

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	learners	completed_1000	completed_10000	completed_1e+05	error_1000	error_10000	error_1e+05
MC	DGP	NA	0	0	0	4	0	0
MC	DGP	glm_true	45	21	11	0	0	0
MC	DGP	glmnet	18	0	0	0	0	0
MC	Marginal	MC	49	21	11	0	0	0
Synthetic	Marginal	glm_true	928	420	220	0	0	0
Synthetic	Marginal	glmnet	409	0	0	0	0	0

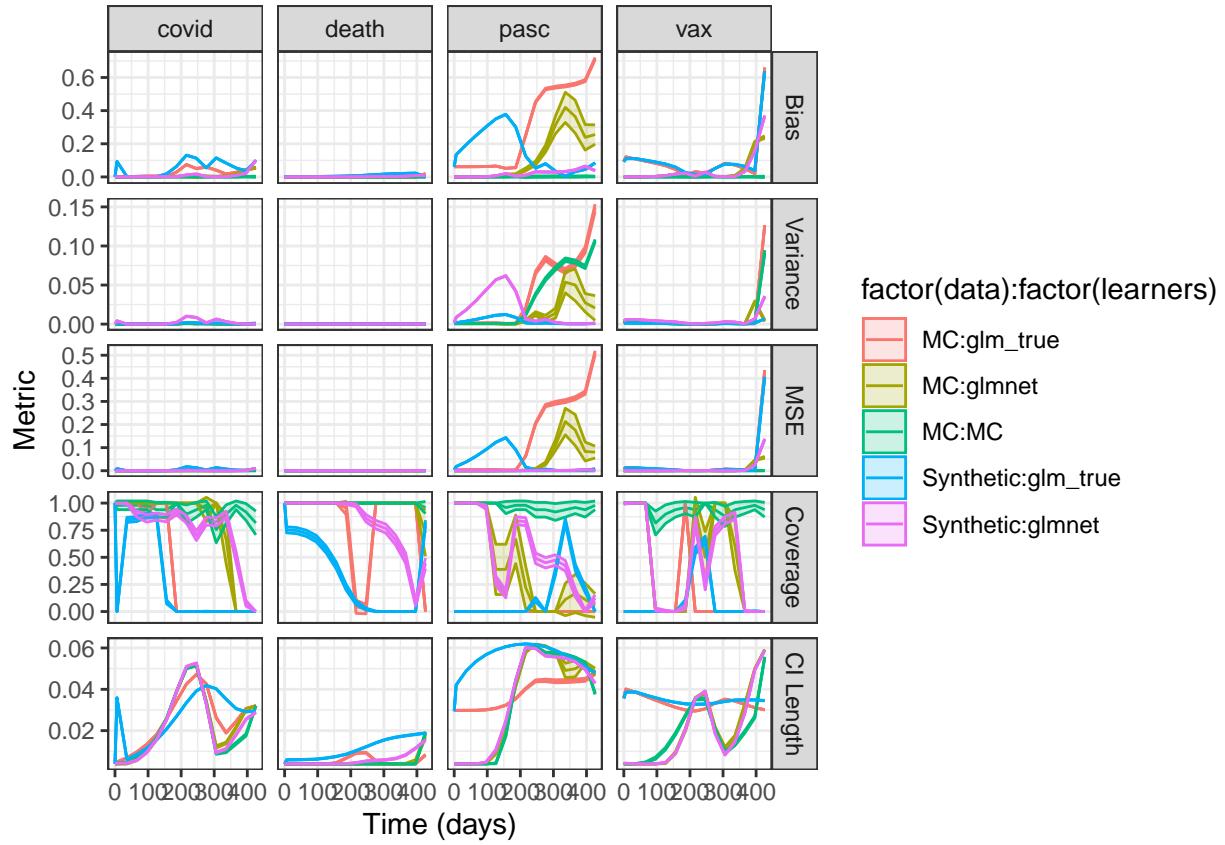
ignore the error counts here (they're for sims we don't expect)

data	estimator	learners	1000	10000	1e+05
MC	DGP	glm_true	00:00:32	00:02:09	00:18:55
MC	DGP	glmnet	00:04:09	NA	NA
MC	Marginal	MC	00:00:03	00:00:16	00:02:13
Synthetic	Marginal	glm_true	00:00:01	00:00:02	00:00:14
Synthetic	Marginal	glmnet	00:00:02	NA	NA

