

Handling missing data in a composite outcome with partially observed components: Application in clustered paediatric routine data.

Susan Gachau (✉ sgachau@kemri-wellcome.org)

KEMRI/ Wellcome Trust <https://orcid.org/0000-0002-2167-857X>

Edmund Njeru Njagi

London School of Hygiene and Tropical Medicine

Nelson Owuor

University of Nairobi

Paul Mwaniki

KEMRI-Wellcome Trust Research Programme, University of Nairobi

Matteo Quartagno

University College London

Mike English

KEMRI-Wellcome Trust and University of Oxford

Philip Ayieko

London School of Hygiene and Tropical Medicine, Mwanza Intervention Trails Unit, Mwanza, Tanzania

Research article

Keywords: Composite outcome, multiple imputation, paediatrics, PAQC score, pneumonia

Posted Date: October 3rd, 2019

DOI: <https://doi.org/10.21203/rs.2.15450/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: In health care settings, composite measures are used to combine information from multiple quality of care measures into a single summary score. Composite scores provide global insights and trends about complex and multidimensional quality of care processes. However, missing data in subcomponents may hinder the overall reliability of the composite measures in subsequent analysis and inferences. In this study we demonstrate strategies for handling missing data in Paediatric Admission Quality of Care (PAQC) score, a composite outcome which summarizes quality of inpatient paediatric care in low income settings.

Methods: We analysed routine data collected in a cluster randomized trial in 12 Kenyan hospitals. Multilevel multiple imputation (MI) within joint model framework was used to fill-in missing values in selected PAQC score subcomponents and partially observed covariates across two levels of hierarchy. We used proportional odds random intercepts and generalized estimating equations (GEE) models to analyse PAQC score before and after multiple imputation. Using a set of simulations scenario, that is, varied proportions of missingness in PAQC score subcomponents of interest under missing at random and missing completely at random mechanisms respectively, we compared the magnitude of bias in parameter estimates obtained under MI and the conventional method in addressing missing data in PAQC score components. Under the conventional method we scored all missing PAQC score components with value 0.

Results: Results from observed data showed that multiple imputation of both PAQC score components and covariates yielded more accurate and precise estimates compared to complete case analysis. From the simulation study, the conventional missing data method led to significantly larger biases in estimated proportional log odds of the outcome compared to MI methods. The amount of bias increased with increase in rate of missingness with substantial variation between the missing data mechanisms under the conventional method.

Conclusion: In comparison with conventional method, MI produce minimally biased estimates regardless of amount of missing data rate and underlying mechanism. We therefore recommend avoiding the conventional method in favour of multiple imputation; more research is needed to compare different ways of performing multiple imputation at the component and composite outcome level.

TRIAL REGISTRATION: US National Institutes of Health-ClinicalTrials.gov identifier (NCT number) NCT02817971 . Registered September 28, 2016-retrospectively registered.

Introduction

Composite measures combine information from multiple measures into a single summary score [1–6]. In health care settings, composite measures have been used as scorecards to measure and benchmark performance and quality of care in neonates [7] and cardiovascular care among adults, [5, 6, 8, 9]. In 2010, Profit et al presented a conceptual framework on composite indicator development in paediatrics care [7]. More recently, Opondo et al., developed and validated the Paediatric Admission Quality of Care (PAQC) score; a 7-point composite score aimed at benchmarking processes of care among children admitted with common childhood illnesses in low income settings [10]. In the validation study, PAQC score was shown to be a good proxy for outcome of care [11]. Besides gain in statistical efficiency, composite scores reduce the amount of data processed thus providing global insights and trends about complex and multidimensional quality of care processes [2, 3, 9]. In addition, the issue of multiple testing is avoided [12, 13]. Although composite outcomes complement single measures, weak theoretical and statistical assumptions may undermine the overall reliability [6, 7]. For instance, use of inappropriate methods to deal with partially observed subcomponents may impede the validity and reliability of the composite measure in subsequent analyses and inferences [2, 4, 6, 9, 14]. In literature, multiple imputation (MI), proposed by Rubin [15], offers a good, often best practice, solution in dealing with partially observed outcomes and covariates [16–19]. In particular, handling missing data in single outcomes (with no subcomponents) is straight forward because the imputation model is usually equivalent to the analyst's model [20]. On the other hand, dealing with missing data in composite outcome context has not received the same level of attention with no consensus on whether to impute at the composite score level or at the missing components level [4, 21].

In previous analysis of PAQC score, partially observed subcomponents were scored with value 0 representing suboptimal care [10]. This approach of dealing with missing PAQC score components is henceforth referred to as "*conventional method*", which will be deemed equivalent to single imputation. A major limitation of the single imputation method is inability to capture uncertainty in the missing data values leading to under estimated standard errors [18, 22]. Using routine paediatric pneumonia data from Kenyan hospitals, we first aim to explore the missing data mechanisms underlying the missing PAQC score components. Depending on the plausible mechanisms, we will explore appropriate strategies of dealing with missing data in the PAQC score components.

In addition, missing data in predictor variables will be addressed, properly accounting for the hierarchical structure of the data (i.e., patients clustered within clinicians who are then nested within hospitals). Through a range of simulation conditions, that is, three missing data rates under two missing data mechanisms, we will assess the implications of the missing data method (MI versus the conventional method)

employed in addressing missing PAQC score subcomponents. Specifically, the amount of bias in regression coefficients and corresponding standard errors attributable to missing PAQC score components across the simulation scenarios will be obtained and compared between MI and the conventional method.

Methods

Case study design

Data analysed here came from a cluster randomized trial conducted in 12 Kenyan county-level hospitals participating in the Clinical Information Network (CIN), an ongoing observational study in routine inpatient paediatric care [23, 24].

The trial objective was to investigate uptake of paediatric pneumonia treatment guidelines as recommended by the World Health Organization (WHO) in 2013 [25]. Detailed on the trial are contained in the trial report [26]. In brief, hospitals were randomly allocated to receive enhanced (6 hospitals) or standard (6 hospitals) audit and feedback. Enhanced audit and feedback constituted a monthly enhanced audit and feedback report on assessment, classification and treatment of pneumonia cases, a bi-monthly standard audit and feedback report on general inpatient paediatric routine care and network intervention strategies [26, 27]. Standard audit and feedback on the other hand constituted a standard audit and feedback report alone and network intervention strategy [26, 27]. Overall, 2299 children aged 2 to 59 months were admitted with childhood pneumonia in 12 study hospitals between March 2016 to November 2016. Data were abstracted by trained data clerks from individual patient medical records after discharge from hospital. The data were entered into an open source data capture tool (Research Electronic Data Capture, (REDCap)) [28] using standard operating procedure manual. For each case record, details of the admitting clinician including a unique clinician code, sex and cadre ("cadre" refers to clinician's qualification depending on the level of training, that is, clinical officers for a clinician with diploma-level training and medical officers for clinician with a bachelor's degree level training) were also abstracted into a separate database. Patients' and clinicians' databases were linked by unique clinician code. The Kenya Ministry of Health and Kenya Medical Research Institute's Scientific and Ethical Review Unit approved the use of de-identified patient data obtained through retrospective review of medical records without individual patient consent.

Outcome: Paediatric Admission Quality of Care (PAQC) score

The outcome of interest was Paediatric Admission Quality of Care (PAQC) score adjusted from its original form to encompass the new pneumonia treatment guidelines. In brief, PAQC score is an ordinal composite measure spanning three quality of paediatric care domains namely assessment, clinical diagnosis and treatment of common childhood illnesses [10]. PAQC score is constructed by summing 6 binary indicators and it ranges between 0 and 6 [10, 11, 29]. A minimum score of zero corresponds to inappropriate pneumonia care and maximum score of 6 represents total adherence to recommended clinical guidelines across domains of care. Details on how we constructed pneumonia PAQC score before and after multiple imputation are provided in subsequent sections.

Covariates

The predictor variables of interest in this analysis included an interaction between the intervention arm and follow up time (in months), hospital level covariates (i.e., malaria prevalence status and paediatric admission workload), clinician level covariates (i.e., gender and cadre). At patient level we considered sex, age categorized into 2–11 months and 12–59 months respectively and the number of comorbid illnesses. Although WHO pneumonia guidelines lines apply for children aged 2 to 50 months [30], we categorized patients in 2 age group because older children have better clinical outcomes compared to infants [31]. To determine the number of comorbidities, we considered common clinical diagnoses documented in patient's medical records besides pneumonia. These included malaria, malnutrition, HIV, Asthma, Tuberculosis (TB), rickets, anaemia, diarrhoea and dehydration. For each diagnosis, we created binary variables with 1 denoting the presence of a disease and 0 denoting absence of a disease. Thereafter, we summed the binary indicators and categorized patients into those with 0, 1, 2, 3 comorbidities. Approximately 48% (1010/2127) of the patient had no comorbidities while 31.2% (663/2127), 16.2% (345/2127) and 4.6% (98/2127) had 1, 2 and 3 comorbidities respectively. There were 11/2127 (0.52%) patients with 4 or more comorbidities. In subsequent analyses, we considered patients with 3 comorbidities and patients with 4 or more comorbidities as one category.

