

Author's Proof

Before checking your proof, please see the instructions below.

- Carefully read the entire proof and mark all corrections in the appropriate place, using the Adobe Reader commenting tools ([Adobe Help](#)).
- Provide your corrections in a single PDF file or post your comments in the Production forum making sure to reference the relevant query/line number. Upload or post all your corrections directly in the Production Forum to avoid any comments being missed.
- We do not accept corrections in the form of edited manuscripts nor via email.
- Do not provide scanned or handwritten corrections.
- Before you submit your corrections, please make sure that you have checked your proof carefully as once you approve it, you won't be able to make any further corrections.
- To ensure the timely publication of your article, please submit the corrections within 48 hours. After submitting, do not email or query asking for confirmation of receipt.

Do you need help? Visit our [Production Help Center](#) for more information. If you can't find an answer to your question, contact your Production team directly by posting in the Production Forum.

Quick Check-List

- Author names** - Complete, accurate and consistent with your previous publications
- Affiliations** - Complete and accurate. Follow this style when applicable: Department, Institute, University, City, Country
- Tables** - Make sure our formatting style did not change the meaning/alignment of your Tables.
- Figures** - Make sure we are using the latest versions.
- Funding and Acknowledgments** - List all relevant funders and acknowledgments.
- Conflict of Interest** - Ensure any relevant conflicts are declared.
- Supplementary files** - Ensure the latest files are published and that no line numbers and tracked changes are visible.
Also, the supplementary files should be cited in the article body text.
- Queries** - Reply to all typesetters queries below
- Content** - Read all content carefully and ensure any necessary corrections are made.

Author Queries Form

Query No.	Details Required	Author's Response
Q1	Confirm that the short running title is correct, making sure to keep it to a maximum of five words.	
Q2	Confirm whether the insertion of the article title is correct.	
Q3	We have noticed a discrepancy between the author list in the submission system and in the accepted manuscript (for authors Eric Forgoston, Rahuman S. Malik Sheriff, James Osborne, Gilberto Gonzalez-Parra, Eric Forgoston, Ruth Bowness, Yaling Liu, Robin Thompson, Winston Garira, Jacob Barhak, John Rice, Marcella Torres, Hana M. Dobrovolny, Tingting Tang, William Waites, James A. Glazier, James R. Faeder and Alexander Kulesza (order mismatch in metadata)). For any authorship changes, Frontiers requires that all authors complete and sign the 'Authorship Change' form which is available at this link . Upload the file as a "Related Article" file type together with your Author's Proof Corrections.	
Q4	There is a discrepancy between the styling of the author names in the submission system and the manuscript. We have used [James R. Faeder, Hana M. Dobrovolny,] instead of [James Faeder, Hana Maria Dobrovolny]. Please confirm that it is correct.	
Q5	The citation and surnames of all of the authors have been highlighted. Check that they are correct and consistent with the authors' previous publications, and correct if need be. Please note that this may affect the indexing of your article in repositories such as PubMed.	

Query No.	Details Required	Author's Response
Q6	Please ask the following authors to register with Frontiers (at https://www.frontiersin.org/Registration/Register.aspx) if they would like their LOOP profile to be linked to the final published version. Please ensure to provide us with the author profile link(s) (not email addresses) when submitting the proof corrections. Non-registered authors and authors with profiles set to private mode will have the default profile image displayed Eric Forgoston John Rice Tingting Tang William Waites James A. Glazier.	
Q7	Confirm that all author affiliations are correctly listed. Note that affiliations are listed sequentially as per journal style and requests for non-sequential listing will not be applied. Note that affiliations should reflect those at the time during which the work was undertaken.	
Q8	Confirm that the email address in your correspondence section is accurate. Please note that any changes to the corresponding authorship would require individual confirmation from all original and added/removed corresponding authors.	
Q9	Confirm that the keywords are correct and keep them to a maximum of eight and a minimum of five. (Note: a keyword can be comprised of one or more words.) Note that we have used the keywords provided at Submission. If this is not the latest version, please let us know.	
Q10	Check if the section headers (i.e., section leveling) were correctly captured.	
Q11	Provide the complete details for reference "Bansal et al., 2013".	
Q12	Congratulations on the acceptance of your manuscript. I would like to draw your attention to the final comments of Reviewer 1 (reproduced below), and we kindly ask that you consider addressing this after receiving the typeset proofs from our Production Office.	
Q13	The author contribution section is mandatory and a standard statement has been inserted. Please edit as needed to accurately reflect the contribution of each author.	
Q14	Ensure to add all grant numbers and funding information, as after publication this will no longer be possible. All funders should be credited and all grant numbers should be correctly included in this section.	
Q15	We have replaced the section head "Acknowledgments" with "Funding." Please confirm that this is correct.	
Q16	Confirm if the text included in the Conflict of Interest statement is correct.	
Q17	Cite the reference "Assaf et al., 2011, Bai et al., 2021, Khain et al., 2014, Wang et al., 2019, Stack et al., 2013" inside the text.	
Q18	Provide the volume number page range and Doi for the following reference. "Cresta et al., 2021, Sarma et al., 2016."	
Q19	Provide the city for "Fujimoto et al., 2017."	
Q20	Provide the journal title volume number and page range for the following reference. "Garfinkel and Leclerc, 2020."	
Q21	Provide the volume number and page range for the following reference. "Hart et al., 2020, Hernandez-Vargas and Velasco-Hernandez, 2020, Ke et al., 2020."	
Q22	Ensure that all the figures, tables and captions are correct, and that all figures are of the highest quality/resolution. Please note that Figures and Tables must be cited sequentially, as per section 2.2 of the author guidelines .	
Q23	TAppendix 1 has not been mentioned in the article. Please add a citation within the text.	
Q24	Check your Appendix and confirm that it is correct.	
Q25	If you decide to use previously published, copyrighted figures in your article, please keep in mind that it is your responsibility, as the author, to obtain the appropriate permissions and licenses and to follow any citation instructions requested by third-party rights holders. If obtaining the reproduction rights involves the payment of a fee, these charges are to be paid by the authors.	



Model Integration in Computational Biology: The Role of Reproducibility, Credibility and Utility

Eric **Forgoston**¹, Rahuman S. Malik **Sheriff**², James **Osborne**³, Gilberto **Gonzalez-Parra**⁴, Eric **Forgoston**⁵, Ruth **Bowness**⁶, Yaling **Liu**⁷, Robin **Thompson**⁸, Winston **Garira**⁹, Jacob **Barhak**^{10*}, John **Rice**¹¹, Marcella **Torres**^{12*}, Hana M. **Dobrovoly**¹³, Tingting **Tang**¹⁴, William **Waites**¹⁵, James A. **Glazier**¹⁶, James R. **Faeder**¹⁷ and Alexander **Kulesza**¹⁸

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States, ²European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom, ³School of Mathematics and Statistics, University of Melbourne, VIC, Parkville, Australia, ⁴Mathematics Department, New Mexico Tech, Socorro, NM, United States, ⁵Department of Applied Mathematics and Statistics, Montclair State University, Montclair, NJ, United States, ⁶Department of Mathematical Sciences, University of Bath, Bath, United Kingdom, ⁷Department of Mechanical Engineering and Mechanics, Department of Bioengineering, Lehigh University, Bethlehem, PA, United States, ⁸Mathematics Institute and the Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Coventry, United Kingdom, ⁹Department of Mathematics and Applied Mathematics, Modelling Health and Environmental Linkages Research Group, University of Venda, Limpopo, South Africa, ¹⁰Jacob Barhak Analytics, Austin, TX, United States, ¹¹Independent Retired Working Group Volunteer, Virginia Beach, VA, United States, ¹²Department of Mathematics and Computer Science, University of Richmond, Richmond, CA, United States, ¹³Department of Physics and Astronomy, Texas Christian University, Fort Worth, TX, United States, ¹⁴Department of Mathematics and Statistics in San Diego State University (SDSU) and SDSU Imperial Valley, Calexico, CA, United States, ¹⁵Department of Computer and Information Sciences, Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, University of Strathclyde, Glasgow, Scotland, ¹⁶Biocomplexity Institute, Indiana University, Bloomington, IND, United States, ¹⁷Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, United States, ¹⁸Novadiscovery SA, Lyon, France

OPEN ACCESS

Edited by:

Shayn Peirce-Cottler,
University of Virginia, United States

Reviewed by:

Giulia Russo,
University of Catania, Italy
Marissa Renardy,
Applied BioMath, United States

*Correspondence:

Jacob Barhak
jacob.barhak@gmail.com
Marcella Torres
mtorres@richmond.edu

Specialty section:

This article was submitted to
Multiscale Mechanistic Modeling,
a section of the journal
Frontiers in Systems Biology

Received: 26 November 2021

Accepted: 04 February 2022

Published: XX XX 2022

Citation:

Forgoston E, Sheriff RSM, Osborne J, Gonzalez-Parra G, Forgoston E, Bowness R, Liu Y, Thompson R, Garira W, Barhak J, Rice J, Torres M, Dobrovoly HM, Tang T, Waites W, Glazier JA, Faeder JR and Kulesza A (2022) Model Integration in Computational Biology: The Role of Reproducibility, Credibility and Utility. *Front. Syst. Biol.* 2:822606. doi: 10.3389/fsysb.2022.822606

During the COVID-19 pandemic, mathematical modeling of disease transmission has become a cornerstone of key state decisions. To advance the state-of-the-art host viral modeling to handle future pandemics, many scientists working on related issues assembled to discuss the topics. These discussions exposed the reproducibility crisis that leads to inability to reuse and integrate models. This document summarizes these discussions, presents difficulties, and mentions existing efforts towards future solutions that will allow future model utility and integration. We argue that without addressing these challenges, scientists will have diminished ability to build, disseminate, and implement high-impact multi-scale modeling that is needed to understand the health crises we face.

Keywords: simulation, reproducibility, crisis, computational modeling, credibility

INTRODUCTION—THE PROMISE OF MODELING

Starting with the development of the SIR model in 1927 (Kermack et al., 1927), there has been a long history of population-level epidemiological modeling. The idea of being able to forecast biological phenomena computationally seemed very promising and as a result it has engaged researchers for a long time. Within time, many promises were made in the name of modeling, and enough effort was spent to warrant evaluating current capabilities, concepts, and hurdles.

Over the years, these studies have included the addition of population compartments, as well as the inclusion of age structure, vaccine, and quarantine, to name just a few advancements. Other studies have moved away from classical deterministic models to better understand the role of

115 stochasticity. Moreover, this wide range of epidemic models has
116 been used to study numerous diseases. Because of the vast
117 amount of knowledge that has been gained through epidemic
118 modeling over approximately the past 100 years, it was possible
119 for researchers to quickly adapt their models to make predictions
120 about the spread and control of the SARS-CoV-2 coronavirus.

121 However, unlike the population-level epidemic modeling
122 effort, much less is known about how viral infections spread
123 throughout the body, including its immune response and the
124 response of different organ systems. Moreover, very little is
125 known about the connection between infection at the
126 individual scale and infection at the population scale.

127 Some examples like (Azmy et al., 2018) and (Ackleh et al.,
128 2013) use bacteria or virus load as a feature parameter in partial
129 differential equation compartment systems, where the
130 progression of the disease among the population is linked to
131 the virus load in infectious individuals.

132 Within-host models can explore how the variance of biology
133 within the body can impact both a disease and its treatment
134 (Zarnitsyna et al., 2021). Such mathematical models of viruses
135 focusing on the in-host dynamics should be developed and
136 utilized to a greater extent.

137 Multi-scale within-host modeling is common, with scales from
138 the molecular and cellular levels integrated successfully with the
139 larger whole-organ or whole-body scales (Powathil et al., 2012;
140 Bowness et al., 2018), (Segovia-Juarez et al., 2004). There are
141 fewer models, however, that successfully combine within-host
142 models with population-level models. One barrier to the
143 development of such models is the potential to lose the
144 within-host granularity that is often seen when integrating to a
145 higher scale. Many researchers may question if these multi-scale
146 models truly offer new insight or if they are more informative
147 when analyzed independently. Such questions emphasize the
148 challenges of multi-scale model integration.

149 As we move from conceptual models (the diagrams and verbal
150 models of biologists) to mathematical models to computer code,
151 we gain in executability, but we lose shareability. Shareability
152 requires multiple concepts: Reusability—someone else can run the
153 same computer code, perhaps with different initial conditions or
154 parameter values; Extensibility—someone can add modules or
155 replace modules within the model without breaking it; Extractability—someone can select model components and use them, and they continue to function independently of their initial context; Portability—the model can be reused in a different computational instantiation from its original implementation. In particular, the knowledge embedded in computer code is generally stranded or lost, since you cannot easily infer the underlying conceptual model from the mathematical model or the mathematical model from its computer code. As a result, an essential aspect of model development is a formal process that begins with a detailed and complete specification of a fully sharable conceptual model, then develops a less sharable, but quantitative mathematical model, which is an interpretation of the biology and physics of the conceptual model, and finally a computer simulation, which implements the mathematical model in the form of specific algorithms and methodologies. At each step we need to define additional parameters and concepts.

