

1 A rubric for assessing conformance to the Ten Rules for  
2 credible practice of modeling and simulation in healthcare

3  
4 Alexandra Manchel<sup>1,10</sup>, Ahmet Erdemir<sup>2,10</sup>, Lealem Mulugeta<sup>3,10</sup>, Joy P. Ku<sup>4,10</sup>, Bruno V. Rego<sup>5,10</sup>,  
5 Marc Horner<sup>6,10</sup>, William W Lytton<sup>7,8,10</sup>, Jerry G. Myers, Jr.<sup>9,10</sup>, Rajanikanth Vadigepalli<sup>1,10,\*</sup>

6  
7 <sup>1</sup> Department of Pathology and Genomic Medicine, Thomas Jefferson University, Philadelphia,  
8 PA, USA

9 <sup>2</sup> Department of Biomedical Engineering, Lerner Research Institute, Cleveland Clinic, OH, USA

10 <sup>3</sup> InSilico Labs LLC, Houston, TX USA

11 <sup>4</sup> Department of Bioengineering, Stanford University, Stanford, CA, USA

12 <sup>5</sup> Department of Biological & Agricultural Engineering, Louisiana State University, Baton Rouge,  
13 LA, USA

14 <sup>6</sup> Ansys Inc. Evanston, IL, USA

15 <sup>7</sup> Downstate Health Sciences University, Brooklyn, NY, USA

16 <sup>8</sup> Kings County Hospital, Brooklyn, NY, USA

17 <sup>9</sup> NASA - John H Glenn Research Center, Cleveland, OH, USA

18 <sup>10</sup> Committee on Credible Practice of Modeling, & Simulation in Healthcare, Interagency  
19 Modeling and Analysis Group and Multiscale Modeling Consortium, Bethesda, MD USA

20

21

22 \* Corresponding Author

23 Email: [rajanikanth.vadigepalli@jefferson.edu](mailto:rajanikanth.vadigepalli@jefferson.edu)

24

25

## 26 **Abstract**

27 The power of computational modeling and simulation (M&S) is realized when the results are  
28 credible, and the workflow generates evidence that supports credibility for the context of use. The  
29 Committee on Credible Practice of Modeling & Simulation in Healthcare was established to help  
30 address the need for processes and procedures to support the credible use of M&S in healthcare  
31 and biomedical research. Our community efforts have led to the Ten Rules (TR) for Credible  
32 Practice of M&S in life sciences and healthcare. This framework is an outcome of a  
33 multidisciplinary investigation from a wide range of stakeholders beginning in 2012. Here, we  
34 present a pragmatic rubric for assessing the conformance of an M&S activity to the TR. This rubric  
35 considers the ability of the M&S to facilitate outreach of the results to a wide range of stakeholders  
36 from context-specific M&S practitioners to policymakers. It uses an ordinal scale ranging from  
37 Insufficient (zero) to Comprehensive (four) that is applicable to each rule, providing a uniform  
38 approach for comparing assessments across different reviewers and different models. We used  
39 the rubric to evaluate the conformance of two computational modeling activities: 1. six viral  
40 disease (COVID-19) propagation models, and 2. a model of hepatic glycogenolysis with neural  
41 innervation and calcium signaling. These examples were used to evaluate the applicability of the  
42 rubric and illustrate rubric usage in real-world M&S scenarios including those that bridge scientific  
43 M&S with policymaking. The COVID-19 M&S studies were of particular interest because they  
44 needed to be quickly operationalized by government and private decision-makers early in the  
45 COVID-19 pandemic and were accessible as open-source tools. Our findings demonstrate that  
46 the TR rubric represents a systematic tool for assessing the conformance of an M&S activity to  
47 codified good practices and enhances the value of the TR for supporting real-world decision-  
48 making.

## 49 Introduction

50 The role of computational modeling and simulation (M&S) in healthcare research and  
51 clinical practice is expanding at a rapid pace. M&S approaches have been integral to the progress  
52 in biomedical sciences and are starting to enable *in silico* and systems medicine efforts (1,2).  
53 Computational modeling is relatively new in clinical and biomedical settings, necessitating the  
54 standardization of M&S efforts. The addition of standardized practices increases the credibility of  
55 the practice of M&S in this area as it has done in other disciplines, such as engineering. This also  
56 increases the M&S usefulness and widespread adaptation. Multiple standards have been  
57 proposed, both in industry and government, for establishing and ensuring credibility of M&S  
58 practices in various engineering fields including medical devices (3–5). Similarly, multiple  
59 standards exist for systems biology applications aimed at addressing conceptual information,  
60 nomenclature, data formats, and representations of biochemical systems, and are intended to  
61 improve the communication and sharing of M&S components (6).

62  
63 In order to promote this standardization process in the biomedical community beyond  
64 those working in systems biology, the IMAG/MSM Committee on Credible Practice of Modeling  
65 and Simulation in Healthcare developed the “Ten Rules for Credible Practice of Modeling and  
66 Simulation in Healthcare” (7). This framework is an outcome of multidisciplinary input from a wide  
67 range of stakeholders (8,9). These rules aim to establish a unified conceptual framework to  
68 design, implement, evaluate, and communicate the activities, products, and outcomes of M&S in  
69 the biomedical sciences and clinical care domain. In application, the unified framework enables  
70 outreach to the entire M&S user community, ranging from model developers to policy makers to  
71 clinicians and other non-M&S practitioners.

72

73           Recently, Tatka et al. (6) reviewed the existing standards for representing and  
74 documenting systems biology models. Current standards are limited to agreed-upon modeling  
75 formats as a means to share information; however, there is no widely utilized standard for  
76 assessing credibility of the practice in this area. As Tatka et al. (6) noted in their review, standards  
77 for model annotation must become more widely accepted such that interoperability, reusability,  
78 comparability, and comprehension can be improved. Credible practice will also be enhanced  
79 when the information needed for simulation and parameter estimation is explicitly defined and  
80 stated. Lastly, reproducibility would not be possible without efficient dissemination of all artifacts  
81 and proper documentation on an open-source repository platform. The authors note that there is  
82 a lack of consensus on quantitative credibility scoring and that a system that addresses this area  
83 would provide the community of practice with a metric for comparing the credibility of models and  
84 a guide for the development of more credible models.

85  
86           The Committee’s Ten Rules for Credible Practice of M&S in Healthcare (Table 1)  
87 establishes initial standards for systems modeling and beyond (7). While every effort was made  
88 to thoroughly describe and define the rules, it lacks a quantitative, rigorous, and repeatable metric.  
89 A consistent application of the rules likely requires a complementary rubric for assessing  
90 conformance to the rules and evaluating the credibility of the M&S practice. Such a rubric would  
91 be used to assess and communicate various aspects of the Ten Rules (TR), including the validity,  
92 level of detail, and overall “correctness” of the M&S practice.

93  
94 **Table 1: The Committee’s Ten Rules of credible practice of M&S in healthcare (7).**

Rule		Description
1.	Define context clearly	Develop and document the subject, purpose, and intended use(s) of the model or simulation
2.	Use contextually appropriate data	Employ relevant and traceable information in the development or operation of a model or simulation
3.	Evaluate within context	Perform verification, validation, uncertainty quantification, and sensitivity analysis of the model or

		simulation with respect to the reality of interest and intended use(s) of the model or simulation
4.	List limitations explicitly	Provide restrictions, constraints, or qualifications for or on the use of the model or simulation for consideration by the users or customers of a model or simulation
5.	Use version control	Implement a system to trace the time history of modeling and simulation activities including delineation of each contributors' efforts
6.	Document appropriately	Maintain up-to-date informative records of all modeling and simulation activities, including simulation code, model mark-up, scope and intended use of modeling and simulation activities, as well as users' and developers' guides
7.	Disseminate broadly	Share all components of modeling and simulation activities, including simulation software, models, simulation scenarios and results
8.	Get independent reviews	Have the modeling and simulation activity reviewed by nonpartisan third-party users and developers
9.	Test competing implementations	Use contrasting modeling and simulation implementation strategies to check the conclusions of different strategies against each other
10.	Conform to standards	Adopt and promote generally applicable and discipline specific operating procedures, guidelines, and regulations accepted as best practices

95

96 In principle, a case can be made that each of the Ten Rules in Table 1 needs its own  
97 assessment approach. For example, Rule 1 on defining the context of use can be assessed  
98 according to factors that quantify the level of detail in the documentation of the M&S subject,  
99 scope of the M&S purpose or results and intended use of the M&S results such as to support  
100 clinical decision making, inform regulatory evidence, or to inform next research steps. Additionally,  
101 Rule 3 on model evaluation requires extensive consideration of how the M&S activity and results  
102 are verified and validated, as well as how the assessment is presented to support the intended  
103 use. This rule is consistent with U.S. Food and Drug Administration (FDA) guidance and American  
104 Society of Mechanical Engineers (ASME) standards for best practices in verification and  
105 validation of medical devices (10,11). However, such a customized, rule-specific assessment

106 approach might become overly complex and unwieldy for consistently evaluating conformance to  
107 the Ten Rules.