Missing data in pneumonia trial data

We linked patients and clinicians' databases using unique clinician code present in both databases with a success rate of 92.5% (2127/2299). This after exclusion of 172/2299 case records lacking admitting clinician's information. This resulted in a hierarchical data set

with three levels of clustering i.e., 2127 patients (level 1) admitted by 378 specific clinicians (level 2) in 12 participating hospitals (level 3). Among the 378 admitting clinicians, gender and cadre were missing in 21.9% (83/378) and 21.7% (82/378) cases respectively. At patient level, all covariates of interest were fully observed except 0.7% (17/2127) case records with missing patient's sex. We also observed missing data in pneumonia care processes used in the construction of PAQC score for the pneumonia trial dataset. A summary of documentation across care domains is presented in Table 1. In the assessment domain, 6 out of 9 signs and symptoms relevant for pneumonia diagnosis and severity classification were not fully documented. The proportion of missingness ranged between 0.2% and 39% (Table 1).

In the second domain, clinical diagnosis was fully documented, however, only 69.3% (1473/2127) were correctly classified. That is, clinical pneumonia diagnosis was in agreement with syndromic pneumonia implied by primary and secondary signs [25]. In the treatment domain, about 3.06% (65/2127) of all eligible pneumonia cases had missing prescription. Of the remaining pneumonia, 50.2% (1036/2062) received oral amoxicillin while 49.8% (1026/2062) were not prescribed with oral amoxicillin contrary to guidelines. Amongst patients with an oral amoxicillin prescription, dose, frequency of administration and patient's weight necessary for calculation of dosage per kilo body weight were missing in 0.4%, 2.6% and 2.9% case records respectively (Table 1).

Insert Table 1

Missing data mechanism underlying pneumonia trial data

Among pneumonia quality of care processes, undocumented (missing) primary and secondary pneumonia signs and symptoms (assessment domain) and missing amoxicillin prescription (treatment domain) were considered as inappropriate pneumonia care and were therefore not of interest in this analysis. However, missing data in amoxicillin dose, frequency of administration and weight of patient among oral amoxicillin recipients (treatment domain) (Table 1) were of interest hence explored further. This was in addition to missing level 1 covariate (patient's sex) and level 2 covariates (clinician's sex and cadre). To investigate plausible missing data mechanisms underlying each partially observed variable of interest, we created a binary missing data indicator and regressed it on fully observed variables. The predictor variables of interest included an interaction between intervention arm and follow up time in months, number of comorbid illnesses, age of patient, hospital malaria prevalence and paediatric admission workload. We also included observed assessment and diagnosis domain components as predictors. When the probability of missing values in a variable was independent of the variable itself or any other observed variable in the data set [32], then we assumed a Missing completely at random (MCAR) mechanism, although an MNAR mechanisms was theoretically still plausible. On the other hand, if the probability of missing values in a variable did not depend on the variable of interest but was conditionally dependent on other observed variables in the data set, we assumed a Missing at random (MAR) mechanism [32]. Preliminary results (Supplementary Table A1) suggested that for clinician's cadre and sex, missingness was dependent on some fully observed variables hence missing at random (MAR). Likewise, the probability of missing amoxicillin dose and frequency of administration and patients' weight were dependent on observed variables (supplementary Table A1).

Multilevel multiple imputation of missing covariates and PAQC score components

Multiple imputation (MI), proposed by Rubin [15] is often the recommended method for obtaining valid parameter estimates from partially observed data [16–18]. MI generally relies on the MAR assumption and it involves three sequential steps: imputation, analysis, and pooling of parameter estimates. In the first step, independent random samples are drawn from the posterior predictive distribution of the missing values given the observed data and a statistical imputation model, thus generating more than one filled-in data sets. Imputed datasets are then analysed using standard statistical methods in step 2. In step 3 final estimates are obtained by averaging over the parameter estimates from all multiply imputed data sets according to Rubin's Rule [15]. In this study, partially observed patients' (level 1) and clinicians' (level 2) variables were imputed within the joint model framework implemented in *jomo* [33] and *mitml* [34] packages in R (version 3.5.4). In the JM framework, replacement values are drawn jointly from a multivariate distribution in a single step [18]. Letting i index patient, j clinician and l hospital, our 2-level MI model corresponded to

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (1)

where the [Due to technical limitations, this equation is only available as a download in the supplemental files section.] is a vector of missing PAQC components in the treatment domain (i.e., patient's weight, oral amoxicillin dose and frequency of amoxicillin administration among oral amoxicillin recipients) and patient's sex. Level 1 predictors [Due to technical limitations, this equation is only available as a download in the supplemental files section.] included an interaction term between follow-up time and intervention arm, hospital workload and malaria prevalence status, patient's age and number of comorbid illnesses. Observed PAQC score components in the assessment and treatment

domains were also included as level predictors variables. Missing clinicians' sex and cadre (target variables) in the second level of the hierarchical structure are denoted by [Due to technical limitations, this equation is only available as a download in the supplemental files section.]. The corresponding predictor variables denoted by [Due to technical limitations, this equation is only available as a download in the supplemental files section.] included an interaction between follow-up time and intervention arm, hospital workload and malaria prevalence status. Column vectors β^1 and β^2 denote level 1 and level 2 fixed effects respectively. A clinician random intercept ($b_{j,i}$) was included to account for clustering at clinicians' level and to ensure compatibility with substantive models of interests. A burn-in of 1000 updates and 100 iterations between each of the 30 imputations were considered. We then assessed chains for individual variables where diagnostic tests [35] indicated satisfactory convergence (Supplementary Figure A1). Final parameter estimates were pooled according to Rubin's rule[32].

PAQC score after multiple imputation of missing treatment domain components

After imputing missing amoxicillin dose, amoxicillin frequency and patient's weight, we constructed PAQC score in each imputed data set in a two-step procedure. First, we created binary indicators with 1 representing adherence to recommended childhood pneumonia guidelines and 0 representing inappropriate care. Specifically, the value zero in three assessment domain constituents corresponded to: - i) lack of documentation of at least of one of the primary signs and symptoms required for pneumonia identification; ii) lack of documentation of at least one of the 7 secondary signs and symptoms required for pneumonia severity classification; iii) incomplete documentation of all primary and secondary pneumonia signs and symptoms (Table 1). For the second PAQC score domain, we created a binary indicator with 1 representing pneumonia diagnosis and classification and 0 for any other classification (e.g. severe pneumonia). The third PAQC score domain (treatment) comprised 2 components; a binary indicator with 1 corresponding to oral amoxicillin prescription and 0 representing inappropriate care either due to missing prescription or documentation in the case record that oral amoxicillin was not prescribed. Among patients prescribed oral amoxicillin, we created a new variable "recommended dose per kilo body" after MI of missing amoxicillin dose, amoxicillin frequency and patient's weight. We then transformed the new variable into binary form with 1 representing recommended oral amoxicillin dose (i.e., dose between 32 and 48 international units (IU) per Kilogram (Kg) every 12 hours) and 0 representing either wrong dosage (under dose for oral amoxicillin < 32 IU/Kg or over dose for oral amoxicillin >48IU/Kg), wrong frequency of drug administration (e.g. administration frequency of once every 24 hours instead of once every 12 hours) or both. In the second and final step of PAQC score construction, we summed all the 6 binary indicators spanning assessment (n = 3), clinical diagnosis (n = 1) and treatment (n = 2) domains to obtain PAQC score. After MI variation in PAQC score on the 7-point scale was attributed to inappropriate inpatient pneumonia care. That is, undocumented primary and secondary signs and symptoms (assessment domain), misclassification of disease severity, failure to prescribe the oral amoxicillin drug or prescription of the drug in the wrong dose or frequency [10].

PAQC score under conventional approach

Under the conventional approach, construction of pneumonia PAQC score proceeded as above except that missing amoxicillin dose, amoxicillin frequency, patients weight among oral amoxicillin recipients in treatment domain were not imputed. Instead, we scored them with value 0 in the binary indicators. As a result, variation in PAQC score on the 7-point scale was due to missing data and/or inappropriate care across the three domains of care. That is, missing data with reference to missing dose per kilo body weight and or frequency among oral amoxicillin recipients. Inappropriate care with reference to undocumented primary and secondary signs and symptoms in the assessment domain, incorrect severity classification, undocumented oral amoxicillin prescription or prescription of the drug in the wrong dose or frequency [10].

