ANALYSIS OF MULTIVARIATE HIERARCHICAL DATA WITH MISSINGNESS: AN APPLICATION TO IN-PATIENT PAEDIATRIC PNEUMONIA CARE

Susan W. Gachau

A thesis submitted to the School of Mathematics, University of Nairobi for the award of Degree of Doctor of Philosophy in Biostatistics.

26 July, 2021

Declaration

This thesis is my original work and has not been presented for a degree in any other University.

Signature..... Date.....26/07/2021.....

Susan W. Gachau

Registration number: I80/52785/2018

This thesis has been submitted for examination with our approval as University supervisors

Signature..... Date.....July 29, 2021.....

Dr. Edmund Njeru Njagi

Department of Non-Communicable Disease Epidemiology

London School of Hygiene and Tropical Medicine, UK.

Signature..... Date.....July 26, 2021.....

Dr. Nelson Owuor

School of Mathematics, University of Nairobi, Kenya.

Signature..... Date26 July 2021.....

Dr. Philip Ayieko

Department of Infectious Disease Epidemiology,

London School of Hygiene and Tropical Medicine, London, United Kingdom.

Mwanza Intervention Trials Unit, Mwanza, Tanzania

Signature..... Date26 July, 2021.....

Dr. Rachel Sarguta

School of Mathematics, University of Nairobi, Kenya.

Abstract

Routine health data are used to monitor quality of care and to inform interventions to improve patient care. However, statistical analysis of such data presents several challenges related to handling missing data and multiple responses in the presence of complex data structures.

In this study we sought to: i) Analyze multilevel clustered data accounting for covariate missingness. ii) Explore appropriate strategies for handling missing data when the outcome is a composite of partially observed components. iii) Examine sensitivity of results to departures from the commonly assumed missing at random (MAR) mechanism. iv) Simultaneously estimate joint covariate effects and association amongst multiple correlated outcomes.

We analysed routine data collected during a cluster randomized trial in 12 Kenyan hospitals between March and November 2016. There were 2127 children admitted by 378 clinicians ascross the study sites. The outcomes of interest were 12 pneumonia quality of care indicators spanning assessment, diagnosis and classification and treatment domains of care. For the first three objectives, we constructed Paediatric Admission Quality of Care (PAQC) score, an ordinal composite outcome using 12 pneumonia care indicators. Covariates of interest included : trial arm and follow-up time, hospital, clinician and patient-level variables. Missing data occurred in patient and clinician level variables. Missing data in covariates were imputed using latent normal joint modelling approach assuming MAR mechanism. Random-effects and marginal models were the substantive models

of interest. To explore appropriate strategies of handling missing PAQC score subcomponents, we conducted a simulation study. Multiple imputation (MI) at subcomponent level versus the conventional method where missing PAQC score subcomponents were scored with value 0. We assessed departure form MAR assumption within pattern mixture models. Elicited experts' opinions were incorporated into the imputation models in the form of prior distributions and delta adjustment parameters to create missing not at random imputed values. In the fourth objective, we analyzed 9 binary pneumonia care indicators under the correlated random effects joint model, by applying pairwise fitting and pseudo-likelihood methods before and after MI of missing covariates.

From results, trial intervention was associated with higher uptake of the paediatric pneumonia guidelines during the trial period. Parameter estimates were precise after MI of covariates compared to complete case analysis. In a range of simulation scenarios, multiple imputation of missing PAQC score elements at item level produced minimally biased estimates compared to the conventional method. Our inferences were insensitive to departures from MAR assumption using either sensitivity analysis approach. Lastly, there was a significant joint interaction effect between intervention arm and follow-up time on pneumonia care indicators. The strength and direction of association amongst outcomes varied within and across domains care.

This study demonstrates the practical utility of advanced biostatistical analyses methods with an aim to promote their use while answering substantive health research questions. Uptake of such methods can improve analysis and report-

ing of health data used to inform policies and in the long run enhance optimal utilization of limited resources while promoting better patients' outcomes.

To my family

Acknowledgments

I am deeply grateful to the Almighty God because without His blessings and unlimited grace, my doctoral journey would not have been possible.

My sincere gratitude goes to my supervisors, Dr. Edmund Njeru Njagi, Dr. Nelson Owuor, Dr. Philip Ayieko and Dr. Rachel Sarguta. Your support, guidance and encouragement were a great source of motivation for me to work better and harder. I was greatly honoured and privileged to travel my PhD. journey together with you.

I am grateful to our collaborators, Prof. Geert Molenberghs and Dr. Matteo Quartagno. I learned a lot from each of you within the specific areas that I have worked with you. Special thanks to Prof. Mike English for providing data which enabled us to perform all the interesting analyses presented in this thesis.

I acknowledge financial support from DELTAS-SSACAB Consortium, as this thesis would not be possible without it. I am grateful to my hosts institutions, KEMRI-Wellcome Trust Research Programme (Training Department and Health Services Unit) and the University of Nairobi (School of Mathematics and Graduate School), for administrative support and facilitation which contributed to a pleasant working atmosphere. I also wish to thank my friends and colleagues for constant encouragement and moral support over the years. Your friendship is a special gift and deeply appreciated.

Last but not the least, I would like to thank my family, for the unconditional and lifelong love. Thank you for encouraging me to pursue my dreams, standing by my side at every step of the journey, and always praying for my success. To my

lovely daughter, Shiku, I am truly blessed to have you and I am sorry for the times I was away.

List of publication/manuscripts

The contents of this thesis are based on the following manuscripts:

1. **Gachau, S**., Owuor, N., Njagi, E. N., Ayieko, P., & English, M. (2019). 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*, 7, 198.

2. **Gachau, S**., Quartagno, M., Njagi, E. N., Owuor, N., English, M., & Ayieko, P. (2020). Handling missing data in modelling quality of clinician-prescribed routine care: Sensitivity analysis of departure from Missing at Random (MAR) assumption. *Statistical Methods in Medical Research*. https://doi.org/10.1177/0962280220918279.

3. **Gachau, S**., Njagi, E.N., Owuor, N., Mwaniki, P., Quartagno, M., Sarguta, R., English, M. and Ayieko, P. (2021). Handling missing data in a composite outcome with partially observed components: simulation study based on clustered paediatric routine data. *Journal of Applied Statistics, pp.1-14*.

4. **Gachau, S**., Njagi, E. N., Molenberghs, G., Owuor, N., Sarguta, R., English, M., & Ayieko, P. (2021). Pairwise joint modelling of clustered and high-dimensional pneumonia care outcomes with covariate missingness. (*Submitted*)

Contents

List of Tables

List of Figures

Acronyms

A&F Audit and Feedback

CCA Complete Case Analysis

CIN Clinical Information Network

EM Expectation-Maximization

FCS Full Conditional Specification

FIML Full Information Maximum Likelihood

GEE Generalized Estimating Equations

GLMM Generalized Linear Mixed Models

G(SPM) Generalized (Shared Parameter Model)

JM Joint Modelling

KEMRI-WT Kenya Medical Research Institute-Wellcome Trust

LOCF Last Observation Carried Forward

LRT Likelihood Ratio Test

MAR Missing at Random

MCAR Missing Completely at Random

MI Multiple Imputation

MNAR Missing Not at Random

MO Medical Officers

MO Multiple Outputation

MSE Mean Squared Error

OR Odds Ratio

PAQC Paediatric Admission Quality of Care

PMM Pattern Mixture Model

SeM Selection Models

WHO World Health Organization

WT Wald Test

Chapter 1

General Introduction

1.1 Background

Data sets containing routine data collected from multiple sites are a common phenomenon in many health research settings and present significant challenges in analysis. For example, in evaluating quality of care provided to patients using data collected during inpatient admission in a set of hospitals,observations from the same hospital tend to be correlated. Further, within each hospital observations from the same clinician will also tend to be correlated. While using data from multiple hospitals enhances generalization of results to a wider population of hospitals, it leads to complex hierarchical data structures. Furthermore, researchers may be interested in multiple outcomes representing different aspects of care processes (Mc Cord et al., 2018) leading to cluster correlated data. In addition to multiple outcomes and complex data structures, routine data are subject to missing information at any level of the hierarchical structure.

When the outcome is affected by missing data, assuming a missing at random (MAR) (Rubin, 1976) missingness mechanism, and assuming separability of the missingness and measurement processes, a likelihood based analysis of the observed data, will be valid. This is not the case with quasi-likelihood based methods(Fitzmaurice et al., 2009a, Chapter 1, p. 3), and only valid under a missing completely at random (MCAR) mechanism.

With missing covariate data as well, unless the probability of the covariate missing does not depend on the outcome variable, an analysis based only on "complete cases" will provide biased inference (Carpenter and Kenward, 2013). In this case, multiple imputation (MI), usually based on an MAR assumption for the missingness mechanism, is applied to provide validity under more realistic assumptions about the missingness mechanism, as well as to mitigate the potential loss of efficiency due to information loss (Carpenter and Kenward, 2013, Chapter 1, p. 9). While MI is a useful tool in this context, there is still a risk of invalid inferences arising from incompatibility between the imputation model and the substantive model (Bartlett et al., 2015). Incompatibility arises whenever the substantive model contains non-linear effects, interactions, and hierarchical structures, yet these are not properly accounted for in the imputation model (Carpenter and Kenward, 2013; Bartlett et al., 2015).

In both missing outcome and the missing covariate contexts, even when an analysis which is valid under the MAR assumption is performed, there is need for sensitivity analyses. This is because the possibility that a missing not at random (MNAR) mechanism could be in operation cannot be discounted (Molen-

berghs et al., 2014, Chapter 1, p. 319). Sensitivity analyses usually take the form of comparing inferences under an MAR assumption with those under an MNAR assumption, or comparing inferences under different models formulated under an MNAR assumption. Sensitivity analyses could be in the framework of the pattern-mixture factorization (usually called the pattern-mixture model, and abbreviated PMM) of the joint model for the missingness and measurement processes (Molenberghs et al., 2014, chapter. 1), the selection models or the generalized shared-parameter factorization (usually called the generalized shared-parameter model, and abbreviated generalized SPM) (Creemers et al., 2010, 2011). Finally, with a multivariate vector of outcomes, joint models (Fieuws and Verbeke, 2004, 2006) are used to simultaneously analyse all the outcomes. Joint models allow one to test hypotheses of joint effects of covariates on the various outcomes simultaneously, and to study the association among the different outcomes (Molenberghs and Verbeke, 2005, Chapter 25, p. 466). Joint models also provide efficiency gain, in case of missing data in some of the outcomes, or in case some fixed effects parameters are shared by the outcomes (Fitzmaurice et al., 2009b). Within the joint modelling framework, correlated random-effects joint models are preferred to shared-parameter joint models (shared random-effects joint models), as the latter may impose too restrictive association structure (Molenberghs and Verbeke, 2005, Chapter 25, P. 468). The cost is that as the dimension of the multivariate response vector increases, so do the number of random-effects in the correlated random-effects joint models, resulting in fitting problems due to increased dimension of the variance covariance matrix. To circumvent this chal-

lenge, fitting strategies such as pairwise fitting combined with pseudo-likelihood methodology (Fieuws and Verbeke, 2006) are used.

1.2 Missing Data Concepts

1.2.1 Notations

Missing data refers to intended information about a study subject but could not be measured or observed for one reason or another. Missing data problems are common in many disciplines and consequently they complicate statistical analysis and inference (Molenberghs et al., 2014, Chapter 1, p. 3). In clinical studies, missing data could occur due to withdrawal, attrition, loss to follow up or lost records. Missingness could also occur due to non-response by study participant or by study design.

Here, we use hypothetical data set to illustrate basic missing data concepts and terminology used throughout this report. Suppose $\boldsymbol{Y}$ (representing both outcome and covariates) is an $N \times P$ matrix containing data values on p variables for N study subjects $i = 1, \ldots, N$. For the i^{th} study subject, the elements of $\boldsymbol{Y}$ denoted by y_{ip} can be grouped into a vector $\boldsymbol{Y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})$. Further, assuming that some of the y_{ip} values are partially observed, a missingness indicator matrix $\boldsymbol{R}$ with same dimension as $\boldsymbol{Y}$ can be defined. The elements of matrix $\boldsymbol{R}$ are defined as follows

$$r_{ip} = \begin{cases} 1, & \text{if } y_{ip} \quad \text{is} \quad \text{observed} \\ 0, & \text{otherwise.} \end{cases}$$

For each study subject, the elements of R can be grouped into a vector $\boldsymbol{R}_i = (r_{i1}, r_{i2}, \ldots, r_{ip})$. Given $\boldsymbol{R}_i$, the vector $\boldsymbol{Y}_i$ can be partitioned into observed and missing data subvectors of denoted by $\boldsymbol{Y}_i^{obs}$ and $\boldsymbol{Y}_i^{miss}$ respectively. When taken together, $(\boldsymbol{Y}_i)$ (i.e., the measurement process) and $(\boldsymbol{R}_i)$ (i.e., the missingness process) constitute the joint density distribution of the full data defined by

$$f(\boldsymbol{Y}_i, \boldsymbol{R}_i, \boldsymbol{\theta}, \boldsymbol{\psi}), \tag{1.1}$$

where $\boldsymbol{\theta}$, and $\boldsymbol{\psi}$ are vector of parameters for the measurement and missingness processes respectively.

1.2.2 Missing Data Mechanisms

Usually, the missing data mechanism underlying a given data set is unknown to the study researchers. Therefore, assumptions are normally made about a plausible mechanism. The validity of inference depends on whether the assumptions hold for the data at hand (Molenberghs et al., 2014, Chapter 1).

Rubin (1976) distinguished three broad classes of missing data mechanisms namely; MCAR, MAR and MNAR. The three mechanisms were defined depending on how $\boldsymbol{R}_i$ is related to $\boldsymbol{Y}_i$. Specifically, the conditional density of the missingness process $\boldsymbol{R}_i$ given $\boldsymbol{Y}_i^{obs}$ and $\boldsymbol{Y}_i^{miss}$ as outlined in the following subsection.

Missing Completely at Random

Missing completely at random (MCAR) mechanism occurs when the missingness process and the measurement process are independent (Rubin, 1976). That is,

$$P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}, \boldsymbol{\psi}) = P(\boldsymbol{R}_i|\boldsymbol{\psi}). \tag{1.2}$$

In other words, the probability of missingness is independent of observed and unobserved components of $\boldsymbol{Y}_i$. When data are MCAR, the observed data can be taken as a random sample of the target population. In this case, restricting analysis to complete cases yields valid but inefficient parameter estimates (Carpenter and Kenward, 2013, Chapter 1, p. 9).

Missing at Random

According to Rubin (1976), data are said to be MAR when the probability of missing values in a variable does not depend on the variable of interest but are conditionally dependent on other observed variables in the data set. In other words, $\boldsymbol{R}_i$ is conditionally independent of $\boldsymbol{Y}_i^{miss}$ given $\boldsymbol{Y}_i^{obs}$, that is,

$$P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}, \boldsymbol{\psi}) = P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{\psi}). \tag{1.3}$$

When data are MAR, complete cases are not as a random sample of the target population and restricting analysis to complete case records yields both biased and inefficient parameter estimates (Molenberghs et al., 2014, Chapter 1, p. 9-10). MAR mechanism further implies that the conditional distribution of $\boldsymbol{Y}_i^{miss}$

given $\boldsymbol{Y}_i^{obs}$ is the same as the distribution of the corresponding observations in the completers and target population.

