

Calibration evaluation in left skewed populations.

Yannick Le Manach, M.D., Ph.D., Gary Collins, Ph.D., Reitze Rodseth, M.B.Ch.B., Ph.D.,
Christine Le Bihan-Benjamin, M.D., M.Sc., Bruce Biccard, M.B.Ch.B., Ph.D., Bruno Riou,
M.D., Ph.D., P.J. Devereaux, M.D., Ph.D., Paul Landais, M.D., Ph.D.

We thank Kappen *et. al.* for their interest and constructive comments on our paper describing the development of the POSPOM model for predicting postoperative mortality. The inconsistencies Kappen et al identified are consequences of an unfortunate error in the text. All tables and figures have been verified and are correct.

We created POSPOM to predict in-hospital mortality in any adult patients following surgery. To achieve this goal, we considered all patients undergoing surgery in France during 2010. As a consequence, the observed mortality reflects the one for the overall surgical population, and is not comparable to what have been reported in selected patients' subgroups with higher risks.^{1,2}

In our validation cohort of 2,789,932 patients, only 22136 (0.79%) had a predicted risk of in-hospital mortality greater than 10%. As a consequence, we intentionally presented a truncated calibration plot focusing on a range of probabilities including more than 99% of the patients. As noted by Kappen *et. al.*, POSPOM overestimates the risk of in hospital mortality in patients with a risk greater than 10%; the observed in hospital mortality was 13.3%, and the average predicted risk in these patients was 17.1%. The figure provided by Kappen *et. al.* is then an acceptable

representation of the agreement between predicted probabilities and observed in-hospital mortality in the validation cohort.

From a clinical point of view, we believe that any preoperative risk of in-hospital mortality greater than 10% (*i.e.* 20 times the average risk in our population) reflects very high risk procedures. The role of a general preoperative assessment tool (*e.g.* POSPOM) is not to distinguish patients with a risk of 23% and a risk of 62% of postoperative mortality. However, both cases are uncommon (409 patients presented a POSPOM equal of greater than 40 in the validation cohort, namely 0.01% of the population) and concerning clinical situations which require further preoperative investigations to determine an appropriate care strategy.

The reporting of calibration for prediction models remains difficult as underlined by the comment of Kappen *et. al.* Graphical analysis of the calibration plot remains the preferred approach,^{3,4} the distribution of the predicted probabilities in the validation cohort and the agreement between predicted probabilities and observed mortality are analyzed simultaneously. Due to a heavy left skewed distribution, whereby most patients were at very low risk of in-hospital mortality following surgery, including a histogram of predicted probabilities was uninformative. However, because the calibration plot included more than 99% of the patients, we maintain our interpretation: in the validation cohort, POSPOM “has good calibration with only a small underestimation of in-hospital mortality in the validation cohort for predicted probabilities ranging from 1 to 10%”. However, we agree with the authors that a closer inspection of those patients with high probabilities of in-hospital mortality is required.

Beyond the discussion about the appropriateness of the calibration of POSPOM in our validation cohort,⁵ we believe that an evaluation of the POSPOM performances in a completely different

cohort (*i.e.* true external validation) is the necessary next step for the implementation of this promising predictive tools.

References:

1. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation Study I, Devereaux PJ, Chan MT, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA*. 2012;307(21):2295-2304.
2. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet*. 2012;380(9847):1059-1065.
3. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Annals of internal medicine*. 2015;162(10):735-736.
4. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*. 2015;162(1):W1-73.
5. Le Manach Y, Collins G, Rodseth R, et al. Preoperative Score to Predict Postoperative Mortality (POSPOM): Derivation and Validation. *Anesthesiology*. 2016;124(3):570-579.