JASSS
THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION

All Models Are Wrong, But Can They Be Useful? Lessons from COVID-19 Agent-Based Models: A Systematic Review

Emma Von Hoene [1], Sara Von Hoene [1], Szandra Péter [1], Ethan Hopson [1], Emily Csizmadia [1], Faith Fenyk [1], Kai Barner [1], Timothy Leslie [1], Hamdi Kavak [3], Andreas Züfle [4], Amira Roess [2], Taylor Anderson [1]

[1] *Department of Geography and Geoinformation Science, George Mason University (GMU), 4400 University Drive, Fairfax, Virginia, 22030*
[2] *Department of Global and Community Health, College of Human and Health Services, George Mason University (GMU), 4400 University Drive, Fairfax, Virginia, 22030*
[3] *Department of Computational and Data Sciences, George Mason University (GMU), 4400 University Drive, Fairfax, Virginia, 22030*
[4] *Department of Computer Science, Emory University, 201 Dowman Drive, Atlanta, Georgia, 30322*
Correspondence should be addressed to *tander6@gmu.edu*

Journal of Artificial Societies and Social Simulation 29(1) 7, 2026
Doi: 10.18564/jasss.5937 Url: http://jasss.soc.surrey.ac.uk/29/1/7.html

Received: 12-09-2025 Accepted: 16-01-2026 Published: 31-01-2026

Abstract: The COVID-19 pandemic prompted a surge in computational models to simulate disease dynamics and guide interventions. Agent-based models (ABMs) are well-suited to capture population and environmental heterogeneity, but their rapid deployment raised questions about utility for health policy. We systematically reviewed 536 COVID-19 ABM studies published from January 2020 to December 2023, retrieved from Web of Science, PubMed, and Wiley on January 30, 2024. Studies were included if they used ABMs to simulate COVID-19, where reviews were excluded. Studies were assessed against nine criteria of model usefulness, including transparency and re-use, interdisciplinary collaboration and stakeholder engagement, and evaluation practices. Publications peaked in late 2021 and concentrated in a few countries. Most models explored behavioral or policy interventions ($n = 294$, 54.85%) rather than real-time forecasting ($n = 9$, 1.68%). While most described model assumptions ($n = 491$, 91.60%), fewer disclosed limitations ($n = 349$, 65.11%), shared code ($n = 219$, 40.86%), or built on existing models ($n = 195$, 36.38%). Standardized reporting protocols ($n = 36$, 6.72%) and stakeholder engagement were less commonly reported (13.62%, $n = 73$). Only 2.24% ($n = 12$) described a comprehensive validation framework, though uncertainty was often quantified ($n = 407$, 75.93%). Over time, reporting of stakeholder engagement and evaluation increased. Studies that claimed policy relevance ($n = 354$, 66.05%) more often included some evaluation ($n = 283$, 79.94% vs. $n = 125$, 68.68%) and stakeholder engagement ($n = 61$, 17.23% vs. $n = 12$, 6.59%), though they were less likely to re-use models or share code. Limitations of this review include under-representation of non-English studies, subjective data extraction, variability in study quality, and limited generalizability. Overall, COVID-19 ABMs advanced quickly, but lacked transparency, accessibility, and participatory engagement. Stronger standards are needed for ABMs to serve as reliable decision-support tools in future public health crises.

Keywords: COVID-19, Agent-Based Models, Infectious Disease Modeling, Systematic Literature Review

Introduction

1.1 The COVID-19 pandemic led to an explosion of computational models ranging from compartmental to agent-based models (ABM) designed to forecast disease trajectories and explore the effect of public health interventions on disease outcomes. However, the push to use models for immediate policy decisions exposed disconnects between modelers' expertise and limited real-world experience as well as user expectations for real-time accuracy and the models' actual capabilities. This frustration was voiced by New York Governor, Andrew Cuomo, who remarked during a press briefing in May 2020: "*Here's my projection model. Here's my projection model. They were all wrong. They were all wrong*" (Cohen 2020). Such moments reveal how modeling outputs shaped not only research but also high-level decisions and public communication, a perspective echoed in academic critiques. Ioannidis et al. (2022) argue bluntly that COVID-19 forecasting "failed", citing the large discrepancies between model projections and observed outcomes while Jewell et al. (2020) caution that overly confident forecasts misled decision-makers when uncertainty was not adequately communicated. Jewell et al. (2020) write "*This appearance of certainty is seductive when the world is desperate to know what lies ahead*". The lack of alignment between COVID-19 projection models with observed outcomes undermined public confidence and sparked a debate about the appropriate use of simulation models in public health decision-making (Holmdahl & Buckee 2020; Ioannidis et al. 2022; Squazzoni et al. 2020).

1.2 While models can support timely interventions, flawed or misunderstood models may appear alarmist, undermine trust, and lead to overcorrections that do more harm than good. For example, the highly influential Institute for Health Metrics and Evaluation (IHME) COVID-19 model was used at both federal and state levels to justify policy interventions that shut down parts of the US economy (Robbins 2021). However, the model assumed uniform compliance with social distancing policies, producing inaccurate estimates with narrow uncertainty bands (Jewell et al. 2020; Schroeder 2021). IHME specifically has long been criticized for lack of transparency in their methods for generating global health metrics (Shiffman & Shawar 2020). Kitching et al. (2006) remind us that this tension is not unique to COVID-19 simulations. A widely cited public health example is the case of the 2001 UK foot-and-mouth disease outbreak, in which a model assumed transmission was possible between adjacent farms (Ferguson et al. 2001). This assumption led the model to forecast a much larger and faster-growing outbreak than occurred, prompting aggressive culling of millions of livestock. The response caused severe economic losses, long-term environmental impacts, and eroded public trust in science-driven policy (Anderson 2021; Kitching et al. 2006).

1.3 Drawing on classic ideas from Box (1979), Holmdahl & Buckee (2020) reiterate that despite being "wrong", models can still be useful. While no model can capture the full complexity of the real world, they have the potential to offer valuable insights from estimating important disease parameters (e.g., R_0, R_E) to providing guidance on interventions. For example, government organizations used models to forecast the size of the 2014 Ebola outbreak and the geographic patterns of spread in West Africa, allowing decision-makers to prioritize where to direct resources and improve surveillance (Fischer 2016; Meltzer et al. 2014; WHO Ebola Response Team 2014). In another example, models were used to estimate the R_0 heterogeneity of a 2008 outbreak of cholera in Zimbabwe, shaping vaccine and clean-water deployment plans (Mukandavire et al. 2011). Despite uncertainty in COVID-19 model forecasts, the *COVID-19 Forecast Hub* (Cramer et al. 2022), provided ensemble forecasts that combined multiple models, resulting in improved accuracy with clearer communication of uncertainty, and informed decisions by organizations like the CDC in their official COVID-19 guidance.

1.4 The COVID-19 pandemic has reignited long-standing questions about what makes models useful for guiding public health policy decisions. Scholars argue that models fail not only when they are inaccurate, but when their assumptions, scope, and intended uses are unclear (Anderson 2021; Ioannidis et al. 2022; Saltelli et al. 2020). Multiple recent commentaries and major calls for reform in the modeling community have highlighted common principles of making disease simulation models truly useful. Saltelli et al. (2020) offer a manifesto: models should be transparent about assumptions and limitations, openly shared, and co-produced with stakeholders; they should clarify uncertainty and avoid false precision. Squazzoni et al. (2020) echo these ideas with a call for open-source software and tools, adoption of standard protocols for model documentation, and the use of permanent online repositories of model code to speed up assessment and model re-use. Eker (2020) stresses that although COVID-19 models can be useful in many ways, they require validation and a clear communication of their uncertainties to be credible for decision-making. Squazzoni et al. (2020) warns that a model that is not rigorously checked could have serious consequences in the findings are used to inform decision making.

1.5 The COVID-19 pandemic offers a rare opportunity to reflect on how simulation models supported public health decision-making, revealing both advances in modeling practice and persistent gaps that limit their utility. Among the different simulation approaches, ABMs are broadly described as being a useful tool for decision-support

(Bui & Lee 1999; DeAngelis & Diaz 2019; Klabunde & Willekens 2016; Sengupta & Bennett 2003). Epidemiological ABMs offer unique insights due to their ability to simulate complex interactions and behaviors between heterogeneous individuals (i.e., agents) and their environments from which disease outcomes emerge and are well-suited for exploring "what if" scenarios, such as alternative policy interventions (Castro & Ford 2021; de Andrade et al. 2023; Von Hoene et al. 2023). In this paper, we conducted a systematic review of COVID-19 ABM studies published between 2020 and 2023 to answer the following research question: "*To what extent did COVID-19 ABMs adhere to best practices of transparency and re-use, interdisciplinary collaboration and stakeholder engagement, and model evaluation – and how did these practices evolve over time*"? Our assessment focuses on published literature in the acute phase of the COVID-19 pandemic, when ABMs were most in demand and uncertainty was greatest, to evaluate their usefulness for policymaking and identify key successes or limitations.