108

109 To address this issue, the Committee formulated a rubric based on the ability of the M&S  
110 to facilitate outreach of the results to a wide range of stakeholders from context-specific M&S  
111 practitioners to policymakers. Since there exist various levels of M&S expertise in the healthcare  
112 domain, the need for direct and clear communication of M&S results is essential. The  
113 development of the Ten Rules facilitates such communication and understanding of  
114 computational modeling implementation and simulation results between stakeholders. This  
115 includes, but is not limited to, communication between the model developers, M&S practitioners,  
116 model end-users, as well as clinicians, policy makers, and other decision makers who depend on  
117 the knowledge generated by the M&S.

118

119 Most recently, the global response to the COVID-19 pandemic highlights the need for a  
120 systematic assessment of credible practice of M&S across this entire spectrum of stakeholders  
121 (12–16). The role of M&S in providing quantitative insight for COVID-19 spread in the general  
122 population was called into question due to a failure to predict early (circa 2020) outbreak dynamics  
123 (12). Nonetheless, the model predictions strongly influenced decision makers due to the ability of  
124 M&S practitioners to quickly generate results with a perceived to be high degree of precision  
125 superior to available observational statistical analyses. In retrospect, although results were  
126 computationally precise, they exhibited lower accuracy than initially anticipated. Model  
127 transparency, which includes explicit documentation of model choices, assumptions, the steps in  
128 the modeling process, and the expectations for the outputs, provides a reasonable defense  
129 against the propagation of misinformation and misunderstanding, such as what occurred during  
130 the pandemic (14),(15). In several instances during the pandemic, a model developed for  
131 population level COVID-19 spread in a large geographical region was applied to a less

132 appropriately applicable region without significant tuning and modification to account for  
133 population-specific demographic, clinical, epidemiological, and other influencing factors (17).  
134 Such activities during the early part of the COVID-19 pandemic illustrate how the lack of  
135 transparency and independent evaluation reduces the utility of models to inform critical decisions  
136 (16). Informing the user of the M&S results of model context and its intended use, such as through  
137 the assessment of its conformance to the “Ten Rules in Healthcare” will greatly minimize the  
138 negative impacts on model utility at all levels of application (7). Such an assessment would not  
139 just improve transparency but would enable communication of credible practice of M&S in a  
140 comprehensive manner (Table 1; (7)).

141  
142 The remainder of the manuscript is organized as follows: First, we describe our process  
143 for developing and assessing the rubric. We then present the rubric and an explanation of its  
144 components. Next, we illustrate the application of the rubric in multiple use cases to evaluate its  
145 utility in assessing the conformance to the Ten Rules in a consistent manner across multiple  
146 reviewers and M&S studies. Finally, we discuss best practices for applying the rubric and possible  
147 future extensions.

148

## 149 **Materials and Methods**

### 150 **Development of the rubric framework**

151

152 Our proposed rubric assesses and communicates the extent of conformance to the Ten  
153 Rules for Credible Practice of M&S based on the capability of outreach to the biomedical and  
154 healthcare community (Table 2). The concept of using a rubric for communicating the credibility  
155 state of an M&S evolved from challenges in communicating the ten simple assessments at

156 different decision-making levels and in different contextual applications. Thus, the intent of the  
 157 rubric development, is to bring a concise communication tool to the M&S healthcare community.

158  
 159 To develop the rubric, the Committee considered requirements of outreach to a wide range  
 160 of stakeholders (Fig 1A), each of whom has their own distinct use cases and priorities in  
 161 evaluating an M&S model. For instance, M&S practitioners may want to conduct granular analysis  
 162 of their own M&S practices, while clinicians are primarily concerned with whether they can trust  
 163 M&S to inform a clinical practice decision.

164

165 **Table 2. rubric for assessing conformance to the Ten Rules.**

<b>Outreach Capability</b>	Outreach to application-domain experts who may not be M&S practitioners	Outreach to M&S practitioners who may not be application-domain experts	Outreach to application-domain specific M&S practitioners	Outreach to application-domain specific M&S practitioners	None or very limited
<b>Conformance Level</b>	Comprehensive	Extensive	Adequate	Partial	Insufficient
	4	3	2	1	0
<b>Description Level</b>	Can be understood by <u>non-M&amp;S practitioners</u> familiar with the application domain and the intended context of use	Can be understood by M&S practitioners <u>not</u> familiar with the application domain and the intended context of use	Can be understood by M&S practitioners familiar with application domain and the intended context of use	<u>Unclear</u> to the M&S practitioners familiar with the application domain and the intended context of use	Missing or grossly incomplete information to properly evaluate the conformance with the rule



167 **Fig 1. Assessing TR rubric conformance to the Ten Rules based on the extent of outreach**  
168 **to the stakeholder’s domain expertise.** (A) The range of stakeholders that may utilize the rubric.  
169 (B) Example distribution of the stakeholders in the stakeholder assessment chart. (C) The  
170 conformance levels to each of the Ten Rules based on the extent of outreach to stakeholders  
171 along the M&S expertise and domain familiarity axes.

172  
173 The Committee developed the rubric framework through an iterative approach. The initial  
174 framework had reviewers assess models qualitatively, ranking the conformance of a model to  
175 each of the Ten Rules as insufficient, partial, adequate, extensive, or comprehensive (Table 2).  
176 The qualitative assessment made it challenging to compare reviewer assessments and derive an  
177 overall rating for the model when there existed variability between the individual reviewers’  
178 assessments, as in the case of the first COVID-19 model (UPenn’s COVID-19 model) to which  
179 the Committee applied the rubric. Therefore, a second development of the Ten Rules rubric was  
180 implemented. In this development, a scoring system was included such that for each rule, the  
181 level of conformance is given a numerical score.

## 182 183 **Application of the rubric for different use cases**

184  
185 We applied the Ten Rules and the rubric to evaluate the M&S practices of several COVID-  
186 19 modeling studies with versions released early in the pandemic and available at the time of this  
187 study: MIT model (18), IHME model (19,20), CU model (21), NE model (22), ICL model (23),  
188 UPenn model (24). Model details can be found in Table 3. Independent reviews by persons with  
189 significant experience in M&S credibility assessment and with some familiarity of the application  
190 of M&S in supporting government and medical industry decision making are used to assess each  
191 model.

192

193 **Table 3. COVID-19 models tested for their conformance to the Ten Rules.**

<b>Location of Model Development</b>	<b>Model Description</b>	<b>Website</b>	<b>Dates Accessed</b>	<b>References</b>
Massachusetts Institute of Technology (MIT)	A novel epidemiological model for predicting detected cases and deaths in the pre-vaccination era of the COVID-19 pandemic	<a href="https://www.covidanalytics.io/">https://www.covidanalytics.io/</a>	June 2020 September 2020 February 2021	(18)
Institute for Health Metrics and Evaluation (IHME)	A model for predicting possible trajectories of COVID-19 infections and the effects of non-pharmaceutical interventions in the United States	<a href="https://www.healthdata.org/covid">https://www.healthdata.org/covid</a>	June 2020 September 2020 February 2021	(19,20)
Columbia University (CU)	A model to infer critical epidemiological characteristics associated with COVID-19	<a href="https://columbia.maps.arcgis.com/apps/webappviewer/index.html?id=ade6ba85450c4325a12a5b9c09ba796c">https://columbia.maps.arcgis.com/apps/webappviewer/index.html?id=ade6ba85450c4325a12a5b9c09ba796c</a>	June 2020 September 2020 February 2021	(21)
Northeastern University (NE)	A model used to study spatiotemporal COVID-19 spread	<a href="https://covid19.gleamproject.org/#model">https://covid19.gleamproject.org/#model</a>	June 2020 September 2020 February 2021	(22)
Imperial College London (ICL)	A model used to study the effect of non-pharmaceutical	<a href="https://www.imperial.ac.uk/mrc-global-infectious-">https://www.imperial.ac.uk/mrc-global-infectious-</a>	June 2020 September 2020 February 2021	(23)

	interventions in controlling the COVID-19 epidemic	<a href="#">disease-analysis/disease-areas/covid-19/covid-19-planning-tools/</a>		
University of Pennsylvania (UPenn)	COVID-19 Hospital Impact Model for Epidemics (CHIME), which was designed to assist hospitals and public health officials with understanding hospital capacity needs during the pandemic	<a href="https://penn-chime.phl.io/">https://penn-chime.phl.io/</a>	April 2020 June 2020 September 2020 February 2021	(24)

194

195           The selection process for the COVID-19 models we evaluated did not take into account  
 196 potential conformance to the Ten Rules, but rather addresses the availability of the model and  
 197 related information present at the time of the study. As none of the models explicitly followed our  
 198 recommended credibility practice, we did not seek to compare the COVID-19 models to determine  
 199 which one is the most conformant to the Ten Rules but rather to express our assessment based  
 200 on the information provided. Finally, this is not an endorsement or criticism of the M&S practices  
 201 utilized for the models tested for conformance to the Ten Rules. Instead, we seek to exemplify  
 202 how the rubric is to be employed when testing for M&S conformance and outreach. Additionally,  
 203 we note that no attempts were made to reproduce any of the results reported by these models,  
 204 nor was any attempt made to assess the scientific validity of the models, assumptions, or  
 205 limitations. Instead, we assessed the outreach capability of the models and were interested in  
 206 their representation and dissemination.