Missing Not at Random

Data are said to be MNAR when the probability of a value being missing depends on unobserved measurements. This is in addition to dependencies on observed covariates and/or outcomes (Molenberghs et al., 2014, Chapter 1, p. 11). That is, the conditional distribution of $\boldsymbol{R}_i$, given $\boldsymbol{Y}_i^{obs}$ is related to $\boldsymbol{Y}_i^{miss}$ as shown below,

$$P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}, \boldsymbol{\psi}) = P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}, \boldsymbol{\psi}). \tag{1.4}$$

Ignorability

With complete data, inference about parameters of the measurement process $(\boldsymbol{\theta})$ can be conducted based on the likelihood of the data given $\boldsymbol{\theta}$, that is, $P(\boldsymbol{Y}|\boldsymbol{\theta})$ (Molenberghs et al., 2014, p. 9). Assuming that $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ denote two distinct sets of parameters, the joint distribution can be simplified to

$$\begin{aligned} P(\boldsymbol{R}_i, \boldsymbol{Y}_i^{obs}|\boldsymbol{\theta}, \boldsymbol{\psi}) &= \int P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{\psi}) P(\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}|\boldsymbol{\theta}) d\boldsymbol{Y}_i^{mis} \\ &= P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{\psi}) \int P(\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}|\boldsymbol{\theta}) d\boldsymbol{Y}^{mis} \\ &= P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{\psi}) P(\boldsymbol{Y}_i^{obs}|\boldsymbol{\theta}). \end{aligned} \tag{1.5}$$

This implies that when data are MAR, the missing data mechanism is ignorable.

When missingness is ignorable, likelihood-based inferences can be obtained by integrating the missing observations from $f(\boldsymbol{Y}_i^{obs}|\boldsymbol{\theta})$, i.e.,

$$L(\boldsymbol{\theta}|\boldsymbol{Y}_i^{obs}) = c \times \prod_{i=1}^{N} \int f(\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}|\boldsymbol{\theta}) d\boldsymbol{Y}_i^{miss}, \tag{1.6}$$

where c is a constant factor that is independent of the missingness process parameters, $(\boldsymbol{\psi})$. In this case, the missing data mechanism need not be known in order to obtain valid statistical inferences about θ from the observed data (Molenberghs et al., 2014, Chapter 1, p. 9). Since MCAR is a special case of MAR, ignorability also holds for MCAR. On the other hand, when data are MNAR and the aim is to make inference about the distribution of the observed data, then missing data mechanism cannot be ignored. The assumed model for $P(\boldsymbol{R}_i|\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss})$ must be accounted for in the analysis model. However, assumptions made about $\boldsymbol{R}_i$ are unverifiable from the data. Therefore, sensitivity analysis of inferences to a variety of plausible MNAR models is recommended (Carpenter and Kenward, 2013, Chapter 10,).

1.2.3 Missing Data Frameworks

To correct for non-ignorability, the joint models for the measurement $(\boldsymbol{Y}_i)$ and the missingness process $(\boldsymbol{R}_i)$ are required to obtain valid estimates (Molenberghs and Verbeke, 2005, Chapter 26, p. 484). In particular, the joint distribution $f(\boldsymbol{Y}_i, \boldsymbol{R}_i|\boldsymbol{\theta}, \boldsymbol{\psi})$ can be factorized into one of the three common modelling frameworks namely; selection models, pattern mixture models and shared parameter models (Rubin, 1976; Molenberghs and Verbeke, 2005; Carpenter and Kenward, 2013).

Selection Models

In the selection model (SeM) factorization, the joint distribution of measurement process $(\boldsymbol{Y}_i)$ and missingness process $(\boldsymbol{R}_i)$ factorizes into

$$f(\boldsymbol{Y}_i, \boldsymbol{R}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{Y}_i|\boldsymbol{\theta}) f(\boldsymbol{R}_i|\boldsymbol{Y}_i, \boldsymbol{\psi}), \tag{1.7}$$

where $f(\boldsymbol{Y}_i|\boldsymbol{\theta})$ is the marginal of the measurement process and $f(\boldsymbol{R}_i|\boldsymbol{Y}_i, \boldsymbol{\psi})$ is the density of the missing data process conditional on the measurement process . SeM impose assumptions about the marginal density of the measurement process for identifiability purpose. Specifically, identifiability is achieved through parametric assumptions about $f(\boldsymbol{Y}_i|\boldsymbol{\theta})$ and unverifiable models for the dependence of the missingness process on unobserved data (Molenberghs et al., 2014, Chapter 1, p. 11).

Pattern Mixture Models

Alternatively, one can consider the pattern mixture models (PMM) factorization. PMM factorization is the reverse of SeM (Little, 1993; Molenberghs and Verbeke, 2005; Molenberghs and Kenward, 2007; Carpenter and Kenward, 2013) in that it specifies the joint distribution in terms of marginal distribution of the missingness process $(\boldsymbol{R}_i)$ and the conditional distribution of the measurement process $(\boldsymbol{Y}_i)$ given the missingness process, that is,

$$f(\boldsymbol{Y}_i, \boldsymbol{R}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{Y}_i|\boldsymbol{R}_i, \boldsymbol{\theta}) f(\boldsymbol{R}_i|\boldsymbol{\psi}), \tag{1.8}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are parameters of the conditional and marginal densities respectively. In PMM, the distribution of $\boldsymbol{Y}_i$ given patterns of missingness process $\boldsymbol{R}_i$ is not completely identifiable. Therefore, unverifiable links are postulated among the distributions of the measurement processes conditional on the patterns of missingness processes for identification purposes (Molenberghs et al., 2014, Chapter 1, p. 11).

Shared Parameter Models

Lastly, the measurement and missingness processes can be modelled jointly within the shared parameter model (SPM) framework (Creemers et al., 2010, 2011). An SPM factorization is defined by

$$f(\boldsymbol{Y}_i, \boldsymbol{R}_i|\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{b}_i) = f(\boldsymbol{Y}_i|\boldsymbol{\theta}, \boldsymbol{b}_i) f(\boldsymbol{R}_i|\boldsymbol{\psi}, \boldsymbol{b}_i). \tag{1.9}$$

In this framework, the measurement process $(\boldsymbol{Y}_i)$ and missingness process $(\boldsymbol{R}_i)$ are assumed to share latent variables, conditional upon which, $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$ are independent. The latent variables are more often considered to be random-effects denoted by $\boldsymbol{b}_i$. The random-effects are assumed to follow a specific parametric distribution (Creemers et al., 2011).

Generalized Share Parameter Models

The SPM framework can be extended to a more general form where the vector of random-effects is defined by $\boldsymbol{b}_i = g_i, h_i, j_i, k_i, l_i, m_i, q_i$. Creemers et al. (2011) constructed a generalized shared parameter model (GSPM) in a missing outcome

context as follows,

$$
\begin{aligned}
&f(\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}, \boldsymbol{R}_i | g_i, h_i, j_i, k_i, l_i, m_i, q_i, \boldsymbol{\theta}, \boldsymbol{\psi}) \\
&= f(\boldsymbol{Y}_i^{obs} | g_i, h_i, j_i, l_i, \boldsymbol{\theta}) f(\boldsymbol{Y}_i^{miss} | \boldsymbol{Y}_i^{obs}, g_i, h_i, k_i, m_i, \boldsymbol{\theta}) f(\boldsymbol{R}_i | g_i, j_i, k_i, q_i, \boldsymbol{\psi}),
\end{aligned} \tag{1.10}
$$

where g_i is a random-effect common to all three factors in the right hand of (2.18), h_i, j_i, and k_i are random-effects shared between pair of factors, and l_i, m_i, and q_i are random-effects restricted to a single factor (Creemers et al., 2011; Njagi et al., 2014).

1.2.4 Missing Data Patterns

Before embarking on a formal analysis, it is important to understand how missing data patterns manifest themselves in each data set (Van Buuren, 2018, Chapter. 3). Such is the case because missing data patterns underlying a given data set can be used to identify auxiliary variables that are predictive of missingness thus enhancing statistical efficiency in subsequent analysis. Figure 2.1 shows a graphical representation of common missing data patterns which include univariate and multivariate missing data patterns. Univariate missing data pattern occurs when only one variable in the data set is partially observed (Figure 2.1, left panel).

When a data set has two or more variables with missing data, then it is said to have a multivariate missing data pattern. A data set is said to have a multivariate monotone missing data pattern if the variables can be ordered such that if Y_p is missing, then all variables Y_k with $k > p$ are also missing (Figure 2.1, middle panel).

Monotone missing data pattern is common in longitudinal studies due to drop-out (Molenberghs and Verbeke, 2005, Chapter 26). On the other hand, if a data set has several variables missing intermittently (Figure 2.1, right panel), then the underlying missingness pattern is said to be multivariate general.

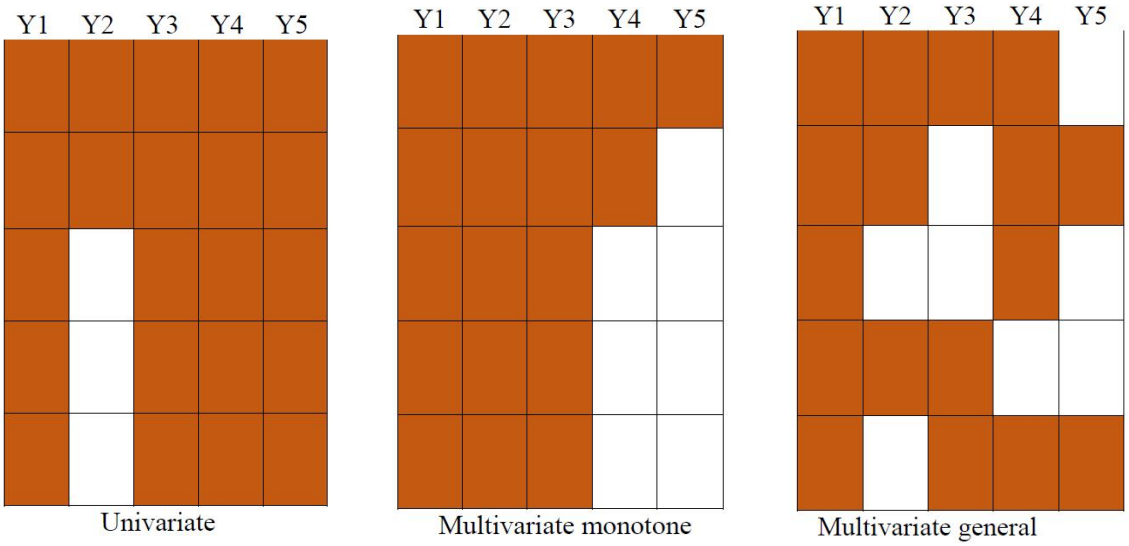

Figure 1.1: Common missing data patterns in multivariate data.

1.3 Motivating Case Study

Study Design

Data to be analysed in this study came from a cluster randomized trial conducted by the Kenya Medical Research Institute-Wellcome Trust Research programme (KEMRI-WTRP) between March 2016 and November 2016. The trial was embedded within the Clinical Information Network (CIN) observational study (Ayieko et al., 2015; English, 2013; Tuti et al., 2015). The trial's objective was to investigate uptake of paediatric pneumonia treatment guidelines following recommendations by the World Health Organization (WHO) in 2013 (Organization, 2013).

Details of the trial are contained in the trial report (Ayieko et al., 2019, 2017). In brief, six hospitals were randomly allocated to the intervention arm while the remaining six were allocated to the control arm. The intervention arm received a monthly enhanced audit and feedback (A&F) 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. Network intervention included peer learning among clinicians across hospitals and follow up visits (emails and phone calls) by the trail co-ordinating paediatrician. The control arm on the other hand received a standard audit and feedback report alone and network intervention strategy (Ayieko et al., 2017, 2019). The Kenya Ministry of Health and KEMRI's Scientific and Ethical Review Unit approved the use of de-identified data without individual patient's consent (Ayieko et al., 2017).

In total, 2299 children aged 2 to 59 months were admitted with childhood pneumonia in all 12 hospitals during the trial period. Of all pneumonia cases, 1084/2299 (47.1%) were admitted in 6 hospitals assigned to the enhanced A&F (intervention) arm.

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)) (Harris et al., 2009) using standard operating procedure manual. For each case record, details of the admitting clinician including a unique clinician code, gender and cadre were also abstracted into a separate database. In this report the term 'cadre' refers to clini-

cian'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.

Patients' and clinicians' databases were linked using unique clinician code present in both databases with a success rate of 92.5% (2127/2299). The remaining 172/2299 case records were excluded from all analyses for lack of admitting clinician's information. This resulted in a hierarchical data set with three levels of clustering that is, 2127 patients (level 1) admitted by 378 specific clinicians (level 2) in 12 hospitals (level 3). The number of paediatric pneumonia admissions per hospital ranged between 42 and 356 patients (Table 1.1).

Characteristic of hospitals, clinicians and patients enrolled in the trial are presented in Table 1.1. Five out of 12 hospitals were drawn from high malaria endemic regions (i.e., three hospitals in the control arm and two hospitals in the intervention arm) while the remaining seven hospitals (i.e., four hospitals in the control arm and three hospitals in the intervention arm) were drawn from regions with low malaria endemicity in Kenya (Ayieko et al., 2015). Furthermore, four in 12 hospitals were high admission workload hospitals, that is, more than 1000 paediatric admissions per annum (i.e., three hospitals in the control arm and one hospital in the intervention arm). On the other hand, 8/12 were low admission workload hospitals, that is, less than 1000 paediatric admissions per year (i.e., three hospitals in the control arm and five hospitals in the intervention arm) irrespective of admission diagnosis (Table 1.1). On average there were 32 clinicians per hospital with a standard deviation of nine clinicians.

Approximately, 21.9% (83/378) and 21.7% (82/378) clinicians had missing data on gender and cadre respectively. Among clinicians with documented cadre, 6 (1.59%) and 184 (48.7%) were clinical officers and clinical officer interns respectively while 6 (1.59%) and 99 (26.19%) were medical officers medical officer interns respectively. The number of patients per clinician ranged between 3 and 46. Overall, 42% (903/2127) of the patients were aged between 2 and 11 months while 45% (950/2127) of the patients were females (Table 1.1). Patient's sex was missing in 0.7% (17/2127) of case records (Table 1.1).

