0.1 Proof-of-concept simulation. Simulation scenario 1 with 3 time-points.

```r
load("/Users/olegsofrygin/GoogleDrive/Alex_SDR/sims/sims_final_04_20_17/sim1_results_t2.Rd")
print(results[["res_name"]])

## [1] "results for E[Y_d(t)] for \bar{A}=1, t=2, N = 500, nsims = 1000"
```

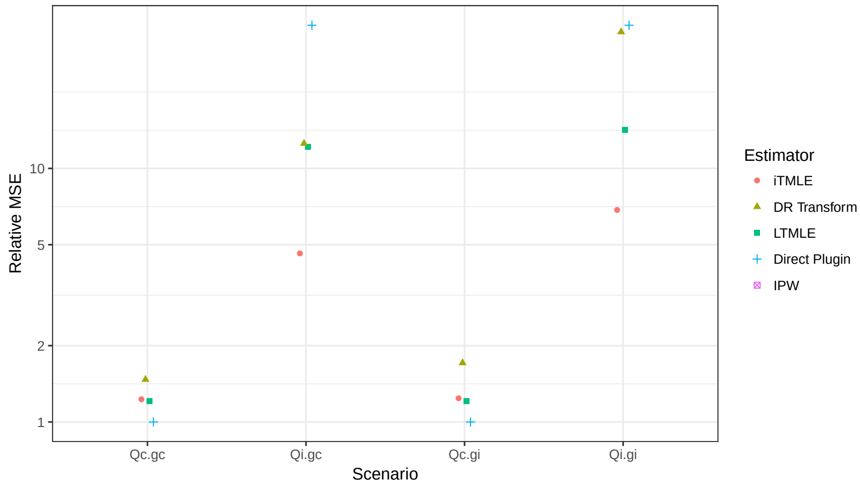

Figure 1: Relative MSE for $\hat{Q}_1$ in simulation scenario 1 with 3 time points.

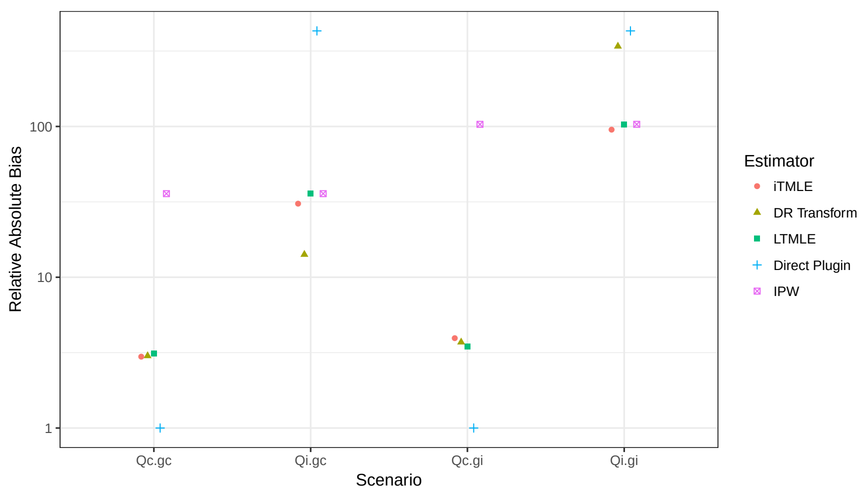

Figure 2: Relative absolute bias for $\hat{Q}_0$ in simulation scenario 1 with 3 time points.

