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

By: Multiscale Modeling and Viral Pandemics Working Group

Abstract:

This white paper summarizes work by two subgroups of the Multiscale Modeling and Viral Pandemics Working Group: the model integration subgroup and the model reproducibility, credibility and standardization subgroup. The discussions exposed the reproducibility crisis that leads to inability to reuse and integrate models. The document presents difficulties, and discusses existing efforts towards future solutions that will allow future model utility and integration.

keywords: simulation, reproducibility, crisis, computational modeling, white paper

Introduction - The Promise of Modeling

Starting with Kermack and McKendrick's [1] development of the SIR model in 1927, there has been a long history of population-level epidemiological modelling. 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 stochasticity. Moreover, this wide range of epidemic models has been used to study numerous diseases. Because of the vast amount of knowledge that has been gained through epidemic modelling over approximately the past 100 years, it was possible for researchers to quickly adapt their models to make predictions about the spread and control of the SARS-CoV-2 coronavirus.

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

Some examples like [2,3] use bacteria or virus load as a feature parameter in partial differential equation compartment systems, the progression of the disease among the population is linked to the virus load in infectious individuals.

Within-host models can explore how the variance of biology within the body can impact both a disease and its treatment. Such mathematical models of viruses focussing on the in-host dynamics should be developed and utilised to a greater extent.

Multi-scale within-host modelling is common, with scales from the molecular and cellular levels integrated successfully with the larger whole-organ or whole-body scales [4,5,6]. There are fewer models, however, that successfully combine within-host models with population-level models. One barrier to the development of such models is the potential to lose the within-host granularity that is often seen when integrating to a higher scale. Many researchers may ask the question, do these multi-scale models truly offer new insight or are they more informative analysed independently?

As we move from conceptual models (the diagrams and verbal models of biologists) to mathematical models to computer code, we gain in executability, but we lose shareability. Shareability requires multiple concepts: Reusability--someone else can run the same computer code, perhaps with different initial conditions or parameter values. Extensibility--someone can add modules or replace modules within the model without breaking it. Extractability--someone can select model components and use them, they continue to function independent 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 which 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.

With so many multiscale modelling methods which are seemingly disjoint and mutually exclusive, recent efforts have sought to bring some order in the discussion of multiscale models of pandemics by providing a complete categorizing of them [7,8]. These publications identified five different categories of multiscale models of diseases which 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 modelling of pandemics. The paper [8] 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 modelling of disease dynamics, that require collaborative research among scientists with different skills to be fully resolved.

And 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 [9]. This group is part of the MultiScale Modeling (MSM) Consortium hosted by the Interagency Modeling and Analysis Group (IMAG) [10].

A main goal of the group is to help create the framework that will allow scientists and mathematicians to advance the state-of-the-art of within host viral modelling to an extent so that

when the next major disease outbreak occurs, researchers can quickly adapt their models to better understand the outbreak.

However, upon formation it became evident that there were many topics of interest, so the large group split into smaller subgroups that tackle specific topics. This paper is a product of unifying the work of two sub groups:

1) The Model Reproducibility, Credibility and Standardization Subgroup
Computational biological models promise to capture knowledge in a way that is more scalable, executable, and shareable than human language. Due to the complexity of biology, this will likely require complex, hybrid models that integrate multiple experimental modalities across multiple biological subsystems and scales. Achieving such models will likely require large collaborative teams of computational and experimental scientists.

Despite significant effort, it remains difficult for investigators to compose each other's models. In fact, many published models are not reproducible, even fewer published models are conducive to reuse beyond their original authors, and yet fewer have been rigorously tested beyond the authors' training data. Together, this deters potential stakeholders from being able to confidently use models to help guide critical medical and public health decisions.

One major reason why many models are not reproducible, reusable, and composable is that the existing community specifications are not standardized and best practices have not been universally adopted. For example, many models are not shared publicly, many models are only shared through custom codes, and many models are shared with insufficient documentation to understand the biology they represent or the behaviors they predict. Because the existing community specifications primarily focus on describing the mathematical meaning of specific types of models and simulations, they also cannot capture all types of models, they are not well-suited to capturing all of the semantic and provenance metadata needed to compose models, and they do not address related topics such as the verification of models.

Together, significant effort remains to develop a technical infrastructure for collaboratively building credible models. This subgroup will explore these barriers and recommend potential remedies.

2) The Model Integration Subgroup
The number of computational medical models are increasing and they come in different scales that model molecules, cells, organs, individuals and populations. Those models take many forms, from differential equations, to stochastic models, and machine learning models. This brings great opportunities for collaboration among modelers. One promising way to coordinate such collaboration is to combine submodels of individual subsystems, each developed, calibrated, and tested by a small group of expert investigators. This structure would also enable investigators to leverage the large number of existing models. Harnessing such composition will require infrastructure that enables teams to build, simulate, and test composable models

systematically, scalably, and transparently. Despite the growth and opportunity, there are still many difficulties to integrate them together. This subgroup deals with the following issues:
- Integration Between Within-host and Population Scales
- Integration Within and Across Scales
- Ensemble modeling

The key concept enabling construction of models of complex phenomena is composition [11]. Decomposition lets us break down a complex problem into simpler problems that can be solved or simulated and composition lets us systematically recombine these solutions into a solution of the original problem. To accomplish this, we need to think in terms of higher-order operations on models: what the models are is less important than what can be done with them. There are two kinds of composition: parallel and serial. Parallel composition means running models concurrently. This is useful for ensemble or consensus approaches that combine multiple models to arrive at a best estimate. Serial composition is when the output of one model becomes the input of another. It is important to think about the type of the model, what input it requires and what output it produces because compatibility is required for serial composition. Serial composition has been used to great effect, for example in whole cell models. Of course, it is possible to compose these combined models.

