TRIPOD: A New Reporting Baseline for Developing and Interpreting Prediction Models

In this issue, the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) investigators propose an annotated checklist for transparent reporting of prediction or prognostic models (1). The accompanying 22 000-word "Explanation and Elaboration" (2), with more than 500 references from statistics, epidemiology, and clinical decision making as well as from the applied clinical literature, should serve as an important resource for model developers. Applications span prediction (for diagnosis from cross-sectional data) and prognosis (for outcomes after longitudinal follow-up) across many clinical settings. These TRIPOD documents represent serious efforts to synthesize best practices for authors and readers.

Because the TRIPOD checklist seeks primarily to improve reporting, the immediate benefit should be more comprehensive expositions of model development and validation. Dissemination is broad: The checklist is being published simultaneously in 10 journals that span the disciplines of general medicine and clinical practice, obstetrics and gynecology, urology, surgery, cardiology, diabetes, cancer, clinical biomedical science, and epidemiology.

Longer-term benefits might flow from TRIPOD's success in leading the research and clinical communities to ask the right questions and consider the challenges in developing, validating, and using prediction models. As TRIPOD suggests, although there is no single correct method for modeling, authors should be able to justify their assumptions and describe their approaches. TRIPOD's comprehensive outline of modeling options should encourage informed choices and transparent presentations to further more general goals of reproducible research (3).

The current version of TRIPOD is not the final word on prediction model development and validation. Methodological advances and critiques appear almost monthly and can quickly render guidelines obsolete. Fortunately, TRIPOD will maintain a Web site for updates.

Complete reporting alone cannot generate clinically useful models. In our review of TRIPOD, we found the following areas of continuing concern for model developers and users.

Why Model? What Is the Question in Clinical Context?

Models must have a clear purpose and should lead to improved decisions or behavior. For example, Smits and colleagues (4) sought to reduce excessive computed tomography scans for minor head injuries (false-positive results) without missing patients who need neurologic intervention. With this clear objective in mind, they could develop a model and select appropriate thresholds for classifying patients. Authors must also consider whether the model will influence decision making. In developing a model to predict bacteremia among hospitalized patients, Lautenbach and colleagues (5) discovered that physicians would not forgo routine blood tests unless the estimated risk for false-negative results from a bacteremia prediction model was well below the level that even a large model could rule out. Regardless of its discrimination or calibration, the model must perform well in the region of predicted risk at which clinicians make treatment decisions or classifications to be of clinical benefit.

Data Quality

Even the best guidance for prognostic or prediction models cannot replace data collection mechanisms, registries, and coding schemes that generate complete and accurate data. Neither statistical expertise nor clinical experience can substitute for missing or misclassified information. With flawed data, unbiased validation becomes problematic, and application to individual patients becomes uncertain.

Moving Beyond the Standard Decision Metrics

Model metrics alone do not translate into clinical decisions. C-statistics of model discrimination and Hosmer–Lemeshow tests for model calibration can depend on context and setting, duration of follow-up, and the distribution of candidate predictors (6). Users should be wary of comparisons based only on these summaries. Although summary metrics might be helpful in reports of model performance, they are not sufficient. For a candidate model, one must also examine subgroups of patients who will benefit and how much (from more true-positive or true-negative decisions), as well as those who might suffer (from false-positive or false-negative decisions). Modeling must move beyond summary measures and into clinically meaningful presentations (7) to evaluate improvement in clinical or financial benefits of one model versus another or any model versus none (8).

Prognosis With Time-Varying Treatments

Individual patient evaluation, treatment, and outcomes do not follow an orderly sequence, as some prognostic models inherently assume. Even for a single patient, treatments change depending on response to prior therapies. Clinical risk factors or markers in turn change and then alter intermediate outcomes. For these realistic settings, authors must develop and validate time-varying decision models that incorporate changes in levels of prognostic factors, patient treatments, and ongoing responses. Analytic options for this

complex but realistic scenario are developing rapidly (9).

