

Letter to the Editor: *CORR* Synthesis: When Should We Be Skeptical of Clinical Prediction Models?

Hung-Kuan Yen MD^{1,2,3}, Hongsen Chiang MD, PhD³ 

To the Editor,

We would like to thank Drs. Karhade and Schwab [8] for a thoughtful *CORR* Synthesis review discussing when and why a reader might trust a clinical prediction model and when such models are most useful.

These models call for careful external validation to ensure that they are generalizable [6, 7, 16]. However, the area under the receiver operating characteristic curve, the Brier score, the calibration analysis, and the decision curves [12], known as the “ABCD methods,” may not be sufficient for certain clinical applications.

A high-quality prediction model provides objective suggestions that can inform the shared decision-making process between patient and physician. Together, with the physician’s expertise,

the best therapeutic strategies can be tailored based on the suggestions delivered by these models [2, 4, 10]. In some clinical situations, several prediction models may be needed to get the necessary information; a common situation involves decision-making for patients with metastatic bone disease, in which different prediction models have evaluated different time points [5, 9, 13], and consulting the correct model could be the difference between a good decision and a bad one. For example, patients with spinal metastasis and a life expectancy of less than 3 months would likely not choose surgery because the postoperative recovery might take multiple months [3]. In contrast, patients with a life expectancy longer than a year could potentially benefit from an aggressive operation that reduces the local

tumor progression and the subsequent revision surgeries [11, 14]. In this example, the timing matters when predicting the potential for therapeutic benefit.

But this only holds true if a prediction model delivers predictions that pass even the most rudimentary logical scrutiny, which is not always the case. For example, if given the same parameters, a prediction model should not estimate survivorship in the long term to be greater than survivorship in the short term. Unfortunately, none of the “ABCD methods” can determine the likelihood of such errant predictions being generated by a particular model. However, we have devised the simple “model consistency (MC)” metric defined as:

An MC = 1 indicates the best consistency, whereas an MC = 0 signifies

(RE: Karhade AV, Schwab JH. *CORR* synthesis: when should we be skeptical of clinical prediction models? *Clin Orthop Relat Res*. 2020;478:2722.)

The authors certify that there are no funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article related to the author or any immediate family members.

This letter was funded by the institutional project of National Taiwan University Hospital (NTUH110-N5000).

H. Chiang ✉, National Taiwan University Hospital, 7 Chungshan South Road, Taipei, 10002, Taiwan, Email: hongsen@ntu.edu.tw

¹Department of Orthopaedic Surgery, National Taiwan University Hospital, Hsin-Chu Branch, Hsin-Chu City, Taiwan

²Department of Medical Education, National Taiwan University Hospital, Hsin-Chu Branch, Hsin-Chu City, Taiwan

³Department of Orthopaedic Surgery, National Taiwan University Hospital, Taipei City, Taiwan

Letter to the Editor

$$MC = \frac{\text{Number of consistent prediction pairs}}{\text{Number of consistent prediction pairs} + \text{Number of inconsistent prediction pairs}}$$

the worst. The MC serves as a performance metric that gives clinicians a more comprehensive idea about the prediction model’s behavior.

We believe that the MC metric has many clinical applications. In particular, it may help to identify when

prediction models are going to have important inconsistencies. For example, Basu et al. [1] proposed prediction models to estimate 5-year cardiovascular and all-cause mortality (Table 1). Naturally, the former is lower than the latter; however, their models give

inconsistent predictions in certain settings. Those kinds of obvious inconsistencies—it’s impossible for 5-year cardiovascular mortality to be higher than 5-year all-cause mortality—should call our attention to the importance of getting this right. We believe that the MC is a tool that can help us to do so.