Statistical analysis

When the responses are ordered and the proportional odds assumptions are upheld (parallel logits) the cumulative logits (proportional odds) model is considered [36]. The proportional odds model expresses the k -category ordered outcome in terms of $k-1$ cumulative logits and estimated covariates effects are assumed common across all $k-1$ cumulative logits [16, 36]

When inference at population level is of interest, a marginal model such as the generalized estimating equation (GEE) model is used [37]. On the other hand, random effects models are useful when interest lies in drawing subject-specific inferences [16, 38]. We used both proportional odds random effects and GEE model families in order to assess stability of parameter before and after MI. Letting i index patient, j clinician and l hospital, the random intercepts model implemented in R's *Ordinal* package [39] corresponded to

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (2)

where α_k , $k = 1,2,3,4,5,6$ are PAQC score specific intercepts, are estimated regression coefficients and b_{ji} are clinician's random intercept. In this analysis, we did not include hospital random effects due to few number of clusters. Similarly, letting i index patient, j clinician and l hospital, the GEE model of interest implemented in R's *Multigeepack* [40] corresponded to

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (3)

where α_k , $k = 1,2,3,4,5,6$ are PAQC score intercepts. In this analysis, we adopted an independent working correlation. Both model families were used to analysis data under complete case analysis and after MI of missing covariates and PAQC score components in the treatment domain. Under complete case analysis, records with missing covariates were discarded while partially observed outcome subcomponents were handled using the conventional approach described in section above. A 5% level of significance was considered in all statistical analyses.

Simulation study

In further analysis, we conducted a simulation study with an objective of examining and comparing bias in parameter estimates (both regression coefficients and standard errors) under multiple imputation and the conventional approach of handling missing data in PAQC score components (treatment domain subcomponents). However, due to the structure of the pneumonia trial data (i.e. mixed variable types of PAQC score components) and the multilevel structure, creating a standard data set based on model parameters while preserving the correlation structure was a challenge. As an alternative, we opted to generate missing data in a complete subset of the pneumonia trial data. However, a limitation of this strategy is that only 16.7% (357/2127) case records in the pneumonia trial data were fully observed with the rest of the observations having missing data either in the covariates or PAQC score components. Using 16.7% of the observed data would affect precision with which parameter were estimated with in subsequent analyses [16, 41]. Consequently, we decided to create a subset of the pneumonia data set complete in PAQC subcomponents of interest, that is, amoxicillin dose, frequency and patient's weight amongst oral amoxicillin recipients. To achieve this, we excluded 65/2127 (3.1%) case records with missing oral amoxicillin prescription. Of the remaining 2062 (96.9%) pneumonia case records, 1036 (50.2%) were prescribed oral amoxicillin while 1026 (49.8%) pneumonia cases were not. Amongst patients prescribed oral amoxicillin, we further excluded 61/1036 (5.9%) cases for whom weight ($n = 30$), amoxicillin dose ($n = 4$) or frequency of amoxicillin administration ($n = 27$) were missing. Therefore, the standard data set we used in the simulation study was a subset of pneumonia trial data and it consisted of 200194.1% observations after exclusions above. Although our standard data set was complete in the target PAQC score components, we still had missing values in patient's sex, clinician's sex and cadre (covariates) as well as assessment domain components. Undocumented signs and symptoms in the assessment domain components were deemed as inappropriate care and were scored 0 in the binary indicators in the PAQC construction stage. To obtain the standard estimates (regression coefficients and standard errors), we multiply imputed missing covariates in the standard data set 10 times with 1000 burn-in and 100 iterations per imputation. Missing covariates were imputed within joint model approach. We fitted a random clinician's intercepts model to each imputed data set and pooled the final parameter estimates according to Rubin's Rule. The pooled (standard) estimates were used as reference estimates against which results from different simulation scenarios were benchmarked.

Simulation scheme

First, we generated missing data in the treatment domain subcomponents (patient's weight, amoxicillin dose and frequency) of the standard data assuming missing completely at random (MCAR) and missing at random (MAR) mechanisms respectively. Binary missing data indicators were generated by sampling random numbers from a random binomial distribution with success probability rates of 3%, 10% and 40%. A 3% missing data rate was selected to mimic the rate of missingness observed in the pneumonia trial data before exclusions (Table 1), while 10% and 40% were chosen to assess the extent of bias in moderate to high rates of missingness. Under the MCAR mechanism, missing values were imposed on treatment domain subcomponents independent of missing and observed PAQC subcomponents (i.e., assessment and clinical diagnosis domains) and covariates (i.e., hospital, clinician and patient characteristics). For the MAR condition, probabilities of missing data were conditionally dependent on fully observed variables associated with probability of missingness in the three variables of interest (based on the real trial dataset) (Supplementary Table 1). In both MAR and MCAR, missing data in weight, oral amoxicillin dose, and frequency of administration were induced independently of each other, such that either one, two or all three variables were missing for any given patient. Overall, we considered 6 simulation scenarios (i.e., two missing data mechanisms by 3 missingness rates). Each scenario was simulated 1,000 times. We chose and maintained random number generators (seeds) for different scenarios to ensure reproducibility of results.

For each scenario, we used two approaches to handle missing data generated in PAQC score subcomponents of interest. In one approach, proposed MI was used to fill-in missing amoxicillin dose and frequency of administration and patient's weight (pneumonia PAQC score

subcomponents) in addition to patient's sex (level 1 covariate) and clinician's sex and cadre (level 2 covariates). In the second approach, we only imputed missing covariates while missing values in assessment and treatment domain components were handled using the conventional method. Specifically, under the conventional method all missing values were scored with a value zero in the binary indicators across all pneumonia PAQC score domains. For each simulated dataset, we imputed missing values 10 times with a burn-in of 1000 updates and 100 iterations per imputation. Thereafter we constructed pneumonia PAQC score before fitting the random intercepts model (see equation 2) to each imputed data set to obtain imputation-specific parameter estimates. Imputation-specific estimates were pooled using Rubin's rules to produce a single estimate for the n^{th} simulation. This procedure was repeated in all the scenarios. Bias in regression coefficients was calculated as differences between the estimates averaged over 1000 simulated datasets $\bar{\beta}$ and the standard estimates [Due to technical limitations, this equation is only available as a download in the supplemental files section.] pooled after MI of covariates in the standard data set. That is,

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (4)

To assess accuracy, we used model based standard errors that is, the average of the estimated within simulation standard errors, that is, [Due to technical limitations, this equation is only available as a download in the supplemental files section.]. Model based standard errors were compared with empirical standard errors calculated as the standard deviation of the estimates of interest [42] across the 1000 datasets, that is,

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (5)

where N is the number of simulations, β_i is the coefficient estimated in the i^{th} simulation and $\bar{\beta}$ is estimator's average over 1000 simulations. We also calculated the mean square error (MSE) for the regression coefficients. The MSE incorporates both measures of bias and variability [19, 42, 43], that is,

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (6)

Similarly, we assessed bias and accuracy of the corresponding standard errors. In this simulation study coverage probability of the 95% confidence intervals were not applicable because missing data were simulated on the same subset of the pneumonia trial dataset. All simulations were conducted in R version 3.5.4

Results: Case Study

Insert Table 2

In Table 2 we present random intercepts and GEE models parameter estimates (in log odds) and the corresponding standard errors obtained under complete case analysis (after deletion of case records with missing clinician's cadre, sex and patient's sex) combined with conventional approach of handling missing PAQC score elements. We also present estimates after multiple imputation missing covariates and missing PAQC score subcomponents in the treatment domain. In both model families, we observed change in the regression coefficients before and after imputing missing covariates and PAQC score subcomponents in the treatments domain. However, the magnitude of change varied across covariates of interest. The largest differences in proportional logs odds were observed in hospital workload regression coefficient with an approximate absolute difference of 0.25 (i.e., from -0.08 to -0.33) in the random effects model and an absolute difference of 0.15 (i.e., from -0.22 to -0.37) in the GEE (Table 2). We further observed model specific shifts in the direction of effect before and after MI. For example, under the random effects model, we observed negative patient's sex effect (log odds = -0.03) under complete case analysis which changed to a positive effect (log odds = 0.01) after MI (Table 1).

With regard to standard errors, MI led to more precise estimates across all variables compared to complete case records methods in both GEE and random intercepts models.

Moreover, we observed changes in statistical significance for some variables after multiple imputation. For instance, hospitals workload effect was not significant at 5% level of significance under complete cases analysis but turned out significant ($p\text{-value} < 0.05$) after multiple imputation. These results were observed in both random effects and GEE models (Table 2). Considering the random effects model, deleting case records with missing data led to inflated variance between clinicians compared to variance estimated under MI (Table 1). This could be explained by the fact that clinicians with missing cadre and sex were discarded under complete case analysis resulting to fewer number of clinicians (clusters) hence increased clinicians' uncertainty. On the other hand, all clinicians were retained after MI, hence leading to more precise estimates.

