Simultaneous Treatment of Missing Data and Measurement Error in HIV Research Using Multiple Overimputation

Michael Schomaker,[a] *Sara Hogger,*[b] *Leigh F. Johnson,*[a] *Christopher J. Hoffmann,*[c,d] *Till Bärnighausen,*[e,f] *and Christian Heumann*[b]

Background: Both CD4 count and viral load in HIV-infected persons are measured with error. There is no clear guidance on how to deal with this measurement error in the presence of missing data.

Methods: We used multiple overimputation, a method recently developed in the political sciences, to account for both measurement error and missing data in CD4 count and viral load measurements from four South African cohorts of a Southern African HIV cohort collaboration. Our knowledge about the measurement error of ln CD4 and $\log_{10}$ viral load is part of an imputation model that imputes both missing and mismeasured data. In an illustrative example, we estimate the association of CD4 count and viral load with the hazard of death among patients on highly active antiretroviral therapy by means of a Cox model. Simulation studies evaluate the extent to which multiple overimputation is able to reduce bias in survival analyses.

Results: Multiple overimputation emphasizes more strongly the influence of having high baseline CD4 counts compared to both a complete case analysis and multiple imputation (hazard ratio for >200 cells/mm^3 vs. <25 cells/mm^3: 0.21 [95% confidence interval: 0.18, 0.24] vs. 0.38 [0.29, 0.48], and 0.29 [0.25, 0.34], respectively). Similar results are obtained when varying assumptions about measurement error, when using p-splines, and when evaluating time-updated CD4 count in a longitudinal analysis. The estimates of the association with viral load are slightly more attenuated when using multiple imputation instead of multiple overimputation. Our simulation studies suggest that multiple overimputation is able to reduce bias and mean squared error in survival analyses.

Conclusions: Multiple overimputation, which can be used with existing software, offers a convenient approach to account for both missing and mismeasured data in HIV research.

(*Epidemiology* 2015;26: 628–636)

It is well known that both CD4 count and viral load in HIV-infected persons are measured with error, due to physiologic and biologic variation and to assay performance.[1,2] Biologic variation includes intrapersonal fluctuations of CD4 cell count over the course of a circadian cycle and from day to day as a result of psychological stresses, intercurrent illnesses, alcohol, caffeine, exercise, and other factors.[1,3,4] Assay variation in CD4 measurements refers to flow cytometry itself and variation attributed to the assays used, their accuracy, specimen preparation techniques, the age of the sample at the time of preparation, and sample conditions during transport to a laboratory.[2,3] Measured CD4 count may therefore not represent the true underlying CD4 count. The same applies to the accuracy of HIV RNA (viral load) measurements: biological variation, related to disease progression, illnesses and lifestyle factors as well as technical variation due to different assays, laboratory standards, technician's skills, and storage temperatures can cause a considerable amount of measurement error.[3,5–8]

Failure to appreciate the extent of measurement error may lead to biased results, for example regression estimates can either be attenuated or strengthened.[9] This makes adjustment for measurement error a topic of considerable interest in the statistical analysis of HIV data.[10] Suggestions for CD4 count measurement error correction include regression calibration,[11–13] and approaches which correct the likelihood function.[14–18] However, these methods have been rarely used in practice because of either their complicated implementation or their construction for a particular regression model or

Submitted 11 July 2014; accepted 19 May 2015.

Editors' note: a commentary on this article appears on page 625.

From the [a]Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Cape Town, South Africa; [b]Institut für Statistik, Ludwig-Maximilans-Universität München, München, Germany; [c]Aurum Institute, Johannesburg, South Africa; [d]School of Medicine, Johns Hopkins University, Baltimore, MD; [e]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA; and [f]Wellcome Trust Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Mtubatuba, South Africa.

Supported by the US National Institute of Allergy and Infectious Diseases (NIAID) through the International epidemiological Databases to Evaluate AIDS, Southern Africa (IeDEA-SA), Grant No 5U01AI069924-05. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

The authors have no conflicts of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com). This content is not peer-reviewed or copy-edited; it is the sole responsibility of the authors.

Correspondence: Michael Schomaker, Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Falmouth Building, Anzio Road, Observatory 7925, Cape Town, South Africa. E-mail:Michael.schomaker@uct.ac.za.

ISSN: 1044-3983/15/2605-0628
DOI: 10.1097/EDE.0000000000000334

study question.[19–21] Furthermore, these approaches require complete data, which limits application in low-income high-burden programmatic care settings where missing data—often related to missed laboratory visits, lost records, or incomplete data capture—is common.[22,23] We are aware of only one approach which accounts for missing data in the presence of measurement error,[15] using a particular nonignorable missingness assumption for the outcome of a nonlinear mixed effect model. However, this specific setting would typically not be relevant to most HIV research as its motivation and assumptions relate only to long-term viral dynamics modeling.

To address the problem of missing data in HIV research several approaches can be considered. A general and relatively straightforward approach to deal with missing data is multiple imputation, which is implemented in many statistical software packages.[24] Based on the user's assumptions about the data distribution (imputation model) missing values can be filled in (imputed) by means of draws from the posterior predictive distribution of the unobserved data given the observed data. This procedure is repeated to create M imputed datasets, the analysis is then conducted on each of these datasets and the M results are combined by a set of simple rules. Multiple imputation yields valid inference under the missing-at-random assumption which states that the probability of any value to be missing from the dataset depends only on the observed data.[25]

We show how multiple overimputation,[19] recently proposed in the political sciences and closely related to multiple imputation, can be used to account for both missing at random data *and* measurement error in HIV research under a general framework. Multiple overimputation treats mismeasured data as an extreme case of missing data: values measured with error are replaced with values obtained from an imputation model that incorporates the mismeasured values, as well as knowledge and assumptions about the measurement error process, in prior distributions on individual measurements. After generating multiple overimputed datasets, standard multiple imputation combining rules can be applied to obtain valid inference under assumptions which are similar to missing at random. The method has the main advantages of (1) being easy to implement with existing software, (2) being applicable to a wide range of analysis models and settings, including longitudinal data analyses, and (3) addressing measurement error and missing data simultaneously.

While the method has been tested in the political sciences and first simulations showed promising results in the context of linear and logistic regression models, little is known about the assumptions, behavior, and success of the method in the context of HIV analyses, particularly survival analyses.

We therefore aim to (1) identify an appropriate measurement error model for CD4 count and viral load, (2) to investigate the implications, assumptions, and challenges related to the implementation of multiple overimputation in HIV research, using South African HIV treatment cohort data from patients starting on highly active antiretroviral treatment (HAART), and (3) to quantify the association of both baseline and follow-up CD4 count and viral load with all-cause mortality and to explore the possible bias resulting from ignoring measurement error and missing data in this illustrative example. In addition, (4) simulations are used to evaluate the extent to which multiple overimputation is able to reduce bias arising from measurement error and missing data in a wide range of survival analysis settings.

METHODS

Framework of Multiple Overimputation in General and for HIV Research

Multiple Overimputation

Multiple Overimputation builds on multiple imputation by interpreting mismeasured values as missing data but including the mismeasured values as prior information in the imputation model. The procedure is as follows:

(1) Multiply impute (say $M = 5$ times) missing values and multiply overimpute (replace, overwrite) mismeasured values based on an appropriate imputation model which uses assumptions about the mismeasured data as prior information.
(2) Conduct any statistical inference (Cox model, Kaplan–Meier estimator,…) on each overimputed set of data.
(3) Combine the M estimates related to the M overimputed sets of data according to standard multiple imputation combining rules ("Rubin's rules").[26]

For example, if we had 1,000 patients and 800 of them had available baseline CD4 counts, we would impute the remaining 200; the 800 measured CD4 counts would be treated as mismeasured, as we know that they do not exactly represent the true CD4 count of a patient, but rather randomly differ from the true value. We would thus overwrite these values from an imputation model which uses our assumptions about the measurement error process as prior information. Subsequently, we would perform our analysis on each overimputed dataset and combine the results accordingly.

Multiple Imputation with Amelia II

It is known from multiple imputation theory that *proper* multiple imputations (yielding valid inference under the missing-at-random assumption) are realized via draws from the posterior predictive distribution of the unobserved data given the observed data.[25] These draws can, for example, be generated by specifying a multivariate distribution of the data and simulating the predictive posteriori distribution with a suitable algorithm. For our analysis, we consider the Expectation Maximization Bootstrap (EMB) algorithm[27] from the *R*-package Amelia II,[28] which assumes a multivariate normal distribution for the data, $D \sim N(\mu, \Sigma)$ (possibly after suitable transformations beforehand). In this algorithm, B bootstrap samples of the data (including missing values) are drawn and in each

bootstrap sample the EM algorithm[29] is applied to obtain estimates of μ and Σ which can then be used to generate proper imputations by means of the sweep-operator.[27,30] Of note, the algorithm can handle highly skewed variables by imposing transformations on variables (log, square root,…) and recodes categorical variables into dummies based on the knowledge that for binary variables the multivariate normal assumption can yield good results.[31]

Multiple Overimputation with Amelia II

We assume (1) a classical measurement error model, meaning that for any observation: $\text{observed}\, x_{ij}^{*} = \text{true}\, x_{ij} + u_{ij}, u_{ij} \sim N\left(0, \sigma_{u_{ij}}^2\right)$ with $\sigma_{u_{ij}}^2$ known or estimated and (2) that the data are mismeasured at random, meaning that the presence of a mismeasured and/or missing value may depend only on observed quantities (and not the unobserved value itself), see eText 1 (http://links.lww.com/EDE/A933) for a formal definition.

Consider the unobserved data to consist of *both* the missing data *and* the true latent values x_{ij}. Blackwell et al.[19] extend the predictive posterior distribution of the unobserved data given the observed data so that both missing and latent values are treated as unobserved. Using this extended predictive posterior distribution, applying the EMB algorithm onto this distribution to obtain imputations, and incorporating the classical measurement error assumptions, $x_{ij}^{*} \sim N\left(x_{ij}, \sigma_{U_{ij}}^2\right)$, into the E-step of the algorithm, allows the use of a multiple imputation framework.[19] Most crucially, the authors show that this modified EMB algorithm leads to identical solutions when compared with using this algorithm implemented in Amelia II when prior distributions on mismeasured values that relate to $x_{ij} \sim N\left(x_{ij}^{*}, \sigma_{U_{ij}}^2\right)$ are used. The reason why multiple overimputation is different from multiple imputation and has the potential to correct for measurement error is because the draws are based on a modified predictive posterior distribution which incorporates the classical measurement error assumptions; see eText 1 (http://links.lww.com/EDE/A933) and the appendices of Blackwell et al. and Honaker et al. for more details.[19,27]

Thus, in summary, using existing software for multiple imputation (Amelia II) and specifying observation level priors for each mismeasured value (normal distribution with the mean relating to the mismeasured value and the variance known or estimated), accounts for both missing and mismeasured data under the above-mentioned assumptions.

After creating M overimputed datasets, the analysis model (e.g., the Cox proportional hazards regression model) can be fitted in each overimputed dataset. The M estimates can then be combined easily either with existing commands contained in most statistical software packages or by hand: the point estimate is just the average of the M point estimates, whereas the variance reflects both the uncertainty in each overimputed dataset and between imputed datasets (eText 1; http://links.lww.com/EDE/A933).

A Measurement Error Model for CD4 Count and Viral Load

Multiple overimputation can be applied to correct for measurement error in both baseline and follow-up HIV RNA and CD4 count with the following assumptions:

(1) A classical measurement error model for both natural logarithm CD4 count (cells/mm^3) as well as $\log_{10}$ viral load (copies/μl):

$$\ln \text{CD}_{4_i}^{*} = \ln \text{CD}_{4_i} + u_i,\ u_i \mid \ln \text{CD}_{4_i} \sim N\left(0, 0.26^2\right)$$

$$\text{Log}_{10}\text{VL}_i^{*} = \log_{10}\text{VL}_i + u_i,\ u_i \mid \log_{10}\text{VL}_i \sim N\left(0, 0.255^2\right)$$

(2) The data are mismeasured at random.