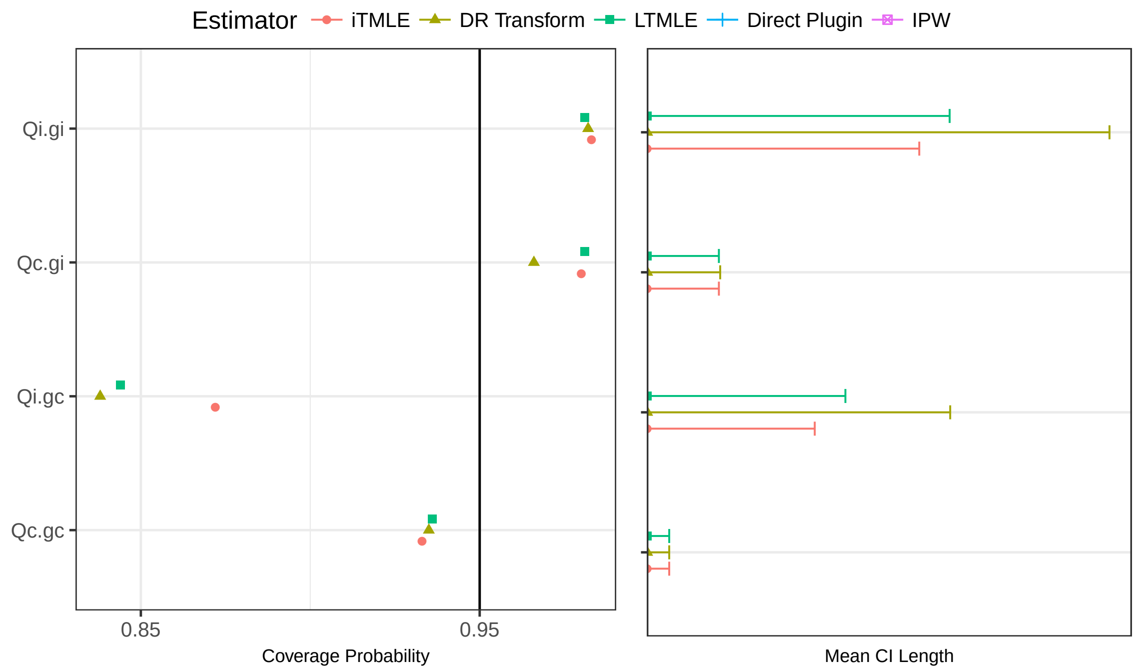

Figure 3: Coverage (left panel) and mean length (right panel) of the two-sided 95% CIs for Q_0 in simulation scenario 1.

restab	Relative Bias
Qc.gc.sdr	2.98
Qc.gc.drtrans	3.01
Qc.gc.tmle	3.11
Qc.gc.gcomp	1.00
Qc.gc.ipw	35.84
Qi.gc.sdr	30.80
Qi.gc.drtrans	14.15
Qi.gc.tmle	35.93
Qi.gc.gcomp	430.36
Qi.gc.ipw	35.84
Qc.gi.sdr	3.94
Qc.gi.drtrans	3.71
Qc.gi.tmle	3.46
Qc.gi.gcomp	1.00
Qc.gi.ipw	103.31
Qi.gi.sdr	95.58
Qi.gi.drtrans	340.11
Qi.gi.tmle	103.22
Qi.gi.gcomp	430.36
Qi.gi.ipw	103.31

Table 1: Simulation scenario 1. Relative absolute bias for estimation of Q_0, over 3 time points and n=500.

restab	Relative MSE
Qc.gc.sdr	1.23
Qc.gc.drtrans	1.47
Qc.gc.tmle	1.21
Qc.gc.gcomp	1.00
Qi.gc.sdr	4.62
Qi.gc.drtrans	12.54
Qi.gc.tmle	12.18
Qi.gc.gcomp	36.68
Qc.gi.sdr	1.24
Qc.gi.drtrans	1.71
Qc.gi.tmle	1.21
Qc.gi.gcomp	1.00
Qi.gi.sdr	6.85
Qi.gi.drtrans	34.52
Qi.gi.tmle	14.19
Qi.gi.gcomp	36.68

Table 2: Simulation scenario 1. Relative MSE for estimation of Q_1, over 3 time points and n=500.

restab	Coverage prob.
Qc.gc.sdr	0.933
Qc.gc.drtrans	0.935
Qc.gc.tmle	0.936
Qc.gc.gcomp	
Qi.gc.sdr	0.872
Qi.gc.drtrans	0.838
Qi.gc.tmle	0.844
Qi.gc.gcomp	
Qc.gi.sdr	0.980
Qc.gi.drtrans	0.966
Qc.gi.tmle	0.981
Qc.gi.gcomp	
Qi.gi.sdr	0.983
Qi.gi.drtrans	0.982
Qi.gi.tmle	0.981
Qi.gi.gcomp	

Table 3: Simulation scenario 1. Coverage of 95% CIs for Q_0, over 3 time points and n=500.

Data generating distribution for simulation scenario 1.

```
require("simcausal")
D <- DAG.empty() +
    node("L", t = 0, distr = "rnorm") +
    node("A", t = 0, distr = "rbern", prob = plogis(L[0])) +
    node("Y", t = 0, distr = "rconst", const = 0) +
    node("L", t = 1, distr = "rnorm") +
    node("A", t = 1, distr = "rbern", prob = plogis(L[1] + A[0])) +
    node("Y", t = 1, distr = "rconst", const = 0) +
    node("L", t = 2, distr = "rnorm", mean = L[0]*A[1] + A[0]*L[1] + L[1]*A[1]) +
    node("A", t = 2, distr = "rbern", prob = plogis(L[2] + A[1])) +
    node("Y", t = 2, distr = "rbern", prob = plogis(L[1]*A[2] + A[1]*L[2] + L[2]*A[2]))
```