Results Simulation study

Insert Table 3

Insert Table 4

Estimated bias in regression coefficients and standard errors across 6 simulation scenarios (missing data rates and missing data mechanisms) under the conventional and proposed multiple imputation methods are presented in Figures 1 and 2 respectively. In Tables 3 and 4, we present bias, empirical standard errors, model based standard errors and MSE for the regression coefficients and standard errors respectively estimated under the MAR mechanism across three missing data rates. Other simulation results under MAR and MCAR are provided in supplementary Tables A2-A7.

From these results, the conventional approach of handling missing PAQC score subcomponents in treatment domain led to biased regression coefficient estimates and the magnitude of bias varied across variables (Table 4 and Supplementary Table A3 and Figure 1 and Figure 2). Additionally, bias in regression coefficients increased with an increase in the proportion of missingness and was somewhat larger when missingness in PAQC score components was generated under MAR mechanism than under MCAR mechanism.

With multiple imputation of missing treatment domain components, we observed smaller magnitude of bias in regression coefficients (Tables 3 and supplementary Table A2). This implied that the parameter estimates averaged over 1000 simulations were closer to the standard data estimates across the missing data rates and missing data mechanisms under consideration (MCAR and MAR).

With regard to standard error, the conventional approach of handling missing PAQC score components in the treatment domain led to larger bias across all variables of interest (Figure 2 and supplementary tables A6 and A7) compared to proposed MI methods (supplementary tables A4 and A5). As expected, under the MAR mechanism all methods consistently exhibited larger bias compared to MCAR mechanism across the missing data rates.

With both conventional and MI approaches, the regression coefficients were either underestimated (negative bias) or overestimated (positive bias), the standard errors across simulation scenarios tended to overestimate the true standard errors thus resulting in positive bias, reflecting the loss of information due to missing data. However, the regression coefficients were more prone to bias across individual variables (Figure 1) compared to standard errors (Figure 2). A possible explanation is that case records with missing PAQC score subcomponents were not discarded and analyses were based on all observations in the standard data set regardless of the approach used hence no major impact on the precision with which the standard errors were estimated across the simulation scenarios. Across simulation scenarios, the estimated empirical standard errors were close to the average of the estimated within simulation SE (model based standard errors). The magnitude of both estimates tended to increase with an increase in the proportion of missing data in PAQC score components. Across the simulation scenarios, MSEs were slightly larger under the conventional method compared to MI approach. Additionally, the MSEs were somewhat larger under MAR mechanism compared with MCAR mechanism.

Discussion

In this study we sought to analyse clustered data with covariate missingness across two levels of a multilevel structure. This was in addition to proposing appropriate strategy for handling missing data in a composite outcome and assessing how the proposed method performs in a simulation study in comparison to simple conventional approach. The composite outcome of interest was PAQC score [10] adjusted to the new pneumonia guidelines, among children aged 2 to 59 months (pocket book) admitted in 12 Kenyan hospitals during a cluster randomized trial. Individual components in PAQC score corresponded to care processes spanning assessment, diagnosis and treatment of inpatient paediatric pneumonia cases. From preliminary analysis, missing data in pneumonia care processes varied within and between PAQC score domains. The rate of missingness could be explained by complexity underlying individual tasks [44]. That is, care processes (PAQC subcomponents) perceived to require more cognitive effort (e.g. assessment and documentation of respiratory rate or oxygen saturation measurement) generally recorded higher rates of missingness compared to tasks that required less effort to perform (e.g. assessment and documentation of difficulty in breathing or cough history from the care giver).

While individual quality of care indicators are important in monitoring specific care processes, composite scores on the other hand summarize multiple individual measures into one single measure providing insight on overall quality of patient care. However, composite measures are complex and their construction should be based on sound conceptual and methodological foundations [3]. In composite measures development guidelines, a required step is strategies for handling missing data to minimize bias and enhance reliability of a composite score [7, 14]. This because missing data in individual items of a composite measure may be magnified when multiple components are combined [22].

In literature, handling missing data in composite components remains a major obstacle in epidemiological studies reporting composite measures [1, 6, 21]. Previously, a systematic review reported that only 2 in 40 (5%) clinical studies adequately and appropriately handled missing data in their composite outcomes [1]. In most studies, researchers avoid missing items in composite scores by conducting complete case analysis [21].

In the construction of PAQC score, missing components do not cause the whole scale score to be missing. Instead, subcomponents with missing values (e.g. undocumented signs and symptoms) and those corresponding to inappropriate care at patient level (e.g., overdose or under dose treatment prescription) are penalized with zero in the binary indicators [10, 11]. This leads to low scores on the 7-point ordinal scale. Another limitation of the conventional approach is that it consists in drawing a single imputation, a missing data method which is known to underestimate variability [18].

In our proposed strategy for handling missing data in pneumonia PAQC score, we coupled MI (a statistical solution) and expert opinion to handle missing data in PAQC score adjusted to new pneumonia guidelines. Specifically, 3 partially observed treatment domain components (oral amoxicillin dose, frequency of administration and patient weight) were multiply imputed taking into account underlying mechanism while documents pneumonia sign and symptoms in the assessment domain and missing oral amoxicillin prescription were treated as inappropriate care hence score 0 in line with the conventional approach. Our decision to treat some PAQC score subcomponents as inappropriate care was based by the study design in addition to expert opinion. By study design, the inclusion criterion was children admitted with pneumonia signs and symptoms (paediatric basic protocol). Therefore, undocumented pneumonia signs and symptoms (primary and secondary) in the assessment domain were treated as inappropriate care and therefore scored 0 in subsequent PAQC score construction. On the other hand, according to expert's belief, a possible explanation for missing oral amoxicillin prescription in the treatment domain could be that patients were prescribed other antibiotics other than oral amoxicillin. Considering this possibility, we scored missing prescription with value 0 because it was contrary to WHO paediatric pneumonia recommendations [25].

In addition to imputing missing PAQC score subcomponents in the treatment domain, we also imputed partially observed covariates, that is, patients' sex at level 1 clinician's sex and cadre at level 2. Multiple imputation of both PAQC score subcomponents (treatment domain) and missing covariates led to slight changes in regression coefficients estimates and standard errors. This was in comparison to results from previous analysis of the trial [45] where we only imputed missing covariates and handled all missing PAQC score subcomponents using the conventional method. Although the proportion of missingness in PAQC score subcomponents of interest (treatment domain) was small, the observed difference is an indication of MI superiority in handling missing PAQC score subcomponents over the conventional approach.

Through a range of simulation conditions, we examined bias in regression coefficients and standard errors associated with missing data in PAQC score treatment domain components. Although increasing the missing data rate had a predictable effect on bias, that is larger bias with increasing proportion of missingness [19] we included this scenario to examine the extent of variation in bias across the missing data mechanisms and between missing data techniques adopted in the study.

Across the missing data rates (3%, 10% and 40%) and the missing data mechanism (MAR and MCAR) underlying PAQC score components in the treatment domain, MI of partially observed treatment subcomponents within PAQC score led to estimates close to the standard estimates hence smaller biases in both regression coefficients and standard errors.

In contrast, using the conventional method to handle all missing PAQC score elements (including treatment domain subcomponents) and imputing missing covariates only led to larger biases across variables of interest. Moreover, the bias in regression coefficients were more pronounced than bias observed under conventional method and more so when missing PAQC score subcomponents were generated assuming a MAR mechanism.

Simulation study results further showed that the magnitude and direction of bias varied across variables. The bias estimated under MAR mechanism was consistently larger in magnitude than bias observed under MCAR and these differences were observed under conventional approach.

Strengths and implications of the study

Through this study we have demonstrated superiority of MI over the conventional method in handling missing data in PAQC score subcomponents. To our knowledge our study is the first study to investigate underlying missingness mechanisms and to propose MI, as a strategy for dealing with partially observed PAQC score domain components. Furthermore, our study is the first to estimate bias in parameter estimates associated with missing PAQC score components through a simulation study based on routine paediatric data. However, we note that even after MI of missing PAQC components in the treatment domain, we still observed some level of bias in the regression coefficients. A possible explanation for these observations could be the lack of compatibility between our imputation model and the analysis model. That is,

our imputation model included PAQC score components in the treatment domain as outcomes and PAQC components in the assessment and diagnosis domains as predictors variables, but the composite outcome was not included. Therefore, further research is needed to compare the performance of our MI method with that of MI including the composite outcome, possibly making use of so-called substantive model compatible imputation [46], in order to guarantee that the relation between component and composite outcome is preserved. MI has been previously used to address missing data in several composite scores assessing quality of patient's care [21, 22, 47]. In one study, Plumpton *et al.*, [22] proposed, MI at component level while another study by Simon *et al.*, [21] proposed MI at index level particularly for smaller samples. In the case of PAQC score, we recommend imputing individual subcomponents before constructing PAQC score. This because there are no possibilities of missing PAQC score at aggregate level (the only possibilities are values between 0 and 6). In consideration of our controlled study inclusion criteria (i.e., inclusion of patients with pneumonia signs and symptoms), we restricted application of multiple imputation in handling missing treatment domain components. We note that for routine paediatrics studies without restricted inclusion criteria, MI can be extended to handle missing PAQC score components in the assessment and diagnosis domains of paediatric care.

