Supplement: Refitting the Model for End-stage Liver Disease for the Eurotransplant region.

B.F.J. Goudsmit MD,[1,2,3] H. Putter PhD,[4] M.E. Tushuizen MD PhD,[2] S. Vogelaar MD,[3] J. Pirenne MD PhD [5,6], I.P.J. Alwayn MD PhD,[1] B. van Hoek MD PhD,[2*] A.E. Braat MD PhD [1*]

*) joint senior authorship

Affiliations:

Division of Transplantation, Department of Surgery,[1] Department of Gastroenterology and Hepatology,[2] Department of Biomedical Data Sciences,[4] Leiden University Medical Centre, The Netherlands.

Eurotransplant International Foundation,[3] Leiden, The Netherlands

Department of Abdominal Transplant Surgery[5], University Hospitals Leuven, Leuven, Belgium.

Representative of the Eurotransplant Liver Intestine Advisory Committee[6].

Supplement 1. Maximum log-likelihood to ensure best fit of distribution to the data

In this study, the optimal bounds for each MELD parameter are sought. The plotted generalized additive models (GAMs) show the relationship between the parameter and risk of 90-day death. The width of the 95% confidence intervals indicate how precise the estimations are, i.e. how many patients fall within that range. The lower and upper bounds are needed to establish an interval in which the linear regression model can be fitted. Thus, between the lower and upper bound, the model fits a straight line that is used for prediction. However, when establishing bounds, we need to know which combination of bounds best describes the 'true' distribution in the data. The log transformation of the likelihood, i.e. log-likelihood, is calculated for each combination of bounds. The combination with the maximum log-likelihood value is chosen as best fit to the data. The figure illustrates the concept of maximum log-likelihood.

In our data, a certain 'true' pattern of e.g. creatinine measurements is present. This pattern can be expressed through e.g. a mean and standard deviation (SD). Then, we choose a combination of a lower and upper bounds for creatinine. We look at the distribution of the mean and SD of creatinine for that combination of bounds. It is estimated how well this combination of bounds represents the 'true' distribution in our data, by calculating the likelihood. The likelihood increases if the tested distribution more closely resembles the 'true' distribution in the data. We repeated this exercise for every combination of lower and upper bounds. Based on the GAM (figure 1 a-b-c) as guidance, the combination with the maximum log-likelihood is calculated and chosen as best fit to the data.

Fit and discrimination

Although the highest c-index possible should be pursued for allocation purposes, the method of deriving this c-index was of importance and had to be evidence-based, like our log-likelihood approach. It is not valid to adhere to the maximum c-index when choosing boundaries. This was because the c-index increased through the inclusion of more extreme data, which was illustrated by the increasing c-indices at the upper bounds of INR and bilirubin (supplement heatmaps). A more diverse population with more extreme values made it easier for the model to discriminate between survival and death, however acknowledging too many of these extreme values resulted in a bad representation of the true distribution in

the population. This true distribution was optimally matched by choosing the boundaries with the maximum log-likelihood. Thus, the discussion on the establishment of refit models should consider at least three aspects: the fit of the model to the data, the discrimination of the model and the number of patient measurements capped. The weighing of maximum log-likelihood, c-index and clinically relevant bounds is difficult and should be done through careful expert-based consensus. Therefore, to facilitate these discussions, the interactive heatmaps were attached for each parameter (*online heatmaps*).