1.6 Drawing on expert critiques from the literature (see references in Table 1), we assess each study against a set of nine criteria widely considered essential for making disease simulation models "useful": (i) use of open source ABM tools for model development, (ii) adoption of standard protocols for model communication, (iii) clear acknowledgment of model assumptions and limitations, (iv) use of permanent model repositories for model code, (v) building upon existing models, (vi) collaboration across multiple disciplines, (vii) inclusion of stakeholders in the modeling cycle, (viii) implementation of model evaluation, and (ix) quantification of model uncertainty.

1.7 The remainder of this paper is structured as follows. We first describe the methodology used to search, retrieve, screen, and assess relevant studies. We then present the results of our systematic review and discuss key trends, gaps, and implications. Finally, we conclude by outlining limitations and directions for future research.

Methodology

2.1 To address our research question, we conducted a systematic review of published COVID-19 ABM studies, following established Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines for transparency and reproducibility (Page et al. 2021) (see the Supplementary Materials, files 7 and 8). The methodology consisted of three stages: (i) literature search and retrieval, (ii) screening and eligibility assessment, and (iii) data extraction and analysis.

2.2 **Assessment Framework and Criteria Definitions**. Nine criteria, grouped under the categories of "Model Transparency and Re-Use", "Interdisciplinary Collaboration and Stakeholder Engagement", and "Model Evaluation", were derived from early expert commentaries that critiqued the application of COVID-19 models in practice. Because no existing framework assesses the usefulness of infectious disease simulation models in practice, we began with a scoping review to identify relevant critiques. We then synthesized common themes, including recurring critiques and suggestions related to model usefulness. Themes that appeared in at least two commentaries were included as criteria in our assessment framework. The selected criteria reflect widely cited indicators of model usefulness and are intended to support transparent and systematic comparison across models, rather than representing a validated or exhaustive list. Finally, related criteria were grouped into three higher-level categories. Table 1 lists and defines the final set of criteria, along with a brief rationale for their inclusion.

2.3 **Search strategy and retrieval**. Based on Zhang et al. (2023), we considered the following academic web portals and databases for our search, including: PubMed, MEDLINE, EMBASE, Web of Science, Scopus, Science-Direct, EBSCO, Wiley, and WHO COVID-19 Database. MEDLINE was excluded since it draws from PubMed. Science-Direct was excluded since it limits the number of Boolean operators in a search. WHO COVID-19 Database was excluded since it stopped operation in June 2023. EMBASE and Scopus were excluded due to lack of organizational access.

2.4 Due to these exclusions, articles were retrieved from Web of Science, EBSCO, PubMed, and Wiley on 1/30/2024 using a four-year date range of 01/01/2020-12/31/2023 to capture all COVID-19 related models. We focus on this period because it represents the acute phase of the pandemic and a historical moment for infectious disease modeling efforts, during which COVID-19 ABM development occurred under conditions of high uncertainty and urgent public health decision-making needs. Following Zhang et al. (2023), the search strategy consisted of two main components. The first component included a comprehensive search of COVID-19 related terms including 'coronavirus disease 2019', 'COVID-19', 'COVID-2019', 'severe acute respiratory syndrome coronavirus 2', 'SARS-CoV2' and 'SARS-CoV-2'. The second component captured articles based on modeling methodologies, using terms such as 'agent-based models', 'agent-based modelling', 'individual-based model', and 'multi-agent system'. Each component was combined internally using OR operators. The final search result was obtained by combining the two components using the AND operator, retrieving articles at the intersection of both COVID-19

and ABM. We include the search history for the databases in the Supplement Materials, file 1. All Supplementary Material files for this study can be found in the Supplementary Materials section.

Category	Criteria	Description and Rationale
Model Transparency and Re-Use	Use of open-source ABM tools for model development	Models implemented in publicly available agent-based modeling platforms (e.g., NetLogo, GAMA, Julia, Repast4Py, Mason) promote access for reviewers and replication (Ioannidis et al. 2022; Squazzoni et al. 2020).
	Adoption of standard protocols for model communication	Use of standardized reporting frameworks (e.g., ODD (Grimm et al. 2010), ODD+D (Müller et al. 2013)) ensures clear and consistent documentation, making models easier to understand, compare, and replicate (Hadley et al. 2021; Lorig et al. 2021; Squazzoni et al. 2020).
	Clear acknowledgment of model assumptions and limitations	Explicitly stating assumptions and limitations defines the scope of a model and increases transparency (Eker 2020; Saltelli et al. 2020).
	Model downloadable from permanent repository	Making model code available in permanent repositories (e.g., GitHub, Zenodo) supports reuse, assessment, and long-term accessibility (Ioannidis et al. 2022; Squazzoni et al. 2020).
	Re-use or extension upon existing models	Extending or reusing prior models reduces redundancy, fosters comparability across studies, and enhances credibility by building on validated frameworks (Schroeder et al. 2022; Squazzoni et al. 2020).
Interdisciplinary collaboration and stakeholder engagement	Interdisciplinary collaboration	Including expertise across multiple disciplines ensures models capture biological, technical, and social dimensions of disease spread (Ioannidis et al. 2022; Squazzoni et al. 2020).
	Inclusion of stakeholders in the modeling cycle	Involving policymakers, public health practitioners, or community representatives ensures model relevance and increases the likelihood of model uptake (Saltelli et al. 2020; Squazzoni et al. 2020).
Model evaluation	Implementation of model evaluation	Verification, calibration, sensitivity analysis, and validation assess credibility and robustness, increasing trust in the model as a decision-support tool (Eker 2020; Squazzoni et al. 2020).
	Model uncertainty quantification	Explicitly measuring and communicating uncertainty (e.g., through confidence intervals, sensitivity testing) acknowledges model limits and avoids false precision (Ioannidis et al. 2022; Saltelli et al. 2020).

Table 1: Assessment framework and criteria definitions.

2.5 Sensitivity testing of the search strings revealed the importance of intentionally including specific disease-related terms (e.g., 'coronavirus disease 2019') rather than broader terms (e.g., 'disease outbreaks'), allowing us to efficiently retrieve articles most directly relevant to our criteria. This specificity minimized the retrieval of irrelevant or overly general records, thereby increasing the precision of our literature search. A total of 902 articles were retrieved, as follows: Web of Science ($n = 893$), PubMed ($n = 3$), Wiley ($n = 6$), EBSCO ($n = 0$). In

general, PubMed, Wiley and EBSCO failed to return articles at the intersection of the ABM approach and COVID-19. For example, PUBMED yielded 17,489 articles using keywords related to COVID-19 and 171 articles using keywords related to ABMs, the intersection of these two searchers returned only 3 articles. Of the 902 articles, 32 were excluded since they were not conference proceedings or journal articles and 10 were marked as duplicates by automation tools, resulting in 860 records for title and abstract screening (Figure 1).

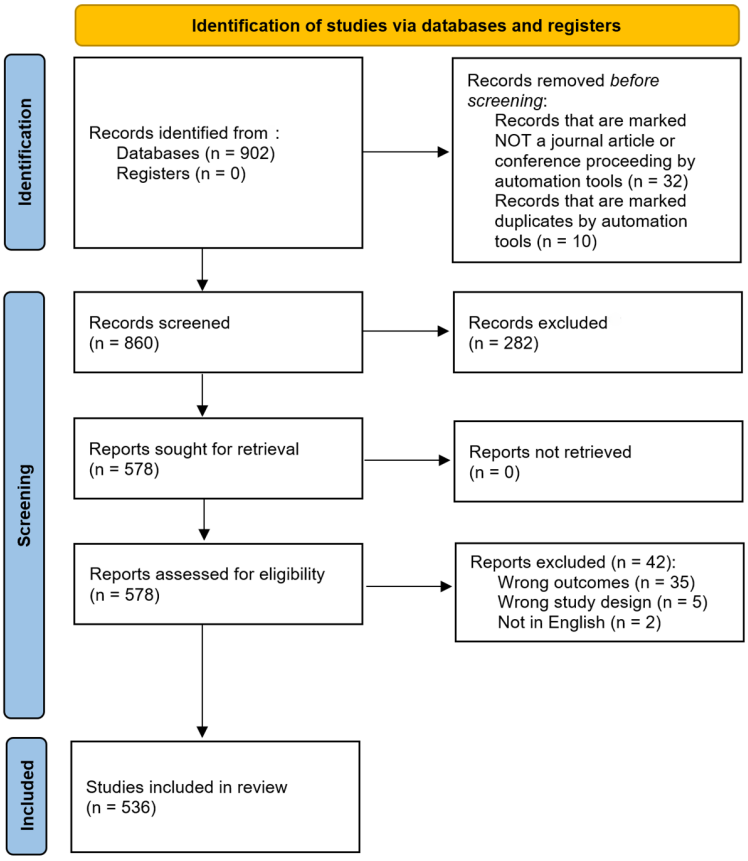

Figure 1: PRISMA 2020 flow diagram outlining our systematic review protocol (adapted from Page et al. 2021).

