Handling missing data in modelling quality of clinician-prescribed routine care: Sensitivity analysis of departure from Missing at Random (MAR) assumption

Susan Gachau[1,2*], Matteo Quartagno[3], Edmund Njeru Njagi[4], Nelson Owuor[2], Mike English[1,5,] Philip Ayieko[6,7].

[1]Health Services Unit, Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi Kenya

[2]School of Mathematics, University of Nairobi, Kenya

[3]Institute of Clinical Trials and Methodology, University College London, London, United Kingdom

[4]Department of Non-Communicable Disease Epidemiology, London School of Hygiene

and Tropical Medicine, London, United Kingdom

[5]Nuffield Department of Medicine, University of Oxford, United Kingdom

[6]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

[7] Mwanza Intervention Trials Unit, Mwanza, Tanzania

Corresponding author:
Susan Gachau: sgachau@kemri-wellcome.org

Abstract

Missing information is a major drawback in analysing data collected in many routine health care settings. Multiple imputation (MI) assuming a missing at random (MAR) mechanism is a popular method to handle missing data. The MAR assumption cannot be confirmed from the observed data alone, hence the need for sensitivity analysis to assess robustness of inference. However, sensitivity analysis is rarely conducted and reported in practice. We analysed routine paediatric data collected during a cluster randomized trial conducted in Kenyan hospitals. We imputed missing patient and clinician-level variables assuming the MAR mechanism. We also imputed missing clinician-level variables assuming a missing not at random (MNAR) mechanism. We incorporated opinions from 15 clinical experts in the form of prior distributions and shift parameters in the delta adjustment method. An interaction between trial intervention arm and follow-up time, hospital, clinician and patient-level factors were included in a proportional odds random-effects analysis model. We performed these analyses using R functions derived from the *jomo* package. Parameter estimates from MI under the MAR mechanism were similar to MI estimates assuming the MNAR mechanism. Our inferences were insensitive to departures from the MAR assumption using either the prior distributions or shift parameters sensitivity analysis approach.

Key words: *Elicitation, Multiple imputation, Missing at random, Missing not at random, Sensitivity analysis, Routine data.*

Word count:

Abstract: 200

Main text: 8148

1 Introduction

Routine health data are increasingly used in monitoring quality of patient care in low and middle income countries.[1-5] However, concerns about quality of routine data including completeness and accuracy limit their use in decision making.[6] To alleviate bias, multiple imputation (MI), a method for handling missing data which repeatedly draws from a model to create multiple completed data sets, is often recommended in the missing data literature.[7-9] Standard MI relies on the assumption that the probability of data being missing is independent of the missing observations conditional on the observed data. This assumption is known as the missing at random (MAR) mechanism.[7, 10] On the other hand, if the probability of a value being missing depends on unobserved data, even after conditioning on all the available information, then data are said to be missing not at random (MNAR).[7, 9] In practice, MAR and MNAR mechanisms cannot be distinguished using observed data only,[9, 11, 12] hence the need for sensitivity analyses.[7, 9, 13] Sensitivity analyses entail scrutinizing plausible models assuming MNAR mechanisms to assess departures from the MAR assumption; the primary analysis model is changed through a number of alterations and the stability of inferences across the alternative settings assessed.[9, 14-16]

Broadly, sensitivity analyses following MI can be conducted within three generic frameworks, namely pattern-mixture models, selection models and shared parameter models. [7, 9, 13, 15-17] Nonetheless, sensitivity analysis within any of these frameworks is rarely reported in practice. This is because it is a computationally complex procedure

which involves defining and examining suitable assumptions for a given data set under analysis.[15, 18] Besides, sensitivity analysis methods are underdeveloped in standard statistical software thus limiting their application in practice.[15]

In health care settings, completeness of routine data depends on an interplay of factors that operate at the patient, clinician and healthcare facility levels.[19] For example, missing data at facility level could result from temporary breakdown of medical devices (e.g. blood pressure machine or pulse oximeter) within a healthcare facility leading to absence of diagnostic investigations in that facility during the breakdown period. At the clinician level, individual attributes such as professional qualification, experience and behaviour can influence quality of care, and its documentation, therefore impacting the quality of routine data.[20] Separately, clinician-level factors are rarely captured within routine health data generated in low income countries and hence clinician effect is often overlooked in studies reporting clinician-prescribed routine care.[4, 5, 21] This problem of missing data at the clinician level is compounded when missing data are handled using inappropriate methods such as complete case analysis (CCA), that increase the risk of obtaining biased and inefficient estimates, hence misleading inference.[12, 22] Furthermore, in most studies for which the primary analysis was based on complete case records, MI assuming MAR mechanism was used as a sensitivity analyses tool .[12] However, similarities between CCA and MI results could lead to false reassurances that data are either Missing completely at random (MCAR)

or missing at random with a mechanism not involving the outcome (i.e. covariate-dependent MAR [23]) whereas a MNAR mechanism could be in operation.[12]

To address this gap, we analysed partially observed paediatric routine data collected in 12 Kenyan hospitals during a cluster randomized trial. Specifically, we imputed missing data assuming MAR while appropriately accounting for the hierarchical structure of the data set. We then conducted sensitivity analyses aimed at assessing robustness of inference under MAR mechanism using two approaches within the pattern-mixture model framework. In one approach, we imputed missing data under the MAR mechanism and then used random draws from prior distributions to create MNAR imputed values.[18] In the second approach, we modified the imputation model assuming MAR mechanism through a range of sensitivity parameters (delta adjustment approach) to ensure multiple imputation of missing data under the MNAR assumption.[7, 24] Missing data were imputed within the joint modelling MI framework.[25]

The remainder of the paper is structured as follows: In Section 2 we introduce the data used in the analysis, before presenting multiple imputation methods under the MAR and MNAR mechanisms in Sections 3 and 4 respectively. Section 5 presents the analysis of imputed data using proportional odds model followed by results in Section 6. We conclude with a discussion in Section 7.

2 Motivating data

2.1 Study design

Data used in the analysis were from a cluster randomized trial conducted in 12 county-level hospitals in Kenya between March 2016 and November 2016. The trial was embedded within an ongoing observational study known as the Clinical Information Network (CIN).[26, 27] Details of the trial are described elsewhere.[28, 29] The trial aimed to examine the effect of an audit and feedback intervention on uptake of recommendations contained in revised World Health Organization (WHO) treatment guidelines for childhood pneumonia.[30] Hospitals were randomly allocated to receive enhanced (six hospitals) or standard (six hospitals) audit and feedback. The six hospitals in the enhanced audit and feedback (A&F) arm received monthly report on assessment, classification and treatment of pneumonia cases in addition to a bi-monthly standard audit and feedback report on general inpatient paediatric routine care and network intervention strategies.[28, 29] On the other hand, the six hospitals in standard A&F arm received bi-monthly standard audit and feedback report on general inpatient paediatric routine care and network intervention strategies.[28, 29]

Children aged between two and 59 months admitted to hospital with pneumonia signs and symptoms were eligible for enrolment into the trial. Overall, 2299 children met the inclusion criteria in all participating hospitals.[28] Trained data clerks abstracted data from individual patient medical records after patients were discharged from hospital. The data were entered directly from the medical record into an open source data capture

tool (Research Electronic Data Capture (REDcap)[31] using a standard operating procedure manual. The details of admitting clinicians including a unique clinician identifier, gender and cadre were entered in a separate database compiled in each hospital. Clinician cadre refers to a clinician's qualification depending on the level of training, that is, clinical officer for a clinician holding diploma-level training, equivalent to physician assistant; and medical officer for medical doctors holding bachelor's degree training. The two databases (patient and clinician database) were linked by unique clinician identifier present in both databases. Of the 2299 pneumonia cases, 2127 (92.4%) were admitted by 378 different clinicians. On average, each hospital had 32 clinicians with a standard deviation of nine. The number of admissions by individual clinician ranged between 3 and 46. The Kenyan 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.