Supplement 2. Heatmaps of calculated boundaries for the refit parameters. For bilirubin, creatinine, INR and sodium the calculated log-likelihood and c-index per combination of lower and upper bounds are shown. A lighter color indicates a higher log-likelihood or c-index. For each parameter, the lower and upper bound combination with the highest log-likelihood (bright yellow) was chosen (figure 1a, 2a, 3a and 4a).
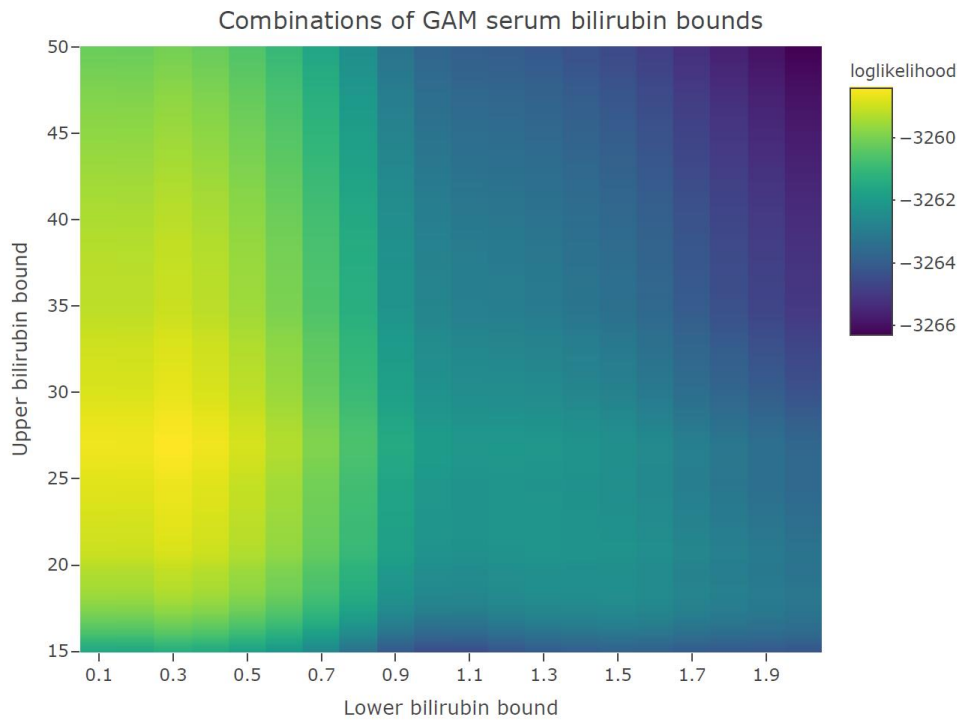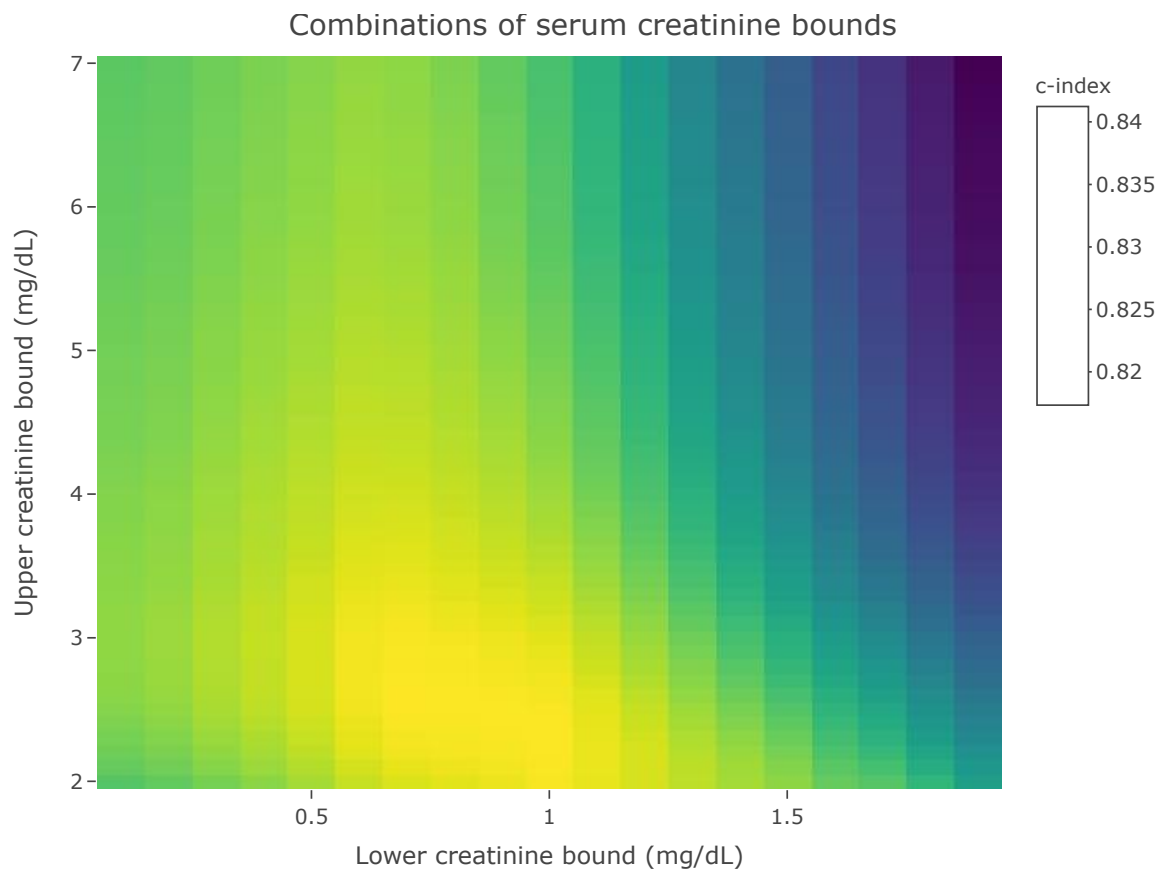