The first assumption is a classical measurement error model. This assumption has been used before in methodological work[12,14,17] and implies increased measurement error for higher absolute (nonlog) CD4 count and (nonlog) viral load measurements which is in line with clinical knowledge.[2,3,6,32] The measurement error variance for the natural logarithm CD4 was obtained from an estimate of a study with a large sample.[3] The estimated variance was similar in studies with smaller samples ($0.275^{2[14]}$ and $0.25^{2[32]}$); other studies report slightly lower estimates but do not necessarily reflect all sources of measurement error.[1,8]

The measurement error variance for $\log_{10}$ viral load is based on Lew et al.[6] who conclude that variation due to biological and technical factors is fairly consistent and in the range of 0.3 to 0.6 $\log_{10}$ copies/ml. Based on this observation, we may assume that the upper and lower limits of a 95% confidence interval for the measured viral load correspond to the true viral load ±0.5. This yields a measurement error variance of approximately 0.255^2 (where 0.255 = 0.5/1.96). This is in line with another report (0.264^2 for viral loads >500 copies/ml).[2]

The second assumption states that the probability of a missing or incorrectly measured value depends only on observed quantities, see eText 1 (http://links.lww.com/EDE/A933) for a detailed definition. We therefore use the term "mismeasured at random" to mean that both the missingness process *and* the measurement error process must not depend on any unobserved values.

In situations where clerical or administrative errors cause a value to be missing, such as in large cohorts where data capturing capacity may be limited, this assumption is certainly fulfilled. If the probability of missingness (or occurrence of measurement error) depends on captured information, such as treatment facility, region, or date of treatment initiation, the assumption would also be fulfilled. In the case where unobserved values determine the probability of missingness (or occurrence of measurement error) the assumption would be violated; for example if particularly high CD4 or viral load measurements were missing or incorrectly measured, or if the missing data relates to a specific healthcare worker and this is not captured. Possible consequences of such situations are described in the discussion.

Simulation Studies

Generation of Data

We generated data of sample size n = 5,000 for the main setting, and n = 1,000, 2,500, 7,500, 10,000 for further settings. Two covariates, representing true log CD4 count and true log viral load, were drawn from log-normal distributions with mean and standard deviations (adapted from the data analysis below) as follows: $X_1 \sim \log N(4.286, 1.086)$ and $X_2 \sim \log N(10.76, 1.8086)$. We used a Clayton Copula (with copula parameter $\theta = 1$ indicating moderate association) to model the dependency between these two variables.[33,34] Survival times Y were simulated as follows:

$$Y = -\frac{\log(U)}{h_0\{\exp(X\beta)\}},$$

where U is drawn from a distribution that is uniform on the interval [0,1], $h_0 = 0.1$, and the linear predictor $X\beta$ is defined as $-0.3 \ln X_1 + 0.3 \log_{10} X_2$. Higher values of X_1 are therefore associated with a lower risk of a (death) event, as is the case for CD4 count, while higher values of X_2 are associated with a higher risk of an event, as is the case with viral load. The censoring times were simulated as

$$C = -\frac{\log(U)}{0.2}$$

The observed survival time T in our simulation was thus $T = \min(Y, C)$.

Measurement Error and Missing Data

To both log-transformed variables, we added measurement error, as in our data, with mean 0 and variances of 0.26^2 and 0.255^2, respectively.

X_1 and X_2 were assumed to be missing at random and the missing indicator was simulated by means of the following missingness function:

$$\pi_X(T) = 1 - \{1 + \exp(1 - 4T)\}^{-1}$$

This yields approximately 9% missing values per variable. Since, in this simulation, the probability of missingness depends on the outcome, one would expect parameter estimates in a regression model of a complete case analysis to be biased.[35]

Estimators and Model

We compare the performance of (1) a complete case analysis (omitting observations with missing values), (2) multiple imputation, and (3) multiple overimputation when estimating the parameters in a Cox proportional hazards model. We also compare (1) the naive estimator and multiple overimputation for the setting without missing data. The multiple (over)imputation model included all variables, but T on a log scale.

Measures of Performance

We evaluate the bias, mean squared error (MSE), and distribution of each estimator of β_i via R = 1,000 runs of the simulation study. The bias is estimated as $R^{-1}\sum_{r=1}^{R}\breve{\beta}_{ir} - \beta_{ir}$, the MSE as $R^{-1}\sum_{r=1}^{R}\left(\beta_{ir} - \breve{\beta}_{ir}\right)^2$.

Sensitivity

To explore the sensitivity of our simulation, we varied our assumptions with respect to the amount of measurement error, the missingness process, the correct specification of the measurement error variance, and the linear predictor.

RESULTS

Results of Simulation Studies

One can see that a complete case analysis yields biased results both when dealing only with mismeasured data (Table, "no missing data" panel) and when dealing with mismeasured and missing data (Table, "missing data" panel). Multiple imputation also yields biased results in our missing-at-random setting when confronted with measurement error (Table, "missing data" panel). Multiple overimputation considerably reduces bias when compared with the two aforementioned approaches (Table). Comparing the distribution of parameter

TABLE. Bias, Variance, and Mean Squared Error in the Main Simulation Study: Results Are Reported for Both β_1 and β_2—for a Naive/CC Analysis, MI, and MO, Respectively

	No Missing Data			Missing Data		
	Naive	MI	MO	CC	MI	MO
Bias β_1	0.034	-	0.023	0.033	0.029	0.017
Bias β_2	-0.048	-	-0.019	-0.045	-0.042	-0.009
Variance β_1	0.0006		0.0009	0.0007	0.0007	0.0010
Variance β_2	0.0010		0.0017	0.0011	0.0011	0.0019
MSE β_1	0.0018	-	0.0014	0.0017	0.0015	0.0012
MSE β_2	0.0033	-	0.0021	0.0031	0.0029	0.0020

The left panel lists results for the setting without missing data; the right panel lists the results for the setting with missing data.
CC indicates complete case; MI, multiple imputation; MO, multiple overimputation.

estimates from the different methods by means of Wilcoxon tests leads to rejection of the null hypotheses of identical distributions, confirming the shift of multiple overimputation estimates toward the true parameter.

The MSE is smaller for multiple overimputation when compared with a naive or complete case analysis and multiple imputation, but the variance is larger (Table). However, the success of the three methods with respect to the MSE depends on the sample size as highlighted in Figure 1: the larger the sample size, the better the performance of multiple overimputation relative to the other methods. If the sample size is small, multiple overimputation does not outperform the two other methods with respect to the MSE. The bias can always be reduced using multiple overimputation, regardless of the sample size (Figure 1).

Sensitivity analyses show that under a correctly specified measurement error variance, changing the assumptions to allow for a higher amount of missing data, a different missingness process, a larger amount of measurement error, a smaller effect of each variable, and the inclusion of more variables, yields similar conclusions (eFigure 2; http://links.lww.com/EDE/A933).

Illustrative Example: HIV-Treatment Data from IeDEA-SA

We used data from the International epidemiological Databases Southern Africa cohort collaboration (IeDEA-SA) to illustrate the practical application of multiple overimputation in HIV treatment data. IeDEA-SA is a collaboration of 19 mostly programmatic cohorts in five southern African countries.[36] Data were collected at each site as part of routine monitoring and were transferred to the coordinating data centre at the University of Cape Town, South Africa. All contributing facilities obtained ethical approval from the institutional review boards before submitting anonymized patient data to the collaboration.

We limited data to four South African cohorts as those were the only ones with routinely assayed viral loads. Our dataset contained data on nearly 30,000 patients, initiating HAART between 1 January 2001 and 1 January 2010; all were followed from the time of first starting HAART (baseline).

Multiple overimputation ($M = 10$) was implemented using the "amelia" function of the *R*-package Amelia II.[28] The (over)imputation model included the mortality outcome, time to event or censoring, cohort, sex, age, year of HAART initiation, baseline ln CD4, and baseline $\log_{10}$ viral load. Our prior knowledge about the measurement error process was specified by means of the "priors" and "overimp" options of the amelia function, adding a prior normal distribution to each measured ln CD4 count and $\log_{10}$ viral load where the mean corresponded to the mismeasured value and the measurement error variance was set to 0.26^2 and 0.255^2, respectively. In sensitivity analyses, the measurement error variance was specified as 0.20^2 and 0.30^2 for CD4 and as 0.15^2 and 0.31^2 for viral load.

We used the Cox proportional hazards model to estimate the association of baseline CD4 count, baseline $\log_{10}$ viral load, year of treatment initiation, sex, cohort, and age with the hazard of death, based on the 10 overimputed datasets and applying multiple imputation-combining rules. Baseline CD4 count and baseline log viral load were included in the model first by categorizing the variables and, second, nonlinearly via p-splines.[37] In addition, as a reference, results from multiple imputation and a complete case analysis were estimated.

This example shows how regression estimates of CD4 count and viral load can vary depending on whether missing data and measurement error are taken into account. We have therefore excluded other variables with high missingness percentages (hemoglobin, WHO stage, creatinine, platelets) and under-reporting (tuberculosis, cryptococcal meningitis, among others) to ensure that comparisons between the different methodological approaches are not complicated by the missingness and measurement error related to these variables.

We also performed a similar analysis for the same data with time-updated CD4 counts and viral loads being included. If a patient did not have a CD4 count/viral load measurement for 6 months, then the respective values were

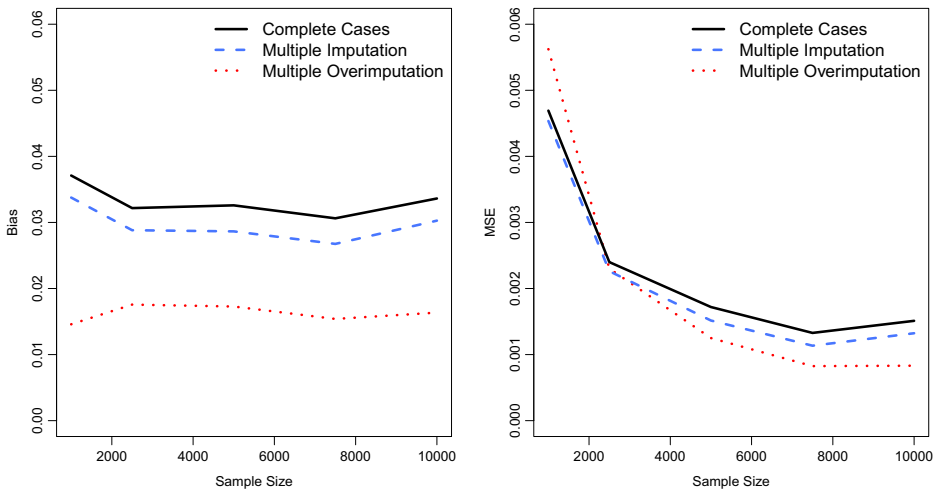


FIGURE 1. Results of the simulation studies: estimated bias and MSE of β_1 depending on the sample size, in the setting where the data are missing at random.

treated as missing. The imputation model included the same variables as above, time-updated CD4 and virologic suppression (viral load <1,000 copies/μl), and took the longitudinal structure of the data into account. The prior information for time-updated CD4 count was specified as for baseline CD4 count. Time-updated virologic suppression was preferred over time-updated viral load due to the bimodal distribution of the latter; the corresponding measurement error variance was assumed to be 0.23^2 which related to an assumption of about 1.5% misclassification (which we assume based on simulated viral loads similar to our data, see eText 2; http://links.lww.com/EDE/A933).

To address the fact that patients lost to follow-up are more likely to die, we linked lost patients to the national South African vital registry to obtain the vital status of these patients. The linkage was performed by a trusted third party, the South African Medical Research Council. Lost patients with recorded IDs could thus be linked (and their outcome ascertained and corrected); these patients were upweighted to represent all patients lost to follow-up: we took the inverse of the modeled probability of having an ID, based on a logistic regression model including age, sex, year of treatment initiation and cohort, to account for any differences between linkable and other patients lost to follow-up (patients with and without ID are known to be very similar though there are typically differences by cohort and year[38,39]). Patients not lost to follow-up received a weight of one, while those lost to follow-up and not linkable received a weight of zero.[39] Alternatively, missing outcome data of patients lost to follow-up could have been imputed with multiple overimputation.

From the 29,256 patients included in our analysis more than 10% had a missing baseline CD4 count and more than 60% had a missing baseline viral load. Median follow-up time (1st; 3rd quartile) was 498 (197; 878) days. The EMB algorithm utilizing multiple overimputation converged successfully for both the cross-sectional and longitudinal data examples.