2.2 Outcome: Paediatric Admission Quality of Care (PAQC) score

The outcome of interest in this study was quality of care measured using an ordinal composite measure known as the Paediatric Admission Quality of Care (PAQC) score.[32, 33] A summary of how we constructed PAQC score based on childhood pneumonia treatment guidelines recommended by the World Health Organization (WHO) in 2013[30] is presented in supplementary TableA1. Specifically, we created and summed 6 binary indicators spanning assessment, diagnosis (and severity

classification) and treatment domains of pneumonia care (supplementary Table A1). The assessment domain had three binary indicators. The first represented assessment and documentation of two primary signs and symptoms required for pneumonia identification. The second binary indicator represented assessment and documentation of seven secondary signs and symptoms required for pneumonia severity classification. The third binary indicator combined assessment and documentation of all primary and secondary signs and symptoms (supplementary Table A1). The second PAQC score domain entailed integration of information on presenting signs and symptoms by admitting clinician to correctly diagnose and classify pneumonia severity (i.e., severe pneumonia or pneumonia) (supplementary Table A1). The third PAQC score domain consisted of two binary indicators. The first one indicated whether oral amoxicillin was prescribed or not. The second one indicated whether oral amoxicillin was prescribed in line with WHO recommended guidelines.[30] To determine correctness of the dose, we first created a new variable "amoxicillin dose per kilogram body weight". That is, the actual dose given at point of care divided by patient's weight. The new variable was then transformed into a binary variable as outlined (supplementary Table A1). After summation of the six binary indicators, pneumonia PAQC score ranged between zero and six. A minimum score of zero corresponded to inappropriate pneumonia care while six represented complete compliance to recommended paediatric pneumonia management guidelines.

2.3 Covariates

Covariates of interest in this analysis included time (counted in months from inception of A&F intervention to time of individual participant's admission) and its interaction with intervention arm, hospital malaria prevalence status and hospital admission workload. At clinician level, gender and cadre were considered (cadre refers to clinician's level of training that is, clinical officers with diploma-level training and medical officers with a bachelor's degree level training). At patient level, we considered gender, number of comorbid illnesses and age at admission.

2.4 Missingness in the data

Missing data occurred both in the covariates as well components of the outcome (PAQC score components).

Approximately, 21.9% (83/378) and 21.7% (82/378) clinicians had missing data on the gender and cadre variables respectively, while patient's gender was missing in 0.7% (17/2127) case records. An assessment of the missing data pattern revealed that nearly all clinicians with observed cadre had gender observed as well. In the PAQC score (outcome) components, missing data occurred in nine subcomponents: six signs and symptoms in the assessment domain (primary and secondary), and three subcomponents in the treatment domain (Supplementary Table A2). The level of missingness in PAQC score components ranged between 0.4% and 39%. Our analysis of the data sought to impute missing covariates and PAQC score components in the treatment domain assuming a MAR assumption. In this analysis, we only addressed

missing PAQC score elements in the treatment domain. This domain had the following specific elements: patients' weight, amoxicillin dose prescribed and frequency of amoxicillin administration. In the pneumonia trial data, patients' weight was missing in 2.9% of case records. Among amoxicillin recipients, dose and frequency of administration were missing in 0.4% and 2.6% of the case records respectively (Supplementary Table A2). Undocumented signs and symptoms in the assessment domain were considered as inappropriate care and therefore scored zero in the construction of PAQC score (*Gachau et al., unpublished data*). Besides addressing missing covariates and outcome components using multiple imputation, we also conducted sensitivity analysis for two partially observed clinician-level variables, that is cadre and gender. Our aim was to evaluate robustness of the inferences through multiple imputation assuming MNAR mechanism.

3 Multiple imputations under MAR assumption

For the pneumonia trial data, we first imputed missing covariate and missing outcome components assuming a MAR mechanism**.** MI was conducted within the joint model imputation framework using *jomo* package in R (version 3.5.0).[34] Joint modelling imputation approach assumes that the data can be described by a multivariate normal distribution from which imputations for all variables are drawn jointly using a single statistical imputation model.[24] The partially observed variables of interest in this study were a mix of categorical and continuous variables. Categorical variables were imputed using the latent normal approach.[7] In a multilevel data context, partially

observed variable at each level of the hierarchy are jointly specified as responses in multilevel structural equations of the imputation model. For instance, considering the i^{th} pneumonia patient attended by clinician *j* in hospital *l*, our multilevel level joint imputation model corresponded to

$$Y_{i,j,l}^{(1)} = X_{ijl}^{(1)}\beta^{(1)} + b_{jl}^{(1)} + e_{ijl}^{(1)} \qquad (1)$$
$$Y_{jl}^{(2)} = X_{jl}^{(2)}\beta^{(2)} + b_{jl}^{(2)}$$

$$e_{ijl} \sim N(0, \sigma_e^2), \;\; and \;\; (b_{jl}^{(1)}, b_{jl}^{(2)}) \sim N(0, \Sigma_b)$$

where $\boldsymbol{Y_{ijl}^{(1)}}$ is a vector of partially observed patient-level variables (i.e., patient's gender, weight, amoxicillin dose prescribed and frequency of amoxicillin administration) and $\boldsymbol{Y_{jl}^{(2)}}$ is a vector of partially observed clinician-level variables (i.e., clinician's gender and cadre). Predictor variables $\left(\boldsymbol{X_{ijl}^{(1)}}\right)$ of missing patient's gender included fully observed follow-up time and its interaction with feedback arm, hospital admission4workload and hospital malaria prevalence status, patient's age, number of comorbid illnesses and PAQC score components in the assessment and diagnosis domains. Besides fully observed covariates above, we also include PACQ score (outcome) subcomponents in the imputation model as level 1 predictors. These included a binary indicator variable representing completeness of documentation of 2 primary signs and symptoms, a binary indicator variable denoting completeness of documentation of 7 secondary signs and symptoms. We also included diagnosis and

classification, amoxicillin prescription indicators in the diagnosis and treatment domains respectively. Level two predictors $\left(\boldsymbol{X}_{\boldsymbol{jl}}^{(\mathbf{2})}\right)$ for missing clinicians' gender and cadre included follow-up time and its interaction with feedback arm, hospital admission workload and hospital malaria prevalence status. Column vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$denote level one and level two fixed effects respectively. Clinician random intercepts (b_{jl}) were included to account for clustering at clinician level and to ensure compatibility with the analysis model of interest. We created 20 imputed data sets under each imputation model.

4 Multiple imputations under MNAR assumption: Sensitivity analyses

We then imputed missing data assuming MNAR mechanism to assess possible departures from MAR mechanism. Our analyses focused on missing clinicians' cadre and gender in the second level of the hierarchical structure using two approaches within the pattern-mixture model (PMM) framework. In this study we considered MNAR imputation in level two variables (i.e., clinician's gender and cadre) while retaining the MAR imputation models for level one variables (patient-level variables) for two reasons. First, we aimed to minimize complexities at analysis stage considering that threes out of four level patient-level variables (i.e., patient's weight, amoxicillin dose prescribed and frequency of amoxicillin administration) were subcomponent of a composite outcome. Secondly, the proportion of missing data in patient-level

variables was much lower (< 4%) compared to the much higher proportion (>20%) of missing data observed in clinician-level variables.

In one approach, we replaced clinicians' gender and cadre imputed assuming MAR mechanism with random draws using appropriate prior distributions creating MNAR imputed data sets.[18] In the second approach, we modified the multiple imputation model assuming MAR mechanism through a range of sensitivity parameters (delta adjustment approach).[7, 24] These changes can be informed by opinion elicited from experts in the subject matter or contextual knowledge.[9]

4.1 Pattern mixture models

Suppose $\boldsymbol{Y}$ (representing both response and independent variables) is an $N \times p$ matrix denoting a hypothetical data set containing p variables $\left(j = 1,..., p\right)$ for the i^{th} study subject, $\left(i = 1, 2, 3, .., N\right)$. For each study subject, $\boldsymbol{Y}_i$ can be partitioned into observed and missing components denoted by $\boldsymbol{Y}_i^{obs}$ and $\boldsymbol{Y}_i^{miss}$ respectively. Further suppose a missingness indicator R_i takes the value 1 when $\boldsymbol{Y}_i$ is observed and 0 when $\boldsymbol{Y}_i$ is missing. When the data are potentially MNAR then the mechanism generating missing data cannot be ignored[17]. In this case, the joint models for (Y_i, R_i) should be considered. The joint model can be factorised within the pattern mixture models (PMM), selection model, or shared-parameter models. In this study, we considered factorization within the PMM framework.