Supplement figure 1a: Heatmap of maximum log-likelihood of bilirubin bounds.
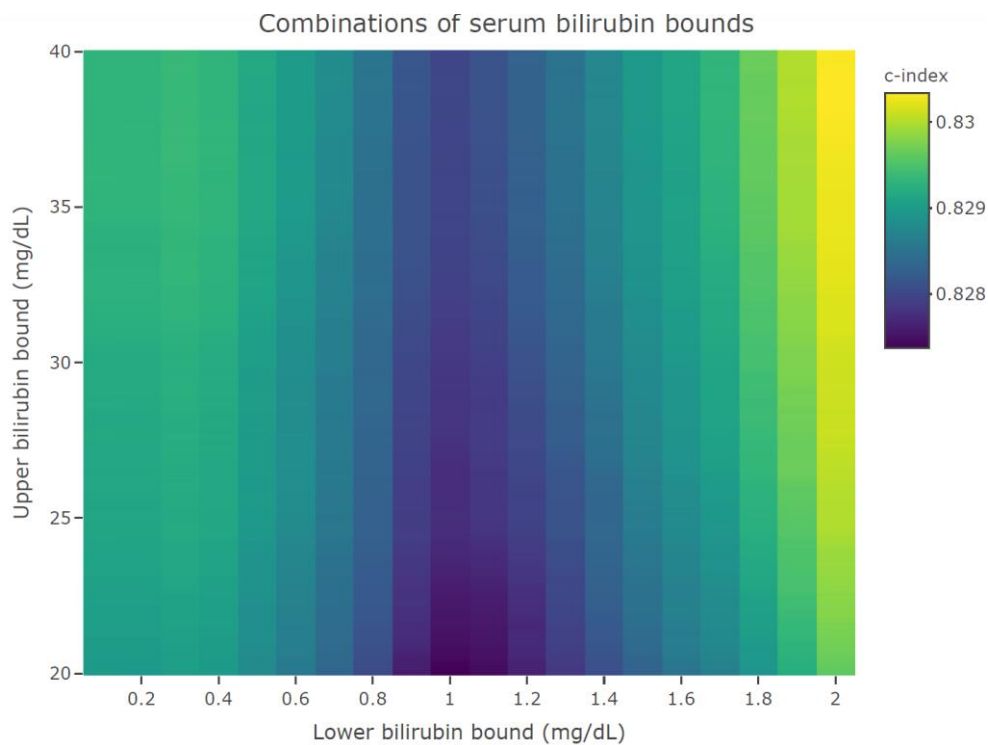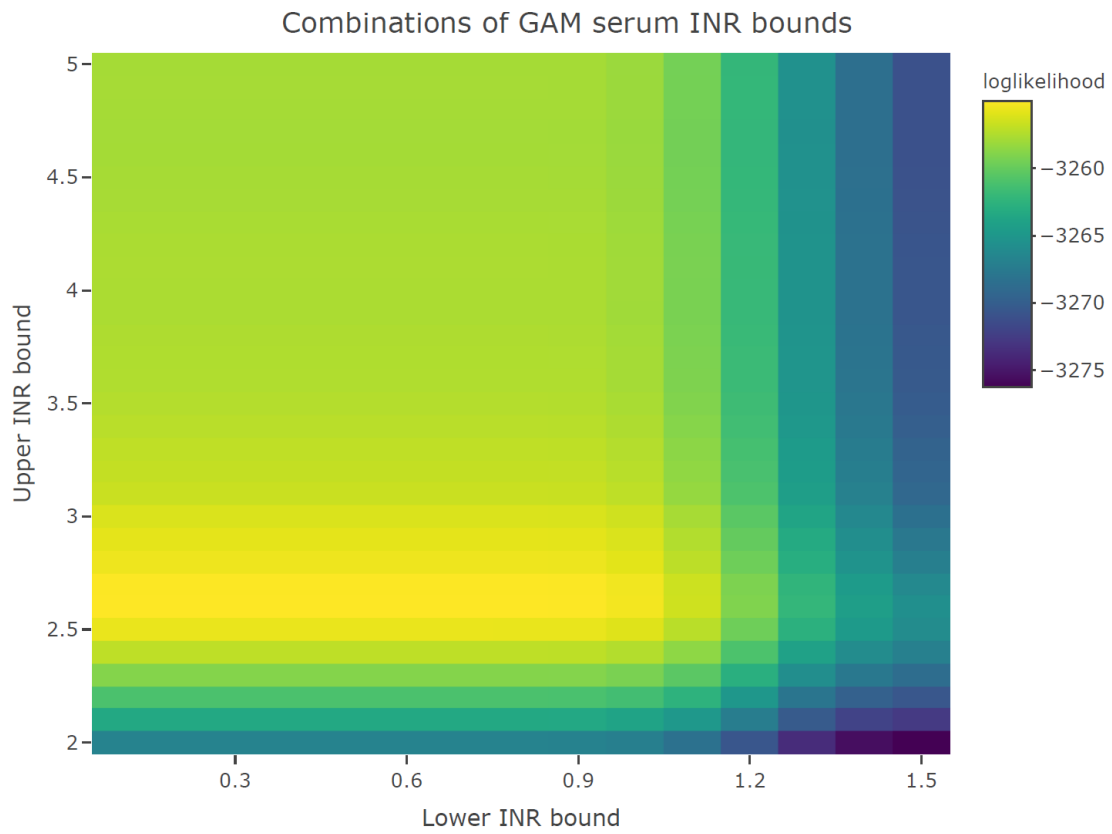

