Summary of the simulation results

These results are based on, and further analyses can be carried out with the result files in the zip archive *simulation_results.zip*.

Convergence

The convergence rate was at least 98% for all studied range combinations and values of Sp for $n = 500$, 99.2% for $n = 1000$, 99.8% for $n = 2000$, and 100% for $n \geq 5000$.

Power to detect that Y depends on X

Missing upper one-third of the probability range for Se and Y, and low Sp influenced the power negatively. For $n \leq 1000$, this resulted in power as low as 40-50%. From $n = 2000$, power exceeded 90% for almost all combinations of probability ranges and Sp values.

For $n = 500$, the power was above 89% when the upper one-third of the probability range for Se was present. When it was missing, the power varied between 41.6% (low Y, Sp≤0.8) and 100% (whole Y, Sp≥0.999).

For n=1000, the power reached 80% for most combinations of ranges, except for low probability range for Se with Sp ≤ 0.9. The smallest values (41.6% and 53.8%) occurred for low Y, low Se, Sp=0.7 and 0.8, respectively.

For n=2000, the power reached 93% for all combinations of ranges, except for three ones: whole Y, low Se, Sp = 0.7 (70.4%), low Y, low Se, Sp = 0.8 (62.6%), low Y, low Se, Sp = 0.7 (52.8).

For n=5000, the power reached 98% for all combinations of ranges, except for three ones: whole Y, low Se, Sp = 0.7 (74.8%), low Y, low Se, Sp = 0.8 (71.8%), low Y, low Se, Sp = 0.7 (70.6%).

For n=10000, the power was at least 98.6% for all combinations of ranges, except for three ones: whole Y, low Se, Sp = 0.7 (77.8%), low Y, low Se, Sp = 0.8 (75.0%), low Y, low Se, Sp = 0.7 (86.0%).

Power to detect that Se depends on Z

The power was negatively influenced by low Sp and missing upper one-third of the probability range for Y. From n = 1000, the power reached 80% for almost all combinations of probability ranges and Sp values.

For n = 500, the power to detect that Se depends on Z exceeded 90% when the upper one-third of the probability range of Y was present, and also when it was missing but the whole probability range for Se was present.

For n=1000, the power reached 80% for most combinations of ranges, except for low Y and low Se with Sp $\leq$ 0.8, when it was 49.0% and 71.6% for Sp=0.7 and 0.8, respectively.

For n=2000, the power reached 88% for all combinations of ranges, except for low Y and low Se with Sp = 0.7, when it was 66.0%.

For n=5000, the power exceeded 97% for all combinations of ranges.

For n=10000, the power was 100% for all combinations of ranges.

The betas

The estimates of the betas were most accurate when the upper one-third for Se was present and Sp was high. If the upper one-third for both Y and Se were present and Sp was at least 0.9, the estimates had acceptable accuracy (RMSE<0.55) from n = 1000.

The estimate of beta1 seems to have an upwards bias (about 0.1 and 0.03 for n=500 and 5000) but estimating it from our simulation study is rather uncertain. Bias of beta0 seems to be smaller and downwards direction (about 0.05 and 0.01 for n=500 and 5000), but estimating it is even more uncertain.

For n = 500, RMSE of beta1 was 0.328 when the whole range was present for both Y and Se, and Sp was 1 (0.331 for Sp=0.999). Beta1 and beta0 were estimable with RMSE≤0.331 and 0.336, respectively, when the whole range for Y and the upper one-third for Se were present, and Sp was at least 0.999.

For n=1000, RMSE of beta1 was 0.155 when the whole range was present for both Y and Se, and Sp was 1 (0.143 for Sp=0.999). Beta1 and beta0 were well estimable with RMSE≤0.343 and 0.546, respectively, when the upper one-third was present for both Y and Se, and Sp was at least 0.9.

For n=2000, RMSE of beta1 was 0.1 when the whole range was present for both Y and Se, and Sp was 1 (0.103 for Sp=0.999). Beta1 was well estimable with RMSE≤0.242 if the upper one-third was present for both Y and Se. If the upper one-third was missing for Y but it was present for Se, and Sp was at least 0.9, the RMSE of beta1 remained under 0.384. Beta0 was well estimable with RMSE≤0.493 if the upper one-third was present for both Y and Se. If the upper one-third was missing for Y but it was present for Se, and Sp was at least 0.9, the RMSE of beta1 remained under 0.558.

For n=5000, RMSE of beta1 was 0.059 when the whole range was present for both Y and Se, and Sp was 1 (0.061 for Sp=0.999). If the upper one-third of the Se probability range was not missing, RMSE was at most 0.206 (at most 0.127 when Sp was at least 0.9). Similarly, beta0 was also well estimable (with RMSE ≤ 0.316) when the upper one-third of the Se probability range was not missing.