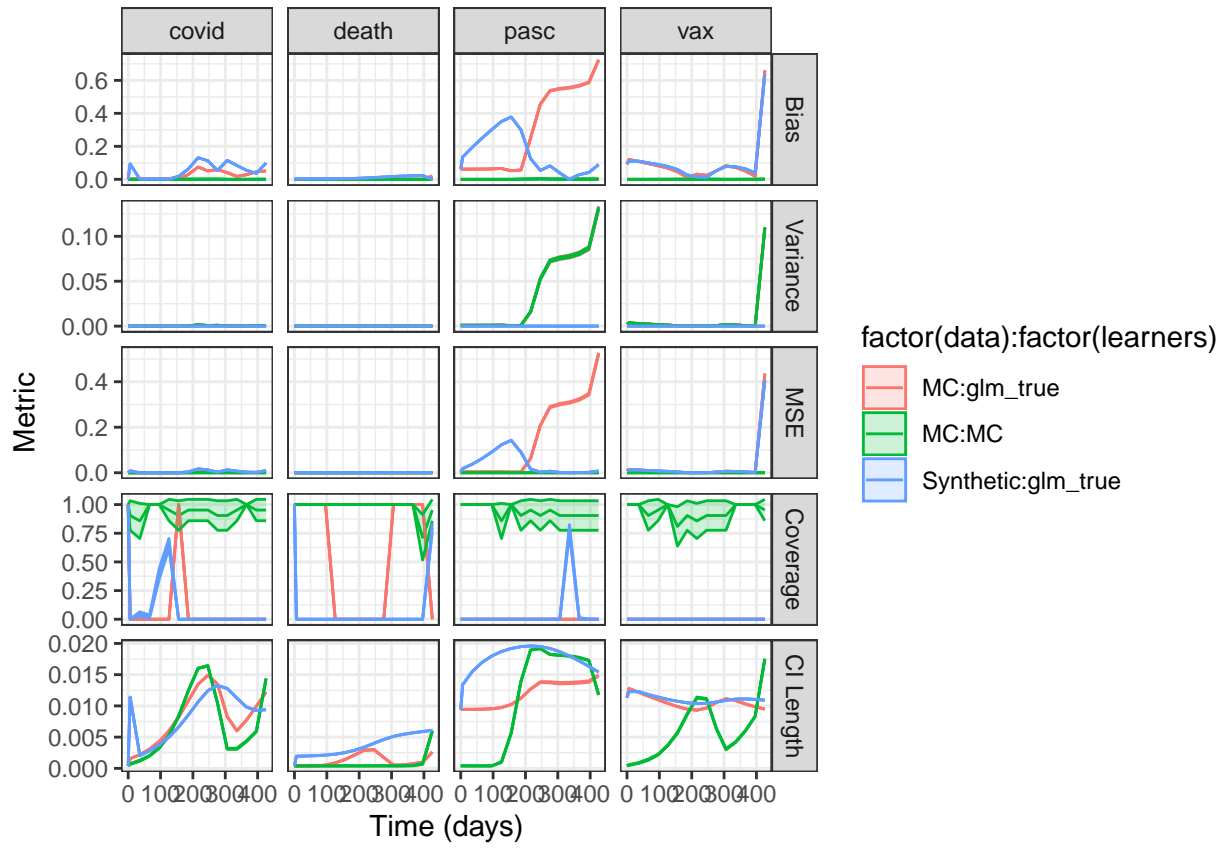
Point Estimate Performance

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

