Summary Findings: 2018-2019 Mid-Term Credibility Plan Review

On behalf of the Committee on Credible Practice of Modeling & Simulation in Healthcare

https://simtk.org/home/cpms

cpmsinhealthcare@gmail.com

Mid-term Credibility Plan Review

- Feedback from the 2018 IMAG-MSM meeting identified the need for more consistent communication of model credibility status

- CPMS proposed a "Mid-term Review" of credibility status
  - Designed as an exercise in communicating credibility
    - Not an assessment of credibility level
  - A means to provide feedback in a timely manner, allowing grantees to prepare for the 2019 IMAG-MSM meeting
  - Targeted for those whose grants required credibility plan submission
  - Open to all U01 grantees

Charge to CPMS

Develop and implement a mid-term review process based on the CPMS Ten Simple Rules

Goals:

- Improve the consistency of communicating progress in each credibility plan
- Give grantees the opportunity to practice communicating how credibility relates to the contextual use of their model
- For CPMS: Evaluate and evolve approaches for tools and methods that promote model and simulation credibility communication

Ten Simple Rules

Ten Simple Rules	
R1 - Define context clearly	*R6 - Document adequately*
R2 - Use appropriate data	*R7 - Disseminate broadly*
R3 - Evaluate within context	*R8 - Get independent reviews*
R4 - List limitations explicitly	*R9 - Test competing implementations*
R5 - Use version control	*R10 -Conform to standards*

Ten Simple Rules Cont.

Rule	Description
R1 - Define context clearly	Develop and document the subject, purpose, and intended use(s) of the model or simulation.
R2 - Use appropriate data	Employ relevant and traceable information in the development or operation of a model or simulation.
R3 - Evaluate within context	Verification, validation, uncertainty quantification, and sensitivity analysis of the model or simulation are accomplished with respect to the reality of interest and intended use(s) of the model or simulation.
R4 - List limitations explicitly	Restrictions, constraints, or qualifications for or on the use of the model or simulation are available for consideration by the users or customers of a model or simulation.
R5 - Use version control	Implement a system to trace the time history of M&S activities including delineation of contributors' efforts.

Ten Simple Rules Cont.

Rule	Description
R6 - Document adequately	Maintain up-to-date informative records of all M&S activities, including simulation code, model mark-up, scope and intended use of M&S activities, as well as users’ and developers' guides.
R7 - Disseminate broadly	Publish all components of M&S activities, including simulation software, models, simulation scenarios and results.
R8 - Get independent reviews	Have the M&S activity reviewed by nonpartisan third-party users and developers.
R9 - Test competing implementations	Use contrasting M&S execution strategies to check the conclusions of the different execution strategies against each other.
R10 -Conform to standards	Adopt and promote generally applicable and discipline specific operating procedures, guidelines, and regulations accepted as best practices.

Request to IMAG-MSM U01 Grantees

1. Sent out 8/28/2018; requested submission by 10/1/2018
2. Details regarding Model Credibility Plan following the [CPMS Ten Simple Rules (TSR)]() format
   A. List of Model Credibility Plan actions
   B. Description of information gained by each action
   C. Summary table of activities classified within the CPMS TSR framework
      1. Explicitly state why a factor is not being implemented
   D. Explain how do the planned activities lead to a credible model
   E. Plans for the next reporting cycle (6 months)
   F. Issues/concerns in achieving the standard of credibility
3. Identify other factors that contribute to credibility

Review Process

Collected all Mid-term updates on IMAG wiki by 10/8/2018

After consultation with committee members, CPMS arrived at a review scoring rubric relating the communication of content in each TSR factor

- Sufficiently Described - Path toward evidence of this factor/rule appears to be sufficient
- Insufficiently Described - Path toward evidence of this factor/rule appears to be insufficient
- Not Available - No path toward evidence is described or an argument is made that the credibility factor did not apply to this model

Ancillary evidence and provided development history could be considered in assessing sufficiency of communicating content in each TSR factor

Clarifying Statement

The reviewers <u>DID NOT assess the CREDIBILITY</u> of the described research project. Please do not associate the scoring with the credibility of individual studies.

