Re: Gene Expression–Based Prognostic Signatures in Lung Cancer: Ready for Clinical Use?

Subramanian and Simon (1) reviewed mRNA-based prognostic signatures for non–small cell lung cancer. They concluded that many studies were poorly executed and that no gene expression signature is ready for clinical application. We agree that guidelines for design, reporting, and analysis are needed but have reservations concerning the accuracy of their review.

We examined the 20 scores that they assigned to the two signatures reported by our group (2,3) and found several errors. The study by Lau et al. (3) received a score of 0 for “statistically significant improvement over standard risk factors,” despite the fact that 95% confidence intervals (CIs) for the difference in concordance index between a clinical model and a model containing clinical and molecular information were presented in Table A5. This table demonstrates that zero is excluded from the 95% CI of the difference in a validation set from Duke (95% CI = 0.018 to 0.162) and was included in validation sets from Harvard (95% CI = −0.018 to 0.079) and Michigan (95% CI = −0.046 to 0.122). The study by Boutros et al. (2) received a score of 0 for “validation in stages IA and IB”; however, Boutros et al. (2) presented a stage-adjusted validation of 345 stage I patients in Figure S3. That study also provided a complete model specification readily replicable by experts in the field: “[Euclidean] distance between the expression profile for each patient and the cluster centers (medians).” Nevertheless, a score of 0 was assigned.

With regard to the validations that Subramanian and Simon performed, we found that our specified procedure was not
followed. They tested our three-gene signature (3) in an independent dataset (4) using data preprocessed with the model-based expression indices with pseudocount addition. Our method requires each Director’s Challenge subset to be separately preprocessed using the robust multiarray average algorithm, with no pseudocount addition. This change alters patient classification and reaffirms the need to follow author-specified procedures exactly.

We also question the appropriateness of the binary scoring criteria used by Subramanian and Simon. For example, a study that provides three of the four specified clinical covariates receives the same score as the one that provides none. Of greater concern, a study that provides summary information receives the same score as the one that provides individual patient annotation. Similarly, although we agree with Subramanian and Simon about the importance of adequate tissue handling, penalizing a study for not including a specific sentence detailing tissue handling presumes that such methods were not followed, when indeed they are standard. As such, the Director’s Challenge study (4), in which a strict tissue handling protocol was enforced and described, would be given a score of 0 for omitting this specific sentence. We also disagree with the authors that studies should be penalized for showing training set performance (ie, resubstitution statistics). Although resubstitution statistics provide an optimistically biased estimate of performance, they still have value at a number of levels. First, it is important that all results are disclosed; second, the direction of effect is important; third, the difference between the training and validation sets is a measure of overfitting and is important for future studies; fourth, demonstration of the training set performance allows evaluation of the signature on its initial patient population. The authors view resubstitution statistics as unnecessary and undesirable. We view them as necessary, but not sufficient, for signature evaluation.

Finally, we believe that all datasets and validations should be published. Publication of failed validations provides datasets that are critical for the development and validation of more mature markers and reduces negative publication bias (5). Given the small sample sizes of existing datasets, validation in many independent patient populations is extremely important. According to the criteria proposed by Subramanian and Simon, only two datasets have appropriate patient characteristics, and for one of these (6), the full data are not publicly available.

There remains a strong clinical need for improved prediction of non–small cell lung cancer prognosis (7). A systematic review is timely, but the authors of such reviews—especially when scoring studies according to inclusion of specific data—have a responsibility to present complete and accurate results.

PAUL C. BOUTROS
MELANDA PINTILIE
THOMAS JOHN
MAUD H. W. STARMANS
SANDY D. DER
FRANCES A. SHEPHERD
MING-SOUND TSAO
IGOR JURISICA

References

Notes
Affiliations of authors: Informatics and Biocomputing Platform, Ontario Institute for Cancer Research, Toronto, ON, Canada (PCB, MHS); Department of Radiation Oncology, Maastricht University Medical Center, Maastricht, the Netherlands (MHWS); Ontario Cancer Institute and the Campbell Family Institute for Cancer Research, University Health Network, Toronto, ON, Canada (IMP, TJ, SDD, FAS, M-ST, IU).
Correspondence to: Paul C. Boutros, PhD, Ontario Institute for Cancer Research, Informatics and Biocomputing Platform, 101 College St, Ste 800, MaRS Centre, South Tower, Toronto, ON, Canada M5G 0A3 (e-mail: paul.boutros@oicr.on.ca) or Igor Jurisica, PhD, The Campbell Family Institute for Cancer Research and Ontario Cancer Institute, 101 College St, TMĐT 9-305, Toronto, ON, Canada M5G 1L7 (e-mail: juris@cs.toronto.edu).
DOI: 10.1093/jnci/djq385
© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org. Advance Access publication on October 4 2010.