Synthetic Data Simulation Results

2023-12-04

Simulation Summary

LTMLE generated quite a few errors, reducing results given available time. Below are counts of completed runs and runs with errors for each simulation setting.

data	estimator	completed_1000	completed_10000	completed_1e+05	error_1000	error_10000	error_1e+05
MC	LTMLE	132	166	173	233	229	226
MC	Marginal	20	42	64	345	353	335
Synthetic	LTMLE	200	420	22	218	420	30
Synthetic	Marginal	418	840	52	0	0	0

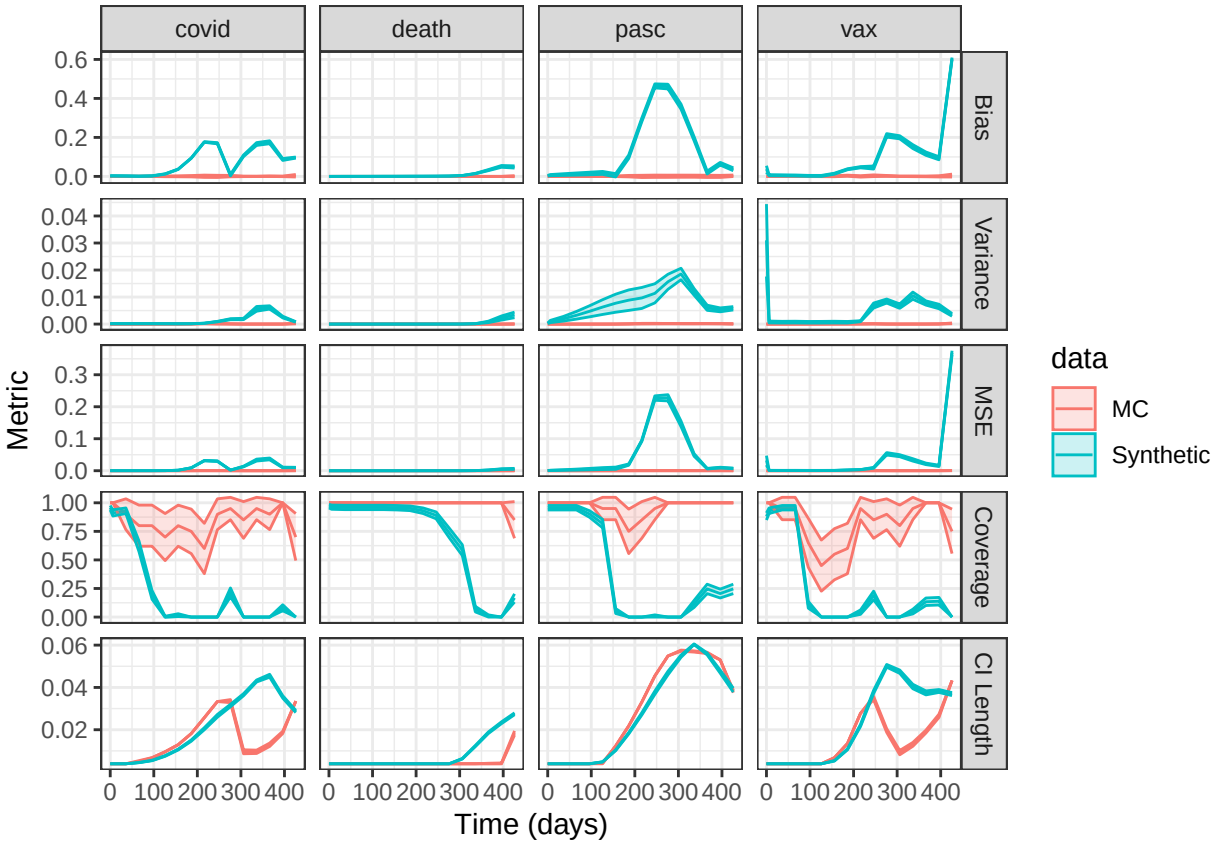
LTMLE was very time consuming, especially at larger sample sizes:

data	estimator	1000	10000	1e+05
MC	LTMLE	00:09:15	01:37:60	23:55:41
MC	Marginal	00:00:04	00:00:27	00:04:49
Synthetic	LTMLE	00:13:14	01:56:07	22:47:25
Synthetic	Marginal	00:00:02	00:00:06	00:00:53