Limitations

This study had several limitations. First, due to the complex nature of the outcome and the data structure we simulated missing data on a complete subset of the pneumonia trial data. This approach has been used in previously simulation studies [21, 48]. Secondly, we did not estimate the true parameters from a complete subset of observed data. This is because most missing covariate data occurred in second level variables (clinician's sex and cadre) and approximately 83.2% case records had missing information. Exclusion of these incomplete records would have led to severe loss of information (i.e., reduced sample size) thus under powering the simulation study [16, 41].

Conclusion

Using pneumonia PAQC score we have demonstrated that missing data in a composite outcome subcomponent should be addressed carefully. In comparison with conventional method, MI produce minimally biased estimates regardless of amount of missing data rate and underlying mechanism. However, more research is needed to compare different ways of performing multiple imputation at the component and composite outcome level.

List Of Abbreviations

CIN: Clinical Information Network:

GEE: Generalized Estimating Equations

GLMM: Generalized Linear Mixed Models

JM: Joint Modelling

MAR: Missing at Random

MCAR: Missing Completely at Random

MNAR: Missing Not at Random

MI: Multiple Imputation

PAQC: Paediatric Admission Quality of Care score

Declarations

Ethics approval and consent to participate

The Kenya Ministry of Health and Kenya Medical Research Institute's Scientific and Ethical Review Unit approved the use of de-identified patient data obtained through retrospective review of medical records without individual patient consent.

Consent for publication

Not applicable

Availability of data and materials

The datasets analysed in this study are not publicly available because they are a property of the Ministry of Health and we do not have authority to share on their behalf.

Competing interests: The authors have declared that no competing interests exist

Funding

This work was supported through the DELTAS Africa Initiative Grant No. 107754/Z/15/Z-DELTAS Africa SSACAB. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant No. 107754/Z/15/Z) and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Funds from the Wellcome Trust (*GrantNo.207522*) awarded to Prof. Mike English as a senior Fellowship together with additional funds from a Wellcome Trust core grant awarded to the KEMRI-Wellcome Trust Research Programme (*GrantNo.092654*) supported CIN data collection.

Authors' contribution

SG conducted the analyses. Feedback on the analytic approach was provided by ENN, NO, PM, MQ, ME and PA. SG drafted the initial manuscript with feedback on subsequent drafts provided by all authors who then approved the final manuscript.

Acknowledgements

We would like to thank the Ministry of Health who gave permission for this work to be developed and have supported the implementation of the CIN together with the county health executives and all hospital management teams. We are grateful to the Kenya Paediatric Association for promoting the aims of the CIN and the support they provide through their officers and membership. We also thank the hospital teams involved in service delivery for the sick child. This work is published with the permission of the Director of KEMRI.

The Clinical Information Network team who contributed to the design of the data collection tools, conduct of the work, collection of data and data quality assurance that form the basis of this report and who saw and approved the report's findings include: Grace Irimu, Samuel Akech, Ambrose Agweyu, Michuki Maina, Jacquie Oliwa, David Gathara, Lucas Malla, Morris Ogero, James Wafula, George Mbevi, Mercy Chepkirui (KEMRI-Wellcome Trust Research Programme); Samuel N'garng'ar (Vihiga County Hospital), Ivan Muroki (Kakamega County Hospital), David Kimutai & Loice Mutai (Mbagathi County Hospital), Caren Emadau & Cecilia Mutiso (Mama Lucy Kibaki Hospital), Charles Nzioki (Machakos Level 5 Hospital), Francis Kanyingi & Agnes Mithamo (Nyeri County Hospital), Margaret Kuria (Kisumu East County Hospital), Samuel Otido (Embu County Hospital), Grace Wachira & Alice Kariuki (Karatina County Hospital), Peris Njiiri (Kerugoya County Hospital), Rachel Inginia & Melab Musabi (Kitale County Hospital), Hilda Odeny (Busia County Hospital), Grace Ochieng & Lydia Thurania (Kiambu County Hospital); Priscilla Oweso (Vihiga County Hospital), Ernest Namayi (Mbale Rural Health and Demonstration Centre), Benard Wambani, Samuel Soita (Kakamega Provincial General Hospital), Joseph Nganga (Mbagathi District Hospital), Margaret Waweru, John Karanja (Kiambu County Hospital), Susan Owano (Mama Lucy Kibaki Hospital), Esther Muthiani (Machakos Level 5 Hospital), Alfred Wanjau (Nyeri Level 5 hospital), Larry Mwallo (Kisumu East District Hospital), Lydia Wanjiru (Embu Provincial General Hospital), Consolata Kinyua (Karatina District Hospital), Mary Nguri (Kerugoya District Hospital) and Dorothy Munjalu (Kitale District Hospital).

References

- 1.Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC: *Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review*. In: *BmJ*. vol. 341; 2010: c3920.
- 2.Profit J, Kowalkowski MA, Zupancic JA, Pietz K, Richardson P, Draper D, Hysong SJ, Thomas EJ, Petersen LA, Gould JB: *Baby-MONITOR: a composite indicator of NICU quality*. In: *Pediatrics*. vol. 134; 2014: 74–82.
- 3.Shwartz M, Restuccia JD, Rosen AK: *Composite measures of health care provider performance: a description of approaches*. In: *The Milbank Quarterly*. vol. 93; 2015: 788–825.
- 4.Ibrahim F, Tom BD, Scott DL, Prevoost AT: *A systematic review of randomised controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes*. *Trials* 2016, 17(1):272.

- 5.Chen LM, Staiger DO, Birkmeyer JD, Ryan AM, Zhang W, Dimick JB: *Composite quality measures for common inpatient medical conditions. Medical care* 2013, *51*(9):832.
- 6.Caldis T: *Composite health plan quality scales. Health care financing review* 2007, *28*(3):95.
- 7.Profit J, Typpo KV, Hysong SJ, Woodard LD, Kallen MA, Petersen LA: *Improving benchmarking by using an explicit framework for the development of composite indicators: an example using pediatric quality of care. Implementation Science* 2010, *5*(1):13.
- 8.Eapen ZJ, Fonarow GC, Dai D, O'Brien SM, Schwamm LH, Cannon CP, Heidenreich PA, Bhatt DL, Peterson ED, Hernandez AF: *Comparison of composite measure methodologies for rewarding quality of care: an analysis from the American Heart Association's Get With The Guidelines program. Circulation: Cardiovascular Quality and Outcomes* 2011, *4*(6):610–618.
- 9.EUnetHTA J: *WP5 methodology guidelines Endpoints used for relative effectiveness assessment of pharmaceuticals HEALTHRELATED QUALITY OF LIFE and UTILITY MEASURES*. In.: February; 2013.
- 10.Opondo C, Allen E, Todd J, English M: *The Paediatric Admission Quality of Care (PAQC) score: designing a tool to measure the quality of early inpatient paediatric care in a low-income setting. Tropical Medicine & International Health* 2016, *21*(10):1334–1345.
- 11.Opondo C, Allen E, Todd J, English M: *Association of the Paediatric Admission Quality of Care score with mortality in Kenyan hospitals: a validation study. The Lancet Global Health* 2018, *6*(2):e203-e210.
- 12.Proschan MA, Waclawiw MA: *Practical guidelines for multiplicity adjustment in clinical trials. Controlled clinical trials* 2000, *21*(6):527–539.
- 13.Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C: *Composite outcomes in randomized trials: greater precision but with greater uncertainty? Jama* 2003, *289*(19):2554–2559.
- 14.Commission JRC-E: *Handbook on constructing composite indicators: methodology and user guide*: OECD publishing; 2008.
- 15.Rubin DB: *Procedures with nonignorable nonresponse. Multiple Imputation for Nonresponse in Surveys* 1987:202–243.
- 16.Molenberghs G, Verbeke G: *Missing data concepts. Models for Discrete Longitudinal Data* 2005:481–488.
- 17.van Buuren S, Groothuis-Oudshoorn K: *Mice: multivariate imputation by chained equations in R. JStat Softw* 45 (3): 1–67. In.; 2011.
- 18.Carpenter JR, Kenward MG: *Multiple imputation and its applications*: Chichester: John Wiley & Sons; 2013.
- 19.Enders CK, Mistler SA, Keller BT: *Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation*. In: *Psychological methods*. vol. 21; 2016: 222.
- 20.Grund S, Lüdtke O, Robitzsch A: *Multiple imputation of multilevel missing data: An introduction to the R Package pan. SAGE Open* 2016, *6*(4):2158244016668220.
- 21.Simons CL, Rivero-Arias O, Yu L-M, Simon J: *Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? Quality of life research* 2015, *24*(4):805–815.
- 22.Plumpton CO, Morris T, Hughes DA, White IR: *Multiple imputation of multiple multi-item scales when a full imputation model is infeasible. BMC research notes* 2016, *9*(1):45.
- 23.English M: *Designing a theory-informed, contextually appropriate intervention strategy to improve delivery of paediatric services in Kenyan hospitals*. In: *Implementation Science*. vol. 8; 2013: 1.
- 24.Tuti T, Bitok M, Paton C, Makone B, Malla L, Muinga N, Gathara D, English M: *Innovating to enhance clinical data management using non-commercial and open source solutions across a multi-center network supporting inpatient pediatric care and research in Kenya*. In: *Journal of the American Medical Informatics Association*. vol. 23; 2016: 184–192.
- 25.Organization WH: *Pocket book of hospital care for children: guidelines for the management of common childhood illnesses*, 2nd edn. Geneva: World Health Organization; 2013.
- 26.Ayieko P, Irimu G, Ogero M, Mwaniki P, Malla L, Julius T, Chepkirui M, Mbevi G, Oliwa J, Agweyu A: *Effect of enhancing audit and feedback on uptake of childhood pneumonia treatment policy in hospitals that are part of a clinical network: a cluster randomized trial. Implementation*