The PMM assumes that observations are stratified based on patterns of missing data, and distinct models formulated to estimate parameters within each pattern.[9, 24, 35] However, since the distribution of the outcome given patterns of non-response is unidentifiable, the conditional distributions under MAR is used as a starting point and then appropriate changes reflecting MNAR assumption are made.[24]

4.2 Elicitation of experts' opinion

In this study, we elicited clinical experts' opinions and used them to define suitable MNAR assumptions about the differences in the distribution of clinicians with observed cadre/gender and clinicians with missing cadre/gender. Our investigations into missing data patterns showed that nearly all clinicians with missing cadre had missing gender (Supplementary file, Figure A1). Further assessment revealed that intervention arm and paediatric admission workload were predictive variables for both missingness and observed values of clinician's cadre and gender. Therefore, we defined

$$k = \begin{cases} 1 \text{ if hospital is in the control arm and has high paeditric admission workload} \\ 2 \text{ if hospital is in the control arm and has low paeditric admission workload} \\ 3 \text{ if hospital is in the intervention arm and has high paeditric admission workload} \\ 4 \text{ if hospital is in the intervention arm and has low paeditric admission workload} \end{cases} \quad (2)$$

For each k, we estimated data predicted probabilities of a clinician belonging to a particular cadre (i.e., clinical officers, clinical officer interns, medical officers or medical officer interns) under the MAR assumption.[18] Specifically, we imputed missing clinicians' cadre and gender jointly assuming MAR mechanism. In this

imputation model, we included trial arm and admission workload as predictor variables. Inclusion of trial arm and admission workload as the only predictor variables in the imputation model followed preliminary results above. Thereafter, we separately regressed clinicians' cadre on trial arm and admission workload using a multinomial logistic model.

The final estimates (log odds) pooled according to Rubin's Rule [10] were then used to determine data predicted probabilities of clinicians belonging to either of the 4 cadre categories for each k (see Supplementary file). Similarly, we fitted a logistic regression model for clinicians' gender with trial arm and admission workload as covariates and determined data predicted probabilities of clinicians being males or females. Data predicted probabilities (P_{jk}) for clinicians' cadre (Supplementary Table A3) and clinicians' gender (Supplementary Table A4) were then presented to experts in the form of questionnaires in face to face interviews. Fifteen clinical experts (three clinical officers, five clinical officer interns, three medical officers, and four medical officer interns) from paediatric wards in two CIN hospitals participated in the elicitation exercise. The experts were briefed about the purpose of the exercise before filling their predicted probability of clinicians with missing cadre being either clinical officers, clinical officer interns, medical officers or medical officer interns. Similarly, they filled in their belief about clinicians with missing gender being males or females in

each k (Supplementary Table A4). Here we denote expert predicted probability for gender/cadre by $(\theta_j k)$.

After the elicitation exercise, we pooled the expert predicted probabilities by calculating the mean $(E[\theta_{jk}])$ and variances $(Var[\theta_{jk}])$ for every cadre/gender category in k. This information was then used to approximate parameters of Dirichlet and beta distributions from which missing clinicians' cadre and gender were imputed assuming a MNAR mechanism. The parameters for the respective prior distributions were approximated using the methods of moments as explained in the following section.

4.2.1 Dirichlet conjugate prior for multinomial distribution

For clinicians' cadre with four categories we chose a Dirichlet distribution as an appropriate conjugate prior distribution.[18] A Dirichlet distribution with four parameters is formulated as

$$f(x_{1k}, x_{2k}, x_{3k}, x_{4k}, \alpha_{1k}, \alpha_{2k}, \alpha_{3k}, \alpha_{4k}) = \frac{\Gamma(\sum_{j=1}^{4} \alpha_{jk})}{\prod_{j=1}^{4} \Gamma(\alpha_{jk})} \prod_{i=1}^{k} x_{jk}^{\alpha_{jk-1}} \qquad (3)$$

$$\alpha_{jk} \text{ and } \sum_{j=1}^{4} x_{jk} = 1$$

where the vector x_{jk} denotes probabilities for different categories in the variable of interest and α_{jk} are concentration parameters. The mean and variance of Dirichlet distribution are denoted by

$$E(x_{jk}) = \frac{\alpha_{jk}}{L_k} \qquad (4)$$

and

$$Var(X_{jk}) = \frac{\alpha_{jk}(L_k - \alpha_{jk})}{L_k^2(L_k+1)} \quad (5)$$

where $Var(X_{jk}) = \frac{\alpha_{jk}(L_k - \alpha_{jk})}{L_k^2(L_k+1)}$.

Using the means and variances of experts' predicted probabilities $(E[\theta_{jk}])$ and $(Var[\theta_{jk}])$ for the j^{th} cadre $(j = 1,2,3,4)$ in each combination of trial arm and admission workload $(k = 1,2,3,4)$, we estimated Dirichlet distribution concentration parameters using the methods of moments[18] as follows:

Step 1: Using a sequence of values between 1 and 50 (L_k) and the mean of experts predicted probabilities $(E[\theta_{jk}])$ to approximate unknown Dirichlet mean $E(x_{jk})$ we estimated the concentration parameters (α_{jk}) of a Dirichlet distribution in (equation 3) using

$$\alpha_{jk} = L_k * E(\theta_{j,k}) \quad (6)$$

Step 2: We substituted α_{jk} values obtained in step 1 in the variance formulae (equation 5) to estimate Dirichlet distribution variances $Var(x_{jk})$ for each value in the sequenceL_k.

Step 3: We plotted Dirichlet distribution variance $Var(x_{jk})$ approximated in step 2 against the sequence L_k and superimposed a horizontal line corresponding to variance of expert predicted probabilities $(Var[\theta_{jk}])$. For instance, in $k = 1$, we had four plots, one for each clinicians' cadre (i.e., clinical officers, clinical officer interns, medical officers

and medical officer interns) (Figure 1). The step was repeated for the other combinations of the trial arm and paediatric admission workload $(k = 2,3,4)$ and the corresponding figures are presented in the Supplementary file (Figures A1-A3).

Step 4: We determined the value in the sequence L_k for which estimated Dirichlet variance $Var(x_{jk})$ (black curve) and variance of experts' predicted probabilities $(Var[\theta_{jk}])$ (red line) intersected (or were approximately equal) for a given cadre. We summed L_k values across the four cadres and divided the total by four. The mean was denoted by $E(L_k)$.

Step 5: We determined Dirichlet distribution parameters for the j^{th} cadre in each by multiplying expert predicted mean probabilities $E(L_k)$.

$$\hat{\alpha}_{jk} = E(L_k) * E(\theta_{j,k}) \quad (7)$$

We used approximated $\hat{\alpha}_{jk}$ vector of parameters to generate random vectors of probabilities from a Dirichlet distribution. Estimated concentration parameters for Dirichlet distribution for a given k are presented in supplementary Table A5. The parameter vectors were used to generate random vectors of probabilities of j^{th} cadre probabilities in each k.

Insert [**Figure 1]**

4.2.2 Beta conjugate prior for the binomial distribution

For clinicians' gender with two levels we considered a beta distribution conjugate prior. A beta distribution is formulated as

$$f(x) = \frac{x^{\alpha_{jk}-1}(1-x)^{\beta_{jk}-1}}{B(\alpha_{jk},\beta_{jk})} \quad (8)$$

where $B(\alpha_{jk},\beta_{jk}) = \frac{\Gamma\alpha_{jk}\Gamma\beta_{j,k}}{\Gamma\alpha_{jk}+\alpha_{j,k}}$, $\alpha_{jk} > 0$ and $\beta_{jk} > 0$.