172 One relatively new key concept enabling construction of
173 models of complex phenomena is composition (Halter et al.,
174 2020). Decomposition lets us break down a complex problem into
175 simpler problems that can be solved or simulated and
176 composition lets us systematically recombine these solutions
177 into a solution of the original problem. To accomplish this, we
178 need to think in terms of higher-order operations on models:
179 what the models are is less important than what can be done with
180 them. There are two kinds of composition: parallel and serial.
181 Parallel composition means running models concurrently. This is
182 useful for ensemble or consensus approaches that combine
183 multiple models to arrive at a best estimate. Serial
184 composition is when the output of one model becomes the
185 input of another. It is important to think about the type of the
186 model, what input it requires and what output it produces
187 because compatibility is required for serial composition. Serial
188 composition has been used to great effect, for example in whole
189 cell models. Of course, it is possible to compose these combined
190 models.

191 Combination of models can be from white box models or
192 black box models. Black box models do not expose much
193 information to the modeler, but even models that are
194 composed of white-box models may suffer from transparency
195 shortcomings due to the composition. In modular models, three
196 different types of submodel couplings can be found: 1) black-box
197 models with code-level coupling using information-hiding
198 interfaces, 2) white-box models with code-level coupling and
199 3) white-box models with biological-level coupling.

200 Compositionality and modeling has been extensively studied
201 theoretically and the primitive operations, parallel and serial
202 composition, explored in detail for certain classes of model
203 (Baez and Master, 2020; Baez et al., 2021) that appear in fields
204 as diverse as electronics engineering, chemistry, molecular
205 biology (Danos and Laneve, 2004), plant biology (Honorato-
206 Zimmer et al., 2018), infectious diseases (Halter et al., 2020;
207 Waites et al., 2021) and economic game theory (Atkey et al.,
208 2020). However, to address practical problems across scales,
209 infrastructure is required. First, it is necessary to be able to
210 discover models; models cannot be composed if they are
211 unknown or unavailable. To do this, a catalogue is needed
212 with metadata about models and how to obtain them. We
213 note the existence of mature software for data catalogues that
214 is easily repurposed (CKAN, 2021). The models must be
215 described sufficiently well to know if they can be composed,
216 annotated with information about their input and output types.
217 Annotations facilitate auxiliary tasks such as searching for
218 appropriate models and ascertaining provenance. Finally,
219 attention is needed to the detail of composition of a broad
220 class of models, recognizing that errors introduced by (de)
221 composition are only well-understood for some cases (Hairer
222 et al., 2006; Blanes et al., 2008).

223 With so many multiscale modeling methods that are
224 seemingly disjoint and mutually exclusive, recent efforts have
225 sought to bring some order in the discussion of multiscale models
226 of pandemics by providing a complete categorization of them
227 (Garira, 2017; Garira, 2018). These publications identified five
228 different categories of multiscale models of diseases that use

different integration frameworks to integrate across scales. While this categorization cannot be claimed to be unique, it constitutes a good starting point, which may be found useful as a basis for further refinement in the discourse of multiscale modeling of pandemics. The paper (Garira, 2018) further identified ten of the most significant challenges that stand in the way of future advances in integration across scales in the development of multiscale modeling of disease dynamics. Collaborative research among scientists with different skills is needed to fully resolve these challenges.

The recent COVID-19 pandemic has highlighted the importance of modeling the disease and its potential vulnerabilities for interventions. Notable examples include (Russo et al., 2020) and (Russo et al., 2021) that develop models for vaccines. Between-host infection simulations helped researchers to make forecasts about the spread and potential control of the coronavirus. These models build from a century of work in population-level epidemiological modeling. While simple models have been attempting to assess potential antiviral drug combinations, very little is still known about the connection between infection at the individual scale and infection at the population scale (Dodds et al., 2021).

Indeed during the COVID-19 pandemic, many scientists working on related issues assembled under the umbrella of the Multiscale Modeling and Viral Pandemics Working Group (Multiscale Modeling and Viral Pandemics, 2021). This group is part of the MultiScale Modeling (MSM) Consortium hosted by the Interagency Modeling and Analysis Group (IMAG) (Interagency Modeling and Analysis Group, 2021). The discussions about modeling and its promise collected many ideas that were represented there, many times representing more than one opinion, creating a choir of voices. Among those voices, the group located many hurdles that might explain why the promise of modeling is not yet fully fulfilled—those are discussed below and organized considering the workflow of reproducing, gaining credibility, reuse, and integration.

FROM REPRODUCIBILITY TO CREDIBILITY TO UTILITY TO INTEGRATION

Multi-scale models are intrinsically complex, and usually are modular, whereby the model is divided into units that interact with each other. Modularity facilitates component reuse and model integration, including the ability to exchange modules during, or between, simulations has many advantages (Petersen et al., 2014), but it depends both upon the validity of each individual module, as well as the ability to connect modules, so that they inter-operate appropriately. The promise of multiscale, modular modeling is that researchers can build from each other, using prior published models and building blocks for new, more accurate or more impactful models. For this vision to be achieved, 1) models must be reproducible, so that researchers are assured the module will perform as expected, 2) models must be credible, so that researchers are confident that reusing a module will be useful and appropriate, 3) models must

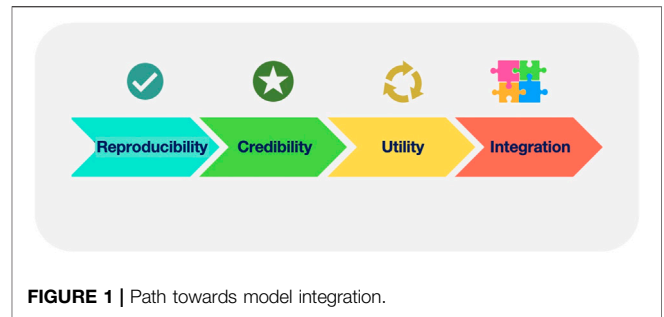


FIGURE 1 | Path towards model integration.

be reusable, meaning not only can they reproduce published results, but also that they can be modified to fit new contexts, and 4) researchers must be able to integrate models with other models. **Figure 1** shows these four steps schematically, including how each step depends on its predecessors.

The discussions in this group raised many issues that prevent model integration that start with inability to reproduce models, which leads to low credibility of those models, which reduces reuse, which leads to inability to combine and integrate larger more complex models. Therefore we address many issues at lower levels that will help reach integration.

The Reproducibility Crisis

Computational biomedical modeling involves mathematical representation of biological processes to study complex system behaviour and was expected to be less affected by the reproducibility crisis. Afterall, computer software should be deterministic and therefore repeatable if designed well, compared to biological processes that have a random nature where experiments are not guaranteed to repeat themselves and many repetitions are required for the mean to converge.

However, models often fail to be reproducible and the reasons for the failure and prevalence are not fully understood. In a recent study (Tiwari et al., 2021), the BioModels group analyzed 455 kinetic models published in 152 peer-reviewed journals, a collective work of about 1,400 scientists from 49 countries. Most of these models were manually encoded from scratch to assess the reproducibility. Their investigation revealed that 49% of the models could not be reproduced using the information provided in the manuscripts. With further effort, they managed to reproduce an additional 12% either by empirical correction or support from authors. The other 37% remained non-reproducible due to missing parameter values, missing initial concentrations, inconsistent model structure, or missing information. Among the corresponding authors they contacted less than 30% responded. Models from many life science journals failed to be reproducible, revealing a common problem in the peer-review process. The group proposed an 8-point reproducibility scorecard to assess each model and address the reproducibility crisis. A similar study that reports similar deficiencies is reported in (Kirouac et al., 2019). The term crisis is not exaggerated and is well justified since the need to combine models together already exists and building blocks should be solid and match expectations. The “promise of modeling” should be fulfilled already, yet we find modelers consistently unable to reproduce basic steps—thus stagnating

343 instead of innovating, or even worse—backtracking progress. The
344 numbers quoted above cannot be ignored or claimed as being
345 normal—thus the term crisis is used.

346 This shows only part of the larger crisis. The ideal we would
347 like to reach is repeatability and reproducibility of models in
348 publications and repositories. Repeatability is the ability to repeat
349 the same experiment with the same system, and reproducibility is
350 the ability to repeat the same experiment by another scientist
351 without the same system. The goal here is to have all aspects of the
352 simulation pipeline (Biological Model, Mathematical Model,
353 Numerical Methods, Computational Implementation) be
354 auditable, i.e., they should be fully described (as appropriate
355 for each component) to enable results to be repeated and
356 reproduced. However, the modeling community is far from this
357 ideal. With the diagnosis of this problem, different groups
358 aim for collaborative efforts to set up best practices.

359 It is important to acknowledge that the situation is much
360 worse for other aspects of modeling. For any other aspect, it
361 would be hard to even conduct such a study to evaluate
362 reproducibility because the information often is not
363 systematically catalogued let alone shared in a common
364 format. Examples include how models were constructed,
365 calibrated, or validated. Compared to the software industry or
366 finite element modeling that have reached well established
367 methods of exchanging information and repositories, such as
368 file exchange formats/git, computational biological modeling has
369 a long way to go and in this paper we will try to address some
370 topics to be dealt with. If we cannot reproduce models, how can
371 those be considered credible by other modelers, stakeholders, or
372 even the public?

373 Credibility of Models

374 In a larger context, model reproducibility is strongly tied to model
375 credibility in light of a given purpose. A model without a purpose
376 is a mere exercise suitable for the classroom and therefore the
377 prerequisite for any realistic application is the design and testing
378 of the model with that specific purpose in mind (often called the
379 “Context of Use”, see below). Often, not only the model
380 developer, but also others will need to confirm or challenge
381 the credibility of a model and thus, a model built for a specific
382 purpose must be at least repeatable and it is highly recommended
383 it be reproducible. In turn, a model that is not repeatable, cannot
384 be reproduced, and therefore cannot be deemed credible by
385 others (who cannot understand the internals, especially if
386 expert modelers cannot reproduce it). Therefore, the proof of
387 model repeatability and reproducibility lies with the modeler who
388 needs to prove the value of the model for a given application.

389 A modeler should consider the model’s purpose from the start
390 of development and consider: for what; by who; what is the level
391 of knowledge and skill of the user; and in what environment. If
392 only the “for what” is specified, this implies the model can be used
393 by anyone on any system, which broadens the scope and reduces
394 the chance of reproducibility and hence reduces credibility by the
395 potential user.

396 This proof of value becomes of utmost importance if critical
397 decisions rely on a model. There are, in fact, many cases where
398 human lives depend (directly or indirectly) on a model.

400 In most cases, users expect a system based on a model to be
401 accredited somehow before use. In analogy, a physician will also
402 not use a medical product, for example a medical device, that is
403 not approved and tested for the anticipated use—because of the
404 risk to harm the patient should the product fail or not work
405 consistently.

406 This accreditation role many times falls on government
407 agencies such as the FDA, or NASA. Those agencies have
408 different approaches towards credibility of models. Guidelines
409 issued by authorities regulating the use of such models give a
410 “gold standard” recipe for how a modeler can ensure
411 reproducibility and establish credibility of a model.

412 NASA takes modeling seriously. After the space shuttle
413 Challenger disaster NASA rewrote a standard (NASA, 2016)
414 and wrote guidelines (NASA, 2019). An interesting component
415 in the NASA approach was a risk adjusted approach that
416 considered both the probability and consequences of a
417 modeled systems failure, in which the level of risk raises or
418 lowers the bar for the data needed to accredit a model. This
419 approach also helps with cost.

420 FDA regulates the use of medical devices and drugs and also
421 assesses computational models submitted as part of the market
422 authorization dossiers. For many years, simulations have been
423 part of these dossiers. If models can systematically shortcut and
424 prevent issues related with long and costly clinical trials still needs
425 proof, but the number of submissions to the FDA under the use of
426 models has been constantly rising over the past years. The FDA
427 has released and adopted guidance documents on the reporting
428 and validation of computational models for regulatory
429 submissions, (FDA, 2016; FDA, 2003), and actually considered
430 the NASA standards when creating those (FDA, 2016). In 2018
431 the American Society of Mechanical Engineers (ASME) issued an
432 important guidance ASME V&V 40 (ASME, 2018) of how to
433 assess credibility of computational models of medical devices
434 through verification and validation (V&V). The guideline is
435 centered around the definition of the context of use (CoU) of
436 the model, which is formulated based on the questions of interest
437 the model will answer. The CoU is then analyzed in terms of the
438 “model risk” - being the influence the model exerts on a decision
439 and the potential consequences these decisions might incur.
440 Commensurate with this model risk, the modeler establishes
441 the credibility goals, performs verification validation and
442 uncertainty quantification actions, and then assesses the
443 outcome of this exercise in order to allow judging of the
444 acceptability of the model CoU. Key to this guidance is its
445 overarching nature that also allows adoption in other (e.g.,
446 drug development) fields irrespective of the model type
447 (Viceconti et al., 2019; Kuemmel et al., 2020). Very recently,
448 an FDA guidance draft has been updated taking into account this
449 standard (FDA 2021b). In the paper by (Viceconti et al., 2019) the
450 verification, validation and uncertainty quantification (VVUQ)
451 pipeline is streamlined to different types of models. It is, perhaps,
452 the closest to score credibility across model types from
453 mechanistic physics driven models to machine learning
454 models. However, it is still short of including very recent
455 developments such as ensemble models, although it touches
456 upon the topic.