The results of the Cox regression analysis are presented in Figure 2 and eTable 1 (http://links.lww.com/EDE/A933). Multiple overimputation emphasizes more strongly the relation between a high baseline CD4 count and a decreased hazard of death compared with the complete case analysis and multiple imputation (hazard ratio for CD4 > 200 cells/mm^3 vs. CD4 < 25 cells/mm^3: 0.21 [95% confidence interval: 0.18, 0.24] vs. 0.38 [0.29, 0.48] for the complete case analysis and 0.29 [0.25, 0.34] for multiple imputation). Looking at the nonlinear association of CD4 count with the hazard of death, or excluding baseline viral load from the analysis, or adding additional variables leads to the same conclusions (Figure 2A, eTables 2 and 3; http://links.lww.com/EDE/A933): the higher the CD4 count, the lower the hazard of death; similarly, the larger the number of viral copies the

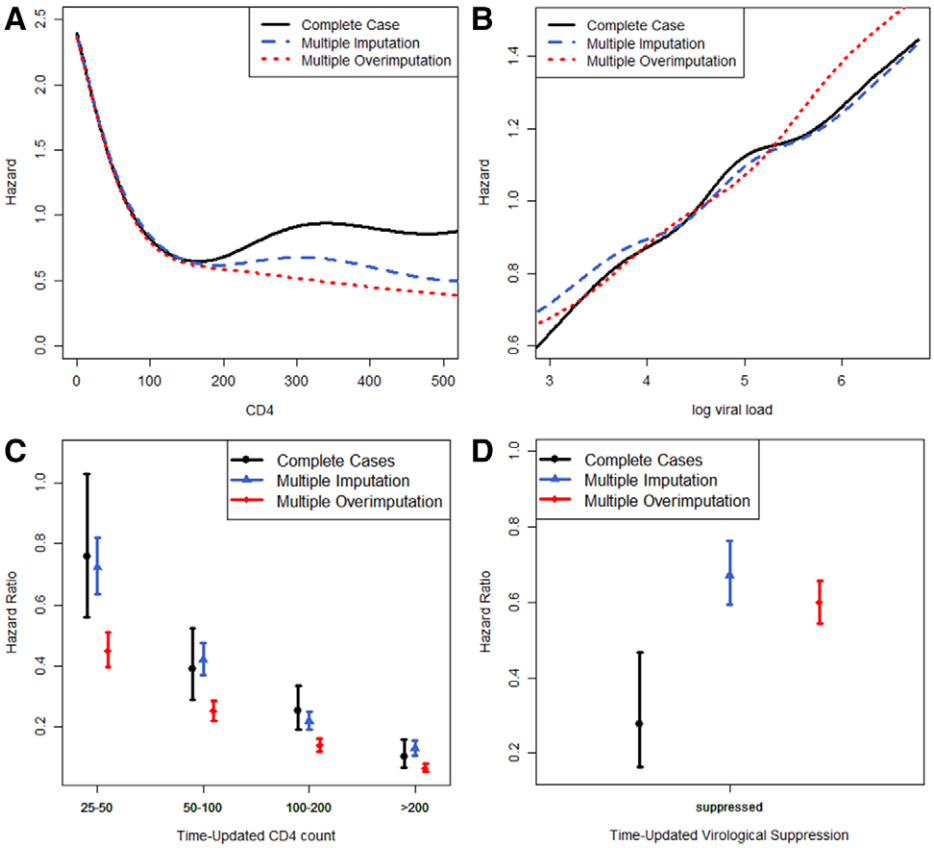


FIGURE 2. Nonlinear association of (A) baseline CD4 and (B) baseline $\log_{10}$ viral load with the hazard of death, modeled via p-splines. The estimates of (C) categorical time-updated CD4 (reference category: <25 cells/mm^3) and (D) categorical time-updated virological suppression (reference category: unsuppressed) are obtained from a Cox model fitted onto the longitudinal data. Results are reported for a complete case analysis, multiple imputation, and multiple overimputation and relate to the illustrative data example. The intervals reported in (C) and (D) are 95% confidence intervals.

greater the risk of death, which is more pronounced for multiple overimputation when compared with multiple imputation (Figure 2B). The sensitivity analyses show that different assumptions about the measurement error variance yield an almost identical nonlinear association of CD4 count with mortality after applying multiple overimputation. Assuming a much smaller measurement error variance for $\log_{10}$ viral load yields similar results for multiple overimputation and multiple imputation (eFigure 1; http://links.lww.com/EDE/A933). Regression coefficients of covariates without measurement error and missing data did not vary much between the three different approaches.

The longitudinal Cox regression analysis (Figure 2C, D, eTable 4; http://links.lww.com/EDE/A933) yields attenuated estimates for CD4 count in the complete case analysis and multiple imputation compared with multiple overimputation (HR for CD4 > 200 cells/mm^3 vs. CD4 < 25 cells/mm^3: 0.10 and 0.13 vs. 0.06); this analysis also shows that a complete case analysis yields very different results from the imputation approaches when assessing the estimates of viral suppression (HR 0.28 vs. 0.67 and 0.60).

In both analyses, the confidence intervals for all point estimates of multiple imputation and multiple overimputation were similarly wide (eTables 1 and 3; http://links.lww.com/EDE/A933). It can also be seen that in all analyses the correction made for measurement error was at least as great as the correction made to account for missing data (Figure 2, eTables 1–3; http://links.lww.com/EDE/A933).

The above analyses demonstrate that multiple overimputation for both baseline and follow-up CD4 count and viral load data can be easily incorporated into existing software (Amelia II for R), that the overimputation algorithm converges successfully for this data, that results may vary depending on whether one adjusts for missing data and measurement error or not, and that attenuation due to measurement error can occur, but this may not always be the case.

DISCUSSION

Statement of Principal Findings

We have demonstrated that multiple overimputation offers a convenient approach to address both measurement error and missing data and can be implemented easily for a variety of situations relevant to HIV research. Our simulation studies suggest that this approach is able to reduce bias and MSE in the context of survival analyses.

Strengths of the Study

This is, to the best of our knowledge, the first attempt to address simultaneously the treatment of missing data and measurement error in HIV research under a general framework. It is fast and easy to implement and, after applying multiple overimputation, many estimators relevant to HIV research can be obtained: for example the Kaplan–Meier estimator, and estimates from survival models such as the Cox proportional hazards model and parametric survival models, among many others. We have demonstrated that existing clinical knowledge about the accuracy of CD4 measurements can be used to specify the measurement error process, model and variance, which we have shown to be crucial for the success of method; moreover, our simulations highlight that not only for generalized linear models (as partially investigated by Blackwell et al.[19]) but also in survival analyses multiple overimputation can be successful.

Limitations

There remain, however, some limitations: as with multiple imputation, multiple overimputation cannot necessarily address situations where data are mismeasured not at random because the overimputed values drawn with the EMB algorithm may not properly reflect the joint distribution of both the data and the missingness/mismeasurement process (eText 1, formula [1]; http://links.lww.com/EDE/A933). In this case, the application of multiple overimputation can lead to biased estimates. Using a complete case analysis (and possibly correcting for measurement error in the respective sample) can also yield biased estimates in this setting, i.e., when the probability of a missing value depends on the outcome or external variables.[31,40] However, if the probability of missingness depends on the unobserved values of the variable itself, a complete case analysis still yields valid inference and may be preferable to multiple overimputation.[40] As we have argued above, in many cases, we would expect CD4 and viral load data to be mismeasured at random; however, time-updated viral load may be missing not at random if unobserved treatment interruptions due to nonadherence predict missingness.

We also have assumed that a successful specification of the imputation model is straightforward. The implementation of multiple overimputation is closely related to the joint modeling approach of Amelia II and thus natural constraints relate to specifying suitable transformations for skewed variables, additional imputation uncertainty with respect to categorical variables, and restrictions regarding complex longitudinal data.[28,41] An inappropriate imputation model or incorrect assumptions about the measurement error process can potentially cause multiple overimputation to be inferior compared with naive estimators.

Meaning of the Results

Our results suggest that regression estimates related to true CD4 count and true viral load may be biased in many studies. Both markers are a cornerstone in HIV research and thus it may be advisable to consider accounting for error in their measurement. Our data example illustrates how the application of multiple overimputation can change regression estimates: for example, the association of CD4 count with the hazard of death was more strongly pronounced under multiple overimputation compared with the approaches which neglected measurement error. This does *not*, however, imply that for any regression analysis the estimates of a complete

case analysis are biased toward the null, nor does it imply that corrected estimates of CD4 count are necessarily better just because their estimates yield stronger associations than naive analyses.

While our data example is illustrative and descriptive in nature, and none of the reported regression coefficients report a causal relation, there are several applications for which our findings are of interest. For example, predictive models can be used to inform mathematical modeling studies that require mortality rates stratified by true CD4 count and viral load.[42,43] It is evident from both our simulation studies and the data example that adjusting for missing data and measurement error can yield different predicted mortality rates; indeed, fitting the predictive model of May et al.[44] to our data shows that the differences between multiple overimputation and naive approaches found in our illustrative example persist in this context (eTable 5; http://links.lww.com/EDE/A933).

Our numerical investigations confirm previous studies showing that even a moderate amount of measurement error and/or missing data can cause bias in regression estimates.[9,45] Multiple overimputation reduced the bias in these estimates and also improved the MSE if the sample size was not too small. The latter observation implies that multiple overimputation may yield estimators with a higher variance compared with a naive analysis (reflecting the underlying uncertainty) and the success with respect to the MSE depends on the sample size. This is in line with the literature on measurement error correction in the case of complete data.[9,45,46] Generally, multiple overimputation yields asymptotically unbiased estimates under the mismeasured at random assumption (eText 1; http://links.lww.com/EDE/A933) given an appropriate imputation model is used, but its performance may vary from context to context.

Research in Context

Methods dealing with measurement error in CD4 count have already been suggested for particular applications and models under the assumption of no missing data.[10–18] One could think of applying these methods in the appropriate context in conjunction with multiple imputation. Since these methods are often very specific, combining the more general simulation extrapolation method[47] with multiple imputation might be a fruitful alternative. An implementation of simulation extrapolation in the statistical software R[48] is already available for (generalized) linear models, allows for both homoscedastic and heteroscedastic measurement error and can be naturally combined with existing multiple imputation procedures in *R*[28]; similar implementations are available for Stata.[49] However, it is an open question whether the imputations generated from mismeasured data yield valid inference or not.

It also remains important to check model assumptions after applying multiple overimputation: for example, when assessing the proportional hazards assumption of a Cox proportional hazards model, one may evaluate graphical diagnostics in each overimputed dataset; or, alternatively, the estimates of an interaction of analysis time with the covariate of interest can be easily combined by means of Rubin's rules.

We have concentrated our investigations on measurement error and missing data in CD4 counts and viral load measurements. There are, however, many more variables prone to measurement error and missing data and relevant in HIV research: CD4 percentage, hemoglobin, creatinine, p24 antigenemia, concentrations of antiretroviral drugs, among others. Existing knowledge can be used to account for both measurement error and missing data in many of these variables.[1,2,8] We stress, however, that an overestimated measurement error variance may yield biased estimates when applying multiple overimputation, see eFigure2 (http://links.lww.com/EDE/A933) and Blackwell et al.[19] Complicated measurement processes such as in pharmacokinetics, where metabolism, concomitant medication, and genetic factors influence measurement error, may, however, require special care and knowledge.

CONCLUSION

Our investigations show that multiple overimputation is a convenient and possibly promising approach to account for both missing and mismeasured data in HIV research. Further studies are needed to explore the implications, feasibility, and challenges of multiple overimputation for other models and applications.

ACKNOWLEDGMENTS

We thank Reneè de Waal, Graeme Meintjes, and Craig Corcoran for their helpful comments regarding the clinical meaning, quantification, and interpretation of measurement error in CD4 count and viral load.