207

208 Two independent reviewers evaluated the conformance of the UPenn COVID-19 model  
209 (details can be found in Table 3) to the Ten Rules in April 2020. They used the initial, qualitative  
210 rubric. A separate independent reviewer (Reviewer 3) assessed the conformance of the remaining  
211 COVID-19 models to the Ten Rules using the numerical scoring version of the rubric. Reviewer  
212 3's original assessment was performed on June 24, 2020 and repeated twice: once on September  
213 7, 2020 and again on February 5, 2021.

214  
215 Finally, the rubric was applied in a clinically relevant M&S study to evaluate the utility of  
216 our approach in assessing the extent of credibility of M&S practices in these contexts of use. In  
217 the study, the authors developed a multi-scale, multi-organ model of hepatic metabolism. The  
218 authors performed a self-assessment of their model's conformance to the Ten Rules prior to their  
219 initial manuscript submission on this model and then reassessed their model during the  
220 manuscript revision process.

221

## 222 **RESULTS**

### 223 **A generalized rubric based on outreach capability**

224

225 The Committee recognized that the rubric needed to account for the different stakeholders  
226 who may be interested in utilizing a model. Assessment of a model's ability to communicate how  
227 and if it satisfied each of the Ten Rules would differ greatly depending on the stakeholder.  
228 Therefore, in the proposed rubric, the key stakeholder traits are distributed along two axes: their  
229 level of M&S expertise and their familiarity with the biological domain (Fig 1B). The stakeholder  
230 communities can have different mixtures of M&S expertise and domain familiarity. Individuals with  
231 expertise in M&S and the biological domain relevant to the context of use are positioned towards  
232 the upper right, while individuals with very little M&S expertise and domain knowledge are

233 positioned towards the lower left. The rubric assesses the conformance to each of the Ten Rules  
234 based on the extent of outreach to each group (Fig 1C). For a given rule, if the M&S practice was  
235 conducted at a level that is primarily accessible to only those with M&S expertise and domain  
236 familiarity, we deem this practice to be conformant to the rule at the Adequate level. If the M&S  
237 practice of a given rule is more broadly understood by individuals with M&S expertise without  
238 familiarity of the specific biological domain, we deem this practice to be conformant to the rule at  
239 the Extensive level. If the M&S practice of a given rule is understood by those familiar with the  
240 biological domain but do not have M&S expertise, we deem this practice to be conformant to the  
241 rule at the Comprehensive level. Lastly, the M&S practice that is unclear to the M&S practitioners  
242 with familiarity of the biological domain is considered as a Partial level of conformance, with  
243 missing information assessed as an Insufficient level.

244  
245 In this rubric, the model with the highest conformance level (Comprehensive) provides  
246 outreach to domain experts who may not be M&S practitioners while the lowest conformance level  
247 (Insufficient) does not provide sufficient outreach to any community level. Taken together, this  
248 rubric provides a generalized and graded approach to assess the conformance to the Ten Rules  
249 (Fig 1C). Table 2 shows a concise representation of the proposed conformance rubric to the Ten  
250 Rules of credible practice of M&S in healthcare. The extended rubric can be found in S1 File.

251  
252 The rubric does not assess the “correctness” (i.e. the validity or accuracy) of the  
253 computational models, but rather analyzes M&S credible practice conformance based largely on  
254 two dimensions: M&S experience and scientific domain expertise (Fig 1). We note that the rubric  
255 for conformance to the Ten Simple Rules is not an M&S practice accreditation process, but rather  
256 a communication tool for analyzing the robustness of the M&S practice employed for a  
257 computational model within specifically stated context of use.

258

## 259 **Assessment criteria for each rule**

260

261           Within each of the Ten Rules (i.e., guiding principles of M&S practice), we specified the  
262 detailed criteria to assess the level of conformance and outreach capability to all stakeholders  
263 across different application contexts (S1 Table). For instance, Rule #1 is to define the context of  
264 use clearly by developing and documenting the application, purpose, and intended uses of the  
265 model and simulation (Table 1). In our proposed rubric, an M&S practice conforms to this rule at  
266 the highest level (Comprehensive) if:

- 267           1. a summary of the context definition can be understood by non-M&S practitioners,
- 268           2. detailed explanation is understandable by experts from the application domain that may  
269           not be M&S practitioners, and
- 270           3. many relevant details are included in the documentation that enable adequate  
271           understanding by both application domain-specific and non-domain M&S experts.

272

273           The next levels of conformance are based on whether the context definition was  
274 communicated at a level that is a) understandable only by M&S experts, even if they are from  
275 outside of the application domain (Extensive); b) restricted to M&S experts with experience in the  
276 specific application domain (Adequate); or c) achieved only partially (Partial).

277

278           As another example, Rule #2 is to use contextually appropriate data by employing relevant  
279 and traceable information in the development or operation of a model or simulation (Table 1). In  
280 our proposed rubric, M&S practice conforms to this rule at the highest level (Comprehensive) if:

- 281           1. all the data used in M&S development and/or operation is traceable to its original source,  
282           and

283 2. application-domain experts that are not M&S practitioners can understand which and how  
284 the data was used.

285

286 The detailed criteria corresponding to all of the Ten Rules can be found in S1 Table.

287

## 288 Numerical scoring used in the rubric

289

290 A numerical scoring system was included in the rubric to quantify the assessments with  
291 each level of conformance. A conformance level of Insufficient is given a score of 0, while a  
292 conformance level of Comprehensive is given a score of 4.

293

294 After the reviewer has completed their assessment of the model's conformance to the  
295 rules, a total numeric score can be calculated, thereby allowing for a higher-level understanding  
296 of the model's conformance and providing a means of easily comparing assessments between  
297 reviewers. Assessment of a model which reaches an overall conformance level of Comprehensive  
298 will have a total score in the range [35, 40], while a model with Insufficient conformance will have  
299 a total score in the range [0, 5) (Table 5).

300

301 **Table 5. Numerical scoring system for assessing conformance to the Ten Rules.**

Conformance Level	Score for each Rule	Score Range for Averaging across Rule	Score Range for Summing all Ten Rules
Comprehensive	4	[3.5, 4]	[35, 40]
Extensive	3	[2.5, 3.5)	[25, 35)
Adequate	2	[1.5, 2.5)	[15, 25)
Partial	1	[0.5, 1.5)	[5, 15)
Insufficient	0	[0, 0.5)	[0, 5)

302

303 If a subset of the TR is not included in a model assessment, the overall scores associated  
304 with a given conformance level will need to be adjusted. For example, if two Rules are omitted in  
305 the assessment, a Comprehensive conforming model will then have a total score in the range [28,  
306 32], rather than [35, 40]. Similarly, Extensive will have a score in the range [20, 28)), Adequate  
307 will have a score in the range [12, 20)), Partial will have a score in the range [4, 12)), and  
308 Insufficient will have a score in the range [0, 4)).

309  
310 Numerical scoring also enables the calculation of statistics, such as averages and  
311 standard deviations, across multiple assessments for a single rule. When averaging scores  
312 across reviewers for a single rule, the score may not be a whole number, which we have  
313 accounted for in Table 5, which details the range of conformance scores for each rule.

314

## 315 **Recommended process for implementing rubric**

316

317 The recommended process for implementing the rubric throughout the M&S life cycle  
318 begins with clearly identifying the M&S intended context of use, including M&S domain of use,  
319 use capacity, and strength of influence (Fig 2) (7). Next, the conformance threshold must be  
320 established according to the rubric and TR. It is expected that throughout the M&S lifecycle there  
321 is to be further development and refinement of the model, thereby necessitating evaluation of the  
322 updated M&S per the Ten Rules and rubric thresholds. Following this assessment, there should  
323 be clear documentation and then implementation of the M&S. Additionally, when implementing  
324 the M&S activities, further reporting and documentation may be needed.