The FDA understood the potential value of models and modeled data to make developments of medical devices in an efficient manner. Early in the collaboration with the device industry about use of model data in trial processes, the FDA suggested having a “library of “reusable”, “regulatory grade” models. The FDA passed on the idea but is revisiting the library idea given models that meet the information guide for first accreditation. The idea is that the FDA understood the model and it proved useful, so they can accredit it much faster (cheaper) for reuse on a very similar application. Time will tell if this approach works.

In order to advance modeling and simulation fit for regulatory application, FDA and pharmaceutical companies engage in a model informed drug development (MIDD) pilot program (FDA, 2021a). This pilot program was released in response to a performance goal agreed to under the sixth iteration of the Prescription Drug User Fee Act (PDUFA VI), included as part of the FDA Reauthorization Act of 2017 and advises how particular MIDD approaches can be used in a specific drug development program (Zineh, 2019), and how to report those complying with existing guidelines for regulatory submissions. While for MIDD, commonly data-driven or phenomenological models are used, more complex and multiscale models are coming of age and get submitted to regulators (Viceconti et al., 2021). Data from relevant MSM tools to refine, reduce or even replace trials, could provide additional economic incentives to sponsors (Galluppi et al., 2021). Likewise, however, especially complex models and MSMs models that do not meet the FDA requirements for credibility of model data, will fail to be considered. Guidelines, especially for complex multiscale models are still lacking and thus adoption of other guidelines e.g., ASME V&V40 is needed and discussion with regulatory agencies should be conducted before submission (Musuamba et al., 2021). Competitors focused on the regulatory side will out-compete those that do not, or cannot, comply. It will therefore be necessary to improve available guidance and standardization efforts with regards to repeatability, reproducibility, and reuse so that guidance can be adopted cross-community and entity, by academic and commercial ones alike.

Despite the importance of developing a model with a question of interest and the respective context of use (CoU) in mind, it is important to note that the past paradigm used towards model acceptance/credibility may change in the future. For example there could be multiple motivations for developing a model, motivations could change over time, and someone else could find a new use for a model that was intended for another purpose. Since some contributors to this manuscript have a less strict opinion regarding models being developed with a purpose in mind, we therefore recorded the range of opinions in this manuscript.

An example of a less strict approach to model credibility are new ensemble techniques, such as in (Barhak, 2016; Barhak, 2017), which allow judging a model by its performance in a group of models. This is similar to building teams in sports, where each individual contributes to a team and the value contributed to the team can be determined. Ensemble models allow assigning

influence to single models and judging their performance by validation in different scenarios. Thus assigning a score to the model and its assumptions compared to others is possible. So the idea of credibility score may evolve through time and government agencies should consider this newer approach towards credibility. However, even if reproducibility and credibility are amenable there are many issues that prevent reuse of models.

While guidelines and concepts of how to establish credibility of models—even for critical applications—do exist, the field is still evolving and lots of work on completing, harmonizing, and adopting these guidelines still exist. One central question is what a minimal requirement might be for a model to be credibly re-usable.

ONGOING DIFFICULTIES IMPEDING THE UTILITY AND INTEGRATION OF MODELS.

Since there is a large variety of known issues that prevent reuse and many solutions, we have divided them by topic. For simplicity **Table 1** describes the difficulties and possible solutions:

We also attempted to spread those difficulties as hurdles that relate to reproducibility, credibility, utility, and integration. **Figure 2** depicts this analogy.

The paper continues elaborating on those topics and expands explanations hereafter.

Built-In Barriers for Evaluating Model Credibility

If one considers models that exist currently, what impedes third parties in assessing their credibility according to “gold standard” guidelines (discussed before)? Most of them are not designed, too little or not transparently documented or supported with material allowing a third party for such credibility assessment.

Models include assumptions that need to be specified. Users need to know, under what conditions the model is appropriate? This is a question asked by any modeler. More provenance information is needed for reuse and composition. Another investigator who wants to expand a model may need to know what the assumptions or design decisions were so they know how to appropriately modify or expand a model. A regulatory body might want to be able to trace a model back to the data sources that informed it. Someone who wants to re-train a model for a different cell type or tissue might want to trace the data back to know what aspects of the training data need to be replaced. Enhancing model credibility can be achieved through enhancing documentation, establishing best practices, and tests.

Examples of information suggested to include in documentation are the design decisions that motivated a model, what the model is designed to explain/forecast, and explanations of data sources that contributed to a model. Ideally, this would include links to data repositories, indicate which assumptions were used to interpret the data, consider the methods/tools/users associated with model calibration, and evaluate if the model fulfils its intended purpose (Parker et al., 2002).

TABLE 1 | Difficulties and possible solutions towards improving model utility.

Difficulty	Potential solutions
Models have built-in barriers to evaluate model credibility	Better modeling practices, documentation, and tests
Models are written in different languages	Common transport specifications such as SBML or CellML, and proper documentation and annotation
Models are hard to locate	Archive web sites such as: BioModels, SimTK, IMAGWiki, and the future modeXchange
Lack of common platforms for executing models and simulations	Platforms such as BioSimulators, and runBioSimulations
Modeling requires adaptation towards integration	Tools for composing models such as SBML-Comp, and SemGen
Unit standardization	Standardization efforts, and machine learning solutions such as ClinicalUnitMapping.com
Data availability and measurement definitions	Models that merge human interpretation, and newer measurement devices
Missing annotations in models	Adoption of policies such as those COMBINE suggests
Models are not consistently licensed to allow for reuse	Abandoning some old school open source licenses and promoting licenses that release to public domain
Different scales and modeling paradigms	Standardization effort and centralization tools
Model application and implementation barriers	Education of modelers, users, and the public
Stochastic modeling difficulties	Development of tools that guarantee repeatability and standards to address stochastic simulations

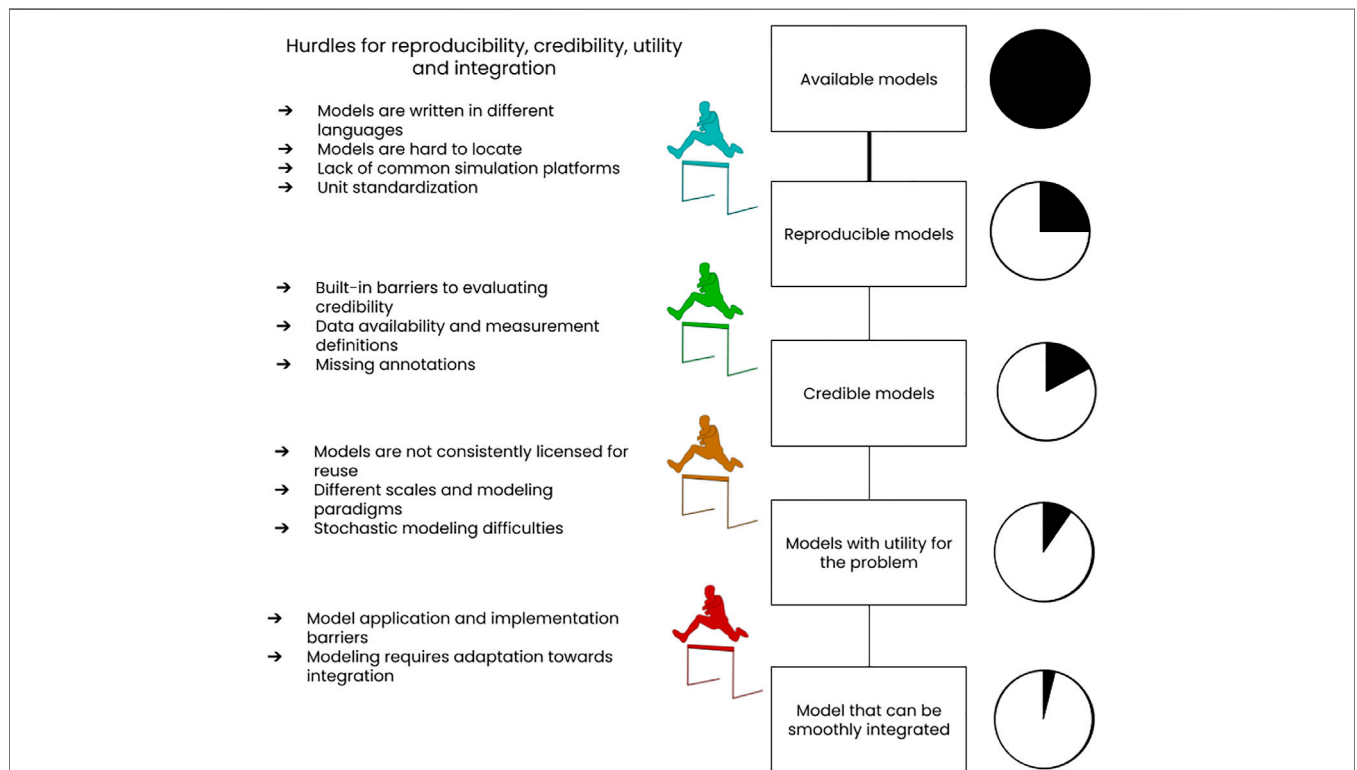


FIGURE 2 | Sketch of difficulties that impede reproducibility, credibility, utility and integration of models, especially in computational biology. An assignment of these difficulties to the four different concepts interpreting them as hurdles is attempted. We would like to point out that the sequential assessment model indicated in the graph is only one of the possibilities a modeler could use to assess the suitability of available models. In the rest of the paper we therefore address each difficulty separately.

Q22

In addition, documentation should describe model limitations. It can be difficult to quickly determine which populations or scenarios a model can be reasonably applied to. This information can usually be teased out by carefully considering the data that has been used for fitting or validation, as well as digging through the discussion or conclusion. However, some doubt often remains because of the natural tendency to promote one’s work, and the, perhaps unrealistic, expectation that publishable work be as widely applicable as possible. If it were standard practice in model

reporting to recommend specific model applications, this could provide clarity for those implementing or extending the model.

Beyond design and implementation, best practices should include reports of tests that describe what was simulated and the experimental or other data that was used to evaluate the test. Unlike software test reports which focus on failures, these reports must also focus on passes because they help establish the domain under which the model has been established to make trustworthy predictions. For example, this establishes the domain under

which their clinical use would be supported. When considering test implementation, some suggestions emphasize the need for a structured approach with unit-test style tests (Sarma et al., 2016; Gerkin et al., 2019) (Lieven et al., 2020) and continuous evaluation of such tests similar to continuous integration of software (Meyer, 2014; Krafczyk et al., 2019; Zhao et al., 2017). To enhance model credibility further, the model description should include validation tests against independent data, uncertainty assessments, and peer reviews (Refsgaard et al., 2005; Jakeman et al., 2006).

Models Are Written in Different Languages

When modelers do use a consistent, declarative language to describe their models, these models can then be stored and searched in readily-available repositories. The BioModels collection is a good example of such a repository for Systems Biology Markup Language (SBML) (Systems Biology Markup Language, 2021) models. As another example, the Physiome Model Repository (PMR) is a collection of CellML models. Although these repositories are a good step forward toward finding and reusing published models, by themselves, they are insufficient.

First, there are often significant differences between modeling languages—e.g., the CellML language and SBML are almost opposite in their approach to capturing the information in a model. Second, even within one modeling language it can be difficult for an outside user to understand the biological and mathematical content of a model written by someone else. As with software engineering, the key to enabling understandability and reuse of models is to provide unambiguous documentation about the intended semantics of the model.

One major problem we face for many kinds of models, which SBML and SBGN (Kitano, 2021) and projects like Biotapestry address partially for biological networks, is that we lack tools and formalism for consistently building, annotating, representing, displaying and manipulating conceptual models of complex biological phenomena with a spatial component. We lack standards for all of the key elements that need to be represented: the objects, the processes (behaviors and interactions) they participate in, the initial and boundary conditions and the dynamics and events that govern their evolution.

In many cases we also lack the scientific understanding of how to convert these conceptual models into mathematical models because we lack the “constitutive relations” that are the equivalent of the standard rate laws for chemical reactions. In this case we don’t have an agreed upon way to parametrize the submodels and to define their inputs and outputs.