REFERENCES

1. Raboud JM, Haley L, Montaner JS, Murphy C, Januszewska M, Schechter MT. Quantification of the variation due to laboratory and physiologic sources in CD4 lymphocyte counts of clinically stable HIV-infected individuals. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1995;10(Suppl 2):S67–S73.
2. Raboud JM, Montaner JS, Conway B, et al. Variation in plasma RNA levels, CD4 cell counts, and p24 antigen levels in clinically stable men with human immunodeficiency virus infection. *J Infect Dis*. 1996;174:191–194.
3. Hoover DR, Graham NM, Chen B, et al. Effect of CD4+ cell count measurement variability on staging HIV-1 infection. *J Acquir Immune Defic Syndr*. 1992;5:794–802.
4. Yu LM, Easterbrook PJ, Marshall T. Relationship between CD4 count and CD4% in HIV-infected people. *Int J Epidemiol*. 1997;26:1367–1372.
5. Best SJ, Gust AP, Johnson EI, McGavin CH, Dax EM. Quality of human immunodeficiency virus viral load testing in Australia. *J Clin Microbiol*. 2000;38:4015–4020.
6. Lew J, Reichelderfer P, Fowler M, et al. Determinations of levels of human immunodeficiency virus type 1 RNA in plasma: reassessment of parameters affecting assay outcome. TUBE Meeting Workshop Attendees. Technology Utilization for HIV-1 Blood Evaluation and Standardization in Pediatrics. *J Clin Microbiol*. 1998;36:1471–1479.
7. van Rensburg EJ, Tait K, Watt A, Schall R. Comparative evaluation of the Roche Cobas AmpliPrep/Cobas TaqMan HIV-1 version 2 test using

the TaqMan 48 analyzer and the Abbott RealTime HIV-1 assay. *J Clin Microbiol*. 2011;49:377–379.
8. Malone JL, Simms TE, Gray GC, Wagner KF, Burge JR, Burke DS. Sources of variability in repeated T-helper lymphocyte counts from human immunodeficiency virus type 1-infected patients: total lymphocyte count fluctuations and diurnal cycle are important. *J Acquir Immune Defic Syndr*. 1990;3:144–151.
9. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu C. Measurement *E*rror in *N*onlinear *Models: A Modern Perspective*. 2nd ed. Boca Raton, FL; London: Chapman & Hall/CRC; 2006.
10. Song X, Ma S. Multiple augmentation for interval-censored data with measurement error. *Stat Med*. 2008;27:3178–3190.
11. Bycott P, Taylor J. A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Stat Med*. 1998;17:2061–2077.
12. Tsiatis AA, Degruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error - applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc*. 1995;90:27–37.
13. Cole SR, Jacobson LP, Tien PC, Kingsley L, Chmiel JS, Anastos K. Using marginal structural measurement-error models to estimate the long-term effect of antiretroviral therapy on incident AIDS or death. *Am J Epidemiol*. 2010;171:113–122.
14. Huang YJ, Wang CY. Cox regression with accurate covariates unascertainable: a nonparametric-correction approach. *J Am Stat Assoc*. 2000;95:1209–1219.
15. Liu W, Wu L. A semiparametric nonlinear mixed-effects model with non-ignorable missing data and measurement errors for HIV viral data. *Comput Stat Data An*. 2008;53:112–122.
16. Nakamura T. Proportional hazards model with covariates subject to measurement error. *Biometrics*. 1992;48:829–838.
17. Song X, Davidian M, Tsiatis AA. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*. 2002;3:511–528.
18. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997;53:330–339.
19. Blackwell M, Honaker J, King G. A unified approach to measurement error and missing data: overview and details and extensions. *Sociol Methods Res*. 2015. In press. Available at: http://j.mp/jqdj72. doi: 10.1177/0049124115585360 and doi:10.1177/0049124115589052
20. Guolo A. Robust techniques for measurement error correction: a review. *Stat Methods Med Res*. 2008;17:555–580.
21. Stefanski LA. Measurement error models. *J Am Stat Assoc*. 2000;95:1353–1358.
22. Braitstein P, Brinkhof MW, Dabis F, et al.; Antiretroviral Therapy in Lower Income Countries (ART-LINC) Collaboration; ART Cohort Collaboration (ART-CC) groups. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet*. 2006;367:817–824.
23. Rosen S, Fox MP, Gill CJ. Patient retention in antiretroviral therapy programs in sub-Saharan Africa: a systematic review. *PLoS Med*. 2007;4:e298.
24. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat*. 2007;61:79–90.
25. Little R, Rubin D. *Statistical Analysis with Missing Data*. New York, NY: Wiley; 2002.
26. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91:473–489.
27. Honaker J, King G. What to do about missing values in time-series cross-section data? *Am J Polit Sci*. 2010;54:561–581.
28. Honaker J, King G, Blackwell M. Amelia II: a program for missing data. *J Stat Softw*. 2011;45:1–47.
29. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B*. 1977;39:1–38.
30. Goodnight JH. Tutorial on the sweep operator. *Am Stat*. 1979;33:149–158.
31. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods*. 2002;7:147–177.
32. Hughes MD, Stein DS, Gundacker HM, Valentine FT, Phair JP, Volberding PA. Within-subject variation in CD4 lymphocyte count in asymptomatic human immunodeficiency virus infection: implications for patient monitoring. *J Infect Dis*. 1994;169:28–36.
33. Yan J. Enjoy the joy of copulas: with package copula. *J Stat Softw*. 2007;21:1–21.
34. Trivedi P, Zimmer D. Copula modeling: an introduction for practitioners. *Found Trends Economet*. 2005;1:1–111.
35. Little RJA. Regression with missing Xs: a review. *J Am Stat Assoc*. 1992;87:1227–1237.
36. Egger M, Ekouevi DK, Williams C, et al. Cohort profile: the international epidemiological databases to evaluate AIDS (IeDEA) in sub-Saharan Africa. *Int J Epidemiol*. 2012;41:1256–1264.
37. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci*. 1996;11:89–102.
38. Boulle A, Schomaker M, May MT, et al. Mortality in patients with HIV-1 infection starting antiretroviral therapy in South Africa, Europe, or North America: a collaborative analysis of prospective studies. *PLoS Med*. 2014;11:e1001718.
39. Schomaker M, Gsponer T, Estill J, Fox M, Boulle A. Non-ignorable loss to follow-up: correcting mortality estimates based on additional outcome ascertainment. *Stat Med*. 2014;33:129–142.
40. Little RJ, Zhang NH. Subsample ignorable likelihood for regression analysis with missing data. *J R Stat Soc C-Appl*. 2011;60:591–605.
41. Schomaker M, Heumann C. Model selection and model averaging after multiple imputation. *Comput Stat Data An*. 2014;71:758–770.
42. Badri M, Cleary S, Maartens G, et al. When to initiate highly active antiretroviral therapy in sub-Saharan Africa? A South African cost-effectiveness study. *Antivir Ther*. 2006;11:63–72.
43. Goldie SJ, Yazdanpanah Y, Losina E, et al. Cost-effectiveness of HIV treatment in resource-poor settings–the case of Côte d'Ivoire. *N Engl J Med*. 2006;355:1141–1153.
44. May M, Boulle A, Phiri S, et al. Prognosis of patients with HIV-1 infection starting therapy in sub-Saharan Africa: a collaborative analysis of scale-up programmes. *Lancet*. 2010;376:449–457.
45. Stefanski LA. The effects of measurement error on parameter-estimation. *Biometrika*. 1985;72:583–592.
46. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. *Int J Epidemiol*. 2006;35:1074–1081.
47. Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. *J Am Stat Assoc*. 1994;89:1314–1328.
48. Lederer W, Kuechenhoff H. simex: SIMEX- and MCSIMEX-Algorithm for measurement error models. R package version 1–5. 2013.
49. Hardin J, Schmiediche H, Carroll RJ. The simulation extrapolation method for fitting generalized linear models with additive measurement error. *Stata J*. 2003;3:373–385.

Supplementary Material

(Schomaker,M., Hogger, S., Johnson, L., Hoffmann, C., Bärnighausen, T., Heumann, C., *Simultaneous Treatment of Missing Data and Measurement Error in HIV Research using Multiple Overimputation,* 2015, Epidemiology)

eTable 1: Mortality in South African patients after starting antiretroviral treatment. Cox regression estimates, reported as hazard ratios, for a complete case analysis, multiple imputation and multiple overimputation. 95% confidence intervals are reported in brackets. All results relate to the data from the illustrative example and should not be interpreted causally.

	Complete Cases	Multiple Imputation	Multiple Overimputation
Baseline CD4			
<25	1	1	1
25-50	0.73 [0.62;0.87]	0.74 [0.67;0.82]	0.63 [0.58;0.70]
50-100	0.47 [0.40;0.56]	0.49 [0.45;0.54]	0.47 [0.43;0.52]
100-200	0.33 [0.28;0.38]	0.33 [0.30;0.36]	0.34 [0.30;0.38]
>200	0.38 [0.29;0.48]	0.29 [0.25;0.34]	0.21 [0.18;0.24]
Baseline $\log_{10}$ viral load			
<4	1	1	1
4 to 5	1.29 [1.07;1.57]	1.18 [1.07;1.31]	1.20 [1.08;1.33]
5 to 6	1.54 [1.26;1.88]	1.41 [1.26;1.57]	1.42 [1.27;1.59]
>6	1.71 [1.27;2.29]	1.60 [1.34;1.91]	1.66 [1.37;2.01]
Sex			
Female	1	1	1
Male	1.34 [1.19;1.51]	1.31 [1.22;1.40]	1.31 [1.22;1.41]
Age			
<25	1	1	1
25-35	1.01 [0.81;1.26]	1.00 [0.87;1.16]	1.04 [0.90;1.20]
35-45	1.05 [0.83;1.32]	1.09 [0.94;1.26]	1.14 [0.99;1.32]
>45	1.22 [0.95;1.57]	1.36 [1.16;1.59]	1.37 [1.17;1.60]
Year			
before 2004	1	1	1
2004-2006	0.94 [0.75;1.18]	1.27 [1.06;1.50]	1.27 [1.06;1.50]
2007 and after	1.01 [0.76;1.34]	1.52 [1.25;1.84]	1.53 [1.26;1.85]
Cohort			
A	1	1	1
B	0.81 [0.67;0.97]	0.93 [0.83;1.04]	0.93 [0.83;1.04]
C	0.70 [0.44;1.13]	0.88 [0.76;1.00]	0.89 [0.78;1.02]
D	0.85 [0.71;1.01]	0.96 [0.88;1.06]	0.95 [0.86;1.04]

eTable 2: Mortality in South African patients after starting antiretroviral treatment. Estimates from a Cox regression model, reported as hazard ratios, if baseline viral load was not included in the analysis. Results are reported for a complete case analysis, multiple imputation and multiple overimputation. 95% confidence intervals are reported in brackets. All results relate to the data from the illustrative example and should not be interpreted causally.

	Complete Cases	Multiple Imputation	Multiple Overimputation
Baseline CD4			
<25	1	1	1
25-50	0.75 [0.63;0.89]	0.73 [0.66;0.81]	0.62 [0.57;0.68]
50-100	0.47 [0.40;0.56]	0.48 [0.44;0.53]	0.46 [0.42;0.50]
100-200	0.32 [0.27;0.37]	0.32 [0.29;0.35]	0.33 [0.29;0.36]
>200	0.36 [0.28;0.46]	0.28 [0.24;0.33]	0.20 [0.17;0.23]
Sex			
Female	1	1	1
Male	1.36 [1.21;1.53]	1.33 [1.24;1.42]	1.33 [1.24;1.43]
Age			
<25	1	1	1
25-35	1.00 [0.80;1.25]	1.00 [0.87;1.15]	1.04 [0.90;1.21]
35-45	1.04 [0.83;1.31]	1.09 [0.94;1.26]	1.15 [1.00;1.33]
>45	1.22 [0.95;1.57]	1.36 [1.16;1.59]	1.39 [1.19;1.62]
Year			
before 2004	1	1	1
2004-2006	0.96 [0.76;1.20]	1.28 [1.07;1.52]	1.27 [1.07;1.52]
2007 and after	0.99 [0.74;1.31]	1.51 [1.24;1.83]	1.52 [1.25;1.85]
Cohort			
A	1	1	1
B	0.86 [0.72;1.03]	0.98 [0.88;1.10]	0.99 [0.88;1.11]
C	0.73 [0.45;1.16]	0.89 [0.77;1.02]	0.90 [0.78;1.03]
D	0.81 [0.69;0.95]	0.91 [0.83;1.00]	0.90 [0.82;0.99]

eTable 3: Mortality in South African patients after starting ART. Estimates from a Cox regression model, reported as hazard ratios, if baseline TB, WHO stage, and haemoglobin are added to the analysis. Results are reported for a complete case analysis, multiple imputation and multiple overimputation. 95% CI's are reported in brackets. All results relate to the data from the illustrative example and should not be interpreted causally.