Model specification for simulation scenario 1.

```
nsims <- 500
nsamp <- 500
tvals <- 2
## g model (same across all time-points):
gform.c <- "A ~ L + Atm1"
gform.i <- "A ~ L"
# Q models
Qforms.c <- c("Qkplus1 ~ L",
              "Qkplus1 ~ A0L1 + L0A0 + L + A + Atm1 + Ltm1",
              "Qkplus1 ~ A0L1 + L0A0 + L + A + Atm1 + Ltm1")
Qinteract.c <- list(c("A", "L"), c("A", "Ltm1"))
Qforms.i <- c("Qkplus1 ~ L", "Qkplus1 ~ Atm1", "Qkplus1 ~ Atm1")
Qinteract.i <- NULL
```

0.2 Demonstrating SDR property. Simulation scenario 2 with 5 time-points.

```
load("/Users/olegsofrygin/GoogleDrive/Alex_SDR/sims/sims_final_04_20_17/sim2_results_t4.Rd")
print(results[["res_name"]])

## [1] "results for E[Y_d(t)] for \bar{A}=1, t=4, N = 5000, nsims = 1000"
```

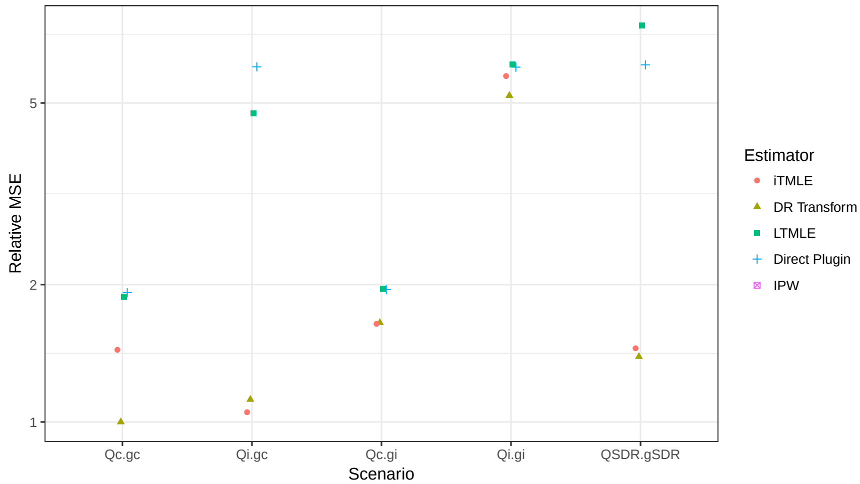

Figure 4: Relative MSE for $\hat{Q}_1$ in simulation scenario 2 with 5 time points.

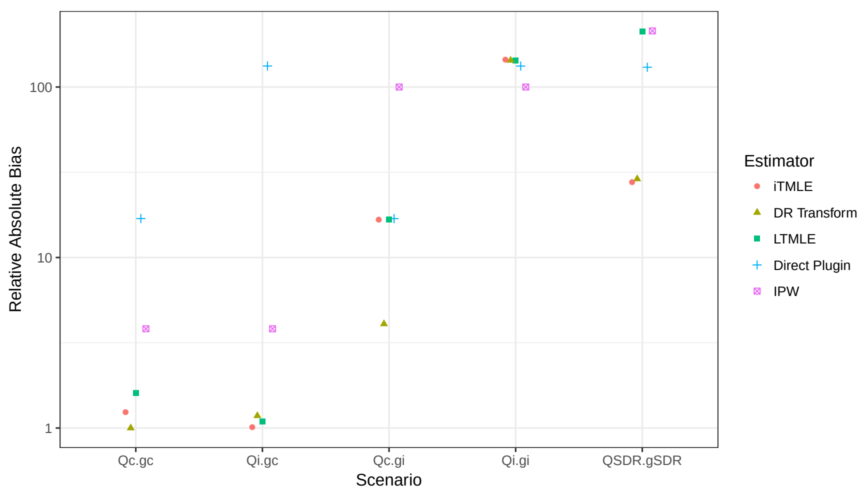

Figure 5: Relative absolute bias for $\hat{Q}_0$ in simulation scenario 2 with 5 time points.