2.6 Title and Abstract Screening. Rayyan systematic review software (rayyan.ai) was used to manage the initial screening of the 860 articles. At each stage, two authors (E.V.H. and T.A) screened each article's titles and abstracts manually and independently, using the inclusion and exclusion criteria as described in Table 2. Rayyan's blind-review feature allowed both reviewers to screen titles and abstracts independently without seeing each other's decisions, reducing bias. There were 97 conflicts (87% agreement), which were resolved through discussion. Through this title and abstract screening, we excluded 282 articles and included 578 articles for full text assessment (Figure 1).

2.7 Full Text Assessment. Covidence systematic review software (app.covidence.org) was used to support full-text screening and conflict tracking. At least two authors (T.A. and several rotating second reviewers including F.F., S.P., K.B., E.C., and E.H.) manually assessed the full text of each article independently. There were 7 conflicts (98% agreement), which were resolved by author E.V.H. Through the full text assessment and the inclusion and exclusion criteria described in Table 2, we excluded 42 articles and included a final 536 articles for data extraction (Figure 1). The list of the 536 articles can be found in the Supplement Materials, file 2.

2.8 Data Extraction. Covidence was also used to coordinate data extraction by assigning articles to reviewers, storing extracted information in a structured form, and automatically flagging discrepancies when paired reviewers entered different responses. Each article was assigned to two independent reviewers for data extraction (F.F., S.P., K.B., E.C., S.V.H., and E.H.). The data extraction form can be found in the Supplementary Materials, file 3. Across all pairs of reviewers, there was an average agreement of 74% (min: 68.8%, max: 83.8%). Disagreements were resolved by a third senior researcher (T.A. or E.V.H.), who used Covidence's interface to compare both reviewers' entries with the source text and make a final decision (see the Supplementary Materials, file 4 for example screenshots illustrating the use of Rayyan and Covidence). The final dataset comprised all data

points agreed upon by reviewer pairs, or when disagreements occurred, the values determined through this consensus process. It is important to note that Rayyan and Covidence facilitated workflow organization, blinded screening decisions, and streamlined conflict resolution, but did not automate inclusion, exclusion, or coding decisions. All judgments were made manually, so any potential bias reflects reviewer interpretation rather than the software itself.

Reason	Include	Exclude
Outcome	The study presents or describes a model that simulates the transmission of COVID-19 from one individual to another. Studies that model the transmission of disease often utilize parameters to model the spread of the disease and should be included (disease staging including susceptible, infectious, recovered, length of time in each stage, transmission likelihood etc.).	The model simulates the transmission of other diseases, outside of COVID-19. Studies that measure risk as a proxy for likelihood of COVID-19 transmission should be excluded. This is typically the case in very small-scale simulations (e.g. simulating the boarding of a plane).
Study Design	The modeling approach is an agent-based model. Approaches with other names equivalent to an "agent-based model" such as "multi-agent", "individual-based model" should be included. If the smallest unit is an individual, it is likely an agent-based model. Hybrid models where one component of the model is an agent-based model should be included.	Other common models that are solely machine learning models or compartmental models should be excluded.
Other Consideration	The article is either a conference article or a journal article. The article should be written in English.	Review articles should be excluded. Articles not in English should be excluded.

Table 2: The inclusion and exclusion criteria.

Results and Discussion

3.1 **Overview**. Our systematic review identified and extracted data from 536 COVID-19 ABM studies published between January 2020 and December 2023. The full dataset is openly accessible in the Supplementary Materials, file 9. The first COVID-19 ABM study in our collection was published on May 20, 2020. Figure 2 shows the quarterly publication trends of COVID-19 ABM studies from 2020 through 2023. Publications peaked in July-September 2021, which may reflect a lag between early pandemic model demand and their appearance in published literature. Although models are needed most urgently at the outset of an epidemic's outset, the time required for development, calibration, peer review and publication processes may delay availability until after the critical decision-making window has passed.

3.2 The COVID-19 ABM studies were published across 261 publication outlets including 219 journals, 2 books, and 40 conference proceedings. The top three most common publication outlets were *Scientific Reports* (36 studies), *PLoS ONE* (31 studies), *PLoS Computational Biology* (16 studies). Among the 219 journals, 103 were fully open access (OA), 112 followed a hybrid model (subscription with OA options), and 4 were subscription only. We note that of the 103 OA journals, 88 were listed in the Directory of Open Access Journals (DOAJ), regarded as a "whitelist" of OA journals, and 15 journals (totaling 30 studies) were not. This may suggest variability in the quality of the studies included in our analysis.

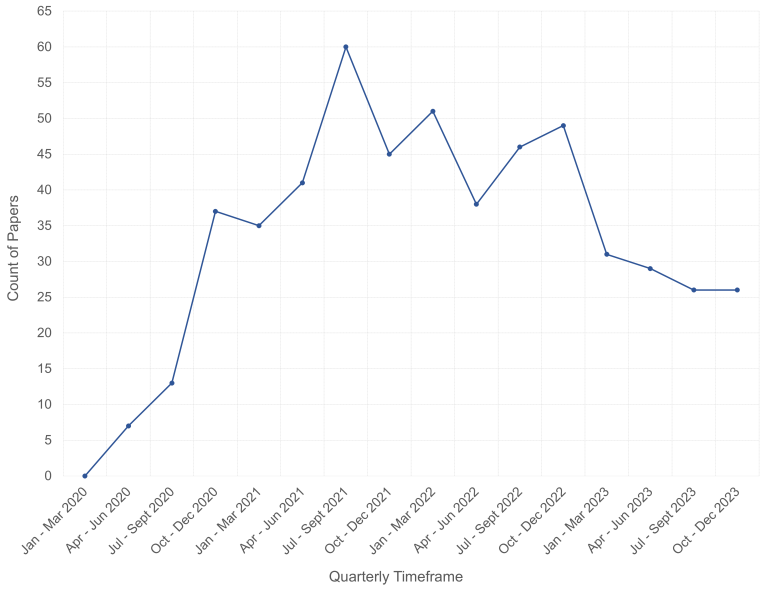

Figure 2: Count of COVID-19 ABM studies per quarter from 2020 to 2023.

3.3 Geographically, the countries where COVID-19 ABM case studies were undertaken spanned all continents excluding the poles. The largest concentrations were models of the United States ($n = 109$, 20.34%), United Kingdom ($n = 50$, 9.33%), China ($n = 49$, 9.14%), Australia ($n = 27$, 5.04%), Canada ($n = 25$, 4.66%), Italy ($n = 25$, 4.66%), France ($n = 18$, 3.36%), and Germany ($n = 18$, 3.36%), which together accounted for a large share of studies (Figure 3).

3.4 Across the 536 studies, the predominant purpose of the COVID-19 ABMs (Figure 4) was to understand the role of specific factors or interventions in disease transmission and control ($n = 383$, 71.46%). A majority of studies assessed the effects of policy guidelines and health behaviors ($n = 294$, 54.85%; e.g., mask mandates, school closures, or vaccine uptake), followed by contact networks ($n = 26$, 4.85%; e.g., spread across different network types, such as workplaces, schools, long-term care facilities, families, etc.), spatial effects ($n = 18$, 3.36%; e.g. urban vs. rural transmission patterns), disease parameters ($n = 16$, 2.99%; e.g., variant-specific dynamics), mobility ($n = 11$, 2.05%; e.g., influence of varying travel patterns), population structure ($n = 11$, 2.05%; e.g., introducing heterogeneity into agent populations), and the spread of misinformation or information ($n = 7$, 1.31%; e.g., how information impacts health behavior uptake). A smaller share of studies ($n = 63$, 11.75%) proposed new modeling frameworks to simulate the spread of COVID-19 in specific locations or settings ($n = 32$, 5.97%; e.g., hospitals, university campuses), to combine ABM with other simulation approaches ($n = 16$, 2.99%; e.g. hybrid ABM-discrete event simulation), and to generically model disease spread ($n = 15$, 2.80%; e.g., FRED (Grefenstette et al. 2013), Covasim (Kerr et al. 2021). Another 15.11% ($n = 81$) developed or refined methods for the following: advancing disease modeling ($n = 57$, 10.63%; e.g., a framework for implementing health behaviors), assessing model outcomes ($n = 17$, 3.17%; e.g., strategies for calibrating, sensitivity analysis and validation), and improving computational efficiency ($n = 7$, 1.31%; e.g., optimizing ABM to simulate faster with more agents). Only 1.68% ($n = 9$) of studies aimed to produce epidemic forecasts under existing conditions, reflecting the challenges of real-time calibration in rapidly evolving conditions with little data availability. These findings highlight an important difference in purpose: forecasting ABMs predict how an epidemic will unfold in the short term from known starting conditions, while policy-assessment ABMs test 'what-if' scenarios to see how hypothetical guidelines or behaviors might affect outcomes. Overall, the distribution of study purposes suggests that ABMs during COVID-19 were developed as exploratory tools to test the impact of interventions, examine behavioral and structural drivers of disease spread, and innovate on ABM methodology.