325

326 **Fig 2. Recommended process for implementing the TR rubric throughout the M&S**  
327 **lifecycle.**

328



## 329 **Illustrative application of the rubric to assess COVID-19 M&S** 330 **practice**

331  
332 We applied the Ten Rules and the rubric to evaluate the M&S practices of several COVID-  
333 19 modeling studies, as described in the Methods. Table 4 illustrates a summary of our two  
334 independent reviewer processes to evaluate the conformance to the Ten Rules of the UPenn  
335 COVID-19 model in April 2020. The complete assessment and conformance testing made by  
336 Reviewer 1 can be found in S2 File, and by Reviewer 2 in S3 File. Briefly, Reviewer 1 and 2  
337 disagreed on the conformance level of the UPenn COVID-19 model for five of the ten rules (Rule  
338 #1, #2, #4, #8, and #10). Despite this variability, the overall conformance of the model as tested  
339 using the numeric scoring system resulted in Reviewer 1's overall score of 21 and Reviewer 2's  
340 overall score of 20. Both reviewers agreed that the overall conformance of the model was  
341 Adequate in that the model can be understood by those with expertise in M&S and the biological  
342 domain. Thus, the scoring system facilitates comparisons between reviewers, assessments for  
343 each individual rule, and also a model's overall conformance. In the present rubric, the total  
344 assessment acts like an average, rather than a sum, of the individual rule assessments and is  
345 presented as representing an overall conformance using the same category scale as used for  
346 each individual rule.

347  
348 **Table 4. Independent reviewer assessment of the UPenn COVID-19 model's conformance**  
349 **to the Ten Rules.**

Rule	Conformance Level		Same Conformance across Reviewers?
	Reviewer 1	Reviewer 2	
1. Define context clearly	Comprehensive	Adequate	No
2. Use contextually appropriate data	Adequate	Partial	No
3. Evaluate within context	Insufficient	Insufficient	Yes
4. List limitations explicitly	Adequate	Partial	No

5.	Use version control	Extensive	Extensive	Yes
6.	Document appropriately	Extensive	Extensive	Yes
7.	Disseminate broadly	Comprehensive	Comprehensive	Yes
8.	Get independent reviews	Insufficient	Partial	No
9.	Test competing implementations	Partial	Partial	Yes
10.	Conform to standards	Adequate	Comprehensive	No

350

351 A separate independent reviewer (Reviewer 3) assessed the conformance of the  
352 remaining COVID-19 models to the Ten Rules. For each of these model assessments, Rule #3  
353 (Evaluate within context) and Rule #9 (Testing competing implementations) were not examined  
354 as these rules posed the greatest uncertainty and required in-depth knowledge of the model  
355 structure and development. Here, for simplicity, we discuss the results from the most recent  
356 assessment on February 5, 2021; however, extensive commentary from each assessment is  
357 documented in S4 File. The most recent assessment of the MIT COVID-19 model's conformance  
358 to the Ten Rules is exemplified in Table 6, and the reviewer's reasoning for each conformance  
359 score is highlighted in the Comments column of the table. Despite the reviewer-assessed overall  
360 model conformance level being Extensive, the total score was 18-19, which would point towards  
361 an overall conformance level of Adequate. The scoring system introduced in the Methods section  
362 provides reviewers with the ability to properly evaluate their complete assessment of model  
363 conformance to the Ten Rules in a more quantitative manner. Reviewer 3 followed the same  
364 protocol for assessing conformance of the remaining COVID-19 models to the Ten Rules. The  
365 models' conformances to the Ten Rules can be found in Table 7 and the detailed assessments  
366 can be found in S4 File. The overall conformance across the COVID-19 models assessed by  
367 Reviewer 3 were between Adequate and Extensive. All of the COVID-19 models have shown  
368 improvement according to the Ten Rules rubric following reassessment.

369

370 **Table 6. Reviewer 3’s assessment of the MIT COVID-19 model’s conformance to the Ten**  
 371 **Rules.**

Rule	Conformance Level	Conformance Level Score	Comments
1. Define context clearly	Extensive	3	- A critical tool for COVID-19 planning that charts the progression of the pandemic across the United States and the world.
2. Use contextually appropriate data	Extensive	3	- Country-level projections are modeled based on historical data to increase the accuracy of future predictions. - Data is aggregated over 100 published clinical studies and preprints released between December 2019 and March 2020. - Declaration of collaborators and partner institutions that provide data and insights to model development team.
4. List limitations explicitly	Adequate	2	- Differences between Johns Hopkins University map and MIT cases reported - Total counts only account for countries in which they have sufficient data and where the pandemic is currently active. - Limitations explicitly stated in the paper
5. Use version control	Extensive	3	- Extensive documentation on differences between versions and updates. - Model codes are available on Github: <a href="https://github.com/COVIDAnalytics/DELPHI">https://github.com/COVIDAnalytics/DELPHI</a>
6. Document appropriately	Extensive	3	- Model documentation contains the system of equations and rates. - Code is well documented and there are detailed instructions on how to reproduce the results.
7. Disseminate broadly	Extensive	3	- Model results regularly published with interactive graphics. - Results, data, models and simulations are openly available to the public and scientific community.
8. Get independent reviews	Insufficient/Partial	0-1	- Paper with scientific findings published in (18)
10. Conform to standards	Partial	1	- Codes are written in Python and Mathematica and data is provided in comma-separated variables (CSV) format.
<b>Overall Conformance</b>	<b>Extensive</b>	<b>18-19</b>	- The epidemiologically based model is disseminated broadly and provides links to

			<p>model descriptions and data sources from their project website.</p> <p>- The code utilized in the research is accessible to the public via a GitHub repository and the model limitations are adequately described.</p>
--	--	--	---

372

373 **Table 7. Reviewer 3 assessment of COVID-19 model conformances to the Ten Rules.**

Rule	Conformance Level (Score)				
	IHME Model	CU Model	NU Model	ICL Model	Average Conformance across models
1. Define context clearly	Extensive (3)	Extensive (3)	Extensive (3)	Adequate/ Extensive (2-3)	Extensive (2.75)
2. Use contextually appropriate data	Extensive (3)	Extensive (3)	Adequate/ Extensive (2-3)	Adequate/ Extensive (2-3)	Extensive (2.5)
4. List limitations explicitly	Partial (1)	Adequate (2)	Adequate (2)	Adequate (2)	Adequate (1.75)
5. Use version control	Adequate/ Extensive (2-3)	Adequate (2)	Insufficient (0)	Extensive (3)	Adequate (1.75)
6. Document appropriately	Adequate (2)	Adequate (2)	Extensive (3)	Extensive (3)	Extensive (2.5)
7. Disseminate broadly	Adequate/ Extensive (2-3)	Adequate/ Extensive (2-3)	Adequate/ Extensive (2-3)	Adequate/ Extensive (2-3)	Adequate (2)
8. Get independent reviews	Adequate/ Extensive (2-3)	Extensive (3)	Adequate (2)	Extensive (3)	Extensive (2.5)
10. Conform to standards	Partial (1)	Partial (1)	Insufficient (0)	Partial (1)	Partial (0.75)
<b>Overall Conformance</b>	<b>Adequate (16)</b>	<b>Adequate (18)</b>	<b>Adequate (14)</b>	<b>Adequate (18)</b>	<b>Adequate (16.5)</b>

374

375 Following Reviewer 3's assessment of the five COVID-19 model conformances to the Ten

376 Rules, we analyzed the results to identify the rules to which the models were least conformant. If

377 a model was between two conformance levels for a given rule, the lower conformance level was  
378 used in the comparative analysis. The average numerical conformance score across all models  
379 for each rule was assessed. Those with an average score lower than 1.5 meant that the  
380 conformance level was at most Partial. The only rule that met this requirement was Rule #10:  
381 Conform to standards. In all COVID-19 models assessed by Reviewer 3, the conformance levels  
382 of the models to Rule #10 was either Partial or Insufficient, meaning the M&S practice of  
383 conforming to standards was incompletely stated (Partial conformance score) or insufficient  
384 evidence was presented to support conformance to standards (Insufficient conformance score)  
385 as assessed by M&S practitioners familiar with the application domain and the intended context  
386 of use. In order for the COVID-19 models to increase their conformance levels to Rule #10 of the  
387 Ten Rules, the models should have been implemented using community standards and formats.  
388 The associated documentation should lay out the details on the standards including version  
389 numbers and any exceptions or deviations that influence the use of the model. For instance, the  
390 IHME COVID-19 model is written in the widely used Python programming language; however, the  
391 model has not yet been configured for use outside of the internal IHME infrastructure. The current  
392 Partial conformance to Rule 10 can be increased to Adequate and possibly Extensive if IHME  
393 provides sufficient evidence for following appropriate standards such as Python style guides and  
394 statistical modeling standards (e.g., The American Statistical Association's Ethical Guidelines for  
395 Statistical Practice).