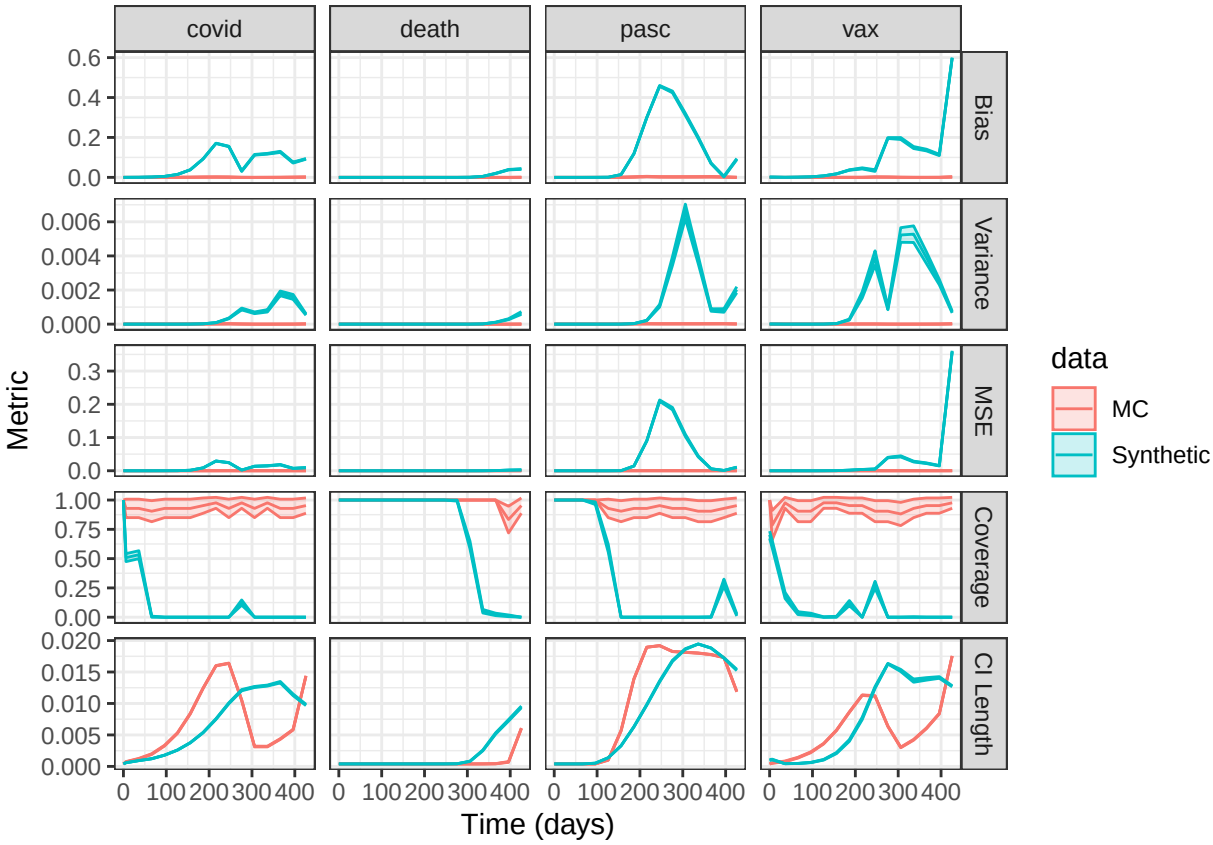
Point Estimate Performance

Point estimate performance was degraded, improving somewhat at larger sample sizes. This suggests a need for estimation with more data adaptive models.