Another big missing piece is a language to describe the possible experimental manipulations or perturbations of a biological system. We have concentrated on building mathematical and computational descriptions of biology, but not on the things we can do to them. Without such a description, classical techniques like perturbation and sensitivity analysis are much less useful. If we want to achieve a desired outcome by manipulating a given biological system, we need to know the constraints in our ability to manipulate that

system. Knowing that we could achieve what we want by increasing the value of k_{xx} by 25% is not actionable unless we can increase k_{xx} . The lack of orthogonality in biology (any perturbation of a biological system affects many aspects simultaneously, is what makes mathematical models so valuable for understanding (we have clean control parameters). But it also reduces their utility in designing experiments or clinical interventions. We need models that combine the model of the biological system with a model of the space of possible experiments. The sensitivity of this combined system is what tells us what is achievable in the lab or clinic.

Understanding the biological content of a model is critical to both reuse and reproducibility. If the model itself is incomprehensible, how can one know what its expected behavior and performance should be under different conditions? Semantic annotation is not necessary for simple repeatability, but if our goals include reproducibility and reusability, then we must make explicit and clear the biology and physics that underlie the model. An ideal modeling language should address this, yet until such a standard language is established we are faced with a need to integrate among different languages.

One simple integration example (Another example of integration and reproduction of a model, 2021) involving two popular languages, python and matlab, demonstrates the problem of transition between languages. There is no real translation between languages. No general compiler exists between multiple languages and human efforts are required. Fortunately there are standardization efforts among languages.

The standardization problem is not new and was considered by modelers a long time ago, resulting in the Systems Biology Markup Language (SBML) (Systems Biology Markup Language, 2021) that is a very helpful format that can help transport models between systems. SBML has a track record of success and allows transporting models between hundreds of systems. However, despite its popularity, it is not an official standard and the community decided not to go in that direction (Recent SDO/COMBINE legal entity issues, 2021). Note that there are many similar community standardization efforts aggregated in the biosimulation modeling community known as COMBINE (Computational Modeling in Biology Network) (COMBINE, Online). COMBINE includes SBML as well as many other specifications, yet those communities are still in the process of standardization and need to organize legally.

Nevertheless, the lack of legal governance does not stop communities from developing even more tools for result handling and analysis like PETab (Schmiester et al., 2021), SED-ML (Waltemath et al., 2011), SESSL (Ewald and Uhrmacher, 2014), KiSAO (Courtot et al., 2011), SBML (Dada et al., 2010), HDF5 (Folk et al., 2011), Vega (Satyanarayan et al., 2014; Satyanarayan et al., 2017), ggplot2 (Wickham, 2011), and others. Those tools show actual needs by the community, but these are much less mature and much less adopted. Their capabilities need to be expanded; they need to be adopted; software tools need to support them; and there needs to be infrastructure to share them, such as a repository. Another

799 piece is that the software tools needed for the above are scattered,
800 plus it is often unclear what subset of the above they support, and
801 tools often become inaccessible. Tools need to be submitted to
802 registries and the capabilities need to be annotated.

803 There is a need to coordinate the various standardization
804 efforts that are needed for the different scales and biology that
805 need to be involved in multi-scale models. The need for multiple
806 standards may be recognized, yet the need to coordinate them to
807 be able to compose multiscale models has received less attention.
808

809 Models Are Hard to Locate

810 Many times model location is a difficult task since models are
811 published in different sources. Despite many repositories
812 available there are many ways models are published including
813 journal papers, conferences, preprint services such as BioArxiv,
814 web sites, and code repositories such as GitHub. In some good
815 cases, there are model archive/linking web sites such as
816 BioModels (BioModels, 2021), SimTK (SimTK, 2021),
817 IMAGWiki (IMAGWiki, 2021), and in the future
818 modeleXchange (Malik-Sheriff et al., 2020). However,
819 currently there is no one aggregator that helps locate all
820 models and many times community members cannot agree on
821 location and attempt to create more repositories rather than
822 centralize efforts.

823 Moreover, simulation workflows are even harder to find. For
824 example, BioModels primarily focuses on models. There has been
825 much less focus on publishing the construction/calibration of
826 models, simulations, their results, analyses of their results, or
827 entire workflows for the above. Sharing all of this needs
828 embracing other repositories and developing some new ones.

829 We recommend that modelers use those repositories since we
830 are at a stage in evolution of modeling where model composition
831 is of interest and availability of modeling components is
832 important. We ask that modelers consider from development
833 to start permissive Intellectual Property for the versions published
834 in those repositories to increase accessibility.
835

836 Lack of Common Platforms for Executing 837 Models and Simulations

838 Even if models can be located, their simulation is a different issue.
839 Due to the existence of many partially supported standardization
840 efforts in this field, it is often difficult to know what tool needs to
841 be used with which model. It can also be difficult to find that tool,
842 download it, install it, learn it, and to use it, especially for large
843 simulations. These issues keep modelers in silos.
844

845 Even formats associated with standardization efforts have
846 difficulties. It is not possible, for example, to load a matlab/
847 SimBiology SBML L2V4 model into COPASI and someone
848 adhering to latest standards implementing L3V2 SBML
849 support for import will find difficulties in importing models.
850 Moreover, it is difficult to find a common platform that supports
851 all SBML versions. This version compatibility gap is not
852 uncommon. However, since biological models take a long time
853 to develop and represent phenomena that will persist for long
854 terms, it is important to have long term stability and support with
855 newer platforms and older models.

856 If the goal is for non-modelers to be able to interact with
857 models (e.g., to analyze data, to contribute data toward a
858 modeling project, or to apply a model for medicine), it needs
859 to be much easier to find and use these tools. Two initiatives that
860 are trying to address this are BioSimulators (BioSimulators, 2021)
861 and runBioSimulations (Shaikh et al., 2021).
862

863 Modeling Requires Adaptation Towards 864 Integration

865 Many times the models as published need some level of
866 manipulation to plug into another model. For example in
867 (Castiglione et al., 2021) the survival function needs
868 adaptation to transform it as can be seen from the public
869 discussion in (About using a multi-scale mortality model in
870 the ensemble, Online). Note that all those models need to be
871 scaled to the same units and scales. Another example is in (Ke
872 et al., 2020) where infectiousness is proportional to max
873 infectiousness while the models in (Hart et al., 2020) are
874 density models. In the model in (Castiglione et al., 2021) the
875 time scale was originally 8 h and it needed to change to daily
876 probability to merge into another model in (COVID19Models/
877 COVID19_Mortality_Castiglione at main, 2021), which required
878 scaling of the probability function. Those examples are relatively
879 simple integrations and in more complex integrations the
880 adaptation effort is more significant and many more
881 obstacles exist.
882

883 One obstacle is lack of standards for describing composite
884 models and software tools for merging models. One specification
885 is SBML-comp, but it is cumbersome and few tools support it.
886 Another tool is SemGen (Neal et al., 2019a), but it focuses on
887 finding mappings between similar models. To the point here,
888 SBML-comp is designed to compose models that were not
889 intended to be composed. Instead, composition needs to be
890 deeply ingrained into the entire community so that models are
891 anticipating the needs of composition from the beginning.

892 Note that adaptation towards standardization also requires
893 matching terminology, and especially matching of units of
894 measure, as well as proper documentation which we will
895 address in the next topics.
896

897 Unit Standardization

898 Unfortunately, units of measure are not yet standardized, an open
899 problem despite many attempts to resolve it by multiple
900 standardization bodies such as IEEE, CDCIC, and NIST. One
901 indication of the severity of the problem is that a Github search
902 for “unit conversion” shows over a thousand results. Another
903 good example of the severity of the problem is ClinicalTrials.Gov
904 that aggregates quantitative data from around the world and this
905 database shows over 24 K different units of measure (Barhak and
906 Schertz, 2019). One attempt at solving this standardization issue
907 using machine learning is ClinicalUnitMapping.com, yet this
908 project requires more effort.
909

910 Unit mismatches become particularly problematic when
911 trying to integrate models across different spatial or time
912 scales. For example, intracellular processes occur on
913 micrometer spatial scales and seconds to minutes time scales.
914

An in-host, tissue-level model of infection processes operates at millimeter to centimeter distances and hour to day time scales. When trying to integrate the two into a single multiscale model, care must be taken to ensure appropriate conversion of units when transferring output of one model as input to the other. This broad range of spatial and temporal scales can also cause computational problems requiring development of new algorithms to make computation more efficient across multiple scales (Jung and Sugita, 2017). While it is impossible to avoid having to convert units, we are advocating for clarity in the use of units. Sometimes models are used (and published) without specifying the units used for simulation parameters—this is a practice that needs to be corrected. Moreover, tools such as ClinicalUnitMapping.com can help overcome standardization difficulties in the future with more development. Once mapping to standardized units is easier, then simulation parameters can be converted appropriately when different scales or units are needed.

Lack of standard units for measurement of infectious virions is particularly problematic when trying to develop stochastic viral models. Stochastic models often require that we track individual infectious viral particles, yet it is not clear how the typical viral titer units of TCID₅₀/ml and pfu/ml convert to individual virions. Two attempts have been made to estimate the conversion factor, both for influenza, resulting in estimates of 1 TCID₅₀/ml of nasal wash corresponding to 10²–10⁵ (Handel et al., 2007) or 3 × 10⁴–3 × 10⁵ (Perelson et al., 2012) virions at the site of infection. Such order of magnitude uncertainty in unit conversion makes it difficult to develop accurate model representations of viral infections. We are not advocating for a standard conversion factor, since the conversion factor likely depends not only on the specific virus, but also conditions such as temperature and pH, which are known to affect viral infectivity (Rowell and Dobrovlny, 2020; Heumann et al., 2021). Rather, we are advocating for development of new viral measurement techniques that can more reliably quantify the number of infectious viruses present in a sample.

Data Availability and Measurement Definitions

When attempting to integrate models, the phenomenon being reproduced by the models or the data they are based on might not be the same. This is especially true when model definitions evolve or can change in many ways. Examples include International Classification of disease (ICD) codes (Wikipedia, 2021a) that went through multiple versions through the years, or even a disease definition that has evolved for sepsis (Gary et al., 2021). Even outcomes of clinical trials change if counted using different definitions as seen in (clinicaltrials.gov/NCT00379769 - GlaxoSmithKline, 2017). Those definitions can hinder connecting different models together. Possible solutions are machine learning techniques that can transfer interpretation or modeling techniques that merge human interpretation from multiple experts into the modeling process (Barhak, 2020a).

Another issue specific to models dealing with viruses may seem like the lack of unit standardization for measurement of

virus. However, it is a measurement definition issue. Infectious virus concentrations are measured using TCID₅₀/ml (50% tissue culture infectious dose) or in pfu/ml (plaque forming units), both of which depend on specific experimental conditions such as temperature, humidity, and measurement time. Studies have shown that even a lab using identical experimental conditions cannot reproduce the same measured experimental values of virus leading to differences in estimated parameters for models (Paradis et al., 2015). There is also an underlying assumption for both units of measurement that an observed plaque was initiated by a single infectious virion, which has never been clearly proven to be true. More recently, non-infectious virus particle concentrations are being measured using PCR. In this technique, the number of segments of a particular piece of RNA are measured. While this unit is more tangible and consistent than the infectious viral titer units, viral kinetics models often consider only infectious virions. Although non-infectious viruses are starting to be incorporated into models, the relationship between infectious and non-infectious virions changes over the course of an infection (Petrie et al., 2013), making it difficult to use these measurements to get at the underlying infectious virus dynamics. New measurement techniques and strategies for more direct measurement of infectious virions are being developed (Cresta et al., 2021).

Data availability to rationalize calibration and validation of models is crucial but often not possible because of data sharing policy and privacy (especially for individual human data). Moreover, undisclosed data from industry sponsored clinical trials used in model building and validation generally excludes many useful models from any assessment by the scientific community. This is a difficult problem but there may exist some partial solutions. Synthetic data that is statistically similar to real-world data without containing information about any real individual can be shared. While the similarity can only be evaluated with access to the original data as in (Bansal et al., 2013), because it can be shared, it can be used for calibration and validation of other models (Barhak, 2017; Ajelle et al., 2018; Liu et al., 2018; Wang et al., 2020). There is a risk of error using synthetic data in this way since, though it may have been similar in some respects to the original data, it might be different in some other respect that matters for a model different from that used for validation. For validation of model results against individual data that cannot be shared, it is conceivable that services could be deployed to query the data. Differential privacy (Dwork 2008; Garfinkel and Leclerc, 2020) to establish a privacy budget for such a service providing an information theoretic bound on how much information is allowed to be revealed in response to queries. This budget can be set to whatever is considered ethically and administratively acceptable. More research is needed to adapt this idea to suit model making needs.