396

## 397 **Illustrative application of the rubric to M&S of liver metabolism**

398

399 The rubric was applied to a clinically relevant M&S study by Verma et al. (25) in which the  
400 authors developed a multi-scale, multi-organ model of hepatic metabolism. The authors  
401 performed a self-assessment of their model's conformance to the Ten Rules prior to their initial

402 manuscript submission and then re-assessed their model during the manuscript revision process.

403 Below is a summary of the author's self-assessment included with the manuscript as a way to

404 illustrate the application of the rubric.

405

406 **Rule #1 (Define context clearly):** The authors provided a detailed description of the

407 model's context written using terminology familiar to non-M&S practitioners who are

408 knowledgeable about the application domain, so the authors described the model's conformance

409 to Rule #1 (Define context clearly) as Comprehensive (score = 4). Briefly, the authors described

410 that the primary goal of the model was to evaluate the role of neural signals in controlling the

411 metabolic functionality of the liver, particularly in regulating the glycogenolysis to maintain

412 appropriate responses to hormonal signals to meet the systemic glucose demands. The biological

413 domain, structures, spatial scales, and time scales are explicitly stated. Additionally, the authors

414 included an explanation of other uses for the model, which include exploration of the effect of

415 dietary intake and insulin resistance in promoting a hepatic steatosis-like phenotype in the context

416 of innervation, calcium signaling and central nervous system (CNS) activation.

417

418 **Rule #2 (Use contextually appropriate data):** The authors believed their model

419 conformed to an Extensive (score = 3) level for Rule #2 since the in vitro, ex vivo, in vivo preclinical

420 and human subject data used for model building and validation was confirmed to meet the detailed

421 data requirements for consistency and explicit description of data heritage.

422

423 **Rule #3 (Evaluate within context):** The authors' self-assessed conformance level was

424 Extensive (score = 3) since verification and validation of the model output was explicitly described

425 and performed by both the developer and a third-party lab member not involved in the study.

426 Additionally, the authors state that during the revision process, there was extensive validation

427 performed as the model was recalibrated based on experimental hepatic calcium dynamics and  
428 catecholamine secretion in humans during periods of increased exercise.

429

430 **Rule #4 (List limitations explicitly):** The model's conformance was considered to be  
431 Comprehensive (score = 4) as all limitations were explicitly provided. In addition, the limitations  
432 were detailed in a manner that is understandable by experts in the liver physiology and pathology  
433 domain, even if they are not M&S experts. An example limitation was that the model was  
434 parameterized the same for simulating human-like and rodent-like extents of innervation and only  
435 differed by the extent of innervation, which does not use M&S terminology but states the issue in  
436 biomedical terms. Note that in the study Verma et al. (25) did not explicitly state the quantitative  
437 levels of M&S prediction error arising from the explicitly stated limitations. Under the rubric, those  
438 details are not required. There just needs to be sufficient information for an individual to  
439 understand under which conditions a model should not be used.

440

441 **Rule #5 (Use version control):** The model's conformance was considered to be  
442 Extensive (score = 3) as the evolution of the model and the various versions are explicitly  
443 documented on GitHub. GitHub is a platform familiar to M&S practitioners but not necessarily to  
444 experts in the liver physiology and pathology domain. Hence, the conformance level was not  
445 considered Comprehensive (score = 4).

446

447 **Rule #6 (Document appropriately):** The model's conformance level is Extensive (score  
448 = 3) as comments were provided in the model code, the scope and intended use were described  
449 in the main text, and a user guide for M&S practitioners was described in the main text and  
450 supplemental files. During the revision process, the model alternative was explained in the main  
451 text and an additional supplemental figure was included to detail the results of the model  
452 alternative. The user guide was utilized by the independent reviewer (see Rule 8 below) with M&S

453 expertise but little domain familiarity, demonstrating the Extensive level of conformance to Rule  
454 6.

455  
456 **Rule #7 (Disseminate broadly):** The conformance level was considered as Extensive  
457 (score = 3) as the simulations, results and implications were described in the main text and the  
458 model files are present in the supplementary material and on GitHub. The software used for this  
459 M&S study (Matlab, XPP and a Matlab/XPP interface) are all publicly available either freely or for  
460 a fee. The links to these resources and code files were included in the manuscript, enabling  
461 potentially Extensive dissemination.

462  
463 **Rule #8 (Get independent reviews):** The self-assessed model conformance was  
464 Extensive (score = 3), as a member of the research group not involved in the study or field  
465 performed a review. We note that in order to minimize the bias in the assessment, an internal  
466 review, even by a member of the group not involved in the study, is more appropriately scored as  
467 a 2 (Adequate). An outside review (outside the primary research groups that conducted the study  
468 or even outside the study authors' institutions) could be considered as a 3 (Extensive), and a  
469 multi-person independent cross-institutional review, particularly by non-M&S practitioners, could  
470 be scored at 4 (Comprehensive).

471  
472 **Rule #9 (Test competing implementations):** The conformance level only reached a  
473 conformance level of Adequate (score = 2) as competing implementations were tested and  
474 compared by the first three authors of the paper during the initial manuscript preparation.  
475 Furthermore, the competing implementations could only be understood by M&S practitioners  
476 familiar with the application domain and the intended context of use, thus justifying the Adequate  
477 conformance level. During the manuscript revision stage, the model was further revised, leading



478 to its refinement and improvement. The main text was also updated to include the changes made  
479 to the model during revision.

480

481 **Rule #10 (Conform to standards):** The model's conformance was considered Adequate  
482 (score = 2) as the model was implemented and simulated in a widely used platform for multiscale  
483 modeling (Matlab, in this case), along with another freely available and popular software, XPP.  
484 Following best coding practices described in Wilson et al. (26) and Hunter-Zinck et al. (27), the  
485 model code is commented at critical locations to aid the reader as well. Although the model was  
486 documented and disseminated using publicly available online platforms such as GitHub and open  
487 access manuscript supplementary material in conformance with rule #7, there was limited  
488 evidence of following the operating procedures, guidelines and standards as described in the  
489 credible practice of M&S in healthcare: ten rules from a multidisciplinary perspective (7).

490

491 The complete self-assessment for this model is included as a supplement to this  
492 manuscript (S5 File).

493

494 The computational modeling and simulation study of hepatic metabolism has an overall  
495 conformance level of Extensive (total numeric score = 30). Therefore, the overall practice of M&S  
496 for this biological scenario can be understood by M&S practitioners not familiar with the  
497 application domain and intended context of use. For this example, M&S practice to reach a  
498 Comprehensive level of conformance to the TR, the authors would need to incorporate additional  
499 features into the study. For example, a detailed step-by-step user's and developer's guide such  
500 that a non-M&S practitioner can replicate the M&S results would improve the score corresponding  
501 to Rule #6. Additionally, the authors could follow a stricter set of operating procedures and  
502 guidelines such that the M&S study appropriately conforms to modeling standards in  
503 representation, software code and documentation (Rule #10). Lastly, the authors could more

504 comprehensively test and formally document competing implementations of their model for  
505 improving the score on Rule #9.

506

## 507 **Discussion**

508

509 We have described a rubric that specifies detailed criteria for assessing the level of  
510 conformance to the Ten Rules for Credible Practice of M&S in Healthcare. The rubric is based on  
511 the outreach capability of an M&S practice across a wide range of stakeholder communities  
512 including program leadership, healthcare providers, policy makers and clinical practitioners. The  
513 rubric establishes a generalized and graded approach to assess levels of conformance from  
514 lowest (Insufficient) to highest (Comprehensive). We have illustrated the application of this rubric  
515 in two contexts of use including COVID-19 studies and a liver metabolism model. In the context  
516 of assessing COVID-19 studies, we evaluated the consistency of applying the rubric across  
517 multiple reviewers. We proposed a scoring scheme that provides a consistent process for M&S  
518 assessments and identification of critical credibility conformance gaps across a range of  
519 reviewers' familiarity levels. The Ten Rules augmented with the rubric aims to provide a  
520 generalized approach for the development and evaluation of the credible practice of M&S in  
521 translational and fundamental research endeavors aimed at *in silico* support of systems medicine  
522 efforts.