	Complete Cases	Multiple Imputation	Multiple Overimputation
Baseline CD4			
<25	1	1	1
25-50	0.74 [0.58;0.94]	0.80 [0.72;0.88]	0.68 [0.62;0.74]
50-100	0.5 [0.39;0.63]	0.56 [0.51;0.62]	0.50 [0.45;0.55]
100-200	0.48 [0.39;0.6]	0.41 [0.37;0.45]	0.37 [0.33;0.42]
>200	0.49 [0.35;0.7]	0.37 [0.32;0.44]	0.24 [0.21;0.28]
Baseline $\log_{10}$ viral load			
<4	1	1	1
4 to 5	1.43 [1.07;1.91]	1.11 [1.01;1.23]	1.14 [1.03;1.27]
5 to 6	1.53 [1.14;2.06]	1.24 [1.11;1.38]	1.26 [1.13;1.42]
>6	1.28 [0.85;1.94]	1.32 [1.11;1.58]	1.43 [1.18;1.73]
Prevalent TB			
no	1	1	1
yes	1.98 [1.38;2.83]	1.11 [0.96;1.28]	1.07 [0.93;1.24]
Baseline WHO stage			
I & II	1	1	1
III	2.45 [1.68;3.56]	1.4 [1.26;1.55]	1.34 [1.21;1.48]
IV	3.40 [2.27;5.1]	1.85 [1.67;2.05]	1.77 [1.60;1.96]
Hemoglobin			
per gm/dL	0.85 [0.82;0.88]	0.9 [0.89;0.91]	0.89 [0.88;0.9]
Sex			
Female	1	1	1
Male	1.34 [1.13;1.59]	1.44 [1.34;1.54]	1.44 [1.34;1.54]
Age			
<25	1	1	1
25-35	0.93 [0.7;1.24]	1 [0.87;1.16]	1.04 [0.91;1.21]
35-45	0.91 [0.68;1.24]	1.1 [0.95;1.27]	1.17 [1.01;1.36]
>45	1.14 [0.81;1.58]	1.39 [1.19;1.62]	1.44 [1.23;1.68]
Year			
before 2004	1	1	1
2004-2006	0.77 [0.56;1.08]	1.32 [1.11;1.57]	1.33 [1.12;1.58]
2007 and after	0.77 [0.51;1.16]	1.56 [1.29;1.89]	1.61 [1.32;1.95]
Cohort			
A	1	1	1
B	0.59 [0.45;0.76]	0.86 [0.77;0.97]	0.87 [0.77;0.98]
C	---[1]	0.87 [0.76;1]	0.88 [0.77;1.01]
D	---[1]	0.97 [0.88;1.07]	0.97 [0.88;1.07]

[1] Cohorts C and D are excluded in the complete case analysis because of missing WHO stage data

eTable 4: Mortality in South African patients after starting antiretroviral treatment. Cox regression estimates, reported as hazard ratios, based on the longitudinal data, stratified by cohort. Results are reported for a complete case analysis, multiple imputation and multiple overimputation. 95% confidence intervals are reported in brackets. All results relate to the data from the illustrative example and should not be interpreted causally.

	Complete Cases	Multiple Imputation	Multiple Overimputation
Time-updated CD4			
<25	1	1	1
25-50	0.76 [0.56;1.03]	0.72 [0.63;0.82]	0.45 [0.40;0.51]
50-100	0.39 [0.29;0.52]	0.42 [0.37;0.48]	0.25 [0.22;0.29]
100-200	0.25 [0.19;0.33]	0.22 [0.19;0.25]	0.14 [0.12;0.16]
>200	0.10 [0.07;0.16]	0.13 [0.11;0.15]	0.06 [0.05;0.08]
Time-updated virological sup.			
unsuppressed	1	1	1
suppressed	0.28 [0.16;0.47]	0.67 [0.59;0.76]	0.60 [0.55;0.66]
Sex			
Female	1	1	1
Male	1.39 [1.13;1.70]	1.21 [1.11;1.33]	1.19 [1.08;1.30]
Age			
<25	1	1	1
25-35	0.92 [0.61;1.36]	0.99 [0.83;1.19]	1.03 [0.85;1.23]
35-45	1.07 [0.72;1.61]	1.19 [0.91;1.32]	1.15 [0.95;1.39]
>45	1.31 [0.85;2.03]	1.45 [1.19;1.76]	1.43 [1.17;1.75]
Year			
before 2004	1	1	1
2004-2006	0.83 [0.58;1.19]	1.20 [0.99;1.45]	1.11 [0.91;1.36]
2007 and after	0.60 [0.38;0.96]	1.08 [0.86;1.35]	0.96 [0.76;1.20]

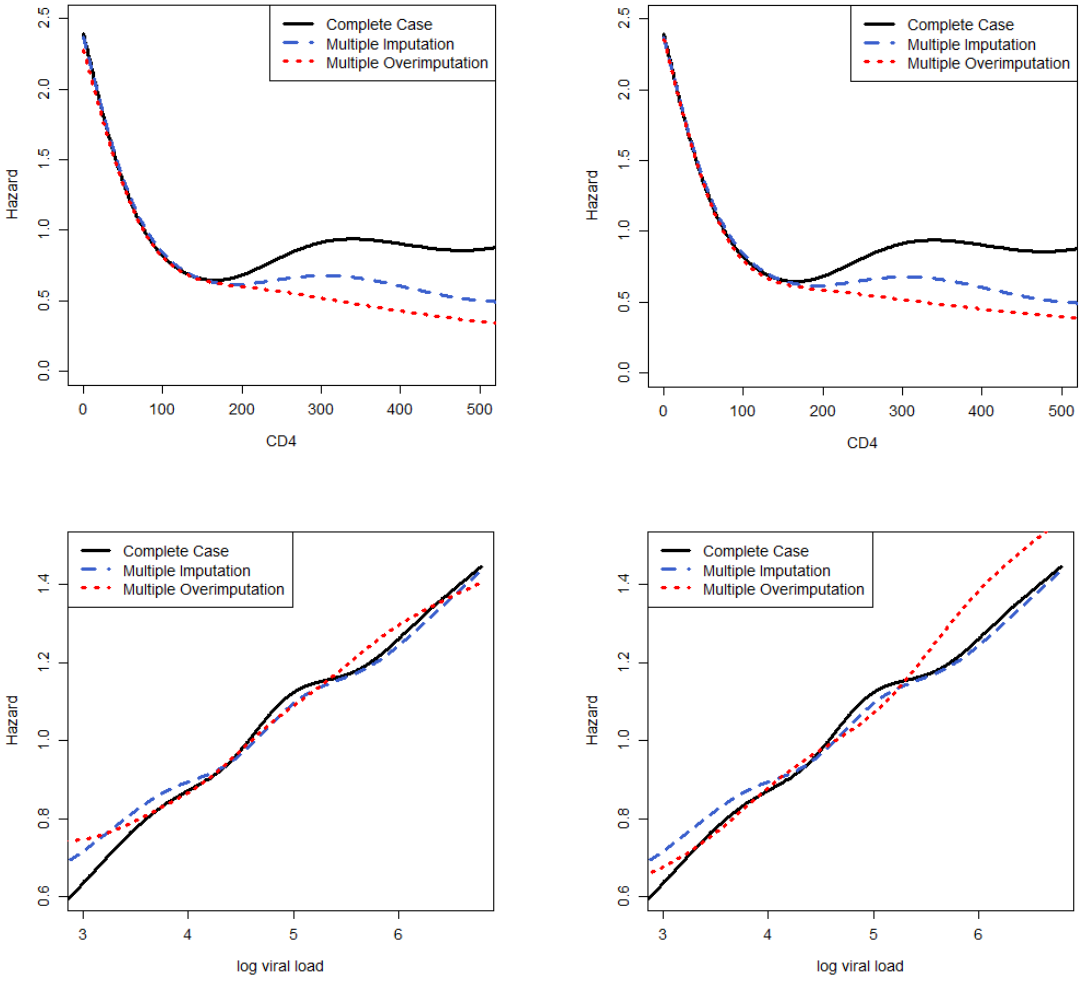
eTable 5: Mortality in South African patients after starting antiretroviral treatment. Estimates from a Cox regression model, reported as hazard ratios, if the variables and categorizations from the predictive model of May et al.[2] (developed in 3 sub-Saharan countries) are used. The analysis is restricted to one year on antiretroviral treatment. 95% confidence intervals are reported in brackets. All results relate to the data from the illustrative example and should not be interpreted causally.

	Complete Cases	Multiple Imputation	Multiple Overimputation	May et al.[2]
Baseline CD4				
<25	1	1	1	1
25-50	0.67 [0.51;0.88]	0.76 [0.68;0.85]	0.61 [0.55;0.67]	0.76 [0.62; 0.94]
50-100	0.39 [0.29;0.52]	0.51 [0.46;0.57]	0.43 [0.39;0.48]	0.46 [0.38; 0.57]
100-200	0.34 [0.26;0.43]	0.34 [0.30;0.38]	0.31 [0.27;0.35]	0.35 [0.28; 0.42]
>200	0.34 [0.22;0.52]	0.32 [0.26;0.38]	0.20 [0.17;0.24]	0.29 [0.22; 0.38]
Sex				
Male	1	1	1	1
Female	0.8 [0.66;0.98]	0.71 [0.66;0.77]	0.71 [0.66;0.77]	0.68 [0.58; 0.79]
Weight (in kg)				
<45	1	1	1	1
45-50	0.65 [0.49;0.87]	0.69 [0.61;0.78]	0.68 [0.61;0.77]	0.59 [0.48; 0.72]
50-60	0.33 [0.26;0.43]	0.41 [0.37;0.46]	0.39 [0.35;0.44]	0.40 [0.33; 0.48]
>60	0.24 [0.18;0.32]	0.30 [0.27;0.34]	0.29 [0.26;0.33]	0.24 [0.19; 0.30]
WHO stage				
I and II	1	1	1	1
III and IV	2.49 [1.58;3.92]	1.61 [1.44;1.81]	1.69 [1.51;1.9]	2.72 [1.87; 3.95]
Age (in years)				
<40	1	1	1	
>40	1.07 [0.86;1.33]	1.23 [1.13;1.34]	1.21 [1.11;1.32]	1.43 [1.23; 1.66]

[2] May M, Boulle A, Phiri S, et al. Prognosis of patients with HIV-1 infection starting therapy in sub-Saharan Africa: a collaborative analysis of scale-up programmes. Lancet. 2010;376:449-457.

eFigure 1: Results of the Cox regression analysis when using different assumptions about the measurement error variance: (a) $\sigma^2_{U_{ij}}=0.20^2$ for CD4 count and $\sigma^2_{U_{ij}}=0.15^2$ for log viral load (b) $\sigma^2_{U_{ij}}=0.30^2$ for CD4 count and $\sigma^2_{U_{ij}}=0.31^2$ for log viral load.

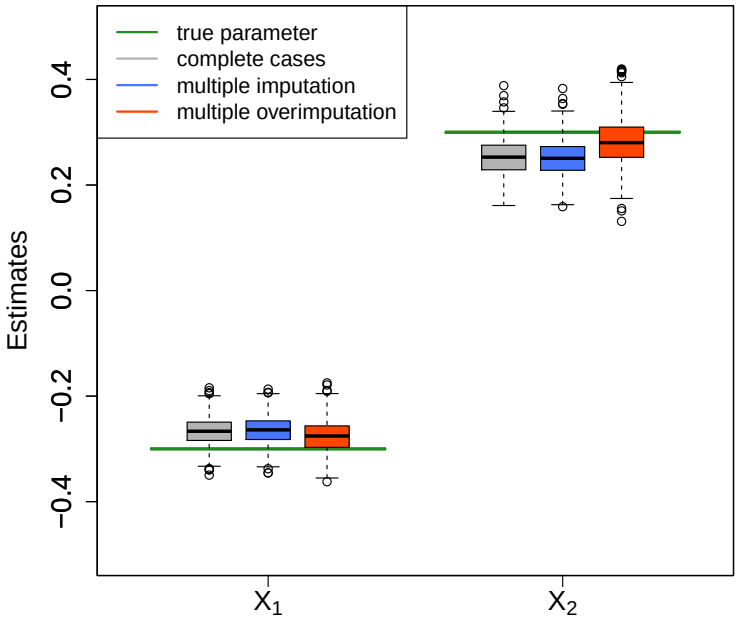
(a) ***(b)***



eFigure 2: Results of variations of the simulation study. The settings specified in the captions refer to changes compared to the main setting in the manuscript.