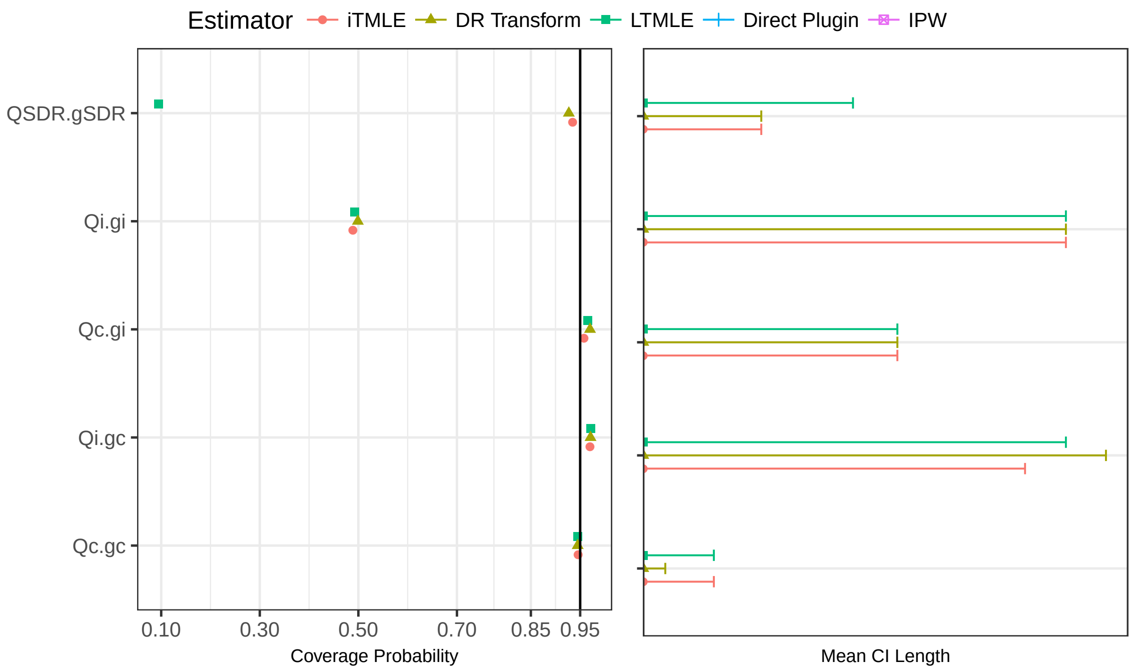

Figure 6: Coverage (left panel) and mean length (right panel) of the two-sided 95% CIs for Q_0 in simulation scenario 2.

restab	Relative Bias
Qc.gc.sdr	1.24
Qc.gc.drtrans	1.00
Qc.gc.tmle	1.61
Qc.gc.gcomp	16.93
Qc.gc.ipw	3.82
Qi.gc.sdr	1.01
Qi.gc.drtrans	1.18
Qi.gc.tmle	1.09
Qi.gc.gcomp	132.82
Qi.gc.ipw	3.82
Qc.gi.sdr	16.71
Qc.gi.drtrans	4.08
Qc.gi.tmle	16.74
Qc.gi.gcomp	16.91
Qc.gi.ipw	99.97
Qi.gi.sdr	144.25
Qi.gi.drtrans	143.26
Qi.gi.tmle	143.12
Qi.gi.gcomp	132.70
Qi.gi.ipw	99.97
QSDR.gSDR.sdr	27.60
QSDR.gSDR.drtrans	28.89
QSDR.gSDR.tmle	211.07
QSDR.gSDR.gcomp	130.58
QSDR.gSDR.ipw	213.37

Table 4: Simulation scenario 2. Relative absolute bias for estimation of Q_0, over 5 time points and n=5,000.

restab	Relative MSE
Qc.gc.sdr	1.44
Qc.gc.drtrans	1.00
Qc.gc.tmle	1.88
Qc.gc.gcomp	1.92
Qi.gc.sdr	1.05
Qi.gc.drtrans	1.12
Qi.gc.tmle	4.74
Qi.gc.gcomp	6.00
Qc.gi.sdr	1.64
Qc.gi.drtrans	1.65
Qc.gi.tmle	1.96
Qc.gi.gcomp	1.95
Qi.gi.sdr	5.73
Qi.gi.drtrans	5.19
Qi.gi.tmle	6.08
Qi.gi.gcomp	5.99
QSDR.gSDR.sdr	1.45
QSDR.gSDR.drtrans	1.39
QSDR.gSDR.tmle	7.40
QSDR.gSDR.gcomp	6.06

Table 5: Simulation scenario 2. Relative MSE for estimation of Q_1, over 5 time points and n=5,000.

restab	Coverage prob.
Qc.gc.sdr	0.946
Qc.gc.drtrans	0.945
Qc.gc.tmle	0.945
Qc.gc.gcomp	
Qi.gc.sdr	0.970
Qi.gc.drtrans	0.971
Qi.gc.tmle	0.971
Qi.gc.gcomp	
Qc.gi.sdr	0.958
Qc.gi.drtrans	0.970
Qc.gi.tmle	0.965
Qc.gi.gcomp	
Qi.gi.sdr	0.489
Qi.gi.drtrans	0.499
Qi.gi.tmle	0.492
Qi.gi.gcomp	
QSDR.gSDR.sdr	0.935
QSDR.gSDR.drtrans	0.927
QSDR.gSDR.tmle	0.095
QSDR.gSDR.gcomp	

Table 6: Simulation scenario 2. Coverage of 95% CIs for Q_0, over 5 time points and n=5,000.

