Transparency and Reproducible Research in Modeling : Why We Need It and How to Get There

Crystal M. Smith-Spangler

Med Decis Making 2012 32: 663
DOI: 10.1177/0272989X12458977

Transparency and Reproducible Research in Modeling: Why We Need It and How to Get There

Crystal M. Smith-Spangler, MD, MS

Decision modeling and cost effectiveness analysis have become important tools to inform clinical decision making and policy development, playing roles in policy decisions for human immunodeficiency virus[1,2] and breast and colon cancer screening,[3,4] as well as clinical decisions in the prevention of cardiac disease,[5,6] for example. The establishment and use of best practices for model development, validation, and reporting are key steps in ensuring that model users have information they can trust.

In this issue of *Medical Decision Making*, Eddy and others[7] describe practices for model transparency and validation recently developed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR)–Society for Medical Decision Making (SMDM) Modeling Good Research Practices Task Force. The authors recommend that modelers 1) provide a freely accessible nontechnical description of the model (VII.1), including a description of the external validation methods and results (VII.1 and VII.6) and verification (internal validation) methods (VII.4); 2) make available (openly or under licensing agreements) sufficiently technical documentation such that expert readers can evaluate and potentially reproduce the model (VII.2); and 3) assess face (VII.3), internal (verification) (VII.4), and external validity (VII.5–VII.11).

Previous guidelines and recommendations have discussed the importance of transparency and validation, urging developers to clearly report model structure, data, equations, and assumptions, as well as methods to validate or check model consistency.[8–12] However, these new ISPOR-SMDM recommendations set a new standard by encouraging the sharing of technical details, including code.

The guidelines arrive at a time of heightened interest in transparency in research communities. Earlier this year, in a report investigating difficulties in evaluating and reproducing key translational omics findings that were used for treatment choice in later clinical trials (and that were later retracted),[13,14] the Institute of Medicine issued a call for greater transparency and sharing of data and code within translational omics and in the broader research community.[15]

Eddy and colleagues[7] appropriately note that concerns about intellectual property, model misuse, and costs may limit sharing of model details. The authors also suggest that if a model is accurate (validated), transparency may be of lesser importance: ''Ultimately, what matters most is whether a model accurately predicts what occurs in reality.'' However, accuracy does not diminish the need for transparency, and there may be ways to address the identified challenges.

First, an accurate model still needs to be transparent. What degree of accuracy is sufficient to eliminate the need for transparency? Which are the exact data against which predictions should be measured? Both are matters of judgment. And, if prospective validation (predictive validation) is the highest standard for validation (as implied by these good practice recommendations), this information may only be available years after publication, limiting its usefulness at the time of decision making. Finally, too narrow a focus on predictive accuracy may unnecessarily discount findings from a model that cannot be thoroughly validated in this manner. The primary purpose of decision

From VA Palo Alto Health Care System, Palo Alto, California, and Stanford School of Medicine, Stanford, California.

Address correspondence to Crystal Smith-Spangler, MD, MS, Division of General Medical Disciplines, Stanford School of Medicine, 1265 Welch Road, MSOB MC 5411, Stanford, CA 94305, USA; e-mail: csmithsp@stanford.edu.

DOI: 10.1177/0272989X12458977

modeling is not to predict future events but rather to provide useful information about choices so as to inform decision making at a particular point in time.[9] Reflecting this aim of decision modeling, Sculpher and others[9] define a valid model differently than do Eddy and colleagues,[7] arguing that a model is valid if *useful;* an invalid (useless) model does not improve decisions (e.g., the model does not assess choices or outcomes of interest to the decision maker). If accuracy or validity cannot be perfectly ascertained or is not the best measure of model usefulness, then transparency and full reporting of technical details remain the foundation of ensuring confidence in decision models.