The reviewers DID:

- Assess the reviewer's opinion on the sufficiency / insufficiency of the communication of plans and accomplishments wrt the CPMS-TSR content
- Identify areas of improvement of the reporting / review process
- Identify credibility topic areas recommended for further discussion in IMAG/MSM community

Participating Reviewers

COMMITTEE EXECUTIVE MEMBERS (EXECUTE & CHARGE)

A. Drach, Ph.D.
Co-Chair
U. of Texas

J. Myers, Ph.D.
Co-Chair
NASA

A. Erdemir, Ph.D.
Past Co-Chair
Cleveland Clinic

L. Mulugeta, M.Sc.
Past Co-Chair
InSilico Labs LLC

W. Lytton, M.D.
Kings County Hosp.
Downstate Med. Cntr

M. Horner, Ph.D.
ANSYS, Inc.

T. Morrison, Ph.D.
FDA

J. Ku, Ph.D.
Stanford U.

R. Vadigepalli, Ph.D.
Thomas Jefferson U.

Review Timeline

10/8/2018 - 10/29/2018 Reviewers self assigned submissions to review with the goal to have 2 reviewers per submission.

Reviewers were not blinded to the submission PI

Reviews were kept in a shared document so reviewers could access all data

Summary scoring statistics updated weekly and open to the reviewers to view

After 10/29/2018 CPMS Co-chairs evaluated the progress and provided reviews to meet a criteria of 100% of submissions with 1 review, and 85% of the submissions with at least 2 reviews (excluding new awardees)