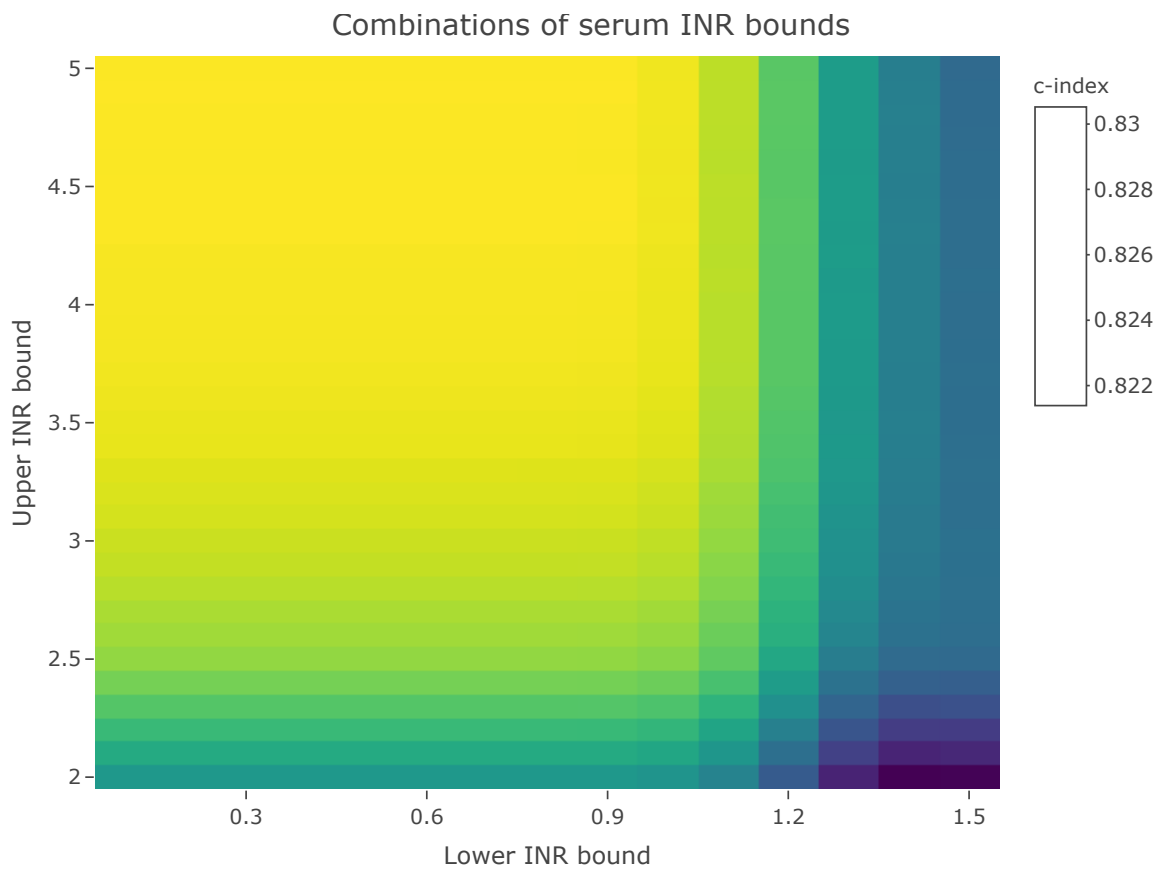Supplement figure 1a. Heatmap of maximum c-index of bilirubin bounds.
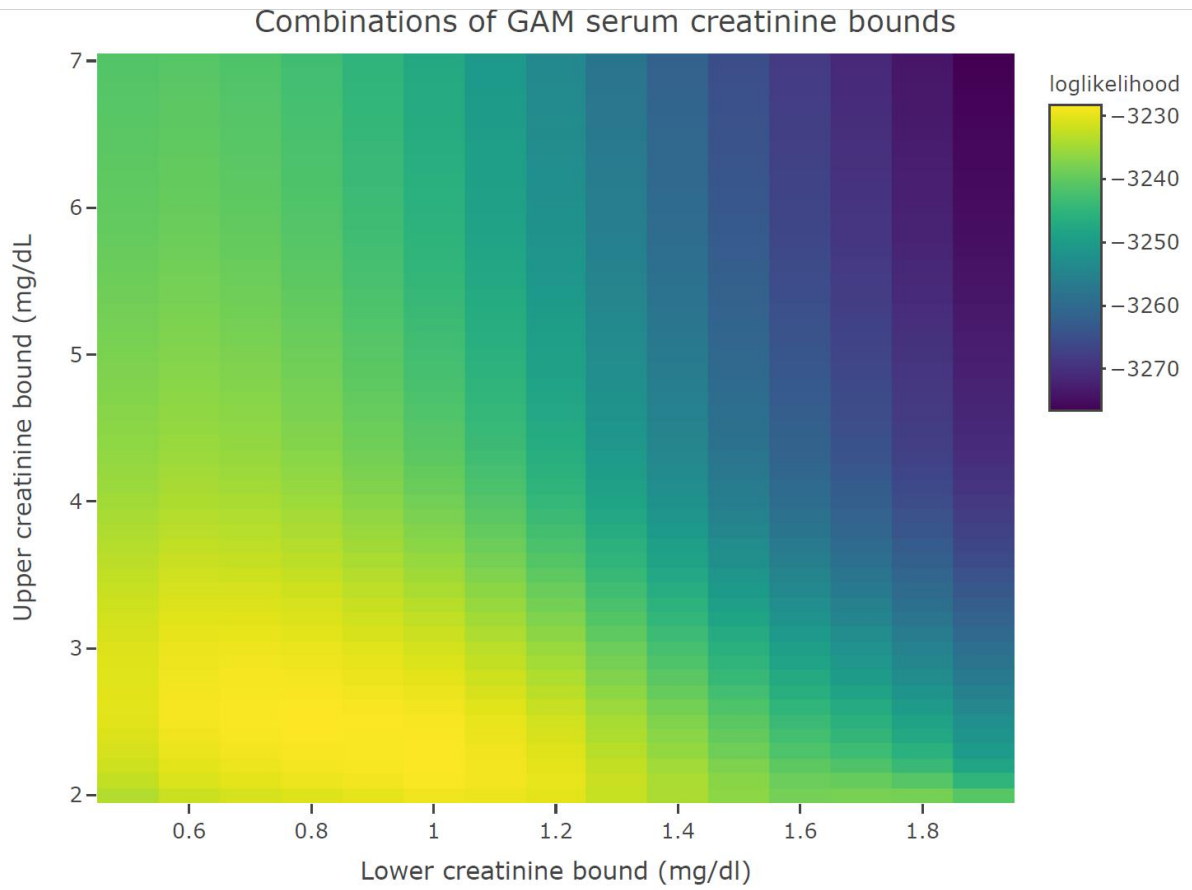