Eddy and colleagues' call[7] for full sharing of technical details is supported by a movement in the greater scientific community to encourage transparency. In 2003, the National Academy of Sciences published a report, ''Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences,'' which describes the ''Uniform Principles for Sharing Publication-Related Materials (UPSIDE).'' The report asserts that

> the act of publishing is a quid pro quo in which authors receive credit and acknowledgement in exchange for disclosure of their scientific findings. An author's obligation is not only to release data and materials to enable others to verify or replicate published findings (as journals already implicitly or explicitly require) but also to provide them in a form on which other scientists can build with further research. All members of the scientific community—whether working in academia, government, or a commercial enterprise—have equal responsibility for upholding community standards as participants in the publication system, and all should be equally able to derive benefits from it.[16]

Investigators wishing to improve transparency and make sharing easier may desire to develop reproducible research habits. *Reproducible research* is the sharing of code, data, metadata, and software such that findings can be easily reproduced by others.[17] A paper may be considered reproducible, for example, if the code submitted by the authors can generate all the figures and tables used in the paper.[18]

Reproducible research habits have additional benefits. Reproducible research can 1) improve teamwork (members of the team can see clearly what is being done, suggest improvements, and find errors), 2) allow greater continuity and cumulative impact (when new people join a team, they can build on previous work rather than spending considerable amounts of time understanding the old model or generating a new one), and 3) increase citations of the original work.[18]

Moreover, the sharing of full technical details of research supported by federal funding can be required under US law. The 1998 Shelby Amendment required all federal funding agencies to ''ensure that all data produced under an award will be made available to the public through the procedures established under the Freedom of Information Act.''[19] The Office of Management and Budget has defined research data as the ''recorded factual material commonly accepted in the research community as necessary to validate research findings.''[20] Technical code for modeling likely meets this definition of data. The Shelby Amendment was enacted in response to lawmakers' anger over a dispute between industry and researchers over research findings pertaining to air quality standards in which federally funded researchers refused to share pertinent patient data.[21] Eventually, researchers released data to an independent third party who confirmed the results.[21] This dispute triggered the development of new platforms and models for sharing both the data and code underlying analyses of air quality and health (see iHAPSS [the Internet-based Health and Air Pollution Surveillance System]: http://www.ihapss.jhsph.edu/).

Eddy and colleagues[7] describe many concerns about sharing model specifications, most of which were also discussed by the Committee on the Responsibilities of Authorship in the Biological Sciences, which published the 2003 UPSIDE principles. Scientists had concerns about 1) commercial interests, 2) the original costs of producing data and materials, 3) protecting young researchers from competition, 4) contractual agreements with industry sponsors, and 5) an investigator's right to mine his or her data before others. However, the committee concluded that those participating in sharing data were as likely to benefit as to be harmed.[16]

Nonetheless, for implementation of increased sharing of models, additional work needs to be done to address these concerns. Tools to support copyright and licensing exist (see Creative Commons licensing options, for example: http://creativecommons.org/licenses/). However, additional tools, repositories, or platforms likely need to be developed to meet the needs of the modeling community. Digital object identifiers (DOIs) could be assigned so that published code can be cited and tracked. Modelers

concerned about misuse could stipulate reproduction only of the model (verification and validation, no reuse). Furthermore, some investigators may wish to share models under restricted access arrangements in which investigators apply for access to the model code through prespecified procedures, similar to the way in which data are released to non–study investigators for many large epidemiologic studies (e.g., Women's Health Initiative: www.nhlbi.nih.gov/whi or Framingham Heart Study: www.framinghamheartstudy.org). For the Women's Health Initiative, other investigators using the data are prohibited from publishing for 12 months after the initial publication, which allows the original authors to complete additional manuscripts. Peng[17] notes that although published code ''does not have to be clean or beautiful,''[22] modelers may wish to share cleaned code or implementable models. Additional funding may be needed to support preparation of models for sharing and model-sharing repositories or platforms. The assistance of bioinformaticians, biostatisticians, and other investigators experienced in reproducible research methods and data sharing will likewise be important to support sharing.

