Author's response to reviews

Title: Stratification of the severity of critically ill patients with classification trees

Authors:

Javier Trujillano (jtruji@cmb.udl.es)
Mariona Badia (marionabadia@wanadoo.es)
Luis Servia (lservia@telefonica.net)
Jaume March (jaume.march@cmb.udl.cat)
Angel Rodriguez-Pozo (rodriguez@medicina.udl.cat)

Version: 2 Date: 28 April 2009

Author's response to reviews: see over
Dear Sir,

Hereby we attach the revision of the original text, so that you may consider if it could be published in your journal.

OUR ANSWER TO THE REVIEWERS

Reviewer 1

1- In Methods a paragraph has been added explaining the data acquisition and score measurement procedures (APACHE II, SAPS II AND MPM II-24):
“Each patient’s medical records and laboratory database files were used to obtain information pertaining to baseline (at ICU admission) demographics, pre-existing comorbidities and scores (APACHE II, SAPS II and MPM II-24). The data were then compiled (manual recording) into single data using a relational database management system (Microsoft Access).”

2- The database gathers information of a ten-year span. A table (Table 3) has been added with mortality tendency, percentage of development/validation sets and score measurement. The table is commented in Results. No significant changes have taken place during that period of time.

Reviewer 2

1- In Discussion a paragraph has been added commenting on the possibility that CTs may split continuous variables into more than two categories:
“The CT software allows to adjust the levels and the number of partitions for each branch in order to get more complex models [7]. In our case, our only restriction (in the 3 CT models) was that the minimum number of subjects in the terminal nodes should be of 50 patients.”

2- The abstract error has been corrected.

3- The reviewer thinks it fit (since it is a model not developed in the work) that the results of the classic scores (APACHE II, SAPS II and MPM II-24) should not be separated into two groups (development and validation), but we believe that if they are placed in another table (making no distinction in the development and validation sets) the differences with the developed methods (LR and CTs) will
not be appreciated. We also believe that seeing a similar behaviour of the classic scores in the two groups somehow reinforces the results (greater homogeneity of the development and validation sets).

4- The reviewer points out the possibility of carrying out a LR model with the continuous variables turned into categorical. This is a possible model, but in order to choose the cut-off points a procedure must be used (there are several methodologies). One of the advantages of the CTs is precisely that the cut-off points of the continuous variables are automatically done. Our aim is not to prove that the CTs are better than models based on LR or scores, but to show some of their advantages. In the Discussion section (Limitations) we have added the following aspect mentioned by the reviewer:

“The LR-based model could have been carried out using the variables as categorical, thus minimizing the possible effect that outlier values (using the variables as continuous) have on the predicted outcome. One of the advantages of CT-based models is that they automatically change the continuous variables into categorical ones and that their cut-off points could also be used to create a LR model with discreet variables.”

5- We have gone deeply into the different selection of variables for the CT and LR models (discussion paragraph):

“We have already seen that the CART CT includes the five general variables. The C4.5 CT adds the MAP (Mean Arterial Pressure) variable and the CHAID CT includes MV (Mechanical Ventilation) variables and the fact of belonging to the trauma group. The LR model uses those of the CHAID CT model, also including the Acute Renal Failure and HR (Heart Rate) variables.”

6- We have added the table of tendency of the data gathered during the ten-year span of the study (in agreement with Reviewer 1)

7- We have added a correlation matrix (Table 7) of the probabilities assigned to each patient for each of the CT and LR models for the development and validation data sets.

Please do not hesitate to contact us for any further queries or inquiries.

Yours sincerely,

Javier Trujillano
jtruji@cmb.udl.es