Compositionality and modelling has been extensively studied theoretically and the primitive operations, parallel and serial composition, explored in detail for certain classes of model [12,13] that appear in fields as diverse as electronics engineering, chemistry, molecular biology [14], plant biology [15], infectious diseases [11,16] and economic game theory [17]. However, to address practical problems at scale, infrastructure is required. First, it is necessary to be able to discover models; models cannot be composed if they are unknown or unavailable. To do this, a catalogue is needed with metadata about models and how to obtain them. We note the existence of mature software for data catalogues that is easily repurposed [18]. The models must be described sufficiently well to know if they can be composed, annotated with information about their input and output types. Annotations facilitate auxiliary tasks such as searching for appropriate models and ascertaining provenance. Finally, attention is needed to the detail of composition of a broad class of models, recognising that errors introduced by (de)composition are only well-understood for some cases [19,20].

Merged Interests

The similarity and overlap between the groups as well as overlap in members and their interests lead to identification of a common theme that is presented in Figure 1. The group discussions 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, that leads to inability to combine and integrate larger more complex models. Therefore we address many issues at lower levels that will help reach integration.

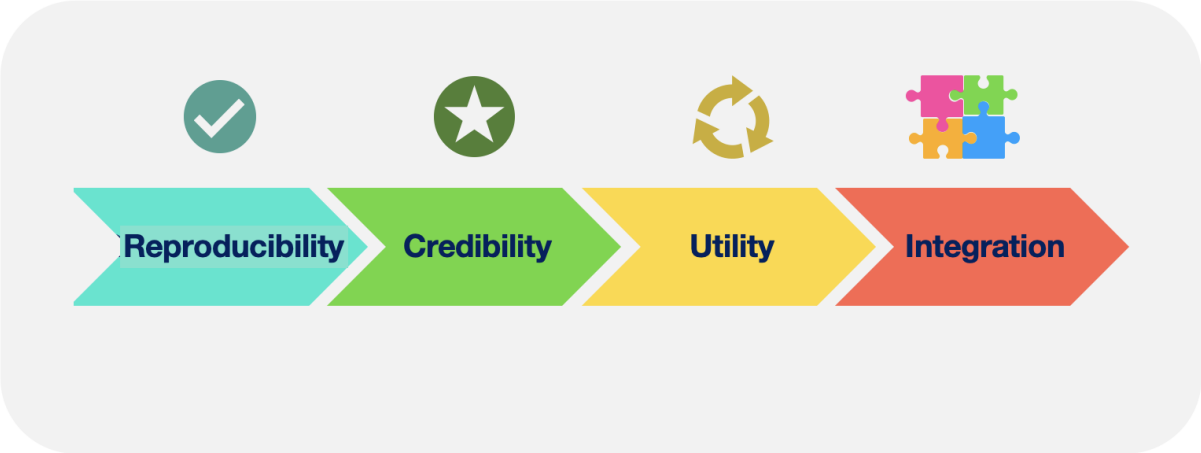

Figure 1: Path towards model integration

The Reproducibility Crisis is a Core Issue:

Computational biomedical modelling involves mathematical representation of biological processes to study complex system behaviour and was expected to be less affected by the reproducibility crisis. However, models often fail to reproduce and the reasons for the failure and prevalence were not fully understood. In a recent study [21], the BioModels group analyzed 455 kinetic models published in 152 peer-reviewed journals, a collective work of about 1400 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, an additional 12% could be reproduced either by empirical correction or support from authors. The other 37% remained non-reproducible due to missing parameter values, missing initial concentration, or inconsistent model structure. Among the corresponding authors contacted <30% responded. Models from many life science journals failed to reproduce, 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.

This shows only part of the larger crisis. The ideal we would like to reach is Repeatability and Reproducibility of models in publications and repositories. Repeatability is the ability to repeat the same experiment with the same system, and Reproducibility is the ability to repeat the same experiment by another scientist without the same system. The goal here is to have all aspects of the simulation pipeline (Biological Model, Mathematical Model, Numerical Methods, Computational Implementation) be auditable, i.e. they should be fully described (as appropriate for each component) to enable results to be repeated and reproduced. However, the modeling community is far from this ideal.

It is important to acknowledge that the situation is much worse for other aspects of modeling. For any other aspect, it would be hard to even conduct such a study to evaluate reproducibility because the information often isn't systematically catalogued let alone shared in a common

format. Examples include how models were constructed, calibrated, or validated. Compared to the software industry that has well established methods of exchanging information and repositories, such as github, computational biological modeling has a long way to go and in this paper we will try to address some topics to be dealt with. If we cannot reproduce models, how can those be considered credible by other models, stakeholders, or even the public?

Credibility of Models

A model should be designed and tested with a specific purpose in mind. Otherwise it cannot be tested by others that can assess its credibility. A model without a purpose is a mere academic exercise suitable for the classroom. A model built for a specific purpose must be repeatable and it is highly recommended it would be reproducible.

When considering reproducibility, it is important to understand the larger context. Even if a model is repeatable in the same environment and even if it can be reproduced by others, it also has to be viewed as credible for use. A model that is not repeatable, cannot be reproduced and is seen as non credible by others who cannot understand the internals if expert modelers cannot reproduce it. For example, a physician will not use a medical device that is not consistent in its functionality since the physician is afraid to harm a patient. Therefore, the proof of model repeatability and reproducibility lies with the modeler that needs to prove the value of the model. However, the modeler cannot assess the credibility of the model - it is in the eyes of the potential customers.

A modeler should consider the model purpose from the start of development and consider: for what, by who, what is the level of knowledge and skill of the user, and in what environment. If only the "for what" is specified, this implies the model can be used by anyone on any system, which broadens the scope and reduces the chance of reproducibility and hence reduces credibility by the potential user.

In most cases, users expect a system based on a model to be accredited somehow before use. This accreditation role many times falls on government agencies such as FDA, or NASA. Those agencies have different approaches towards credibility of models. There are clear examples of what federal agencies require for model use.

NASA takes modeling seriously. After the space shuttle Challenger disaster NASA rewrote a standard [22] and wrote guidelines [23]. An interesting component in the NASA approach was a risk adjusted approach that considered both the probability and consequences of a modeled systems failure, in which the level of risk raises or lowers the bar for the data needed to accredit a model. This approach helps with cost.

FDA has released documents [24,25]. FDA considered the NASA standards when creating [24]. FDA understood the potential value of models and modeled data to make developments of

medical devices in an efficient manner. If a model indeed can make shortcuts and prevent issues related with long and costly clinical trials there may be benefit. Early in the collaboration with the device industry about use of model data in trial processes, FDA suggested having a “library of “reusable”, “regulatory grade” models. 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.