Using the mean ($E[\theta_{jk}]$) and variances ($Var[\theta_{jk}]$) of experts predicted probabilities for jth (j=1,2) gender category in the kth stratum ($k = 1,2,3,4$), we estimated α_{jk} and β_{jk} using the moments method [36] as shown below

$$\hat{\beta}_{jk} = \frac{E[\theta_{jk}](1-E[\theta_{jk}])^2}{Var[\theta_{jk}]} + E[\theta_{jk}] - 1 \quad (9)$$

$$\hat{\alpha}_{jk} = \frac{E[\theta_{jk}]*\hat{\beta}_{jk}}{(1-E[\theta_{jk}])} \quad (10)$$

The approximated $\hat{\alpha}_{jk}$ and $\hat{\beta}_{jk}$ parameters for each k $(k = 1,2,3,4)$ are presented in Supplementary Table A6. The parameters were used to generate random probabilities for female clinicians in k^{th} stratum. We drew 20 random probabilities of a clinician being female. In each draw, the probability of being a male clinician was 1 minus the probability of being a female clinician.

4.3 Multiple imputations from MNAR prior distributions

Using estimated Dirichlet and beta prior parameters vectors (Supplementary Tables A4 and A5), we generated 20 random probability vectors for each k $(k=1,2,3,4)$. The number of random draws i.e., 20 corresponded to the number of imputations. Each imputed data set was split into four mutually exclusive strata defined by k (k=1,2,3,4). The j^{th} probability value in the i^{th} random vector (i=1, 2,…,20) was then used to determine the proportion of occurrence of clinicians' cadre/gender category in the k^{th} stratum (here j= 1 denotes clinical officers, j= 2 for clinical officer interns, j= 3 for medical officers and j= 4 for medical officer interns while for clinicians' gender, j=1 denotes females and j=2 denotes males). After drawing values for clinician gender/cadre from the probability vectors, the four strata (k=1,2,3,4) were merged into one data set. This step was repeated for all the imputed data sets before fitting the analysis model of interest.

4.4 Multiple imputation with shift parameters (delta adjustment method)

Multiple imputation with delta adjustment involves adding a fixed quantity δ to the linear predictor of the imputation model.[7, 22, 24, 37] For continuous target variables, δ represents the difference in mean between non-respondents and respondents.[17] When the variable of interest is categorical, addition of shift parameter δ in the imputation model modifies the predicted probabilities for the classification levels [7, 17, 22] thus producing MNAR imputed values.[24] In this study, we conducted separate MI-MNAR

analyses for clinicians' gender and clinicians' cadre rather than two dimensional sensitivity analysis. In the first multilevel joint imputation model, we modified the probability of classification among clinicians with missing gender while missing clinicians' cadre was imputed without any modifications (i.e., multiple imputation assuming MAR). In the second imputation model, the shift parameter modified the probability of classification in the imputation of clinicians with missing cadre while missing clinicians' gender was imputed without any modification. We performed these analyses using R functions derived from the *jomo* package in R (version 3.5.0).[34] These functions are not yet available in the version of the package available in CRAN, but will be included in the near future. Our modified multilevel joint imputation model is formulated as follows:

$$\begin{aligned} Y_{ijl}^{(1)} &= X_{ijl}^{(1)}\beta^{(1)} + b_{jl}^{(1)} + e_{ijl}^{(1)} \\ Y_{jl}^{(2)} &= X_{jl}^{(2)}\beta^{(2)} + \delta\left(1 - R_{jl}\right) + b_{jl}^{(2)} \end{aligned} \tag{11}$$

$$e_{ijl} \sim N(0, \sigma_e^2), \;\; and \;\; (b_{jl}^{(1)}, b_{jl}^{(2)}) \sim N(0, \Sigma_b)$$

where $\boldsymbol{Y_{ijl}^{(1)}}$ is a vector of partially observed level 1 variables (i.e., patient's gender, weight, amoxicillin dose prescribed and frequency of amoxicillin administration) at level one of the hierarchical structure. The vector of clinicians' gender and cadre at level two of the hierarchical structure is denoted by $\boldsymbol{Y_{jl}^{(2)}}$ while R_{jl} is a binary indicator

with value 1 if clinicians' gender/cadre is observed and 0 if missing. When δ is 0, a MAR mechanism is implied.[7]

To determine a set of shift parameters for clinicians' gender with two levels, we used latent normal variables which is equivalent to modelling binary data with a probit link. Specifically, we obtained the quartiles of the prior distribution for the proportion of female clinicians, and chose values of the latent normal corresponding to these quartiles values. We chose three shift parameters (i.e., δ = -0.2, -0.3, -0.5) to alter probability of classification in the imputation of clinicians' gender. The negative shift parameters decreased the latent normal for female clinicians on the probit scale. As such clinicians with missing gender were more likely to be imputed as males. The same values used to alter classification probabilities for clinicians' gender were also used to alter classification probabilities among clinicians with missing cadre.

In this case, negative shift parameters increased the probability of being medical officers and medical officer interns, by decreasing latent normal for clinical officer (interns) on the probit scale. Therefore, clinicians with missing cadre were more likely to be imputed as medical officers (interns). The MI-MNAR analysis under the delta-adjusted approach was repeated for different shift parameters. The differences in proportion of classification increased with an increase in the magnitude of shift parameters.

5 Statistical analysis

After MI assuming MAR and MNAR mechanism (i.e., with delta adjustment and from appropriate prior distribution), we constructed PAQC score in each imputed data set following the procedure outlined in section 2.2. For each imputed data set, we fitted the proportional odds random intercepts[13] model below

$$logit\left[P\left(Y_{(PAQC\ Score;\ i,j,l)} \leq m\right)\right] = \alpha_m + \beta_1 x_{(age\ group;\ ijl)} + \beta_2 x_{(patient\ sex;\ ijl)} + \beta_3 x_{(commobidity;\ ijl)} + \beta_4 x_{(clinician\ cadre;\ jl)} + \beta_5 x_{(clinician\ sex;\ jl)} + \beta_6 x_{(admission\ workload;\ l)} + \beta_7 x_{(malaria\ prevalence;\ l)} + \beta_8 x_{(time\ in\ months;\ l)} * x_{(trial\ arm;\ l)} + b_{jl} \quad (12)$$

where α_m, m=1,2,3,4,5,6 are PAQC score specific intercepts, *i* indexes the patient, and *j* and *l* index clinician and hospital respectively. The intercepts denote thresholds distinguishing adjacent PAQC score levels. The fixed effect parameters $\beta's$, are common across all m-1 cumulative logits[13] and they denote proportional odds ratios of individual variables on PAQC score holding all other variables in the model constant. Clinician random intercepts are denoted by b_{jl}. The analysis models were fitted using *ordinal package* [38] functions in R version 3.5.0. We combined MI estimates using Rubin's rules and compared inferences under MAR and MNAR mechanisms.

We also compared MI results with those obtained under complete case analysis which was based on 77.1 % (1639/2127) observations after deletion of case records with missing data in patient and clinician level variables.

6 Results

Table 1 presents a summary of both data predicted probabilities and experts' predicted probabilities (mean and variance) for the four cadre categories in each combination of trial arm and admission workload. Experts' opinions predicted higher probabilities of medical officers and clinical officers compared to data predicted probabilities. Furthermore, elicited opinion suggested that medical officers were more likely in hospitals with high paediatric admission workload compared to hospitals with low admission workload (Table 1). With regard to clinicians' gender, experts' opinions suggested that among clinicians with missing gender, males were more likely in high workload hospitals than in low admission hospitals in each k (Table 1). In both clinicians' gender and cadre, experts' responses did not vary widely across stratification groups $(k = 1, 2, 3, 4)$.