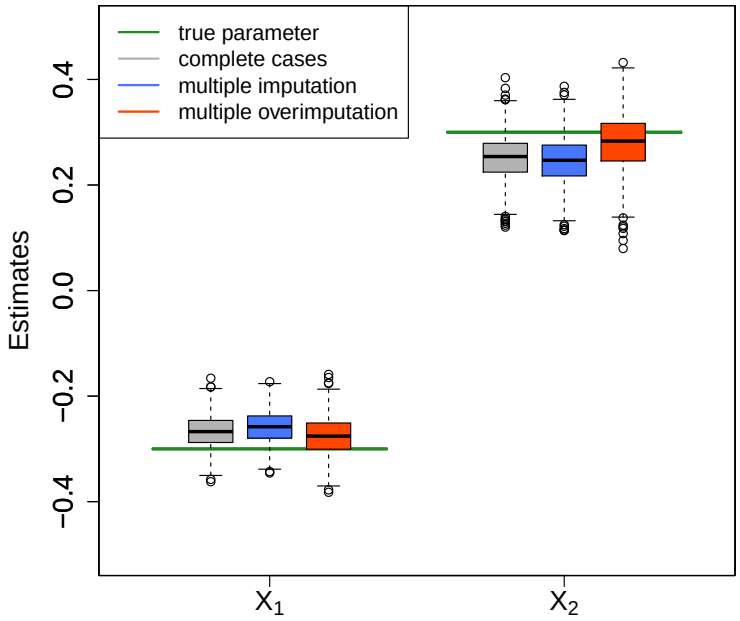
Variation of missingness assumption

(a) Data missing completely at random, with 10% missingness for both X_1 and X_2



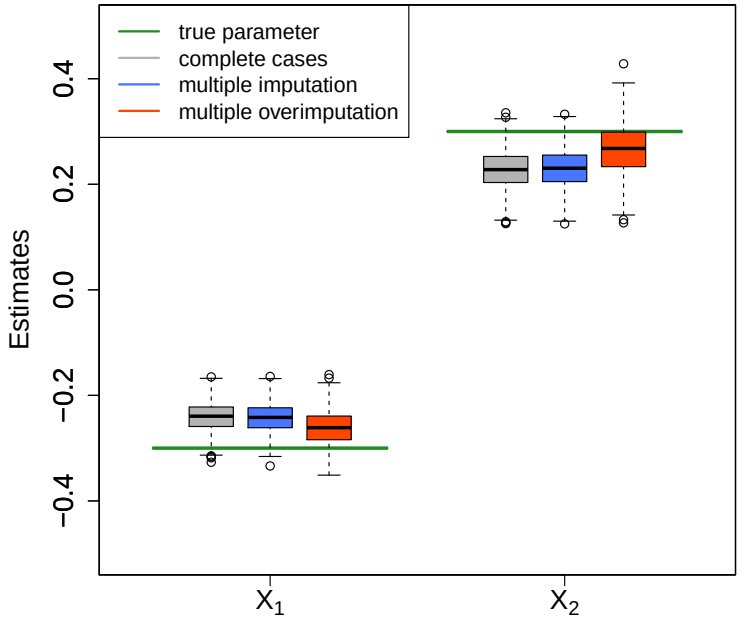
	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.033	-0.048	0.0018	0.0034
Multiple imputation	0.035	-0.049	0.0019	0.0036
Multiple overimputation	0.024	-0.019	0.0014	0.0023

(b) Data missing completely at random, with 40% missingness for both X_1 and X_2



	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.033	-0.048	0.0020	0.0040
Multiple imputation	0.042	-0.054	0.0028	0.0047
Multiple overimputation	0.024	-0.018	0.0019	0.0033

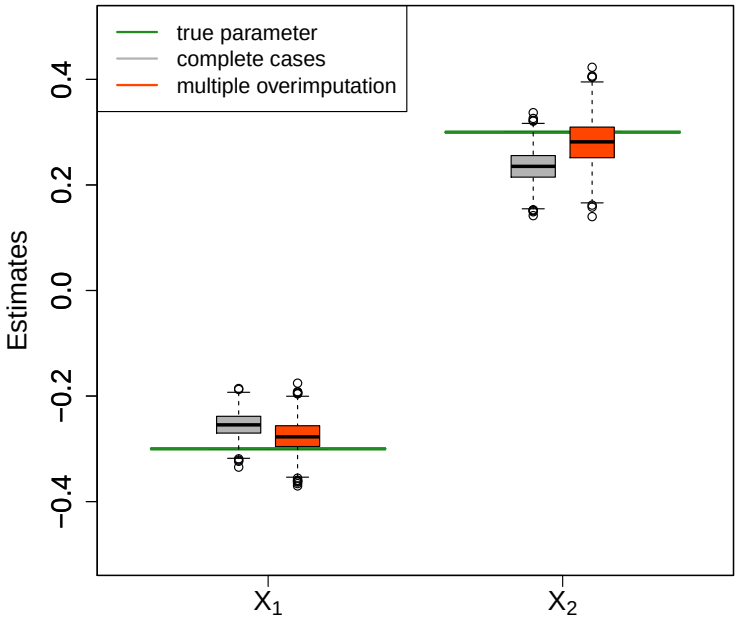
(c) Data missing at random, but with higher missingness probability, ca. 20% for both X1 and X2, defined via $\pi_X(T) = 1 - \{0.02T^2 + 1\}^{-1}$



	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.060	-0.073	0.0043	0.0065
Multiple imputation	0.058	-0.069	0.0041	0.0062
Multiple overimputation	0.038	-0.032	0.0025	0.0033

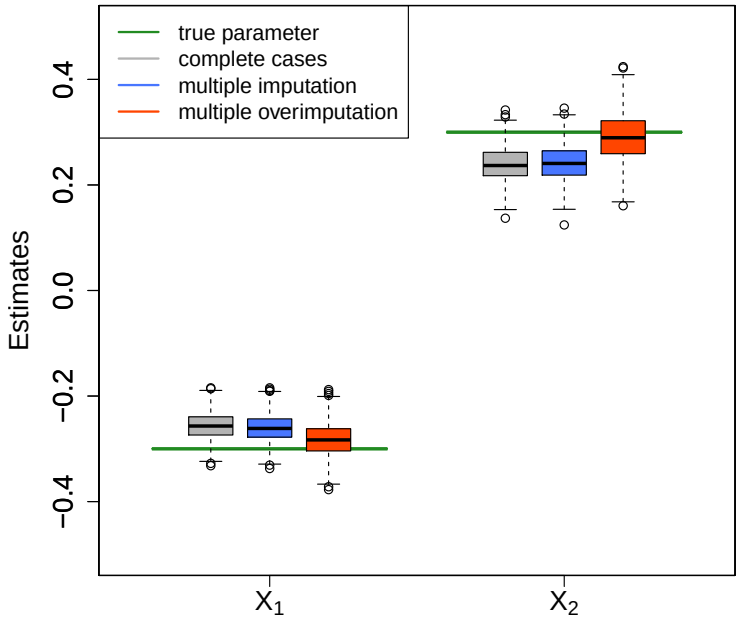
Variation of the measurement error variance

(d) Larger measurement error variance, 0.3^2 and 0.31^2 for both X_1 and X_2 respectively, no missing data



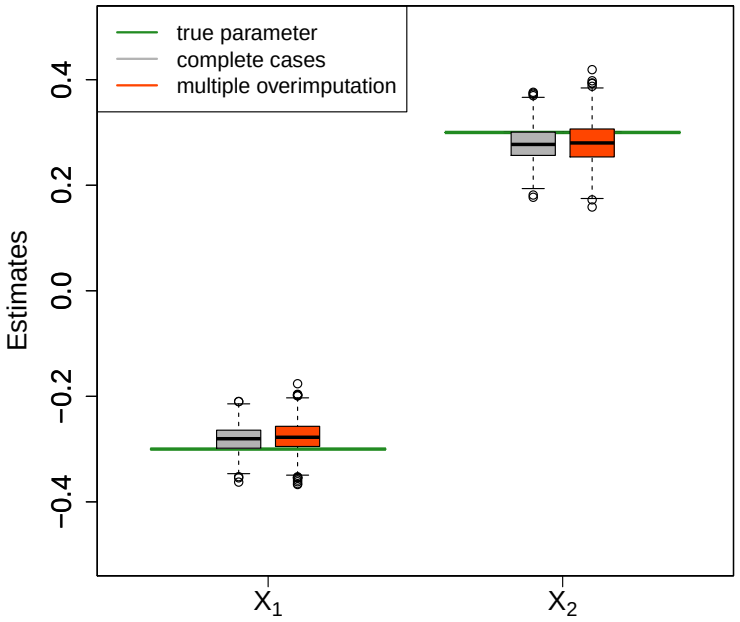
	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.045	-0.065	0.0026	0.0052
Multiple imputation	–	–	–	–
Multiple overimputation	0.023	-0.019	0.0015	0.0022

(e) Larger measurement error variance, 0.3^2 and 0.31^2 for both X_1 and X_2 respectively, with missing data



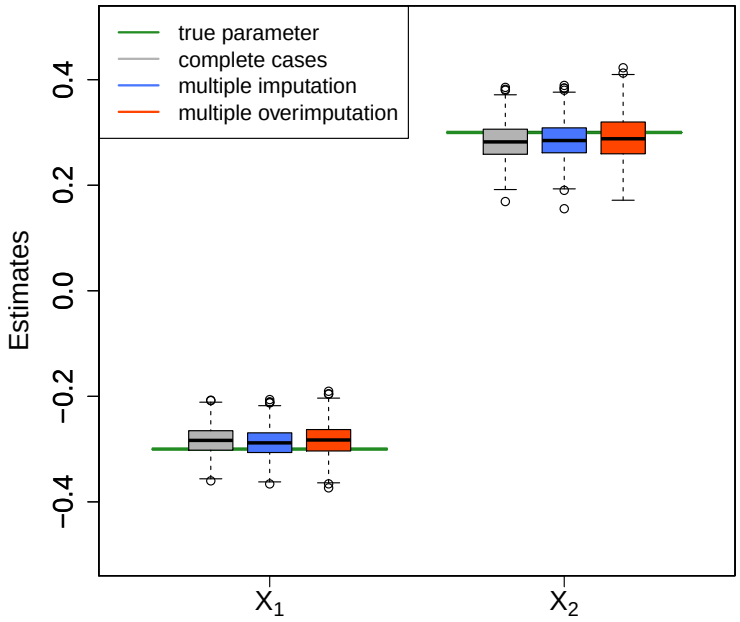
	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.043	-0.062	0.0025	0.0048
Multiple imputation	0.040	-0.059	0.0022	0.0045
Multiple overimputation	0.017	-0.009	0.0013	0.0022

(f) Smaller measurement error variance, 0.2^2 and 0.15^2 for both X_1 and X_2 respectively, no missing data



	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.019	-0.022	0.0010	0.0016
Multiple imputation	–	–	–	–
Multiple overimputation	0.023	-0.019	0.0014	0.0019

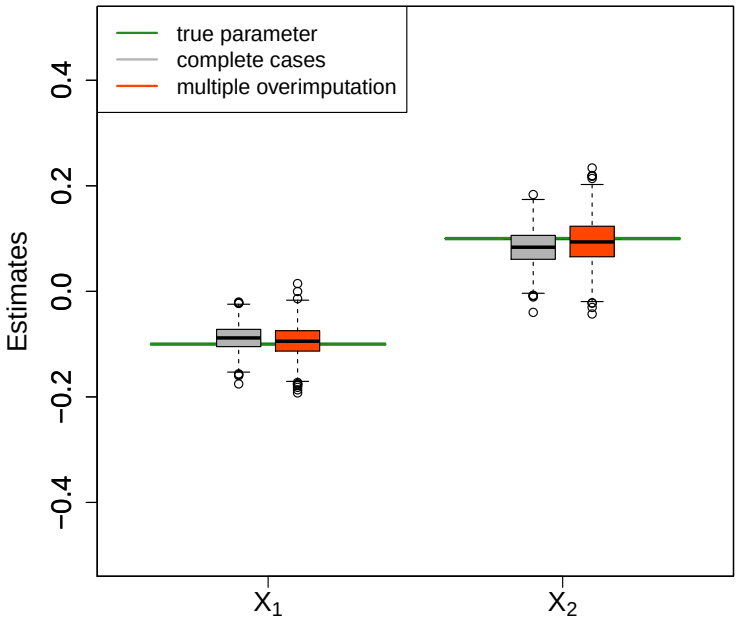
(g) Smaller measurement error variance, 0.2^2 and 0.15^2 for both X_1 and X_2 respectively, with missing data



	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.017	-0.018	0.0009	0.0015
Multiple imputation	0.012	-0.015	0.0009	0.0015
Multiple overimputation	0.017	-0.009	0.0012	0.0019