Supplement figure 2a. Heatmap of maximum log-likelihood of creatinine bounds.
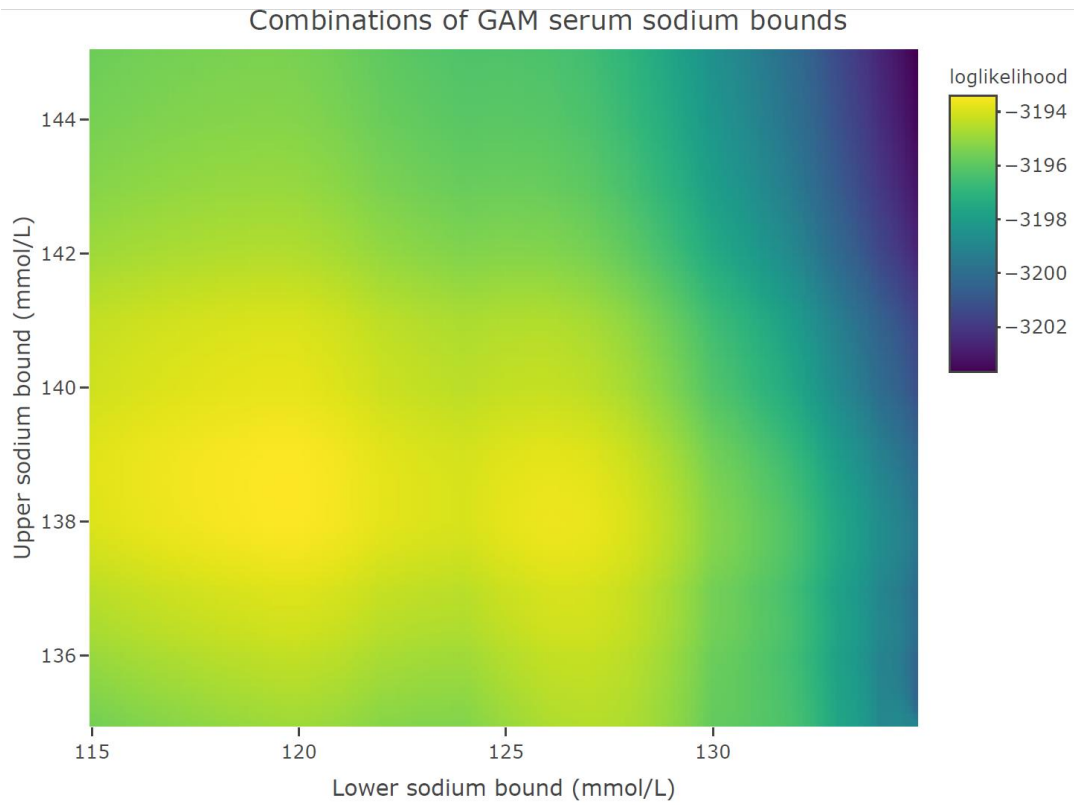