Even data that are publicly available has limitations. In the case of within-host viral kinetics models, sampling of viral loads and immune responses is often not done frequently enough or long enough to ensure parameter identifiability (Miao et al., 2011). During the recent SARS-CoV-2 pandemic, several attempts were made to parameterize within-host viral kinetics models using viral loads measured from patients, but these measurements were

often collected only after a patient was hospitalized, so the crucial viral growth phase is missing (Hernandez-Vargas and Velasco-Hernandez., 2020; Wang et al., 2020). Additionally, viral loads were measured via nasal swab, though it is not clear that the viral load in the nose is correlated to the viral load in the respiratory tract, which is the infection location simulated in the models. Other viruses can also infect internal organs that are difficult to access for frequent measurements without invasive procedures. Immune responses are often measured using levels in the blood, which is typically not the site of infection and often not the location of viral load measurements either. Since models are attempting to replicate virus and immune dynamics at the site of infection, these data collection limitations make it difficult to collect the data needed to accurately parameterize such models for humans.

A further methodological issue with some of the available *in vitro* and *in vivo* experimental data is that it often does not represent the infection time course in a single individual or single experimental well. Clinical trial data is often presented as medians or means taken over all patients. A recent study of influenza infections showed that parameter estimates based on fits of models to a single median viral titer curve do not match estimated parameter values based on fits to individual patients (Hooker and Ganusov, 2021). This also masks patient-to-patient variability in infection. Pre-clinical animal studies and *in vitro* studies can be even worse as animals are often sacrificed and infections in individual wells are stopped to make measurements at each time point. In this case, experimental data consist of an average of measurements from several animals/wells at each time point that differ from the set of animals/wells at other time points.

Missing Annotations in Models

In biosimulation models, documentation about the intended semantics of the model is captured by annotations - additional information that describes the model, and the biological entities included in that model. Further, these annotations can leverage the rich resources of bio-ontologies-consistent nomenclatures and terminologies that describe the biological world in great detail.

COMBINE has recognized these challenges for understandability and reuse of models and is working hard to disseminate best practices around semantic annotation. COMBINE consensus around annotation is described in (Neal et al., 2019b). This paper describes some key tenants for improved semantic annotation: First, these annotations should be written using a standard format, and one that is independent of modeling languages. Thus, COMBINE recommends RDF as a simple triple-based representation to connect model elements to annotations and knowledge resources (e.g., ontologies). Next, COMBINE recommends that annotations should be stored externally from the source code of the model. Obviously, the annotations should be linked to elements within the model source code, but in order to be language independent, they should be stored separately. Finally, COMBINE recommends that modelers and model building communities provide policies and rationale for choosing which knowledge resources to use for which types of annotations. Otherwise, the same biological entity may look

different if different modelers annotate the entity against different bioontologies.

Annotations can be useful for multiple tasks such as: annotation of semantic meaning where biology or real world relevance is explained, annotation of provenance where the origins of the model and its creators are referenced, and annotation of verification indicating tests the model should undertake and pass. However, there is a lack of sufficient annotation about the components of models. This is particularly because modelers choose not to provide annotation and because tools for describing the semantic meaning of components are just starting to emerge. For example, for biochemical models there's HELM and BpForms. The lack of such annotation makes it hard to determine the points of overlap between models. In addition, there is a lack of annotation about data sources and assumptions which makes it hard to determine whether models are compatible or what needs to be done to make them compatible. For example, do two models represent the same cell type, tissue, or gender? Hopefully policies will be adopted to resolve this issue.

Models Are Not Consistently Licensed in an Easy Way That Allows Reuse

Different institutions have different approaches towards licensing, as can be seen from this discussion (Licensing issues, 2021). Therefore, model creators may not be aware of the implications of licensing many times when they publish their models. Moreover, some licenses are incompatible with each other or other forms of Intellectual Property (IP) such as patents (Wikipedia, 2021b). Even open source licenses are quite restricted since they are based on copyright laws, which give the owner rights to restrict usage (Barhak, 2020b). In this sense open source licenses resemble patents and in some cases are more restrictive since patents become public domain quicker. Moreover, community members take different sides with regards to licensing issues as can be seen in this discussion (Issues with regard to Call for transparency of COVID-19 models, 2021). Specifically, one license that will make reuse much easier is Creative Commons Zero (CC0) (CC0 Creative Commons, 2021). This license uses the term "No rights reserved" and makes it easier for models and text to be reused with less restrictions. In fact model repositories such as BioModels require releasing the models uploaded there under CC0 (Malik-Sheriff et al., 2020). However, CC0 license has not been adopted by some (Wikipedia, 2021c). To eliminate the licensing problem, modeling communities will have to abandon old school open source licenses that are based on copyright and create conflicts and recommend releasing models to the public domain using licenses such as CC0.

We therefore recommend that models and their associated data should be published under permissive terms. For maximizing reproducibility and integration, we suggest that the most permissive license possible should be chosen. In that regard the CC0 license would be a good choice, effectively waiving interests of the creator in their works and therefore emulating the public domain in jurisdictions where this is necessary.

Different Scales and Modeling Paradigms

Models are operating on different spatial scales (population or individual) with different modeling paradigms (continuous vs discrete). A tissue could be modeled as a continuum leading to Partial Differential Equations (PDEs) or as a collection of individual interacting cells leading to an agent-based model. Specification of these two models would probably require different languages.

The fact that models capture different scales or that they don't consistently capture any single scale creates challenges for composition. One opinion is that the challenge is that the scale of a model is not clearly annotated. To compose models, this forces the composing investigator to try to figure out the scales of each model and how to mesh them. Typically this is combined with lack of annotation of units. When developing a standard for specifying models, developers will probably need standards specific for each modelling approach. One possibility is that a family of specification standards may be created. The need for different formats for different domains and scales will probably create the need for a central place where, especially non-modelers, can find information about these various future standards and which tools support them. Ideally, there would also be a central place where these tools can be obtained and executed so that even non-modelers can easily explore models without having to figure out what software is needed, install it, etc. However, it is still unclear what common practices might facilitate composition across scales and how the various component standards should be architected to facilitate integration.

Model Application and Implementation Barriers

Models are difficult to be used by a community or government. Scientific, regulation, and social communities have different sets of models and different understanding and standards in models. It is hard to convince and establish a common popular model widely acceptable by a wide range of communities and even adopted by the government. The long term validation and approval process may delay the cycle from model application to implementation.

Many models can increase their utility to the scientific community if their applicability and implementation is easy. These aspects can be improved by means of several features. One crucial feature is the reproducibility of the model since usually this is necessary before the model can be applied by the scientific community. The likelihood that a model is applied to different problems increases if their results are reproducible. The reproducibility is difficult to test if there are implementation barriers. Some of these barriers have been pointed out in this article. Thus, we can infer that implementation issues of the models affect the reproducibility of the models and therefore a broad utility of the models to the scientific community.

Currently many models are difficult to implement and therefore unable to make a real impact. Many of the existing models that are used by decision makers are used because those were implementable. More sophisticated models are many times not used due to a need for proper tools or proper expertise. Therefore, many good ideas remain unused due to

implementation difficulties. The solution to this problem is long term and requires education of developers, users, and the public.

Stochastic Modeling Difficulties

Biological systems are exceptionally complex, involving a multitude of interactions among a large number of components at different spatial and temporal scales. Over the years, much work has been performed wherein deterministic models have been developed to understand dynamics from the cellular level to the population level (Murray, 2002). Although these works have provided much insight, it is known that the mean-field dynamics of these deterministic models do not always capture important phenomena (Forgoston and Moore, 2018). For example, disease population models often have a stable endemic state for reproduction numbers greater than one, and therefore it is not possible for the disease to go extinct in the models. This is in direct contrast to the local extinctions of disease that occur all the time in the real world (Assaf and Meerson, 2010; Bauver et al., 2016), (Doering et al., 2005; Dykman et al., 2008; Ovaskainen, and Meerson, 2010; Forgoston et al., 2011; Schwartz et al., 2011; Nieddu et al., 2017; Billings and Forgoston, 2018). Similarly, at the within-host level, new infections may or may not establish. This phenomenon can be captured by stochastic models, but is not realised by a deterministic model with a single set of model parameters. Furthermore, deterministic models do not account for the random interactions of cells or individuals, nor do they account for the changes in the model's rates which are related to random events.

To properly model real-world multiscale dynamics, it is often necessary to use stochastic approaches that allow one to make quantitative, statistical predictions, while simultaneously providing qualitative descriptions of system dynamics. The ability to generate stochastic simulations that provide quantitative statistics for the emergence of new dynamics is increasing with advances in computational power. However, the inclusion of stochasticity leads to a variety of issues related to reproducibility. One concern is associated with noise-induced transitions (Assaf and Meerson, 2010; Forgoston and Moore, 2018) or stochastic resonance in which the deterministic system is qualitatively different from the stochastic system. The output of a stochastic model is a distribution or time-series of distributions. In particular, it is possible to have different outcomes for the same model or set of parameter values (e.g., a multimodal equilibrium distribution). In this case, one must use more sophisticated techniques such as Kullback-Leibler divergence or Wasserstein's distance to make quantitative comparisons with data or other models. Moreover, while it may be possible to compare distributions generated by different stochastic models, it is often not possible to generate identical, individual realisations.

Stochastic models are critically important. Indeed, unlike deterministic models, stochastic models give rise to probabilistic predictions based on ensembles of realisations. However, in general, these types of models present difficulties that include: 1) how one validates a stochastic simulation; and 2) how one can ensure the repeatability of a stochastic simulation. It is worth noting that the latter issue becomes more problematic

when software libraries that support modern high-performance computation (including standard parallel computation as well as GPU computation meant to accelerate simulation), cannot guarantee deterministic reproducibility (DOC, 2021). Potential solutions include the development of tools that guarantee repeatability such as ways to set and record pseudo-random number generator seeds as used by the Micro Simulation Tool (MIST) (Barhak, 2013) and developing standards to address stochastic simulations.

Open Discussion

During the work on the paper, a few other items were raised that were not resolved and we assembled those here so those can be addressed in a future version. Some of those issues are visible, yet out of reach, at least for this group of authors and the group invites correspondence on resolving the issues we list below.

How existing Multi-Scale frameworks could be made more transparent with respect to the models they encode and potentially more interoperable. How are most multiscale models encoded? Do they use general modeling frameworks that have been developed for this purpose or do they rely on custom-built code? What general purpose multiscale frameworks exist and what types of model integration and linking do they support? Are there opportunities to develop standards that would make such models more reproducible and understandable? What intermediate steps might be possible? For example, some of these frameworks support integration of SBML models (or models encoded in other somewhat standardized languages). Is it beneficial to leverage existing standardization efforts and to develop new ones for the description and implementation of MSM's going forward? Right now, it is a bit of a wild west where these models are being developed with submodels as generic pieces of code in general purpose languages like C++ and *Python* that rely only on unstructured comments for documentation.

It is still unclear how various component standards should be architected to facilitate integration. Many of the solutions discussed will require dedicated tools and standards and there will be many of those. It is unclear how distributed and decentralized components will be governed. A resource such as modeXchange (Honorato-Zimmer et al., 2018) may help investigators navigate this landscape for models, and similar tools may be needed to navigate tools and standards.

There is no common methodology on how to deal with the gap between model parameters, data collection, and standards. This topic should be discussed in the future in light of a potential solution of standardized model development explanation. However, standards have to evolve to handle this issue in proper standard organizations.

The COMBINE (COMBINE - Coordinating standards for modeling in biology, 2021) organization has been helpful to start the standardization process. However, it is not a Standards Development Organization (SDO) and its members rejected joining SISO, which was an established SDO. Therefore the products of this work may not be widely accepted unless the organization matures and adopts a legal entity standing with all regulations involved. However, will the community behind this organization mature enough to adopt legal bindings and regulations?

Common formats for results and visualizations still have not been established despite importance. When models produce results during simulation those should be archived and visualized somehow to help user interactions. A common format to represent results and how to generate graphics to represent will help with credibility and integration efforts.

For models to mature, there is a need to establish testing paradigms. Testing included all general software development good testing practices such as verification of calculation code, regression testing, plus model validation and usability testing. One of the authors recommended that a future model testing practices paper be developed. Another voice mentioned that the ensemble modeling approach includes tests within it. Yet the group reached no conclusion.

Barriers preventing validation reduce credibility and therefore have a negative impact on model credibility, which prevents reuse. One barrier is model validation at different scales: Models at different scales from molecular to population scale are usually validated at different standard and testing samples. Cross-scale validation is very difficult since there are multiple factors involved that influence the outcome of different scales. Another barrier is model validation for practical use cases: Real world prediction from the developed model is challenging because of the complexity of the pathogen spreading process. The real spreading process always has a lot of random social and physiological variables that are hard to be included in any model. With more advanced models and availability of more data, practical forecasts will get more accurate.

Another, difficult topic is spatial models. It seems we are pretty much at the beginning in terms of defining spatial models. We don't have a quantitative language to specify cell shapes, cell behaviors, or tissue architecture. In many cases we don't even have a qualitative language to do this.