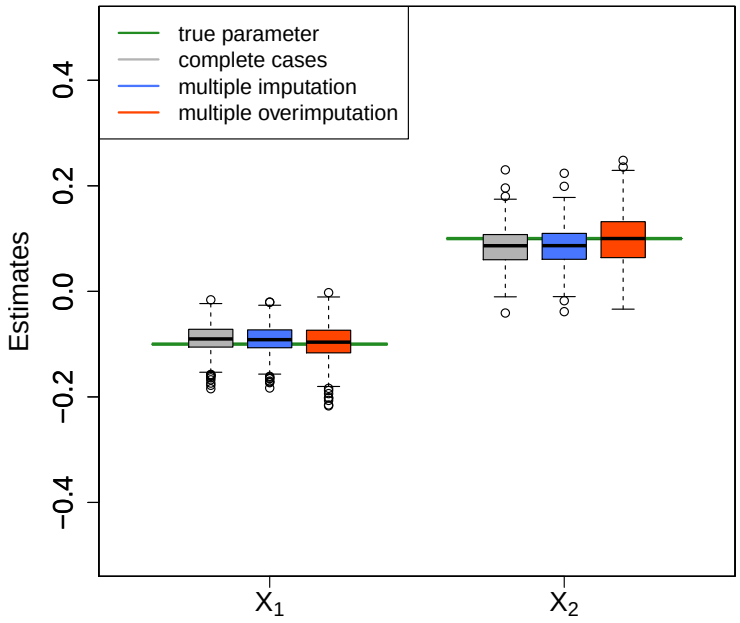
Variation of the linear predictor

(h) The linear predictor $X\beta$ is defined as $-0.1 \ln X_1 + 0.1 \log_{10} X_2$, no missing data



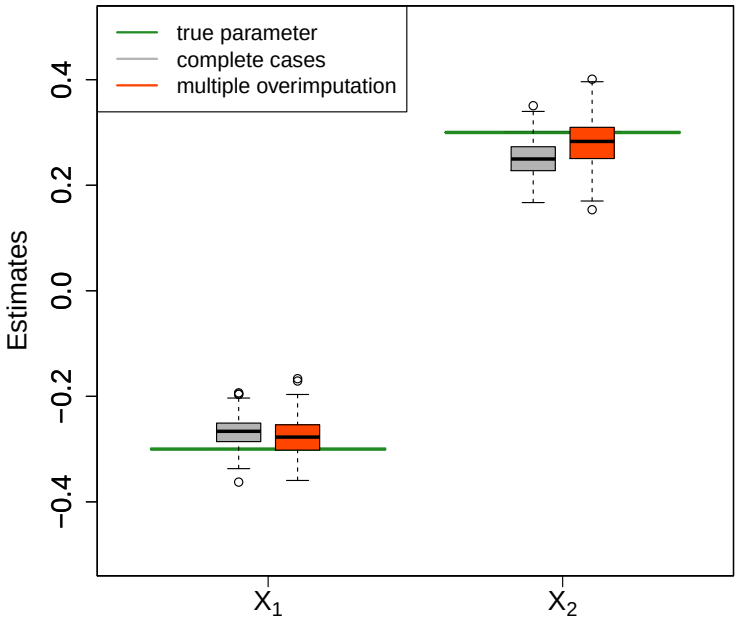
	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.012	-0.017	0.0007	0.0014
Multiple imputation	–	–	–	–
Multiple overimputation	0.005	-0.005	0.0009	0.0018

(i) The linear predictor $X\beta$ is defined as $-0.1 \ln X_1 + 0.1 \log_{10} X_2$, with missing data



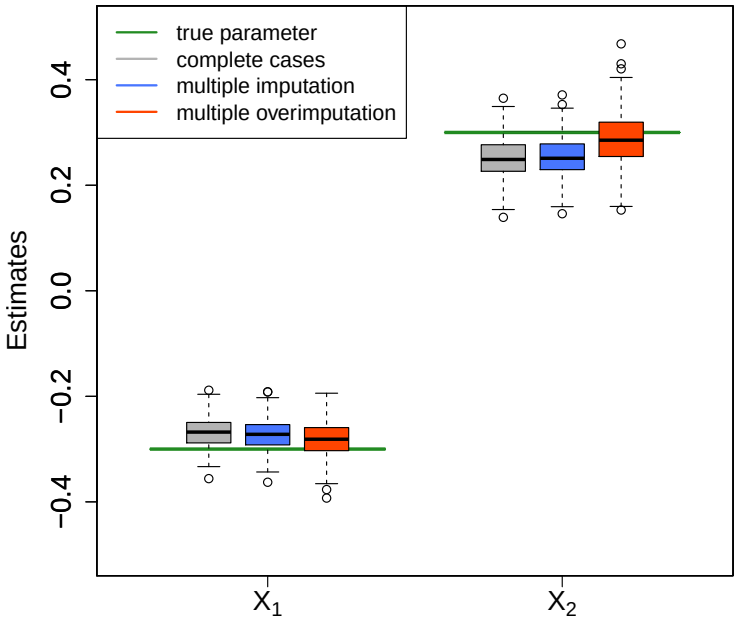
	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.011	-0.016	0.0008	0.0015
Multiple imputation	0.010	-0.015	0.0008	0.0015
Multiple overimputation	0.004	-0.001	0.0011	0.0023

(j) Additional 4 covariates: $X_3 \sim Binom(0.65)$, $X_4 \sim Weibull(1.75, 1.9)$, $X_5 \sim Exp(1)$, $X_6 \sim Gamma(0.25, 2)$, $\beta = (-0.3, 0.3, 0, 0, 0, 0)$, no missing data



	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.033	-0.050	0.0018	0.0035
Multiple imputation	–	–	–	–
Multiple overimputation	0.023	-0.018	0.0015	0.0023

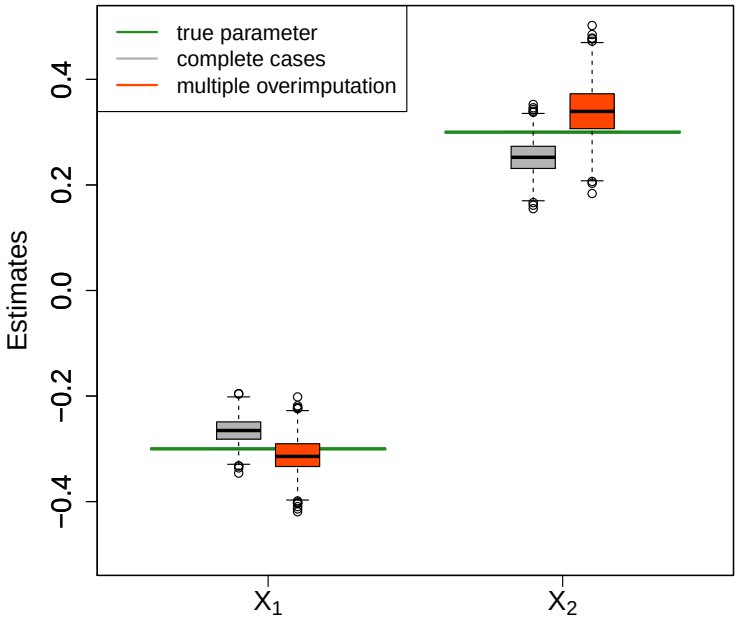
(k) Additional 4 covariates: $X_3 \sim Binom(0.65)$, $X_4 \sim Weibull(1.75, 1.9)$, $X_5 \sim Exp(1)$, $X_6 \sim Gamma(0.25, 2)$, $\beta = (-0.3, 0.3, 0, 0, 0, 0)$, with missing data



	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.031	-0.049	0.0018	0.0037
Multiple imputation	0.027	-0.047	0.0016	0.0035
Multiple overimputation	0.018	-0.014	0.0015	0.0027

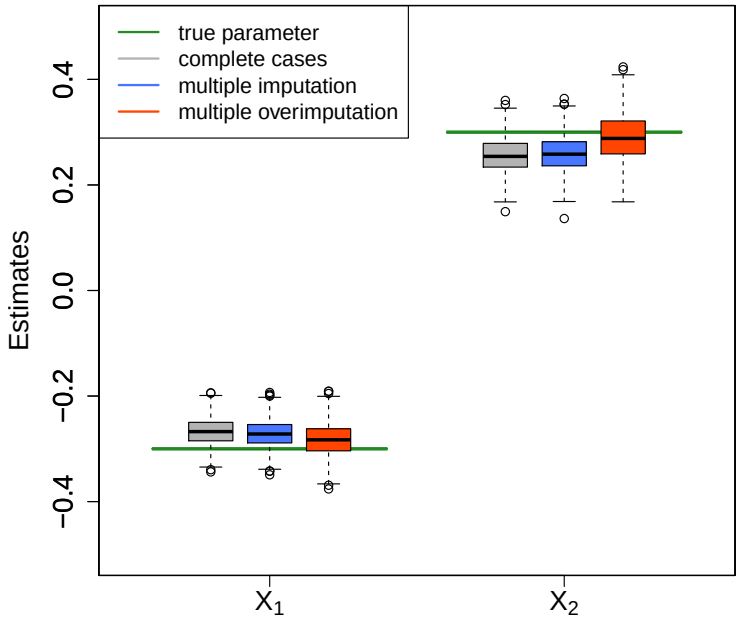
Wrong assumption for measurement error variance used

(l) Wrong assumption used for measurement error variance: 0.36^2 and 0.355^2 for X_1 and X_2 respectively, no missing data



	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.034	-0.048	0.0018	0.0033
Multiple imputation	–	–	–	–
Multiple overimputation	-0.013	0.040	0.0012	0.0039

(m) Wrong assumption used for measurement error variance: 0.36^2 and 0.355^2 for X_1 and X_2 respectively, with missing data



	Bias β_1	Bias β_2	MSE β_1	MSE β_2
Complete cases	0.033	-0.045	0.0017	0.0030
Multiple imputation	0.029	-0.042	0.0015	0.0028
Multiple overimputation	0.017	-0.009	0.0012	0.0021

eText 1: Outline of the technical background of multiple overimputation. More details can be found in Blackwell et al. (2015b), Blackwell et al. (2015a) and Honaker and King (2010). For a better understanding of the technical details the reader may also wish to consult Rubin (1996) for more insight on multiple imputation, King et al. (2001) for useful technicalities of an algorithm similar to EMB, Dempster et al. (1977) for the EM algorithm, and Goodnight (1979) for the details of the sweep operator.

1. *Data and notation:* Consider a data set $\mathbf{X}$ consisting of observations $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{ip})$. Let e_{ij} be an indicator whether x_{ij} was measured with error and m_{ij} be an indicator if x_{ij} is missing. The data may consist of perfectly measured values $\mathbf{x}_i^{obs} = \{x_{ij}; e_{ij} = m_{ij} = 0\}$, values which are missing, $\mathbf{x}_i^{mis} = \{x_{ij}; m_{ij} = 1\}$, and values measured with error (w_{ij}) as a proxy to the latent 'true' unobserved values x_{ij}^{err}, $\mathbf{x}_i^{err} = \{x_{ij}; e_{ij} = 1\}$, $\mathbf{w}_i = \{w_{ij}; e_{ij} = 1\}$. Thus, the observed data for any observation is $\mathbf{d}_i = (\mathbf{x}_i^{obs}, \mathbf{w}_i)$ while the true underlying data is $\mathbf{x}_i = (\mathbf{x}_i^{obs}, \mathbf{x}_i^{err}, \mathbf{x}_i^{mis})$.

2. *Observed data probability density function:* The probability density function for the *observed* data equates to

$$p(\mathbf{d}_i, \mathbf{m}_i, \mathbf{e}_i|\theta, \gamma, \phi) \quad = \quad \int\int p(\mathbf{x}_i|\theta)p(\mathbf{w}_i|\mathbf{x}_i, \gamma)p(\mathbf{m}_i, \mathbf{e}_i|\mathbf{d}_i, \mathbf{x}_i, \phi)\, d\mathbf{x}_i^{err} d\mathbf{x}_i^{mis} \tag{1}$$

   whereby θ refers to the parameterization of the true underlying data, γ to the error distribution, and ϕ to the joint distribution of $\mathbf{m}_i$ and $\mathbf{e}_i$. Using the mismeasured at random (MMAR) assumption, which is $p(\mathbf{m}_i, \mathbf{e}_i|\mathbf{d}_i, \mathbf{x}_i, \phi) = p(\mathbf{m}_i, \mathbf{e}_i|\mathbf{d}_i, \phi)$, (1) can be written as

$$p(\mathbf{d}_i, \mathbf{m}_i, \mathbf{e}_i|\theta, \gamma, \phi) \quad = \quad p(\mathbf{m}_i, \mathbf{e}_i|\mathbf{d}_i, \phi)p(\mathbf{d}_i|\theta, \gamma) \tag{2}$$

   which is proportional to

$$p(\mathbf{d}_i|\theta, \gamma) \quad = \quad \int\int p(\mathbf{x}_i|\theta)p(\mathbf{w}_i|\mathbf{x}_i, \gamma)\, d\mathbf{x}_i^{err} d\mathbf{x}_i^{mis}\,. \tag{3}$$

   Note that from a Bayesian perspective, for a given prior on (θ, γ), this gives us a posterior distribution for $p(\theta, \gamma|\mathbf{d}_i)$ which makes use of only observed quantities.