However, FDA and Pharmaceuticals are making progress in their drug development pilot program [26]. Whatever they come up with to make creditable use of bio-science models towards design and towards in-silico contributions to trials. When they formalize their decisions, developers will want to use model data from relevant MSM tools to reduce trial cost. This opens opportunities. However, If academic models don’t meet the FDA requirements to determine credibility of model data, they will not be useful. Industry is very focused on the regulations and will out-compete academics. Therefore, it is important that academic practices improve with regards to repeatability, reproducibility, and reuse. This will create the path towards credibility and reuse.

Despite the importance of developing a model with a purpose 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. So some contributors to this manuscript have a less strict opinion regarding models being developed with a purpose in mind.

An example of a less strict approach to model credibility are new ensemble techniques, such as in [27, 28], 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. 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.

Utility of Models

There is a large variety of known issues that prevent reuse and many solutions, we will try to divide those by topic. For simplicity we will arrange those in a table describing difficulties and possible solutions:

Difficulty	**Potential Solutions**
Evaluating model credibility	Better modeling practices, documentation, tests.
Model validation barriers	No solution suggested
Models are written in different languages	Common transport specifications such as SBML , and proper documentation and annotation
Models are hard to locate	Archive web sites such as: BioModels, SimTK, IMAGWiki, and modeleXchange
Lack of common platforms for executing models and simulations	Platforms such as BioSimulators , runBioSimulations
Modeling requires adaptation towards integration	Tools for composing models such as SBML-Comp
Unit standardization	Standardization efforts, machine learning solutions such as ClinicalUnitMapping.com
Data and measurement definitions	Models that merge human interpretation, and newer measurement devices
Missing annotations in models	Adoption of policies such as COMBINE
Model application and implementation barriers	Education of modelers, users, and the public
Models are not consistently licensed in an easy way that allows reuse	Abandoning old school open source licenses and promoting licenses that release to public domain such as CC0- Creative Commons Zero
Different scales and modeling paradigms	Standardization effort and centralization tools
Stochastic modelling difficulties	Development of tools that guarantee repeatability such as MIST and standards to address stochastic simulations.

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

Evaluating Model Credibility

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 which 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. Examples of information suggested to include are:

- The design decision that motivated a model: what is the model designed to predict?
- What data sources contributed to a model? Ideally, this would be links to data repositories.
- What assumptions were used to interpret this data.
- What methods/tools/users calibrated the model
- Does the model fulfil its intended purpose? [29].
  Reports of these tests which 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.
- 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.
- Model credibility can be enhanced by validation tests against independent data, uncertainty assessments, and peer reviews [30,31].
  Some suggestions emphasize the need for a structured approach with:
  - Unit-test style tests [32,33, 34]
  - Continuous evaluation of such tests similar to continuous integration of software [35,36, 37]

Model Validation Barriers

Barriers preventing validation reduce credibility and therefore have a negative impact on model credibility which prevents reuse. Here are some difficulties:

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.

Model validation for practical prediction: 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, the practical prediction will get more accurate.

Unfortunately the working group did not find potential tools to help remedy this issue partially due to the issues discussed below.

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) [38] 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 model, which SBML and SBGN [39] 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” which 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 the 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 which 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.

One simple integration example [40] 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) [38] 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 [41]. Note that there are many similar community standardization efforts aggregated in the biosimulation modeling community known as COMBINE (Computational Modeling in Biology Network) [42]. COMBINE includes SBML as well as many other specifications, yet those communities are still in the process of standardization and need to organize legally [41]. Nevertheless, the lack of legal governance does not stop communities from developing even more tools for result handling and analysis like PETab [43], SED-ML [44], SESSL [45], KiSAO [46], SBRML [47], HDF5 [48], Vega [49, 50], ggplot2 [51], and others. Those tools show actual needs by the community, on the other hand 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 piece is that the software tools needed for the above are scattered, plus it is often unclear what subset of the above they support, and tools often become inaccessible because they're built by academic groups. Tools need to be submitted to registries and the capabilities need to be annotated.

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

Models are Hard to Locate

Many times model location is a difficult task since models are published in different sources. Despite many repositories available there are many ways models are published including: journal papers, conferences, preprint services such as BioArxiv, web sites, and code repositories such as GitHub. In some good cases, there are model archive/linking web sites such as: BioModels [52], SimTK [53], IMAGWiki [54], and in the future modeleXchange [55]. However, currently there is no one aggregator that helps locate all models and many times community members cannot agree on location and attempt to create more repositories rather than centralize efforts.
Moreover, simulation workflows are even harder to find. For example, BioModels primarily focuses on models. There has been much less focus on publishing the construction/calibration of models, simulations, their results, analyses of their results, or entire workflows for the above. Sharing all of this needs embracing other repositories and developing some new ones.

Lack of Common Platforms for Executing Models and Simulations

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

Especially, if the goal is for non-modelers to be able to interact with models (e.g., to analyze data, to contribute data toward a modeling project, or to apply a model for medicine), it needs to be much easier to find and use these tools. Two initiatives that are trying to address this are BioSimulators [56] and runBioSimulations [57].

Modeling Requires Adaptation Towards Integration

Many times the models as published need some level of manipulation to plug into another model. For example in [58] the survival function needs adaptation to transform it as can be seen

from the public discussion in [59]. Note that all those models need to be scaled to the same units and scales. Another example is in [60] where infectiousness is proportional to max infectiousness while the models in [61] are density models. In the model in [58] the time scale was originally 8 hours and it needed to change to daily probability to merge into another model in [62], which required scaling of the probability function. Those examples are relatively simple integrations and in more complex integrations the adaptation effort is more significant and many more obstacles exist.

One obstacle is lack of standards for describing composite models and software tools for merging models. One specification is SBML-comp, but it i's cumbersome and few tools support it. Another tool is SemGen [63], but it focuses on finding mappings between similar models. To the point here, SBML-comp is designed to compose models that weren't intended to be composed. Instead, composition needs to be deeply ingrained into the entire community so that models are anticipating the needs of composition from the beginning.