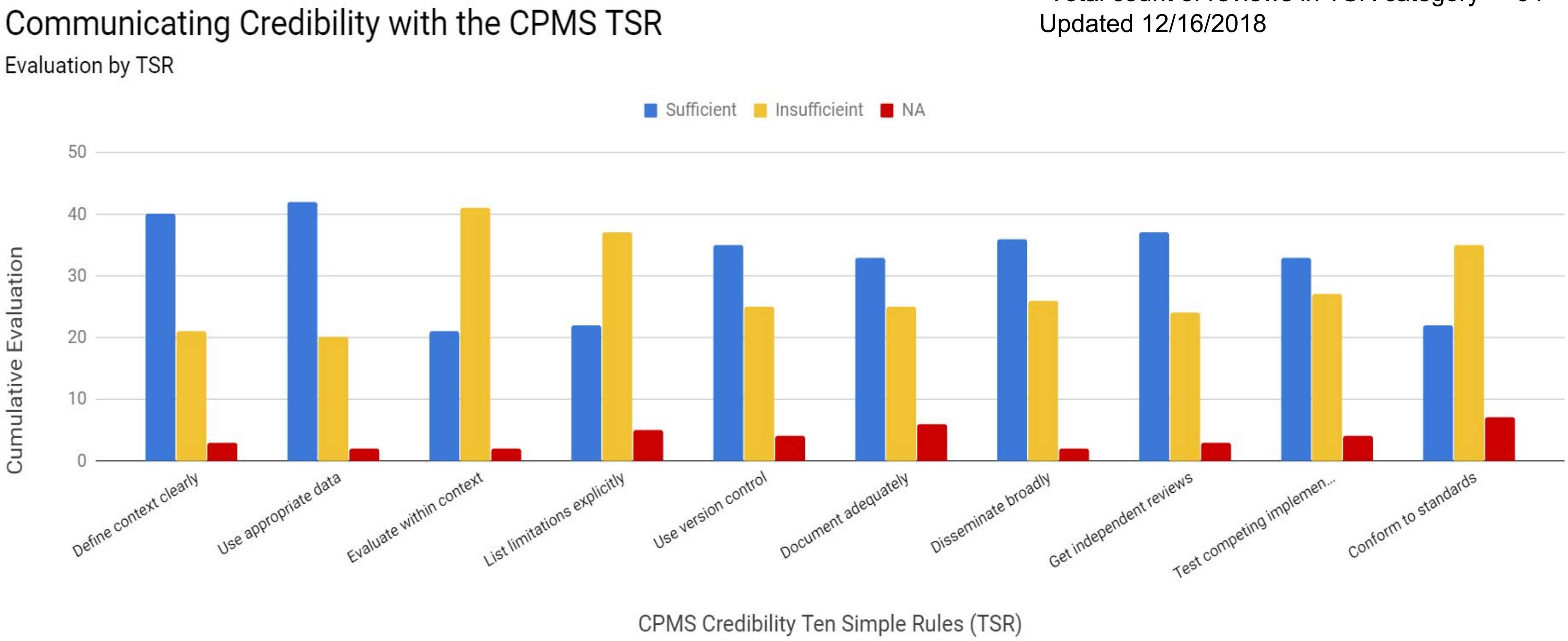
EXERCISE RESULTS

Review Statistics

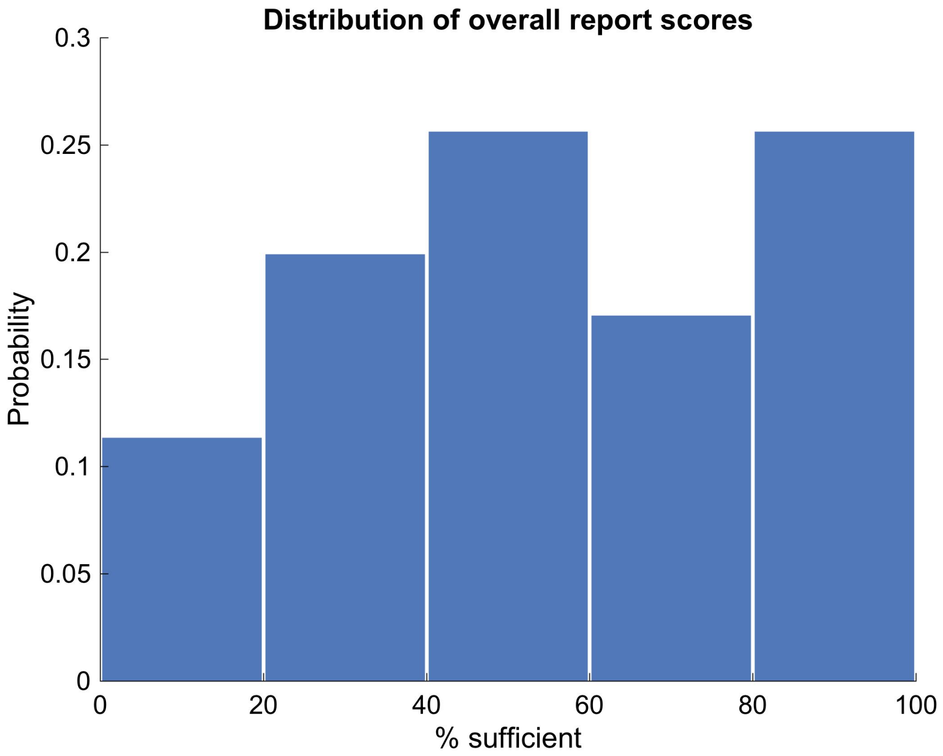
- 35 Credibility Plan Mid-Term Updates Submitted
    - 6 after the stated deadline
    - 3 were noted to be new grant awards
- 100% Received at least 1 review
- 85% Received 2 reviews as of 11/15/2018

