The external validation of the EuroSCORE II risk stratification model

Dusko Nezic a,*, Milorad Borzanovic b and Tatjana Spasic a

a Department of Cardiac Surgery, ‘Dedinje’ Cardiovascular Institute, Belgrade, Serbia
b Department of Cardiology, ‘Dedinje’ Cardiovascular Institute, Belgrade, Serbia

* Corresponding author. Department of Cardiac Surgery I, ‘Dedinje’ Cardiovascular Institute, M. Tepica 1, 11000 Belgrade, Serbia. Tel: +381-11-3601631/3601724; fax: +381-11-2666392; e-mail: nezic@eunet.rs; nezic@ikvbd.com (T. Spasic).

Received 12 May 2013; accepted 28 May 2013

Keywords: EuroSCORE II • Risk modelling

We read with great interest the article by Chalmers et al. [1] regarding external validation of the EuroSCORE II performances. Since the EuroSCORE (European System for Cardiac Operative Risk Evaluation) was introduced in daily cardiac surgical practice [2], it has been used as a measure of operative risk of adult cardiac surgery patients in more than 1300 formal citations in the medical literature [3]. Although both the additive and the logistic versions of the EuroSCORE have remained very good discriminatory power, suspicions were raised that the model may now be inappropriately calibrated for current cardiac surgery [3]. Therefore, the old EuroSCORE was renewed into EuroSCORE II (database of 22 381 consecutive patients undergoing adult cardiac surgery in 154 hospitals in 43 countries over a 3 month period in 2010) in order to optimize its discriminatory power and especially its calibration [3].

Although the Hosmer–Lemeshow test was used in 95% of published manuscripts to test the calibration of the old EuroSCORE [4], the calibration of the EuroSCORE II was assessed using the observed/expected (O/E) ratio of mortality. Ideally, this ratio equals one (the observed mortality equals expected mortality, thus the predictive model is perfectly calibrated). A value above one means that the model underestimates mortality, whereas a value below one means that model overestimates mortality. If the 95% confidence interval (CI) of the O/E ratio excludes the value 1.0, it may be considered statistically significant [4]. Chalmers et al. [1] used Hosmer–Lemeshow statistics to validate the calibration ability of the EuroSCORE II (a single-institution experience; 5576 patients, including coronary surgery, valve surgery, mixed cases, aortic surgery and miscellaneous cases). The Hosmer–Lemeshow test confirmed good calibration of the EuroSCORE II model for all subgroups of cardiac surgery patients (P-values of 0.052–0.99), while overall calibration (all 5576 patients) was poor (P < 0.001). The calculated O/E ratio (using the data presented in Table 1 of Chalmers et al. [1]) and its 95% CI confirmed good calibration for all subgroups, except for miscellaneous cases (the model underestimated mortality). Observed mortality for all patients was 2.2%, and predicted mortality (EuroSCORE II) was 2.0% (112 patients). The O/E ratio confirmed good calibration. This would be a nice argument to support the opponents of Hosmer–Lemeshow statistics, were these data not misleading. In fact, all the data of categorical variables showed in the second column (all patients) of Table 1 of Chalmers et al. [1] were miscalculated (i.e. hypercholesterolaemia, 3866 of 5576, as 84.3% instead of 69.3%; urgent operation, 1295 of 5576 as 28.3%, instead of 23.2% etc.). In-hospital mortality has also been presented incorrectly. Neither is 101 of 5576 patients 2.2% (it is 1.8%), nor is the total number of all patients who died 101 (the correct overall number of patients who died is 191; then corrected mortality appears to be 3.43%). That leads us to the conclusion that the O/E ratio (191/112) is 1.71 (95% CI 1.47–1.95), thus not including the value of 1.0. Therefore, overall poor calibration has been confirmed with the O/E ratio of mortality, as well as with Hosmer–Lemeshow statistics.

REFERENCES