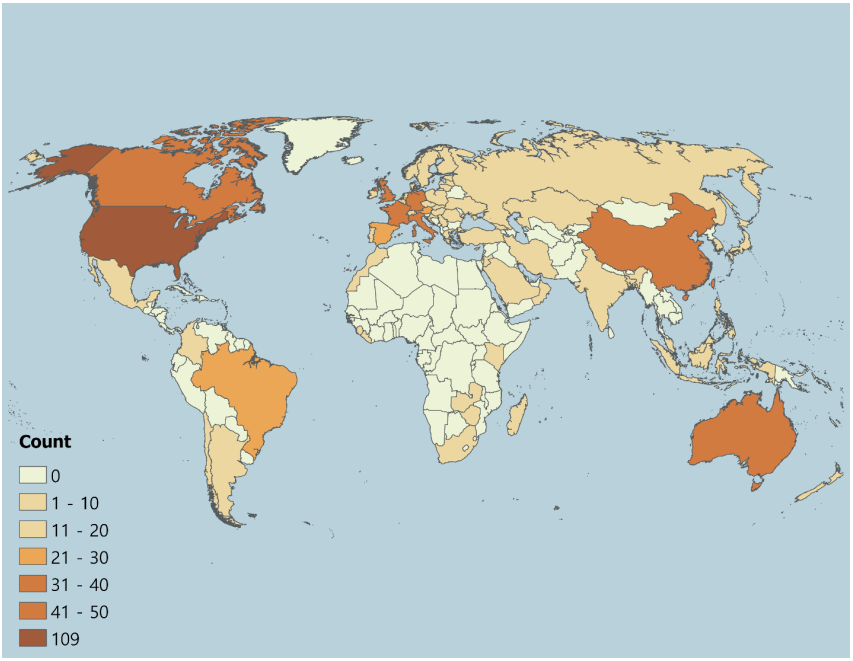

Figure 3: Geographic distribution of COVID-19 ABMs by study country. Counts represent the number of studies in which the model was applied to simulate disease spread in that country. The map was created in ArcGIS Pro, and the generalized world boundary was obtained from ESRI's open-source ArcGIS Online Portal (`https://www.arcgis.com/home/item.html?id=2b93b06dc0dc4e809d3c8db5cb96ba69`).

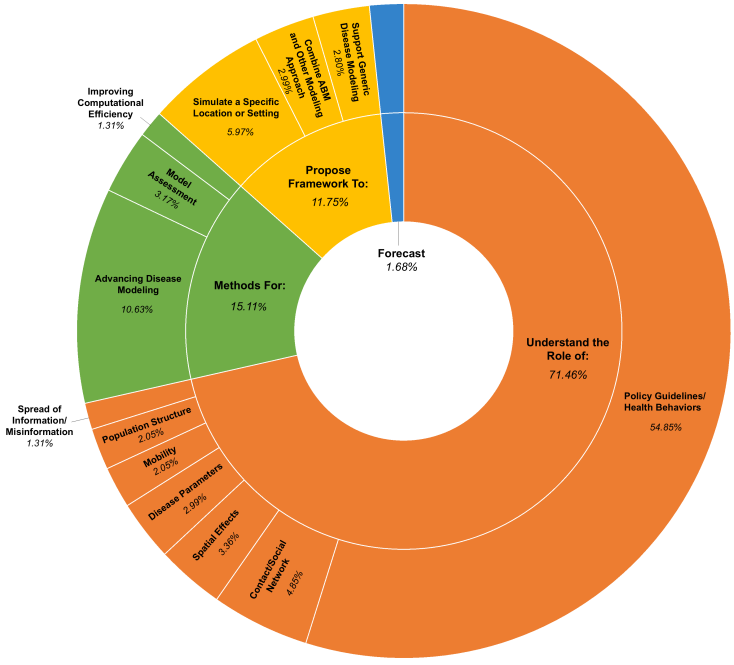

Figure 4: Percentages of COVID-19 ABM studies by purpose (inner pie chart) and by sub-purpose (outer pie chart).

3.5 Two-thirds of the studies ($n = 354$, 66.05%) explicitly stated that their models or model outputs are "useful" for

informing public health policy making or other researchers, highlighting the field's emphasis on practical application. The implication that COVID-19 ABMs will be used for decision making support may place a responsibility on the modeling community to use best practices in model development, documentation, and evaluation. Figure 5 provides an overview of the assessment of the COVID-19 ABM studies against a set of nine criteria that are widely considered essential for translating models into useful tools for decision-making, falling more broadly under the categories of "Model Transparency and Re-Use", "Interdisciplinary Collaboration and Stakeholder Engagement", and "Model Evaluation". Analyses are presented for all studies, for studies claiming potential policy usefulness versus non-claiming studies (policy-claiming, $n = 354$, 66.05%; non-claiming, $n = 182$, 33.95%), and by publication year from 2020 to 2023 (2020: $n = 58$, 10.82%; 2021: $n = 181$, 33.77%; 2022: $n = 184$, 34.33%; 2023: $n = 113$, 21.08%). Percentages shown in Figure 5 are also reported in a table available in the Supplementary Materials, file 5. These percentages are calculated relative to the number of papers in each subset: for all studies, percentages are out of 536; for policy-claiming versus non-claiming studies, percentages are out of 354 and 182, respectively; and for each publication year, percentages are out of the total papers published that year. Below, we present and discuss these results. Supplementary Materials, file 6 provides a more detailed sub-analysis of the criteria.