For n=10000, RMSE of beta1 was 0.040 when the whole range was present for both Y and Se, and Sp 0.999 or 1. If the upper one-third of the Se probability range was not missing, RMSE was at most 0.126 (at most 0.086 when Sp was at least 0.9). Similarly, beta0 was also well estimable (with RMSE ≤ 0.211) when the upper one-third of the Se probability range was not missing.

The gammas

The estimates of the gammas were most accurate when the upper one-third of Y was present, and Sp was high. If the upper one-third was present for both Y and Se, the estimates had acceptable accuracy from n = 1000.

Bias of gamma1 seems to be upwards (about 0.1 and 0.01 for n=500 and 5000), but estimating is rather uncertain from the simulation. Bias of gamma0 seems to be smaller and downwards (about 0.01 and 0.005 for n=500 and 5000) but estimating it is even more uncertain.

For n = 500, RMSE of gamma1 was 0.231 when the whole range was present for both Y and Se, and Sp was 1 (0.877 for Sp=0.999). RMSE of the gammas remained below 0.877 if the whole range for Se and the upper one-third for Y were present.

For n=1000, RMSE of gamma1 was 0.132 when the whole range was present for both Y and Se, and Sp was 1 (0.136 for Sp=0.999). RMSE of the gammas remained below 0.251 if the upper one-third was present both for Y and Se.

For n=2000, RMSE of gamma1 was 0.093 when the whole range was present for both Y and Se, and Sp was 1 (0.092 for Sp=0.999). RMSE of gamma1 remained below 0.227 if the upper one-third of Y was present and Sp was higher than 0.7. RMSE of gamma0 was below 0.365 if the upper one-third of Y was present. If the upper one-third of Se was also present, it remained below 0.186.

For n=5000, RMSE of gamma1 was 0.055 when the whole range was present for both Y and Se, and Sp was 1 (0.059 for Sp=0.999). Gamma1 was well estimable with RMSE ≤ 0.121 if the upper one-third of Y was present and Sp was greater than 0.7. RMSE of gamma0 was at most 0.219 if the upper one-third of Y was present.

For n=10000, RMSE of gamma1 was 0.038 when the whole range was present for both Y and Se, and Sp was 1 (0.040 for Sp=0.999). Gamma1 was well estimable with RMSE ≤ 0.081 if the upper one-third of Y was present and Sp was greater than 0.7. RMSE of gamma0 was at most 0.168 if the upper one-third of Y was present.

Sp

The estimate of Sp was most accurate when the lower one-third for Y and the whole range of Se were present, and second most accurate when the lower one-third for Y and the upper one-third for Se were present. Accuracy was acceptable from n = 2000.

Bias of Sp depends on its true value. For true Sp<0.9, the estimate is biased upwards, for Sp>0.9 downwards. Estimation of bias was rather uncertain.

For n=500, RMSE of Sp was smallest (0.037 to 0.041 for Sp≤0.9 and 0.009 to 0.010 for Sp≥0.999) when the whole range for both Y and Se was present.

For n=1000, RMSE of Sp was smallest (0.021 to 0.035 for Sp≤0.9 and 0.004 to 0.006 for Sp≥0.999) when the lower one-third for Y and the whole range for Se was present. With missing lower one-third for Se it was higher (0.034 to 0.062 for Sp≤0.9 and 0.010 to 0.011 for Sp≥0.999).

For n=2000, RMSE of Sp was smallest (0.013 to 0.019 for Sp≤0.9 and 0.003 to 0.004 for Sp≥0.999) when the lower one-third for Y and the whole range for Se was present. With missing lower one-third for Se, it was a little higher (0.023 to 0.032 for $Sp \leq 0.9$ and 0.006 to 0.008 for $Sp \geq 0.999$).

For n=5000, RMSE of Sp was smallest (0.008 to 0.019 for $Sp \leq 0.9$ and 0.001 to 0.004 for $Sp \geq 0.999$) when the lower one-third for Y and the whole range for Se was present. With missing lower one-third for Se, it was a little higher (0.013 to 0.019 for $Sp \leq 0.9$ and 0.004 for $Sp \geq 0.999$).

For n=10000, RMSE of Sp was smallest (0.006 to 0.009 for $Sp \leq 0.9$ and 0.001 to 0.002 for $Sp \geq 0.999$) when the lower one-third for Y and the whole range for Se was present. With missing lower one-third for Se, it was a little higher (0.009 to 0.013 for $Sp \leq 0.9$ and 0.003 for $Sp \geq 0.999$).