Final Scores by TSR

*Total count of reviews in TSR category = 64
Updated 12/16/2018

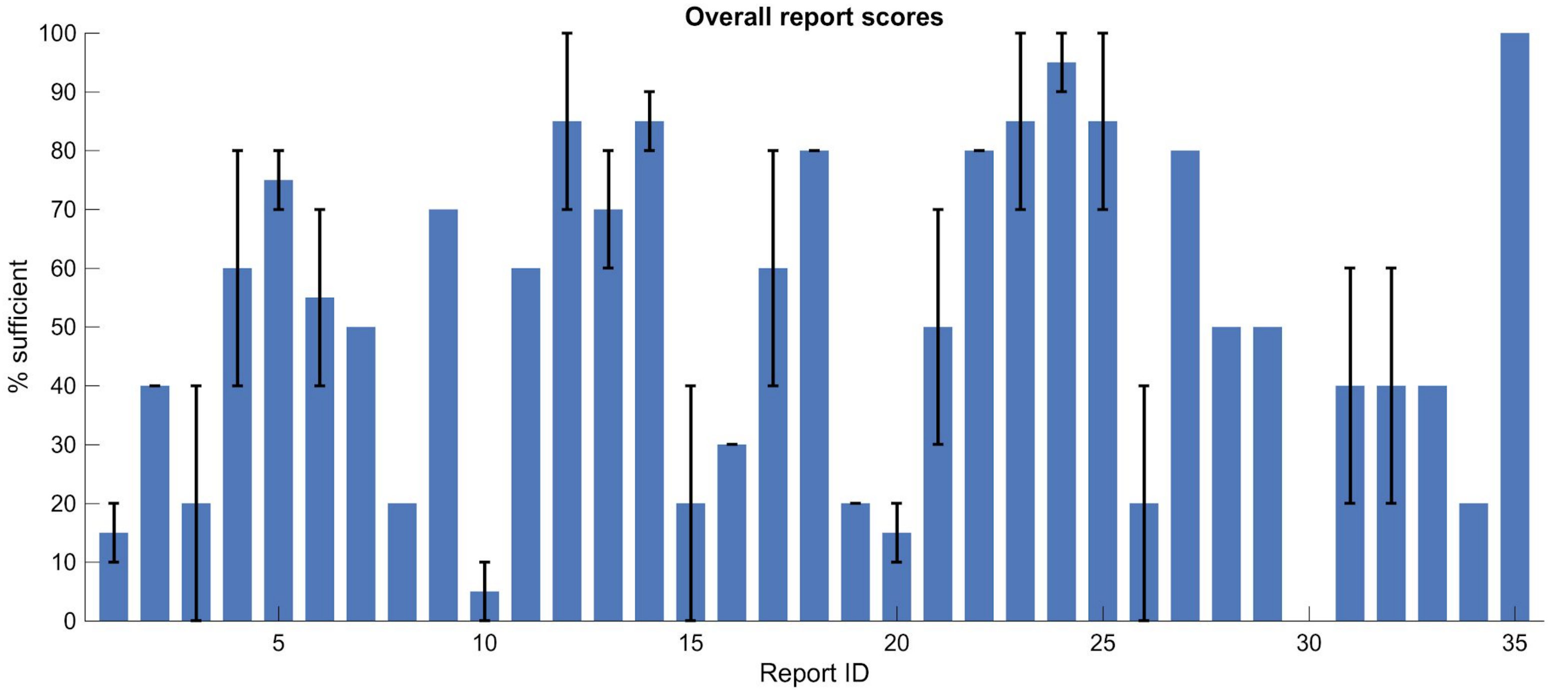