3. *Posterior predictive distribution of the unobserved data:* To obtain valid inference with multiple imputation (MI), one needs to draw from the posterior predictive distribution of the unobserved data. If one were to omit mismeasured data and thus define $x_{ij}^{err} = x_{ij}^{mis}$ MI would already yield valid inference but omit important information. Given that both missing *and* latent values are unobserved, draws from the predictive posterior distribution of this unobserved data relate to:

$$p(\mathbf{x}_i^{err}, \mathbf{x}_i^{mis}) = \int p(\mathbf{x}_i^{err}, \mathbf{x}_i^{mis}|\mathbf{d}_i, \theta, \gamma)p(\theta, \gamma|\mathbf{d}_i)d\theta d\gamma\,. \tag{4}$$

4. *Multiple imputation with EMB:* To draw values from (4) one needs to (i) draw $(\theta_{(i)}, \gamma_{(i)})$ from its posterior distribution $p(\theta, \gamma|\mathbf{d}_i)$ and then (ii) draw $(\mathbf{x}_i^{err}, \mathbf{x}_i^{mis})$ from $p(\mathbf{x}_i^{err}, \mathbf{x}_i^{mis}|\mathbf{d}_i, \theta, \gamma)$. The EMB algorithm utilizes this (i) by means of the EM algorithm to obtain an unbiased estimate $\hat{\theta}$ in the presence of unobserved data, and (ii) by repeating this for different bootstrap samples of $\mathbf{d}$. Specifically, under the assumption of a multivariate normal distribution for the data, $\mathbf{X} \sim N(\mu, \mathbf{\Sigma})$, and under the assumption of no measurement error, EMB does the following:

   i) The Expectation-Maximization (EM) algorithm estimates $\theta = (\mu, \mathbf{\Sigma})$ in the presence of unobserved data. In the E(xpectation)-Step the algorithm fills in estimates for the missing values using conditional expectations; in the M(aximazation)-Step the complete data parameters are estimated (from the sufficient statistics) using the available and filled-in data. These two steps are repeated until the parameter estimates converge and one obtains $(\hat{\mu}, \hat{\mathbf{\Sigma}})$, see Dempster et al. (1977) for the technical details. Thus, an estimate of θ can be drawn from $N(\hat{\mu}, \hat{\mathbf{\Sigma}})$. This step simulates draws from $p(\theta, \gamma|\mathbf{d}_i)$ related to (3).

   ii) The draws from $N(\hat{\mu}, \hat{\mathbf{\Sigma}})$ are used to obtain $\tilde{\beta}$ (an estimate of β) via the sweep operator and impute missing values via $x_{ij} = x_{i,-j}^{obs}\tilde{\beta} + \tilde{\epsilon}$. We refer the reader to Goodnight (1979) and the appendix of Honaker and King (2010) for the details on how $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$ relate to $\tilde{\beta}$.

iiii) Repeating this procedure for M bootstrap samples of $\mathbf{d}$ yields M different imputations adequately reflecting estimation uncertainty. They can be seen as draws from (4) for $\mathbf{x}_i^{err} = \emptyset$.

5) *Incorporating measurement error into EMB via prior distributions:* To simulate proper draws from (4) one needs to first simulate proper draws from the posterior relating to (3). Blackwell et al. (2015a) show that under the setting of (3) the EM-algorithm needs to estimate

$$E(T(\mathbf{x}_i)|\mathbf{d}_i, \theta^{(t)}, \gamma) \quad = \quad \int\int T(\mathbf{x}_i) \underbrace{p(\mathbf{x}_i^{err}, \mathbf{x}_i^{mis}|\mathbf{x}_i^{obs}, \theta^{(t)})}_{\text{imputation}} \underbrace{p(\mathbf{w}_i|\mathbf{x}_i, \gamma)}_{\text{mismeasurement}} \, d\mathbf{x}_i^{err} d\mathbf{x}_i^{mis} \tag{5}$$

in the E-Step. Note that $\theta^{(t)}$ refers to the t^{th} updated estimate of θ and $T(\mathbf{x}_i)$ to the complete data sufficient statistic (from which θ can be derived; under multivariate normality $T(\mathbf{x}) = \mathbf{X'X}$). Now, if we assume a *classical measurement error* model we implicitly specify

$$w_{ij} \quad \sim \quad N(x_{ij}, \lambda_{ij}^2) \qquad \forall w_{ij} \in \mathbf{w}_i \, . \tag{6}$$

Putting (6) into (5) and using the normality assumptions $\mathbf{x}_i^{err}|\mathbf{x}_i^{obs}, \theta \sim N(\mu_{e|o}, \mathbf{\Sigma}_{e|o})$, $\mathbf{w}_i|\mathbf{x}_i^{err}, \lambda_i^2 \sim N(\mathbf{x}_i^{err}, \mathbf{\Lambda}_i)$ yields the following distribution

$$(\mathbf{x}_i^{err}|\mathbf{d}_i, \theta^{(t)}, \lambda_i^2) \sim N(\mu^*, \mathbf{\Sigma}^*) \quad \text{with} \quad \mathbf{\Sigma}^* = (\mathbf{\Lambda}_i^{-1} + \mathbf{\Sigma}_{e|o}^{-1})^{-1}, \, \mu^* = \mathbf{\Sigma}^*(\mathbf{\Lambda}_i^{-1}\mathbf{w}_i + \mathbf{\Sigma}_{e|o}^{-1}\mu_{e|o}) \, , \tag{7}$$

as demonstrated by Blackwell et al. (2015a). Thus, to calculate the expectation on the left hand side of (5) for each cell with error, the E-Step needs simply make use of (7). This will result in overall proper multiple overimputations drawn from (4).

6) *Implications for the Implementation with* `Amelia II`*:* The standard EMB algorithm is implemented in the *R*-package `Amelia II` (Honaker et al., 2011). It allows the incorporation of prior distribution for single cells, i.e $x_{ij} \sim N(\mu_{ij,0}, \kappa_{ij,0}^2)$. If using $\mu_{ij,0} = w_{ij}$ and $\kappa_{ij,0}^2 = \lambda_{ij}^2$ one obtains the same results as in (7), see the appendix of Honaker and King (2010); and thus, using priors for mismeasured cells which equal $x_{ij} \sim N(w_{ij}, \lambda_{ij}^2)$ yields draws from the modified EMB algorithm described in step 5, and therefore *proper* multiple overimputations. To specify the mismeasured cells one needs the *overimp* option of the function `amelia`, and to specify the priors for the respective cells one needs the *priors* option.

7) *Combining estimates after multiple overimputation:* After generating M overimputed datasets by means of multiple overimputation, the analysis model (e.g. any regression model) can be fitted on each overimputed dataset and the M results will be combined as follows: the point estimate of θ (here implicitly referring to the parameters in the analysis model) is

$$\hat{\theta}_{\text{MI}} = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}^{(m)} \tag{8}$$

where $\hat{\theta}^{(m)}$ refers to the estimate of θ in the mth overimputed set of data $\mathcal{D}^{(m)}$, $m = 1, \ldots, M$. Based on the average within imputation covariance $\widehat{\mathbf{W}} = M^{-1} \sum_m \widehat{\text{Cov}}(\hat{\theta}^{(m)})$ and the between imputation covariance $\hat{\mathbf{B}} = (M-1)^{-1} \sum_m (\hat{\theta}^{(m)} - \hat{\theta}_{\text{MI}})(\hat{\theta}^{(m)} - \hat{\theta}_{\text{MI}})'$ one obtains variance estimates via

$$\widehat{\text{Cov}}(\hat{\theta}_{\text{MI}}) = \widehat{\mathbf{W}} + \frac{M+1}{M}\hat{\mathbf{B}} = \frac{1}{M} \sum_{m=1}^{M} \widehat{\text{Cov}}(\hat{\theta}^{(m)}) + \frac{M+1}{M(M-1)} \sum_{m=1}^{M} (\hat{\theta}^{(m)} - \hat{\theta}_{\text{MI}})(\hat{\theta}^{(m)} - \hat{\theta}_{\text{MI}})' \tag{9}$$

To construct confidence intervals for $\hat{\theta}_{\text{MI}}$ in the scalar case, it may be assumed that $\widehat{\text{Var}}(\hat{\theta}_{\text{MI}})^{-\frac{1}{2}}(\hat{\theta}_{\text{MI}} - \theta)$ follows a t_R-distribution with approximately $R = (M-1)[1 + \{M\hat{W}/(M+1)\hat{B}\}]^2$ degrees of freedom. Instead of computing these quantities by hand the function `mi.inference` in the *R*-package `norm` can be used; or, alternatively, the functionalities in the *R*-package `Zelig` or the Stata commands `mi estimate` or `mim` can be used.

eText 2: Simulation of viral loads and the misclassification proportion related to measurement error.

R-code:

```
# Generating true data
n=30000
VL            <- rlnorm(n, meanlog=10.760, sdlog=1.808607)

# Generating mismeasured data
VL_measured   <- 10^(log10(VL)+rnorm(n,0,0.255))

# Virological suppression if VL<1000
VL_supp            <- as.numeric(VL<1000)
VL_measured_supp   <- as.numeric(VL_measured<1000)

# Evaluating misclassification
VL_total       <- cbind(VL_supp,VL_measured_supp)
misclass_FN    <- as.numeric(VL_total[,1]==1 &  VL_total[,2]==0)
misclass_FP    <- as.numeric(VL_total[,1]==0 &  VL_total[,2]==1)
mean(misclass_FN)+mean(misclass_FP)  # overall misclassification
```

With $z_{0.985} = 2.17$ and a standard deviation of 0.23 one obtains about 1.5% misclassification by evaluating the confidence intervals related to the prior distributions used during multiple overimputation for the mismeasured values. For example, if a patient had a virological failure ($\mathrm{VL}_{\mathrm{supp}} = 0$) we impose a prior normal distribution, $N(0, 0.23)$, on the mismeasured value which implies that the upper limit of a 98.5% confidence interval corresponds to $0+2.17\cdot 0.23 = 0.499$ and thus 1.5% of values from this normal distribution are > 0.499 and therefore get rounded off to 1 which relates to virological failure and thus misclassification. The motivation and more details on how to use prior normal distributions for categorical variables can be found in Blackwell et al. (2015b) and Blackwell et al. (2015a).

References

Blackwell, M., Honaker, J., King, G., 2015a. A unified approach to measurement error and missing data: Details and extensions. Sociological Methods and Research, in press, copy at http://j.mp/1BQAvH9[1].

Blackwell, M., Honaker, J., King, G., 2015b. A unified approach to measurement error and missing data: Overview and applications. Sociological Methods and Research, in press, copy at http://j.mp/jqdj72[1].

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39, 1–38.

Goodnight, J. H., 1979. Tutorial on the sweep operator. American Statistician 33 (3), 149–158.

Honaker, J., King, G., 2010. What to do about missing values in time-series cross-section data? American Journal of Political Science 54, 561–581.

Honaker, J., King, G., Blackwell, M., 2011. Amelia II: A program for missing data. Journal of Statistical Software 45 (7), 1–47.

King, G., Honaker, J., Joseph, A., Scheve, K., 2001. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. American Political Science Review 95, 49–69.

Rubin, D. B., 1996. Multiple imputation after 18+ years. Journal of the American Statistical Association 91 (434), 473–489.

[1]Please note that eText1, as well as parts of the main body of the manuscript, uses terminology from a technical report of Blackwell at el. which resulted in two publications, Blackwell et al. (2015a) and Blackwell et al. (2015b). While the content of the publications and the content of the technical report is similar, the language used may differ. For example, we refer to the 'mismeasured at random' assumption which relates to the 'Ignorable Measurement Mechanism Assignment (IMMA)' assumption in Blackwell et al. (2015a).