University of Groningen

External Validation Is Required for Comparisons of the Diagnostic Performance of Several Predictive Scores reply

Hiemstra, Bart; Eck, Ruben J.; Keus, Frederik; van der Horst, Iwan C. C.

Published in:
Critical Care Medicine

DOI:
[10.1097/CCM.00000000000004121](https://doi.org/10.1097/CCM.00000000000004121)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Hiemstra, B., Eck, R. J., Keus, F., & van der Horst, I. C. C. (2020). External Validation Is Required for Comparisons of the Diagnostic Performance of Several Predictive Scores reply. *Critical Care Medicine*, 48(2), E154. <https://doi.org/10.1097/CCM.00000000000004121>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

2. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297-1310
3. Zhang Z: Model building strategy for logistic regression: Purposeful selection. *Ann Transl Med* 2016; 4:111
4. 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:i140
5. Zhang Z, Zhang H, Khanal MK: Development of scoring system for risk stratification in clinical medicine: A step-by-step tutorial. *Ann Transl Med* 2017; 5:436

DOI: 10.1097/CCM.0000000000004037

The authors reply:

We thank Dr. Chen (1) for his interest in our recently published study (2) in *Critical Care Medicine*, and for his comments. Unfortunately, it was unclear to Dr. Chen (1) which cohort was used to compare the performance (i.e., area under the curve [AUC]) of each predictive score. The AUCs of all predictive scores were based on patients included in the Simple Intensive Care Studies-I: that is, all consecutive adults acutely admitted to the ICU and expected to stay for at least 24 hours (3). We fully agree that prognostic models tend to perform better in the cohort in which they were developed when compared with external cohorts (4). This is a limitation of our study, which we also emphasized in the discussion; it was conducted in a single-center setting. We hope to validate our findings in a new, comparable cohort within our center (2), but preferably, our findings should be validated in a multicenter external setting. However, we currently lack such data. As suggested by Dr. Chen (1), we have now reconstructed the models by recalibrating the variables used in each score. This indeed led to better predictive performances (**Fig. 1**). Nevertheless, the performance of all four models remained only reasonable.

Most important, we would like to emphasize that the aim of our study was not to develop a new risk score to replace established prognostic scores. We intended to assess the prognostic value of a comprehensive clinical examination and to determine which clinical examination findings herald a worse prognosis. Second, we intended to assess whether the prognostic value of clinical examination would approach the performance of complex models. Physicians may use and can rely on their clinical examination to obtain a rough estimate of prognosis at the bedside. In our view, such a rough estimation currently is “the best available practice” in daily clinical care. This is illustrated by the comparison to existing complex prognostic risk scores because they perform either worse or slightly better, but are still far from perfect. Each patient is subjected to constant alterations regarding their clinical status and treatment regimens and neither clinical examination nor prognostic scores can fully compensate for all these dynamic changes. Our study showed that a single clinical examination already has reasonable prognostic value; repeated clinical examinations might predict the individual patient prognosis more accurately.

The authors have disclosed that they do not have any potential conflicts of interest.

Figure 1. The performance of clinical examination compared with established prognostic scores using multivariable models to distinguish 90-d survivors from nonsurvivors. Figure legend presents the area under the receiver-operating curves with 98.5% CI. The Delong method was used to compare our clinical examination model to three prognostic models commonly used in the intensive care. The prognostic models were recalibrated with the variables used in each prognostic score. APACHE = acute physiology and chronic health evaluation, SAPS = Simplified Acute Physiology Score, SOFA = Sequential Organ Failure Assessment.

Bart Hiemstra, MD, PhD, Department of Anesthesiology, University of Groningen, University Medical Center Groningen, The Netherlands; **Ruben J. Eck, MD**, Department of Critical Care, University of Groningen, University Medical Center Groningen, The Netherlands, and Department of Internal Medicine, University of Groningen, University Medical Center Groningen, The Netherlands; **Frederik Keus, MD, PhD**, Department of Critical Care, University of Groningen, University Medical Center Groningen, The Netherlands; **Iwan C. C. van der Horst, MD, PhD**, Department of Intensive Care, University of Maastricht, Maastricht University Medical Center+, Maastricht, The Netherlands, and SICS Study Group, Department of Critical Care, University of Groningen, University Medical Center Groningen, The Netherlands

REFERENCES

1. Chen J: External Validation Is Required for Comparisons of the Diagnostic Performance of Several Predictive Scores. *Crit Care Med* 2019; 48:e153–e154
2. Hiemstra B, Eck RJ, Wiersema R, et al; SICS Study Group: Clinical examination for the prediction of mortality in the critically ill: The Simple Intensive Care Studies-I. *Crit Care Med* 2019; 47:1301–1309
3. Hiemstra B, Eck RJ, Koster G, et al; SICS Study Group: Clinical examination, critical care ultrasonography and outcomes in the critically ill: Cohort profile of the Simple Intensive Care Studies-I. *BMJ Open* 2017; 7:e017170
4. Labarère J, Bertrand R, Fine MJ: How to derive and validate clinical prediction models for use in intensive care medicine. *Intensive Care Med* 2014; 40:513–527

DOI: 10.1097/CCM.0000000000004121