Distribution of “Sufficient” Rate Scores



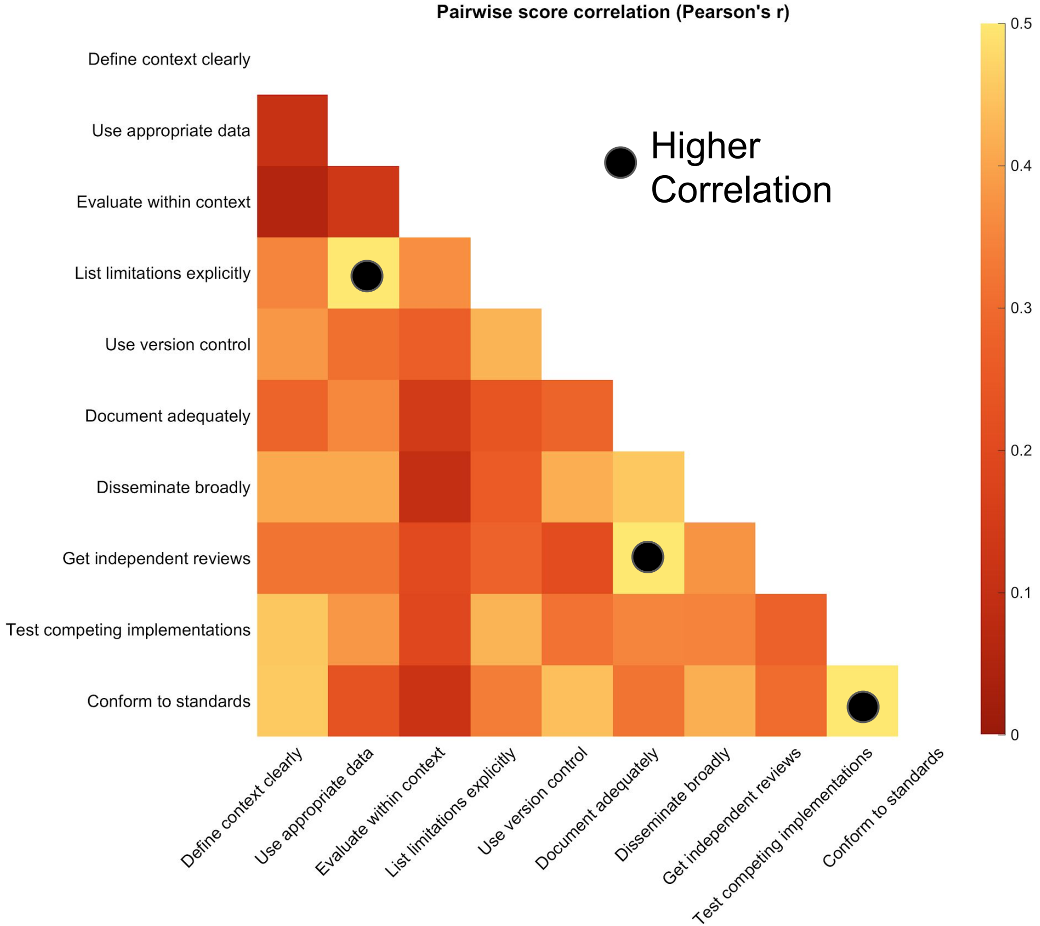
Scores: % of Rules Marked as “Sufficient”