Data generating distribution for simulation scenario 2.

```
nsims <- 500
nsamp <- 5000
tvals <- 4
require("simcausal")
D <- DAG.empty() +
    node("UL",    t = 0, distr = "rnorm") +
    node("W",     t = 0, distr = "rnorm", mean = 0) +
    node("L",     t = 0, distr = "rconst", const = abs(UL[t])) +
    node("phiRsk", t = 0, distr = "rconst", const = L[0]) +
    node("hiRsk", t = 0, distr = "rconst", const = plogis(phiRsk[t]) > 0.8) +
    node("A",     t = 0, distr = "rbern", prob = plogis(L[0])) +
    node("Y",     t = 0, distr = "rconst", const = 0) +

    node("UL",t = 1, distr = "rnorm") +
    node("L", t = 1, distr = "rconst", const = abs(UL[t])) +
    node("phiRsk", t = 1, distr = "rconst", const = -2 + 0.5*L[t-1] + 0.5*2*L[t]) +
    node("hiRsk", t = 1, distr = "rconst", const = plogis(phiRsk[t]) > 0.9) +
    node("A", t = 1, distr = "rbern", prob = A[t-1]*plogis(1.7 - 2.0*hiRsk[t])) +
    node("Y", t = 1, distr = "rbern",
      prob = plogis(-3 + 0.5*L[t-1]*A[t] + 0.5*A[t-1]*L[t] + 0.5*L[t]*A[t])) +

    node("UL",t = 2, distr = "rnorm", mean = A[0]*L[1] + L[1]*A[1]) +
    node("L", t = 2, distr = "rconst", const = abs(UL[t])) +
    node("phiRsk", t = 2, distr = "rconst", const = -2 + 0.5*L[t-1] + 0.5*2*L[t]) +
    node("hiRsk", t = 2, distr = "rconst", const = plogis(phiRsk[t]) > 0.85) +
    node("A", t = 2, distr = "rbern", prob = A[t-1]*plogis(1.7 - 2.0*hiRsk[t])) +
    node("Y", t = 2, distr = "rbern",
      prob = plogis(-3*Y[t-1] + 0.5*L[t-1]*A[t] + 0.5*A[t-1]*L[t] + 0.5*L[t]*A[t])) +

    node("UL",t = 3, distr = "rnorm", mean = L[1] * A[t-1] + A[1]*L[t-1] + L[t-1]*A[t-1]) +
    node("L", t = 3, distr = "rconst", const = abs(UL[t])) +
    node("phiRsk", t = 3, distr = "rconst", const = -2 + 0.5*L[t-1] + 0.5*2*L[t]) +
    node("hiRsk", t = 3, distr = "rconst", const = plogis(phiRsk[t]) > 0.80) +
    node("A", t = 3, distr = "rbern", prob = A[t-1]*plogis(1.7 - 2.0*hiRsk[t])) +
    node("Y", t = 3, distr = "rbern",
      prob = plogis(-3*Y[t-1] + 0.5*L[t-1]*A[t] + 0.5*A[t-1]*L[t] + 0.5*L[t]*A[t])) +

    node("UL",t = 4, distr = "rnorm", mean = L[1] * A[t-1] + A[1]*L[t-1] + L[t-1]*A[t-1]) +
    node("L", t = 4, distr = "rconst", const = abs(UL[t])) +
    node("phiRsk", t = 4, distr = "rconst",
      const = -1 + 0.25*L[t-1] + 0.25*2*L[t] - 0.1*L[t]*L[t-1] + 1.5*W[0]*L[t-1]) +
    node("hiRsk", t = 4, distr = "rconst", const = plogis(phiRsk[t]) > 0.80) +
    node("A", t = 4, distr = "rbern", prob = A[t-1]*plogis(2 - 2.0*hiRsk[t])) +
    node("Y", t = 4, distr = "rbern",
      prob = plogis(-1*Y[t-1] + A[t] + phiRsk[t]*A[t] + 0.20*A[t-1]*L[t]) )
```