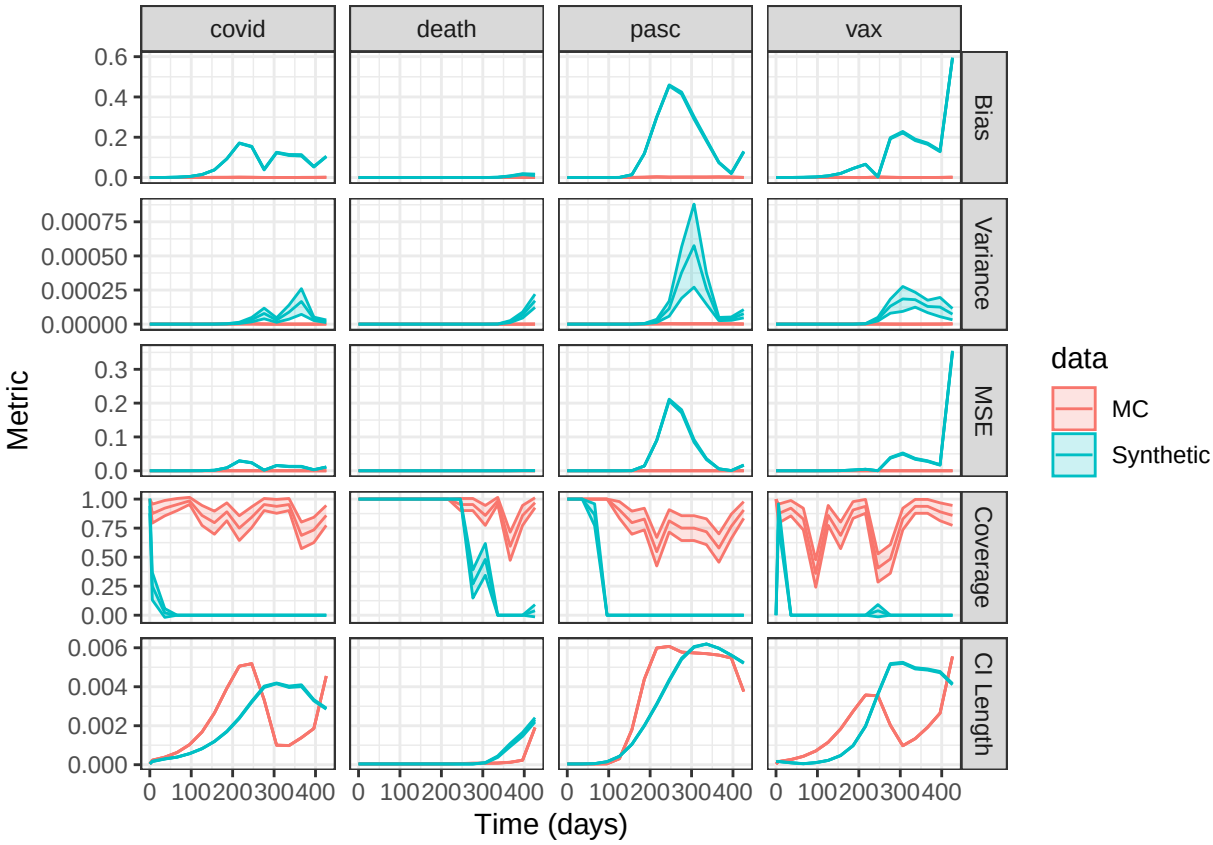
n = 1000



$\mathbf{n = 10000}$

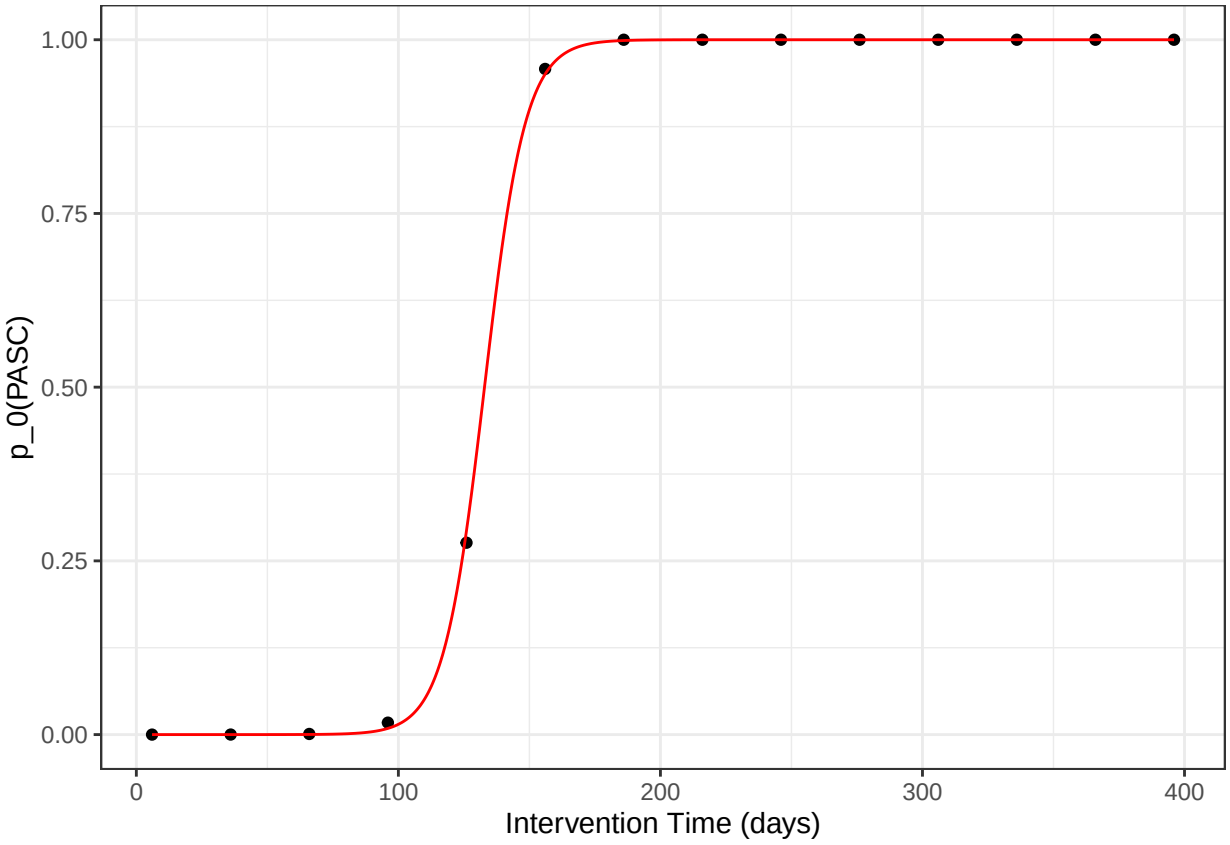


$\mathbf{n = 100000}$



Causal Estimate Performance

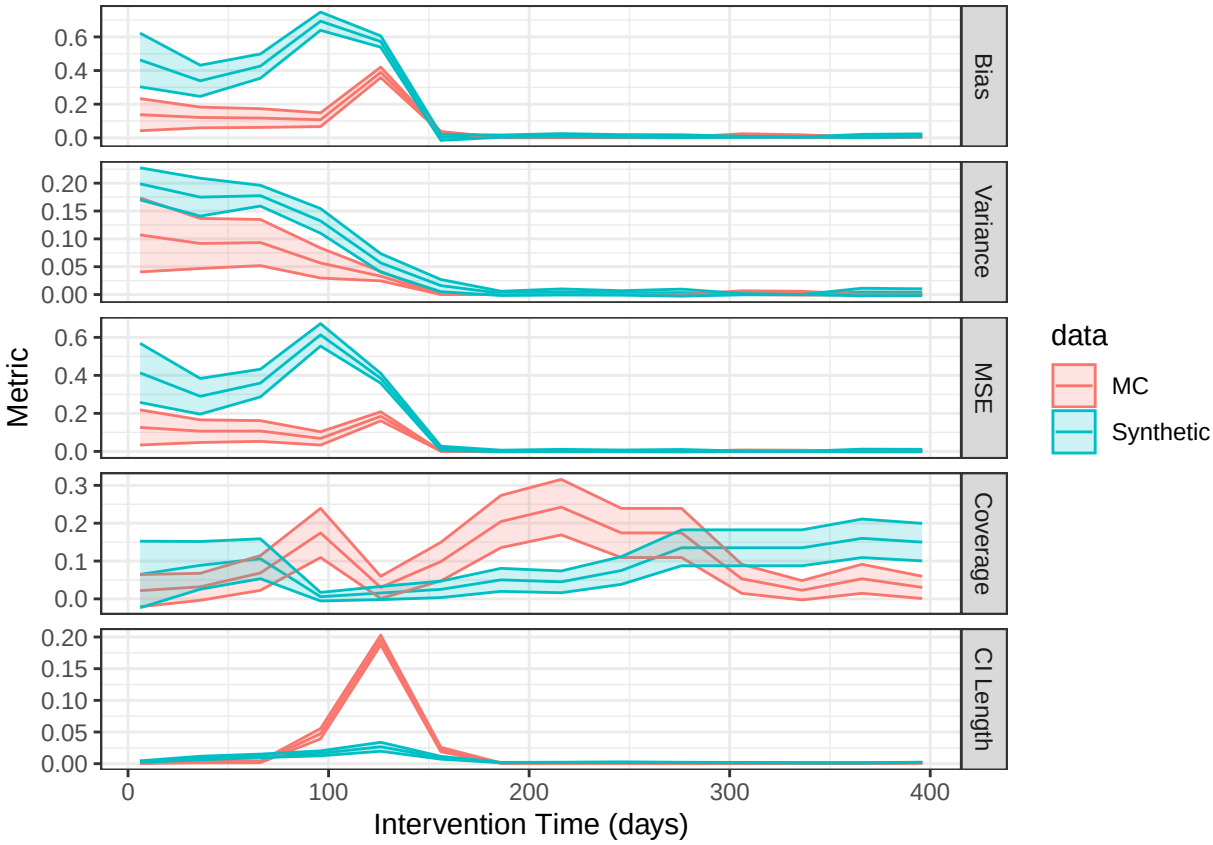
The true data generating process includes the probability of PASC as a logistic function of intervention time, meaning that the estimated marginal structural model (MSM) is a true model.