Table 1.1: Descriptive characteristic of pneumonia trial data at hospital, clinician and patients level. Denominator for proportions exclude case with missing values

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	All hospitals combined
Enhanced A&F[a] arm	No	Yes	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	
Admission workload	Low	Low	High	Low	Low	High	Low	Low	Low	High	High	Low	
Malaria prevalence	High	Low	High	Low	Low	Low	High	High	Low	Low	Low	High	
Pneumonia admissions, n (%)	132 (6.2)	215 (10.1)	210 (9.9)	243 (11.4)	110 (5.2)	356 (16.7)	63 (2.9)	167 (7.9)	88 (4.1)	172 (8.1)	329 (15.6)	42 (1.9)	2127 (100)
Patients aged 2-11 months,n (%)	44 (33.3)	79 (36.7)	71 (33.8)	89 (36.6)	49 (44.6)	193 (54.5)	22 (34.9)	70 (41.9)	45 (51.1)	99 (57.6)	129 (39.2)	13 (30.9)	903 (42.5)
Patients aged 12-59 months,n (%)	88 (66.7)	136 (63.3)	139 (66.2)	154 (63.4)	61 (55.5)	162 (45.5)	41 (65.1)	97 (58.1)	43 (48.9)	73 (42.4)	200 (60.8)	29 (69.1)	1224 (57.5)
Male patients,n (%)	80 (60.6)	118 (54.9)	103 (49.1)	138 (56.8)	55 (50.0)	194 (54.5)	35 (55.6)	100 (59.9)	42 (47.7)	95 (55.2)	181 (55.1)	23 (54.8)	1164 (54.7)
Female patients, n (%)	52 (39.4)	97 (45.1)	107 (50.9)	101 (41.6)	55 (50.0)	162 (45.5)	27 (42.9)	67 (40.1)	46 (52.3)	76 (44.2)	141 (42.9)	19 (45.2)	950 (44.6)
Missing patients sex, n (%)	0 (0.0)	0 (0.0)	0 (0.0)	4 (1.7)	0 (0.0)	0 (0.0)	1 (1.6)	0 (0.0)	0 (0.0)	1 (0.6)	7 (2.1)	0 (0.0)	13 (0.6)
Number of clinicians	31	36	43	33	25	36	24	39	32	20	44	15	378
Female clinicians, n (%)	15 (54.6)	11 (30.6)	15 (34.9)	13 (39.4)	2 (8.0)	14 (38.9)	13 (54.2)	16 (41.0)	0 (0.0)	0 (0.0)	24 (54.6)	5 (33.3)	128 (33.9)
Male clinicians,n (%)	16 (45.45)	18 (50.0)	28 (65.2)	20 (60.6)	8 (32.0)	10 (27.8)	11 (45.8)	23 (59.0)	3 (9.4)	1 (5.0)	20 (45.4)	10 (66.7)	168 (44.4)
Clinicians with missing sex, n (%)	0 (0.0)	7 (19.4)	0 (0.0)	0 (0.0)	15 (60.0)	12 (33.3)	0 (0.0)	0 (0.0)	29 (90.6)	19 (95.0)	0 (0.0)	0 (0.0)	82 (21.7)
Cadre: CO[b],n (%)	0 (0.0)	0 (0.0)	0 (0.0)	2 (6.1)	3 (12.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (3.1)	0 (0.0)	0 (0.0)	0 (0.0)	6 (1.6)
Cadre: CO interns, n (%)	20 (64.5)	18 (50.0)	31 (72.1)	20 (60.6)	2 (8.0)	14 (38.9)	16 (66.7)	29 (74.4)	1 (3.1)	0 (0.0)	25 (56.82)	8 (53.3)	184 (48.7)
Cadre: MO[c], n (%)	1 (3.2)	12.8 (2.8)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.8)	0(0.0)	1 (2.6)	0 (0.0)	1 (5.0)	1 (2.3)	0 (0.0)	6 (1.6)
Cadre: MO interns, n (%)	10 (32.3)	10 (27.8)	12 (27.9)	11 (33.3)	5 (20.0)	9(25.0)	7 (29.2)	9 (23.1)	1 (3.1)	0 (0.0)	18 (40.9)	7 (46.7)	99 (26.2)
Clinicians with missing cadre, n (%)	0 (0.0)	7 (19.4)	0 (0.0)	0 (0.0)	15(60.0)	12 (33.3)	1 (4.1)	0 (0.0)	29 (90.6)	19 (95.0)	0 (0.0)	0 (0.0)	83 (21.9)

$A\&F^a$:-Audit and feedback,CO^a:- Clinical officer,MO^b:-Medical Officer

Childhood Pneumonia

Pneumonia is an infection of the lungs caused by bacteria, viruses or fungi. The most common causes of pneumonia in low-and middle-income countries (LMICs) are bacteria and viruses, but in these settings diagnosis and treatment of pneumonia is syndromic (based on clinical signs and symptoms). Globally pneumonia continues to be a leading cause of mortality among children under five years of age with nearly 1 million deaths every year (Organization et al., 2014). Half of these deaths occur in Sub-Saharan Africa (UNICEF et al., 2016). Although pneumonia is preventable and manageable with antibiotics, it is estimated to causes more deaths than HIV/AIDS, diarrhoea and malaria combined (Amouzou et al., 2016).

In 2013 the World Health Organization (WHO) revised the classification of pneumonia to include only two categories of pneumonia. That is, “pneumonia” with fast breathing and/or chest indrawing, which is treatable with oral amoxicillin with a dose of at least $40mg/kg/dose$ twice daily $(80mg/kg/day)$ for five days. On the other hand, “severe pneumonia”, with any general danger sign (i.e., oxygen saturation < 90, cyanosis, inability to drink/breast feed, AVPU= "verbal", ’pain’, or "unresponsive, and grunting)is treated with oxygen, injectable penicillin and gentamicin (Organization et al., 2014). Figure 1.1 summarizes pneumonia assessment, classification and treatment guidelines recommended by the WHO in the year 2013.

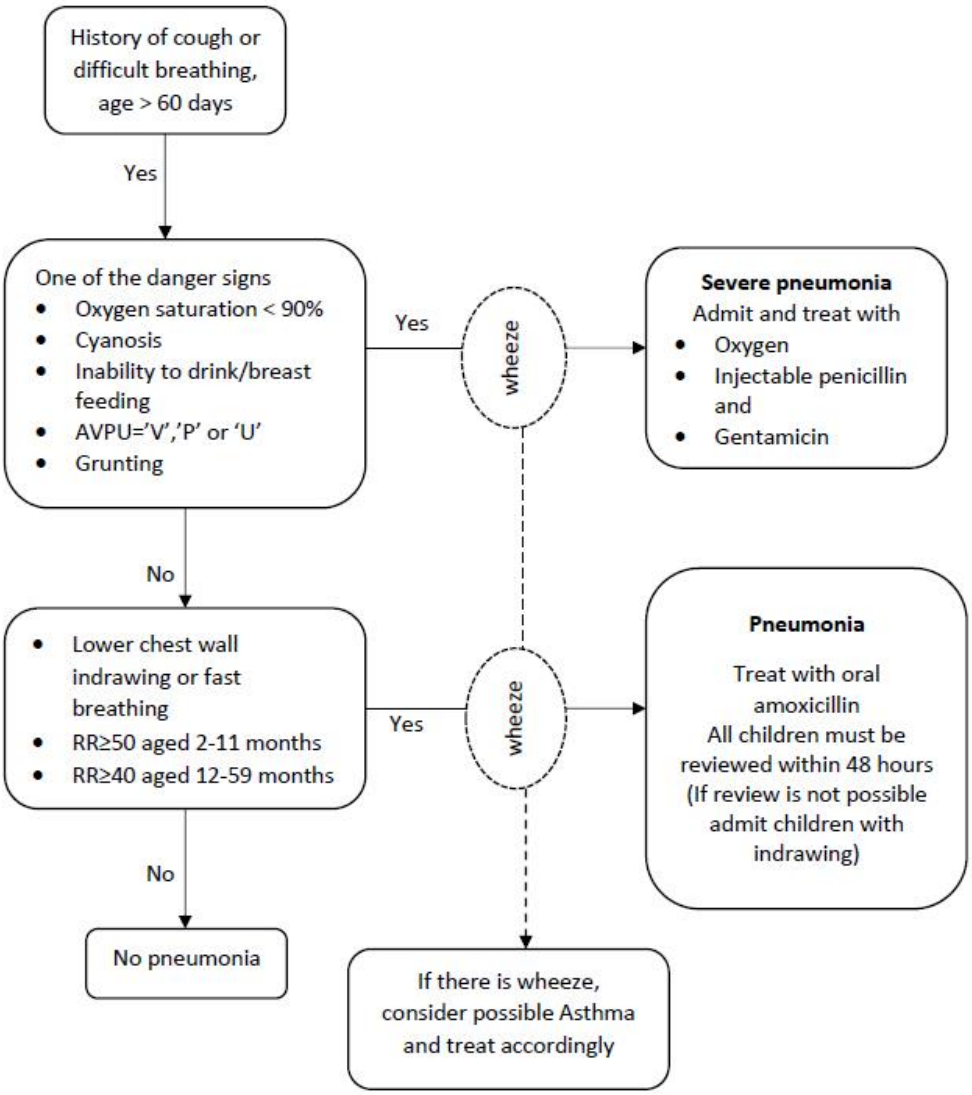

Figure 1.2: Assessment, diagnosis and treatment of pneumonia cases for children aged 2-59 months

Pneumonia Care Processes

In this research study, we focused on 12 pneumonia care processes spanning assessment, diagnosis and classification, and treatment domains of paediatric care (Table 1.2). In the assessment domain, nine pneumonia signs and symptoms (two primary and seven secondary signs & symptoms) relevant for diagnosis and classification of pneumonia of severity were considered as per WHO guidelines (Organization, 2013). Diagnosis domain entailed clinical diagnosis and classification of disease severity by admitting clinician while treatment domain contains two indicators of interest, that is, prescription of correct treatment depending on pneumonia severity and correctness of treatment dosage (Table 1.2). In this case, the severity of interest was 'pneumonia' treatable with oral amoxicillin. Correctness of dose was calculated using patient's weight, prescribed oral amoxicillin dose and frequency of oral amoxicillin administration.

Missing data occurred in 6/9 primary and secondary signs and symptoms within assessment domain of care (Organization, 2013). The proportion of missingness ranged between 0.2% and 39% (Table 1.2). In the diagnosis domain, all patients had a clinical diagnosis documented in the medical record by the admitting clinician. However, only 69.3% (1473/2127) patients had correct severity classification (Organization, 2013). That is, pneumonia severity classification abstracted from the medical record corresponded to the severity implied by secondary signs and symptom documented at point of admission.

In the treatment domain, only 50.2% (1036/2062) of all pneumonia patients received oral amoxicillin during the trial period as recommended. Among oral

amoxicillin recipients, 4/1036 (0.4%), 27/1036 (2.6%) and 30/1036(2.9%) had missing oral amoxicillin dose, frequency of oral amoxicillin administration and patient's weight respectively (Table 1.2). Missing information in either of the 3 indicators made it impossible to calculate correctness of the dose as per the guidelines (Organization, 2013). Table 1.2 presents a summary of the level of documentation within domain of interest among children admitted with childhood pneumonia during the trial period.

Table 1.2: Documentation of pneumonia care processes in the assessment, diagnosis and treatment domains

Quality of care domain	Pneumonia care indicator	Documented cases (%)	Missing cases (%)	Variable type
1. Assessment				
Primary signs & symptoms	Cough	2118/2127 (99.6)	9/2127 (0.4)	binary
	Difficult breathing	2114/2127 (99.4)	13/2127 (0.6)	binary
Secondary sign & symptoms	Oxygen saturation	1297/2127 (60.9)	830/2127 (39.1)	continuous
	Ability to drink	2127/2127 (100)	0/2127 (0.0)	binary
	Central cyanosis	2127/2127 (100)	0/2127 (0.0)	binary
	AVPU[b]	2112/2127 (99.3)	15/2127 (0.7)	categorical (4 levels)
	Grunting	2127/2127 (100)	0/2127 (0.0)	binary
	Respiratory rate	1889/2127 (88.8)	238/2127 (11.2)	continuous
	Lower chest wall indrawing	2123/2127 (99.8)	4/2127 (0.2)	binary
2. Diagnosis and classification	Classified pneumonia cases	2127/2127 (100)	0/2127 (0.0)	binary
	Correct classification	1473/2127 (69.3)		
3. Treatment domain	Oral amoxicillin prescribed	2062/2127 (96.9)	65/2127 (3.1)	binary
	Yes	1036/2062 (50.2)		
	No	1026/2062 (49.8)		
Amoxicillin indicators	Amoxicillin dose prescribed	1032/1036 (99.6)	4/1036 (0.4)	continuous
	Patients weight	1006/1036 (97.1)	30/1036 (2.9)	continuous
	Frequency of administration	1009/1036 (97.4)	27/1036 (2.6)	categorical (4 levels)

AVPU[b]: A for Alert, V for Verbal response, P for pain , U for unresponsive

1.4 Statement of the problem

Statistical analysis of multivariate hierarchical data with both missing outcome and covariates poses several challenges. The first challenge is imputation of the missing covariates, ensuring that the hierarchical structure is properly accounted for. Secondly, when some of the multiple outcomes are partially observed and one is interested in combining them into a single composite measure, then the challenge is whether to address missing data at item level or at composite score level. The third challenge is assessing sensitivity of results to departures from the commonly used MAR assumption especially in multilevel data contexts. The fourth is circumventing computational burden when jointly analysing a multivariate vector of outcomes.

1.5 Objectives of the Study

General Objective

The aim of this research study is to develop a robust analysis framework for inpatient data measuring quality of care received by children admitted to hospitals in Kenya. The proposed analysis framework will address the challenges of analysing multivariate vectors of clustered outcomes (quality of care outcomes that are captured using more than a single measure with such measures often showing correlation), and holistically handling problems of missing data.

Specific objectives

The specific objectives of this study are to:

(i) Analyse a composite outcome accounting for covariate missingness and clustering at clinician and hospital level respectively.

(ii) Conduct multiple imputation when the study outcome is a composite of partially observed components.

(iii) Examine sensitivity of results to departures from the commonly assumed missing at random mechanism.

(iv) Jointly analyse multiple clustered outcomes under the correlated random-effects joint model applying pairwise fitting and pseudo-likelihood methods.

1.6 Significance of the Study

Through application and extension of existing biostatistics methods, this study will provide proper statistical analysis framework of partially observed multivariate hierarchical data often encountered by researchers in paediatric health research. Specifically, through objectives (i) and (ii), the study will demonstrate practical utility of advanced biostatistical analyses methods with an aim to promote their use in reporting of paediatric routine care studies. Through objectives (iii) and (iv), the study will extend existing methods thus adding new knowledge to the biostatistical tool kit.

These methods are not only be applicable to paediatric routine care context but also generalizable to other multivariate data contexts with missing data.

1.7 Literature Review

1.7.1 Routine Paediatric Care

In the recent past, there has been a steady growth of paediatric routine care literature on mortality and burden of common childhood diseases in LMIC settings. In 2013, CIN was established in Kenya by the KEMRI-WTRP in collaboration with the ministry of health and 14 county level hospitals and other partner institutions (Tuti et al., 2015). CIN is among the largest inpatient paediatric databases in the Sub-Saharan Africa region and was established with an aim to promote adoption and delivery of recommended paediatric clinical guidelines by health care provider for better patients' outcomes (English, 2013).

Out of CIN, several studies on management of common childhood illnesses such as malaria (Amboko et al., 2016), severe acute malnutrition (Gachau et al., 2018), pneumonia (Agweyu et al., 2018a; Tuti et al., 2017), rickets (Karuri et al., 2017) diarrhoea and dehydration (Akech et al., 2018) among children aged 1-59 months have been reported. Other studies have examined adherence to guidelines in documentation of clinical indicators (Gachau et al., 2017), monitoring of vital signs (Ogero et al., 2018) blood transfusion (Thomas et al., 2017) and management of shock (Mbevi et al., 2016) using CIN database. Other examples related to paediatric care studies in Kenyan hospitals include Ayieko et al. (2012),Mwaniki et al.

(2014), Opondo et al. (2016) and Muthee et al. (2018). Elsewhere Hau et al. (2018) studied post-hospital mortality among children in Tanzanian hospitals, while Gordon et al. (2013) studied the prevalence and burden of diseases in general paediatric wards in Ethiopia.

1.7.2 Missing Data in Routine Care

Missing data is a common problem in epidemiological studies (Bartlett et al., 2015). However, consequences of missing data on inference were neglected until Rubin introduced three missing data mechanisms through a seminar paper (Rubin, 1976). Since then, there has been a tremendous growth of robust missing data methods to mitigate the effects of incomplete data on statistical analysis and inference (Molenberghs et al., 2014). Some of the work are by (Molenberghs and Kenward, 2007) with a focus on clinical studies and (Daniels and Hogan, 2008) with an emphasis on longitudinal studies. More recently, Carpenter and Kenward (2013) and Van Buuren (2018) provided an overview of multiple imputation and its application in a range of complex data structures in medical and social sciences. A review of paediatric literature revealed that the choice of missing data handling methods by various researchers varied across studies. For instance, some of researchers did not acknowledge missing data in their studies e.g. Gachau et al. (2017) and Thomas et al. (2017). Others including Gordon et al. (2013); Mwaniki et al. (2014); Mbevi et al. (2016); Gachau et al. (2018) and Hau et al. (2018), used complete case analysis to handle missing data. Despite its limitations, complete case analysis method is the default methods in most standard statistical software such as R, SAS, STATA, and SPSS, among others. In other

studies including Opondo et al. (2016); Bohlius et al. (2016); Hooli et al. (2016); Malla et al. (2017); Tuti et al. (2017); Agweyu et al. (2018a,b); Ogero et al. (2018); Akech et al. (2018); Agweyu et al. (2018a) and Agweyu et al. (2018b), MI assuming a MAR mechanism was used to handle missing data in the respective studies. However, details of imputation models used are rarely reported in most of the above-mentioned studies thus, hindering replication and verification of MI methods used. For instance, compatibility between the imputation model and substantive models could not be ascertained in multilevel study contexts due to lack of sufficient imputation model details. Additionally, tests and assessment of imputation model convergence were hardly reported.