Table 1: Data predicted and expert predicted probabilities (mean and variance) for clinicians' cadre.

k	Data predicted probabilities under MAR[a] ($p_{j,k}$)	Mean(variances) of experts predicted probabilities $E(\theta_{j,k})$ ($Var(\theta_{j,k})$
Clinicians' cadre		
1: Control arm and high workload		
Clinical officer interns	0.38	0.12 (0.08)
Clinical officers	0.01	0.14 (0.10)
Medical officer interns	0.60	0.49 (0.12)
Medical officer	0.01	0.25 (0.09)
2: Control arm and Low workload		
Clinical officer interns	0.45	0.17 (0.12)
Clinical officers	0.03	0.39 (0.11)
Medical officer interns	0.50	0.29 (0.10)
Medical officer	0.02	0.15 (0.05)
3: Intervention arm and high workload		
Clinical officer interns	0.42	0.23 (0.05)
Clinical officers	0.01	0.23 (0.09)
Medical officer interns	0.55	0.22 (0.06)
Medical officer	0.02	0.31 (0.08)
4: Intervention arm and low workload		
Clinical officer interns	0.50	0.25 (0.04)
Clinical officers	0.01	0.25 (0.12)
Medical officer interns	0.47	0.31(0.06)
Medical officer	0.02	0.19 (0.05)
Clinicians' gender		
1: Control arm and high workload		
Females	0.47	0.45(0.02)
Males	0.53	0.55 (0.06)
2: Control arm and Low workload		
Females	0.36	0.54(0.04)
Males	0.64	0.46(0.07)
3: Intervention arm and high workload		
Females	0.57	0.44(0.06)
Males	0.46	0.56(0.08)

4: Intervention arm and low workload		
Females	0.42	0.52(0.05)
Males	0.58	0.48(0.10)

MAR: -Missing at Random

Table 2 shows the distribution of clinicians' cadre and gender under complete case analysis and under MAR and MNAR mechanisms. When clinicians' cadre was the variable of interest in the sensitivity analysis, we observed a systematic increase in the proportion of clinicians imputed as medical officers and medical officer interns. On the other hand, when clinician gender was the variable of interest, more clinicians were imputed as males compared to females. For clinicians' cadre, the proportions of medical officer tended to increase with an increasing magnitude of sensitivity parameter (delta values). Similarly, the proportion of male clinicians increased with an increasing magnitude of sensitivity parameter. Furthermore, we observed similarities in the proportions of clinicians' gender and clinicians' cadre after multiple imputation from prior distributions and delta adjustment with a sensitivity parameter equal to -0.2 (Table 2). Considering the small number of clinical officers and medical officers in comparison to interns in the respective cadres, we grouped clinicians into two categories in subsequent analysis, i.e. clinical officers and clinical officer interns as one group, and medical officers and medical officer interns as the other group.

1 Table 2: Percentage of clinicians' cadre and gender in complete records and under multiple imputation under MAR and MNAR mechanisms.

			Sensitivity analysis variable: clinicians' cadre				**Sensitivity analysis variable: clinicians' gender**			
	Complete records	MI-MAR[a]	MI-MNAR[b]				MI-MNAR			
			δ = -0.2	δ = -0.3	δ = -0.5	Dirichlet prior	δ = -0.2	δ = -0.3	δ = -0.5	Beta prior
Clinician cadre										
Clinical officers	0.52	1.05	0.55	0.60	0.69	1.58	0.69	0.68	0.88	0.64
Clinical officer interns	39.80	43.58	40.31	39.59	36.59	39.19	44.47	43.53	44.38	45.34
Medical officers	2.62	2.62	3.62	4.17	4.51	4.71	2.87	2.88	2.62	2.63
Medical officer interns	57.05	52.74	55.53	55.64	58.33	54.53	51.97	52.91	52.11	51.38
Clinician gender										
Males	58.61	57.34	58.31	56.44	55.79	57.33	60.21	61.26	63.7	60.34
Females	41.39	42.66	41.69	43.56	44.21	42.67	39.79	38.74	36.3	39.66

2 [a]MI-MNAR- multiple imputation assuming Missing Not at Random, [b]MI-MAR- multiple imputation assuming Missing at Random.

Complete case analysis (CCA), MI results assuming MAR mechanism and MI results assuming MNAR mechanism (i.e. MI with delta adjustment over a range of parameters and MI from appropriate conjugate prior distributions) for clinicians' cadre and gender are presented in Table 3 and Table 4 respectively.

After multiple imputation assuming MAR mechanism, enhanced audit and feedback led to improve uptake of new pneumonia paediatric guideline over time. For example, considering a patient admitted in an intervention hospital (enhanced audit and feedback arm), the odds of PAQC score=1 versus PAQC score ≥ 2 were 1.22 (*95% CI*: 1.04-1.358) times higher the odds of a patients admitted in a control hospital, for a unit increase in follow-up time and holding other variables at reference levels (Table 3/Table 4). Similar observations were made under complete case analysis but the magnitude of effect was smaller and characterized by a slightly wider 95% confidence interval.

The study results also exhibited contrasting results before and after multiple imputation for selected variables. For instance, adjusting for other variables, the odds of PAQC score=1 versus PAQC score ≥ 2 for a patient admitted by female clinician were 1.52 (95% CI: 1.05 to 2.18) times higher the odds of patient admitted by a male clinician (Table 3/Table 4). However, after MI assuming MAR mechanism, the odds ratio and the corresponding 95% confidence interval (i.e., 0R=1.37 (95% CI: 0.977 to 1.912))

did not suggest difference between male and female clinicians in the odds of PAQC score=1 versus PAQC score ≥ 2.

To assess stability of parameter estimates under MI assuming MAR mechanism, we imputed missing clinicians' cadre (Table 3) and clinicians' gender (Table 4) assuming MNAR mechanism. Our study results showed that the odds ratios and the corresponding 95% CI under MI assuming MNAR mechanism were close to those obtained under MI assuming MAR mechanism. Moreover, the magnitude and direction of effects were comparable after multiple imputation with the delta adjustment method and multiple imputation based on appropriate prior distributions. The similarities in parameter estimates were more apparent for $\delta = -0.2$.

When we added shift parameters in the imputation of missing clinicians' cadre (delta adjustment method) we observed some changes in clinicians' cadre effect (adjusting for other variables) whereas the odds ratios and the 95% CI for other variables remained more or less the same. Specifically, the effect of clinicians' cadre (adjusted odd ratio) changed from 1.05 (*95% CI*: 0.735 to 1.421) under MI assuming MAR mechanism to 1.02 (*95% CI*: 0.740 to 1.460) and 1.01 (*95% CI*: 0.741 to 1.461) for $\delta = -0.3$ and $\delta = -0.5$ respectively (Table 3). Similarly, replacing imputed clinicians' cadre with random draws from a prior Dirichlet distribution, the adjusted odds ratio decreased to 1.04 (*95% CI*: 0.719 to 1.464) (Table 3). Nevertheless, the observed shifts changes in magnitude did not change the conclusion.

After imputing clinicians' gender with shift parameters (i.e., delta adjustment), the estimated clinicians' gender effect remained close to that observed under MI assuming MAR except for MI-MNAR with $\delta = -0.5$ where the odds ratio changed from 1.37 (*95% CI:* 0.977 to 1.912) to 1.46 (*95% CI:* 0.989 to 2.313). Likewise, replacing imputed clinicians' gender with random draws from a prior beta distribution, the adjusted odds ratio for clinicians' gender changed to 1.37 (*95% CI:* 0.975 to 1.857) (Table 4). Despite the changes in magnitude of effect, the inference remained the same. With regard to variability between admitting clinicians, complete case analysis led to larger variance between clinicians compared to that estimated under MI assuming MAR and MNAR respectively (Table 3 and 4).

Table 3: Adjusted odds ratios and corresponding 95% confidence intervals under complete case analysis and under MI assuming MAR and MNAR mechanisms respectively: Clinicians' cadre probabilities adjusted using shift parameters (δ) under delta adjustment methods. MAR imputed clinicians' cadre replaced with draws from a Dirichlet prior distribution.