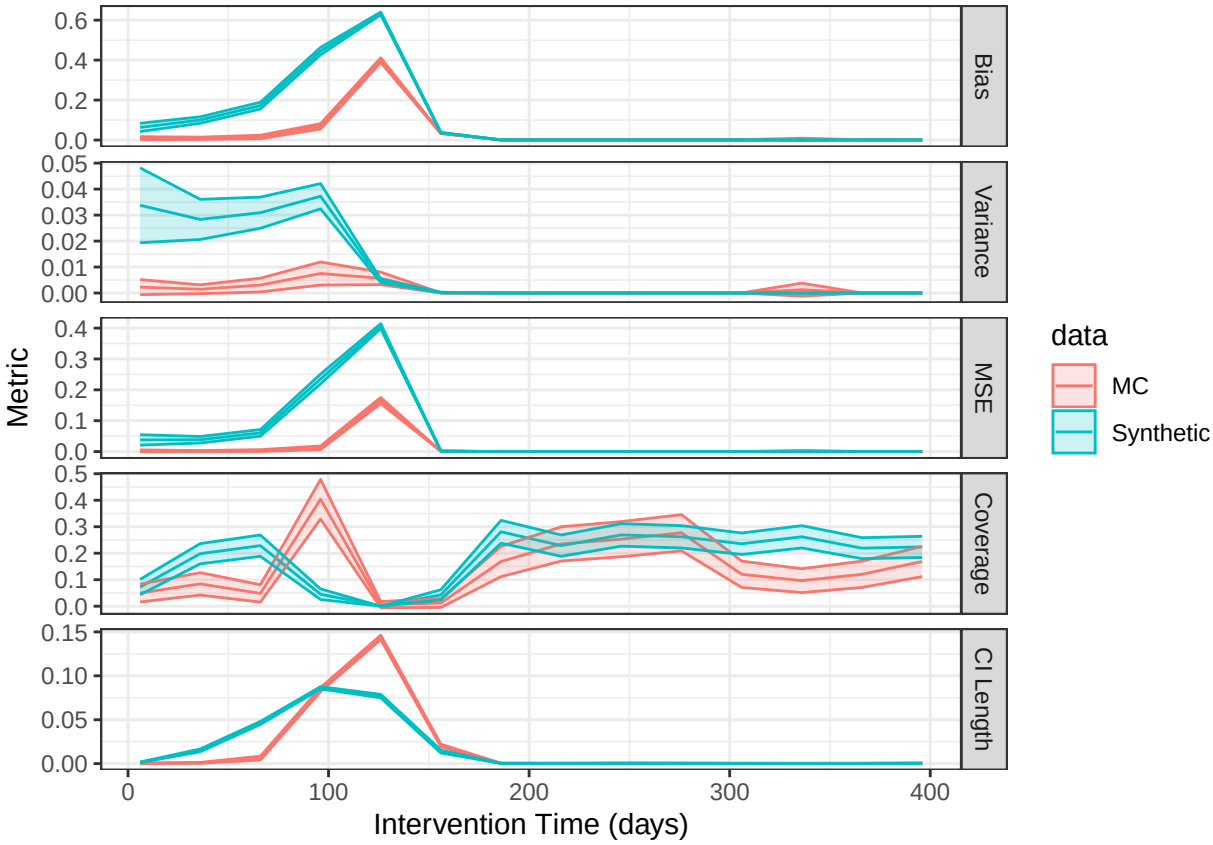
Treatment Specific Mean

It is unclear why all estimates are biased at a particular timepoint, perhaps it's an issue with targetting the MSM and not the individual TSMs. Performance is again degraded for the Synthetic data.