As already mentioned, MI in its standard application assumes a MAR mechanism; an assumption that cannot be confirmed using observed data alone, hence the need for sensitivity analysis (Carpenter and Kenward, 2013; Molenberghs et al., 2014). In missing data literature, sensitivity analysis can be implemented within selection models, pattern mixture and shared parameter model framework, respectively (Molenberghs et al., 2014). Nevertheless, sensitivity analysis to assess robustness of inference under MAR assumption are rarely conducted and reported in practice (Mackinnon, 2010; Smuk et al., 2017). For instance, among reviewed paediatric routine care studies, only three reported assessing departures from MAR assumption (Agweyu et al., 2018a,b; Gathara et al., 2017). In these studies, observations were partitioned according to patterns of missingness (i.e., observations with no missing data, observations with one to three incomplete variables and observations with more than three partially variables). Missing

values were then imputed multiple times and the substantive model of interest fitted to each pattern independently. Thereafter, final estimates were pooled across the three patterns weighted by the proportions of individuals in each pattern per variable (Gathara et al., 2017; Agweyu et al., 2018a). Besides splitting observations along the patterns of missingness, no further details were provided on how uncertainty reflecting MNAR mechanism was incorporated in the imputation model.

1.7.3 Multiple Outcomes in outine Care

As earlier mentioned, researchers may be interested in multiple outcomes spanning several domains of care in order to assess trends in adherence to recommended clinical practices. This section presents a brief review of common approaches of handling multiple outcomes in clinical studies.

1.7.4 Multiple Univariate Analysis

In the literature, the most common approach of handling such outcomes is multiple univariate analysis. This approach involves estimating the effect of a set of covariates on each outcome separately and is most suitable when the outcomes of interest are independent (Huberty and Morris, 1989). In paediatric literature, multiple univariate analysis approach has been previously used to analyze multiple quality of care indicators by several authors including (Gachau et al., 2017, 2018; Gathara et al., 2017). Although this approach is straightforward, analyses and reporting can be time consuming and cumbersome when the number of outcomes is large (Eapen et al., 2011).

Alternatively, a single primary outcome is specified while all other outcomes are treated as secondary. The primary outcome is analysed and interpreted formally while the secondary outcomes are analysed for exploratory purposes (Pocock et al., 1987). This strategy was employed in the initial analysis of pneumonia trial data where the trial investigators defined correct diagnosis and classification of pneumonia cases as the primary outcome while the assessment and treatment domain outcomes were treated as secondary outcomes (Ayieko et al., 2017, 2019). A major limitation of this approach is the potential loss of power owing to selection and analysis of one primary outcome among several other outcomes that address/describe different aspects of care (Pocock et al., 1987; Bebu and Lachin, 2018).

Composite Outcomes

In health care settings, composite scores which combine multiple outcomes into a single summary measures have been used as scorecards to measure and benchmark performance and quality of care in neonatal (Profit et al., 2010), peadiatrics, as indicated by Opondo et al. (2016, 2018) and adult cardiovascular studies (Caldis, 2007; Chen et al., 2013; Eapen et al., 2011; EUnetHTA, 2013). A key advantage of composite outcomes over individual outcomes is increased statistical efficiency (Freemantle et al., 2003). However, missing data at item level (composite sub-components) or at score level may undermine their reliability (Caldis, 2007; Profit et al., 2010).

Joint Modelling

With a multivariate vector of outcomes, joint models can be used to test hypotheses of joint effects of covariates on the various outcomes simultaneously, and to study the association among the different outcomes (Fieuws and Verbeke, 2004, 2006; Fitzmaurice et al., 2009a; Molenberghs and Verbeke, 2005). In the literature, joint modelling is commonly used to analyse two outcomes, but it can be extended to jointly model three or more outcomes. However, computationally complexity arises in high dimensional joint modelling settings (i.e., when the number of outcomes is greater than three) (Molenberghs and Verbeke, 2005, Chapter 25). In our literature review of paediatric routine, we did not come across any study using high dimensional joint models to analyze multiple outcomes.

1.8 Thesis outline

The remainder of this report is organized as follows: Notations, fundamental concepts and general methods used throughout this report are outlined in Chapter 2. In Chapter 3, we construct and analyze pneumonia paediatric admission quality of care (PAQC) score while addressing missing covariate in a multilevel data context using MI method. Strategies for handling incomplete components of a composite outcome is the subject of Chapter 4. The relative merits of the different approaches are highlighted via a simulation study conducted. In Chapter 5, the study integrates sensitivity analysis within the pattern-mixture modelling framework to assess departures from MAR missing data mechanism assumed in

Chapters 3 and 4, respectively. In Chapter 6, we analyze high dimensional pneumonia outcomes using correlated random-effects joint modelling approach at the same time addressing covariate missingness. We conclude in Chapter 7 with a discussion and recommendations for further research.

Chapter 2

Fundamental Concepts and Research Methodology

2.1 Introduction

This chapter introduces important concepts and methods that will be used throughout this thesis report. In particular, an overview of model families for clustered data is presented in Section 2.2 while basic missing data concepts and terminologies are presented in Section 2.3. This is followed by a review of methods used to handle missing data in Section 2.4. Section 2.5 provides a detailed account of multiple imputation.

2.2 Model Families for Clustered Data

Clustered data are common in epidemiological studies. Such data arise as a result of natural clustering, such as children from the same household. Cluster-

correlated data may also occur by study design. For instance, one can consider systolic/diastolic blood pressure measurements collected on the same patients over a period of time. Consequently, measurements from the same subjects/cluster tend to be more alike than measurements from different subjects/clusters (Agresti, 2002).

In the literature, there are several approaches of analyzing cluster-correlated data and the choice of the method to use largely depends on the nature of the outcome and study objectives (Molenberghs and Verbeke, 2005, Chapter 5.). This section briefly introduces two model families commonly used in the analysis of clustered data. These include the random effects and marginal family of models.

2.2.1 Random-effects Models

Letting Y_{ij} be the j^{th} outcome measured for cluster (subject) i, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, it is assumed that conditional on the random-effects (b_i), the outcomes, Y_{ij} are independent and belong an exponential distribution of the form

$$f_i(y_{ij}|b_i) = exp\left\{\phi^{-1}\left[y_{ij}\theta_{ij} - \psi(\theta_{ij})\right] + c(y_{ij}, \phi)\right\}. \tag{2.1}$$

It further follows that

$$E(Y_{ij}|bi) = \mu_{ij} = h^{-1}\left\{E(Y_{ij}|\boldsymbol{b}_i, \boldsymbol{X}_{ij}, Z_{ij})\right\} = \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{'}b_i, \tag{2.2}$$

where $h^{-1}(\cdot)$ is a known link function, $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ are vectors of known covariate values with fixed and random-effects respectively, $\boldsymbol{\beta}$ is vector of unknown fixed

regression coefficients, ϕ is a scale (dispersion) parameter, and θ_{ij} is the canonical (natural parameter) which is a function of the linear predictor (Molenberghs and Verbeke, 2005, p. 27-28). Random-effects models are useful when research interest lies in drawing inference with respect to subject specific parameters denoted by b_i. These models are also useful when subject specific predictions are of interest (Molenberghs and Verbeke, 2005, Chapter. 13).

Linear Mixed Model

When the outcome is continuous, an identity link function is used and the random-effects model is commonly known as linear mixed model (LMM). A LMM is formulated as

$$\begin{aligned} Y_{ij} &= \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{'}\boldsymbol{b}_i + e_{ij}, \\ b_i &\sim N\left(\mathbf{0}, \boldsymbol{D}\right), \quad e_{ij} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \end{aligned} \tag{2.3}$$

The random-effects b_i are assumed to be sampled from a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{D}$ while the residual errors are multivariate normal with mean $\mathbf{0}$ and variance covariance $\boldsymbol{\Sigma}$.

Generalized linear mixed model

When the response of interest is discrete, generalized linear mixed model (GLMM) is the most frequently used random-effects models (Fitzmaurice et al., 2009a, Chapter. 4). In general, a GLMM is formulated as

$$h^{-1}\left\{E(Y_{ij}|\boldsymbol{b}_i, \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij})\right\} = \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{'}\boldsymbol{b}_i. \tag{2.4}$$

The choice of the link function $h^{-1}(\cdot)$ in GLMM depends on the nature of the outcome (Molenberghs and Verbeke, 2005, p. 27). For instance, when the responses are binary, a mixed logit model is considered appropriate, that is,

$$logit(\mu_{ij}) \equiv log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{'}b_i,$$

$$Y_{ij}|\mu_{ij} \sim Bernoulli(\mu_{ij}).$$

Alternatively, a probit mixed model defined below is considered

$$\Phi^{-1}(\mu_{ij}) = \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta} + Z_{ij}^{'}b_i,$$

where $\Phi^{-1}(\cdot)$ is the probit link or the inverse standard normal cumulative distribution function. When the responses are ordinal, a cumulative logit/probit link is considered appropriate (Agresti, 2002, Chapter 7, p. 275). On the other hand, count data are modelled using a mixed Poisson model with a log link (Agresti, 2002) as defined below

$$log(\mu_{ij}) = \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta} + Z_{ij}^{'}b_i,$$

$$Y_{ij}|\mu_{ij} \sim Poisson(\mu_{ij}).$$

2.2.2 Marginal Models

Letting Y_{ij} be the j^{th} outcome measured for cluster (subject) i, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, a marginal model for clustered data is specified as

$$h^{-1}(\mu_{ij}) = E(Y_{ij}|\boldsymbol{X}_{ij}) = \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta}. \tag{2.5}$$

In this case, the vector of regression parameters ($\boldsymbol{\beta}$) describe population averaged means (Molenberghs and Verbeke, 2005, Chapter 5, p. 48). For marginal models with discrete outcomes, specification of the joint multivariate distribution of the Y_{ij} is computationally challenging. As an alternative, semi-parametric marginal models are used to circumvent computational complexity associated with full likelihood. A common semi-parametric estimation method for marginal models is the generalized estimating equations (Fitzmaurice et al., 2009a, Chapter 3) and (Molenberghs and Verbeke, 2005, Chapter 8).

Generalized estimating equations

GEE model proposed by Liang and Zeger (1986) is quasi-likelihood marginal model applicable to both continuous (Gaussian) and discrete outcomes (non-Gaussian) settings. To circumvent computation challenges highlighted above, the mean response and within subject association (the association among responses) are modelled separately. The association is considered as a nuisance characteristic of the data that must considered in order to make correct inferences about changes in the population mean response. The quasi-likelihood estimator of

marginal regression parameters, β can be obtained by solving the following quasi-likelihood score equations,

$$S(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta'} (A_i^{1/2} R_i(\alpha) A_i^{1/2})^{-1} (Y_i - \mu_i) = 0, \tag{2.6}$$

where A_i is a diagonal matrix with the marginal variances $Var(Y_i) = \phi v(\mu_i)$ along the diagonal, ϕ is a dispersion parameter and $R_i(\alpha)$ is an $n_i \times n_i$ working correlation matrix. The working correlation is a function of the nuisance association parameters, α (Liang and Zeger, 1986; Molenberghs and Verbeke, 2005; Fitzmaurice et al., 2009a). Some of the commonly used working correlation structures include:

i) Independence working correlation structure defined by

$$\rho_{ijk}(\alpha) = corr(Y_{ij}, Y_{ik}; \alpha) = 0 \ \ for \ \ all \ \ j \neq k.$$

ii) First-order auto regressive $(AR(1))$ working correlation structure which is defined by

$$\rho_{ijk}(\alpha) = Corr(Y_{ij}, Y_{ik}; \alpha) = \alpha^{|j-k|}, \ \ for \ \ all \ \ j \neq k,$$

where the α lies in the interval $[0, 1]$. The correlation decreases with an increase in time between measurements, that is $(|j - k|)$ (Fitzmaurice et al., 2009a).

iii) Exchangeable working correlation structure defined by,

$$\rho_{ijk}(\alpha) = Corr(Y_{ij}, Y_{ik}; \alpha) = \alpha, \ \ for \ \ all \ \ j \neq k.$$

iv) Unstructured working correlation structure which is defined by

$$\rho_{ijk}(\alpha) = Corr(Y_{ij}, Y_{ik}; \alpha) = \alpha_{jk} \ \ for \ \ all \ \ j \neq k.$$

In practice, exchangeable and independence working correlation structures are applicable in a wide range of clustered correlated data settings compared to $AR(1)$ and unstructured working correlation structures (Molenberghs et al., 2014, Chapter 3 , p. 50). In the event that adopted working correlation structure strongly differs from the true underlying structure in a given setting, the consequences are loss of efficiency. But even then, the estimator $\hat{\beta}$ obtained by solving (2.6) remains consistent and asymptotically normally distributed with mean β. Moreover, point estimates and empirically corrected (robust) standard errors based on the sandwich variance estimator are asymptotically correct whether the working correlation structure is correct or not (Liang and Zeger, 1986).

2.2.3 Relationship between Marginal and Random-effects Models

The underlying differences between random effects and marginal models reflect the distinct targets of inference associated with the two model families. Specifically, random effects models are appropriate when one is interested in the effects of covariates on changes in an individual$'$s response, while marginal models are more useful when population-averaged covariates' effects are of interest (Molenberghs and Verbeke, 2005, Chapter . 6). For linear mixed models with identity link (i.e., models continuous outcomes), a random-effects model implies marginal model. That is, the mean of marginal models can be obtained by conditioning the random vector $\boldsymbol{b}_i$ to zero or by marginalizing over the distribution of random-effects. In this case, the vector of fixed-effects denoted $\boldsymbol{\beta}$ has both population-averaged and subject-specific interpretation (Molenberghs and Verbeke, 2005, p. 48).

In contrast, when the outcomes are discrete, the relationship between marginal and random-effects models is not straightforward because it involves taking the mean of a non-linear function of b_i, (Fitzmaurice et al., 2009a, p.34). That is,

$$\begin{aligned} \mu_{ij} &= E(\boldsymbol{Y}_i|\boldsymbol{X}_i) \\ &= E([E(\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{b}_i)] \\ &= E\left[h(\boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i)\right. \\ &= \int_{-\infty}^{\infty} h(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i) f(b_i) db_i, \end{aligned} \tag{2.7}$$

where $f(\boldsymbol{b}_i)$ is the joint probability density function for $\boldsymbol{b}_i$, and

$$h^{-1}\left\{E(\boldsymbol{Y}_i|\boldsymbol{X}_i,\boldsymbol{b}_i)\right\} \neq X_i\beta,$$

for any β. To illustrate this, consider a binary response modelled via a logistic regression model with random intercepts,

$$logit(E(Y_{ij}|x_{ij},b_i) = \beta_0 + \beta_1 x_i + b_i,$$

where $b_i \sim N(0,\sigma_b^2)$. The corresponding model for the marginal probability of success is

$$\begin{aligned}
\mu_{ij} &= E(Y_{ij}|x_{ij}) \\
&= E(\left[E(Y_{ij}|x_{ij},b_i)\right] \\
&= E\left[\frac{exp(\beta_0+\beta_1 x_{ij}+b_i)}{1+exp(\beta_0+\beta_1 x_{ij}+b_i)}\right] \\
&= \int_{-\infty}^{\infty} \frac{exp(\beta_0+\beta_1 x_{ij}+b_i)}{1+exp(\beta_0+\beta_1 x_{ij}+b_i)} \times \frac{1}{\sqrt{2\pi\sigma_b^2}} exp\left(-\frac{1}{2}\frac{b_i^2}{\sigma_b^2}\right) db_i.
\end{aligned} \tag{2.8}$$

The marginal mean in (2.8) is insolvable (Fitzmaurice et al., 2009a, p. 34) and therefore,

$$E\left[\frac{exp(\beta_0+\beta_1 x_{ij}+b_i)}{1+exp(\beta_0+\beta_1 x_{ij}+b_i)}\right] \neq \frac{exp(\beta_0+\beta_1 x_{ij})}{1+exp(\beta_0+\beta_1 x_{ij})},$$

for any β. This shows that conditioning the random vector $\boldsymbol{b}_i$ to zero in a GLMM does not lead to marginal mean (Fitzmaurice et al., 2009a, p. 34). Therefore,

regression parameters are not comparable to regression parameter in marginal models. In practice, the (G)LMMs and GEE models are commonly used in analysing one outcome at a time (i.e., univariate analysis). However, (G)LMMs are also applicable in the joint modelling of two or more outcomes. In joint modelling context, it is assumed that the outcomes share or have correlated random-effects $\boldsymbol{b}_i$ (Molenberghs and Verbeke, 2005, Chapters. 24-25).