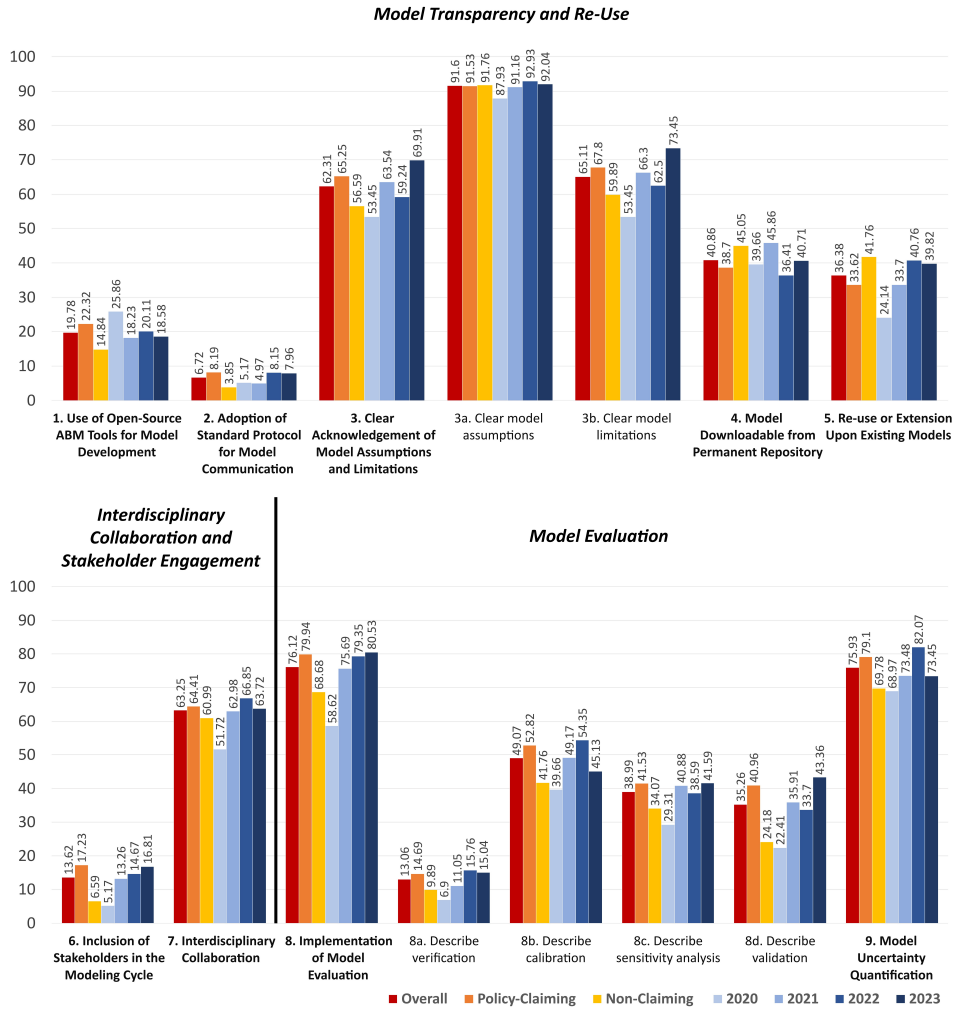

Figure 5: Assessment of COVID-19 ABM studies against nine criteria essential for decision-support modeling. Criteria are analyzed across all articles, articles claiming potential usefulness, and by publication year (2020–2023). Percentages are calculated relative to the total papers in each subset (see text for details).

3.6 Model Transparency and Re-Use. Overall, 19.78% ($n = 106$) of models were implemented with open-source ABM software, with NetLogo being the most common platform ($n = 57$, 10.63% of total studies), followed by GAMA ($n = 13$, 2.43%), and Julia ($n = 10$, 1.87%). In contrast, most studies relied on custom-built code in general-purpose programming languages such as Python, C++, and Java. As Squazzoni et al. (2020) note, the use of open-source platforms and tools minimizes barriers to replication and re-use, which may help explain

why only 36.38% (n = 195) of studies explicitly built upon existing models. One exception is Covasim (Kerr et al. 2021), developed in Python and openly shared with extensive documentation, which emerged as one of the most frequently reused frameworks ($n = 30$, 5.60%). This suggests that re-use may depend not only on whether a model is implemented in open-source ABM software, but also on accessibility and transparency. We also found that studies that did not extend existing models were more likely to be published in open access journals not listed in the DOAJ (6.88%) compared to studies that did (3.88%). This may suggest that model re-use could be associated with publication in higher-quality venues, potentially reflecting available time or resources to align with stronger modeling and reporting practices.

3.7 On that note, adoption of standard protocols for model documentation was reported in 6.72% of studies ($n = 36$), with the Overview Design and Details (ODD) protocol (Grimm et al. 2010, 2020) accounting for most ($n = 31$, 5.78%). More detailed frameworks such as ODD+D (Decisions) (Müller et al. 2013) and International Society for Pharmacoeconomics and Outcomes Research Society for Medical Decision Making (ISPOR-SMDM) Modeling Good Research Practices (Caro et al. 2012) were almost entirely absent ($< 1\%$). Clear articulation of model assumptions was common ($n = 491$, 91.60%), yet explicit statements of limitations were provided in only 65.11% ($n = 349$) of studies. Although 62.31% ($n = 334$) of studies included a description of both assumptions and limitations, indicating variation in how comprehensively studies described model scope. Less than half ($n = 219$, 40.86%) made their model code available for download, most frequently via GitHub ($n = 169$, 31.53%), followed by Zenodo ($n = 18$, 3.36%) and Gitlab ($n = 10$, 1.87%). These findings are consistent with those from Zavalis & Ioannidis (2022) which find that only 25% of all disease models published in 2021 (including ABM) share coded, suggesting a systematic issue. The lack of sharing code likely has implications for reproducibility and limits the ability of researchers and public health practitioners to adapt existing models for other applications. Establishing shared community standards for code archiving and documentation, whether through journal policies or collaborative initiatives, could strengthen the transparency and reuse of ABMs in future public health challenges (see Janssen et al. 2020 for discussion).

3.8 To promote transparency, facilitate reproducibility, and encourage model reuse, we provide a compiled list of available model code links from the 536 reviewed articles in the Supplementary Materials, file 10 at: `https://www.jasss.org/29/1/7/Supplement_10.xlsx`. Given repeated calls in the literature for increased model transparency, the low uptake of open-source ABM software, standard protocols for model documentation, and availability of model code may reflect systematic issues related to time pressure, the cost burden (e.g., money, time, effort), or lack of awareness. Furthermore, while several journals recommend using such tools, it is not typically a requirement.

3.9 Interdisciplinary Collaboration and Stakeholder Engagement. Most studies ($n = 339$, 63.25%) included authors from more than one discipline, with 28.92% ($n = 155$) involving 3 or more (Supplementary Materials, file 6), which likely reflects some alignment with best-practice recommendations for interdisciplinary collaboration. Disciplines were classified according to the U.S. National Science Foundation (NSF) Codes for Classification for Research (The University of New Mexico, Office of Sponsored Projects 2025). We also added a category for industry/non-academic fields (e.g., government organizations or contractors, hospitals, churches). The x-axis of Figure 6 shows each discipline alongside the number and percentage of papers (out of 536) that included at least one author from that discipline. The most represented disciplines included life sciences ($n = 252$, 47.01%), industry/non-academic fields ($n = 208$, 38.81%), computer and information sciences ($n = 167$, 31.16%), and engineering ($n = 137$, 25.56%). The interdisciplinary collaborations between discipline pairs (i.e., number of papers with at least one author from the two disciplines) are shown in each cell in Figure 6. For each pair, a chi-square test (χ^2) compared the observed number of papers (the first bolded value in each cell) coauthored by the two disciplines to the expected number ($\hat{n}$) under random collaboration. Cells colored in gray indicate that there is no more or less collaboration than there would be at random (p-value greater than 0.05). Colored cells (also denoted with an $*$ after each cell's parentheses) indicate statistically significant deviations: orange represents more collaborations than expected, and blue indicates fewer.

3.10 Collaborations that occurred more frequently than expected and were statistically significant included life sciences (which include biomedical sciences and health sciences) and industry/non-academic fields ($n = 120$, $\hat{n} = 97.8$, $\chi^2 = 15.6$) followed by life sciences and mathematics and statistics ($n = 67$, $\hat{n} = 52.7$, $\chi^2 = 9.3$). Notably, while authors from engineering appeared in about 25% of studies (more than mathematics and statistics), collaborations with life sciences occurred less frequently than expected ($n = 46$, $\hat{n} = 64.4$, $\chi^2 = 13.3$), although possibly engineering has less of a stake in disease simulations. It is important to note that the chi-square analysis relies on an implicit assumption: the expected number of collaborations is calculated as if each paper had exactly two authors, with each author randomly paired with one co-author. In practice, papers may have only one author or many co-authors, which can increase or decrease the observed number of collaborations relative to this expectation. As a result, some observed patterns, particularly in fields with many multi-author

papers (e.g., life sciences), may reflect differences in author count rather than true disciplinary preference.

3.11 Only 13.62% ($n = 73$) of studies reported direct stakeholder involvement in the modeling cycle. Of these studies, the most common stakeholder roles were to inform model assumptions ($n = 23$, 4.29%), guide model intervention scenarios ($n = 21$, 3.92%), use model results for decision making ($n = 16$, 2.99%), and provide expert validation ($n = 12$, 2.24%). Less frequent roles included model testing and calibration ($n = 3$, 0.56%), providing data ($n = 3$, 0.56%), supporting dissemination of model results ($n = 2$, 0.37%), further analyzing the model data ($n = 2$, 0.37%), and defining performance metrics ($n = 1$, 0.19%). Stakeholders ranged from institutional actors including city, county, and state governments (e.g., "City of Chicago, Cook County, and State of Illinois"), public health agencies (e.g., "National Health Service in the United Kingdom", "Health Canada", "Polish Ministry of Health"), schools or university administration (e.g., "UC San Diego Return to Learn Program") to individual actors including public health officials, educators, healthcare providers, business owners, local planners, and leaders of worship.

3.12 Yet, stakeholder participation was reported in 13.26% of studies, while two-thirds of studies claimed potential policy relevance. Yet, this aligns with literature that points to a longstanding gap between modelers and policy makers (Mihaljevic et al. 2024). The lack of engagement may be attributed to several practical and political challenges: models can seem like "black boxes" that are hard to trust, policies often change quickly and unpredictably, decisions must be made on shorter timelines than models can support, government contracting processes can be slow and inflexible, political values can outweigh evidence, and there are usually many different stakeholders to involve (Gilbert et al. 2018). As Edmonds et al. (2019) states: a model can only be validated relative to a clearly defined purpose, and if stakeholders are not involved in setting that purpose, it is difficult to claim validity for policy or operational decision-making.

