Published in final edited form as:

Med Decis Making. 2015 October ; 35(7): 815–816. doi:10.1177/0272989X15598742.

Validation is necessary but insufficient

R. Scott Braithwaite, MD, MS^{1,*} and Mark S Roberts, MD, MPP²

¹Division of Comparative Effectiveness and Decision Sciences, Department of Population Health, New York University School of Medicine, New York, N.Y

²Department of Health Policy and Management, Graduate School of Public Health, University of Pittsburgh

Dear Editor

At first read, Dr. Caro's editorial (1) along with its companion articles (2,3) makes a compelling point – predictive models could well benefit from more rigorous and standardized validation. But at second read, a distinct insight emerges.

Would you care if a financial advisor is able to accurately predict market gains to within some statistically defined interval of certainty? It certainly would be appealing, but you would likely base your decision regarding whether to choose the financial advisor on whether her predictions were sufficiently more accurate than yours (or her competition) to result in greater financial return over a designated time interval.

To apply the same idea to medical decision making, would you care if diagnostic software for medical diagnoses was accurate to within some statistically defined interval? Or would you use the software if it reduced the number of diagnostic errors that are made without unduly increasing time and resource burden?(4) Suppose you are not faced with a complex model but rather a simple scenario: comparing two RCTs for a new cancer drug. Drug A compared to standard therapy increased life expectancy by 1 month at a p of 0.01; Drug B increased life expectancy by 10 years but at a p of 0.06 – Does satisfying statistical criteria really impact which drug you should choose?

A prediction model by itself has no inherent value – a prediction model is only useful when it improves a specific decision, and in that case, it is the overall value that the model-based decision brings that is important to understand.

Few model users would base decisions regarding models on whether the validation procedure used was based on Bayesian evidence synthesis (2), Pareto Frontiers (3), minimizing a least difference parameter, or the “eyeball method.” Comparing these methods is extremely important to the state of science to identify and codify methods of validating models that are robust, rigorous, and reproducible. But all of these approaches are rooted in

*Corresponding Author: 227 East 30th street, Floor 6 Room 615, New York, NY 10016, Phone: 212-263-4964, Fax: 646-501-2706, Scott.Braithwaite@nyumc.org.

Financial disclosure/conflict of interest: The authors have no conflicts of interest of financial disclosures to declare.

statistics, whereas arbitrating among these approaches is not something that can be accomplished in a statistical domain. Rather, arbitrating among these methods can only be addressed within a decision analytic domain, by asking which method yields the greatest expected value under a plausible distribution of uses.

Validation needs to be supplemented by efforts to describe whether the model improves decisions compared with decisions that would have been made without exposure to the model, along with a comparison of which predictions lead to decisions that yield the greatest expected value. Of course, one might argue that any model with informative “signal” (and, certainly, which has passed a validation test against an external criterion) will narrow one’s posterior probability distribution regarding events of interest, offer greater certitude compared to no model, and consequently improve prediction. However, models need to accommodate users who are not expert in Bayesian methods, and who may substitute the model’s predictions for their own rather than combining them “in the field” in a correctly specified Bayesian manner.

Clearly, validating models is necessary for the simple reason that predictive models should be able to predict what they model. Validation asks a question that is compelling, satisfying, and vital to the art of modeling. However, it is different from the question that is most important for model users.

References

1. Caro JJ. Psst, Have I Got a Model for You. *Med Decis Making*. Feb.2015 35:136–138. [PubMed: 25510875]
2. Jackson CH, Jit M, Sharples LD, De Angelis D. Calibration of Complex Models Through Bayesian Evidence Synthesis: A Demonstration and Tutorial. *Medical Decis Making*. Feb.2015 35:148–161.
3. Enns EA, Cipriano LE, Simons CT, Kong CY. Identifying Inputs in Health-Economic Model Calibration: A Pareto Frontier Approach. *Med Decis Making*. Feb.2015 35:170–182. [PubMed: 24799456]
4. Braithwaite RS, Scotch M. Using value of information to guide evaluation of decision supports for differential diagnosis. Is it time for a new look? *BMC Medical Informatics and Decision making*. 2013; 13:105. [PubMed: 24020989]