Learning of Cluster-based Feature Importance for Electronic Health Record Time-series

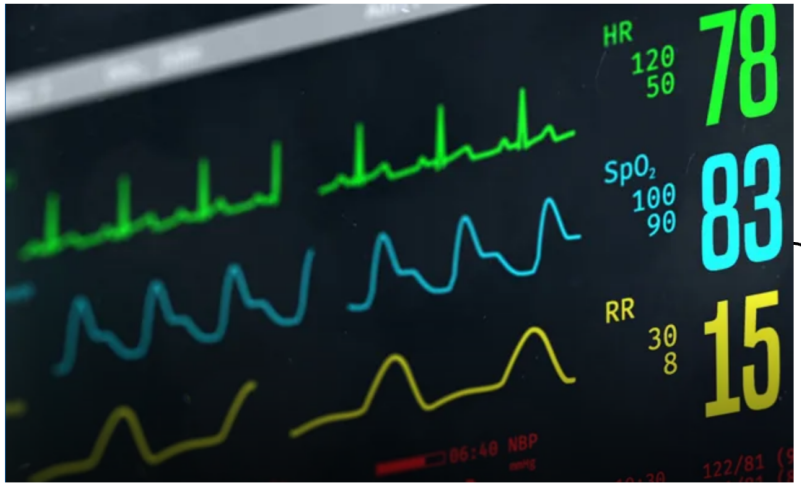
H. Aguiar [1] M. Santos [1] P. Watkinson [2] T. Zhu [1]

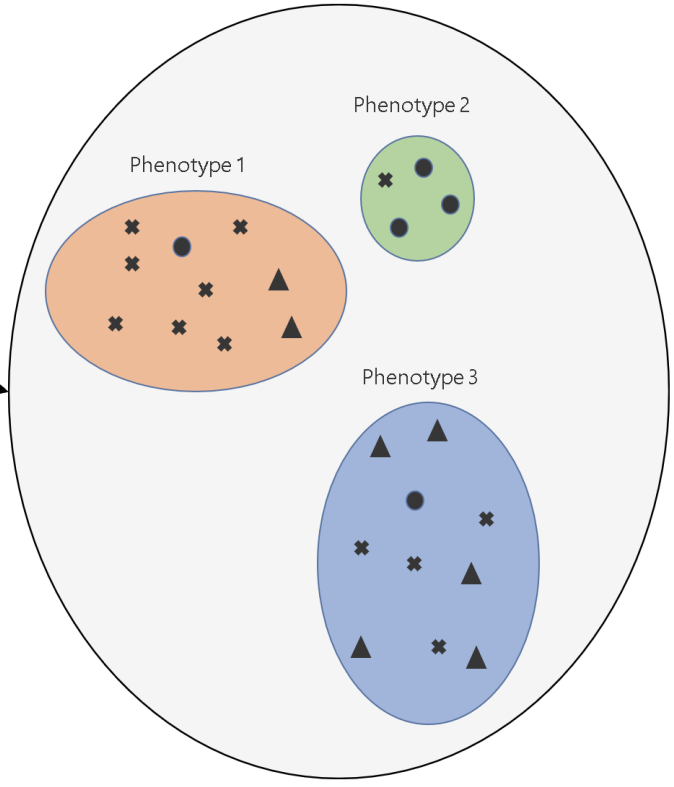
[1]Department of Engineering Science, University of Oxford