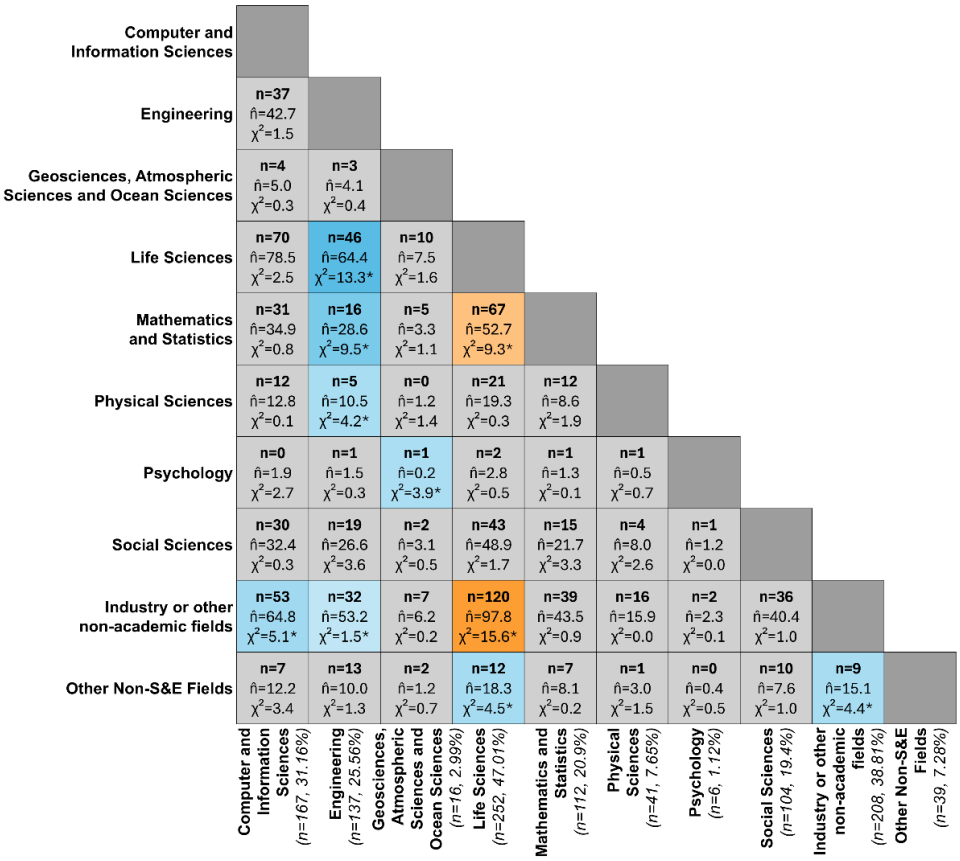

Figure 6: Matrix of interdisciplinary collaboration across the 536 papers assessed in this study. Within each cell, bold values indicate the observed number of papers with at least one author from both disciplines. $\hat{n}$ denotes the expected number of papers under random collaboration, while χ^2 represents the corresponding chi-square statistic. Colored cells and an asterisk ($*$) denote statistically significant differences ($p < 0.05$): blue indicates fewer collaborations than expected, orange indicates more collaborations than expected, and grey indicates non-significant results. The x-axis includes the count and percentages of papers with at least one author from the given discipline.

3.13 **Model Evaluation**. Overall, 76.12% ($n = 408$) of studies mention at least one evaluation method: verification, calibration, sensitivity analysis, or validation. Calibration was the most commonly reported evaluation practice ($n = 263$, 49.07%), followed by sensitivity analysis ($n = 209$, 38.99%) and validation ($n = 189$, 35.26%). Common approaches included comparing model outcomes with observed data, or comparing results with another modeling approach. Model verification was reported less frequently ($n = 70$, 13.06%), possibly reflecting its integration into routine model development rather than explicit reporting. Verification practices included running the model under extreme conditions to ensure logical consistency and systematically adjusting key parameters. While most COVID-19 ABM studies incorporated some form of evaluation, only 2.24% ($n = 12$) described a systematic framework encompassing verification, calibration, sensitivity analysis, and validation together. Communication of either aleatory or epistemic model uncertainty was more frequent ($n = 407$, 75.93%), typically through repeated runs with confidence intervals or sensitivity analysis, respectively. Taken together, these findings highlight a tension between the urgency of producing models and the time-intensive process of rigorous evaluation. The limited use of evaluation methods, particularly multi-method evaluation, likely reflects constraints on available data, the absence of clear standards for evaluating ABMs in the community, and the time pressures associated with model development during the pandemic. Given that two-thirds of studies positioned themselves as potentially useful for public health policy, the uneven application of evaluation methods likely raises concerns about the credibility and reproducibility of their insights. Strengthening evaluation practices remains essential if ABMs are to serve as reliable decision-support tools in future pandemics.

3.14 **Trends Over Time**. Figure 5 shows the patterns of adherence to best practices in COVID-19 ABMs over time (2020-2023). Note that percentages are calculated relative to the total number of studies published each year. Because modeling standards were evolving rapidly during the early stages of the pandemic, we hypothesized that earlier studies may be less likely to meet the criteria. Transparency indicators remained relatively low overall: the use of open-source ABM tools fluctuated between 18-25% without a clear upward trajectory, while the adoption of standard communication protocols were consistently adopted by less than 9% of studies each year. Reporting of model assumptions or limitations strengthened over time. Assumptions remained consistently well documented (> 87%), but explicitly describing the model limitations rose from 53.45% ($n = 31$) in 2020 to over 73.45% ($n = 83$) in 2023. Availability of models for download hovered around 36-46% without consistent gains. Evidence of re-use improved modestly whereby studies building on existing models increased from about 24% in 2020 to nearly 40% by 2022-23.

3.15 Interdisciplinary collaboration remained steady (about 63-67%) between 2021-2023 after a weaker start in 2020 (51.72%). Stakeholder engagement, while limited overall, grew from 5% in 2020 to almost 17% in 2023, suggesting gradual recognition of the value of participatory modeling. Evaluation practices showed the clearest progress. The share of studies applying verification, calibration, sensitivity analysis, or validation rose approximately from 59% in 2020 to 81% in 2023. Calibration generally remained consistent over time, but increased 10% between 2020 and 2021. Validation more than doubled over the study period, from 22.41% ($n = 13$) to 2020 to 43.36% ($n = 49$) in 2023. Communication of uncertainty was reported in 69-82% of studies, peaking in 2022. In general, we find that while COVID-19 ABM development show increased attention to model evaluation and stakeholder engagement over time, gaps in transparency, documentation, and open accessibility persist. We note that there is often a lag between submission and publication dates, and analyses based on submission dates could produce slightly different temporal patterns than those reported here.

3.16 **Policy-Claiming vs. Non-Claiming Studies**. We classified studies as policy-claiming if they included an explicit statement that the model or its outputs were intended to be used by policymakers or other decision-makers. Such claims were identified based on the study's language/text indicating potential policy use (e.g., "inform", "support", "assist", or "guide" policy-making). Studies that presented models and results without an explicit claim regarding policy use were classified as non-claiming. Of the 536 COVID-19 ABM studies, 354 (66.05%) explicitly claimed policy relevance, while 182 (33.95%) did not. Percentages reported below are calculated relative to the number of studies in each group. Comparing these groups shows that, although policy-claiming studies were somewhat more likely to meet best-practice criteria, the differences were modest and inconsistent. For example, policy-claiming studies were slightly more likely to use open-source ABM tools ($n = 79/354$, 22.32% vs. $n = 27/182$, 14.84%), follow standard communication protocols ($n = 29$ 8.19% vs. $n = 7$ 3.85%), and acknowledge assumptions and limitations ($n = 231$, 65.25% vs. $n = 103$, 56.59%), although overall adoption of existing models remained low. Notably, non-claiming studies were more likely to make code available ($n = 82$, 45.05% vs. $n = 137$, 38.70%) and to build upon existing models ($n = 76$, 41.76% vs. $n = 119$, 33.62%). Interdisciplinary authorship was common in both groups ($n = 228$, 64.41% vs. $n = 111$, 60.99%). Stakeholder involvement, however, was more than twice as frequent among policy-claiming studies ($n = 61$, 17.23% vs. $n = 12$, 6.59%). The clearest differences emerged in evaluation. Policy-claiming studies more often reported verification, calibration, sensitivity analysis, or validation ($n = 283$, 79.94% vs. $n = 125$, 68.68%), particularly validation ($n = 145$, 40.96% vs. $n = 44$, 24.18%), which is critical for model credibility. These findings

suggest that while many policy-oriented studies took additional steps to strengthen credibility, there remain opportunities to better align claims of usefulness with best-practice standards.