	Complete case analysis	**MI-MAR[a]**	**MI-MNAR[b] δ =-0.2**	**MI-MNAR δ =-0.3**	**MI-MNAR δ =-0.5**	**MI-MNAR (Dirichlet prior)**
Effect	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
PAQC score intercept 0	Ref	Ref	Ref	Ref	Ref	Ref
PAQC score intercept 1	0.002 (0.001, 0.003)	0.002 (0.001, 0.004)	0.002 (0.001, 0.004)	0.002 (0.001, 0.004)	0.002 (0.001, 0.004)	0.002 (0.001, 0.004)
PAQC score intercept 2	0.20 (0.092, 0.458)	0.03 (0.01, 0.076)	0.03 (0.01, 0.079)	0.03 (0.01, 0.079)	0.03 (0.01, 0.079)	0.02 (0.007, 0.062)
PAQC score intercept 3	0.63 (0.283, 1.397)	0.08 (0.028, 0.221)	0.08 (0.029, 0.229)	0.08 (0.029, 0.229)	0.08 (0.029, 0.229)	0.06 (0.021, 0.171)
PAQC score intercept 4	1.94 (0.874, 4.325)	0.27 (0.097, 0.759)	0.28 (0.101, 0.785)	0.28 (0.101, 0.785)	0.28 (0.101, 0.785)	0.21 (0.074, 0.599)
PAQC score intercept 5	5.99 (3.567, 7.935)	1.02 (0.364, 2.864)	1.06 (0.376, 2.964)	1.06 (0.376, 2.964)	1.06 (0.376, 2.964)	0.77 (0.27, 2.196)
PAQC score intercept 6	2.16 (9.342, 7.916)	2.56 (0.909, 7.194)	2.64 (0.937, 7.444)	2.64 (0.937, 7.444)	2.64 (0.937, 7.444)	1.83 (0.641, 5.24)
Age:12-59 months	1.20 (0.991, 1.464)	1.19 (1.010, 1.410)	1.19 (1.011, 1.411)	1.19 (1.011, 1.411)	1.19 (1.011, 1.411)	1.20 (1.011, 1.428)
Child gender: Males	0.97 (0.806, 1.174)	0.99 (0.842, 1.166)	0.99 (0.844, 1.168)	0.99 (0.844, 1.168)	0.99 (0.844, 1.168)	0.97 (0.820, 1.15)
Comorbidities: 0	1.59 (1.015, 2.513)	1.51 (1.029, 2.219)	1.51 (1.029, 2.22)	1.51 (1.029, 2.22)	1.51 (1.029, 2.22)	1.50 (1.016, 2.226)
Comorbidities :1	1.59 (1.005, 2.498)	1.34 (0.91, 1.974)	1.34 (0.911, 1.977)	1.34 (0.911, 1.977)	1.34 (0.911, 1.977)	1.33 (0.877, 1.928)
Comorbidities :2	1.61 (1.001, 2.591)	1.38 (0.929, 2.076)	1.39 (0.93, 2.078)	1.39 (0.93, 2.078)	1.39 (0.93, 2.078)	1.35 (0.897, 2.033)
Clinician gender: female	1.52 (1.057, 2.183)	1.37 (0.977, 1.912)	1.37 (0.981, 1.931)	1.39 (0.985, 2.11)	1.35 (0.892, 1.951)	1.37 (0.973, 1.937)
Clinician Cadre: MO[c]	1.02 (0.709, 1.468)	1.05 (0.735, 1.421)	1.04 (0.741, 1.462)	1.02 (0.740, 1.460)	1.01 (0.740, 1.461)	1.04 (0.719, 1.464)
Hospital workload: low	0.93 (0.624, 1.376)	0.73 (0.531, 1.02)	0.74 (0.535, 1.025)	0.74 (0.535, 1.025)	0.74 (0.535, 1.025)	0.74 (0.526, 1.04)

Malaria prevalence: low	0.95 (0.644, 1.40)	0.87 (0.588, 1.151)	0.87 (0.606, 1.185)	0.87 (0.606, 1.185)	0.84 (0.606, 1.185)	0.86 (0.61, 1.226)
Time (months)	1.05 (0.969, 1.145)	1.01 (0.941, 1.083)	1.01 (0.943, 1.085)	1.01 (0.943, 1.085)	1.01 (0.943, 1.085)	0.99 (0.927, 1.074)
Enhanced A&F[d] arm	0.18 (0.095, 0.349)	0.19 (0.109, 0.345)	0.19 (0.108, 0.340)	0.19 (0.108, 0.340)	0.19 (0.108, 0.341)	0.18 (0.101, 0.334)
Time* Enhanced A&F	1.15 (1.018, 1.307)	1.22 (1.104, 1.358)	1.23 (1.107, 1.362)	1.23 (1.107, 1.362)	1.23 (1.107, 1.362)	1.24 (1.112, 1.379)
Variance between random clinician's intercepts	1.32(1.151)	1.16(1.07)	1.16(1.07)	1.16(1.07)	1.16(1.07)	1.16(1.07)

[a]MI-MAR: -Multiple imputation assuming Missing at Random, [b]MI-MNAR: -Multiple imputation assuming Missing at Not Random,
[c]MO: -Medical officers, [d]A&F: -Audit and feedback

Table 4: Adjusted odds ratios and corresponding 95% confidence intervals under complete case analysis and under MI assuming MAR and MNAR mechanisms respectively: Clinicians' gender probabilities adjusted using shift parameters (δ) under delta adjustment methods and imputed clinicians' gender (under MAR) replaced with draws from a beta prior distribution.

	Complete case analysis	**MI-MAR[a]**	**MI-MNAR[b] δ =-0.2**	**MI-MNAR δ =-0.3**	**MI-MNAR δ =-0.5**	**MI-MNAR (Beta prior)**
Effect	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
PAQC score intercept 0	Ref	Ref	Ref	Ref	Ref	Ref
PAQC score intercept 1	0.002 (0.001, 0.003)	0.002 (0.001, 0.004)	0.002 (0.001, 0.004)	0.002 (0.001, 0.004)	0.002 (0.001, 0.004)	0.002 (0.001, 0.004)
PAQC score intercept 2	0.20 (0.092, 0.458)	0.03 (0.01, 0.076)	0.03 (0.010, 0.076)	0.03 (0.01, 0.077)	0.03 (0.01, 0.079)	0.02 (0.008, 0.061)
PAQC score intercept 3	0.63 (0.283, 1.397)	0.08 (0.028, 0.221)	0.08 (0.028, 0.221)	0.08 (0.028, 0.223)	0.08 (0.029, 0.229)	0.07 (0.024, 0.178)
PAQC score intercept 4	1.94 (0.874, 4.325)	0.27 (0.097, 0.759)	0.27 (0.097, 0.758)	0.274(0.098, 0.766)	0.28 (0.101, 0.785)	0.23 (0.083, 0.611)
PAQC score intercept 5	5.99 (3.567, 7.935)	1.02 (0.364, 2.864)	1.02 (0.364, 2.861)	1.03 (0.368, 2.892)	1.06 (0.376, 2.964)	0.85 (0.313, 2.304)
PAQC score intercept 6	2.16 (9.342, 7.916)	2.56 (0.909, 7.194)	2.56 (0.909, 7.186)	2.58 (0.918, 7.264)	2.64 (0.937, 7.444)	2.12 (0.779, 5.787)
Age:12-59 months	1.20 (0.991, 1.464)	1.19 (1.010, 1.410)	1.19 (1.010, 1.411)	1.19 (1.010, 1.411)	1.19 (1.011, 1.411)	1.19 (1.011, 1.412)
Child gender: Males	0.97 (0.806, 1.174)	0.99 (0.842, 1.166)	0.99 (0.843, 1.168)	0.99 (0.843, 1.168)	0.99 (0.844, 1.168)	0.99 (0.842, 1.167)
Comorbidities: 0	1.59 (1.015, 2.513)	1.51 (1.029, 2.219)	1.51 (1.028, 2.218)	1.51 (1.031, 2.223)	1.51 (1.029, 2.22)	1.51 (1.03, 2.222)
Comorbidities :1	1.59 (1.005, 2.498)	1.34 (0.91, 1.974)	1.34 (0.909, 1.973)	1.34 (0.911, 1.977)	1.34 (0.911, 1.977)	1.34 (0.910, 1.975)
Comorbidities :2	1.61 (1.001, 2.591)	1.38 (0.929, 2.076)	1.38 (0.928, 2.074)	1.39 (0.93, 2.079)	1.39 (0.93, 2.078)	1.38 (0.929, 2.076)
Clinician gender: female	1.52 (1.057, 2.183)	1.37 (0.977, 1.912)	1.37 (0.962, 1.891)	1.37 (0.971, 2.026)	1.46 (0.989, 2.313)	1.37 (0.975, 1.857)
Clinician Cadre: MO[c]	1.02 (0.709, 1.468)	1.05 (0.735, 1.421)	1.03 (0.729, 1.453)	1.04 (0.718, 1.402)	1.04 (0.741, 1.461)	1.03 (0.741, 1.423)
Hospital workload: low	0.93 (0.624, 1.376)	0.73 (0.531, 1.02)	0.73 (0.53, 1.016)	0.74 (0.533, 1.022)	0.74 (0.535, 1.025)	0.73 (0.527, 1.012)
Malaria prevalence: low	0.95 (0.644, 1.40)	0.87 (0.588, 1.151)	0.87 (0.597, 1.169)	0.86 (0.603, 1.181)	0.86 (0.606, 1.185)	0.86 (0.578, 1.139)

