The surgeon, the statistics and the data

Studies which evaluate the outcome of surgical activity are important to increase knowledge, to improve quality of care and of course to organize the planning and economics of health care systems.

Ranking health care providers can contribute to improving the quality of care delivered. Nevertheless, there is debate in the scientific community on whether such a ranking should be openly published or analyzed by peer review in a select professional environment. It has been clearly demonstrated that the publication of an outcome registry with ranking of medical care providers reduces the published mortality rates, but this may be achieved partly by undertreatment of sicker patients in an attempt to achieve a better ranking[1—3].

It goes without saying that when such a registry is published, the quality and reliability of the data collection must be near perfect. Without this, results are inconsistent, misleading and potentially dangerous. Actual mortality results can be stratified by risk and analyzed by comparison to expected mortality (mortality index) but, once again, such an index is only as precise as the database from which it is generated.

This issue of the European Journal of Cardio-thoracic Surgery features an article by D’Errico and co-workers about the correlation between the EuroSCORE model[4] and the results of the Italian CABG outcome study published last year in the same journal[5]. The new article is a different analysis of the same database. In the conclusion of this paper D’Errico and co-workers affirm that ‘the EuroSCORE model can be exported to the Italian population and used to rank hospital performance and evaluate preoperative risk of patients undergoing open heart surgery’.

This paper is well written but unfortunately contains important methodological and conceptual errors, and it can generate dangerous consequences if it were to be considered as a correction tool for EuroSCORE. The EuroSCORE could be considered obsolete and therefore we can speculate the Italian Risk Calculator must be used to define the risk of our patients.

For this reason we must be sure that the Italian Risk Calculator is sure, realistic and reliable.

The major concern about the Italian CABG outcome study was the quality of the data collected, and an Editorial Comment[6], published in the same issue of the journal, emphasized the bias present in the study. The same bias is evidently present in this new analysis of the same data. Briefly, the concerns about the data collection in the Italian CABG outcome study were:

1. The study covered a period between 1st January 2002 and 30th September 2004. Centers were allowed to present data which did not cover the entire period: it was enough to collect data for a minimum period of 6 months, and therefore the range of data collection was very different between centers (range 6—29 months). This opens the possibility for selective reporting and is clearly against any guideline for observational studies.
2. The reported incidence of some risk factors was too wide between centers and goes beyond what is normally observed in other multi-institutional studies like EuroSCORE[7] or the Italian Society of Cardiac Surgery data[8]. For example, some centers reported an incidence of unstable angina of 62% or cardiogenic shock of 7.2%. It seems that, at least for some risk factors, adherence to definition was variable.

This anomaly was emphasized in the discussion that followed the publication of the study, and the authors affirmed that ‘prevalence of emergency and cardiogenic shock is consistent with the literature’[9]. This is true if we consider the mean, but this paper is a ranking analysis and it is evident the adjusted mortality for a center with 7.2% incidence of cardiogenic shock is completely different from a center with 0.5% incidence.

It is surprising that in this new analysis[4] the authors recognize that ‘there is a great discrepancy between the data of EuroSCORE and the data we found for unstable angina, respiratory disease and critical preoperative status’. They justify this discrepancy by citing possible differences in the Italian population and differences in the periods covered by the two studies. One explanation is that today’s patients are worse because in the angioplasty era, more critical patients are sent to the surgery. This is not entirely correct. Although it is true that the surgical population is older than before, but the critical operative status and the comorbidities are seen by many as strong indicators for angioplasty and not for surgery. Coronary anatomy is a far stronger determinant of surgical treatment than comorbidities for clinical conditions.

Another indicator for the possible presence of bias in the data is that the best surgical results were achieved in low
workload hospitals. All high workload hospitals were in the low or lowest ranks. This is inconsistent with many reports examining the relationship between volume and outcome in complex surgery [10] and, if true, has serious implications for the planning and management of cardiac surgical services in our country.

The Italian surgical community is small and of course we know each other well. My hospital, with 2000 procedures per year is one of the most active in Italy and belongs to the lowest rank group. We know that the highest rank hospitals are excellent organizations and have outstanding surgeons, but the fact that one of the best performing centers has no emergency department and during the period of observation was closed for 24 days in summer and 18 days over Christmas, indicates that there may be something amiss in the data collection and raises the possibility that judicious patient selection can play a certain role.

With this in mind, the comparison between EuroSCORE and an empirically derived Italian risk-model seems to me a very difficult statistical exercise which reminds me of Dr Akins when proposed at the AATS meeting his First Law of Statistics:

‘Torture the numbers and they will admit everything’.

This new publication is retrospective and not a prospective analysis with all limits of these analyses.

We know because it was stressed by the authors themselves in the previous paper that ‘there is a substantial difference between ‘predictive models’ aimed to predict the occurrence of outcome, and ‘explicative models’ aimed to control confounding in outcome comparison between centers or populations’ [5] (p. 60); but if the two models are so different, why compare apples to pears? The authors probably need to find a method to justify in some way a wrong ‘explicative model’.

The authors recognize that they did not collect the whole set of variables necessary for the EuroSCORE calculation [4], indeed the information about the ‘Critical preoperative state’ was not available. The weight of this risk factor is 3 in the additive EuroSCORE and higher in the logistic EuroSCORE if other risk factors are present. It is therefore an important contributor to expected mortality. The definition of ‘respiratory dysfunction’ was slightly different.

Adherence to the definition of variables is a basic requirement for such a study as it may influence the final score calculation. It is impossible to compare accuracy and predictive power of the risk model if the exact EuroSCORE data are not available.

As surgeons who use EuroSCORE on a daily basis, we know that it is not perfect and needs refinement and adjustment, but this paper is not the solution.

Last but not least, in the authorship of this paper, there is not one surgeon belonging to the Italian Society of Cardiac Surgery or to European Association for Cardio-Thoracic Surgery, or to the EuroSCORE project group. This raises the question: who is the owner of the data?

Any other comment is up to the readers.

References