Total count of reviews in TSR category = 60



Average % sufficient scores, error bars denoting the full range

Cross-Correlation between TSR



Distribution of Scores Based on Reviewer Agreement

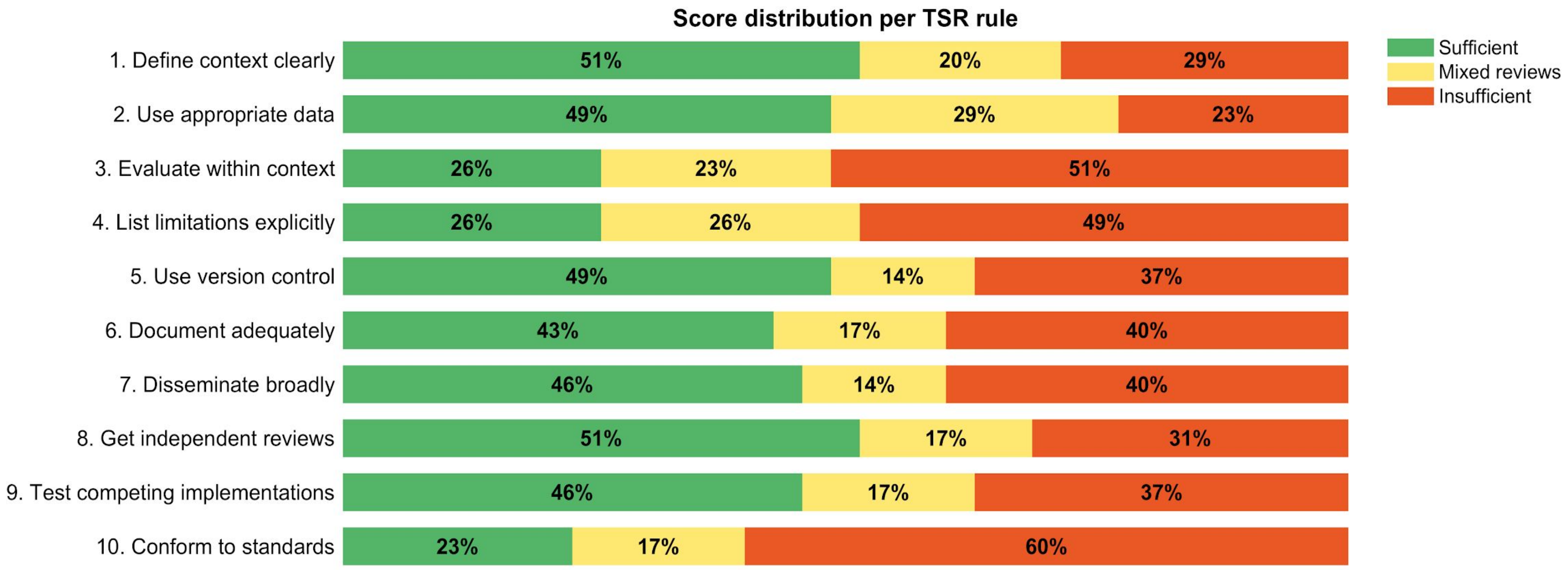


Chart illustrates agreement between reviewers in evaluating the TSR credibility category. Yellow indicates the 2 reviewers split on evaluation

Examples of TSR Satisfactory Description

General comments: The more detailed the information provided in each TSR category, the more likely a reviewer would score a provided description as “sufficient to allow credibility to be assessed.” The following examples received a “sufficient” score from both reviewers, although several were accompanied by comments and caveates from the reviewers.

R1 - Define context clearly

“We are currently developing multiscale mathematical and computational models of bone that integrate cellular and molecular scales. The aim is to elucidate the bone biology, the ecology of metastatic prostate cancer and design new treatment options for bone metastatic prostate cancer patients”

R2 - Use appropriate data

“We ensure that all parameters and input variables are based on published and in-house in vitro observations. If any parameters cannot be validated (due to lack of available data or techniques), other model variables are monitored to ensure accurate reflection of platelet biology”

Examples of TSR Satisfactory Description

R3 - Evaluate within context

“We test the code under conditions for which the correct behavior is known or can be calculated independently. For example, our Greens function method for oxygen transport was tested by comparing its solution with corresponding solutions using the Krogh cylinder model.

• We continuously generate graphical output during program execution, to check for Inconsistent or unexpected behavior. Graphics files showing network structure, hemodynamic variables, oxygen fields on slices through 3D domains, histograms of relevant variables, etc., are generated and monitored.

• In Specific Aim 2, the models are used to test hypotheses regarding the mechanisms of flow regulation in the brain, by generating multiple models in which specific mechanisms are turned on or off. We anticipate that many of these models will be unable to predict behavior consistent with observations, regardless of assumed parameter values. These “failures” will guide the choice of mechanisms to be included in the eventual model.
Comparisons with observed responses to several types of experimental conditions will aid in establishing the credibility of these models.