Time (months)	1.05 (0.969, 1.145)	1.01 (0.941, 1.083)	1.01 (0.942, 1.084)	1.01 (0.942, 1.084)	1.01 (0.943, 1.085)	1.01 (0.94, 1.082)
Enhanced A&F[d] arm	0.18 (0.095, 0.349)	0.19 (0.109, 0.345)	0.19 (0.108, 0.342)	0.19 (0.108, 0.339)	0.19 (0.108, 0.340)	0.19 (0.11, 0.347)
Time* Enhanced A&F	1.15 (1.018, 1.307)	1.22 (1.104, 1.358)	1.22 (1.106, 1.361)	1.22(1.107, 1.362)	1.23 (1.107, 1.362)	1.22 (1.103, 1.357)
Variance between random clinician's intercepts	1.32(1.151)	1.16(1.07)	1.16(1.07)	1.16(1.07)	1.16(1.07)	1.16(1.07)

[a]MI-MAR: -Multiple imputation assuming Missing at Random, [b]MI-MNAR: -Multiple imputation assuming Missing at Not Random, [c]MO: -Medical officers

[d]A&F: -Audit and feedback

7 Discussion

In this study we sought to address missing data in a multilevel routine data context and to conduct sensitivity analyses to assess stability and robustness of inference under assumed MAR mechanism. This work was motivated by data collected among paediatric inpatient admission receiving routine paediatric care in a group of Kenyan hospitals. Missing data occurred in patient and clinician-level covariates, as well as pneumonia care indicators used to construct a composite measure for quality of care - PAQC score. To handle missingness, we used complete case analysis and multiple imputation methods. As expected, CCA analysis led to estimates with wider 95% confidence intervals (due to larger standard errors) compared to MI under MAR mechanism given that MI makes use of all the available information. Complete case analysis or list wise deletion is the default technique for handling missing data in most statistical software hence its wide use in practice. [12] A major drawback of CCA is loss of power particularly for data sets with multiple partially observed variables.[15] Furthermore, there is potential for biased estimates when complete case records are not a random sample of the population being studied.[9] For this reason, inference under MI assuming MAR mechanism is often preferred. However, the MAR assumption cannot be ascertained using the data alone. Therefore, we conducted sensitivity analyses within the pattern mixture models.[7, 9] The focus of our sensitivity analyses was clinician-level variables in the second level of the hierarchical structure. In order to

87 define suitable assumptions reflecting MNAR missing data mechanism[9] in the two
88 variables of interest, we elicited and incorporated experts' opinions into the analysis.
89 Specifically, we interviewed 15 clinical experts in paediatrics wards in two study
90 hospitals and incorporated their opinions into our sensitivity analysis using two
91 approaches. In the first approach, we incorporated uncertainty about the missing data
92 mechanism in the form of conjugate prior distributions. gender
93 In the second approach, we incorporated experts' opinion in the form of shift
94 parameters within the delta adjustment method. Although this approach is a transparent
95 and flexible means by which to impute data under MNAR mechanisms, the choice of
96 appropriate sensitivity parameters is less straightforward.[7, 24] In this study, we utilized
97 elicited probabilities combined with additional information probed from experts during
98 interview sessions in the choice of sensible shift parameters. According to experts'
99 contextual knowledge, hospitals with high workload were more likely to be teaching
100 and referral hospitals, hence more medical officers and medical officer interns.
101 Furthermore, experts' opinions indicated that there are more male medical
102 officers/interns than female medical officers/interns, compared to the observed data.
103 Therefore, clinicians with missing information in high workload hospitals were more
104 likely to be male medical officers/interns than female medical officers/interns. In our
105 analysis, we implemented experts' opinion over a range of 3 shift parameters (i.e., -0.2,
106 -0.3 and -0.5). The shift parameters altered the probabilities with which the multilevel
107 joint imputation model imputed missing clinicians' cadre and gender. Furthermore, the

degree of departure from MAR assumption was the same for individuals with missing clinicians' cadre and gender. This was in consideration of experts' beliefs that departures from MAR assumptions would be similar for the two clinician level variables.

From the study results, parameter estimates (i.e., odds ratios and corresponding 95% confidence intervals) estimates under MI assuming MNAR scenarios were close to those from the analysis under MAR. The similarities were an indication of robust inferences under MAR assumptions. For delta adjusted over a range of parameter we observed slight increase/decrease in magnitude of clinicians' cadre and gender effects. However, these changes did not lead to changes in inference and conclusions. More importantly, the effect of enhanced A&F over follow-up time remained stable across a range of MNAR scenarios. In the event that conclusions differ between CCA and MI-MAR, it could mean that either CCA is wrong (outcome dependent MAR) or that MI is wrong (covariate dependent MNAR) or both are wrong (outcome dependent MNAR). When the mechanism is covariate-dependent MNAR (i.e., it does not depend on the outcome), then CCA is valid and in this case it can be better than MI assuming MAR mechanism.[39]

Strengths and implications of the study

Through this study, we have demonstrated application of two sensitivity analysis approaches in multilevel routine data contexts incorporating experts' opinion. The sensitivity analyses methods adopted in this study have been used and reported in

previous studies.[7, 15, 18, 22, 24] In our case we apply the approaches to multilevel data compared to single level data used in previous analyses. A key difference between the two sensitivity analyses methods is that one provides several inferences based on specified sensitivity parameters (i.e., MI with delta adjustment method) while the other provides a single inference based on informative prior distributions (i.e., MI from prior distribution). In spite of these differences, parameter estimates were comparable between the two sensitivity analyses methods. A possible explanation for the similarities could be the fact that both methods utilized same experts' opinions to create differences between MAR and MNAR imputed values in the variables of interest. Therefore, we recommend both methods as guiding examples for conducting sensitivity analyses within the pattern mixture model framework, rather than prescribe how every sensitivity analysis in the multilevel data setting should be conducted. Moreover, more studies are needed to examine the performance of the two methods in a range of simulation scenarios.

In this study, we elicited experts' opinions in face to face interviews, which allowed us to probe for additional information and clarifications not captured in the questionnaires. We therefore recommend face to face interviews. In instances where face to face interviews are impractical, telephone discussions or electronic questionnaires can be considered.[9] When imputing from prior distributions, the choice of a conjugate prior should be informed by the distribution of the variable under analysis. However, in situations where prior knowledge is difficult to elicit, delta adjustment method with

tipping-point analysis can be a valuable alternative.[22, 40] Tipping-point analysis allows one to explore sensitivity parameters across a wide range of values in order to determine a set of sensitivity parameters for which inference and conclusions change.[37] In this study, we applied the delta adjustment method within the pattern mixture framework and combined estimates across the imputed data sets using Rubin's rule. A recent study by Tang (2017) evaluated the extent of bias associated with used of Rubin's variance estimator under the delta-adjusted pattern mixture models (PMMs) and control-based PMM. From the study results, bias of MI variance was found to be negligible in the delta-adjusted PMM but substantial in the control-based PMM context.[41] The study results further showed that inference based on Rubin's rule in the delta-adjusted PMM was approximately valid. [41] For this reason, we only reported estimates based on Rubin's rule.[10] The alternative asymptotic sampling variance estimator suggested by Tang (2017) can be considered in future studies.[41]

