

## Fair comparisons in the arena of risk scores

Junfeng Wang<sup>1</sup>, Ronald B. Geskus<sup>2,3,4</sup>, Mariska M.G. Leeflang<sup>4</sup>, Cyriel Y. Ponsioen<sup>5,\*</sup>

1. Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

2. Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

3. Nuffield Department of Medicine, University of Oxford, United Kingdom

4. Amsterdam UMC, University of Amsterdam, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health, Amsterdam, the Netherlands

5. Amsterdam University Medical Centers, location AMC, Department of Gastroenterology & Hepatology, Amsterdam, the Netherlands.

\* Corresponding author: Prof. Dr. Cyriel Y. Ponsioen

Mailing address: Meibergdreef 9, 1105AZ, Amsterdam, the Netherlands

Tel: +31 20 5666012

Fax: +31 20 6917033

E-mail: c.y.ponsioen@amsterdamumc.nl

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/HEP.30964](https://doi.org/10.1002/HEP.30964)

This article is protected by copyright. All rights reserved

## Footnote Page

### Contact Information:

Prof. Dr. Cyriel Y. Ponsioen

Mailing address: Meibergdreef 9, 1105AZ, Amsterdam, the Netherlands

Tel: +31 20 5666012

Fax: +31 20 6917033

E-mail: c.y.ponsioen@amsterdamumc.nl

### List of Abbreviations:

PSC: Primary sclerosing cholangitis

AST: Aspartate aminotransferase

APRI: AST-to-platelet ratio index

Financial Support: This work was supported by the Wellcome Trust (grant number 106680/Z/14/Z)

Disclosures: dr. Ponsioen has received research grants from Takeda, has received speaker's fees from Takeda, Tillotts, and Abbvie, and served as an adviser for Takeda, Pliant, and Shire.

To the Editor:

We congratulate Goode and colleagues on their work(1). We are happy to see several research groups are working on this topic, underscoring the importance of the need for predicting outcome of PSC. The authors mention the Amsterdam-Oxford PSC score, which contains similar variables and evaluates almost the same outcomes and state that it has a moderate C statistic of 0.68, and suggest the UK-PSC score is superior due to its C-statistic of 0.80.

However, a C-statistic is not an inherent feature of a model, it also depends on the features of the cohort: e.g. censoring distribution, correlation between predictors and outcome, even correlation between predictors, these can all influence the best possible C-statistic that can be achieved.

Therefore, a fair way to compare two models predicting the same outcome is a direct comparison in the same external validation dataset. Unfortunately, the authors only compared the performance of the UK-PSC score with the Mayo score and AST-to-platelet ratio index (APRI), and did not assess the Amsterdam-Oxford PSC score as well.

For an indirect comparison of the UK-PSC score with the Amsterdam-Oxford score we would like to raise the following points: The short-term 2 year risk score ( $RS_{ST}$ ) takes advantage of the fact that early events are easier to predict. If we apply the Amsterdam-Oxford score only for the first two years, the C-statistic is 0.76 (95% CI 0.58-0.94) in our external validation instead of 0.68. For the 10-year risk of outcome score ( $RS_{LT}$ ), the longest prediction is 10 years (prediction at 2 years post diagnosis over the 8 following years), whereas the Amsterdam-Oxford PSC score was evaluated with its prediction up to 30 years from time of diagnosis(2). If we evaluate the Amsterdam-Oxford score only for events within 10 years, the C-statistic is 0.71 (95% CI 0.62-0.80) in our external validation.

Furthermore, in Figure 3 calibration is not assessed by comparing predicted survival versus observed survival probability, it is actually between observed survival in the derivation cohort versus the validation cohort(1). From this figure, one cannot conclude the calibration of  $RS_{LT}$  is good.

From a practical point of view,  $RS_{LT}$  needs information both at T0 and T2, thus prediction is only possible at T2. This is undesirable when counselling patients as to their prognosis after establishing a diagnosis of PSC. Instead of building a model at 2 years, an alternative method is to apply a so-called dynamic prediction model, which can better reflect the dynamics of the disease progress than a baseline model(3).

## References:

1. Goode EC, Clark AB, Mells GF, Srivastava B, Spiess K, Gelson WTH, Trivedi PJ, et al. Factors Associated With Outcomes of Patients With Primary Sclerosing Cholangitis and Development and Validation of a Risk Scoring System. *Hepatology* 2019;69:2120-2135.
2. **de Vries EM, Wang J**, Williamson KD, Leeflang MM, Boonstra K, Weersma RK, Beuers U, et al. A novel prognostic model for transplant-free survival in primary sclerosing cholangitis. *Gut* 2018;67:1864-1869.
3. Rizopoulos D. Joint models for longitudinal and time-to-event data: With applications in R: Chapman and Hall/CRC, 2012.

Author names in bold designate shared co-first authorship