Table 1. Examples of inconsistent prediction results from two widely used models

Parameters entered into the SORG prediction model [11, 15]	Parameters entered into the RECODE prediction model [1]
Primary tumor: slow-growing ECOG PS: 3-4	Age: 60, Gender: Male
ASIA impairment score: A	Ethnicity: African-American
Charlson comorbidities: 2	Tobacco use: No
Visceral metastases: No	Blood pressure: 160 mmHg
Brain metastases: No	Cardiovascular disease history: Yes
Number of spine metastases: 2	Antihypertensive agents: Yes
Previous systemic therapy: Yes	Statins use: Yes
BMI: 27 kg/m ²	Anticoagulants use: Yes
Hemoglobin: 7 g/dL	HbA1c: 9.5 %
Platelet count: 127×10 ³ /uL	Total cholesterol: 180 mg/dL
Absolute lymphocyte count: 0.91×10 ³ /uL	High density lipoprotein: 65 mg/dL
Absolute neutrophil count: 0.8×10 ³ /uL	Creatinine: 2.0 mg/dL
Creatinine: 0.5 mg/dL	Albumin to creatinine ratio: 300 mg/g
International normalized ratio: 1.1	
Albumin: 3.9 g/dL	
Alkaline Phosphatase: 63 IU/L	
3-month survival as estimated by SORG: 46%*	All-cause mortality as estimated by RECODE: 30%**
1-year survival as estimated by SORG: 70%*	Cardiovascular mortality as estimated by RECODE: 32%**

*This necessarily is an inconsistent prediction; 3-month survival in the same clinical circumstances should always be the same or higher than 1-year survival, not lower.
 **This, too, is an inconsistent prediction by definition; in the same patient, cardiovascular mortality should not exceed all-cause mortality at the same time point.
 SORG = Skeletal Oncology Research Group; RECODE = Risk Equations for Complications Of type 2 Diabetes; ASIA = American Spinal Injury Association Impairment Scale; ECOG PS = Eastern Cooperative Oncology Group performance status.

References

1. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODE) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol.* 2017;5:788-798.
2. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med.* 2017;376:2507.
3. Dea N, Versteeg AL, Sahgal A, et al. Metastatic spine disease: should patients with short life expectancy be denied surgical care? An international retrospective cohort study. *Neurosurgery.* 2020;87:303-311.
4. Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG; Cochrane IPD Meta-analysis Methods Group. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med.* 2015;12:e1001886.
5. Forsberg JA, Eberhardt J, Boland PJ, Wedin R, Healey JH. Estimating survival in patients with operable skeletal metastases: an application of a Bayesian belief network. *PLoS One.* 2011;6:e19956.
6. Groot OQ, Bindels BJJ, Ogink PT, et al. Availability and reporting quality of external validations of machine-learning prediction models with orthopedic surgical outcomes: a systematic review. *Acta Orthop.* 2021;92:385-393.
7. Karhade AV, Ogink PT, Thio Q, et al. Development of machine learning algorithms for prediction of prolonged

Letter to the Editor

- opioid prescription after surgery for lumbar disc herniation. *Spine J.* 2019;19:1764-1771.
8. Karhade AV, Schwab JH. CORR synthesis: when should we be skeptical of clinical prediction models? *Clin Orthop Relat Res.* 2020;478:2722.
 9. Karhade AV, Thio Q, Ogink PT, et al. Predicting 90-day and 1-year mortality in spinal metastatic disease: development and internal validation. *Neurosurgery.* 2019;85:E671-E681.
 10. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ.* 2016;353:i13140.
 11. Shah AA, Karhade AV, Park HY, et al. Updated external validation of the SORG machine learning algorithms for prediction of ninety-day and one-year mortality after surgery for spinal metastasis. *Spine J.* 2021;21:1679-1686.
 12. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J.* 2014;35:1925-1931.
 13. Thio Q, Karhade AV, Bindels BJJ, et al. Development and internal validation of machine learning algorithms for pre-operative survival prediction of extremity metastatic disease. *Clin Orthop Relat Res.* 2020;478:322-333.
 14. Tseng TE, Lee CC, Yen HK, et al. International validation of the SORG machine-learning algorithm for predicting the survival of patients with extremity metastases undergoing surgical treatment. *Clin Orthop Relat Res.* 2022;480:367-378.
 15. Yang JJ, Chen CW, Fourman MS, et al. International external validation of the SORG machine learning algorithms for predicting 90-day and one-year survival of patients with spine metastases using a Taiwanese cohort. *Spine J.* 2021;21:1670-1678.
 16. Yen HK, Ogink PT, Huang CC, et al. A machine learning algorithm for predicting prolonged postoperative opioid prescription after lumbar disc herniation surgery. An external validation study using 1,316 patients from a Taiwanese cohort. *Spine J.* 2022;22:1119-1130.