2.3 Methods of Handling Data with Missing Values

In missing data literature, there are several missing data handing methods ranging between simple and complex methods. In this section, we review some of the methods highlighting their strengths and limitations.

2.3.1 Simple Methods

Listwise Deletion

In listwise deletion also known as complete case analysis (CCA), analysis is based on a subset of complete records after exclusion of case all records with missing information. CCA is widely used in practice due to its ease of application. Besides, it is the default missing data handling method in most statistical software such as R, SAS and STATA. When the underlying mechanism is MCAR or covariates dependent MAR, CCA yields valid but inefficient parameter estimates (White and Carlin, 2010). The loss of efficiency is as a result of estimation based on a reduced sample size and it is characterized by inflated standard errors . On

the other hand, when the underlying missing data mechanism is outcome dependent MAR, the expected results are both biased and inefficient (Molenberghs et al., 2014, Chapter 1). Bias refers to lack of generalization of sample estimates to target populations parameters.

Single Imputation

This method involves imputing missing data values once yielding a single complete data set (Eekhout, 2015). Single imputation method can be implemented in one of the following approaches:

1. *Marginal mean imputation*

   Missing values are replaced with the mean of the observed values for that variable ignoring all other variables. Mean imputation is more relevant in the imputation of continuous variables (Molenberghs et al., 2014, Chapter 2, p. 36).

2. *Regression mean imputation*

   In this approach, variables with complete observations are used to predict the values of the missing observations. For example, considering two variables, X (a fully observed covariate) and Y (partially observed response). The regression mean is found by regressing Y on X that is, $Y_i = \beta_0 + \beta_1 X_i$. The estimates of β_0 and β_1 denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively are obtained and used to impute missing $Y_i's$ using $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. In this method, imputed values fall on a regression line without random variation (error). Consequently, correlation between predictor variables and the missing out-

come is overestimated (Eekhout, 2015).

3. *Stochastic Regression Imputation*

   Stochastic regression imputation reduces bias in regression mean imputation by adding an error term to each predicted value (Eekhout, 2015). That is, imputing missing observations with $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \varepsilon_i$ where ε_i is the normally distributed error term with mean 0 and a variance equal to the residual variance from the regression of the predictor on the outcome (Eekhout, 2015). Inclusion of the error term preserves variability in the data and parameter estimates are unbiased with MAR data. Nonetheless, the method does not account for uncertainty about the imputed values and the standard errors tends to be underestimated (Enders, 2010).

4. *Matching Method (Hot Deck Imputation)*

   This method matches missing data records with observed data records. Specifically, records are stratified into separate homogenous groups and missing values are imputed by the observed values of the closest match within the subsets (Andridge and Little, 2011). This method is common in survey research (Little and Rubin, 2002; Eekhout, 2015).

5. Last Observation Carried Forward (LOCF)

   In this method the last observation is substituted whenever a value is missing. This method is mainly applicable in longitudinal studies exhibiting monotone missing data pattern due to attrition/dropout. In LOCF, it is assumed that the observation of the individual does not change after the last measured observation. This assumption is often unrealistic (Molenberghs

and Verbeke, 2005, Chapter 27, p. 493).

A common limitation of the single imputation methods above is that imputed values are treated as if they were the original observations ignoring variability due to imputations. This issue leads to substantial underestimation of the standard errors and biased estimates especially when data are not MCAR or covariate dependent MAR (Little and Rubin, 2002; Eekhout, 2015).

2.3.2 Maximum-likelihood Methods

With incomplete data, likelihood inference is based on maximizing the likelihood of the observed data (Little and Rubin, 2002), that is,

$$L(\boldsymbol{\theta}, \boldsymbol{\psi} | \boldsymbol{Y}_i^{obs}, \boldsymbol{R}_i) = c \times \prod_{i=1}^{N} \int f(\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss}, \boldsymbol{R}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) d\boldsymbol{Y}_i^{miss}, \tag{2.9}$$

where c is a constant factor that is independent of the measurement process parameter vector ($\boldsymbol{\theta}$) and the missingness process parameter vectors ($\boldsymbol{\psi}$). However, when ignorability is assumed, then maximum likelihood estimates are obtained by maximizing (2.14), where the likelihood contribution of the i^{th} study subject is $f(\boldsymbol{Y}_i^{obs} | \boldsymbol{\theta})$ (Molenberghs et al., 2014). It is further assumed that missing values can be validly predicted using the conditional mean, that is, $E(\boldsymbol{Y}_i^{miss} | \boldsymbol{Y}_i^{obs}, \theta)$ and covariance model of the observed data (Molenberghs et al., 2014). In this report, we briefly discuss two methods for obtaining estimates using maximum-likelihood estimation, that is, the expectation-maximization (EM) algorithm and full information maximum likelihood $(FIML)$.

Expectation-Maximization Method

The Expectation Maximisation (EM) is a two-step iterative process developed to compute maximum likelihood (ML) estimates in parametric models in the presence of missing data (Dempster et al., 1977; Molenberghs and Verbeke, 2005; Molenberghs and Kenward, 2007; Fitzmaurice et al., 2009a). The algorithm assumes that data are MAR and it iterates between the E and M step as follows.

E-step

The E-step calculates the conditional expectation of the complete-data log-likelihood given the observed data and parameter estimates from previous iteration. Complete-data likelihood function is calculated by filling in the missing elements in the likelihood with their expected values, given $\boldsymbol{Y}_i^{obs}$ and a current set of parameter estimates $\boldsymbol{\theta}^r$ (Molenberghs and Verbeke, 2005), that is,

$$L(\boldsymbol{\theta}|\boldsymbol{\theta}^r) = \int log(\boldsymbol{\theta}, \boldsymbol{Y}_i) f(\boldsymbol{Y}_i^{miss}|\boldsymbol{Y}_i^{obs}, \boldsymbol{\theta}^r) d\boldsymbol{Y}_i^{miss} = E[log(\boldsymbol{\theta}|\boldsymbol{Y}_i)|\boldsymbol{Y}_i^{obs}, \boldsymbol{\theta}^r]. \qquad (2.10)$$

M-step

Given the complete-data log-likelihood, the M-step then finds the parameter estimates θ^{r+1}, to maximize the complete-data log-likelihood from the E step (Molenberghs and Verbeke, 2005; Fitzmaurice et al., 2009a).

$$L(\boldsymbol{\theta}^{r+1}|\boldsymbol{\theta}^r) \geq L(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) \qquad for\ all \quad \boldsymbol{\theta}. \qquad (2.11)$$

The iteration process is repeated between the E and M steps until convergence yields the ML estimate of θ (Molenberghs and Verbeke, 2005; Fitzmaurice et al., 2009a).

Although the EM algorithm always converges to a final solution, it is rarely used in practice due to computational complexities and slow convergence. Moreover, the algorithm does not provide standard errors for the parameter estimates (Molenberghs and Verbeke, 2005; Fitzmaurice et al., 2009b) and considerable model specifications are needed to obtain them (Louis, 1982).

Full Information Maximum Likelihood/Direct Maximum Likelihood Method

In full information maximum likelihood (FIML)/ direct maximum likelihood method, missing data are handled directly within the analysis model during estimation (Enders, 2010). FIML assumes that data are MAR (Allison, 2012). For illustration, suppose there are N independent observations on p fully observed variables. If the first variable y_1 is missing for a particular observation i, then the joint probability for observation y_{i2} to y_{ip} is the probability of observing y_{i2} to y_{ip} (Allison, 2012). When y_{i1} is continuous, then the joint probability is integrated over all possible values of the partially observed variable as follows

$$f_i^*(y_{i2}, y_{i3}, \ldots, y_{ip}; \boldsymbol{\theta}) = \int_{y_{i1}} f_i(y_{i1}, y_{i2}, \ldots, y_{ip}; \boldsymbol{\theta}) dy_{i1}. \tag{2.12}$$

When y_1 is discrete, then the joint probability is summed over all possible values as shown below,

$$f_i^*(y_{i2}, y_{i3}, \ldots, y_{ip}; \boldsymbol{\theta}) = \Sigma_{y_1} f_i(y_{i1}, y_{i2}, \ldots, y_{ip}; \boldsymbol{\theta}). \tag{2.13}$$

When there are m fully observed cases and $N - m$ partially observed cases on y_1, then the likelihood function for the full data set is the product of the likelihoods for all the observations. That is,

$$L = \prod_{i=1}^{m} f_i(y_{i1}, y_{i2}, \ldots, y_{ip}; \boldsymbol{\theta}) \prod_{i=m+1}^{N} f_i^*(y_{i2}, y_{i3}, \ldots, y_{ip}; \boldsymbol{\theta}). \tag{2.14}$$

The main limitation of FIML method is that it does not allow inclusion of auxiliary variables and cannot handle missing covariates data in second and higher levels of multilevel data (Grund et al., 2018).

2.3.3 Multiple Imputation

Multiple imputation (MI), proposed by Rubin (1987) is the most recommended method for obtaining valid parameter estimates from partially observed data (Molenberghs and Verbeke, 2005; Molenberghs and Kenward, 2007; van Buuren and Groothuis-Oudshoorn, 2011; Carpenter and Kenward, 2013; Enders et al., 2016; Grund et al., 2018). In its standard application, MI assumes a MAR mechanism. However, missing data can also be imputed assuming a MNAR mechanism (Carpenter and Kenward, 2013, Chapter. 10). MI involves three sequential steps (Rubin, 2004; Carpenter and Kenward, 2013).

Step 1: Independent random samples are drawn from the posterior predictive distribution (Bayesian framework) of the missing values given the observed data and a statistical imputation model, thus generating more than one filled-in data sets (Rubin, 1976). In the Bayesian perspective, missing data ($\boldsymbol{Y}_i^{miss}$) are treated as an additional set of nuisance parameters (Carpenter and Kenward, 2013). Thus, a joint posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{Y}_i^{miss}$ is given by

$$P(\boldsymbol{\theta}, \boldsymbol{Y}_i^{miss}|\boldsymbol{Y}_i^{obs}) = P(\boldsymbol{\theta}|\boldsymbol{Y}_i^{miss}, \boldsymbol{Y}_i^{obs})P(\boldsymbol{Y}_i^{miss}|\boldsymbol{Y}_i^{obs}). \tag{2.15}$$

The corresponding marginal posterior distribution for $\boldsymbol{\theta}$ given $\boldsymbol{Y}^{obs}$ is

$$P(\boldsymbol{\theta}|\boldsymbol{Y}_i^{obs}) = \int P(\boldsymbol{\theta}, \boldsymbol{Y}_i^{miss}|\boldsymbol{Y}^{obs})d\boldsymbol{Y}_i^{miss}, \tag{2.16}$$

and it can be regarded as the Bayesian equivalent to the observed-data likelihood in (2.14) (Little and Rubin, 2002). At iteration r $(r = 1, 2, \dots)$, the Markov Chain Monte Carlo (MCMC) algorithm behind MI simulates from the joint posterior distribution by iterating between a posterior step and an imputation step (Carpenter and Kenward, 2013) as follows,

$$\begin{aligned} \boldsymbol{\theta}^{(r+1)} &= P(\boldsymbol{\theta}|\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{miss(r)}) && (Posterior \quad step) \\ \boldsymbol{Y}_i^{miss,(r+1)} &= P(\boldsymbol{Y}_i^{miss}|\boldsymbol{Y}_i^{obs}, \boldsymbol{\theta}^{(r+1)}) && (Imputation \quad step). \end{aligned} \tag{2.17}$$

The sequence of iterations converges in distribution to $P(\boldsymbol{\theta}, \boldsymbol{Y}_i^{miss}|\boldsymbol{Y}_i^{obs})$ for a large number of iterations. The algorithm is repeated a number of times, say M, resulting in M copies of the original data with missing data filled in by the imputed

values (Rubin, 1976; Van Buuren, 2018; Carpenter and Kenward, 2013).

Step 2: Imputed data sets are analyzed using standard statistical methods. The choice of statistical method depends on several factors such as the type of outcome, study design and research question for a given data set.

Step 3: Pooled estimates are obtained by averaging over the parameter estimates from all multiply imputed data sets according to Rubin's rules (Rubin, 1976). The pooled MI estimator for $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_{MI} = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\beta}}_m, \tag{2.18}$$

with variance estimator

$$\hat{\boldsymbol{V}}_{MI} = \boldsymbol{W} + \left(\frac{M+1}{M} \right) \times \boldsymbol{B},$$

where

$$\boldsymbol{W} = \frac{1}{M} \sum_{m=1}^{M} \hat{\sigma}_m^2$$

is the average within imputation variance and

$$\boldsymbol{B} = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\boldsymbol{\beta}}_m - \hat{\boldsymbol{\beta}}_{MI})^2$$

is the between imputation variance.

A key advantage of MI over other missing data handling methods is that uncertainty about the missing values is taken into account. Besides, MI is flexible in that, the imputation phase is separate from the analysis phase. This allows

inclusion of auxiliary variables in the imputation model that are predictive of missing variables and the missingness mechanism (Meng, 1994; van Buuren and Groothuis-Oudshoorn, 2011; Carpenter and Kenward, 2013; Bartlett et al., 2015; Grund et al., 2018). In the literature and statistical software, MI is implemented in two broad frameworks, that is, joint modelling imputation framework (Schafer, 1997) and fully conditional specification imputation framework (Van Buuren, 2018). The following subsections provide a general introduction to MI within the two frameworks.

Joint Modelling Imputation Framework

In the joint modelling (JM) imputation framework, it is assumed that data can be described by a multivariate normal distribution of the form.

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i, \tag{2.19}$$

where $\boldsymbol{Y}_i$ denote a vector of incomplete variables, $\boldsymbol{X}_i$ is a vector of fully observed predictor variables and $\boldsymbol{e}_i$ is a vector of residuals which follows a multivariate normal distribution with mean zero and covariance matrix, Σ (Carpenter and Kenward, 2013; Molenberghs et al., 2014; Van Buuren, 2018). Suppose that $\boldsymbol{Y}_i = (\boldsymbol{Y}_i^{obs}, \boldsymbol{Y}_i^{obs})$ is any partially observed variable with an arbitrary missing data pattern, then imputations are generated in two steps as earlier noted. In the first step, the model parameters $(\boldsymbol{\theta} = \boldsymbol{\beta}, \boldsymbol{\Sigma})$, are drawn from their posterior distributions, given $\boldsymbol{Y}^{obs}$ and current imputations for $\boldsymbol{Y}_i^{miss}$. Secondly, new imputations for $\boldsymbol{Y}_i^{miss}$ are generated based on θ and $\boldsymbol{Y}_i^{obs}$ (Grund et al., 2017, 2018). Initial

values for $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{Y}_i^{miss}$ are estimated using the observed data (Carpenter and Kenward, 2013).