Bioscience modelers are not alone dealing with utility and reuse related issues. A 2016 report on complex systems engineering challenges (Fujimoto et al., 2017) identified other non technical barriers in the form of social, behavioral and programmatic barriers that were not addressed among the technical issues in this paper.

These and many other topics may be issues for the group to discuss in the future and readers are welcome to join the discussion.

CONCLUSION

This manuscript discussed the reproducibility crisis in biological computational models. Many issues and difficulties and barriers have been presented. Nevertheless, some efforts towards solutions already are in progress and have been mentioned. We can categorize those challenges to scientific problems, and cultural and community based challenges.

Examples of scientific problems include the need to build good model integration environments, and to establish how to integrate models across paradigms. We also need to resolve many of the stochastic modeling challenges.

Examples of cultural and community barriers include education towards standardization of units, education towards proper annotation of models. Good tools may help with education by helping do those tasks semi automatically.

It is expected that many solutions will have both scientific and cultural aspects. The list of issues should not discourage modelers from developing models. Instead modelers should view this list as a reference of issues to be solved in the future and issues to avoid. The first step in solving the problem is admitting it exists. With this paper the authors recognize the challenges and admit the current state of modeling needs fixing. Hopefully fixing those issues, starting with reproducibility, will increase model credibility and will facilitate reuse and later integration of models. The long term goal of this group is improving models to achieve better human and machine comprehension of biological processes.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

About using a multi-scale mortality model in the ensemble (Online) (2021). MSM Viral Pandemic Integration Subgroup Mailing List Archives. Available at: <https://lists.simtk.org/pipermail/vp-integration-subgroup/2021-January/000011.html> (Accessed October 24, 2021).

Ackleh, A. S., Ma, B., and Thibodeaux, J. J. (2013). A Second-Order High Resolution Finite Difference Scheme for a Structured Erythropoiesis Model Subject to Malaria Infection. *Math. Biosciences*. 245, 2–11. doi:10.1016/j.mbs.2013.03.007

Ajelli, M., Zhang, Q., Sun, K., Merler, S., Fumanelli, L., Chowell, G., et al. (2018). The RAPIDD Ebola Forecasting challenge: Model Description and Synthetic Data Generation. *Epidemics*. 22, 3–12. doi:10.1016/j.epidem.2017.09.001

Another example of integration and reproduction of a model (2021). MSM Viral Pandemic Integration Subgroup Mailing List Archives. Available at: <https://lists.simtk.org/pipermail/vp-integration-subgroup/2021-January/000012.html> (Accessed Nov 1, 2021).

ASME (2018). Assessing Credibility of Computational Modeling through Verification & Validation: Application to Medical Devices - ASME. Available at: <https://www.asme.org/codes-standards/find-codes-standards/v-v-40-assessing-credibility-computational-modeling-verification-validation-application-medical-devices> (Accessed October 24, 2021).

Assaf, M., and Meerson, B. (2010). Extinction of Metastable Stochastic Populations. *Phys. Rev. E*. 81, 021116. doi:10.1103/PhysRevE.81.021116

Assaf, M., Roberts, E., and Luthey-Schulten, Z. (2011). Determining the Stability of Genetic Switches: Explicitly Accounting for mRNA Noise. *Phys. Rev. Lett.* 106, 248102. doi:10.1103/PhysRevLett.106.248102

Atkey, R., Gavranović, B., Ghani, N., Kupke, C., Ledent, J., and Forsberg, F. N. (2020). Compositional Game Theory, Compositionally. Available at: <https://pureportal.strath.ac.uk/en/publications/compositional-game-theory-compositionally> (Accessed October 24, 2021).

AzmyAckleh, S., and KarynSuttonLihong Zhao, L. (2018). A Second Order Finite Difference Scheme for a Variable Infection-Structured Model of mycobacterium Marinum Dynamics in Aquatic Animals. *J. Nonlinear Variational Anal.* 2, 2. doi:10.23952/jnva.2.2018.2.06

Baez, J. C., Genovese, F., Master, J., and Shulman, M. (2021). Categories of Nets. arXiv:2101.04238 [cs, Math]. Available at: <http://arxiv.org/abs/2101.04238> (Accessed October 24, 2021).

Baez, J. C., and Master, J. (2020). Open Petri Nets. *Math. Struct. Comp. Sci.* 30, 314–341. doi:10.1017/S0960129520000043

FUNDING

Jonathan Karr was supported by Grant NIH-NIBIB P41EB023912 “Center for Reproducible Biomedical modeling”. WW is supported by UK Medical Research Council MRC grant MR/V027956/1. RS is supported by EMBL-core funding from the member states of the European Molecular Biology Laboratory. RB was supported by a fellowship funded by the Medical Research Council MR/P014704/1. JG acknowledges funding support from grants NSF 188553, NSF 186890, NSF 1720625, NIH U24 EB028887 and NIH R01 GM122424.

ACKNOWLEDGMENTS

John Gennari has significantly contributed to this manuscript, yet chose not to be included on the author list. The group thanks him for his efforts. Additional thanks to Tomas Helikar who participated in discussions. Thanks to the MSM viral pandemic working group.

Bai, J. P. F., Earp, J. C., Florian, J., Madabushi, R., Strauss, D. G., Wang, Y., et al. (2021). Quantitative Systems Pharmacology: Landscape Analysis of Regulatory Submissions to the US Food and Drug Administration. *CPT Pharmacom Syst. Pharma.* 10, 1479–1484. doi:10.1002/psp4.12709

Barhak, J. (2013). MIST: Micro-Simulation Tool to Support Disease Modeling: SciPy 2013 Presentation. Available at: <https://www.youtube.com/watch?v=AD896WakR94> (Accessed October 24, 2021).

Barhak, J. (2016). The Reference Model for Disease Progression Combines Disease Models. In (Orlando Florida). Available at: http://sites.google.com/site/jacobbarhak/home/IITSEC2016_Upload_2016_11_05.pptx.

Barhak, J. (2017). The Reference Model: A Decade of Healthcare Predictive Analytics (PyTexas 2017). Available at: https://www.youtube.com/watch?v=Pj_N4izLmsI (Accessed October 24, 2021).

Barhak, J. (2020a). The Reference Model for Disease Progression Handles Human Interpretation. Available at: https://www.modsimworld.org/papers/2020/MODSIM_2020_paper_42_.pdf.

Barhak, J. (2020b). Open Source and Sustainability. Available at: https://jacobbarhak.github.io/COMBINE2020_OpenSource_upload_2020_10_04.odp.

Barhak, J., and Schertz, J. (2019). Clinical Unit Mapping with Multiple Standards. Available at: https://jacobbarhak.github.io/Poster_CDISC2019.html.

Bauer, M., Forgoston, E., and Billings, L. (2016). Computing the Optimal Path in Stochastic Dynamical Systems. *Chaos*. 26, 083101. doi:10.1063/1.4958926

Billings, L., and Forgoston, E. (2018). Seasonal Forcing in Stochastic Epidemiology Models. *Ricerche mat.* 67, 27–47. doi:10.1007/s11587-017-0346-8

BioModels (Online) (2021). Available at: <https://www.ebi.ac.uk/biomodels/> (Accessed October 24, 2021).

BioSimulators (Online) (2021). Available at: <https://biosimulators.org/> (Accessed October 24, 2021).

Blanes, S., Casas, F., and Murua, A. (2008). Splitting and Composition Methods in the Numerical Integration of Differential Equations. arXiv:0812.0377 [math]. Available at: <http://arxiv.org/abs/0812.0377> (Accessed October 24, 2021).

Bowness, R., Chaplain, M. A. J., Powathil, G. G., and Gillespie, S. H. (2018). Modelling the Effects of Bacterial Cell State and Spatial Location on Tuberculosis Treatment: Insights from a Hybrid Multiscale Cellular Automaton Model. *J. Theor. Biol.* 446, 87–100. doi:10.1016/j.jtbi.2018.03.006

Castiglione, F., Deb, D., Srivastava, A. P., Liò, P., and Liso, A. (2021). From Infection to Immunity: Understanding the Response to SARS-CoV2 through In-Silico Modeling. Available at: <https://www.biorxiv.org/content/10.1101/2020.12.20.423670v2> (Accessed October 24, 2021).

CC0 Creative Commons (2021). Available at: <https://creativecommons.org/share-your-work/public-domain/cc0/> (Accessed October 24, 2021).

1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
Q14 437
Q15 438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482

1483 CKAN (Online) (2021). The Open Source Data Management System ckan.Org. Available at: <http://ckan.org/> (Accessed October 24, 2021).

1484 clinicaltrials.gov/NCT00379769 (Online) - GlaxoSmithKline (2017). A Long Term, Open Label, Randomised Study in Patients With Type 2 Diabetes, Comparing the Combination of Rosiglitazone and Either Metformin or Sulfonylurea With Metformin Plus Sulfonylurea on Cardiovascular Endpoints and Glycaemia. clinicaltrials.Gov. Available at: <https://clinicaltrials.gov/ct2/show/results/NCT00379769> (Accessed October 24, 2021).

1489 COMBINE (Online) - Coordinating standards for modeling in biology (2021). COMBINE. Available at: <http://co.mbine.org/> (Accessed October 24, 2021).

1491 Courtot, M., Juty, N., Knüpfner, C., Waltemath, D., Zhukova, A., Dräger, A., et al. (2011). Controlled Vocabularies and Semantics in Systems Biology. *Mol. Syst. Biol.* 7, 543. doi:10.1038/msb.2011.77

1494 COVID19Models/COVID19_Mortality_Castiglione at main (Online) (2021). Jacob-Barhak/COVID19Models GitHub. Available at: <https://github.com/Jacob-Barhak/COVID19Models> (Accessed October 24, 2021).

1495 Cresta, D., Warren, D. C., Quirouette, C., Smith, A. P., Lane, L. C., Smith, A. M., et al. (2021). Time to Revisit the Endpoint Dilution Assay and to Replace TCID₅₀ and PFU as Measures of a Virus Sample's Infection Concentration. *arXiv:2101.11526 [q-bio]*. Available at: <http://arxiv.org/abs/2101.11526> (Accessed October 24, 2021).

1499 Q18 Dada, J. O., Spasic, I., Paton, N. W., and Mendes, P. (2010). SBRML: a Markup Language for Associating Systems Biology Data with Models. *Bioinformatics*. 26, 932–938. doi:10.1093/bioinformatics/btq069

1500 Danos, V., and Laneve, C. (2004). Formal Molecular Biology. *Theor. Computer Sci.* 325, 69–110. doi:10.1016/j.tcs.2004.03.065

1501 DOC (2021). Clearly Indicate Deterministic Algorithms for Reproducible Science · Issue #2685 · Rapidsai/cuml GitHub (Online). Available at: <https://github.com/rapidsai/cuml/issues/2685> (Accessed October 24, 2021)

1502 Dodds, M. G., Krishna, R., Goncalves, A., and Rayner, C. R. (2021). Model-informed Drug Repurposing: Viral Kinetic Modelling to Prioritize Rational Drug Combinations for COVID-19. *Br. J. Clin. Pharmacol.* 87, 3439–3450. doi:10.1111/bcp.14486

1503 Doering, C. R., Sargsyan, K. V., and Sander, L. M. (2005). Extinction Times for Birth-Death Processes: Exact Results, Continuum Asymptotics, and the Failure of the Fokker-Planck Approximation. *Multiscale Model. Simul.* 3, 283–299. doi:10.1137/030602800

1504 Dwork, C. (2008). “Theory and Applications of Models of Computation,” in *Theory and Applications of Models of Computation Lecture Notes in Computer Science*. Editors M. Agrawal, D. Du, Z. Duan, and A. Li (Berlin, Heidelberg: Springer), 1–19. doi:10.1007/978-3-540-79228-4

1505 Dykman, M., Schwartz, I., and Landsman, A. (2008). Disease Extinction in the Presence of Random Vaccination. *Phys. Rev. Lett.* 101, 078101. doi:10.1103/PhysRevLett.101.078101

1506 Ewald, R., and Uhrmacher, A. M. (2014). Sessl. *ACM Trans. Model. Comput. Simul.* 24 (11), 1–25. doi:10.1145/2567895

1507 FDA (2003). Guidance for Industry Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications. Available at: <https://www.fda.gov/media/71277/download>.

1508 FDA (2016). Reporting of Computational Modeling Studies in Medical Device Submissions - Guidance for Industry and Food and Drug Administration Staff. Available at: <https://www.fda.gov/media/87586/download>.

1509 FDA (2021a). Model-Informed Drug Development Pilot Program. FDA. Available at: <https://www.fda.gov/drugs/development-resources/model-informed-drug-development-pilot-program> (Accessed October 24, 2021).

1510 FDA (2021b). Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions. FDA. Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/assessing-credibility-computational-modeling-and-simulation-medical-device-submissions> (Accessed January 2, 2022).