• We carry out sensitivity analyses of model results to key unknown parameters. These analyses are used to assess model robustness, to obtain estimates of uncertainty of model predictions where key parameters are not precisely known, and to predict the effects of parameter changes that occur in various physiological and pathological conditions. For example, we will examine the dependence of tissue oxygen distribution and hypoxic fraction on oxygen consumption rate and on perfusion.”

Examples of TSR Satisfactory Description

R4 - List limitations explicitly

“The Virtual Cell platform has a strict limitation on the number of grid points in the model that limit simulation size at the resolution needed. This problem is more severe for 3D simulations, which are the most realistic.
The crowding and juxtaposition of membranes in the interior of mitochondria and the inherently noisy nature of the tomograms, combined with directional resolution loss in the tomograms, often has made parts of the organelles more difficult to distinguish. Therefore, the automation must be combined with manual segmentation for accuracy.”

R5 - Use version control

“We are currently using GitHub for our version control. This extends to three major thrusts in this project. (1) With respect to the post-processing of DENSE MRI data, we currently have a repository for updates to the MATLAB based code. (2) The source code for the new FEniCS implementation of the finite element code, with the general-state contraction law, are be updated (privately for now). (3) The cellular level code is also be tracked with version control.”

Examples of TSR Satisfactory Description

R6 - Document adequately

“While developing model code, we simultaneously write drafts of publications that precisely describe our eccentric and concentric growth models. This allows us to iterate between the mathematical model and the computational simulation tool. To document the simulation tools, we utilize the documentation control system provided by Github.”

R7 - Disseminate broadly

Besides publications, research performed in this project has been discussed at several symposia, seminars, and national meetings, including:

- The Kavli Institute of Theoretical Physics, Santa Barbara, CA (July 2018)
- A symposium hosted by the Department of Pharmaceutical Sciences, University at Buffalo (July 2018)
- A preconference workshop sponsored by the American College of Clinical Pharmacology (September 2018)
- The American Conference on Pharmacometrics (upcoming October, 2018)

Several of these presentations have led to new collaborations that allow for independent testing of model code, both additional simulations to test reproducibility and new experiments to test whether predictions are accurate

Examples of TSR Satisfactory Description

R8 - Get independent reviews

“So far, all the models were independently reproduced by laboratory colleagues not directly involved in the specific project. New members to the lab routinely review prior models as part of their initial training.”

R9 - Test competing implementations

“We implement our models into different finite element platforms and perform independent simulations to compare individual implementations. We test our algorithms using simple model problems and benchmark our codes against established model problems with known solutions to identify discrepancies and estimate numerical error tolerances.”

R10 -Conform to standards

“We have adopted accepted standards for the programming of computational schemes using standard libraries in C++ in addition to highly optimized math and linear algebra libraries such as BLAS and ScaLapack. Data types and formats are all binary and can be used by popular visualization software such as ParaView and Tecplot for postprocessing.”

PI Observations

- “Model credibility is best evaluated by the unbiased user who needs the information coming out of the model the most. If there is a way to identify these people during the model building process and solicit their feedback more regularly, that would be very beneficial to ensuring model credibility.”
- There seems to be gaps between the standard developed and that of medical doctors for clinical applications.
    - Involve MDs to bridge the gaps between modelers and MDs
- Face-to-face visit beyond annual IMAG meeting

Reviewer Observations

- Generally appreciative of the effort from each PI team
  - Still areas where improvement can be achieved
- PI experienced challenges in providing detail in the report:
  - Project started recently or plan was focused on recently
  - PI had difficulty with template and/or instructions
  - PIs reported on project progress not model credibility activities
  - PIs had difficulty distinguishing between what activities have been planned versus completed

Reviewer Observations cont.

