**Data S2.** Model specifications for imputation by multiple imputation chained equation (MICE).

Suppose is the -th variable in the incomplete data set with a total of *K* partially observed variables, . Let denote the set of all variables in the data set except . Each is modeled conditional on and all other parameters. The conditional model can be chosen in various ways (e.g. predictive mean matching, Bayesian parametric regression, logistic regression). The algorithm imputes the missing values variable-by-variable for all ’s and iterates until convergence. The initial values to start the chain can be chosen as the observed mean for all variables with missing values.

For each of the 50 imputed datasets, missing values for continuous variables (e.g. pre-pregnancy BMI, gestational age) and binary variables (e.g. use of alcohol, insurance type) were imputed using predictive mean matching and logistic regression. These algorithms are described in detail in the documentation for the package (<https://cran.r-project.org/web/packages/mice/mice.pdf>).1 For example, predictive mean matching calculates predictive values for both observed and missing values of . An observed value is then randomly drawn from a group of candidates whose predictive values are close to the predictive value of the missing entry. This observed value is then taken as the imputed value for the missing entry. The maximum number of iterations for each draw was set to be ten. Imputation diagnostics and convergence checks were carried out following standard guidelines for multiple imputation.2

REFERENCES

1. Pinheiro J BD, DebRoy S, Sarkar D and R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. 2017; R package version 3.1-131, [https://cran.r-project.org/package=nlme](https://CRAN.R-project.org/package=nlme).
2. Bondarenko I, Raghunathan T. Graphical and numerical diagnostic tools to assess suitability of multiple imputations and imputation models. *Statistics in medicine.* 2016;35(17):3007-3020.