1511 Folk, M., Heber, G., Koziol, Q., Pourmal, E., and Robinson, D. (2011). *An Overview of the HDF5 Technology Suite and its Applications AD '11*. Uppsala, Sweden: Association for Computing Machinery, 36–47. doi:10.1145/1966895.1966900

1512 Forgoston, E., Bianco, S., Shaw, L. B., and Schwartz, I. B. (2011). Maximal Sensitive Dependence and the Optimal Path to Epidemic Extinction. *Bull. Math. Biol.* 73, 495–514. doi:10.1007/s11538-010-9537-0

1513 Forgoston, E., and Moore, R. O. (2018). A Primer on Noise-Induced Transitions in Applied Dynamical Systems. *SIAM Rev.* 60, 969–1009. doi:10.1137/17M1142028

1514 Fujimoto, R., Bock, C., Chen, W., Page, E., and Panchal, J. H. (2017). *Research Challenges in Modeling and Simulation for Engineering Complex Systems*. Cham. Springer International Publishing. Q19 1544

1515 Galluppi, G. R., Brar, S., Caro, L., Chen, Y., Frey, N., Grimm, H. P., et al. (2021). Industrial Perspective on the Benefits Realized From the FDA's Model-Informed Drug Development Paired Meeting Pilot Program. *Clin. Pharmacol. Ther.* 110, 1172–1175. doi:10.1002/cpt.2265

1516 Garfinkel, S. L., and Leclerc, P. (2020). Randomness Concerns When Deploying Differential Privacy. doi:10.1145/3411497.3420211 Q20 1549

1517 Garira, W. (2017). A Complete Categorization of Multiscale Models of Infectious Disease Systems. *J. Biol. Dyn.* 11, 378–435. doi:10.1080/17513758.2017.1367849

1518 Garira, W. (2018). A Primer on Multiscale Modelling of Infectious Disease Systems. *Infect. Dis. Model.* 3, 176–191. doi:10.1016/j.idm.2018.09.005

1519 Gary, T., Mingle, D., and Yenamandra, A. (2021). The Evolving Definition of Sepsis. Available at: <https://arxiv.org/ftp/arxiv/papers/1609/1609.07214.pdf>.

1520 Gerkin, R. C., Birgiolas, J., Jarvis, R. J., Omar, C., and Crook, S. M. (2019). NeuronUnit: A Package for Data-Driven Validation of Neuron Models Using SciUnit. Available at: <https://www.biorxiv.org/content/10.1101/665331v1> (Accessed October 24, 2021).

1521 Hairer, E., Wanner, G., and Lubich, C. (2006). *Geometric Numerical Integration*. Berlin/Heidelberg: Springer-Verlag.

1522 Halter, M., Patterson, E., Baas, A., and Fairbanks, J. (2020). Compositional Scientific Computing with Matlab and SemanticModels. arXiv:2005.04831 [cs, Math]. Available at: <http://arxiv.org/abs/2005.04831> (Accessed October 24, 2021).

1523 Handel, A., Longini, I. M., and Antia, R. (2007). Neuraminidase Inhibitor Resistance in Influenza: Assessing the Danger of Its Generation and Spread. *Plos Comput. Biol.* 3, e240. doi:10.1371/journal.pcbi.0030240

1524 Hart, W. S., Maini, P. K., and Thompson, R. N. (2020). High Infectiousness Immediately before COVID-19 Symptom Onset Highlights the Importance of Contact Tracing. *Epidemiology*. doi:10.1101/2020.11.20.20235754 Q21 1567

1525 Hernandez-Vargas, E. A., and Velasco-Hernandez, J. X. (2020). In-host Modelling of COVID-19 in Humans. *Infect. Dis. (except HIV/AIDS)*. doi:10.1101/2020.03.26.20044487

1526 Honorato-Zimmer, R., Millar, A. J., Plotkin, G. D., and Zardilis, A. (2018). Chromar, a Rule-Based Language of Parameterised Objects. *Electron. Notes Theor. Computer Sci.* 335, 49–66. doi:10.1016/j.entcs.2018.03.008

1527 Hooker, K. L., and Ganusov, V. V. (2021). Impact of Oseltamivir Treatment on Influenza A and B Virus Dynamics in Human Volunteers. *Front. Microbiol.* 12, 631211. doi:10.3389/fmicb.2021.631211

1528 IMAGWiki (2021). MODELS, TOOLS & DATABASES. Available at: <https://www.imagwiki.nibib.nih.gov/resources/models-tools-databases>.

1529 Interagency Modeling and Analysis Group (2021). Available at: <https://www.imagwiki.nibib.nih.gov/> (Accessed October 24, 2021).

1530 Issues with regard to Call for transparency of COVID-19 models (2021). CoMSES Net Discourse Forums. Available at: <https://forum.comses.net/t/issues-with-regard-to-call-for-transparency-of-covid-19-models/8433> (Accessed October 24, 2021).

1531 Jakeman, A. J., Letcher, R. A., and Norton, J. P. (2006). Ten Iterative Steps in Development and Evaluation of Environmental Models. *Environ. Model. Softw.* 21, 602–614. doi:10.1016/j.envsoft.2006.01.004

1532 Jung, J., and Sugita, Y. (2017). Multiple Program/multiple Data Molecular Dynamics Method with Multiple Time Step Integrator for Large Biological Systems. *J. Comput. Chem.* 38, 1410–1418. doi:10.1002/jcc.24511

1533 Ke, R., Zitzmann, C., Ribeiro, R. M., and Perelson, A. S. (2020). Kinetics of SARS-CoV-2 Infection in the Human Upper and Lower Respiratory Tracts and Their Relationship with Infectiousness. *Infect. Dis. (except HIV/AIDS)*. doi:10.1101/2020.09.25.20201772

1534 Kermack, W. O., McKendrick, A. G., and Walker, G. T. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proc. R. Soc. Lond. A*. 115, 700–721. doi:10.1098/rspa.1927.0118

1535 Khain, E., Khasin, M., and Sander, L. M. (2014). Spontaneous Formation of Large Clusters in a Lattice Gas above the Critical point. *Phys. Rev. E*. 90, 062702. doi:10.1103/PhysRevE.90.062702

- 1597 Kirouac, D. C., Cicali, B., and Schmidt, S. (2019). Reproducibility of Quantitative
1598 Systems Pharmacology Models: Current Challenges and Future Opportunities.
1599 *CPT Pharmacometrics Syst. Pharmacol.* 8, 205–210. doi:10.1002/psp4.12390
- 1600 Kitano, H. (2021). The Standard Graphical Notation for Biological Networks.
1601 Available at: [https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.541.
1602 4169&rep=rep1&type=pdf](https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.541.4169&rep=rep1&type=pdf).
- 1603 Krafczyk, M., Shi, A., Bhaskar, A., Marinov, D., and Stodden, V. (2019). *Scientific
1604 Tests and Continuous Integration Strategies to Enhance Reproducibility in the
1605 Scientific Software Context P-RECS '19*. Phoenix, AZ, USA: Association for
1606 Computing Machinery, 23–28. doi:10.1145/3322790.3330595
- 1607 Kuemmel, C., Yang, Y., Zhang, X., Florian, J., Zhu, H., Tegenge, M., et al. (2020).
1608 Consideration of a Credibility Assessment Framework in Model-Informed
1609 Drug Development: Potential Application to Physiologically-Based
1610 Pharmacokinetic Modeling and Simulation. *CPT Pharmacometrics Syst.
1611 Pharmacol.* 9, 21–28. doi:10.1002/psp4.12479
- 1612 Licensing issues (2021). MSM Viral Pandemic Integration Subgroup Mailing List
1613 Archives. Available at: [https://lists.simtk.org/pipermail/vp-integration-
1614 subgroup/2021-January/000022.html](https://lists.simtk.org/pipermail/vp-integration-subgroup/2021-January/000022.html).
- 1615 Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P.,
1616 et al. (2020). MEMOTE for Standardized Genome-Scale Metabolic Model
1617 Testing. *Nat. Biotechnol.* 38, 272–276. doi:10.1038/s41587-020-0446-y
- 1618 Liu, Q.-H., Ajelli, M., Aleta, A., Merler, S., Moreno, Y., and Vespignani, A. (2018).
1619 Measurability of the Epidemic Reproduction Number in Data-Driven Contact
1620 Networks. *Proc. Natl. Acad. Sci. USA.* 115, 12680–12685. doi:10.1073/pnas.
1621 181115115
- 1622 Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier,
1623 A., et al. (2020). BioModels-15 Years of Sharing Computational Models in Life
1624 Science. *Nucleic Acids Res.* 48, D407–D415. doi:10.1093/nar/gkz1055
- 1625 Meyer, M. (2014). Continuous Integration and Its Tools. *IEEE Softw.* 31, 14–16.
1626 doi:10.1109/ms.2014.58
- 1627 Miao, H., Xia, X., Perelson, A. S., and Wu, H. (2011). On Identifiability of
1628 Nonlinear ODE Models and Applications in Viral Dynamics. *SIAM Rev.* 53,
1629 3–39. doi:10.1137/090757009
- 1630 Multiscale Modeling and Viral Pandemics (2021). Available at: [https://www.
1631 imagwiki.nibib.nih.gov/working-groups/multiscale-modeling-and-viral-
1632 pandemics](https://www.imagwiki.nibib.nih.gov/working-groups/multiscale-modeling-and-viral-pandemics).
- 1633 Murray, J. D. (2002). *Mathematical Biology: I. An Introduction*. New York, NY:
1634 Springer.
- 1635 Musuamba, F. T., Skottheim Rusten, I., Lesage, R., Russo, G., Bursi, R., Emili, L.,
1636 et al. (2021). Scientific and Regulatory Evaluation of Mechanistic In Silico Drug
1637 and Disease Models in Drug Development: Building Model Credibility. *CPT
1638 Pharmacometrics Syst. Pharmacol.* 10, 804–825. doi:10.1002/psp4.12669
- 1639 NASA (2016). NASA-STD-7009 | NASA Technical Standards System (NTSS).
1640 Available at: [https://standards.nasa.gov/standard/nasa/nasa-std-7009
1641 \(Accessed October 24, 2021\)](https://standards.nasa.gov/standard/nasa/nasa-std-7009).
- 1642 NASA (2019). Nasa Handbook For Models And Simulations: An Implementation
1643 Guide For Nasa-Std-7009a. Available at: [https://standards.nasa.gov/sites/
1644 default/files/standards/NASA/PUBLISHED/A/nasa-hdbk-7009a.pdf](https://standards.nasa.gov/sites/default/files/standards/NASA/PUBLISHED/A/nasa-hdbk-7009a.pdf).
- 1645 Neal, M. L., Thompson, C. T., Kim, K. G., James, R. C., Cook, D. L., Carlson, B. E.,
1646 et al. (2019a). SemGen: a Tool for Semantics-Based Annotation and
1647 Composition of Biosimulation Models. *Bioinformatics.* 35, 1600–1602.
1648 doi:10.1093/bioinformatics/bty829
- 1649 Neal, M. L., König, M., Nickerson, D., Misırlı, G., Kalbasi, R., Dräger, A., et al.
1650 (2019b). Harmonizing Semantic Annotations for Computational Models in
1651 Biology. *Brief. Bioinform.* 20, 540–550. doi:10.1093/bib/bby087
- 1652 Nieddu, G. T., Billings, L., Kaufman, J. H., Forgoston, E., and Bianco, S. (2017).
1653 Extinction Pathways and Outbreak Vulnerability in a Stochastic Ebola Model.
1654 *J. R. Soc. Interf.* 14, 20160847. doi:10.1098/rsif.2016.0847
- 1655 Ovaskainen, O., and Meerson, B. (2010). Stochastic Models of Population
1656 Extinction. *Trends Ecol. Evol.* 25, 643–652. doi:10.1016/j.tree.2010.07.009
- 1657 Paradis, E. G., Pinilla, L. T., Holder, B. P., Abed, Y., Boivin, G., and Beauchemin, C.
1658 A. A. (2015). Impact of the H275Y and I223V Mutations in the Neuraminidase
1659 of the 2009 Pandemic Influenza Virus *In Vitro* and Evaluating Experimental
1660 Reproducibility. *PLoS ONE.* 10, e0126115. doi:10.1371/journal.pone.0126115
- 1661 Parker, P., Letcher, R., Jakeman, A., Beck, M. B., Harris, G., Argent, R. M., et al.
1662 (2002). *Progress in Integrated Assessment and modellingA Summary of a
1663 Workshop on Integrated Assessment and Modelling, Held at EcoSummit
1664 2000*. Halifax: Integrating the Sciences. June 18–22, 2000. See Costanza and
1665 Jorgensen (2001) for a further report on Ecosummit.1. *Environmental
1666 Modelling & Software* 17, 209–217. 10.1016/S1364-8152(01)00059-7.
- 1667 Perelson, A. S., Rong, L., and Hayden, F. G. (2012). Combination Antiviral Therapy
1668 for Influenza: Predictions From Modeling of Human Infections. *J. Infect. Dis.*
1669 205, 1642–1645. doi:10.1093/infdis/jis265
- 1670 Petersen, B. K., Ropella, G. E., and Hunt, C. A. (2014). Toward Modular Biological
1671 Models: Defining Analog Modules Based on Referent Physiological
1672 Mechanisms. *BMC Syst. Biol.* 8, 95. doi:10.1186/s12918-014-0095-1
- 1673 Petrie, S. M., Guarnaccia, T., Laurie, K. L., Hurt, A. C., McVernon, J., and McCaw,
1674 J. M. (2013). Reducing Uncertainty in Within-Host Parameter Estimates of
1675 Influenza Infection by Measuring Both Infectious and Total Viral Load. *PLOS
1676 ONE.* 8, e64098. doi:10.1371/journal.pone.0064098
- 1677 Powathil, G. G., Gordon, K. E., Hill, L. A., and Chaplain, M. A. J. (2012). Modelling
1678 the Effects of Cell-Cycle Heterogeneity on the Response of a Solid Tumour to
1679 Chemotherapy: Biological Insights from a Hybrid Multiscale Cellular
1680 Automaton Model. *J. Theor. Biol.* 308, 1–19. doi:10.1016/j.jtbi.2012.05.015
- 1681 Recent SDO/COMBINE legal entity issues (2021). Available at: [https://groups.
1682 google.com/g/sbml-discuss/c/8-ljMxJkMqA](https://groups.google.com/g/sbml-discuss/c/8-ljMxJkMqA).
- 1683 Refsgaard, J. C., Henriksen, H. J., Harrar, W. G., Scholten, H., and Kassahun, A.
1684 (2005). Quality Assurance in Model Based Water Management - Review of
1685 Existing Practice and Outline of New Approaches. *Environ. Model. Softw.* 20,
1686 1201–1215. doi:10.1016/j.envsoft.2004.07.006
- 1687 Russo, G., Di Salvatore, V., Sgroi, G., Parasiliti Palumbo, G. A., and Pappalardo, F.
1688 (2021). A Multi-step and Multi-Scale Bioinformatic Protocol to Investigate
1689 Potential SARS-CoV-2 Vaccine Targets. *Brief Bioinform.* 23, bbab403. doi:10.
1690 1093/bib/bbab403
- 1691 Russo, G., Pennisi, M., Fichera, E., Motta, S., Raciti, G., Viceconti, M., et al. (2020).
1692 In Silico trial to Test COVID-19 Candidate Vaccines: a Case Study with UISS
1693 Platform. *BMC Bioinformatics.* 21, 527. doi:10.1186/s12859-020-03872-0
- 1694 Sarma, G. P., Jacobs, T. W., Watts, M. D., Ghayoomie, S. V., Larson, S. D., and
1695 Gerkin, R. C. (2016). Unit Testing, Model Validation, and Biological
1696 Simulation. *F1000Research*. Available at: [https://f1000research.com/articles/
1697 5-1946](https://f1000research.com/articles/5-1946) (Accessed October 24, 2021).
- 1698 Satyanarayan, A., Moritz, D., Wongsuphasawat, K., and Heer, J. (2017). Vega-lite:
1699 A Grammar of Interactive Graphics. *IEEE Trans. Vis. Comput. Graphics.* 23,
1700 341–350. doi:10.1109/tvcg.2016.2599030
- 1701 Satyanarayan, A., Wongsuphasawat, K., and Heer, J. (2014). *Declarative Interaction
1702 Design for Data Visualization UIST '14*. Honolulu, Hawaii, USA: Association
1703 for Computing Machinery.
- 1704 Schmiester, L., Schälte, Y., Bergmann, F. T., Camba, T., Dudkin, E., Egert, J., et al. (2021).
1705 PETab-Interoperable Specification of Parameter Estimation Problems in Systems
1706 Biology. *Plos Comput. Biol.* 17, e1008646. doi:10.1371/journal.pcbi.1008646
- 1707 Schwartz, I. B., Forgoston, E., Bianco, S., and Shaw, L. B. (2011). Converging
1708 towards the Optimal Path to Extinction. *J. R. Soc. Interf.* 8, 1699–1707. doi:10.
1709 1098/rsif.2011.0159
- 1710 Segovia-Juarez, J. L., Ganguli, S., and Kirschner, D. (2004). Identifying Control
1711 Mechanisms of Granuloma Formation during *M. tuberculosis* Infection Using an
1712 Agent-Based Model. *J. Theor. Biol.* 231, 357–376. doi:10.1016/j.jtbi.2004.06.031
- 1713 Shaikh, B., Marupilla, G., Wilson, M., Blinov, M. L., Moraru, I. L., and Karr, J. R.
1714 (2021). runBioSimulations: an Extensible Web Application that Simulates a
1715 Wide Range of Computational Modeling Frameworks, Algorithms, and
1716 Formats. Available at: [https://www.biorxiv.org/content/10.1101/2021.03.05.
1717 433787v1](https://www.biorxiv.org/content/10.1101/2021.03.05.433787v1) (Accessed October 24, 2021).
- 1718 SimTK (2021). Welcome (Online). Available at: <https://simtk.org/>(Accessed
1719 October 24, 2021).
- 1720 Stack, J. C., Bansal, S., Kumar, V. S. A., and Grenfell, B. (2013). Inferring
1721 Population-Level Contact Heterogeneity from Common Epidemic Data.
1722 *J. R. Soc. Interf.* 10, 20120578. doi:10.1098/rsif.2012.0578
- 1723 Systems Biology Markup Language (SBML) (2021). Available at: <http://sbml.org/>.
- 1724 Tiwari, K., Kananathan, S., Roberts, M. G., Meyer, J. P., Sharif Shohan, M. U.,
1725 Xavier, A. P., et al. (2021). Reproducibility in Systems Biology Modelling. *Mol.
1726 Syst. Biol.* 17 (2), e9982. doi:10.15252/msb.20209982
- 1727 Viceconti, M., Emili, L., Afshari, P., Courcelles, E., Curreli, C., Famaey, N., et al.
1728 (2021). Possible Contexts of Use for In Silico Trials Methodologies: A
1729 Consensus-Based Review. *IEEE J. Biomed. Health Inform.* 25 (10),
1730 3977–3982. doi:10.1109/jbhi.2021.3090469
- 1731 Viceconti, M., Juárez, M. A., Curreli, C., Pennisi, M., Russo, G., and Pappalardo, F.
1732 (2019). POSITION PAPER: Credibility of In Silico Trial Technologies: A
1733 1710