Model specification for simulation scenario 2.

```r
## g model (same across all time-points):
gform.c <- c(rep.int("A ~ hiRsk + L + Atm1 + Ytm1", tvals), "A ~ hiRsk + Atm1 + Ytm1")
gform.i <- "A ~ 1"

## Q models:
Qforms.c <- c("Qkplus1 ~ L + W",
              "Qkplus1 ~ L + Ltm1 + A + Atm1 + Ytm1 + W",
              "Qkplus1 ~ L + Ltm1 + Ltm2 + A + Atm1 + Atm2 + Ytm1 + W",
              "Qkplus1 ~ L + Ltm1 + Ltm2 + A + Atm1 + Atm2 + Ytm1 + W",
              "Qkplus1 ~ L + Ltm1 + A + Atm1 + Ytm1 + W"
              )
Qinteract.c <- NULL

Qforms.i <- c("Qkplus1 ~ L + W",
              "Qkplus1 ~ W",
              "Qkplus1 ~ W",
              "Qkplus1 ~ W",
              "Qkplus1 ~ W")
Qinteract.i <- NULL

## g and Q model to demonstrate SDR property (only last g is correctly specified)
gform.SDR <- c(rep.int("A ~ 1", tvals-1), "A ~ 1", "A ~ hiRsk + Atm1 + Ytm1")
Qforms.SDR <- c("Qkplus1 ~ L + W",
                "Qkplus1 ~ L + Ltm1 + A + Atm1 + Ytm1 + W",
                "Qkplus1 ~ L + Ltm1 + A + Atm1 + Ytm1 + W",
                "Qkplus1 ~ L + Ltm1 + A + Atm1 + Ytm1 + W",
                "Qkplus1 ~ W")
Qinteract.SDR <- NULL
```

0.3 Combined results for simulation scenarios 1 and 2.

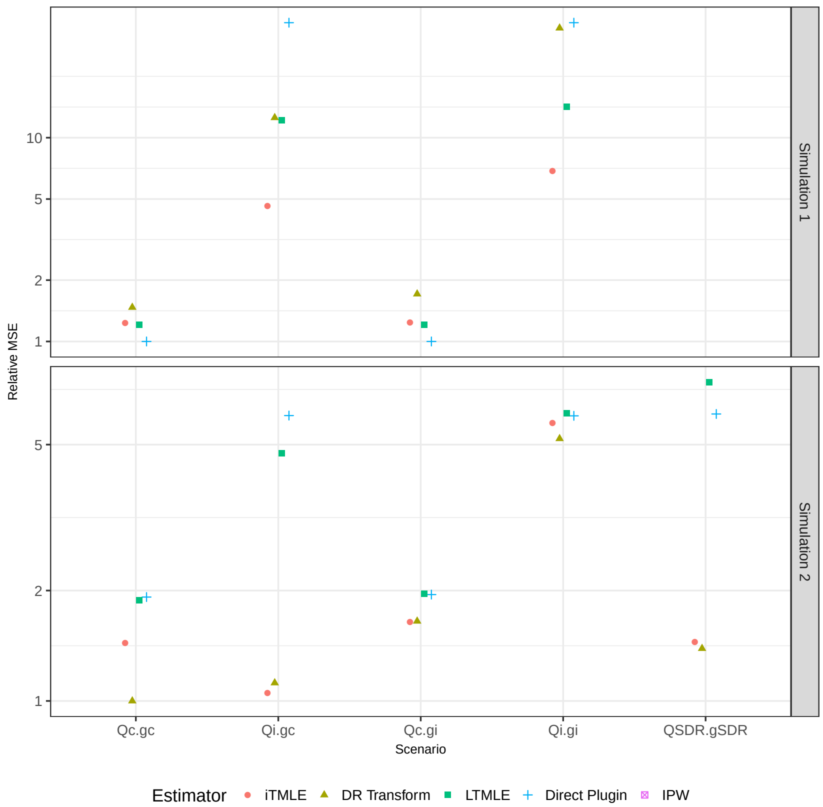

Figure 7: Relative MSE for $\hat{Q}_1$ for simulation scenario 1 (top panel) and simulation scenario 2 (bottom panel). Simulation 1 is based on longitudinal data with 3 time-points and n=500 observations. Simulation 2 is based on longitudinal data with 5 time-points and n=5,000 observations. The iTMLE and DR Transform typically outperform or perform comparably to both competitors.