Science 2019, 14(1):20.

27. Ayieko P, Irimu G, English M: *Effect of enhanced feedback to hospitals that are part of an emerging clinical information network on uptake of revised childhood pneumonia treatment policy: study protocol for a cluster randomized trial*. In: *Trials*. vol. 18; 2017: 416.
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG: *Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support*. *Journal of biomedical informatics* 2009, 42(2):377–381.
29. Opondo C, Allen E, Todd J, English M: *The Paediatric Admission Quality of Care (PAQC) score: designing a tool to measure the quality of early inpatient paediatric care in a low-income setting*. In: *Tropical medicine & international health*. vol. 21; 2016: 1334–1345.
30. Organization WH: *Updates on the management of severe acute malnutrition in infants and children*. Geneva: WHO 2013:p60.
31. UNICEF U: *Levels and trends in child mortality. Estimates developed by the UN inter-agency group for child mortality estimation* 2011.
32. Rubin DB: *Inference and missing data*. *Biometrika* 1976, 63(3):581–592.
33. Quartagno M, Carpenter J: *jomo: A package for multilevel joint modelling multiple imputation*. *R package version* 2016:2.2–0.
34. Grund S, Robitzsch A, Luedtke O: *mitml: Tools for Multiple Imputation in Multilevel Modeling*. *R package version* 0.3–3. In.; 2016.
35. Gelman A, Rubin DB: *Inference from iterative simulation using multiple sequences*. In: *Statistical science*. 1992: 457–472.
36. Agresti A: *Categorical data analysis*, vol. 482: John Wiley & Sons; 2003.
37. Molenberghs G, Verbeke G: *Models for discrete longitudinal data*: New York Springer; 2005.
38. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G: *Longitudinal data analysis*. New York: Chapman&Hall /CRC Press; 2009.
39. Christensen RHB, Christensen MRHB: *Package 'ordinal'*. *Stand* 2015, 19:2016.
40. Touloumis A: *R package multgee: A generalized estimating equations solver for multinomial responses*. *arXiv preprint arXiv:14105232* 2014.
41. Nakai M, Ke W: *Review of the methods for handling missing data in longitudinal data analysis*. *International Journal of Mathematical Analysis* 2011, 5(1):1–13.
42. Burton A, Altman DG, Royston P, Holder RL: *The design of simulation studies in medical statistics*. *Statistics in medicine* 2006, 25(24):4279–4292.
43. Grund S, Lüdtke O, Robitzsch A: *Multiple imputation of missing data for multilevel models: Simulations and recommendations*. In: *Organizational Research Methods*. vol. 21; 2018: 111–149.
44. Gachau S, Ayieko P, Gathara D, Mwaniki P, Ogero M, Akech S, Maina M, Agweyu A, Oliwa J, Julius T: *Does audit and feedback improve the adoption of recommended practices? Evidence from a longitudinal observational study of an emerging clinical network in Kenya*. In: *BMJ Global Health*. vol. 2; 2017: e000468.
45. Gachau S, Owuor N, Njagi EN, Ayieko P, English M: *Analysis of hierarchical routine data with covariate missingness: Effects of Audit & Feedback on clinicians' prescribed paediatric pneumonia care in Kenyan hospitals*. *Frontiers in Public Health* 2019, 7:198.
46. Bartlett JW, Seaman SR, White IR, Carpenter JR, Initiative* AsDN: *Multiple imputation of covariates by fully conditional specification: accommodating the substantive model*. *Statistical methods in medical research* 2015, 24(4):462–487.
47. Blough DK, Ramsey S, Sullivan SD, Yusen R: *The impact of using different imputation methods for missing quality of life scores on the estimation of the cost-effectiveness of lung-volume-reduction surgery*. *Health economics* 2009, 18(1):91–101.
48. O'Keefe AG, Farewell DM, Tom BD, Farewell VT: *Multiple imputation of missing composite outcomes in longitudinal data*. *Statistics in biosciences* 2016, 8(2):310–332.

Tables

Table 1: Documentation of pneumonia care process among children admitted with pneumonia during the trial period.

PAQC score domain	Indicator	Documented cases (%)	Missingness (%)	Variable type
1. Assessment domain				
Primary signs & symptoms	cough	2118/2127 (99.6)	9/2127(0.4)	factor (2 levels)
	difficult breathing	2114/2127(99.4)	13/2127(0.6)	factor (2 levels)
Secondary sign & symptoms	oxygen saturation	1297/2127(60.9)	830/2127(39.1)	continuous
	ability to drink	2127/2127 (100)	0 (0.0)	factor (2 levels)
	central cyanosis	2127/2127 (100)	0 (0.0)	factor (4 levels)
	AVPU ^b	2110/2127 (99.3)	17/2127(0.7)	factor (4 levels)
	grunting	2127/2127 (100)	0 (0.0)	factor (2 levels)
	respiratory rate	1889/2127(88.8)	238/2127(11.19)	continuous
	indrawing	2123/2127(99.8)	4/2127(0.2)	factor (2 levels)
2. Diagnosis and classification	classified pneumonia cases	2127/2127(100)	0 (0.0)	factor (2 levels)
	correct classification (pneumonia)	1473/2127(69.3)		
3. Treatment domain	amoxicillin prescribed	2062/2127(96.94)	65/2127 (3.06)	factor (2 levels)
	yes	1036/2062(50.2)		
	no	1026/2062(49.8)		
	amoxicillin dosage	1032/1036(99.6)	4/1036(0.4)	continuous
	weight	1006/1036(97.1)	30/1036(2.9)	continuous
	amoxicillin frequency of administration	1009/1036(97.4)	27/1036 (2.6)	factor (4 levels)

^aPAQC score: Paediatric Admission Quality of Care Score,

^bAVPU :-Alert, Voice, Pain, Unresponsive

Table 2: Parameter estimates before and after multiple imputation of missing covariates and PAQC^a score components.