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

Unit Standardization

Unfortunately units of measure are not yet standardized an open problem despite many attempts to resolve it by multiple standardization bodies such as IEEE, CDCIC. NIST. One indication of the severity of the problem is that a Github search for "unit conversion" shows over a thousand results. Another good example of the severity of the problem is ClincialTrials.Gov that aggregates quantitative data from around the world and this database shows over 24K different units of measure [64]. One attempt at solving this standardization issue using machine learning is ClinicalUnitMapping.com, yet this project requires more effort.

Data and Measurement Definitions

When attempting to merge models, the phenomenon modeled by the model or the data it is based on may not be the same. This is especially true when model definitions evolve or can be denied in many ways. Examples include International Classification of Disease (ICD) codes [65] that went through multiple versions through the years, or even a disease definition that has evolved for sepsis [66]. Even outcomes of clinical trials change if counted using different definitions as seen in [67]. 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 [68].

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 [69]. 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 [70], 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 [71].

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 [72]. In this paper, we describe 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.

And note that annotations can be useful for multiple tasks such as:
- Annotation of semantic meaning
- Annotation of provenance
- Annotation of verification

However, despite the intention, there is a lack of use of annotations as discussed below:

Lack of sufficient annotation about the components of models: This is particular 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.

Lack of annotation about data sources and assumptions: Lack of annotation 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 [73]. Therefore, model creators may not be aware of the implications of licensing many times, even if they publish their models. Moreover, some licenses are incompatible with each other or other forms of Intellectual Property (IP) such as patents [74] . Even open source licenses are quite restricted since they are based on copyright laws which give the owner rights to restrict usage [75]. 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 [76]. Specifically, one license that will make reuse much easier is Creative Commons Zero (CC0) [77]. 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 [55]. However, CC0 license has not been adopted by some [78].

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.

Different Scales and Modeling Paradigms

Models are operating on different spatial scales (population or individual) with different modelling paradigms (continuous vs discrete):

A tissue could be modelled as a continuum leading to Partial Differential Equations (PDE)s 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.

Models are difficult to implement 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 Modelling 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 ordinary differential equation models have been

developed to understand viral dynamics at the cellular level as well as disease dynamics at the population level [79]. 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 [80]. For example, disease population models typically 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 [81,82,83,84,85,86,87,88].

As an alternative, one can employ stochastic modelling approaches which allow one to make quantitative, statistical predictions, while simultaneously providing qualitative descriptions of system dynamics. Moreover, another major advantage in considering stochastic models lies in their ability to capture specific dynamics observed in nature. Deterministic models are based on mean behavior and do not account for the random interactions of cells or individuals, nor do they account for the changes in growth/death rates or interaction rates related to random events. While the ability to generate stochastic simulations that provide quantitative statistics for the emergence of new dynamics is increasing with advances in computational power, there remains a need for new methods to analyze the underlying stochastic models [80].

Recent years have seen an increase in the use of stochastic systems to model a wide variety of biological phenomena, including subcellular processes and tissue dynamics [89], large-scale population dynamics [90], and genetic switching [91]. One often sees rare transition events in these systems that are induced by noise which may be internal or external to the system. These noise-induced rare events may be associated with a desirable outcome, such as the extinction of an infectious disease outbreak [81,83] or eradication of a pest [92], or an undesirable outcome, such as the sudden clustering of cancerous cells [93], or the outbreak of an infectious disease [87].

In these stochastic systems, noise can affect the system in a variety of ways. Assessing the full impact of noise is rarely possible, and therefore analysis and computations often concentrate on the most important noise-induced events, which include spontaneous switching between coexisting stable states or escape/extinction from a stable state. One important feature of interest when studying noise-induced transitions is the optimal transition pathway of escape from a metastable state either to another metastable state or to a stable (absorbing) state. The optimal path is the path that is most likely to occur among all possible paths, recognizing that this path may not be unique. Knowledge of the optimal path enables the computation of the mean switching time between states or the mean time to exit/extinction [80].

Beyond understanding extinction of a viral infection within the host or the population level extinction of a disease outbreak, it is important to consider stochastic effects to understand the onset of infection. For example, work on HIV transmission has suggested that most sexually transmitted infections are started by a single virus or infected cell. This observation coupled with the fact that successful HIV transmission only occurs in 1 per 100 to 1 per 1000 coital acts suggests that early events in infection are stochastic [94]. In a similar manner, to understand the vulnerability of a population to a zoonotic spillover event, one should consider a stochastic epidemic model. The latter, of course, is crucially important to understanding how a virus leaves

an infected individual to the external environment, and then onto causing an infection in a susceptible individual.

Despite the importance of stochastic models, they present difficulties that include:

- How one validates a stochastic simulation
- How to ensure stochastic simulation repeatability - this problem increases when software libraries that support modern GPU computation hardware that is supposed to accelerate simulation, cannot guarantee deterministic compassion [95]

Potential solutions include development of tools that guarantee repeatability such as MIST [96] and developing standards to address stochastic simulations.

Open Discussion Issues

During the work on the paper, a few other items were raised and we assembled those here. Here are some open ended topics that were raised yet not discussed:

- How should the various component standards be architected to facilitate integration?
Many of the solutions discussed will require dedicated tools and standards and there are many of those. It is unclear how distributed and decentralized components will be governed. A resource such as modeleXchange [15] may help investigators navigate this landscape for models, and similar tools may be needed to navigate tools and standards.

- Common formats for results and visualizations.
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.

- 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.

- Is common governance needed beyond the current COMBINE [42] organization. This organization handles standardization efforts, however, it is not a Standards Development Organization (SDO) and its members rejected joining SISO that 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?

- How should models be tested?
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.

- Spatial models
It is really important to say that 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, tissue architecture. In many cases we don't even have a qualitative language to do this.

- Next step forward
How existing Multi-Scale frameworks could be made more transparent with respect to the models they encode and potentially more interoperable:

1. 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?
2. 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).
3. Is it beneficial to leverage existing standards and to develop new ones for the description and implementation of MSM's going forward? Right now, it's 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.

- Other barriers
Bioscience modelers are not alone dealing with utility and reuse related issues. A 2016 report on complex systems engineering challenges [97] 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.