[2]Nuffield Department of Clinical Neurosciences, University of Oxford

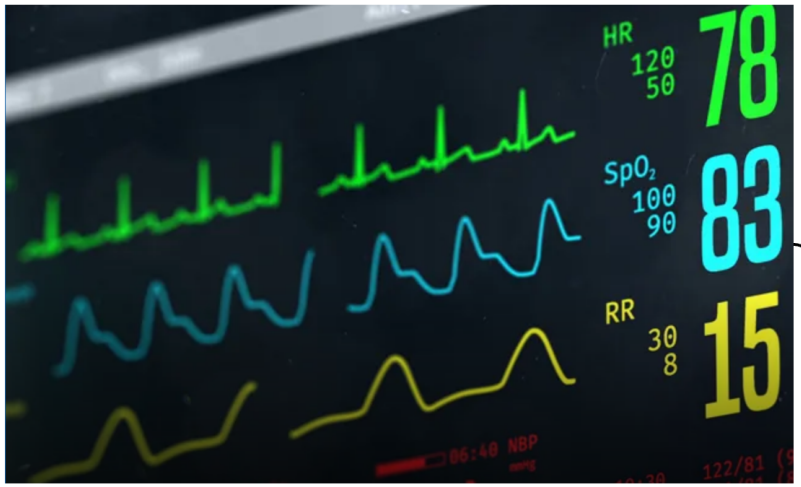
July 18, 2022

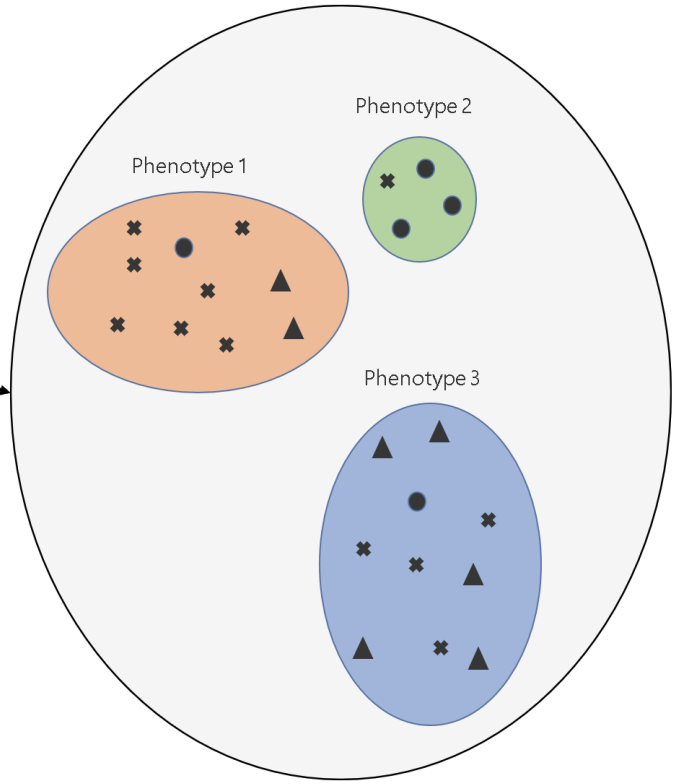
Introduction

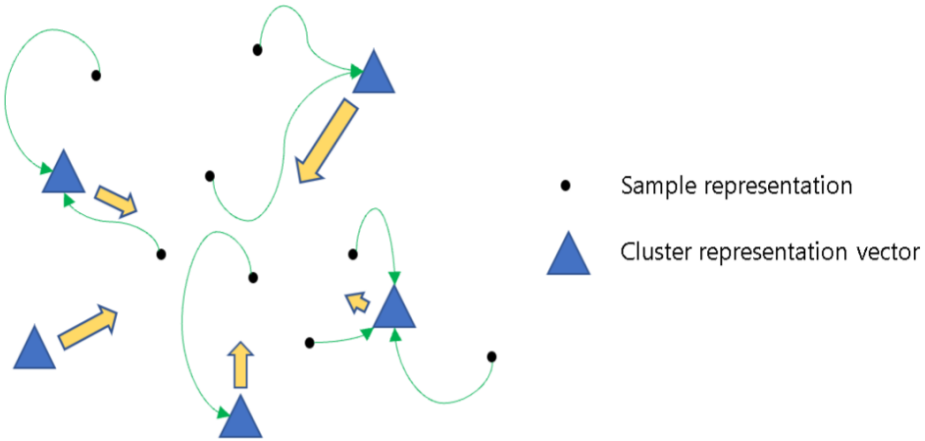




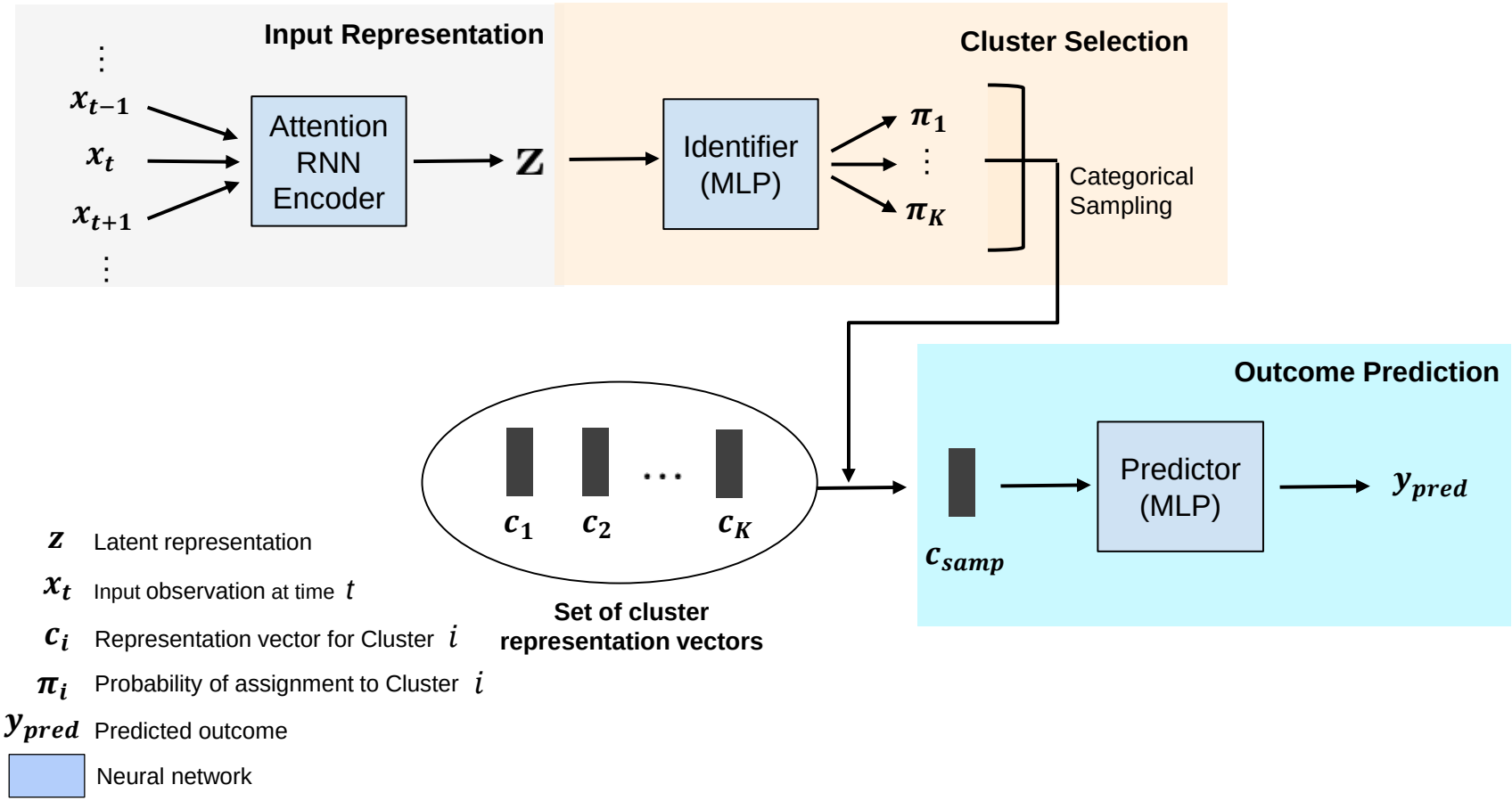