Effect	Random intercepts model				GEE ^b Model			
	Complete case analysis		Multilevel MI ^c		Complete case analysis		Multilevel MI	
	N=1619 (76.1%)	N=2127 (100%)	N=1619 (76.1%)	N=2127 (100%)	N=1619 (76.1%)	N=2127 (100%)	N=1619 (76.1%)	N=2127 (100%)
	Estimate (SE ^d)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
PAQC score intercept 0	Reference	-	Reference	-	Reference	-	Reference	-
PAQC score intercept 1	-7.77(1.076)	<0.001	-7.74 (0.829)	0.000	-7.41 (1.032)	<0.001	-7.19(0.794)	0.00
PAQC score intercept 2	-1.77(0.383)	<0.001	-2.2(0.341)	0.000	-1.65 (0.332)	<0.001	-2.03(0.314)	0.00
PAQC score intercept 3	-0.65(0.379)	0.03	-1.14(0.336)	0.001	-0.72 (0.329)	0.03	-1.07(0.308)	0.00
PAQC score intercept 4	0.48(0.379)	0.12	0.11(0.334)	0.740	0.19 (0.336)	0.561	-0.04(0.306)	0.91
PAQC score intercept 5	1.89(0.384)	<0.001	1.41(0.337)	0.000	1.3 (0.334)	<0.001	1.03 (0.308)	0.00
PAQC score intercept 6	2.86(0.388)	<0.001	2.38(0.339)	0.000	2.15 (0.342)	<0.001	1.87(0.308)	0.00
Age-group:12-59 months	0.19(0.099)	0.04	0.20(0.086)	0.019	-0.18(0.086)	0.04	0.24(0.079)	<0.001
Child sex: Males	-0.03(0.096)	0.773	0.01(0.084)	0.925	0.01(0.084)	0.958	-0.02(0.073)	0.82
Comorbidities: 0	0.47(0.231)	0.042	0.42(0.201)	0.034	0.31(0.209)	0.136	0.39(0.186)	0.03
Comorbidities :1	0.46(0.232)	0.047	0.30(0.201)	0.132	0.29(0.213)	0.174	0.22(0.187)	0.23
Comorbidities :2	0.48(0.243)	0.049	0.33(0.211)	0.116	0.30(0.209)	0.145	0.26(0.187)	0.16
Clinicians' sex: female	0.42(0.184)	0.023	0.31(0.169)	0.068	0.45(0.179)	0.011	0.33(0.168)	0.07
Clinicians' cadre: MO ^e	0.020(0.186)	0.913	0.05(0.167)	0.787	-0.19(0.161)	0.242	-0.21(0.151)	0.16
Hospital workload: low	-0.08(0.201)	0.705	-0.33(0.166)	0.045	0.22(0.178)	0.226	0.37(0.153)	0.01
Malaria prevalence: low	-0.05(0.198)	0.793	-0.20(0.172)	0.25	-0.19(0.189)	0.322	0.02(0.167)	0.91
Time (months)	0.05(0.043)	0.223	0.01(0.036)	0.75	-0.01(0.037)	0.856	-0.03(0.034)	0.44
Enhanced A&F ^f arm	-1.71(0.333)	<0.001	-1.59(0.294)	<0.001	-2.07(0.334)	<0.001	-1.96(0.304)	<0.001
Time× Enhanced A&F	0.14(0.063)	0.025	0.19(0.053)	<0.0001	0.25(0.06)	<0.001	0.27(0.052)	<0.001
Variance between random clinician's intercepts	1.328(1.151)		1.161(1.073)					

^aPaediatric Admission Quality of Care Score

^bGeneralized Estimating Equations

^cMultiple imputation

^dStandard error,

^eMedical Officer

^fAudit and feedback

Table 3: Performance measures of regression coefficients after multiple imputation of covariates and outcome elements: MAR^a mechanism.

Effect	True est ^b .	Proportion Missing											
		3%				10%				40%			
		Bias	Model-based SE	Emp SE ^c	MSE ^d	Bias	Model-based SE	Emp SE	MSE	Bias	Model-based SE	Emp SE	MSE
PAQC score intercept 1	-7.825	0.018	0.015	0.017	0.005	0.021	0.023	0.024	0.001	0.022	0.036	0.036	0.002
PAQC score intercept 2	-2.253	-0.616	0.030	0.030	0.380	-0.707	0.449	0.450	0.701	-0.739	0.153	0.154	0.569
PAQC score intercept 3	-1.189	-0.327	0.080	0.080	0.113	-0.375	0.319	0.321	0.242	-0.392	0.414	0.414	0.325
PAQC score intercept 4	0.083	-0.241	0.582	0.583	0.400	-0.277	0.186	0.188	0.111	-0.289	0.142	0.143	0.104
PAQC score intercept 5	1.371	-0.264	0.237	0.237	0.126	-0.303	0.373	0.374	0.231	-0.317	0.249	0.249	0.162
PAQC score intercept 6	2.246	-0.041	0.131	0.135	0.020	-0.047	0.203	0.203	0.043	-0.049	0.229	0.228	0.055
Age-group:12-59	0.154	0.038	0.083	0.088	0.008	0.044	0.147	0.148	0.023	0.046	0.170	0.172	0.032
Child sex: males	-0.046	-0.027	0.216	0.216	0.047	-0.031	0.259	0.260	0.068	-0.032	0.276	0.277	0.078
Comorbidities: 0	0.474	-0.130	0.228	0.230	0.069	-0.149	0.014	0.015	0.022	-0.156	0.070	0.071	0.029
Comorbidities: 1	0.309	-0.134	0.330	0.333	0.129	-0.154	0.557	0.558	0.334	-0.161	0.644	0.645	0.442
Comorbidities: 2	0.335	-0.111	0.384	0.386	0.163	-0.127	0.570	0.570	0.341	-0.133	0.643	0.644	0.431
Clinicians' sex: female	0.337	-0.03	0.020	0.020	0.002	-0.05	0.027	0.019	0.002	-0.08	0.016	0.018	0.003
Clinicians' cadre: MO ^e	0.038	0.062	0.155	0.156	0.028	0.071	0.054	0.055	0.008	0.074	0.014	0.016	0.006
Hospital workload: low	-0.367	-0.063	0.147	0.150	0.025	-0.072	0.250	0.252	0.068	-0.075	0.290	0.292	0.090
Malaria prevalence: low	-0.189	0.159	0.306	0.302	0.119	0.183	0.170	0.171	0.063	0.191	0.275	0.276	0.112
Enhanced A&F ^f	-0.002	-0.065	0.192	0.193	0.041	-0.053	0.189	0.190	0.038	-0.060	0.180	0.182	0.036
Time (months)	-1.754	0.015	0.720	0.721	0.518	0.017	0.696	0.698	0.484	0.018	0.686	0.687	0.470
Time* Enhanced A&F	0.226	-0.028	0.160	0.163	0.026	-0.032	0.106	0.108	0.012	-0.034	0.126	0.126	0.017

^aMissing at Random

^bTrue estimate

^cEmpirical standard error

^dMSE:- Mean Square Error

^eMO:- Medical Officer

^fA&F:-Audit and feedback

Table 4: Performance measures of regression coefficients after multiple imputation of covariates: MAR^a mechanism

Effect		Proportion Missing												
		True est ^b .	3%				10%				40%			
			Bias	Model-based SE	Emp SE ^c	MSE ^d	Bias	Model-based SE	Emp SE	MSE	Bias	Model-based SE	Emp SE	MSE
PAQC score intercept 1	-7.825	0.141	0.015	0.016	0.020	0.171	0.023	0.024	0.03	0.484	0.036	0.037	0.236	
PAQC score intercept 2	-2.253	-0.386	0.031	0.031	0.150	-0.697	0.450	0.451	0.688	-0.808	0.154	0.154	0.677	
PAQC score intercept 3	-1.189	-0.736	0.080	0.081	0.548	-0.8	0.322	0.323	0.743	-0.9	0.414	0.415	0.982	
PAQC score intercept 4	0.083	-0.542	0.584	0.586	0.637	-0.665	0.187	0.188	0.477	-0.798	0.143	0.145	0.657	
PAQC score intercept 5	1.371	-0.594	0.236	0.238	0.409	-0.727	0.370	0.374	0.669	-0.805	0.254	0.255	0.71	
PAQC score intercept 6	2.246	-0.092	0.135	0.136	0.027	-0.113	0.204	0.205	0.054	-0.186	0.232	0.233	0.088	
Age-group:12-59	0.154	0.086	0.084	0.085	0.015	0.106	0.146	0.148	0.033	0.175	0.173	0.173	0.061	
Child sex: males	-0.046	-0.061	0.216	0.217	0.051	-0.074	0.260	0.261	0.074	-0.122	0.278	0.279	0.092	
Comorbidities: 0	0.474	-0.293	0.230	0.230	0.139	-0.358	0.015	0.017	0.128	-0.593	0.071	0.071	0.357	
Comorbidities: 1	0.309	-0.302	0.334	0.335	0.203	-0.37	0.557	0.558	0.448	-0.612	0.643	0.646	0.791	
Comorbidities: 2	0.335	-0.25	0.389	0.390	0.215	-0.305	0.571	0.571	0.419	-0.505	0.642	0.644	0.67	
Clinicians' sex: female	0.337	-0.007	0.021	0.022	0.001	-0.007	0.017	0.018	0.011	-0.011	0.017	0.018	0.001	
Clinicians' cadre: MO ^e	0.038	0.14	0.156	0.157	0.044	0.17	0.055	0.056	0.032	0.281	0.015	0.016	0.079	
Hospital workload: low	-0.367	-0.142	0.148	0.149	0.042	-0.173	0.251	0.252	0.093	-0.285	0.291	0.292	0.166	
Malaria prevalence: low	-0.189	0.358	0.307	0.307	0.222	0.439	0.172	0.172	0.222	0.726	0.276	0.277	0.603	
Enhanced A&F ^f	-0.002	-0.005	0.193	0.194	0.038	-0.008	0.189	0.190	0.036	-0.017	0.181	0.182	0.033	
Time (months)	-1.754	0.034	0.720	0.721	0.521	0.041	0.696	0.697	0.488	0.068	0.687	0.688	0.477	
Time* Enhanced A&F	0.226	-0.063	0.161	0.162	0.030	-0.077	0.110	0.111	0.017	-0.129	0.127	0.127	0.033	

