This document describes additional simulation results empirically illustrating the analytical results described in: Bartlett JW, Harel O, Carpenter JR. Asymptotically unbiased estimation of exposure odds ratios in complete records logistic regression. Am J Epidemiol. 2015; 182(8):730–736, DOI: 10.1093/aje/kwv114

To illustrate the analytical derivations provided in the aforementioned paper, here we present results of a simple simulation study. Datasets of size 1,000 were simulated. A binary exposure X was simulated with prevalence 50%. A continuous confounder Z was generated, with distribution $N(X,1)$. The binary outcome was generated from a logistic regression with log odds of disease equal to -3+X+Z, such that β_0=-3, and $\beta_X=\beta_Z=1$. Estimates were obtained based on full data, and also based on a CRA under five missingness mechanisms, each of which gave an approximate 50% missingness rate (except mechanism 4). The missingness mechanisms were given by $P(R=1|X,Y,Z)$ equal to (where $\text{expit}(a)=\exp(a)/(1+\exp(a))$):

1. 0.9Y+(1-Y)0.41 (outcome dependent missingness)
2. $s(X,Z) = (X+2(1-X))/3$ (covariate dependent missingness)
3. $t(Y,Z) = \text{expit}(-0.6+Y+Z)$ (confounder and outcome dependent missingness)
4. $s(X,Z)t(Y,Z)$ (exposure, confounder and outcome dependent missingness)
5. $\text{expit}(-1.1+x+y+z)$ (exposure, confounder and outcome dependent missingness)

The table shows the results of the simulation study. As expected, the estimates of all three parameters were unbiased with full data. With outcome dependent missingness (missingness mechanism 1), as expected the intercept was estimated with bias, but the exposure and confounder effects were estimated without bias. With missingness dependent on the exposure (mechanism 2), estimates of all parameters were unbiased. With missingness dependent jointly on the outcome and confounder (mechanism 3), as implied by the paper's analytical results, while the estimate of the exposure effect was unbiased, the confounder (and intercept) effect was estimated with bias. When missingness depended jointly on exposure, confounder and outcome, but such that $P(R=1|X,Y,Z)=s(X,Z)t(Y,Z)$ (mechanism 4), again the exposure effect was estimated without bias, but the intercept and confounder parameters were estimated with bias. Lastly, with missingness jointly dependent on all three variables but in such a way that the probability of observation does not factorise into the form $s(X,Z)t(Y,Z)$ (mechanism 5), estimates of all parameters were biased. In summary, the biases found are in line with the analytical results of the paper.

As expected, the variability of CRA was larger than full data analysis. The variability of CRA was fairly similar across mechanisms 1, 2, 3, and 5, in which 50% of records were complete, with increased variability for mechanism 4, in which approximately 25% of records were complete.

Table: Simulation Results for Complete Records Analysis Logistic Regression with Full Data and Five Missing Data Mechanisms. Mean (SD) of log Odds Ratio Estimates Across 1,000 Simulations are Shown.

Missingness mechanism: $P(R=1\|X,Y,Z)$	$\beta_0=-3$	$\beta_X=1$	$\beta_Z=1$
Full data	-3.02 (0.20)	1.01 (0.22)	1.00 (0.11)
1. 0.9Y+(1-Y)0.41	-1.23 (0.23)	1.01 (0.26)	1.02 (0.13)
2. s(X,Z) = (X+2(1-X))/3	-3.04 (0.26)	1.02 (0.30)	1.01 (0.17)
3. t(Y,Z) = expit(-0.6+Y+Z)	-2.54 (0.26)	1.01 (0.26)	0.83 (0.13)
4. s(X,Z)t(Y,Z)	-2.56 (0.34)	1.02 (0.37)	0.84 (0.21)
5. expit(-1.1+x+y+z)	-2.47 (0.27)	0.84 (0.27)	0.86 (0.13)