Conclusions
This white paper 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. 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 multiscale viral pandemic working group

recognizes the challenges and admits the current state of modeling needs fixing. Hopefully fixing those issues starting with reproducibility will increase model credibility and will facilitate towards 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.

Contributing Authors:

The multiscale viral pandemic working group collaborated on writing this paper. Contributing authors in no particular order are:

1. Jonathan Karr, Icahn Institute for Data Science and Genomic Technology and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai USA
2. Rahuman Sheriff, The European Bioinformatics Institute (EMBL-EBI), UK
3. James Osborne, School of Mathematics and Statistics, University of Melbourne, Australia
4. Gilberto Gonzalez Parra, Mathematics Department, New Mexico Tech, USA
5. Eric Forgoston, Department of Applied Mathematics and Statistics, Montclair State University, USA
6. Ruth Bowness, Department of Mathematical Sciences, University of Bath, UK
7. Yaling Liu, Department of Mechanical Engineering and Mechanics , Department of Bioengineering, Lehigh University, USA
8. Robin Thompson, Mathematics Institute & The Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, UK
9. Winston Garira - Department Of Mathematics And Applied Mathematics , Modelling Health and Environmental Linkages Research Group. University of Venda, South Africa
10. Jacob Barhak - Barhak, Jacob, USA
11. John Rice - Independent Retired Working Group Volunteer, USA
12. Marcella Torres, School of Arts and Sciences, University of Richmond, USA
13. Hana M. Dobrovolny , Department of Physics & Astronomy, Texas Christian University, Fort Worth, USA
14. Tingting Tang, Department of Mathematics and Statistics in San Diego State University (SDSU) and SDSU Imperial Valley, USA
15. William Waites, Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, UK
16. James Glazier, Biocomplexity Institute, Indiana University, Bloomington, USA
17. James R. Faeder, Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, USA

Original drafts and discussions are available online at:

https://docs.google.com/document/d/1cqwXAjBWEiJZ1tUBnf66QVHdHd2fKq_W0py7t4PNVLo/edit?usp=sharing

https://docs.google.com/document/d/1voUSrSpv3AZlC1T-BLa3W4wzHQ5vEdJCVrBbwMUTDiQ/edit?usp=sharing

https://docs.google.com/document/d/1IMEgmdNkx-EsnOjGuegpenSIMmKIkK00Lc8Gred3QxM/edit?usp=sharing

Email Discussions:

https://lists.simtk.org/pipermail/vp-integration-subgroup/

https://lists.simtk.org/pipermail/vp-reproduce-subgroup/

For correspondence and communications with the group, please send an email to:

vp-reproduce-subgroup@lists.simtk.org

Acknowledgements:
Jonathan Karr was supported by Grant NIH-NIBIB P41EB023912 "Center for Reproducible Biomedical modeling". William Waites is supported by UK Medical Research Council MRC grant MR/V027956/1 .

Conflict of Interest Statement:

Dr. Jacob Barhak: Payment/services info: Dr. Barhak reports non-financial support and other from Rescale, and MIDAS Network, other from Amazon AWS, Microsoft Azure, MIDAS network, other from The COVID tracking project at the Atlantic, other from John Rice and Jered Hodges, other from Anaconda. Financial relationships: Jacob Barhak declare(s) employment from MacroFab, United Solutions, B. Well Connected health and Anaconda. Intellectual property info: Dr. Barhak holds US Patent 9,858,390 - Reference model for disease progression issued to Jacob Barhak, and a US patent 10,923,234 - Analysis and Verification of Models Derived from Clinical Trials Data Extracted from a Database. Other relationships: During the conduct of the study; personal fees from MacroFab, personal fees from United Solutions, personal fees from B. Well Connected health, personal fees and non-financial support from Anaconda, outside the submitted work.

References

1. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character. 1927 Aug 1;115(772):700-21.

2. A.S. Ackleh, K.L. Sutton, T. Tang and L. Zhao, A Second Order Finite Difference Scheme for a Variable Infection-Structured Model of Mycobacterium Marinum Dynamics in Aquatic Animals, Journal of Nonlinear and Variational Analysis, 2(2018), 177-202.
3. A.S. Ackleh, B. Ma, J. Thibodeaux, A Second Order High Resolution Finite Difference Scheme for a Structured Erythropoiesis Model Subject to Malaria Infection, Mathematical Biosciences, 245(2013), 2-11.

4. Powathil GG, et al. Modelling the effects of cell-cycle heterogeneity on the response of a solid tumour to chemotherapy: biological insights from a hybrid multiscale cellular automaton model. J Theor Biol 2012; 308: 1–19.

5. Bowness R, et al. Modelling the effects of bacterial cell state and spatial location on tuberculosis treatment: insights from a hybrid multiscale cellular automaton model. J Theor Biol 2018; 446: 87–100.

6. Segovia-Juarez JL, Ganguli S, Kirschner D. Identifying control mechanisms of granuloma formation during M. tuberculosis infection using an agent-based model. J Theor Biol 2004; 231: 357–376.

7. Garira, W. (2017). A complete categorization of multiscale models of infectious disease systems. Journal of biological dynamics, 11(1), 378-435.

8. Garira, W. (2018). A primer on multiscale modelling of infectious disease systems. Infectious Disease Modelling, 3, 176-191.

9. IMAG: Multiscale Modeling and Viral Pandemics. Online: https://www.imagwiki.nibib.nih.gov/working-groups/multiscale-modeling-and-viral-pandemics

10. IMAG: Interagency Modeling and Analysis Group. Online: https://www.imagwiki.nibib.nih.gov/

11. Halter, Micah, Evan Patterson, Andrew Baas, and James Fairbanks. “Compositional Scientific Computing with Catlab and SemanticModels.” ArXiv:2005.04831 [Cs, Math], June 29, 2020. http://arxiv.org/abs/2005.04831.
12. Baez, John C., and Jade Master. “Open Petri Nets.” Mathematical Structures in Computer Science 30, no. 3 (March 2020): 314–41. https://doi.org/10.1017/S0960129520000043.