523

524 Assessing the outreach capability of an M&S study is useful for those within and outside  
525 of a specific scientific discipline. It enables clear communication and application across various  
526 stakeholder groups. For example, through the use of the Ten Rules and TR rubric, those working  
527 in an industrial setting can easily understand and implement the M&S practices undertaken by  
528 the academic research community. Additionally, these parties can communicate to policy makers

529 and higher-level stakeholders that can take action and employ a new development of the Ten  
530 Rules and TR rubric to suit their needs. The continuous evolution of the rubric as seen with the  
531 implementation of a numerical scoring system for conformance illustrates a framework that is  
532 driven by refinement and improvement by the healthcare community.

533  
534 The TR rubric was introduced to expand the reviewer's utilization of the Ten Rules.  
535 Specifically, the introduction of the rubric concept is aimed at expanding the focus of the reviewer  
536 from solely evaluating a model based on its validity and accuracy, to including the assessments  
537 of how supporting information regarding the M&S credibility engages the community beyond those  
538 who are familiar with M&S and the context of use. It is important to note, however, that not every  
539 M&S needs to meet a score of Comprehensive to be acceptable. For example, for a  
540 Comprehensive conformance level, the outreach is to non-M&S practitioners familiar with the  
541 application, while a conformance level of Extensive can be understood by M&S practitioners not  
542 familiar with the domain and context of use. Therefore, depending on the use of the model, an  
543 Extensive conformance level may be more appropriate than a Comprehensive conformance level.

544  
545 Assessment of the five discussed COVID-19 model conformances to the Ten Rules shows  
546 the value of utilizing such a rubric that prioritizes outreach capability. Specifically, it shows the  
547 Ten Rules and TR rubric can establish a cumulative assessment of the TR that has improved  
548 consistency in evaluation at each competency level, which was a critical need for decision making  
549 support as illustrated by the application to COVID-19 models. Multiple reviewers assessed the  
550 conformance of the UPenn COVID-19 model to the Ten Rules. There were notable differences in  
551 how the reviewers viewed the supporting credibility evidence, which illustrated that the reviewer's  
552 experience level, and their understanding of the context of use relative to the models' intended  
553 use, can influence the evaluation. This influence appeared to be nearly orthogonal to the

554 underlying credibility factor domains, leading to the investigation into a more granular and defined  
555 TR rubric.

556

557         Following the updated rubric application, the consistency of findings between reviewers of  
558 similar backgrounds was improved, especially if we consider the consolidated or summed  
559 conformance scoring where both reviewers' scores correspond to an overall model conformance  
560 level of Adequate.

561

562         The assumption that each rule's contribution is equally weighted with respect to the global  
563 conformance introduces a limitation in the assessment scheme. For instance, it is possible to  
564 accumulate an overall score in the Adequate or Extensive range and still have conformance to  
565 one or more individual TR be characterized as Insufficient. This suggests a comprehensive  
566 reporting that is more representative of the individual scores may be necessary to communicate  
567 the complete M&S credibility outreach picture. An option is to use tailored decision ranking tools  
568 such as pairwise comparison and analytical hierarchical processes (AHP) to capture specific  
569 community best practice principles by effectively weighting the individual credibility rules.  
570 Although the pairwise and AHP approaches may provide domain specific consistency, it is a  
571 recommended best practice to provide the set of conformance scores for individual rules as well  
572 as the global conformance score when delivering these assessments to decision makers in order  
573 to ensure appropriate communications levels. In this case, the rubric assessed 10 rules that can  
574 be grouped into representations tailored for the technical or decision-making community. A  
575 proposed method is illustrated in Table 8 representing a grouping of the Ten Rules to derive  
576 categorical scores for Development, Application and Supporting Evidence aspects for use in  
577 regulatory applications.

578

579 **Table 8. Example of categorizing the TR and rubric assessments to support regulatory**  
 580 **applications according to Reviewer 1’s assessment of the UPenn COVID-19 model’s**  
 581 **conformance to the Ten Rules.**

Development		Application		Supporting Evidence		Overall
Rule 1	4	Rule 2	2	Rule 6	3	Summary Score 2.1
Rule 3	0	Rule 4	2	Rule 7	4	
Rule 5	3			Rule 8	0	
Rule 10	2			Rule 9	1	
Development Score 2.25		Application Score 2		Supporting Evidence Score 2		

582

583           There is additional need and opportunity for streamlining the assessment of M&S activities

584 using the Ten Rules and TR rubric in addition to other associated frameworks. For instance, it

585 may take a significant amount of time to perform the assessment manually. Therefore, automating

586 components of the assessment may provide a capability of assessing the M&S results and

587 associated literature in an unbiased manner. This would be a boon to many communities of

588 practice, especially the healthcare community. A more systematic approach could be taken such

589 that the wider scientific community and stakeholders of the Ten Rules and TR rubric can be

590 included. The Interagency Modeling and Analysis Group (IMAG) and the Multiscale Modeling

591 (MSM) consortium are examples of two groups with significant roles in formulating and developing

592 the Ten Rules and TR rubric. As they both serve a joint purpose of serving the scientific

593 community and providing a collaborative platform to outline good practice of simulation-based

594 medicine, it may be possible to look to their leadership and guidance in systematizing and

595 automating unbiased assessment processes (8).

596

597           While a community effort is valuable to progress and implement the ideologies of the Ten

598 Rules and TR rubric, a specific set of guidelines must be established to ensure proper

599 employment. An excellent example of successful first steps in this direction lies with The

600 Physiome scientific journal. The Physiome is an open access journal that, for a small fee, confirms  
601 the reproducibility and reusability of the models according to the Ten Rules. By adopting the Ten  
602 Rules and TR rubric for M&S credibility, journal curators established that published models  
603 generally only conform to an Adequate level of outreach. Implementing an additional guideline in  
604 which the model must meet an overall conformance level of Extensive for publication into the  
605 journal may promote the benefits of M&S outreach capability to the scientific community.

606  
607 Future implementations of the Ten Rules and TR rubric could consider how the credible  
608 practice for assessment of an M&S practice may be transferable from one context to another. For  
609 example, another context of use was noted for the model of liver metabolism. However, the  
610 reviewers did not assess the model in this alternate context. It is an open question as to how the  
611 assessments of the Ten Rules can be applied to the alternate contexts of use and under what  
612 conditions this can occur. It may be the case where the previous assessments of only some of  
613 the Rules can be transferred while others may be “non-transferable”. Additionally, M&S practices  
614 may be altered at different user levels. For instance, it may not be appropriate to use a model  
615 built on data from a local hospital system and apply the model at the national scale. Furthermore,  
616 the transferability issue has implications for assessing the conformance of ensemble models, or  
617 a single model that contains multiple diverse models, to the Ten Rules. One potential solution  
618 may be to provide reasoning for including each of the models into the greater ensemble model in  
619 the same way that a single equation within an ordinary differential equation (ODE)-based model  
620 would be explained.

621  
622 In an additional future implementation, we propose that the current rules and rubric can  
623 be adjusted to more explicitly account for patient-specific/digital twin models as they begin to be  
624 utilized in the clinical setting. An updated and extended set of rules and practices may be  
625 developed for assessing and ensuring the credibility of these models. The need for an updated

626 list of rules is essential in the personalized modeling realm as the current methods lack  
627 consistency and credibility, especially within the clinic. Additionally, the Ten Rules may not be  
628 adequate in assessing the complexity and detail required for digital twin modeling. The updated  
629 and extended rules for digital twin modeling can then be used as a guide during the developmental  
630 stages of model development to avoid the problems seen with current digital twin efforts as  
631 previously discussed. Future modeling efforts that are guided by future implementations of the  
632 rules may also establish more trust and interaction between the modeler and clinician, thereby  
633 bridging the gap that currently exists in translating computational models from research into the  
634 healthcare field.

635

636 The TR rubric is used to assess a model's conformance to the Ten Rules for credible  
637 practice in M&S in healthcare. It is highly recommended that the M&S activity in the healthcare  
638 domain reaches either a conformance level of Comprehensive or Extensive. Both conformance  
639 levels have their own intended outreach capability as Comprehensive models can be understood  
640 by **non**-M&S practitioners familiar with the application domain and the intended context of use  
641 while Extensive models can be understood by M&S practitioners **not** familiar with the application  
642 domain and the intended context of use. Thus, defining which group must use the M&S results to  
643 support their decision is of utmost importance. The outreach goal for a given model is to be as  
644 clear and comprehensible to as broad an audience as possible such that the model can be widely  
645 adopted.

