



Figure 1: The relationship between the NEWS2 and MAP indicators with cost values.

Table 1: Comparison across policies on the sepsis test set using NEWS2 reward. The best algorithms are highlighted in red.  $RMSE_{IV}$  and  $RMSE_{VASO}$  mean the RMSE loss for the IV fluid treatment and vasopressor treatment. P.F1 and S.F1 denote the patient-wise F1 and sample-wise F1.

Metric	DQN	CQL	IQL	BCQ	CDT+CT (without attention)	CDT+CT
$RMSE_{IV}$	$638.51 \pm 8.63$	$541.67 \pm 5.74$	$578.96 \pm 10.06$	$626.2 \pm 9.56$	$435.89 \pm 9.60$	<b><math>433.55 \pm 7.20</math></b>
$RMSE_{VASO}$	$0.44 \pm 0.07$	<b><math>0.30 \pm 0.01</math></b>	$0.31 \pm 0.01$	0.31	1.14	$1.13 \pm 0.01$
WIS	$-3.79 \pm 0.01$	$-4.10 \pm 1.43$	-5.83	-4.58	$-5.38 \pm 2.13$	<b><math>-3.51 \pm 0.11</math></b>
$WIS_b$	$-3.88 \pm 0.73$	$-4.48 \pm 0.77$	-5.310.06	$-5.41 \pm 0.17$	$-5.37 \pm 1.73$	<b><math>-3.52 \pm 0.17</math></b>
$WIS_t$	$-3.84 \pm 0.11$	$-4.10 \pm 1.43$	-5.83	-4.58	$-5.75 \pm 2.13$	<b><math>-3.51 \pm 0.11</math></b>
$WIS_{bt}$	$-3.87 \pm 0.67$	$-4.38 \pm 0.98$	$-5.27 \pm 0.05$	$-5.55 \pm 0.19$	$-5.27 \pm 1.72$	<b><math>-3.52 \pm 0.17</math></b>
DR	<b><math>-0.14 \pm 0.04</math></b>	$-0.71 \pm 0.05$	$-0.51 \pm 0.04$	$-1.54 \pm 0.01$	$-3.40 \pm 0.36$	-3.08
P.F1	$0.06 \pm 0.02$	$0.33 \pm 0.01$	<b><math>0.34 \pm 0.01</math></b>	$0.23 \pm 0.01$	$0.19 \pm 0.04$	$0.17 \pm 0.02$
S.F1	$0.06 \pm 0.02$	$0.32 \pm 0.01$	<b><math>0.33 \pm 0.01</math></b>	$0.22 \pm 0.01$	$0.18 \pm 0.04$	$0.16 \pm 0.02$

Table 2: The impact of generative world models with different target rewards on policy estimation.

Target Reward	IV DIFF	VASO DIFF	ACTION DIFF
1	$51.60 \pm 1.78$	$58.8 \pm 2.74$	$54.25 \pm 1.79$
5	$52.50 \pm 1.46$	$58.84 \pm 3.24$	$54.45 \pm 1.65$
10	$52.25 \pm 1.33$	$56.85 \pm 4.20$	$55.00 \pm 1.80$
40	$52.05 \pm 1.30$	$56.75 \pm 3.13$	$55.80 \pm 1.76$
50	$52.00 \pm 1.31$	$57.35 \pm 2.09$	$55.05 \pm 1.93$