Introduction



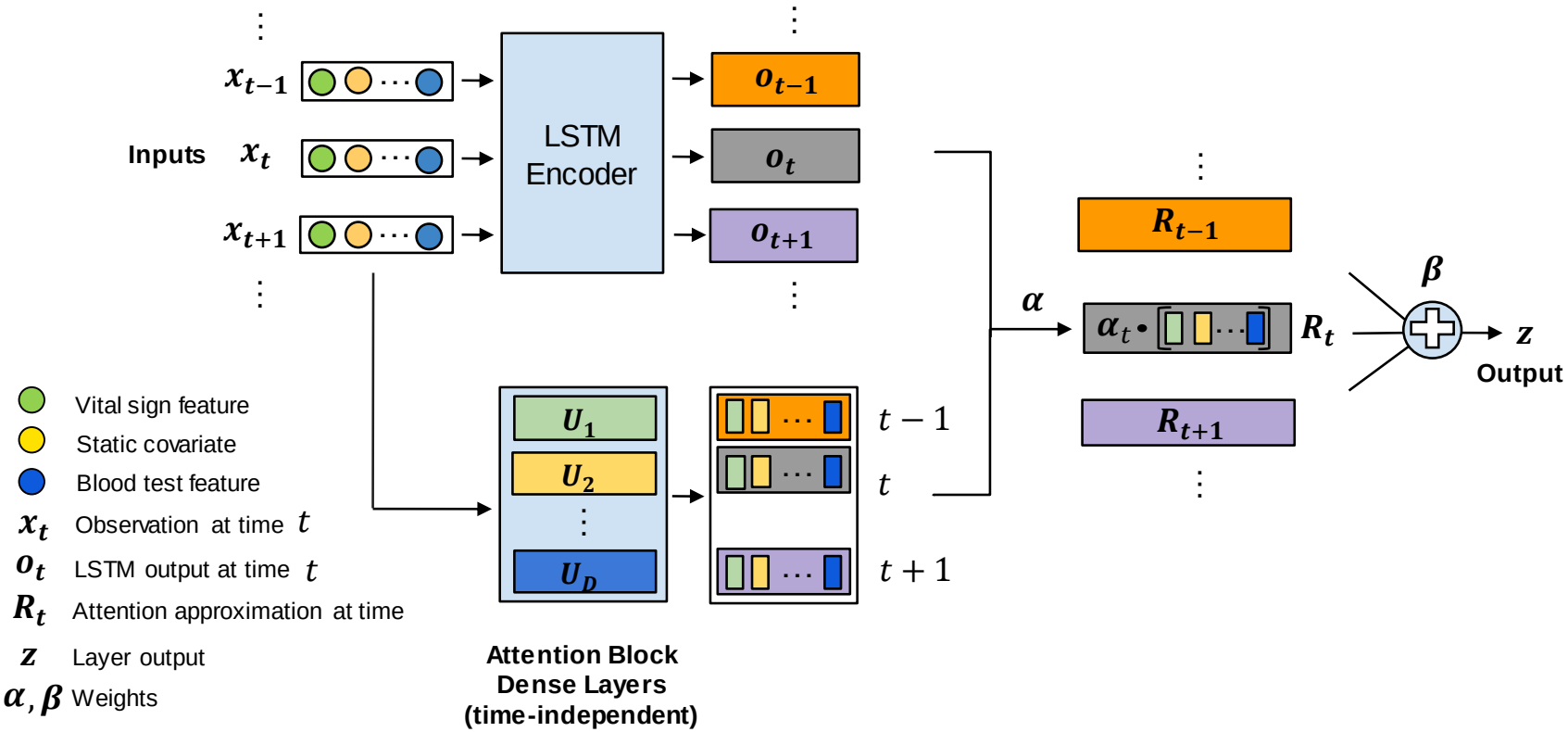




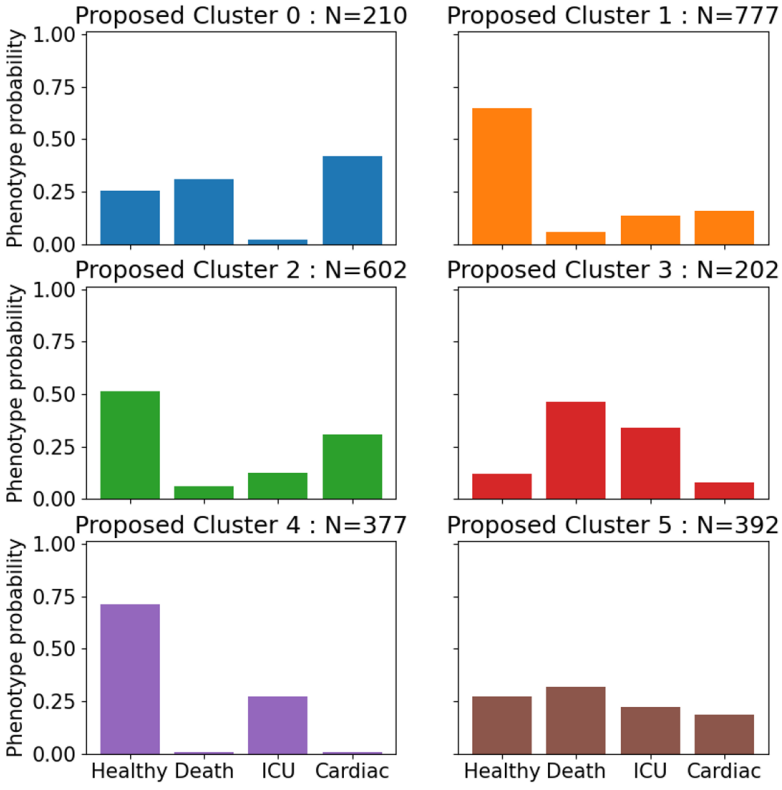
CAMELOT


Input Representation
x_{t-1}
x_t
x_{t+1}
Attention RNN Encoder
z
Cluster Selection
Identifier (MLP)
π_1
π_K
Categorical Sampling
Outcome Prediction
c_1 c_2 ... c_K
Set of cluster representation vectors
c_{samp}
Predictor (MLP)
y_{pred}
z Latent representation
x_t Input observation at time t
c_i Representation vector for Cluster i
π_i Probability of assignment to Cluster i
y_{pred} Predicted outcome
Neural network