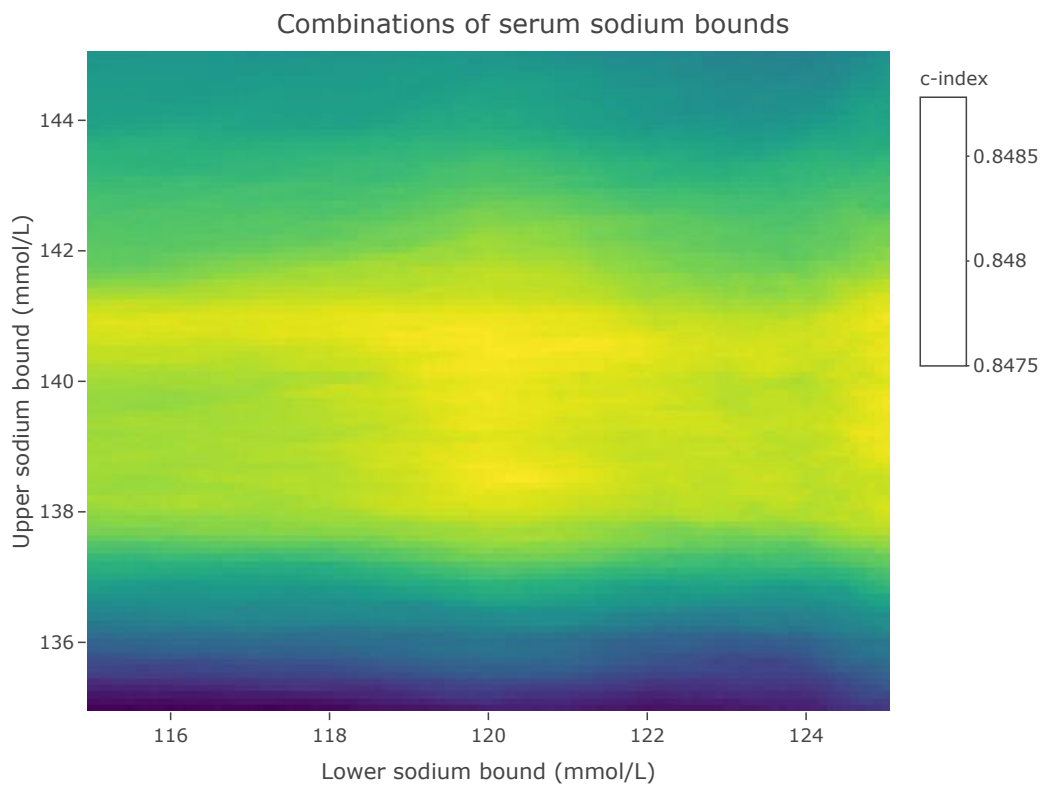Supplement figure 2b. Heatmap of maximum c-index of creatinine bounds.

Supplement figure 3a. Heatmap of maximum log-likelihood of INR bounds.

Supplement figure 3b. Heatmap of maximum c-index of INR bounds.

Supplement figure 4a. **Heatmap of maximum log-likelihood of sodium bounds.**

Supplement figure 4b. **Heatmap of maximum c-index of sodium bounds.**

Supplement 3: table showing the MELD mortality equivalents on which the ET exception point system is currently based.

MELD score	3-mo mortality equivalent
20	10%
22	15%
24	20%
25	25%
26	30%
28	35%
29	40%
29	45%
30	50%
31	55%
32	60%
33	65%
33	70%
34	75%
35	80%
36	85%
37	90%
39	95%
40	100%