Multilevel MI in JM framework

In multilevel data context, partially observed variable at each level of the hierarchy are jointly specified as responses in a multilevel structural equation of the imputation model regardless of the missing data pattern (Carpenter and Kenward, 2013; Grund et al., 2017, 2018; Quartagno and Carpenter, 2018). Suppose we have two levels of hierarchy in the data at hand where i denotes level 1 units nested within level 2 units denoted by j, then a two level JM imputation model is defined by

$$\begin{aligned} \boldsymbol{Y}_{ij}^{(1)} &= \boldsymbol{X}_{ij}^{(1)}\boldsymbol{\beta}^{(1)} + \boldsymbol{Z}_{ij}^{(1)}\boldsymbol{b}_j^{(1)} + \boldsymbol{e}_{ij}^{(1)} \\ \boldsymbol{Y}_j^{(2)} &= \boldsymbol{X}_j^{(2)}\boldsymbol{\beta}^{(2)} + \boldsymbol{b}_j^{(2)}, \end{aligned} \tag{2.20}$$

where $\boldsymbol{Y}_{ij}^{(1)}$ and $\boldsymbol{Y}_j^{(2)}$ are vectors of partially observed level 1 and level 2 variables respectively, with corresponding vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ fixed effects. $\boldsymbol{X}_{ij}^{(1)}$ and $\boldsymbol{X}_j^{(2)}$ are vectors of fully observed variables used to predict partially observed level 1 and level 2 variables (Carpenter and Kenward, 2013, Chapter. 9, p.212). Covariates with random-effects are denoted by $\boldsymbol{Z}_{ij}$. Jointly, random-effects $\boldsymbol{b}_j^{(1)}$ and $\boldsymbol{b}_j^{(2)}$ are assumed to follow multivariate normal distributions with mean zero and covariance matrices $\boldsymbol{\Omega}$. Lastly, level 1 residuals $(\boldsymbol{e}_{ij}^{(1)})$ also follow a multivariate normal distribution with mean zero and variance denoted by $\boldsymbol{\Sigma}$ (Carpenter and Kenward, 2013).

Multilevel JM accounts for both the between-and within-cluster relations among variables (Grund et al., 2017). Incomplete level 1 variables are imputed conditionally on the observed data at level 1. On the other hand, partially observed level 2 variables are imputed conditionally on the observed data at level 2, in addition to the random-effects of the variables at level 1. More details on sampling algorithm are provided by Grund et al. (2018). Although JM assumes multivariate normal model, it extends easily to accommodate partially observed categorical data (ordered and nominal) (Carpenter and Kenward, 2013, p. 99). Specifically, JM imputation framework handles a nominal variable with S levels by including $S-1$ latent normal background variables. The $S-1$ latent normal variables correspond to different levels of the variable under consideration. For an ordinal variable with S levels, JM imputation model includes a single background variable, where the differences between categories are represented by a set of $S-1$ threshold parameters corresponding to different levels of the ordered variable. For more details on the computational aspects of different types of variables see (Carpenter and Kenward, 2013, Chapter 3-5).

Full Conditional Specification Framework

In fully conditional specification (FCS) framework, incomplete variables are imputed on a variable-by-variable basis (Van Buuren, 2018),(Molenberghs et al., 2014, Chapter 13,p. 275). Suppose that Y_{ip}, the p^{th} variable for the i^{th} subject is partially observed.

Further, if Y_{ip} follows a normal distribution, then the imputation model of interest corresponds to

$$Y_{ip} = \boldsymbol{Y}_{i(-p)}\boldsymbol{\beta}_p + e_{ip}, \tag{2.21}$$

where $\boldsymbol{Y}_{i(-p)}$ is the vector of predictor variables in the p^{th} imputation model, excluding the variable being imputed, $\boldsymbol{\beta}_p$ is a vector of regression coefficients and e_{ip} denotes normally distributed residual with mean zero and variance σ_p^2 (Grund et al., 2018). In this case, imputations for Y_{ip}^{miss} are drawn from the conditional distribution of missing data, given the observed data $\boldsymbol{Y}_{i(-p)}^{obs}$ and the most recent imputations for the missing data in other variables $\boldsymbol{Y}_{i(-p)}^{imputed}$.

The FCS approach accommodates relationships between variables by repeatedly conditioning them on one another. Specifically, it iterates back and forth between variables. When the target variable is continuous, a linear regression model is used with an assumption that the residuals are normally distributed. The FCS imputation framework also accommodates other data types such as ordered and unordered categorical data using generalized linear models. For example, if Y_{ip} is ordinal, an ordinal logistic imputation model is used while a multinomial logistic imputation model is specified for an unordered (nominal) variable. A Poisson regression model is considered an appropriate choice for a variable with count data (van Buuren and Groothuis-Oudshoorn, 2011).

Multilevel Multiple Imputation in FCS Framework

In multilevel data settings, the joint distribution of the variables can be approximated with a sequence of conditional models by multilevel FCS. For example, considering multilevel data set with 2 levels of clustering, missing data in a continuous variable can be addressed with univariate random-effects models (van Buuren and Groothuis-Oudshoorn, 2011; Grund et al., 2018). Specifically, imputation of p^{th} variable with missing data at level 1 using multilevel FCS approach can be based on the following set of models.

$$Y_{ijp}^{(1)} = \boldsymbol{Y}_{ij(-p)}^{(1)} \boldsymbol{\beta}_p^{(1)} + b_{jp}^{(1)} + e_{ijp}, \tag{2.22}$$

where $\boldsymbol{Y}_{ij(-p)}^{(1)}$ denotes all level 1 variables except $Y_{ijp}^{(1)}$ (or a subset of these) as well as the between-group components of the variables at level 1, $\boldsymbol{\beta}_p^{(1)}$ is a vector of regression coefficients. The random intercepts $b_{jp}^{(1)}$ and residuals e_{ijp}, are each assumed to follow a normal distribution with mean zero and variances $\boldsymbol{\Omega}_{1p}^2$ and σ_p^2 respectively (Grund et al., 2018). The FCS imputation model for q^{th} variable at level 2 is given by,

$$Y_{jq}^{(2)} = \boldsymbol{Y}_{j(-q)}^{(2)} \boldsymbol{\beta}_q^{(2)} + b_{jq}^{(2)}, \tag{2.23}$$

where $\boldsymbol{Y}_{j(-q)}^{(2)}$ represents all level 2 variables except the variable being imputed, that is, $Y_{jq}^{(2)}$ and $\boldsymbol{\beta}_q^{(2)}$ is a vector of regression coefficients. The random intercepts $b_{jq}^{(2)}$ are assumed to be normally distributed with mean zero and variance $\boldsymbol{\Omega}_q^2$.

The multilevel FCS approach iterates across all variables with missing data to address multivariate patterns of missing data. More details on the sampling algorithm are provided by Grund et al. (2018) and van Buuren and Groothuis-Oudshoorn (2011).

Considerations between FCS and JM Imputation Frameworks

For single level data with multivariate normal distribution, the FCS and JM imputation frameworks are equivalent (Enders et al., 2016; Meng, 1994). However, standard tasks such as the specification of the imputation model tend to be simpler under the JM framework. Such is the case because JM uses a single imputation model for all partially observed variables while FCS uses a separate imputation model for each incomplete variable (Grund et al., 2017, 2018). In multilevel data context with a mixture of variables types at second and higher levels of hierarchy, setting up an appropriate conditional models accounting for multilevel structures is more difficult in FCS framework compared to JM framework (Carpenter and Kenward, 2013; Grund et al., 2017, 2018; Enders et al., 2016). In subsequent chapters of this thesis, multilevel joint imputation framework was used to handle missing data in the pneumonia trial data.

Potential Drawbacks of MI

While MI is a useful tool in a wide range of missing contexts, there is risk of biased estimates leading to invalid inferences regardless of the imputation framework. Bias arises from poor or lack of convergence at imputation stage hence the need for monitoring imputed values using appropriate diagnostic tools/tests (Gelman

et al., 1992). Besides convergence issues, bias may also arise due to incompatibility between the imputation model and the substantive model of interest for a given data set. Specifically, incompatibility can occur when the imputation model omits some variables present in the analysis model of interest or when it incorrectly handles nonlinear effects, interactions and multilevel structures present in the analysis model (Bartlett et al., 2015). In this case, the imputation model is said to be poorer than the analysis model. The consequence of a poor imputation model is invalidity of Rubin's rules variance formula in addition to inconsistent parameter estimates in subsequent analyses and inferences (Carpenter and Kenward, 2013, p. 64).

Incompatibility can also occur when the imputation model is richer than the substantive model. That is, the imputation model has more variables than the analysis model interest. The additional variables (auxiliary variables) are included if they are thought to be predictive of the missingness mechanism. The consequences of a richer imputation model is overestimation of the sampling variability of the MI estimators though negligible in practice (Meng, 1994).

Multiple Imputation in Statistical Software

In practice, MI can be implemented in several standard statistical software including SAS, STATA, R and REALCOM, among others. In this study, we used R, which is an open source statistical software, to handle missingness in pneumonia trial data. Some of the MI packages in R, include *mice* (van Buuren and Groothuis-Oudshoorn, 2011) and *mi* (Su et al., 2011) packages, which are implemented in the FCS framework. Other MI packages in R include *jomo* (Quartagno

et al., 2019) and *mitml* (Grund et al., 2019), both implemented within the joint model framework.

Among FCS based packages, *mice* is the most commonly used package. However, the package does not have functionalities to impute categorical variables in the second and higher level of multilevel data structures. For this reason, the recently developed *jomo* and *mitml* packages, which allows MI of different variable types at any level of the hierarchical structure were preferred. A detailed review on capabilities and limitations of various imputation packages in multilevel data context is provided by (Grund et al., 2017).

Besides MI assuming MAR, incomplete variables can also be imputed assuming a MNAR mechanism for sensitivity analysis purposes (Carpenter and Kenward, 2013, p. 229). An example is MI with shift parameters (commonly known as the delta adjustment method) (Carpenter and Kenward, 2013; Tsiatis et al., 2014; Leacy et al., 2017). In software, MI with delta adjustment has been previously implemented in SAS (Yuan, 2014) with a generic function for the same in *mice* package in R (Van Buuren, 2018; Galimard et al., 2018). However, these generic functions are limited to single level data contexts thus hindering sensitivity analysis in multilevel data contexts.

2.4 Summary

This chapter provides a general background and justification of methods used in subsequent chapters of this thesis. In Chapter 3, an ordinal composite outcome will be constructed using 12 pneumonia care indicators. Thereafter, multilevel

joint imputation framework will be used to handle missing covariates across two levels of pneumonia trial data. In Chapter 4, multilevel joint imputation method will be used to handle missing pneumonia care indicators used to construct an ordinal composite outcome. In Chapters 3 and 4, MI will be conducted assuming a MAR mechanism and parameter estimates compared to those obtained from complete case analysis. At analysis stage, GLMM and GEE models with a cumulative logit link will be used to analyse an ordinal composite outcome. In Chapter 5, robustness of inference obtained in Chapters 3 and 4 will be assessed using two sensitivity analyses approaches within the pattern mixture models framework. In this case, missing data covariates will be imputed assuming a MNAR mechanism. In Chapter 6, nine out of 12 pneumonia care indicators previously used to construct the ordinal composite outcome will be modelled jointly using correlated random effects models. The choice of link functions will depend on the type of an individual outcome. Joint modelling will be conducted under complete case analysis and after MI of missing covariates across two levels of pneumonia trial data. Specific details on methods application and extensions are contained in the respective chapters.

Chapter 3

Analysis of Ordinal Hierarchical Data with Covariate Missingness

3.1 Introduction

Routine data are widely used in many health care settings to monitor the quality of care and to inform intervention programmes for better patients' health outcomes (Harries et al., 2013). Routine data can also be used to highlight areas of concern in clinical performance; thus, prompting actions and strategies to improve practice at individual or institutional levels (Omore et al., 2016). Prior studies show that quality of care vary across place and time despite standard clinical guidelines (Gachau et al., 2017). These variations can be attributed to multiple factors including changes in clinical guidelines, degree of task complexity and patient's characteristics, clinician characteristics in addition to organisational and contextual factors at hospital level.

While data from multiple sites enhance generalization of results to wider a population, it leads to complex hierarchical data structures, for instance, patients clustered within clinicians, who are then clustered within hospitals. Besides complex structures, routine data are subject to missing information at any level of hierarchy. Missing information may occur due to lack of documentation of care processes by health care providers, poor record keeping or limited health care technology at facility level (Harries et al., 2013; Lloyd et al., 2013; Houngbo et al., 2017). In the occurrence of missing data, appropriate missing data methods at analysis stage are recommended to avoid biased results (Carpenter and Kenward, 2013) informing clinical policies and ultimately leading to poor patients' care and outcomes (Rombach et al., 2018).

In the recent past, there has been an increase in literature on quality of care among children admitted with common childhood illnesses in LMICs, (Gathara et al., 2017; Opondo et al., 2016; Gachau et al., 2017; Thomas et al., 2017; Agweyu et al., 2018a). However, majority of these studies account for variation at patient and hospital levels, ignoring variation due to clinicians' characteristics in spite of their critical role in delivery of routine care (Rowe et al., 2005). Besides, missing data is a common problem in paediatric routine care and researchers use complete case analysis and MI to handle missingness. A major limitation of complete case records is biased, and inefficient parameter estimates due to information loss. Among studies reporting MI to handle missingness, the nature and details of the imputation model are rarely reported, which pose uncertainty about conclusions and barriers for replicate analyses. Furthermore, when missing data occur

in multilevel data context, incompatibility between the imputation model and the analysis models potentially leads to biased estimates, underestimated cluster level variances and overestimated individual level variances (Enders et al., 2016; Grund et al., 2018; Drechsler, 2015).

The main aim of this chapter is to construct and analyse pneumonia PAQC score, adapted to new WHO recommendations on assessment and treatment of inpatient paediatric pneumonia cases. This is in addition to addressing missing covariates while properly accounting for hierarchical structure in inpatient routine data set, that is, patients nested within clinicians who are in hospitals. Specifically, we analysed pneumonia trial date introduced in Chapter 1, Section 1.2.

The remainder of this chapter is structured as follows. Section 3.2 presents an outline of pneumonia PAQC score construction steps. This is followed by missing data methods used to handle missing covariates in pneumonia trial data. The methods section also presents statistical approaches for analysing ordinal responses followed by a review of Wald tests and likelihood ratio tests. Results are presented in Section 3.3 and we conclude with a discussion in Section 3.4.

3.2 Methods

3.2.1 Paediatric Admission Quality of Care (PAQC) score

PAQC score is an ordered composite measure developed to benchmark quality of care among children admitted with common childhood illnesses in LMIC settings (Opondo et al., 2016, 2018).

Table 3.1 presents a summary of the procedures used to construct PAQC score based on childhood pneumonia treatment guidelines (Organization, 2013). In the first step, we created binary indicators with one representing adherence to recommended paediatric pneumonia guidelines and zero 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 seven secondary signs and symptoms required for pneumonia severity classification; iii) incomplete documentation of all primary and secondary pneumonia signs and symptoms (Table 3.1).

The second PAQC score domain entails integration of information on presenting signs and symptoms by admitting clinician to correctly diagnose and classify pneumonia severity (i.e., severe pneumonia or pneumonia). For example, pneumonia was the correct diagnosis for a child who, in addition to cough and/or difficult breathing (primary signs), presented with lower chest indrawing or respiratory rate greater than 50 for patients aged 2-11 months (or respiratory rate less than 40 for patients aged 12-59 months) in the absence of all other secondary signs and symptoms. We created a binary indicator with one representing correct pneumonia severity classification (i.e., pneumonia severity documented in the medical record by the admitting clinician was in line with severity implied by presenting signs and symptoms) and zero representing misclassified pneumonia severity.

The third PAQC score domain (treatment) comprised two components; a binary

indicator with one corresponding to oral amoxicillin prescription and zero representing inappropriate care due to missing prescription or documentation in the case record that oral amoxicillin was not prescribed. For patients prescribed oral amoxicillin, we created a new variable "recommended dose per kilo body". That is, the actual dose given at point of care divided by patient's weight. We then transformed the new variable into a binary form with one representing correct oral amoxicillin dose (i.e., dose between 32 and 48 international units (IU) per Kilogram (Kg) every 12 hours) and zero representing either missing or wrong oral amoxicillin dosage (under dose for < 32 IU/Kg or over dose for $>$48IU/Kg), missing or wrong frequency of oral amoxicillin administration (e.g., administration frequency of once every 24 hours instead of once every 12 hours) or both.

In the second and final step, we summed all the six binary indicators spanning assessment (n=3), clinical diagnosis (n=1) and treatment (n=2) domains to obtain PAQC score. The score ranges between zero and six where a minimum score of zero corresponds to inappropriate pneumonia care and maximum score of six represents total adherence to recommended clinical guidelines across domains of care.