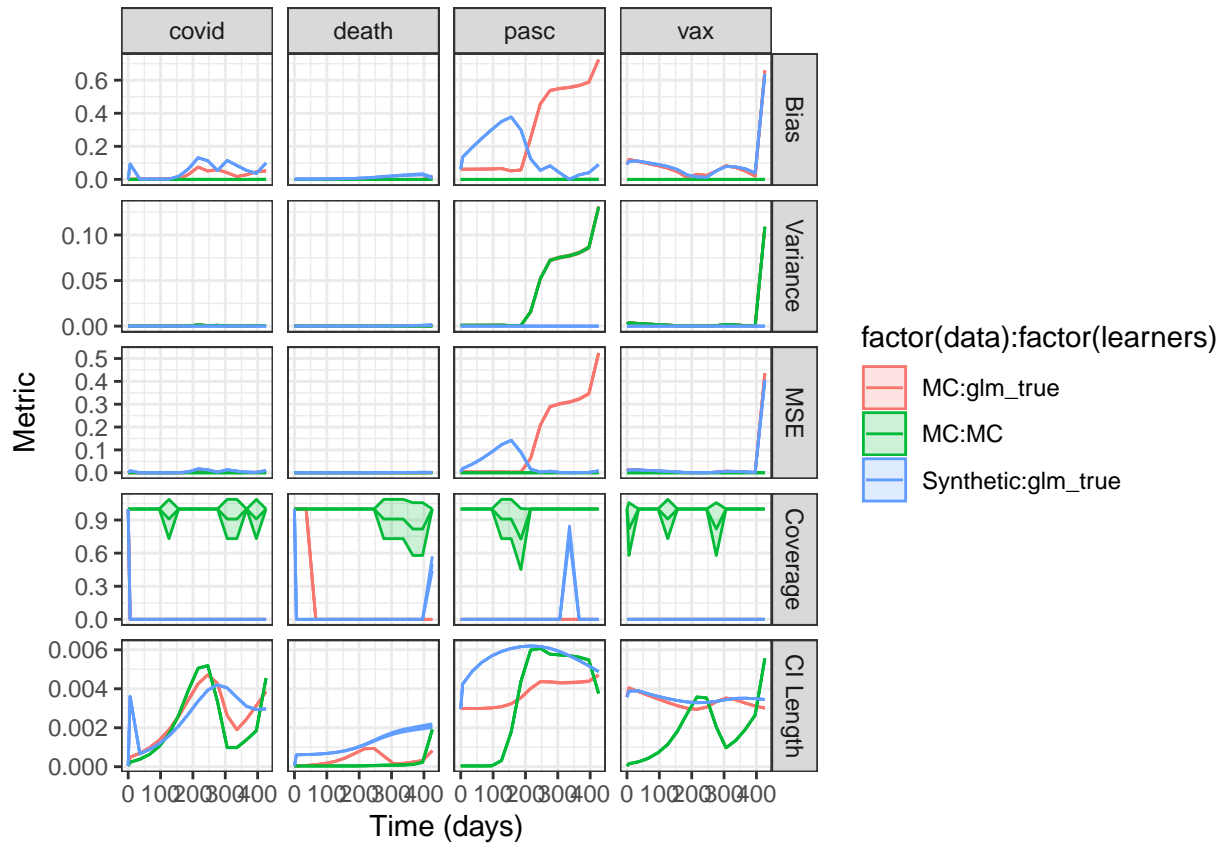
$n = 1000$



n = 10000

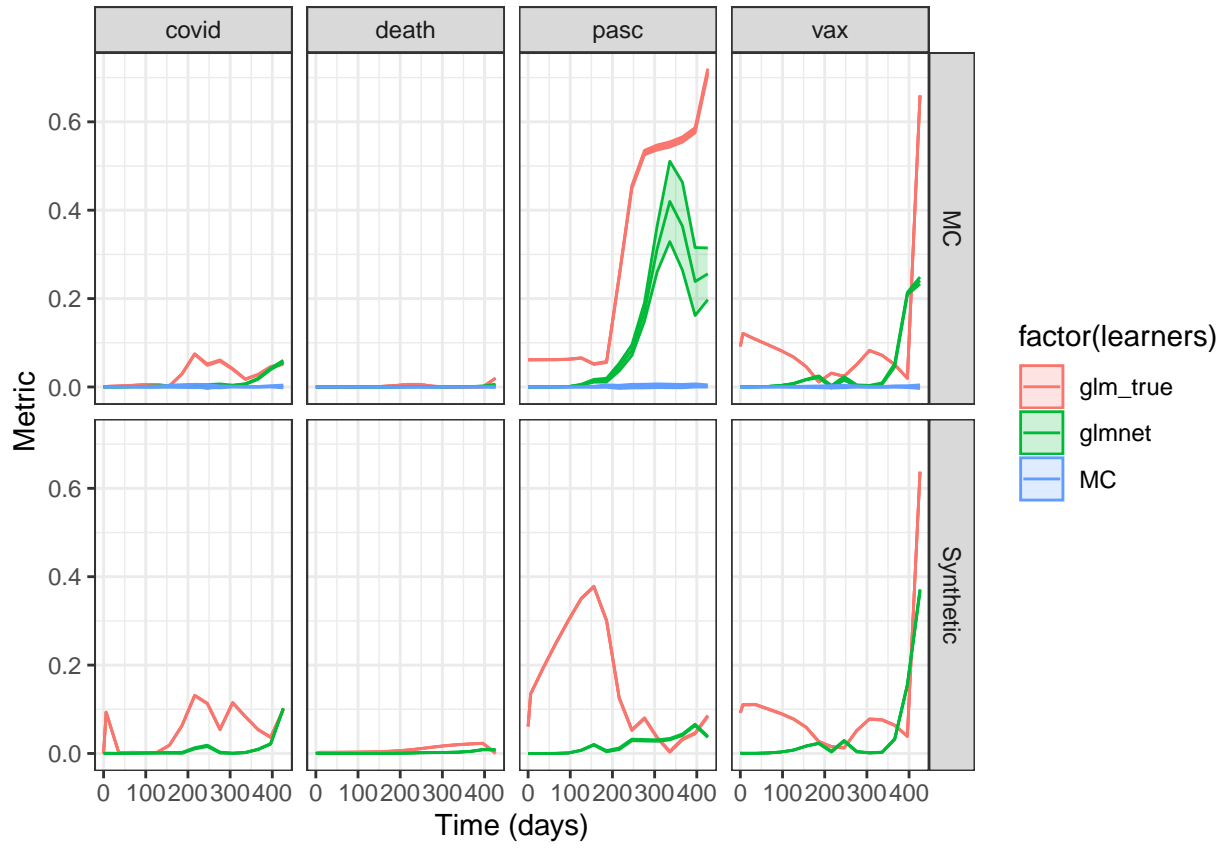