13. Baez, John C., Fabrizio Genovese, Jade Master, and Michael Shulman. “Categories of Nets.” ArXiv:2101.04238 [Cs, Math], January 11, 2021. http://arxiv.org/abs/2101.04238.
14. Danos, Vincent, and Cosimo Laneve. “Formal Molecular Biology.” Theoretical Computer Science, Computational Systems Biology, 325, no. 1 (September 28, 2004): 69–110. https://doi.org/10.1016/j.tcs.2004.03.065.
15. Honorato-Zimmer, Ricardo, Andrew J. Millar, Gordon D. Plotkin, and Argyris Zardilis. “Chromar, a Rule-Based Language of Parameterised Objects.” Electronic Notes in Theoretical Computer Science, 7th International Workshop on Static Analysis and Systems Biology (SASB 2016), 335 (April 10, 2018): 49–66. https://doi.org/10.1016/j.entcs.2018.03.008.
16. Waites, William, Matteo Cavaliere, David Manheim, Jasmina Panovska-Griffiths, and Vincent Danos. “Rule-Based Epidemic Models.” ArXiv:2006.12077 [q-Bio], February 26, 2021. http://arxiv.org/abs/2006.12077.
17. Atkey, Robert, Bruno Gavranović, Neil Ghani, Clemens Kupke, Jérémy Ledent, and Fredrik Nordvall Forsberg. “Compositional Game Theory, Compositionally,” 2020. https://pureportal.strath.ac.uk/en/publications/compositional-game-theory-compositionally.
18. CKAN https://ckan.org/
19. Hairer, Ernst, Christian Lubich, and Gerhard Wanner. Geometric Numerical Integration - Structure-Preserving Algorithms for Ordinary Differential Equations, 2006. http://www.springer.com/mathematics/computational+science+%26+engineering/book/978-3-540-30663-4.
20. Blanes, Sergio, Fernando Casas, and Ander Murua. “Splitting and Composition Methods in the Numerical Integration of Differential Equations.” ArXiv:0812.0377 [Math], December 1, 2008. http://arxiv.org/abs/0812.0377.
21. Tiwari, K., Kananathan, S., Roberts, M. G., Meyer, J. P., Shohan, M. U. S., Xavier, A., ... & Sheriff, R. S. M. (2020). Reproducibility in systems biology modelling. bioRxiv.

22. NASA Technical Standards System. Standard for Models and Simulations https://standards.nasa.gov/standard/nasa/nasa-std-7009

23. Office of the NASA Chief Engineer. NASA HANDBOOK FOR MODELS AND SIMULATIONS: AN IMPLEMENTATION GUIDE FOR NASA-STD-7009A https://standards.nasa.gov/sites/default/files/standards/NASA/PUBLISHED/A/nasa-hdbk-7009a.pdf

24. FDA. Reporting of Computational Modeling Studies in Medical Device Submissions. Guidance for Industry and Food and Drug Administration Staff. September 21, 2016

25. FDA. Guidance for Industry Exposure-Response Relationships — Study Design, Data Analysis, and Regulatory Applications. https://www.fda.gov/media/71277/download . April 2003

26. FDA. Model-Informed Drug Development Pilot Program. https://www.fda.gov/drugs/development-resources/model-informed-drug-development-pilot-program

27. J. Barhak, The Reference Model: A Decade of Healthcare Predictive Analytics with Python, PyTexas 2017, Nov 18-19, 2017, Galvanize, Austin TX. Presentation: http://sites.google.com/site/jacobbarhak/home/PyTexas2017_Upload_2017_11_18.pptx Video: https://youtu.be/Pj_N4izLmsI

28. J. Barhak, The Reference Model for Disease Progression Combines Disease Models. I/ITSEC 2016 28 Nov – 2 Dec Orlando Florida. Paper: http://www.iitsecdocs.com/volumes/2016 Presentation: http://sites.google.com/site/jacobbarhak/home/IITSEC2016_Upload_2016_11_05.pptx

29. Parker, P., Letcher, R., Jakeman, A., Beck, M. B., Harris, G., Argent, R. M., ... & Bin, S. (2002). Progress in integrated assessment and modelling. Environmental Modelling & Software, 17(3), 209-217.

30. Refsgaard, J. C., Henriksen, H. J., Harrar, W. G., Scholten, H., & Kassahun, A. (2005). Quality assurance in model based water management–review of existing practice and outline of new approaches. Environmental Modelling & Software, 20(10), 1201-1215.
31. Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software, 21(5), 602-614.

32. Sarma, G. P., Jacobs, T. W., Watts, M. D., Ghayoomie, S. V., Larson, S. D., & Gerkin, R. C. (2016). Unit testing, model validation, and biological simulation. F1000Research, 5.
33. Gerkin, R. C., Birgiolas, J., Jarvis, R. J., Omar, C., & Crook, S. M. (2019). NeuronUnit: A package for data-driven validation of neuron models using SciUnit. bioRxiv, 665331.

34. Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., ... & Zhang, C. (2020). MEMOTE for standardized genome-scale metabolic model testing. Nature biotechnology, 38(3), 272-276.

35. Meyer, M. (2014). Continuous integration and its tools. IEEE software, 31(3), 14-16.

36. Krafczyk, M., Shi, A., Bhaskar, A., Marinov, D., & Stodden, V. (2019, June). Scientific tests and continuous integration strategies to enhance reproducibility in the scientific software context. In Proceedings of the 2nd International Workshop on Practical Reproducible Evaluation of Computer Systems (pp. 23-28).

37. Zhao, Y., Serebrenik, A., Zhou, Y., Filkov, V., & Vasilescu, B. (2017, October). 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) (pp. 60-71). IEEE.

38. Systems Biology Markup Language (SBML), Online: http://sbml.org/
39. Hiroaki Kitano, The Standard Graphical Notation for Biological Networks. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.541.4169&rep=rep1&type=pdf

40. MSM Viral Pandemic Integration Subgroup mailing list archives. Another example of integration and reproduction of a model. https://lists.simtk.org/pipermail/vp-integration-subgroup/2021-January/000012.html

41. Google Groups - SBML discuss. Recent SDO/COMBINE legal entity issues. https://groups.google.com/g/sbml-discuss/c/8-ljMxJkMqA

42. COMBINE. The Computational Modeling in Biology Network . http://co.mbine.org/

43. Schmiester, L., Schälte, Y., Bergmann, F. T., Camba, T., Dudkin, E., Egert, J., ... & Weindl, D. (2021). PEtab—Interoperable specification of parameter estimation problems in systems biology. PLoS computational biology, 17(1), e1008646.