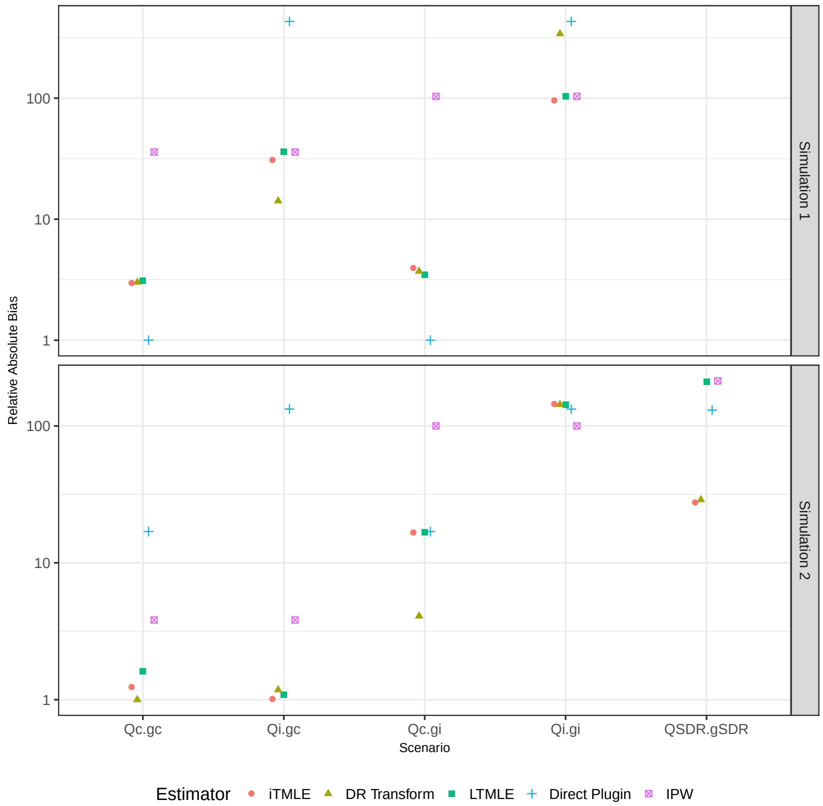

Figure 8: Relative absolute bias for $\hat{Q}_0$ for simulation scenario 1 (top panel) and simulation scenario 2 (bottom panel). Simulation 1 is based on longitudinal data with 3 time-points and n=500 observations. Simulation 2 is based on longitudinal data with 5 time-points and n=5,000 observations. The performance of LTMLE, iTMLE, and DR Transform is similar. The only exception for Simulation 1 is under $Qi.gc$, where DR Transform outperforms other methods. The only exceptions for Simulation 2 are for $Qc.gi$, where DR Transform outperforms other methods, and $QSDR.gSDR$, where both SDR methods outperform LTMLE.

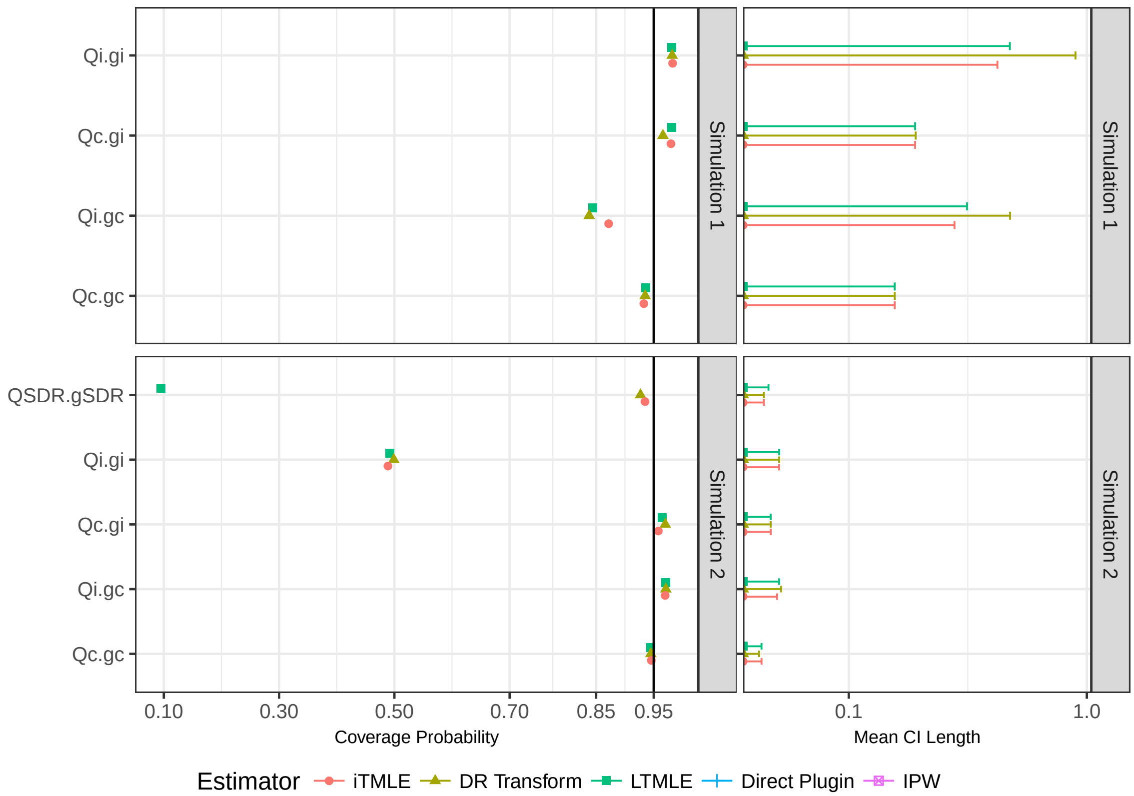

Figure 9: Coverage (left panels) and mean length (right panels) of the two-sided 95% CIs for Q_0 in simulation scenario 1 (top panels) and simulation scenario 2 (bottom panels). Confidence interval coverage and width appear to be comparable between the two SDR methods and the LTMLE. The only exception is for the $QSDR.gSDR$ scenario, where the LTMLE has roughly 10% coverage, whereas the SDR approaches nearly achieve the nominal coverage level.