^aMAR:-Missing at Random

^bTrue estimate

^cEmpirical standard error

^dMSE:- Mean Square Error

^eMO:- Medical Officer

^fA&F:-Audit and feedback

Figures

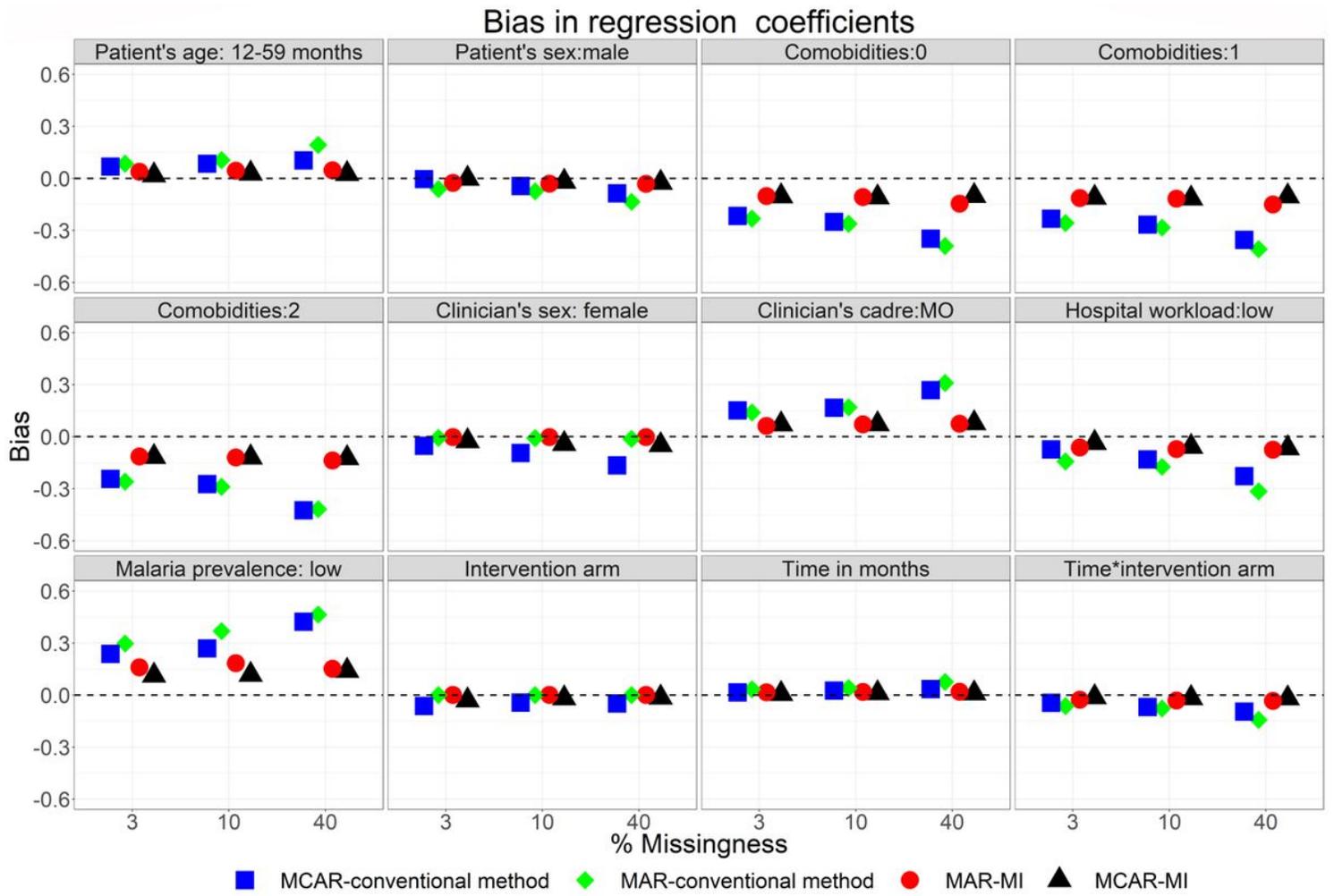


Figure 1

Bias in regression coefficients under the conventional approach of handling missing PAQC score components and after multiple imputation of missing PAQC score subcomponents in the treatment domain and missing covariates imputed across missing data rates and missing data mechanisms.

Bias in standard errors

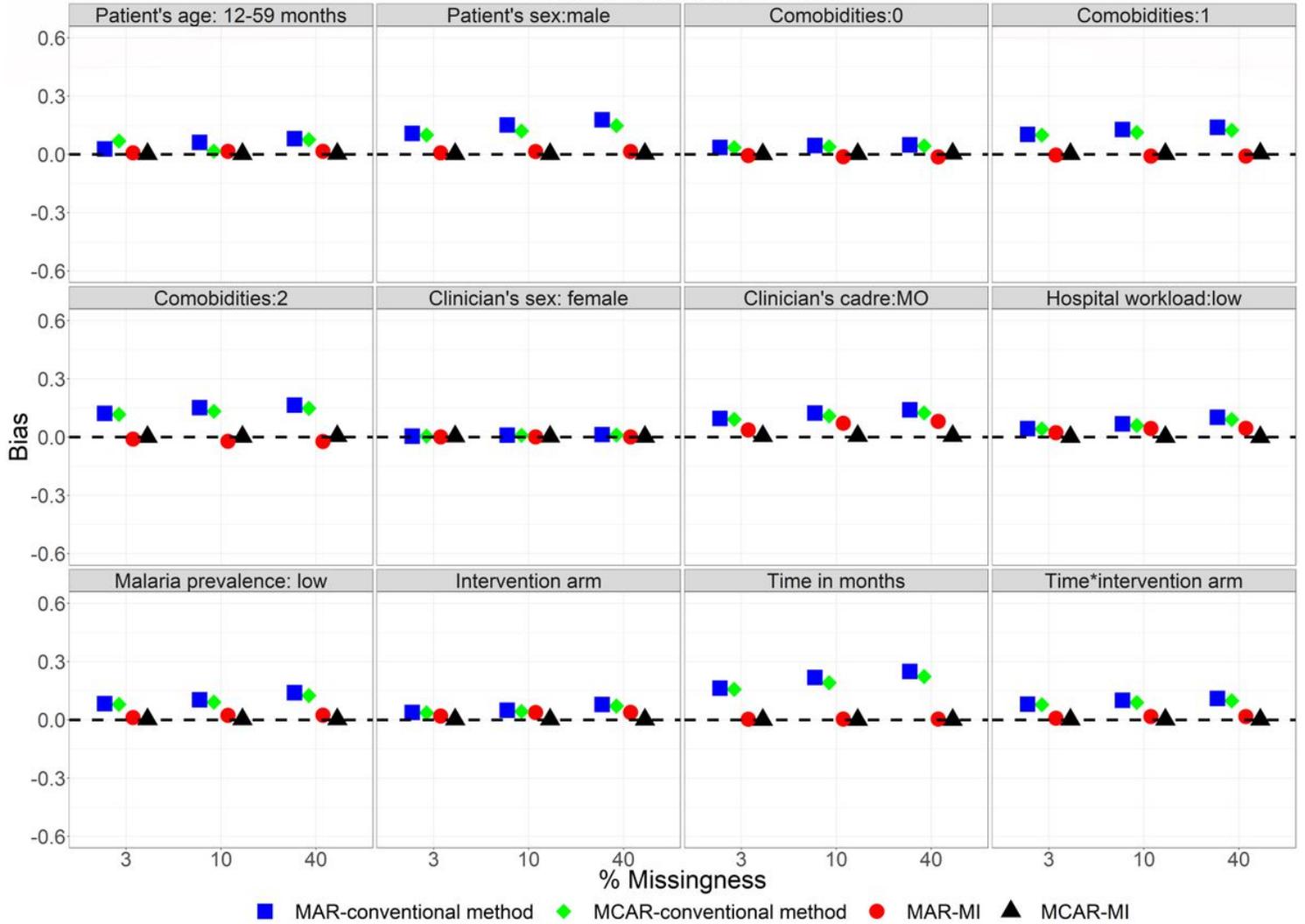


Figure 2

Bias in standard errors under the conventional approach of handling missing PAQC score components and after multiple imputation of missing PAQC score subcomponents in the treatment domain and missing covariates imputed across missing data rates and missing data mechanisms.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [eq3.5.jpg](#)
- [eq2.jpg](#)
- [eq3.jpg](#)
- [eq4.5.jpg](#)
- [Supplementaryfile.docx](#)
- [eq5.jpg](#)
- [eq1.2.jpg](#)
- [eq6.jpg](#)
- [eq1.4.jpg](#)
- [eq1.3.jpg](#)
- [eq1.jpg](#)

- eq4.jpg