

# Diabetes-dementia updated simulation results

Andrew Mertens

2022-11-17

## Summary

A simulation study was performed to identify the best estimator performance for a longitudinal TMLE analysis of the effect of second-line diabetes drugs on dementia risk in the Danish National Registry data with four key features: many timepoints (10), a rare outcome (dementia prevalence: 1.9%), competing risks from death, and a high degree of administrative censoring. Three simulations were completed. 1) A simple simulation without positivity violations, rare outcomes, long-term followup, or competing risks as a sanity check that estimators were implemented correctly, especially as we modified the LTMLE package code, 2) a realistic simulation in terms of dementia prevalence and diabetes drug patterns, but with scrambled outcomes and competing risks to check estimator performance with a known null association, and 3) a realistic simulation with a protective effect of GLP1 usage on dementia and death, with the truth calculated as the counterfactual 5 year risk of dementia prior to death when continuously on GLP1 versus not, with the effect of GLP1 on death removed to remove the competing risk.

## Scenario 1: Realistic simulation, null outcome

True RD: 0

### Notes:

- Outcome, death and censoring all jointly scrambled
- Oracle coverage is pretty good but a little too high for all estimators, but GLM estimators have more variance

estimator	Qint	DetQ	bias	variance	mse	bias_se_ratio	oracle.coverage
LASSO	No	No	-0.00006	2e-05	2e-05	-0.01181	96.0
GLM	Yes	No	-0.00004	2e-05	2e-05	-0.00754	96.5
LASSO	Yes	No	-0.00005	2e-05	2e-05	-0.01101	96.5
GLM	No	Yes	0.00061	3e-05	3e-05	0.10875	97.0
GLM	No	No	0.00103	9e-05	9e-05	0.10823	98.0

## Performance of difference variance estimators on null data

### Notes:

- Only showing LASSO estimator results-all estimator performances assessed in the realistic simulated data below.

- Sanity-check on estimation performance on data with a known null association between GLP1 and dementia.
- The IC variance estimator is anti-conservative and the TMLE variance estimator is conservative.
- The bootstrap is anti-conservative but less so than the IC variance estimator.
- The TMLE estimator is very conservative, with CI widths 8-10X that of the bootstrap.
- The IPTW estimator is uniformly biased with overly-wide confidence intervals in all simulations (not shown).

variance_estimator	coverage	mean_ci_width
ic	51.0	0.00722
tmle	100.0	0.11535
bootstrap	91.5	0.01288

## Scenario 2: Realistic simulation, protective effect of GLP1 on dementia

True RD: -0.009683665

True RR: 0.5148661

### Comparison of different estimators' performance

#### Notes:

- Based on these results, we chose the LASSO estimator with Q-prediction and no deterministic Q function
- Several of the estimators have comparable performance, but the chosen estimator performs best in both RR and RD estimation
- Ridge regressions have lower MSE but not perfect 95% oracle coverage
- Including the deterministic Q function marginally decreases bias/variance, so we should use in the bootstrap estimator

#### Risk difference

estimator	bias	variance	mse	oracle.coverage
LASSO, Det-Q, AUC fit	-0.002080	6.0e-06	1.0e-05	84.50000
LASSO, Det-Q, AUC fit	-0.002080	6.0e-06	1.0e-05	84.50000
LASSO, Lambda: 1se	-0.001631	1.0e-05	1.3e-05	91.50000
Elastic Net, Lambda: 1se	-0.001450	9.0e-06	1.2e-05	92.00000
GLM, LASSO prescreen	0.002793	4.9e-05	5.7e-05	92.78351
LASSO, Q-intercept	-0.001583	1.1e-05	1.3e-05	93.00000
LASSO, Det-Q, Lambda: 1se	-0.001109	8.0e-06	9.0e-06	93.50000
GLM	0.002819	5.6e-05	6.4e-05	93.50000
GLM, LASSO prescreen, Det-Q	0.002795	5.1e-05	5.9e-05	93.87755
Ridge, Det-Q	0.000446	1.1e-05	1.1e-05	94.00000
Elastic Net, Det-Q, Lambda: 1se	-0.000899	8.0e-06	8.0e-06	94.50000
LASSO, Det-Q	0.000267	1.4e-05	1.4e-05	94.50000

estimator	bias	variance	mse	oracle.coverage
Ridge, Lambda: 1se	-0.000978	8.0e-06	9.0e-06	94.50000
Ridge	-0.000118	1.3e-05	1.3e-05	94.50000
LASSO, AUC fit	-0.001365	1.2e-05	1.4e-05	95.00000
LASSO	-0.000265	1.7e-05	1.7e-05	95.00000
Ridge, Det-Q, Lambda: 1se	-0.000536	6.0e-06	7.0e-06	95.50000

## Comparison of different variance estimators

### Notes:

- Showing LASSO estimator results with modeled Q (rather than intercept-only)
- The IC variance estimator is anti-conservative and the TMLE variance estimator is conservative
- The bootstrap is anti-conservative but less so than the IC variance estimator
- The IPTW estimator is uniformly biased with overly-wide confidence intervals in all simulations (not shown)

### Risk difference coverage

variance_estimator	coverage	mean_ci_width	power	bias_se_ratio_emp
ic, Det-Q	67.0	0.00736	92.0	0.14223
tmle	99.5	0.02129	49.0	-0.05020
ic	62.0	0.00737	91.0	-0.14089
Bootstrap, Det Q function	87.0	0.01346	68.5	NA
Bootstrap, Det Q function, 500 iterations	89.0	0.01338	69.5	NA
Bootstrap-Ridge	87.5	0.01289	72.0	NA
Bootstrap	85.5	0.01454	68.5	NA