44. Waltemath, D., Adams, R., Bergmann, F. T., Hucka, M., Kolpakov, F., Miller, A. K., ... & Le Novère, N. (2011). Reproducible computational biology experiments with SED-ML-the simulation experiment description markup language. BMC systems biology, 5(1), 1-10.

45. Ewald, R., & Uhrmacher, A. M. (2014). SESSL: A domain-specific language for simulation experiments. ACM Transactions on Modeling and Computer Simulation (TOMACS), 24(2), 1-25.
46. Courtot, M., Juty, N., Knüpfer, C., Waltemath, D., Zhukova, A., Dräger, A., ... & Le Novère, N. (2011). Controlled vocabularies and semantics in systems biology. Molecular systems biology, 7(1), 543.

47. Dada, J. O., Spasić, I., Paton, N. W., & Mendes, P. (2010). SBRML: a markup language for associating systems biology data with models. Bioinformatics, 26(7), 932-938.

48. Folk, M., Heber, G., Koziol, Q., Pourmal, E., & Robinson, D. (2011, March). An overview of the HDF5 technology suite and its applications. In Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases (pp. 36-47).

49. Satyanarayan, A., Wongsuphasawat, K., & Heer, J. (2014, October). Declarative interaction design for data visualization. In Proceedings of the 27th annual ACM symposium on User interface software and technology (pp. 669-678).

50. Satyanarayan, A., Moritz, D., Wongsuphasawat, K., & Heer, J. (2016). Vega-lite: A grammar of interactive graphics. IEEE transactions on visualization and computer graphics, 23(1), 341-350.

51. Wickham, H. (2011). ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics, 3(2), 180-185.

52. BioModels. http://www.ebi.ac.uk/biomodels/

53. SimTK. https://simtk.org/

54. IMAGWiki . MODELS, TOOLS & DATABASES https://www.imagwiki.nibib.nih.gov/resources/models-tools-databases

55. Rahuman S Malik-Sheriff, Mihai Glont, Tung V N Nguyen, Krishna Tiwari, Matthew G Roberts, Ashley Xavier, Manh T Vu, Jinghao Men, Matthieu Maire, Sarubini Kananathan, Emma L Fairbanks, Johannes P Meyer, Chinmay Arankalle, Thawfeek M Varusai, Vincent Knight-Schrijver, Lu Li, Corina Dueñas-Roca, Gaurhari Dass, Sarah M Keating, Young M Park, Nicola Buso, Nicolas Rodriguez, Michael Hucka, Henning Hermjakob, BioModels—15 years of sharing computational models in life science, Nucleic Acids Research, Volume 48, Issue D1, 08 January 2020, Pages D407–D415, https://doi.org/10.1093/nar/gkz1055

56. BioSimulators, Reproducing and Reusing Biomodels and Simulations, Registry of Biosimulation Tools, https://biosimulators.org

57. Shaikh B, Marupilla G, Wilson M, Blinov ML, Moraru II & Karr JR. runBioSimulations: an extensible web application that simulates a wide range of computational modeling frameworks, algorithms, and formats (2021). https://run.biosimulations.org

58. Filippo Castiglione, Debashrito Deb, Anurag P. Srivastava, Pietro Liò, Arcangelo Liso From infection to immunity: understanding the response to SARS-CoV2 through in-silico modeling. bioRxiv 2020.12.20.423670; doi: https://doi.org/10.1101/2020.12.20.423670
59. MSM Viral Pandemic Integration Subgroup mailing list archives. About using a multi-scale mortality model in the ensemble. Online: https://lists.simtk.org/pipermail/vp-integration-subgroup/2021-January/000011.html

60. Ruian Ke, Carolin Zitzmann, Ruy M. Ribeiro, Alan S. Perelson. Kinetics of SARS-CoV-2 infection in the human upper and lower respiratory tracts and their relationship with infectiousness. medRxiv 2020.09.25.20201772; doi: https://doi.org/10.1101/2020.09.25.20201772

61. W.S. Hart, P.K. Maini, R.N. Thompson , High infectiousness immediately before COVID-19 symptom onset highlights the importance of contact tracing. medRxiv 2020.11.20.20235754; doi: https://doi.org/10.1101/2020.11.20.20235754

62. Jacob Barhak Github - COVID-19 mortality model by Filippo Castiglione et. al. https://github.com/Jacob-Barhak/COVID19Models/tree/main/COVID19_Mortality_Castiglione

63. Neal ML, Thompson CT, Kim KG, James RC, Cook DL, Carlson BE, and Gennari JH (2019). SemGen: a tool for semantics-based annotation and composition of biosimulation models. Bioinformatics. doi:10.1093/bioinformatics/bty829.

64. Jacob Barhak, Joshua Schertz, Clinical Unit Mapping with Multiple Standards, 2019 CDISC U.S. Interchange, San Diego, CA, 16 - 17 October 2019. Poster: https://jacob-barhak.github.io/Poster_CDISC2019.html

65. Wikipedia: International Classification of Diseases , Online https://en.wikipedia.org/wiki/International_Classification_of_Diseases

66. Gary T., Mingle D., Yenamandra A.(2016) The Evolving Definition of Sepsis. arXiv:1609.07214v1. https://arxiv.org/ftp/arxiv/papers/1609/1609.07214.pdf

67. ClinicalTrials.gov - NCT00379769: Rosiglitazone Evaluated for Cardiac Outcomes and Regulation of Glycaemia in Diabetes (RECORD) (Online) https://clinicaltrials.gov/ct2/show/results/NCT00379769?view=results

68. Jacob Barhak, The Reference Model for Disease Progression Handles Human Interpretation, MODSIM World 2020. Paper: https://www.modsimworld.org/papers/2020/MODSIM_2020_paper_42_.pdf Interactive Results: https://jacob-barhak.netlify.app/thereferencemodel/results_2020_03_21_visual_2020_03_23/CombinedPlot.html