Limitations

This study was limited in several ways. Firstly, we interviewed 15 clinical experts in two study sites due to time and cost constraints, on top of refusal by some of the respondents to fill in the questionnaires. Secondly, we only imputed clinicians' cadre and gender under MNAR mechanism while patient-level variables were imputed assuming MAR mechanism. Moreover, we conducted separate MI-MNAR analysis for clinicians' cadre and gender instead of a two dimensional sensitivity analysis. This because eliciting experts' opinions for the two variables jointly would have been

complicated and more difficult to implement. Thirdly, although our data had clustering at hospital (n=12) and clinician level (n=378), we only accounted for clinicians' random effect in our analysis model while hospital characteristics were included as fixed effects. This was because, while we wanted to ensure compatibility between analysis and imputation models, statistical software used could accommodate random effects only at the second level of hierarchy. Moreover, our outcome variable (the PAQC score) was a composite outcome, and we imputed for it by imputing and combining its components. This approach may not be fully compatible with the analysis model. To the best of our knowledge, more work is still needed on the best way to impute for composite outcomes in multilevel settings, to assure compatibility between imputation and substantive models in that setting. Nevertheless, multiple imputation of missing PAQC score components at item level has been shown to produce less biased estimates compared to the conventional approach where all missing PAQC score components are scored with zero at construction stage *(Gachau et al., unpublished data*).

Conclusion

In conclusion, sensitivity analysis is useful in ascertaining robustness of inference under MAR assumption. We have demonstrated that eliciting and incorporating experts' opinions in form of prior distribution and shift parameters provides transparent and flexible means of assessing departures from the MAR assumption following multilevel MI. After multilevel MI of clinician level variables assuming MNAR, our

inferences were insensitive to departures from the MAR mechanism. These observations were made using both sensitivity analysis methods. That is, incorporating uncertainty about the missing data mechanism in the form of conjugate prior distributions and in the form of shift parameters within the delta adjustment method.

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 with support from MQ. Feedback on the analytic approach was provided by ENN, NO, 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 Thuranira (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. Gachau S, Ayieko P, Gathara D, et al. Does audit and feedback improve the adoption of recommended practices? Evidence from a longitudinal observational study of an emerging clinical network in Kenya. *BMJ Global Health* 2: e000468 (2017).
2. Akech S, Chepkirui M, Ogero M, et al. The clinical profile of severe paediatric malaria in an area targeted for routine RTS, S/AS01 malaria vaccination in Western Kenya. *Clinical Infectious Diseases*. 2019.
3. Oliwa JN, Gathara D, Ogero M, et al. Diagnostic practices and estimated burden of tuberculosis among children admitted to 13 government hospitals in Kenya: An analysis of two years' routine clinical data. *PloS one*. 2019; 14: e0221145.
4. Thomas J, Ayieko P, Ogero M, et al. Blood Transfusion Delay and Outcome in County Hospitals in Kenya. *Am J Trop Med Hyg* 96: 511-7 (2017, accessed Feb 08).
5. Gachau S, Irimu G, Ayieko P, Akech S, Agweyu A and English M. Prevalence, outcome and quality of care among children hospitalized with severe acute malnutrition in Kenyan hospitals: A multi-site observational study. *PloS one*. 2018; 13: e0197607.
6. Organization WH. Improving data quality: a guide for developing countries. Manila: WHO Regional Office for the Western Pacific, 2003.
7. Carpenter JR and Kenward MG. *Multiple imputation and its applications*. Chichester: John Wiley & Sons, 2013.
8. Molenberghs G and Kenward M. *Missing data in clinical studies*. John Wiley & Sons, 2007.
9. Tsiatis AA, Kenward MG, Fitzmaurice G, Verbeke G and Molenberghs G. *Handbook of missing data methodology*. Chapman and Hall/CRC, 2014.
10. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63: 581-92.
11. Kenward MG and Carpenter J. Multiple imputation: current perspectives. *Statistical methods in medical research*. 2007; 16: 199-218.
12. Mackinnon A. The use and reporting of multiple imputation in medical research–a review. *Journal of internal medicine*. 2010; 268: 586-93.
13. Molenberghs G and Verbeke G. *Models for discrete longitudinal data*. New York Springer, 2005.
14. Fitzmaurice G, Davidian M, Verbeke G and Molenberghs G. *Longitudinal data analysis*. New York: Chapman&Hall /CRC Press, 2009.
15. Héraud-Bousquet V, Larsen C, Carpenter J, Desenclos J-C and Le Strat Y. Practical considerations for sensitivity analysis after multiple imputation applied to epidemiological studies with incomplete data. *BMC medical research methodology*. 2012; 12: 73.
16. Liublinska V and Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. *Statistics in medicine*. 2014; 33: 4170-85.
17. Little RJ, D'agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*. 2012; 367: 1355-60.

18. Smuk M, Carpenter J and Morris T. What impact do assumptions about missing data have on conclusions? A practical sensitivity analysis for a cancer survival registry. *BMC medical research methodology*. 2017; 17: 21.
19. Yelin E, Yazdany J, Tonner C, et al. Interactions between patients, providers, and health systems and technical quality of care. *Arthritis care & research*. 2015; 67: 417-24.
20. Rowe AK, de Savigny D, Lanata CF and Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *The Lancet* 366: 1026-35 (2005).
21. Tuti T, Agweyu A, Mwaniki P, Peek N and English M. An exploration of mortality risk factors in non-severe pneumonia in children using clinical data from Kenya. *BMC medicine*. 2017; 15: 201.
22. Tompsett DM, Leacy F, Moreno-Betancur M, Heron J and White IR. On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice. *Statistics in medicine*. 2018.
23. Fitzmaurice G, Davidian M, Verbeke G and Molenberghs G. *Longitudinal data analysis*. CRC Press, 2008.
24. Van Buuren S. *Flexible imputation of missing data*. Chapman and Hall/CRC, 2018.
25. Quartagno M and Carpenter J. jomo: A package for multilevel joint modelling multiple imputation. *R package version*. 2016: 2.-0.
26. English M. Designing a theory-informed, contextually appropriate intervention strategy to improve delivery of paediatric services in Kenyan hospitals. *Implementation Science* 8: 1 (2013).
27. Tuti T, Bitok M, Paton C, et al. 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. *Journal of the American Medical Informatics Association* 23: 184-92 (2016).
28. Ayieko P, Irimu G and 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. *Trials* 18: 416 (2017).
29. Ayieko P, Irimu G, Ogero M, et al. 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: 20.
30. Organization WH. *Pocket book of hospital care for children: guidelines for the management of common childhood illnesses*. 2nd ed. Geneva: World Health Organization, 2013.
31. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N and 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: 377-81.
32. Opondo C, Allen E, Todd J and 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: 1334-45.

33. Opondo C, Allen E, Todd J and 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: e203-e10.
34. Quartagno M, Carpenter J and Quartagno MM. Package 'jomo'. 2019.
35. Carpenter JR and Kenward MG. Multiple Imputation of Unordered Categorical Data. *Multiple Imputation and its Application*. 2013: 112-26.
36. Lunn D, Jackson C, Best N, Spiegelhalter D and Thomas A. *The BUGS book: A practical introduction to Bayesian analysis*. Chapman and Hall/CRC, 2012.
37. Yuan Y. Sensitivity analysis in multiple imputation for missing data. *Proceedings of the SAS Global Forum 2014 Conference:[http://support sas com/resources/papers/proceedings14/SAS270-2014 pdf]*. 2014.
38. Christensen RHB. ordinal—regression models for ordinal data. *R package version*. 2015; 28: 2015.
39. White IR and Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in medicine*. 2010; 29: 2920-31.
40. Leacy FP, Floyd S, Yates TA and White IR. Analyses of sensitivity to the missing-at-random assumption using multiple imputation with delta adjustment: Application to a tuberculosis/HIV prevalence survey with incomplete HIV-status data. *American journal of epidemiology*. 2017; 185: 304-15.
41. Tang Y. On the multiple imputation variance estimator for control-based and delta-adjusted pattern mixture models. *Biometrics*. 2017; 73: 1379-87.