Key takeaways

3.17 The rapid development of COVID-19 ABMs sought to support timely exploration of policy scenarios under conditions of uncertainty throughout the early stages of the pandemic. *The findings from this review ultimately reveal that early urgent calls to action for increased transparency, collaboration and stakeholder engagement, and model evaluation were not consistently met by the research community*. This may be explained by a combination of technical, institutional, and structural barriers that widened the gap between simulation modeling and policymaking during COVID-19. Generally, extreme values observed for certain criteria, such as low rates of forecasting ABMs or use of standardized protocols, and high rates of reporting model assumptions, likely reflect inherent features of ABMs (e.g., stochasticity requiring clear assumptions and their strengths in exploratory analysis). Furthermore, given the intense time pressure and uncertainty surrounding the pandemic, it is possible that some COVID-19 ABMs were produced with a focus on timeliness over extensive documentation or code preparation for public release. Publishing norms have only recently begun to mandate open data, long-term code repositories, and standardized documentation protocols. Stakeholder engagement requires trusted partnerships and communication protocols, which can be difficult to initiate and build during crisis conditions, especially for newly funded research teams. Likewise, model evaluation was most likely constrained by the limited availability of reliable and comprehensive datasets at the onset of the pandemic. Collectively, these pressures likely contribute to explaining the uneven adoption of best practices across COVID-19 ABMs.

3.18 What is striking, however, is that these challenges are not new. Commentaries and calls to action over the past decade repeatedly identified the same obstacles in infectious disease modeling. For example, Metcalf et al. (2015) discussed the importance of transparency and integration of modelers and model building into the policy cycle. Weston et al. (2018) highlighted the importance of interdisciplinary collaboration, particularly the inclusion of behavioral scientists, to ensure that models reflect social complexity. Outside the infectious disease community, Heppenstall et al. (2021) observed that even after decades of methodological advances, ABMs continue to face enduring difficulties in validation and in establishing credibility for policy application.

3.19 Unfortunately, this study does not claim to offer definitive solutions to these challenges. As we have pointed out, there are already many commentaries that put forward best practices towards improving the utility of models in health policy and decision-making (Eker 2020; Hadley et al. 2021; Saltelli et al. 2020; Squazzoni et al. 2020). Rather, this study contributes in two ways. First, it synthesizes recurrent themes from existing calls to action and operationalizes them as criteria to assess the practical usefulness of COVID-19 ABMs. Second, it provides a snapshot of the field during an unprecedented real-world crisis, identifying both progress and persistent shortcomings. While the proposed assessment framework could be further expanded, we hope that it offers a starting point to encourage modelers to reflect on their own practices and to work toward a shared, transparent framework for evaluating model usefulness in future public-health emergencies. We also note that this review identifies studies that were successful in model transparency, stakeholder engagement and collaboration, and model evaluation that the research community may learn from.

Limitations and Future Work

4.1 This review has several limitations that should be acknowledged. First, by focusing exclusively on published journal and conference articles, we may have excluded relevant models described in preprints, reports, or those that remain unpublished. Second, geographic and linguistic biases may exist given our focus on English-language publications, potentially underrepresenting work from non-English-speaking regions and low- and middle-income countries (Figure 3). Third, we examined only ABMs, meaning that findings cannot be generalized to the broader COVID-19 modeling community. Additionally, data extraction involved some subjective judgments. While discrepancies between the two reviewers were resolved through consensus with a senior researcher, the potential for bias remains and may influence the reported results. The scope of this review is limited to 2020–2023, a period that marks the acute stage of the pandemic and the rapid development of COVID-19- ABMs. Because modeling objectives and the focus on COVID-19 in public health likely shifted after 2023, this offers an opportunity to conduct an additional comparative study to analyze COVID-19 ABM practices beyond the crisis context.

4.2 Future work could extend this review to include other modeling approaches (e.g., compartmental, machine learning) to determine whether the observed gaps are unique to ABMs or common across simulation methodologies. 30 studies included in our analysis were published in journals that were both OA and not listed in the DOAJ, suggesting variability in study quality. Future work may further investigate whether OA journals participate in activities considered "predatory" and conduct a sub-analysis of these studies. Additionally, other indicators of study quality such as journal impact factor or number of citations could be explored and the nuance between model usefulness and model quality could be more deeply explored. To our knowledge, no framework for assessing the usefulness of infectious disease simulation models currently exists. Future research could build on the foundation of our assessment framework to develop a more formal and systematic approach. Furthermore, while our nine criteria were grounded in expert commentary, the data extraction process inevitably involved some subjectivity, with variability across reviewers. Despite these limitations, this review offers a longitudinal analysis on ABM modeling practices during the COVID-19 pandemic, highlighting both important progress and persistent gaps. Future research could build on these insights by identifying barriers that undermine best practices and by continued development and promotion of frameworks, tools, and participatory modeling approaches that strengthen them.

Conclusion

5.1 This review assessed the extent to which COVID-19 ABMs published between 2020 and 2023 adhere to best practices of transparency and re-use, interdisciplinary collaboration and stakeholder engagement, and model evaluation. Among more than 500 studies, models were widely used to explore the effects of interventions with some studies reporting that their models were used directly in practice. However, in general, models often fell short of maximizing their potential policy utility. While reporting of limitations, model re-use, evaluation practices, and stakeholder engagement improved over time, there is limited adoption overall. Overall, policy-claiming studies tended to demonstrate stronger evaluation practices and somewhat higher levels of stakeholder engagement, but model code is only publicly available in 38.70% of studies. These findings suggest that while ABMs proved valuable as exploratory tools, their potential as decision-support systems may have been only partially realized. Strengthening standards for transparency, evaluation, and interdisciplinary collaboration and stakeholder engagement remains essential for future pandemic preparedness.

Acknowledgments

This research was funded by National Science Foundation's Division of Environmental Biology (Award #2109647) and Division of Information and Intelligent Systems (Award #2302970).

Supplementary Materials

The Supplementary Material files, including direct access links, are provided below.

- **Supplementary Material 1**. Search histories: `https://www.jasss.org/29/1/7/Supplement_1.zip`
- **Supplementary Material 2**. List of papers included in the study: `https://www.jasss.org/29/1/7/Supplement_2.docx`
- **Supplementary Material 3**. Data extraction form: `https://www.jasss.org/29/1/7/Supplement_3.docx`
- **Supplementary Material 4**. Screenshots of screening software used: `https://www.jasss.org/29/1/7/Supplement_4.docx`
- **Supplementary Material 5**. Table describing COVID-19 ABM studies against nine criteria (corresponds to Figure 5): `https://www.jasss.org/29/1/7/Supplement_5.docx`
- **Supplementary Material 6**. Additional sub-analysis of criteria: `https://www.jasss.org/29/1/7/Supplement_6.docx`

- **Supplementary Material 7**. PRISMA abstract checklist: https://www.jasss.org/29/1/7/Supplement_7.docx
- **Supplementary Material 8**. PRISMA article checklist: https://www.jasss.org/29/1/7/Supplement_8.docx
- **Supplementary Material 9**. Full dataset from extraction: https://www.jasss.org/29/1/7/Supplement_9.csv
- **Supplementary Material 10**. List of models and links: https://www.jasss.org/29/1/7/Supplement_10.xlsx

References

Anderson, W. (2021). The model crisis, or how to have critical promiscuity in the time of Covid-19. *Social Studies of Science*, *51*(2), 167–188

Box, G. E. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in Statistics*, (pp. 201–236). New York, NY: Academic Press

Bui, T. & Lee, J. (1999). An agent-based framework for building decision support systems. *Decision Support Systems*, *25*(3), 225–237

Caro, J. J., Briggs, A. H., Siebert, U. & Kuntz, K. M. (2012). Modeling good research practices - Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Medical Decision Making*, *32*(5), 667–677

Castro, D. A. & Ford, A. (2021). 3D agent-based model of pedestrian movements for simulating COVID-19 transmission in university students. *ISPRS International Journal of Geo-Information*, *10*(8)

Cohen, S. (2020). "We All Failed"- The real reason behind NY Governor Andrew Cuomo's surprising confession. Forbes