Multicenter Settings

Multicenter data sources are essential, especially for rare outcomes. Although TRIPOD notes this reality, readers must appreciate that when risk factors are collected and outcomes are monitored across institutions, whether from established registries or from ad hoc data consortia, common modeling metrics require special methods for estimation (10). Of more concern, institutions differ not only in methods of patient care and management but also in how and when they collect and code clinical information. These variations in data collection can confound observed associations of patient-level risk factors and outcomes. The multicenter solution to sample size and generalizability has risks, and the ability of clinical research networks to generate data of adequate quality and completeness for prediction modeling should remain a focus of network design and evaluation.

Transparency

As with any reporting guidance, TRIPOD should supplement rather than replace the standard admonition from the International Committee of Medical Journal Editors to authors: "Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to judge its appropriateness for the study and to verify the reported results" (www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html). Technical appendices will probably become the norm for reports of multivariable prediction models, especially when they use complex statistical methods.

Better prognostic and prediction models cannot succeed unless we do a better job of specifying the clinical question at the start and characterizing the consequences to patients for whom the model will guide treatments or interventions. TRIPOD can shepherd prediction models, but successful efforts increasingly must overcome the challenges of multi-institutional data sources, data misclassification, measurement error, unobserved factors, and time-varying risks and treatments.

A. Russell Localio, PhD
Perelman School of Medicine, University of Pennsylvania
Philadelphia, Pennsylvania

Catharine B. Stack, PhD
Deputy Editor, Statistics

Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M14-2423.

Requests for Single Reprints: A. Russell Localio, PhD, Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, 635 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021.

Current author addresses are available at www.annals.org.

Ann Intern Med. 2015;162:73-74. doi:10.7326/M14-2423

References

1. **Collins GS, Reitsma JB, Altman DG, Moons KG.** Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162:55-63. doi:10.7326/M14-0697
2. **Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al.** Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1-73. doi:10.7326/M14-0698
3. **Laine C, Goodman SN, Griswold ME, Sox HC.** Reproducible research: moving toward research the public can really trust. Ann Intern Med. 2007;146:450-3. [PMID: 17339612] doi:10.7326/0003-4819-146-6-200703200-00154
4. **Smits M, Dippel DW, Steyerberg EW, de Haan GG, Dekker HM, Vos PE, et al.** Predicting intracranial traumatic findings on computed tomography in patients with minor head injury: the CHIP prediction rule. Ann Intern Med. 2007;146:397-405. [PMID: 17371884] doi:10.7326/0003-4819-146-6-200703200-00004
5. **Lautenbach E, Localio R, Nachamkin I.** Clinicians required very high sensitivity of a bacteremia prediction rule. J Clin Epidemiol. 2004;57:1104-6. [PMID: 15528062]
6. **Vickers AJ, Cronin AM.** Everything you always wanted to know about evaluating prediction models (but were too afraid to ask). Urology. 2010;76:1298-301. [PMID: 21030068] doi:10.1016/j.urology.2010.06.019
7. **Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, et al.** Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case–control and cohort validation study. Ann Intern Med. 2012;157:242-50. [PMID: 22910935] doi:10.7326/0003-4819-157-4-201208210-00004
8. **Pepe MS, Janes H, Li CI.** Net risk reclassification *P* values: valid or misleading? J Natl Cancer Inst. 2014;106:dju041. [PMID: 24681599] doi:10.1093/jnci/dju041
9. **Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA.** Methods for dealing with time-dependent confounding. Stat Med. 2013;32:1584-618. [PMID: 23208861] doi:10.1002/sim.5686
10. **van Klaveren D, Steyerberg EW, Perel P, Vergouwe Y.** Assessing discriminative ability of risk models in clustered data. BMC Med Res Methodol. 2014;14:5. [PMID: 24423445] doi:10.1186/1471-2288-14-5

Current Author Addresses: Dr. Localio: Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, 635 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021.
Dr. Stack: American College of Physicians, 190 N. Independence Mall West, Philadelphia, PA 19106.