Many if not most of the challenges of reproducible research are not unique to the modeling community. We can learn from others' experiences and work together to improve the reproducibility of our work. Funders, institutions, journals, and individual investigators can all play a role in encouraging transparent, reproducible research practices. Error is likely ubiquitous in research (although much of it may be unintentional).[18,23] Reproducible research practices can help us avoid error, enhance confidence in our findings, and improve teamwork and the cumulative impact of the decision model. The ISPOR-SMDM Good Practices Guidelines for Model Transparency and Validation are a welcome set of recommendations to improve our research methods and promote confidence in modeling research.

REFERENCES

1. Sanders GD, Bayoumi AM, Sundaram V, et al. Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. N Engl J Med. 2005;352(6):570–85.

2. Paltiel AD, Weinstein MC, Kimmel AD, et al. Expanded screening for HIV in the United States—an analysis of cost-effectiveness. N Engl J Med. 2005;352(6):586–95.

3. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Ann Intern Med. 2009;151(10):738–47.

4. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. Ann Intern Med. 2008;149(9):659–69.

5. Eddy DM, Adler J, Patterson B, Lucas D, Smith KA, Morris M. Individualized guidelines: the potential for increasing quality and reducing costs. Ann Intern Med. 2011;154(9):627–34.

6. Hayward RA, Krumholz HM, Zulman DM, Timbie JW, Vijan S. Optimizing statin treatment for primary prevention of coronary artery disease. Ann Intern Med. 2010;152(2):69–77.

7. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7. Med Decis Making. 2012;32(5):733–743.

8. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. PharmacoEconomics. 2006;24(4):355–71.

9. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models: a suggested framework and example of application. PharmacoEconomics. 2000;17(5):461–77.

10. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value Health. 2003;6(1):9–17.

11. Karnon J, Goyder E, Tappenden P, et al. A review and critique of modelling in prioritising and designing screening programmers. Health Technol Assess. 2007;11(52):iii–iv, ix–xi, 1–145.

12. Brandeau ML, McCoy JH, Hupert N, Holty JE, Bravata DM. Recommendations for modeling disaster responses in public health and medicine: a position paper of the Society for Medical Decision Making. Med Decis Making. 2009;29(4):438–60.

13. Baggerly KA, Coombes KR. What information should be required to support clinical ''omics'' publications? Clin Chemist. 2011;57(5):688–90.

14. Potti A, Dressman HK, Bild A, et al. Retraction: genomic signatures to guide the use of chemotherapeutics. Nat Med. 2011;17(1):135.

15. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Board on Health Care Services, Board on Health Sciences Policy, Institute of Medicine. Evolution of Translational Omics: Lessons Learned and the Path Forward. Washington, DC: National Academies Press; 2012.

16. Committee on the Responsibilities of Authorship in the Biological Science, National Research Council. Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences. Washington, DC: National Academic Press; 2003.

17. Peng RD. Reproducible research in computational science. Science. 2011;334(6060):1226–7.

18. Donoho DL. An invitation to reproducible computational research. Biostatistics. 2010;11(3):385–8.

19. Senate and House of Representatives of the United States of America. Public Law 105-277. Washington, DC, 1998. Available from: http://www.gpo.gov/fdsys/pkg/PLAW-105publ277/pdf/PLAW-105publ277.pdf. Accessed 6 July 2012.

20. Office of Management and Budget. OMB Circular A-110, ''Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations.'' 1999. Available from: http://www.whitehouse.gov/omb/fedreg_a110-finalnotice. Accessed 23 January 2012.

21. Dresser R. Accountability in science and government: is access the answer? Hastings Cent Rep. 2000;30(3):29–30.

22. Barnes N. Publish your computer code: it is good enough. Nature. 467(7317):753.

23. Steen RG. Retractions in the scientific literature: is the incidence of research fraud increasing? J Med Ethics. 2011;37(4):249–53.