant results in simulation modeling.

For an in-depth analysis of the 100 articles, we devised a scoring system assigning one point to each binary criteria met, following other examples of binary assessment without limiting the analysis to a single application of simulation modeling [29]. This assignment only focuses on the existence of a set of criteria and does not offer a more nuanced understanding of quality. We also do not include differential weights for these criteria in assessing the overall quality. We acknowledge that those criteria vary in their importance to overall quality and impact of research, yet, in the absence of an objective method for aggregating them, we preferred to avoid imposing our own subjective weights on various criteria.

Further reviews can focus on ways to systematically analyze and improve the reproducibility of simulation modeling. For example, they could investigate reproducibility across application domains and levels of analysis (e.g., cell, individual, or society) to inform concrete suggestions for specific communities of research. Additionally, future studies can compare our results on the reproducibility of simulation models in health policy with other application domains of simulation modeling. Tracking the same metrics over time provides another measure of progress. Finally, using our dataset, machine learning methods may be trained to identify reproducibility and quality metrics more efficiently and apply those criteria over larger bodies of research.

Conclusions

Our analysis highlights the changes in simulation modeling studies topics, methods, and quality over the last half-century and suggests several areas for improvement. Regardless of the quality of underlying model, lack of reproducibility is a major challenge that erodes confidence in the policies and decisions these papers inform and their broader impact, and thus deserves more attention in simulation research. We hope that this study facilitates conversation around research gaps that can benefit from simulation modeling and motivates modelers to collaborate in

addressing those gaps and increasing the diversity of research areas while enhancing the rigor and reproducibility of simulation research in health policy.

Acknowledgments: We would like to express our gratitude to Professors Donald Burke (University of Pittsburgh), John Sterman (MIT), and Gary King (Harvard) who shared their suggestions and thoughts. We thank participants at Medical Decision Making and System Dynamics Conferences who provided constructive feedback on initial versions of this report. We also thank Meera Gregerson, Yikang Qi, and Yuan Yuan who assisted in data extraction and analysis at MIT.

Authors' Contributors: MSJ designed and conceived the project. AG and KL searched and reviewed the literature. CD and MSJ analyzed the data extracted from the reviewed articles. HR advised the review process. MSJ and CD wrote the first draft of the manuscript, and HR contributed to further development of the analysis and content. All authors reviewed the draft of the manuscript and provided comments and critical review.

Ethical approval: Not needed.

Funding: No funding source was used to conduct this study.

Competing interests: The authors declare that there is no conflict of interest.

References:

1. Sterman, J.D., *Learning from Evidence in a Complex World.* 2006. **96**(3): p. 505-514.
2. Glied, S. and N. Tilipman, *Simulation modeling of health care policy.* Annu Rev Public Health, 2010. **31**: p. 439-55.
3. Fone, D., et al., *Systematic review of the use and value of computer simulation modelling in population health and health care delivery.* Journal of Public Health, 2003. **25**(4): p. 325-335.
4. Goldie, S.J., J.D. Goldhaber-Fiebert, and G.P. Garnett, *Chapter 18: Public health policy for cervical cancer prevention: The role of decision science, economic evaluation, and mathematical modeling.* Vaccine, 2006. **24**: p. S155-S163.
5. Stehlé, J., et al., *Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees.* BMC Medicine, 2011. **9**(1): p. 87.
6. Grimm, V., et al., *Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE.* Ecological Modelling, 2014. **280**: p. 129-139.
7. Monks, T., et al., *Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines.* Journal of Simulation, 2019. **13**(1): p. 55-67.
8. Taylor, S.J.E., et al., *Applying the stress guidelines for reproducibility in modeling & simulation: application to a disease modeling case study*, in *Proceedings of the 2018 Winter Simulation Conference*. 2018, IEEE Press: Gothenburg, Sweden. p. 739–748.
9. Goodman, S.N., D. Fanelli, and J.P.A. Ioannidis, *What does research reproducibility mean?* Science Translational Medicine, 2016. **8**(341): p. 341ps12-341ps12.
10. Begley, C.G. and J.P.A. Ioannidis, *Reproducibility in Science.* Circulation Research, 2015. **116**(1): p. 116-126.
11. Collins, F.S. and L.A. Tabak, *Policy: NIH plans to enhance reproducibility.* Nature News, 2014. **505**(7485): p. 612.

12. Taylor, S.J.E., et al., *Crisis, what crisis: does reproducibility in modeling & simulation really matter?*, in *Proceedings of the 2018 Winter Simulation Conference*. 2018, IEEE Press: Gothenburg, Sweden. p. 749–762.
13. Eddy, D.M., et al., *Model Transparency and Validation:A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7.* Medical Decision Making, 2012. **32**(5): p. 733-743.
14. Rahmandad, H. and J.D. Sterman, *Reporting guidelines for simulation-based research in social sciences.* System Dynamics Review, 2012. **28**(4): p. 396-411.
15. Galea, S., M. Riddle, and G.A. Kaplan, *Causal thinking and complex system approaches in epidemiology.* Int J Epidemiol, 2010. **39**(1): p. 97-106.
16. Speybroeck, N., et al., *Simulation models for socioeconomic inequalities in health: a systematic review.* Int J Environ Res Public Health, 2013. **10**(11): p. 5750-80.
17. Adams, J. and K.A. Gurney, *Bilateral and Multilateral Coauthorship and Citation Impact: Patterns in UK and US International Collaboration.* 2018. **3**(12).
18. Caro, J.J., et al., *Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1.* Med Decis Making, 2012. **32**(5): p. 667-77.
19. Dahabreh, I.J., et al., *Guidance for the Conduct and Reporting of Modeling and Simulation Studies in the Context of Health Technology Assessment*, in *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. 2008, Agency for Healthcare Research and Quality (US): Rockville (MD).
20. Grimm, V., et al., *Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE.* Ecological modelling, 2014. **280**: p. 129-139.
21. Brak, K., et al., *Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for Chagas disease chemotherapy.* Journal of medicinal chemistry, 2010. **53**(4): p. 1763-1773.
22. Patz, J.A., et al., *Effects of environmental change on emerging parasitic diseases.* International Journal for Parasitology, 2000. **30**(12): p. 1395-1405.
23. Trtanj, J., et al., *Ch. 6: Climate Impacts on Water-Related Illness*, in *The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment*. 2016, U.S. Global Change Research Program: Washington, DC. p. 157–188.
24. Hui, D., J. Arthur, and E. Bruera, *Palliative Care for Patients With Opioid Misuse.* Jama, 2019. **321**(5): p. 511.
25. Wei, Y.J., et al., *Performance of the Centers for Medicare & Medicaid Services' Opioid Overutilization Criteria for Classifying Opioid Use Disorder or Overdose.* Jama, 2019. **321**(6): p. 609-611.
26. Laurichesse Delmas, H., et al., *Congenital unilateral renal agenesis: Prevalence, prenatal diagnosis, associated anomalies. Data from two birth-defect registries.* Birth Defects Res, 2017. **109**(15): p. 1204-1211.
27. Mårtensson, P., et al., *Evaluating research: A multidisciplinary approach to assessing research practice and quality.* Research Policy, 2016. **45**(3): p. 593-603.
28. Emerson, J., et al., *Publication of Decision Model Source Code: Attitudes of Health Economics Authors.* PharmacoEconomics, 2019. **37**(11): p. 1409-1410.
29. Health, C.f.t.E.o.V.a.R.i., *The Cost-Effectiveness Analysis Registry [Internet]*.