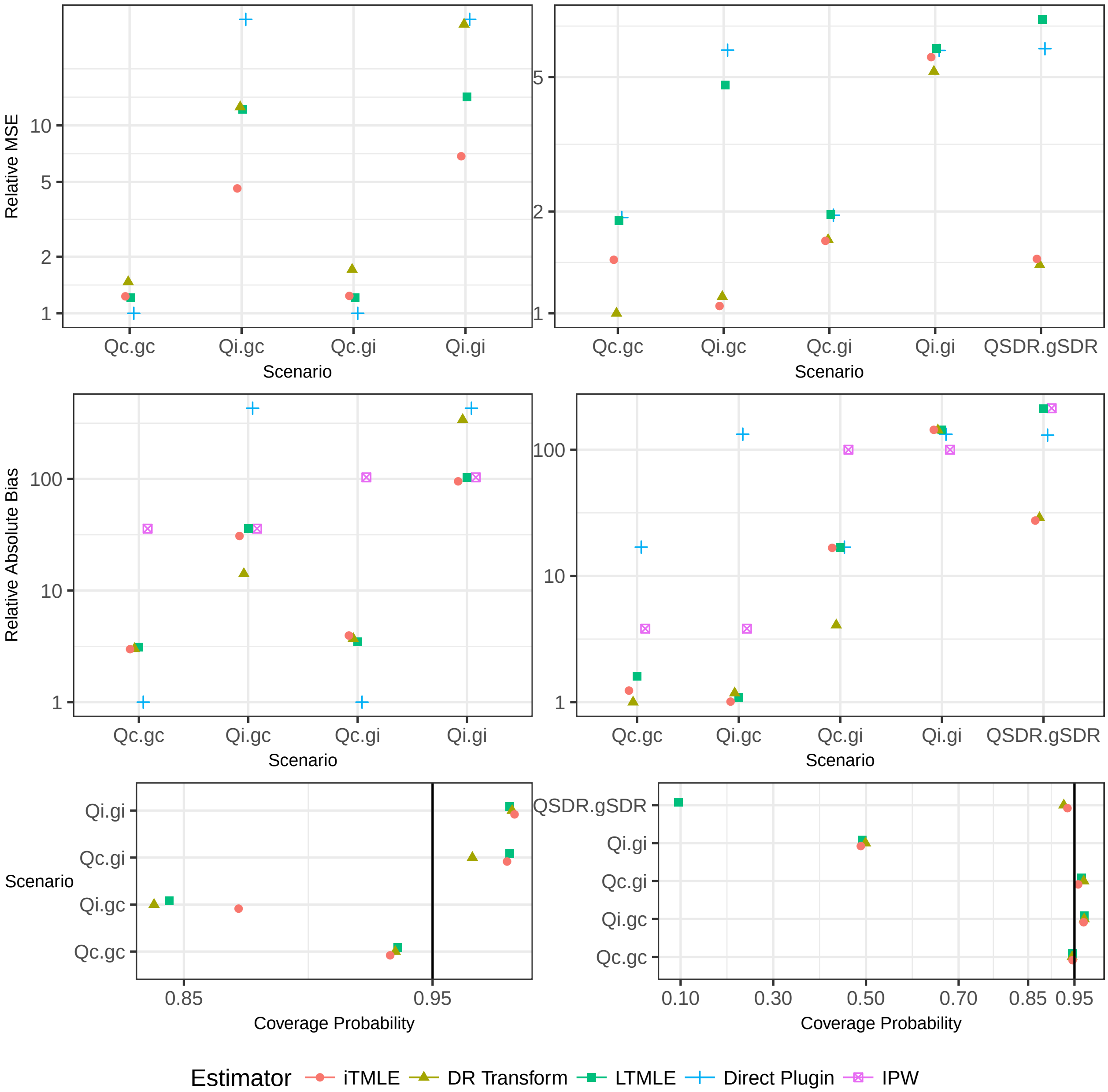

Figure 10: Top to bottom: MSE for estimation of Q_1, bias for estimation of Q_0 and 95% CI coverage for estimation of Q_0. Left plot: simulation scenario 1. Right plot: simulation scenario 2.