69. E.G. Paradis and L.T. Pinilla and B.P. Holder and Y. Abed and G. Boivin and C.A.A. Beauchemin (2015). Impact of the {H275Y} and {I223V} Mutations in the Neuraminidase of the 2009 Pandemic Influenza Virus In Vitro and Evaluating Experimental Reproducibility", PLoS ONE 10(5):e0126115 DOI:10.1371/journal.pone.0126115

70. Stephen M Petrie and Teagan Guarnaccia and Karen L Laurie and Aeron C Hurt and Jodie McVernon and James M McCaw (2013). Reducing Uncertainty in Within-Host Parameter Estimates of Influenza Infection by Measuring Both Infectious and Total Viral Load. PLoS One 8(5):e64098 DOI:10.1371/journal.pone.0064098

71. Daniel Cresta, Donald C. Warren, Christian Quirouette, Amanda P. Smith, Lindey C. Lane, Amber M. Smith, Catherine A. A. Beauchemin (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

72. Neal ML, König M, Nickerson D, Mısırlı G, Kalbasi R, Dräger A, Atalag K, Chelliah V, Cooling M, Cook DL, Crook S, de Alba M, Friedman SH, Garny A, Gennari JH, Gleeson P, Golebiewski M, Hucka M, Juty N, Myers C, Olivier BG, Sauro HM, Scharm M, Snoep JL, Touré V, Wipat A, Wolkenhauer O, Waltemath D (2018) Harmonizing semantic annotations for computational models in biology, Briefings in Bioinformatics, Volume 20, Issue 2, March 2019, Pages 540–550, https://doi.org/10.1093/bib/bby087

73. MSM Viral Pandemic Integration Subgroup mailing list archives. Licensing issues. Online: https://lists.simtk.org/pipermail/vp-integration-subgroup/2021-January/000022.html

74. Wikipedia. License Compatibility. https://en.wikipedia.org/wiki/License_compatibility

75. Jacob. Barhak, Open Source and Sustainability, COMBINE 2020 October 5-9. Presentation: https://jacob-barhak.github.io/COMBINE2020_OpenSource_upload_2020_10_04.odp

76. ComSES Net discourse forums. Issues with regard to Call for transparency of COVID-19 models. Online: https://forum.comses.net/t/issues-with-regard-to-call-for-transparency-of-covid-19-models/8433

77. Creative Commons. CC0. Online: https://creativecommons.org/share-your-work/public-domain/cc0/

78. Wikipedia. Public-domain-equivalent license. https://en.wikipedia.org/wiki/Public-domain-equivalent_license

79. Murray JD. Mathematical biology: I. An introduction. Springer Science & Business Media; 2007 Jun 12.

80. Forgoston E, Moore RO. A primer on noise-induced transitions in applied dynamical systems. SIAM Review. 2018;60(4):969-1009.

81. M. Assaf and B. Meerson, Extinction of metastable stochastic populations, Phys. Rev. E, 81 (2010), art. 021116.

82. M. Bauver, E. Forgoston, and L. Billings, Computing the optimal path in stochastic dynamical systems, Chaos, 26 (2016), art. 083101.

83. L. Billings and E. Forgoston, Seasonal forcing in stochastic epidemiology models, Ric. Mat., 67 (2018), pp. 27-47.

84. C. R. Doering, K. V. Sargsyan, and L. M. Sander, Extinction times for birth-death processes: Exact results, continuum asymptotics, and the failure of the Fokker-Planck approximation, Multiscale Model. Simul., 3 (2005), pp. 283-299

85. M. I. Dykman, I. B. Schwartz, and A. S. Landsman, Disease extinction in the presence of random vaccination, Phys. Rev. Lett., 101 (2008), art. 078101.

86. E. Forgoston, S. Bianco, L. B. Shaw, and I. B. Schwartz, Maximal sensitive dependence and the optimal path to epidemic extinction, Bull. Math. Biol., 73 (2011), pp. 495-514.

87. G. T. Nieddu, L. Billings, J. H. Kaufman, E. Forgoston, and S. Bianco, Extinction pathways and outbreak vulnerability in a stochastic Ebola model, J. Roy. Soc. Interface, 14 (2017), art. 20160847.

88. I. B. Schwartz, E. Forgoston, S. Bianco, and L. B. Shaw, Converging towards the optimal path to extinction, J. Roy. Soc. Interface, 8 (2011), pp. 1699-1707.

89. L. S. Tsimring, Noise in biology, Rep. Progr. Phys., 77 (2014), art. 026601.

90. O. Ovaskainen and B. Meerson, Stochastic models of population extinction, Trends Ecol. Evol., 25 (2010), pp. 643-652.

91. M. Assaf, E. Roberts, and Z. Luthey-Schulten, Determining the stability of genetic switches: Explicitly accounting for mRNA noise, Phys. Rev. Lett., 106 (2011), art. 248102.

92. G. Nieddu, L. Billings, and E. Forgoston, Analysis and control of pre-extinction dynamics in stochastic populations, Bull. Math. Biol., 76 (2014), pp. 3122-3137.

93. E. Khain, M. Khasin, and L. M. Sander, Spontaneous formation of large clusters in a lattice gas above the critical point, Phys. Rev. E, 90 (2014), art. 062702.

94. Pearson JE, Krapivsky P, Perelson AS. Stochastic theory of early viral infection: continuous versus burst production of virions. PLoS Comput Biol. 2011 Feb 3;7(2):e1001058.

95. Github Issue: rapidsai / cuml. Clearly indicate deterministic algorithms for reproducible science . online: https://github.com/rapidsai/cuml/issues/2685

96. J. Barhak, MIST: Micro-Simulation Tool to Support Disease Modeling. SciPy, 2013, Bioinformatics track, https://github.com/scipy/scipy2013_talks/tree/master/talks/jacob_barhak , Presentation also available at: http://sites.google.com/site/jacobbarhak/home/SciPy2013_MIST_Presented_2013_06_26.pptx Video of Talk: http://www.youtube.com/watch?v=AD896WakR94

97. Research Challenges in Modeling & Simulation Engineering Complex Systems . Richard Fujimoto Conrad Bock Wei Chen Ernest Page Jitesh H. Panchal (ed): 2017, 100. 10.1007/978-3-319-58544-4