third time after the second application, but no further CoSeal was applied if an air leak persisted at this stage.

REFERENCES


LETTER TO THE EDITOR

The EuroSCORE: are we contributing to its overprediction of mortality in cardiac surgery nowadays?

Dusko Nezic*, Milorad Borzanovicb, Aleksandar Knezevic and Miomir Jovic
d

* Department of Cardiac Surgery, ‘Dedinje’ Cardiovascular Institute, Belgrade, Serbia
b Department of Cardiology, ‘Dedinje’ Cardiovascular Institute, Belgrade, Serbia
c Department of Anaesthesiology, ‘Dedinje’ Cardiovascular Institute, Belgrade, Serbia
d Institute of Faculty of Medicine, University of Belgrade, Belgrade, Serbia

† Corresponding author. Department of Cardiac Surgery, ‘Dedinje’ Cardiovascular Institute, Milana Tepića 1, 11000 Belgrade, Serbia. Tel: +381-11-3601631; fax: +381-11-2666392; e-mail: nezic@eunet.rs; nezic@ikvbd.com (D. Nezic).

Received 28 June 2011; accepted 18 August 2011

Keywords: EuroSCORE • Calibration • Adequate statistical testing

We read with great interest the article by Akar et al. [1] regarding validation of the EuroSCORE risk models in the Turkish adult cardiac surgical population. This manuscript is in agreement with previous reports [2–4] that additive and logistic EuroSCORE models overpredict mortality in cardiac surgery nowadays. However, in our opinion there are a few issues that should be clarified. The first one is a definition of mortality. The original EuroSCORE model [5] defined operative mortality as death within 30 days from operation or even later than 30 days if still in hospital (not 30-day mortality as the authors stated in Discussion section of their manuscript [1]). On the other side, their calculation and statistical observations were based only on in-hospital mortality. It is possible that, with the inclusion of all patients who should be included as perioperative deaths (30-day mortality plus in-hospital mortality), the findings of the study [1] would no longer be statistically significant.

When we validate a risk stratification model on a population out of its original database, we have to check its discriminatory power and calibration. Model discrimination was analysed by determining the area under the receiver operating characteristic (ROC) curve [1–5]. However, several statistical procedures have been used to determine calibration. The Hosmer–Lemeshow (H–L) test has been used in >90% of all manuscripts to test the EuroSCORE calibration in different patient population (the H–L test confirms good calibration if the P-value is >0.05), including the EuroSCORE working group [5]. In a few manuscripts calibration plots, unpaired t-test [1], chi-square test [4] or observed to expected (O/E) ratios were used to prove that calibration of the EuroSCORE models was adequate. Thus, we are curious as to what would be the results of the EuroSCORE calibration checking in these articles if the H–L test was used. On the contrary, Yap et al. [2] wrote an excellent manuscript with intention of validating the EuroSCORE model in Australia. Mortality definition was correct (in-hospital mortality plus 30-day mortality), areas under the ROC curves (0.82) confirmed an excellent discrimination for both additive and logistic EuroSCORE models, but calibration using the H–L test was poor (P < 0.05) for both EuroSCORE variants. Although D’Errigo et al. [3] reported the logistic EuroSCORE to be inadequate to predict mortality in the Italian CABG population (area under ROC curve, 0.77; H–L test; P < 0.0001), the additive EuroSCORE has confirmed very good discrimination (area under the ROC curve of 0.77) and calibration (H–L test; P = 0.228) for the same cohort. An O/E ratio of 1 means that the score predicts mortality perfectly. An O/E ratio of 0.30 means that the score overpredicts mortality by 70% [6]. Badreldin et al. [6] have shown that although the O/E ratio in their CABG population was 0.32 (the logistic EuroSCORE overpredicts mortality by 68%), the H–L test confirmed good calibration (P > 0.05) of the logistic EuroSCORE model.

Finally, our question is—should we use the H–L test any more, or we are looking for a statistical procedure that will confirm poor calibration of the EuroSCORE models?
REFERENCES


LETTER TO THE EDITOR RESPONSE

Reply to Nezic et al.
Ahmet Ruchan Akar*, Murat Kurtcepeh*, Serkan Durdu* and Halil Altay Güvenir*

a Department of Cardiovascular Surgery, Ankara University School of Medicine, Ankara, Turkey
b Department of Computer Science, Case Western Reserve University, Cleveland, USA
c Department of Computer Engineering, Engineering Faculty, Bilkent University, Ankara, Turkey
* Corresponding author. Department of Cardiovascular Surgery, Heart Center, Ankara University School of Medicine, Dikimevi, Ankara 06340 Turkey. Tel: +90-533-6460684; fax: +90-312-3625639; e-mail: akarruchan@gmail.com (A. Ruchan Akar).
Received 15 August 2011; accepted 18 August 2011

Keywords: Cardiac surgery · Risk prediction model · Mortality · EuroSCORE

Thank you for the opportunity to reply to the letter from Nezic et al. [1]. In their letter, Nezic et al. first expressed concerns about the potential differences in the definition of mortality between our validation study [2] and original EuroSCORE models [3]. In the discussion section of our article, we only aimed to point out the most commonly used risk prediction models in current Turkish cardiac surgical practice. In fact, in our study, we defined operative mortality as any death, regardless of cause, occurring within 30 days after surgery, in or out of the hospital, and after 30 days during the same hospitalization subsequent to the original operation. Thus, our mortality definition is similar to original EuroSCORE definitions and our findings regarding mortality over-prediction of EuroSCORE is valid.

Second, as suggested by Nezic et al. [1] we used Hosmer-Lemeshow (H-L) goodness-of-fit statistics to test EuroSCORE calibration in our patient population. For the entire cohort, the calibration values were 5.5 (P = 0.357) and 53.0 (P < 0.05) for additive and logistic models, respectively. Furthermore, for isolated CABG subgroup, we obtained similar results with the calibration values of 6.9 (P = 0.231) and 19.0 (P < 0.05) for additive and logistic EuroSCORE models, respectively. Thus, H-L tests confirmed good calibration for the additive EuroSCORE model in our patient cohort, unlike the logistic EuroSCORE model. Nevertheless, a significant H-L test does not necessarily mean that a predictive model is not useful. As Nezic et al. [1] mention the contradictive findings regarding the usefulness of H-L statistics, we agree that further studies are needed to clarify the usefulness and necessity of H-L statistics.

Furthermore, we believe that a statistical prediction model, namely linear regression, trained on a certain population cannot operate ideally on a different population target, which we showed in our previous work [2]. There can be two different approaches to overcoming this handicap. The first one would be using the same predictors as EuroSCORE does and build up a linear model by using these predictors on the Turkish population. Current cardiac surgical mortality in our cohort has been running 0.49–0.67 of EuroSCORE prediction for different risk subgroups. The second approach would be to come up with a better model, probably with different predictors and specifically trained and tested on Turkish population. Our group is currently working on the second approach.

Conflict of interest: none declared.

REFERENCES