To visualize adherence to paediatric pneumonia guidelines during the trial period, we first calculated the mean monthly PAQC score for each trial arm and thereafter plotted the LOESS smoothing curves and the corresponding 95% confidence bands.

Table 3.1: Pneumonia care indicators used in PAQC[a] score construction

Quality of care domain	Pneumonia care indicators	Binary indicators
1. Assessment		
Primary S&S[b]	Cough, difficult breathing	1: if both primary S&S are documented 0: if at least one is not documented
Secondary S&S	Oxygen saturation, AVPU[c] ability to drink,central cyanosis grunting, respiratory rate lower chest wall indrawing	1: if all secondary S&S are documented 0: if at least one is not documented
Complete assessment	All primary and secondary S&S	1: if all are documented 0: if at least one S&S is not documented
2. Diagnosis and classification	Pneumonia diagnosis and classification	1 : for correct diagnosis classification 0: incorrect diagnosis classification
3. Treatment	Amoxicillin prescription	1: if oral amoxicillin was prescribed 0 : if amoxicillin was not prescribed
	Amoxicillin dosage	1 : if dose ranges between 32-48 (IU/Kg)every 12 hours. 0 :if dose is missing or $<$ 32 IU/Kg (under dose) or $>$48 IU/Kg (overdose) or wrong frequency or missing frequency or missing patient's weight.

PAQC[a]:- Paediatric Admission Quality of Care, S&S[b]:- sign & symptoms, AVPU[c]: A for Alert, V for Verbal response, P for pain, U for unresponsive

3.2.2 Covariates

The predictor variables of interest included an interaction between the trial arm and follow up time (in months), hospital level covariates (i.e., malaria prevalence status and paediatric admission workload), and clinician level covariates (i.e., gender and cadre). At patient level, we considered gender, age categorized into 2-11 months and 12-59 months respectively and the number of comorbid illnesses. Although WHO pneumonia guidelines apply for children aged 2 to 50 months (Organization, 2013), we categorized patients in two age groups because older children have better clinical outcomes compared to infants (Lopez, 2014).

To determine the number of comorbidities, we considered common clinical diagnoses documented in patient's medical records besides pneumonia. This included malaria, malnutrition, HIV, Asthma, Tuberculosis (TB), rickets, anaemia, diarrhoea and dehydration. For each diagnosis, we created binary variables with one denoting the presence of a disease and zero denoting absence of a disease. Thereafter, we summed the binary indicators and categorized patients into those with 0, 1, 2, 3 or more comorbidities. Clinically, 46.8% (995/2127) of the patients had no comorbidities, 29.8% (633/2127) had one comorbidity, 17.9% (381/2127) had two comorbidities, and 5.5% (118/2127) had at least three comorbidities.

3.2.3 Investigating Missing Data Mechanism Underlying Pneumonia Trial Data

To explore plausible missing data mechanism underlying pneumonia trial data set, we created binary missingness indicators for each partially observed variables, that is, patient's gender, clinician's cadre and gender respectively. We

created binary indicators such that $R_{pijl} = 1$ if the p^{th} variable is observed and $R_{pijl} = 0$ if the p^{th} variable is missing for the i^{th} patient admitted by j^{th} clinician in hospital l. We regressed the binary indicators separately on fully observed variables in pneumonia data set using multivariable logistic regression model below

$$logit(P[R_i = 1|\boldsymbol{X}_i]) = \boldsymbol{X}_i\boldsymbol{\beta}, \tag{3.1}$$

The predictor variables of interest $(\boldsymbol{X}_i)$ included: fully observed PACQ score, an interaction between intervention arm and follow up time in months, number of comorbid illnesses, age of the patient, hospital malaria prevalence and paediatric admission workload. The vector $\boldsymbol{\beta}$ denotes fixed regression parameters to be estimated. We also used graphical methods to explore missing data patterns underlying pneumonia trial data.

3.2.4 Multilevel Multiple Imputation of Pneumonia Trial Data

To handle missingness in the trial data set, we imputed missing covariate components assuming a MAR mechanism. MI was conducted within the latent normal joint model imputation framework using *jomo* (Quartagno et al., 2019) and *mitml* (Grund et al., 2019) packages in R (version 3.5.4). Considering the i^{th} pneumonia patient attended by clinician j in hospital l, our multilevel level joint imputation model corresponded to

$$\begin{aligned} Y_{ijl}^{(1)} &= \boldsymbol{X}_{ijl}^{(1)}\boldsymbol{\beta}^{(1)} + \boldsymbol{b}_{jl}^{(1)} + e_{ijl}^{(1)} \\ \boldsymbol{Y}_{jl}^{(2)} &= \boldsymbol{X}_{jl}^{(2)}\boldsymbol{\beta}^{(2)} + \boldsymbol{b}_{jl}^{(2)}, \end{aligned} \tag{3.2}$$

$$e_{ijl}^{(1)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \ \ and \ \ \left(\boldsymbol{b}_{jl}^{(1)}, \boldsymbol{b}_{jl}^{(2)}\right) \sim N(\mathbf{0}, \boldsymbol{\Omega}_b),$$

where $Y_{ijl}^{(1)}$ and $\boldsymbol{Y}_{jl}^{(2)}$ denote partially observed level 1 variables (patient's gender) and level 2 variables (clinician's gender and cadre) respectively. Level 1 predictors $(\boldsymbol{X}_{ijl}^{(1)})$ included fully observed covariates (i.e., an interaction term between follow-up time and trial arm, hospital workload and malaria prevalence status, patient's age and number of comorbid illnesses). Besides, we included pneumonia PAQC score (outcome) as a predictor in the first level of the imputation model. On the other hand, level 2 predictors $\boldsymbol{X}_{jl}^{(2)}$ included an interaction term between follow-up time and intervention arm, hospital admission workload and hospital malaria prevalence status.

A random intercept $(\boldsymbol{b}_{jl})$ 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. Trace plots, auto-correlation functions, and the Gelman and Rubin diagnostic tests were used to assess convergence (Gelman et al., 1992). An example trace plot in Appendix Figure A.2 indicated satisfactory convergence. Final estimates were pooled according Rubin's rules (Rubin, 1976).

3.2.5 Cumulative-logit Models for Ordinal Responses

When the outcomes of interest is ordinal with S levels, $s = 1, 2, \ldots, S$, the associated probabilities correspond to $\pi_1 + \pi_2 + \cdots + \pi_S$. Furthermore, the outcome can be expressed in terms of $S - 1$ cumulative logits (Agresti, 2002, p. 275). The cumulative probability of the response for subject i being in category s or below

is given by

$$P(Y_i \leq s) = \pi_1 + \pi_2 + \cdots + \pi_S, \tag{3.3}$$

while the cumulative logit describing the log-odds of two cumulative probabilities is defined by

$$\log\left(\frac{P(Y_i \leq s)}{P(Y_i > s)}\right) = \log\left(\frac{P(Y_i \leq s)}{1 - P(Y_i \leq s)}\right) = \log\left(\frac{\pi_1 + \pi_2 + ... + \pi_s}{\pi_{s+1} + \pi_{s+2} + ... + \pi_S}\right). \tag{3.4}$$

That is, the probability that a response is in category s or below versus the probability that a response is in a category higher than s. Thus, the corresponding sequence of cumulative logits is

$$\begin{aligned} L_1 &= \log\left(\frac{\pi_1}{\pi_2 + \pi_3 + \cdots + \pi_s}\right) \\ L_2 &= \log\left(\frac{\pi_1 + \pi_2}{\pi_3 + \pi_4 + \cdots + \pi_s}\right) \\ &\vdots \\ L_{S-1} &= \log\left(\frac{\pi_1 + \pi_2 + \cdots + \pi_{S-1}}{\pi_S}\right). \end{aligned} \tag{3.5}$$

When the ordered outcome is regressed on a set of fixed effects, a proportional-odds cumulative logit model defined by

$$logit[P(Y_i \leq s)] = \alpha_s + X_i\beta, \tag{3.6}$$

is commonly used (Agresti, 2002, Chapter 7. 326). When the proportional odds assumptions are upheld (i.e., parallel logits), the slope for each variable stays the same across different cumulative logits. That is, the β regression coefficient of each covariate is assumed identical across all $S-1$ logit equations while the intercepts α_s can differ. The intercepts describe the log-odds of being in category s or below when all the fixed effects are fixed to zero (i.e., $X_1 = X_2 = \cdots = X_p = 0$) for continuous variables or held at reference levels for categorical variables (Agresti, 2002). For ordinal responses in multilevel data contexts, model families that account for clustering can be used as appropriate. In this study, we used both proportional odds random-effects and proportional odds generalized estimating equation (GEE) model introduced in Section 2.2. Specifically, letting i index patient, j clinician and l hospital, the proportional odds random intercepts model implemented in R's *Ordinal* package (Christensen, 2015) corresponded to

$$\begin{aligned} &logit[P(Y_{PAQC\ score;ijl} \leq s)] = \alpha_s + \beta_1 X_{age\ group;ijl} + \beta_2 X_{patient\ gender;ijl} + \\ &\beta_3 X_{comorbidity=0;ijl} + \beta_4 X_{comorbidity=1;ijl} + \beta_5 X_{comorbidity=2;ijl} + \\ &\beta_6 X_{clinician\ cadre;jl} + \beta_7 X_{clinician\ gender;jl} + \beta_8 X_{admission\ workload;l} + \\ &\beta_9 X_{malaria\ prevalence;l} + \beta_{10} X_{time\ in\ months;l} + \beta_{11} X_{trial\ arm;l} + \\ &beta_{12} X_{time\ in\ months;l} * X_{trial\ arm;l} + b_{jl}, \end{aligned} \tag{3.7}$$

where $\alpha_s, s = 1,2,3,4,5,6$ are PAQC score specific intercepts, β are estimated regression coefficients and b_{jl} are clinician's random intercept. PAQC score $= 0$ was considered as a reference category. Hospital random-effects were not considered because the number of hospitals (n=12) was low to consider random-effects at

that level. This was in addition to ensuring compatibility with the two-level MI model.

Similarly, letting i index patient, j clinician and l hospital, the proportional odds GEE model of interest implemented in R's *Multgee* package (Touloumis, 2014) corresponded to

$$\begin{aligned} &logit[P(Y_{PAQC\ score;ijl} \leq s)] = \alpha_s + \beta_1 X_{age\ group;ijl} + \beta_2 X_{patient\ gender;ijl} + \\ &\beta_3 X_{comorbidity=0;ijl} + \beta_4 X_{comorbidity=1;ijl} + \beta_5 X_{comorbidity=2;ijl} + \\ &\beta_6 X_{clinician\ cadre;jl} + \beta_7 X_{clinician\ gender;jl} + \beta_8 X_{admission\ workload;l} + \\ &\beta_9 X_{malaria\ prevalence;l} + \beta_{10} X_{time\ in\ months;l} + \beta_{11} X_{trial\ arm;l} + \\ &beta_{12} X_{time\ in\ months;l} * X_{trial\ arm;l}, \end{aligned} \tag{3.8}$$

where α_s, $s = 1, 2, 3, 4, 5, 6$ are PAQC score intercepts. We adopted an exchangeable working correlation. In this thesis, both model families were used to analyse PAQC score with an aim of assessing stability of parameter estimates within the models before and after MI of missing covariates. Under complete case analysis, records with missing covariates were discarded. Before analyses (imputed data sets and complete case records), clinicians were grouped into two cadres from the initial four cadres, that is, clinical officers (combining clinical officers and clinical officer interns) and medical officers (combining medical officers and medical officer interns). Re-grouping was due to the small number of clinical officer and medical officers (Table 1.1).

3.2.6 Hypotheses Testing for Regression Coefficients Associated with Ffixed Effects

To determine covariates with statistically significant effect on pneumonia PAQC score, Wald tests and likelihood-ratio tests were used. The tests entailed comparing full (saturated) model containing all the covariates and a reduced (null) models which dropped one covariate at a time. Supposing that the full model had p estimated regression parameters, removing one of the fixed effects with a regression coefficient β_p, $i = 1, 2, \ldots, p$ resulted to a reduced model with fewer parameters than the saturated model, say $q = 1, 2 \ldots, p-1$. In this case, the null and the alternative hypotheses correspond to

$$H_0 : \boldsymbol{\beta}_p = 0 \tag{3.9}$$
$$H_1 : \boldsymbol{\beta}_p \neq 0.$$

The likelihood-ratio test was used to test for statistical significance of covariates in the random-effects models while the Wald tests was used for the GEE model. The tests were conducted on complete case records and after MI as outlined below.

Likelihood Ratio Test

To perform likelihood-ratio test (LRT) under complete case analysis, we obtained the log-likelihoods for the saturated model ($l(\hat{\beta}_{saturated})$) and the reduced model ($l(\hat{\beta}_{reduced})$) respectively. We then calculated the test statistic as follows

$$\boldsymbol{G}^2 = 2(l(\hat{\boldsymbol{\beta}}_{saturated}) - l(\hat{\boldsymbol{\beta}}_{reduced})). \tag{3.10}$$

The LRT statistic follows a chi-squared distribution with degrees of freedom being the difference in the number of fixed-effects parameters between the saturated and the reduced model (i.e., $p - q$). If the LRT statistic is greater than a critical value, the null hypothesis is rejected conclude the alternative hypothesis. However, if the LRT statistic is less than the critical value, we fail to reject the null. In this case the covariate of interest is not statistically significant.

Extending LRT to imputed data sets, first we fitted the saturated model to each imputed data set ($m = 1, 2 \ldots M$) and obtained the corresponding log-likelihood functions. Likewise, we fitted the reduced model to each imputed data set (removing one covariate at a time) to obtain the log-likelihood functions. We then calculated

$$\boldsymbol{G}_m^2 = 2(l(\hat{\boldsymbol{\beta}}_{m,saturated}) - l(\hat{\boldsymbol{\beta}}_{m,reduced})) \quad for \quad m = 1, \ldots, M$$

for each imputed data set before computing the average of the likelihood ratios across the M imputed data sets as follows

$$\boldsymbol{G}^* = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{G}_m^2. \tag{3.11}$$

We also obtained the averages of regression parameters estimated in the saturated and reduced models using $\bar{\boldsymbol{\beta}}_{saturated} = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\beta}}_{m,saturated}$ and $\bar{\boldsymbol{\beta}}_{reduced} = \frac{1}{M} \sum_{m=1}^{M} \hat{\boldsymbol{\beta}}_{m,reduced}$ respectively. The saturated and reduced models were then re-estimated with model parameters fixed to $\bar{\boldsymbol{\beta}}_{saturated}$ and $\bar{\boldsymbol{\beta}}_{reduced}$ respectively. For

each imputed data set, we obtained

$$\bar{G}_m^2 = 2(l_m(\bar{\boldsymbol{\beta}}_{saturated}) - l_m(\bar{\boldsymbol{\beta}}_{reduced})) \quad for \quad m = 1, \ldots, M,$$

and thereafter calculated the average of log-likelihood functions across imputed data sets using

$$\bar{G} = \frac{1}{M} \sum_{m=1}^{M} \bar{G}_m^2.$$

The LRT statistic under MI is defined by

$$F_{LR} = \frac{\bar{G}}{q(1+r)}, \tag{3.12}$$

where

$$r = \frac{M+1}{q(M-1)}(\boldsymbol{G}^* - \bar{\boldsymbol{G}}),$$

estimates the average relative increase in variance due to missingness (Van Buuren, 2018; Carpenter and Kenward, 2013; Meng and Rubin, 1992). The LRT statistics F_{LT} is compared to a reference F distribution with q and v_l degrees of freedom where

$$v_l = \begin{cases} 4 + (t-4)\left[1 + \frac{(1-2t^{-1})}{r}\right]^2 & \text{for} \quad t = q(M-1) \; > \; 4 \\ t(1+1/q)(1+1/r)^2/2 & \text{otherwise.} \end{cases}$$

The $p-$value for F_{LT} is given by

$$P_l = Pr[F_{q,v_l} > F_{LT}]. \tag{3.13}$$

Wald Test

Under complete case analysis the Wald test statistics is defined by

$$W = \frac{(\hat{\beta} - \beta_0)^2}{Var(\hat{\beta})}, \tag{3.14}$$

which follows a chi-squared distribution with $p - q$ degrees of freedom. If the Wald test statistic is greater than a critical value, the null hypothesis is rejected and conclude the alternative. However, if the statistic is less than the critical value, we fail to reject the null. In this case the covariate of interest is not statistically significant. Extending to imputed data sets, the Wald test statistic is defined by

$$WT = \frac{(\hat{\boldsymbol{\beta}}_{MI} - \boldsymbol{\beta})^T \hat{\boldsymbol{V}}_{MI}^{-1} (\hat{\boldsymbol{\beta}}_{MI} - \boldsymbol{\beta})}{q(1 + r^*)}, \tag{3.15}$$

where $\hat{\boldsymbol{V}}_{MI} = \hat{\boldsymbol{W}} + (1 + 1/M)\hat{\boldsymbol{B}}$ is the estimate of the total variance and $r^* = \frac{1}{q}(1 + \frac{1}{M})tr(\boldsymbol{B}\boldsymbol{W}^{-1})$ is the average fraction of missing information (Carpenter and Kenward, 2013; Van Buuren, 2018).