646

647 In conclusion, we formulated a rubric that promotes consistent and continuous evolution  
648 and testing of M&S practices such that one can reach the appropriate outreach level. In addition  
649 to the evolution of individual models, the TR rubric may evolve to meet the needs of its users as  
650 one continues to test its conformance to the Ten Rules. The development of the TR rubric has

651 initiated a large community effort to assess the outreach, reproducibility, replicability, and  
652 credibility of M&S studies in the scientific healthcare domain.

653

## 654 **Acknowledgments**

655 The authors would like to acknowledge the Interagency Modeling and Analysis Group and the  
656 Multiscale Modeling Consortium, who enabled activities of the Committee on Credible Practice  
657 of Modeling & Simulation in Healthcare. We would also like to thank the individual contributions  
658 of the Committee members who continue to help advance the efforts of the Committee but were  
659 not able to contribute to this manuscript.



## 660 References

- 661 1. VPH Institute | Virtual Physiological Human - International non-profit organisation  
662 [Internet]. [cited 2023 Feb 14]. Available from: <https://www.vph-institute.org/>
- 663 2. European Virtual Human Twin [Internet]. [cited 2023 Apr 21]. Available from:  
664 <https://www.edith-csa.eu/>
- 665 3. National Research Council of National Academies. Assessing the Reliability of Complex  
666 Models. National Research Council. 2012;
- 667 4. National Aeronautics and Space Administration. Standard for Models and Simulations.  
668 NASA. 2016;Report No.: NASA-STD-7009.
- 669 5. Pathmanathan P, Aycock K, Badal A, Bighamian R, Bodner J, Craven BA, et al.  
670 Credibility assessment of in silico clinical trials for medical devices. PLoS Comput Biol.  
671 2024 Aug 8;20(8):e1012289.
- 672 6. Tatka LT, Smith LP, Hellerstein JL, Sauro HM. Adapting modeling and simulation  
673 credibility standards to computational systems biology. J Transl Med. 2023 Jul  
674 26;21(1):501.
- 675 7. Erdemir A, Mulugeta L, Ku JP, Drach A, Horner M, Morrison TM, et al. Credible practice  
676 of modeling and simulation in healthcare: ten rules from a multidisciplinary perspective. J  
677 Transl Med. 2020 Sep 29;18(1):369.
- 678 8. Mulugeta L, Erdemir A. Committee on credible practice of modeling and simulation in  
679 healthcare. ASME 2013 Conference on Frontiers in Medical Devices: Applications of  
680 Computer Modeling and Simulation. ASME; 2013. p. V001T10A015.
- 681 9. Erdemir A, Guess TM, Halloran JP, Modenese L, Reinbolt JA, Thelen DG, et al.  
682 Commentary on the integration of model sharing and reproducibility analysis to scholarly  
683 publishing workflow in computational biomechanics. IEEE Trans Biomed Eng. 2016  
684 Oct;63(10):2080–5.
- 685 10. Assessing the credibility of computational modeling through verification and validation:  
686 application to medical devices. New York: American Society of Mechanical Engineers;  
687 2018.
- 688 11. Strategic Plan for Regulatory Science | FDA [Internet]. [cited 2023 Oct 16]. Available  
689 from: [https://cacmap.fda.gov/science-research/advancing-regulatory-science/strategic-](https://cacmap.fda.gov/science-research/advancing-regulatory-science/strategic-plan-regulatory-science)  
690 [plan-regulatory-science](https://cacmap.fda.gov/science-research/advancing-regulatory-science/strategic-plan-regulatory-science)
- 691 12. Kuhl E. Data-driven modeling of COVID-19-Lessons learned. Extreme Mechanics  
692 Letters. 2020 Oct;40:100921.
- 693 13. Panovska-Griffiths J. Can mathematical modelling solve the current Covid-19 crisis?  
694 BMC Public Health. 2020 Apr 24;20(1):551.
- 695 14. Barton CM, Alberti M, Ames D, Atkinson J-A, Bales J, Burke E, et al. Call for  
696 transparency of COVID-19 models. Science. 2020 May 1;368(6490):482–3.

- 697 15. Bunker D. Who do you trust? The digital destruction of shared situational awareness and  
698 the COVID-19 infodemic. *Int J Inf Manage.* 2020 Dec;55:102201.
- 699 16. Reps JM, Kim C, Williams RD, Markus AF, Yang C, Duarte-Salles T, et al.  
700 Implementation of the COVID-19 Vulnerability Index Across an International Network of  
701 Health Care Data Sets: Collaborative External Validation Study. *JMIR Med Inform.* 2021  
702 Apr 5;9(4):e21547.
- 703 17. Newton AJH, Chartash D, Kleinstein SH, McDougal RA. A pipeline for the retrieval and  
704 extraction of domain-specific information with application to COVID-19 immune  
705 signatures. *BMC Bioinformatics.* 2023 Jul 20;24(1):292.
- 706 18. Li ML, Bouardi HT, Lami OS, Trikalinos TA, Trichakis N, Bertsimas D. Forecasting  
707 COVID-19 and Analyzing the Effect of Government Interventions. *Oper Res.* 2022 Jun  
708 10;
- 709 19. IHME COVID-19 health service utilization forecasting team, Murray CJ. Forecasting the  
710 impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for  
711 the USA and European Economic Area countries. *medRxiv.* 2020 Apr 26;
- 712 20. IHME COVID-19 Forecasting Team. Modeling COVID-19 scenarios for the United  
713 States. *Nat Med.* 2021 Jan;27(1):94–105.
- 714 21. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented  
715 infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science.*  
716 2020 May 1;368(6490):489–93.
- 717 22. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of  
718 travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak.  
719 *Science.* 2020 Apr 24;368(6489):395–400.
- 720 23. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating  
721 the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.* 2020  
722 Aug;584(7820):257–61.
- 723 24. Weissman GE, Crane-Droesch A, Chivers C, Luong T, Hanish A, Levy MZ, et al. Locally  
724 Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic.  
725 *Ann Intern Med.* 2020 Jul 7;173(1):21–8.
- 726 25. Verma A, Manchel A, Narayanan R, Hoek JB, Ogunnaike BA, Vadigepalli R. A spatial  
727 model of hepatic calcium signaling and glucose metabolism under autonomic control  
728 reveals functional consequences of varying liver innervation patterns across species.  
729 *Front Physiol.* 2021 Nov 26;12:748962.
- 730 26. Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, et al. Best practices  
731 for scientific computing. *PLoS Biol.* 2014 Jan 7;12(1):e1001745.
- 732 27. Hunter-Zinck H, de Siqueira AF, Vásquez VN, Barnes R, Martinez CC. Ten simple rules  
733 on writing clean and reliable open-source scientific software. *PLoS Comput Biol.* 2021  
734 Nov 11;17(11):e1009481.

735

## 736 **Supporting Information**

737 **S1 File.** The extended Ten Simple Rules rubric with detailed criteria.