n = 100000

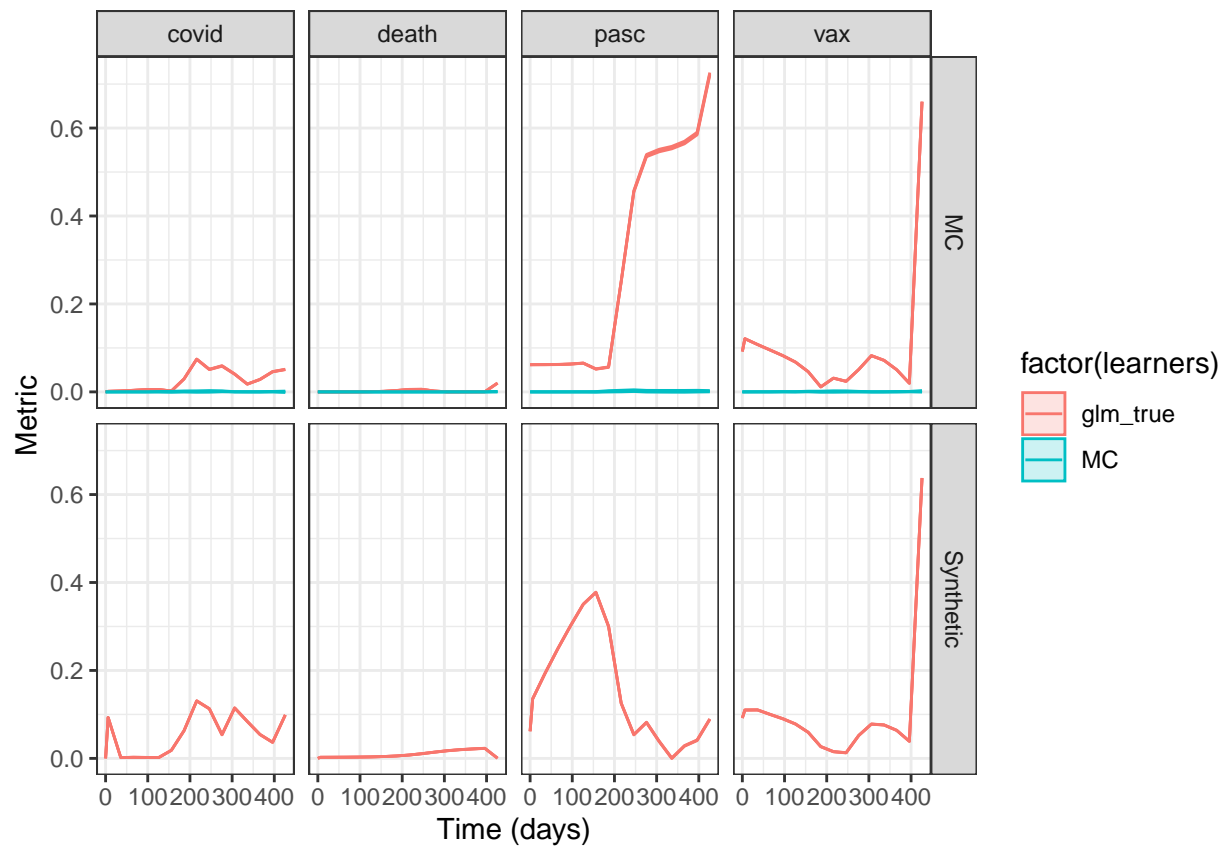


Bias

n = 1000



n = 10000



n = 100000