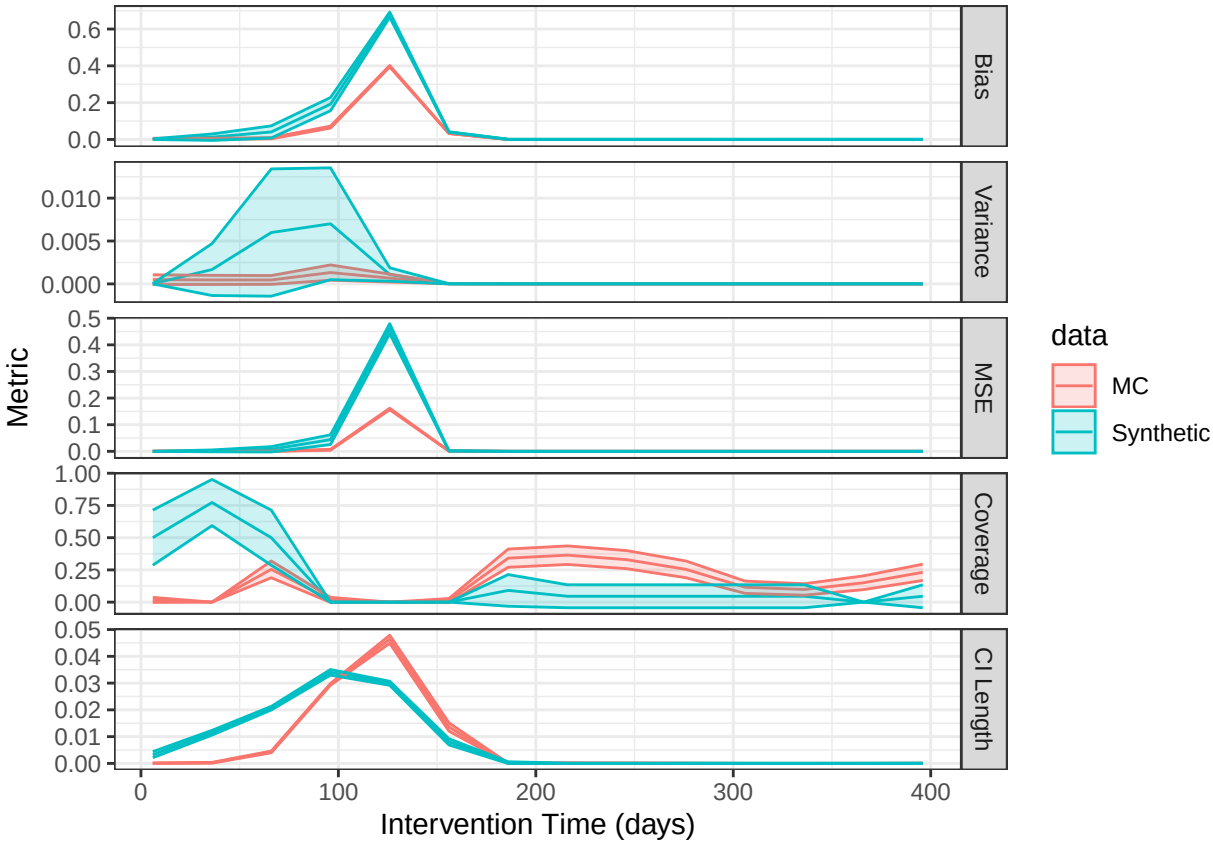
n = 1000



n = 10000


Bias
Variance
MSE
Coverage
CI Length
Metric
Intervention Time (days)
data
MC
Synthetic

n = 100000


Bias
Variance
MSE
Coverage
CI Length
Metric
Intervention Time (days)
data
MC
Synthetic

Marginal Structural Model

Performance is again degraded for the Synthetic data, although not as badly.