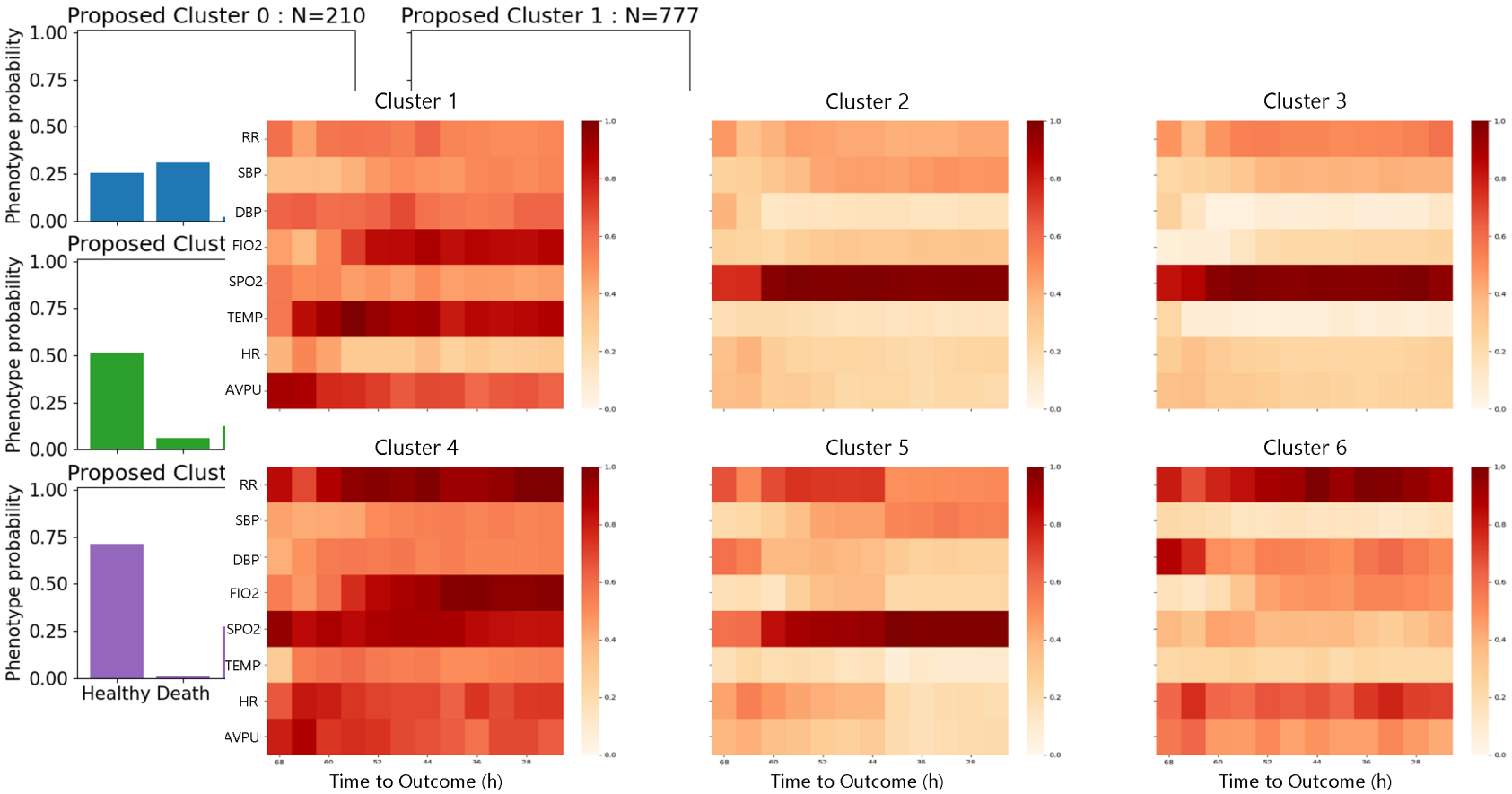
Deeper Look into Encoder Block



Results


Proposed Cluster 0 : N=210
Proposed Cluster 1 : N=777
Proposed Cluster 2 : N=602
Proposed Cluster 3 : N=202
Proposed Cluster 4 : N=377
Proposed Cluster 5 : N=392
Phenotype probability
1.00
0.75
0.50
0.25
0.00
Healthy
Death
ICU
Cardiac

Results


Proposed Cluster 0 : N=210
Proposed Cluster 1 : N=777
Phenotype probability
1.00
0.75
0.50
0.25
0.00
Proposed Clust
Proposed Clust
Healthy Death
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
RR
SBP
DBP
FIO2
SPO2
TEMP
HR
AVPU
68
60
52
44
36
28
Time to Outcome (h)