Cramer, E. Y., Huang, Y., Wang, Y., Ray, E. L., Cornell, M., Bracher, J., Brennen, A., Castro Rivadeneira, A. J., Gerding, A., House, K., Jayawardena, D., Kanji, A. H., Khandelwal, A., Le, K., Mody, V., Mody, V., Niemi, J., Stark, A., Shah, A., Wattanchit, N., Zorn, M. W., Reich, N. G. & US COVID-19 Forecast Hub Consortium (2022). The United States COVID-19 Forecast Hub dataset. *Scientific Data*, *9*(1), 462

de Andrade, B. S. S., Espindola, A. L. L., Faria Junior, A. & Penna, T. J. P. (2023). Agent-based model for COVID-19: The impact of social distancing and vaccination strategies. *International Journal of Modern Physics C*, *34*(10)

DeAngelis, D. L. & Diaz, S. G. (2019). Decision-making in agent-based modeling: A current review and future prospectus. *Frontiers in Ecology and Evolution*, *6*, 237

Edmonds, B., Le Page, C., Bithell, M., Chattoe-Brown, E., Grimm, V., Meyer, R., Montañola-Sales, C., Ormerod, P., Root, H. & Squazzoni, F. (2019). Different modelling purposes. *Journal of Artificial Societies and Social Simulation*, *22*(3), 6

Eker, S. (2020). Validity and usefulness of COVID-19 models. *Humanities and Social Sciences Communications*, *7*(1)

Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. (2001). The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science*, *292*(5519), 1155–1160

Fischer, L. S. (2016). CDC grand rounds: Modeling and public health decision-making. *Morbidity and Mortality Weekly Report*, *65*

Gilbert, N., Ahrweiler, P., Barbrook-Johnson, P., Narasimhan, K. P. & Wilkinson, H. (2018). Computational modelling of public policy: Reflections on practice. *Journal of Artificial Societies and Social Simulation*, *21*(1), 14

Grefenstette, J. J., Brown, S. T., Rosenfeld, R., DePasse, J., Stone, N. T., Cooley, P. C., Wheaton, W. D., Fyshe, A., Galloway, D. D., Sriram, A., Guclu, H., Abraham, T. & Burke, D. S. (2013). FRED (A Framework for Reconstructing Epidemic Dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Public Health*, *13*(1), 940

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

Grimm, V., Railsback, S. F., Vincenot, C. E., Berger, U., Gallagher, C., DeAngelis, D. L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A. S. A., Milles, A., Nabe-Nielsen, J., Polhill, J. G., Radchuk, V., Rohwäder, M.-S., Stillman, R. A., Thiele, J. C. & Ayllón, D. (2020). The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. *Journal of Artificial Societies and Social Simulation*, *23*(2), 7

Hadley, L., Challenor, P., Dent, C., Isham, V., Mollison, D., Robertson, D. A., Swallow, B. & Webb, C. R. (2021). Challenges on the interaction of models and policy for pandemic control. *Epidemics*, *37*, 100499

Heppenstall, A., Crooks, A., Malleson, N., Manley, E., Ge, J. & Batty, M. (2021). Future developments in geographical agent-based models: Challenges and opportunities. *Geographical Analysis*, *53*(1), 76–91

Holmdahl, I. & Buckee, C. (2020). Wrong but useful - What Covid-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine*, *383*(4), 303–305

Ioannidis, J. P. A., Cripps, S. & Tanner, M. A. (2022). Forecasting for COVID-19 has failed. *International Journal of Forecasting*, *38*(2), 423–438

Janssen, M. A., Pritchard, C. & Lee, A. (2020). On code sharing and model documentation of published individual and agent-based models. *Environmental Modelling & Software*, *134*, 104873

Jewell, N. P., Lewnard, J. A. & Jewell, B. L. (2020). Caution warranted: Using the Institute for Health Metrics and Evaluation Model for predicting the course of the COVID-19 pandemic. *Annals of Internal Medicine*, *173*(3), 226–227

Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., Nunez, R. C., Cohen, J. A., Selvaraj, P., Hagedorn, B., George, L., Jastrzebski, M., Izzo, A., Fowler, G., Palmer, A., Delport, D., Scott, N., Kelly, S. L., Bennette, C. S., Wagner, B. G., Chang, S. T., Oron, A. P., Wenger, E. A., Panovska-Griffiths, J., Famulare, M. & Klein, D. J. (2021). Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*, *17*(7)

Kitching, R. P., Thrusfield, M. V. & Taylor, N. M. (2006). Use and abuse of mathematical models: An illustration from the 2001 foot and mouth disease epidemic in the United Kingdom. *Revue Scientifique et Technique de l'OIE*, *25*(1), 293–311

Klabunde, A. & Willekens, F. (2016). Decision-making in agent-based models of migration: State of the art and challenges. *European Journal of Population*, *32*(1), 73–97

Lorig, F., Johansson, E. & Davidsson, P. (2021). Agent-based social simulation of the Covid-19 pandemic: A systematic review. *Journal of Artificial Societies and Social Simulation*, *24*(3), 5

Meltzer, M. I., Atkins, C. Y., Santibanez, S., Knust, B., Petersen, B. W., Ervin, E. D., Nichol, S. T., Damon, I. K., Washington, M. L. & Centers for Disease Control and Prevention (CDC) (2014). Estimating the future number of cases in the Ebola epidemic - Liberia and Sierra Leone, 2014-2015. *MMWR Supplements*, *63*(3), 1–14

Metcalf, C. J. E., Edmunds, W. J. & Lessler, J. (2015). Six challenges in modelling for public health policy. *Epidemics*, *10*, 93–96

Mihaljevic, J. R., Chief, C., Malik, M., Oshinubi, K., Doerry, E., Gel, Y., Hepp, C., Lant, T., Mehrotra, S. & Sabo, S. (2024). An inaugural forum on epidemiological modeling for public health stakeholders in Arizona. *Frontiers in Public Health*, *12*, 1357908

Mukandavire, Z., Liao, S., Wang, J., Gaff, H., Smith, D. L. & Morris, J. G. (2011). Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in Zimbabwe. *Proceedings of the National Academy of Sciences*, *108*(21), 8767–8772

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H. & Schwarz, N. (2013). Describing human decisions in agent-based models - ODD + D, an extension of the ODD protocol. *Environmental Modelling & Software*, *48*, 37–48

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P. & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71

Robbins, T. R. (2021). An assessment of the IHME Covid-19 model: US fatalities in 2020. *American Journal of Management*, *21*(2), 10–38

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke Jr, R., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., Stirling, A., van der Sluijs, J. & Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. *Nature*, *582*(7813), 482–484

Schroeder, S. A. (2021). How to interpret Covid-19 predictions: Reassessing the IHME's model. *Philosophy of Medicine*, *2*(1)

Schroeder, S. A., Vendome, C., Giabbanelli, P. J. & Montfort, A. M. (2022). Towards reusable building blocks to develop COVID-19 simulation models. 2022 Winter Simulation Conference (WSC)

Sengupta, R. R. & Bennett, D. A. (2003). Agent-based modelling environment for spatial decision support. *International Journal of Geographical Information Science*, *17*(2), 157–180

Shiffman, J. & Shawar, Y. R. (2020). Strengthening accountability of the global health metrics enterprise. *The Lancet*, *395*(10234), 1452–1456

Squazzoni, F., Polhill, J. G., Edmonds, B., Ahrweiler, P., Antosz, P., Scholz, G., Chappin, É., Borit, M., Verhagen, H., Giardini, F. & Gilbert, N. (2020). Computational models that matter during a global pandemic outbreak: A call to action. *Journal of Artificial Societies and Social Simulation*, *23*(2), 10

The University of New Mexico, Office of Sponsored Projects (2025). NSF codes for classifications for research. Available at: `https://osp.unm.edu/pi-resources/nsf-research-classifications.html`

Von Hoene, E., Roess, A., Achuthan, S. & Anderson, T. (2023). A framework for simulating emergent health behaviors in spatial agent-based models of disease spread. Proceedings of the 6th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation

Weston, D., Hauck, K. & Amlôt, R. (2018). Infection prevention behaviour and infectious disease modelling: A review of the literature and recommendations for the future. *BMC Public Health*, *18*(1), 336

WHO Ebola Response Team (2014). Ebola virus disease in West Africa - The first 9 months of the epidemic and forward projections. *New England Journal of Medicine*, *371*(16), 1481–1495

Zavalis, E. A. & Ioannidis, J. P. A. (2022). A meta-epidemiological assessment of transparency indicators of infectious disease models. *PLoS One*, *17*(10), e0275380

Zhang, W., Liu, S., Osgood, N., Zhu, H., Qian, Y. & Jia, P. (2023). Using simulation modelling and systems science to help contain COVID-19: A systematic review. *Systems Research and Behavioral Science*, *40*(1), 207–234