- PI's showed some confusion on reporting
    - limitations of the modeling and limitations of data acquisition
    - documentation of assumptions and key decisions, not just code and data
    - conforming to standards vs. internal best practices
    - evaluate within context (V, V, & UQ) and test competing implementations
    - test competing implementations vs. redeveloping the model in multiple coding languages
    - how some rules apply to their projects

Initial Recommendations

Recommendations to IMAG-MSM Community

Provided PI guidance / opportunities on approaches to “Documentation” and “Dissemination”

- Several PIs illustrated well thought out documentation and dissemination approaches.
- Many PIs rely on peer publications to meet documentation and dissemination credibility factors
    - Others seem to have a ”post it online” approach, without discussing curation

CPMS concern: Reliance on peer publication or uncurated posting as the primary means of providing evidence in these categories may be unsatisfactory to the user community

Recommendations to IMAG-MSM Community

PIs should be given criteria and opportunity for implementing third-party reviewer processes. Specifically, what exactly it means to sufficiently “include” them (ex. Researchers, MD’s) in the model development process

Range of PI Approaches

- Teaming of independent research groups, providing continuous reviews
- Seeking internal reviews within one’s organization but not on the project
- Hiring external consultants to provide the review
- Journal publication peer review

CPMS Concern: If not addressed early in grant it may become unachievable at a level commensurate with community’s credibility expectations.

Recommendations to CPMS

Implement process improvements of mid-term credibility review exercise

- Standardize input format to improve the PI ability to provide applicable information
- Improve directions to reduce PI confusion on breadth and depth of requested information
- Enhance definition of CPMS-TSR, to improve relating credibility plan content to TSR factors
    - Soften/ Improve description of “engineer or computer scientist jargon”
- Provide examples communicating model credibility information in each TSR
    - Include illustration of how credibility actions may contribute to informing multiple TSR factors
    - Illustrate it’s acceptable to communicate explicitly when credibility actions are not planned

CPMS Applying What We Have Learned

Efforts Prior to IMAG-MSM Meeting

Mid-Dec. 2018: Finalize global findings and recommendations as a short report or slide deck to post on the IMAG wiki

Mid-Dec. 2018: Consolidate individual grantee findings and provide a short outline of individualized recommendations

Feb 2019: Development of a set of representative examples of each TSR factor to be used for guidance in providing TSR related information

March 2019: Development of Poster and / or Presentation of the process and findings for IMAG-MSM Meeting

After the IMAG-MSM meeting

4/2019: TSR Papers - outlining the TSR development and implementation process

5/2019: Develop a more formal means to request and review mid-term credibility plan updates

7/2019: Expand and clarify definitions, with expanded examples, for each TSR rule

7/2019: Expand reviewer directions and number of reviewers.

9/2019: Add TSR Rubric for consideration in the evaluation process

CPMS TSR Rubric - Draft

Outreach Capability

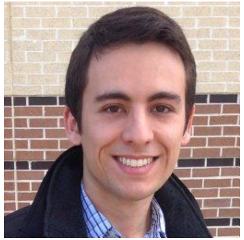
Outreach to Application-domain experts that may not be M&S practitioners	Outreach to M&S practitioners that may not be Application-domain experts	Outreach to Application-domain specific M&S practitioners	Outreach to Application-domain specific M&S practitioners	None/Too little
Comprehensive	Extensive	Adequate	Partial	Insufficient

Compliance Level

Note: Specific interpretation being tailored to each TSR rule

Acknowledgements

Invaluable contributions to meta-analysis of all collected data



Bruno Rego, PhD candidate
University of Texas at Austin

Amir Khalighi, PhD
University of Texas at Austin

QUESTIONS?


OH, YOU DID NOT FOLLOW
THE TEN SIMPLE RULES
TELL ME MORE WHO DID NOT
FOLLOW A CREDIBLE PRACTICE
imgflip.com

Backup

Final Score by Study

*Total count of reviews in TSR category = 64
Updated 12/16/2018