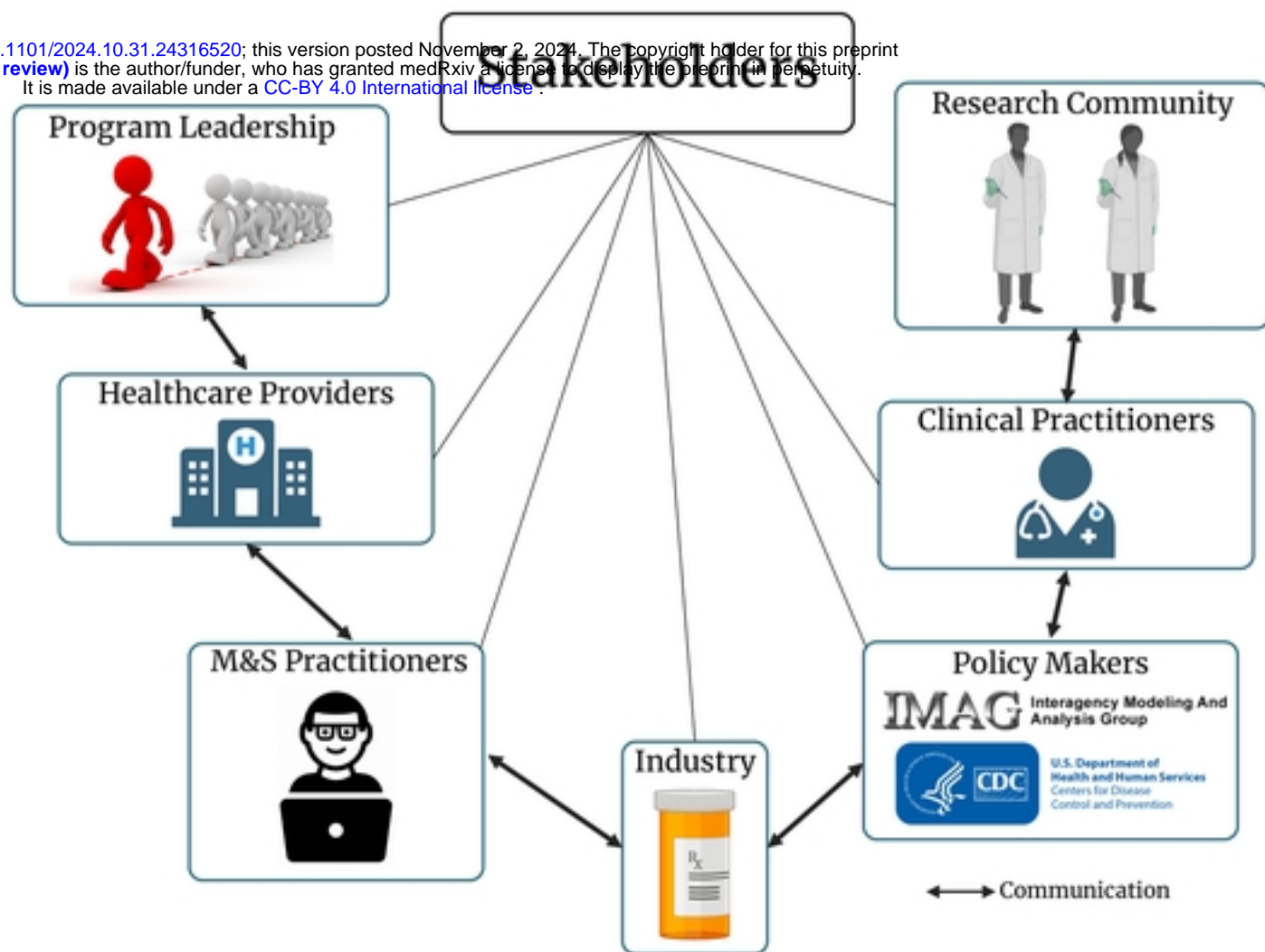
738 **S2 File.** Reviewer 1's complete assessment and conformance testing to the Ten Simple Rules  
739 using the rubric applied to the COVID-19 modeling studies.

740 **S3 File.** Reviewer 2's complete assessment and conformance testing to the Ten Simple Rules  
741 using the rubric applied to the COVID-19 modeling studies.

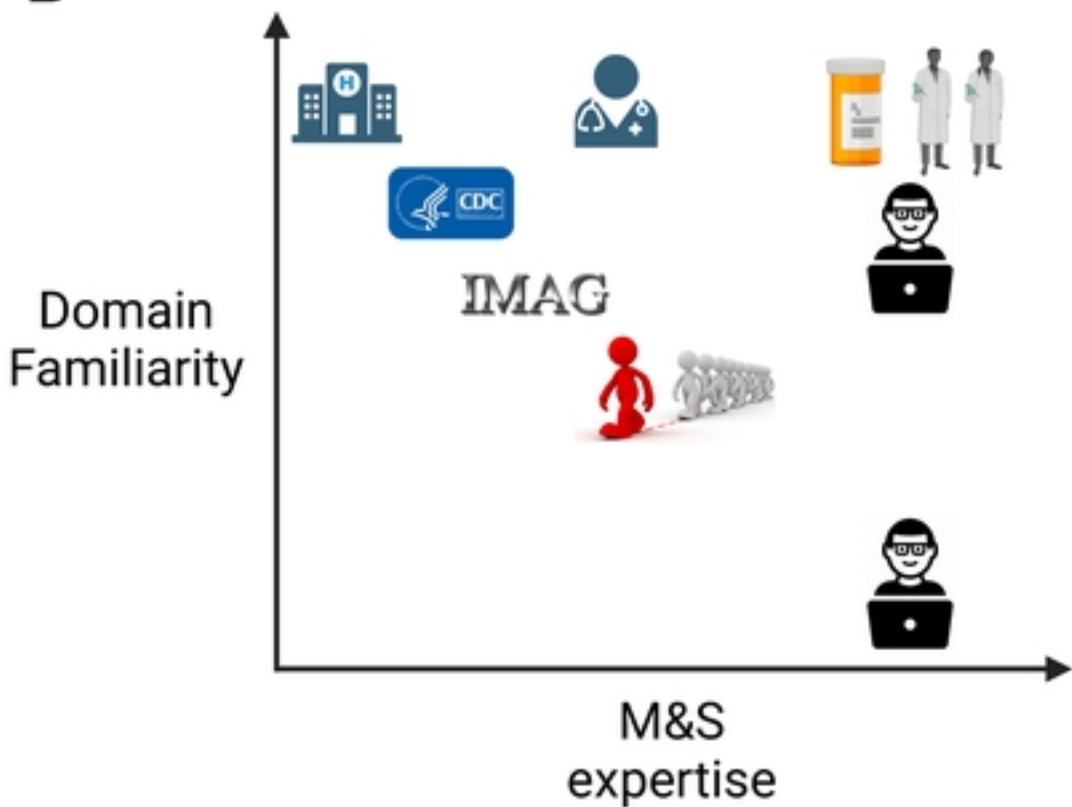
742 **S4 File.** Reviewer 3's complete assessment and conformance testing to the Ten Simple Rules  
743 using the rubric applied to the COVID-19 modeling studies.

744 **S5 File.** The complete self-assessment and conformance testing to the Ten Simple Rules using  
745 the rubric applied to the Verma et al. (25) multi-scale, multi-organ model of hepatic metabolism.

A



B



C

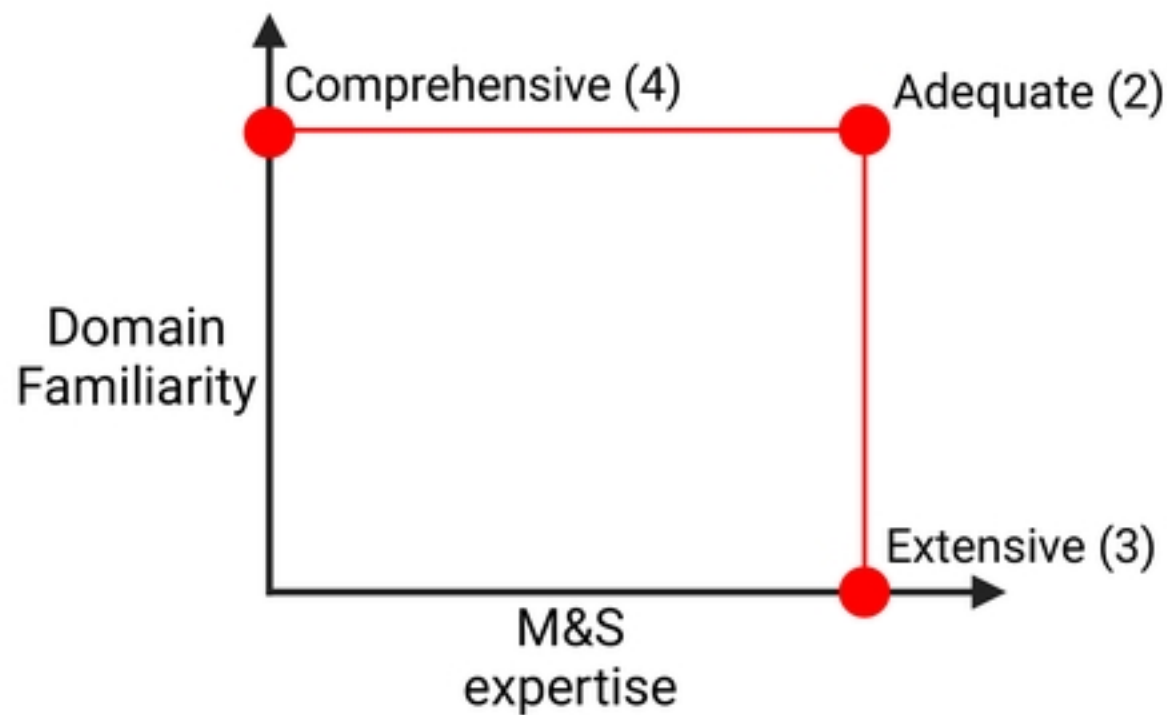


Figure 1

Define context of use clearly, including the M&S

- *Domain of Use*
- *Use Capacity*
- *Strength of Influence*

medRxiv preprint doi: <https://doi.org/10.1101/2024.10.31.24316520>; this version posted November 2, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Establish Conformance Thresholds Accordingly for the Ten Rules of Credible Practice

## M&S Practice Lifecycle

Develop & Refine M&S

Implement M&S

Evaluate M&S Practice per Ten Rules & Rubric Thresholds

Report / Document

Figure 2