Theoretical Framing. arXiv:1909.04660 [cs, Q-Bio]. Available at: <http://arxiv.org/abs/1909.04660> (Accessed October 24, 2021).

Waites, W., Cavaliere, M., Manheim, D., Panovska-Griffiths, J., and Danos, V. (2021). Rule-based Epidemic Models. *J. Theor. Biol.* 530, 110851. doi:10.1016/j.jtbi.2021.110851

Waltemath, D., Adams, R., Bergmann, F. T., Hucka, M., Kolpakov, F., Miller, A. K., et al. (2011). Reproducible Computational Biology Experiments with SED-ML - The Simulation Experiment Description Markup Language. *BMC Syst. Biol.* 5, 198. doi:10.1186/1752-0509-5-198

Wang, L., Chen, J., and Marathe, M. (2019). DEFSI: Deep Learning Based Epidemic Forecasting with Synthetic Information. *Aaai.* 33, 9607–9612. doi:10.1609/aaai.v33i01.33019607

Wang, S., Pan, Y., Wang, Q., Miao, H., Brown, A. N., and Rong, L. (2020). Modeling the Viral Dynamics of SARS-CoV-2 Infection. *Math. Biosciences.* 328, 108438. doi:10.1016/j.mbs.2020.108438

Wickham, H. (2011). ggplot2. *Wires Comp. Stat.* 3, 180–185. doi:10.1002/wics.147

Wikipedia (2021a). International Classification of Diseases (Online). Available at: https://en.wikipedia.org/wiki/International_Classification_of_Diseases.

Wikipedia (2021b). License Compatibility (Online). Available at: https://en.wikipedia.org/w/index.php?title=License_compatibility&oldid=1051142395 (Accessed October 24, 2021).

Wikipedia (2021c). Public-domain-equivalent License (Online). Available at: https://en.wikipedia.org/w/index.php?title=Public-domain-equivalent_license&oldid=1051493543 (Accessed October 24, 2021).

Zarnitsyna, V. I., Gianlupi, J. F., Hagar, A., Segó, T., and Glazier, J. A. (2021). Advancing Therapies for Viral Infections Using Mechanistic Computational Models of the Dynamic Interplay between the Virus and Host Immune Response. *Curr. Opin. Virol.* 50, 103–109. doi:10.1016/j.coviro.2021.07.007

Zhao, Y., Serebrenik, A., Zhou, Y., Filkov, V., and Vasilescu, B. (2017). “The Impact of Continuous Integration on Other Software Development Practices: A Large-Scale Empirical Study,” in 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE), 60–71.

Zineh, I. I. (2019). Quantitative Systems Pharmacology: A Regulatory Perspective on Translation. *CPT Pharmacometrics Syst. Pharmacol.* 8, 336–339. doi:10.1002/psp4.12403

Conflict of Interest: AK was employed by the Novartis SA.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Forgoston, Sheriff, Osborne, Gonzalez-Parra, Forgoston, Bowness, Liu, Thompson, Garira, Barhak, Rice, Torres, Dobrovolny, Tang, Waites, Glazier, Faeder and Kulesza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
Q16 1778
1779
1780
1781
1782
1783
1784
1785
Q25 1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824

APPENDIX–TRACEABILITY

Draft versions of this paper are available online at:

- 1825 [https://docs.google.com/document/d/1voUSrSpv3AZIC1T-](https://docs.google.com/document/d/1voUSrSpv3AZIC1T-BLa3W4wzHQ5vEdJCVrBbwMUTDiQ/edit?usp=sharing)
- 1826 [BLa3W4wzHQ5vEdJCVrBbwMUTDiQ/edit?usp=sharing](https://docs.google.com/document/d/1voUSrSpv3AZIC1T-BLa3W4wzHQ5vEdJCVrBbwMUTDiQ/edit?usp=sharing)
- 1827 [https://docs.google.com/document/d/1IMEgmdNkx-](https://docs.google.com/document/d/1IMEgmdNkx-EsnOjGuegpenSIMmKIkK00Lc8Gred3QxM/edit?usp=sharing)
- 1828 [EsnOjGuegpenSIMmKIkK00Lc8Gred3QxM/edit?usp=sharing](https://docs.google.com/document/d/1IMEgmdNkx-EsnOjGuegpenSIMmKIkK00Lc8Gred3QxM/edit?usp=sharing)
- 1829
- 1830
- 1831
- 1832
- 1833
- 1834
- 1835
- 1836
- 1837
- 1838
- 1839
- 1840
- 1841
- 1842
- 1843
- 1844
- 1845
- 1846
- 1847
- 1848
- 1849
- 1850
- 1851
- 1852
- 1853
- 1854
- 1855
- 1856
- 1857
- 1858
- 1859
- 1860
- 1861
- 1862
- 1863
- 1864
- 1865
- 1866
- 1867
- 1868
- 1869
- 1870
- 1871
- 1872
- 1873
- 1874
- 1875
- 1876
- 1877
- 1878
- 1879
- 1880
- 1881

- 1882 <https://docs.google.com/document/d/1Ag4ipuybjtthxgV0YjXqYP7AwwNSYcWh/edit>
- 1883 https://docs.google.com/document/d/1VvyP3YZQdQYjj8DFKOpQ4pn_0pdDGgiT/edit
- 1884 [https://drive.google.com/file/d/1U_](https://drive.google.com/file/d/1U_IThrV6STXWNT3GiCepvsLk1WdYgzN5/view)
- 1885 [IThrV6STXWNT3GiCepvsLk1WdYgzN5/view](https://drive.google.com/file/d/1U_IThrV6STXWNT3GiCepvsLk1WdYgzN5/view)
- 1886
- 1887
- 1888 Many Email discussions that constructed this manuscript are
- 1889 made public and can be accessed through:
- 1890 <https://lists.simtk.org/pipermail/vp-integration-subgroup/> **Q23**
- 1891 <https://lists.simtk.org/pipermail/vp-reproduce-subgroup/> **Q24**
- 1892
- 1893
- 1894
- 1895
- 1896
- 1897
- 1898
- 1899
- 1900
- 1901
- 1902
- 1903
- 1904
- 1905
- 1906
- 1907
- 1908
- 1909
- 1910
- 1911
- 1912
- 1913
- 1914
- 1915
- 1916
- 1917
- 1918
- 1919
- 1920
- 1921
- 1922
- 1923
- 1924
- 1925
- 1926
- 1927
- 1928
- 1929
- 1930
- 1931
- 1932
- 1933
- 1934
- 1935
- 1936
- 1937
- 1938