**Supplement A**

**Continuous prediction of mortality in the PICU: a recurrent neural network model in a single center dataset**

**M. Aczon, D. Ledbetter, E. Laksana, L. Ho, R. Wetzel**

**Data Pre-processing**

*Observation Aggregation and Normalization.* Values from different measurement methods of the same variable were aggregated into a single feature when deemed medically appropriate. For example, non-invasive and invasive measurements of systolic blood pressures were grouped together into a single systolic blood pressure variable (1). The time of and the source from which cultures were ordered were used as inputs. Results unknown at the time each prediction was made were not used. If a therapy (drug or intervention) was administered to less than 1% of patient episodes in the training set, then it was excluded from the list of model inputs. The variable aggregation and exclusion process resulted in 430 different demographic, physiologic, and therapy variables used as model inputs (for the list, see Tables S-1 [http://links.lww.com/PCC/B672], Table S-2 [http://links.lww.com/PCC/B673], Table S-3 [http://links.lww.com/PCC/B674], and Table S-4 [http://links.lww.com/PCC/B675]; acronyms can be found in Table S-5, http://links.lww.com/PCC/B676). Minimum and maximum values were used to filter measurements that were considered incompatible with life (for example, heart rates exceeding 400 beats per minute). Physiologic variables (vitals and laboratory measurements) were transformed to zero-mean and unit-variance features (z-normalization), where the means and standard deviations used for this transformation were derived from the training set. Therapies administered to the patient were scaled to the interval [0,1] using clinically defined upper limits. The z-normalization and [0,1]-scaling were done for computational stability. Binary indicators of continuous therapies (1 - present, 0 - absent) are maintained from the time an intervention begins to the time of its discontinuation. Since age was one of the inputs, no variables were normalized by age. Diagnoses were used for descriptive analyses but not as RNN input features. PELOD scores from the first 10 days, PIM2 scores, and PRISM3 (12-hour variant) scores, all collected for VPS (2), were available for most episodes but were used only for analyses; they were not used as model inputs.

*Imputation.* EMR measurements are sparsely, asynchronously, and irregularly charted, where time between measurements ranged from a minute to several hours. At any time when at least one variable had a recorded value, the missing values for other variables were imputed. The imputation process depended on the variable. Any missing measurement of a drug or an intervention variable indicates actual absence of treatment; therefore, it was set to zero. When a physiologic observation or lab measurement was available, it was propagated forward until another measurement was recorded. This choice was based on an assumption about clinical practice: measurements are recorded more frequently when the patient is unstable, and less frequently when the patient appears stable (3)[.](#_heading=h.4d34og8) If a physiologic or laboratory variable had no recorded value throughout the entire episode, then the training set population mean was used as the value for that variable at all time points in that episode. Table S-1 (http://links.lww.com/PCC/B672) and Table S-2 (http://links.lww.com/PCC/B673) show for each demographic, vital sign and lab variable the percentage of patients with measurements and the fraction of timesteps with imputed values. On average, there were two distinct timesteps per hour with a measurement. Depending on the vital sign, charting was done every 48-63 minutes on average over the cohort. Overall, 48% of the data were imputed.

**REFERENCES**

1. Imholz, BP, Settels JJ, van der Meiracker AH, et al: Non-invasive continuous finger blood pressure measurement during orthostatic stress compared to intra-arterial pressure. *Cardiovasc Res.* 1990; 24(3):214-221
2. VPS Data Collection and Definitions Manual - VPS 7. Web Version 7.2, 2016.
3. Schulman CS, Staul L: Standards for frequency of measurement and documentation of vital signs and physical assessments. *Crit Care Nurse.* 2010; 30(3):74-76

**SUPPLEMENTAL FIGURE LEGENDS**

**Figure S-1.** RNN AUROC, as a function of time relative to ICU admission (left) or ICU discharge (right), of two test set subcohorts: episodes whose length of stay (LOS) was between 4 and 8 days, and whose LOS was greater than or equal to 10 days.

**Figure S-2.** Receiver Operator Characteristic Curve (left) and Precision-Recall Curve (right) of RNN (12th hour predictions), PELOD Day 1, PIM2 and PRISM3 (12th hour version).