Components $\boldsymbol{B}$ and $\boldsymbol{W}$ denote the between and within imputation variances defined in Section 2.3.3. The corresponding p-value for the test statistic WT is characterized by

$$P_w = Pr[F_{q,v_w} > WT],$$

where F_{q,v_w} is an F distribution with q and v_w degrees of freedom with

$$v_w = \begin{cases} 4 + (t-4)\left[1 + \frac{(1-2t^{-1})}{r^*}\right]^2 & \text{if} \quad t = q(M-1) \; > \; 4 \\ t(1+1/q)(1+1/r^*)^2/2 & \text{otherwise.} \end{cases}$$

More details on multi-parameter hypothesis tests after MI using Wald tests and likelihood-ratio tests are available in (Carpenter and Kenward, 2013, p. 52-54) and (Van Buuren, 2018, 157-158). All analyses were conducted in R version 3.5.4. A 5% level of significance was considered under complete case analysis and after MI of missing covariates.

3.3 Results

Examining pneumonia PAQC score over time graphically, hospitals in the standard A&F (control) arm (dashed red curve) exhibited a higher mean PAQC score at baseline with no significant fluctuations over time (Figure 3.1). On the other hand, hospitals assigned to enhanced A&F (intervention) arm (solid blue curve) had a lower mean PAQC score at baseline which rapidly improved towards higher score in the first 6 months of follow-up. Although enhanced A&F arm's trend line surpassed that of standard A&F arm after six months of follow-up, the 95% confidence bands of the two trial arms overlapped substantially (Figure 3.1).

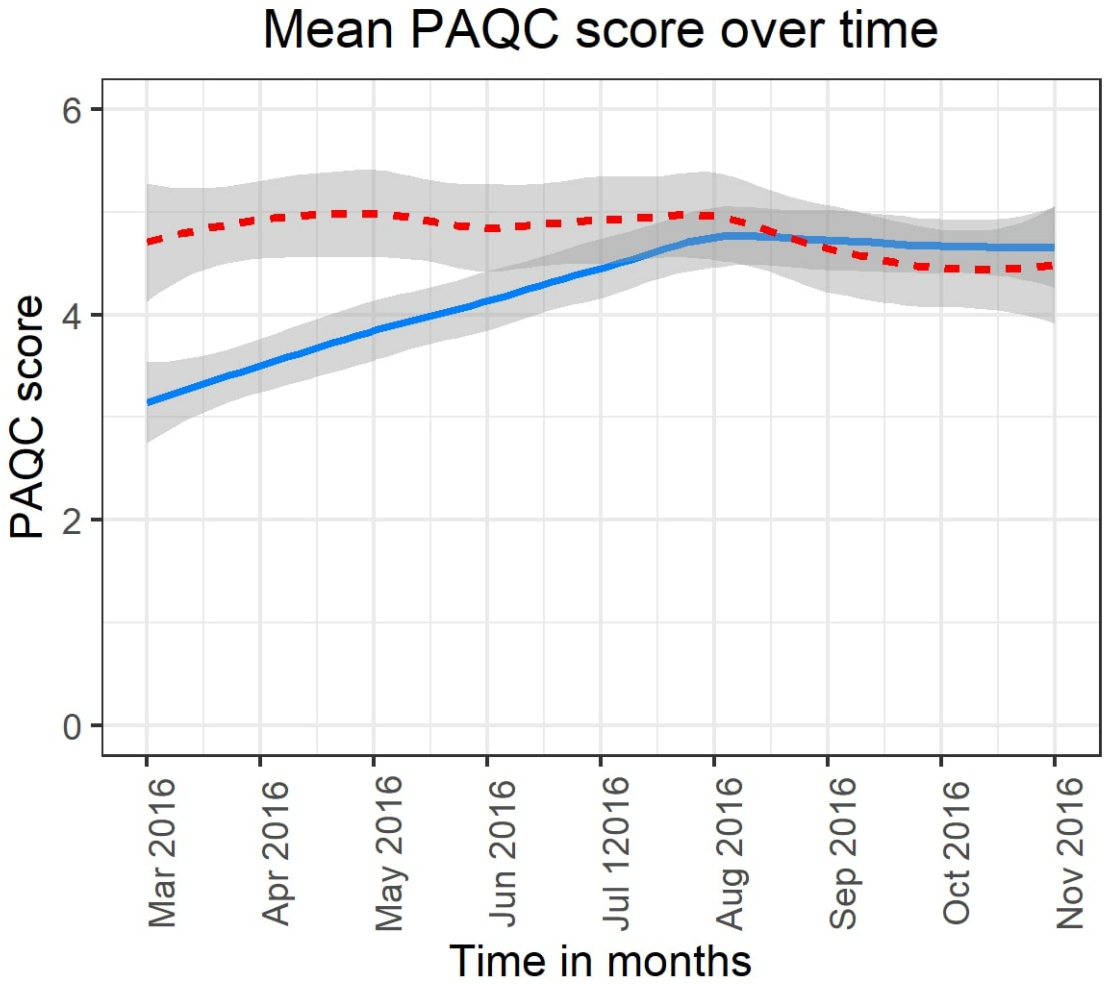

Figure 3.1: Loess curves for mean PAQC score over time (dashed curve represent the average for 6 hospitals in the standard A&F arm and solid curve represent the mean for 6 hospitals in the enhanced A&F arm) and corresponding 95% confidence bands.

An assessment of missing data patterns suggested a multivariate missing data pattern (Appendix Figure A.1). The missing data pattern were similar between clinician's cadre and gender. That is, nearly all clinicians with missing gender had missing cadre as well. Further investigations into missing data patterns showed that missing clinicians' cadre and gender only occurred in 6 out of 12 hospitals (Figure 3.2).

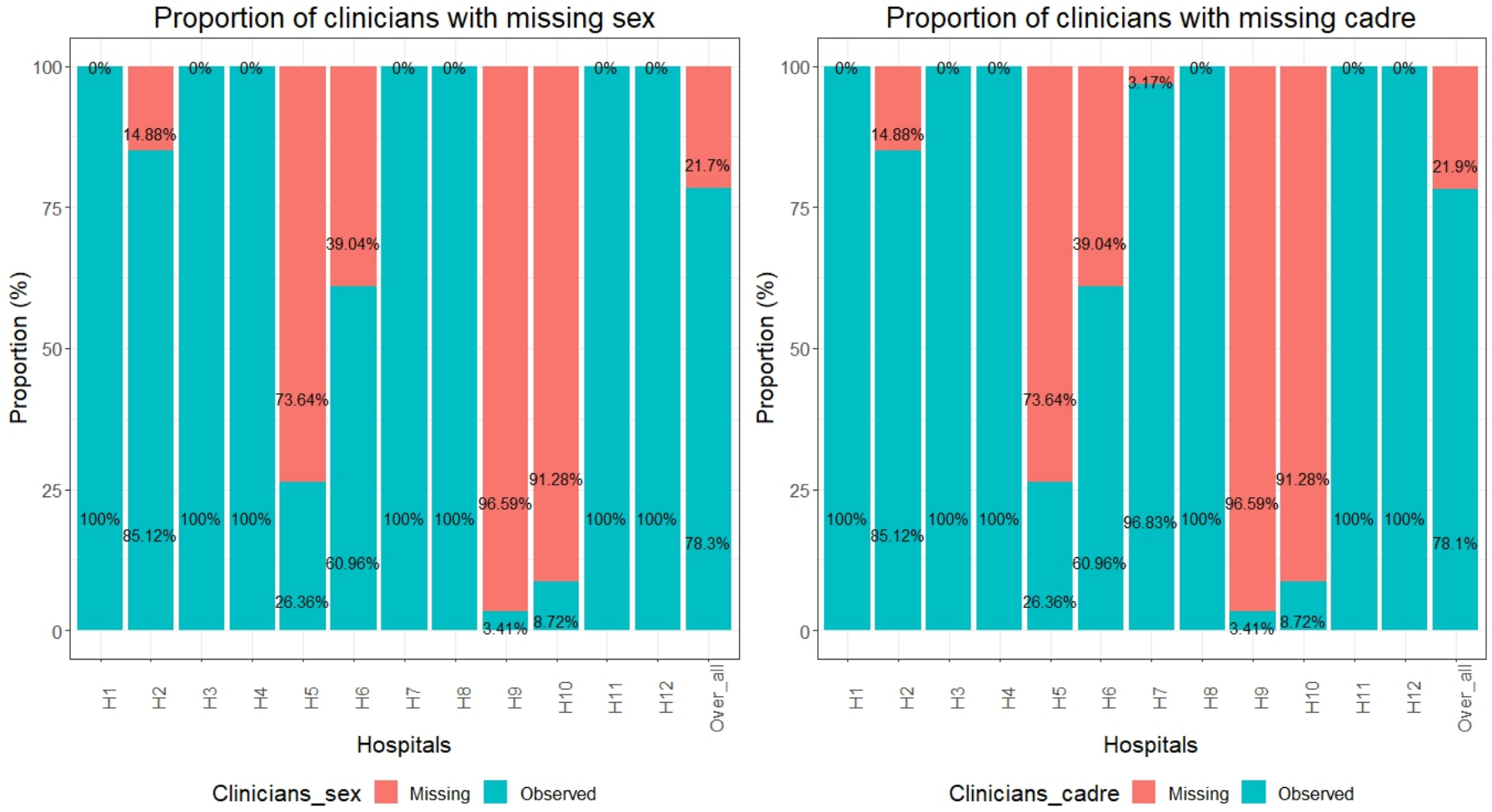

Figure 3.2: Proportion of missing clinicians' cadre and gender at hospital level and across all hospitals combined.

Logistic regression results on plausible mechanisms underlying pneumonia trial data indicated that the probability of missing patient's gender was neither dependent on the outcome (PAQC score) nor fully observed covariates (interaction between intervention arm and follow up time in months, hospital admission workload and malaria prevalence, patient's age group and the number of presenting comorbid illnesses). On the other hand, the probabilities of missing clinician's cadre and gender were dependent on both the outcome and fully observed covariates suggesting evidence against MCAR (Appendix Table A.1).

Random-effects and GEE Model Results

Test for proportional odds assumption was not statistically significant at 5% level (P-value =0.17). Therefore, we assumed parallel logits and fitted proportional odds models to complete case records and imputed data sets. Table 3.2 presents the likelihood ratio test and Wald test results for proportional odds random-effects and GEE model respectively. After MI of missing covariates, we observed consistent results between the random-effects model and the GEE model in terms of statistical significance of covariates of interest (Table 3.2). Specifically, we found statistically significant interaction effect between intervention arm and follow-up time. Similarly, admission workload at hospital level was significant at 5% level. At patient's level, age and the number of comorbidities were statistically significant while at clinician's level, gender showed significant effect on pneumonia PAQC score (Table 3.2).

Table 3.2: Likelihood ratio test and Wald test results for random-effects model and GEE model under complete case analysis and after multilevel multiple imputation of missing covariates

	Random-effects model				Marginal model: GEE[a]			
	CCA[b]		Multilevel MI[c]		CCA		Multilevel MI	
	N=1619 (76.1%)		N=2127 (100%)		N=1619 (76.1%)		N=2127 (100%)	
Effect	LRT[d]	P value	LRT	P value	Wald test	P value	Wald test	P value
Patient's age	3.49	0.06	4.66	0.03	4.18	0.04	7.81	0.01
Patient's gender	0.08	0.77	0.01	0.92	0.003	0.96	0.02	0.88
Comorbidities	4.46	0.02	4.83	0.03	2.42	0.49	5.48	0.02
Clinician's gender	5.06	0.02	4.02	0.04	6.32	0.01	4.47	0.03
Clinician's cadre	0.01	0.91	0.23	0.63	1.36	0.24	2.96	0.08
Hospital workload	0.14	0.71	3.39	0.04	1.46	0.23	4.95	0.03
Malaria prevalence	0.07	0.79	1.35	0.25	0.98	0.32	0.01	0.91
Time (months)	11.98	< 0.001	14.16	< 0.001	11.37	0.003	11.16	< 0.001
Enhanced A&F [e] arm	28.58	$<$0.001	17.51	< 0.001	28.86	< 0.001	17.76	$<$0.001
Time $\times$ Enhanced A&F	14.92	0.02	14.16	< 0.001	17.85	< 0.001	9.45	$<$0.001

GEE[a]: -Generalized estimating equations, CCA[b]:-Complete case analysis, MI[c]: -Multiple imputation, LRT[d]: - Likelihood ratio test, A&F[e]: -Audit and feedback

Table 3.3: Parameter estimates (standard errors) from random-effects and marginal GEE models: complete case analysis and after multiple imputation.

	Random-effects model		Marginal model: GEE[a]	
	Complete case analysis N=1619 (76.1%)	Multilevel MI[b] N=2127 (100%)	Complete cases analysis N=1619 (76.1%)	Multilevel MI N=2127 (100%)
Effect	Odds Ratios (95% CI[c])	Odds Ratios (95% CI)	Odds Ratios (95% CI)	Odds Ratios (95% CI)
PAQC[d] score intercept 0	Reference	Reference	Reference	Reference
PAQC score intercept 1	0.06 (0.031,0.140)	0.18 (0.071, 0.375)	6.13 (3.188, 9.456)	4.18 (2.010,6.128)
PAQC score intercept 2	0.07 (0.036, 0.138)	0.17 (0.082, 0.365)	7.73 (3.258, 12.31)	4.98 (2.056, 2.078)
PAQC score intercept 3	0.22 (0.110, 0.420)	0.53 (0.251, 1.105)	3.12 (1.345, 7.23)	2.02 (0.852, 4.809)
PAQC score intercept 4	0.67 (0.342, 1.294)	1.63 (0.779, 3.427)	1.29 (0.561, 2.981)	0.84 (0.354, 1.987)
PAQC score intercept 5	2.74 (1.401, 5.347)	6.69 (3.166, 14.14)	0.44 (0.192, 1.012)	0.29 (0.122, 0.678)
PAQC score intercept 6	7.24 (3.678, 4.253)	7.79 (8.336, 3.964)	0.21 (0.089, 0.501)	0.14 (0.057, 0.336)
Age-group:12-59	1.20 (0.991, 1.464)	1.19 (0.986, 1.454)	1.15 (0.922, 1.432)	1.16 (0.932, 1.454)
Patient's gender: males	0.97 (0.806, 1.174)	0.97 (0.805, 1.173)	0.95 (0.759, 1.185)	0.95 (0.760, 1.183)
Comorbidities: 1	0.99 (0.783,1.267)	0.99 (0.782,1.253)	1.02 (0.810,1.295)	1.03 (0.815,1.304)
Comorbidities: 2	1.01 (0.766,1.327)	1.01 (0.767,1.326)	1.01 (0.779,1.304)	1.01 (0.781,1.312)
Comorbidities: $\geq$3	0.63 (0.398,0.985)	0.61 (0.387,0.955)	1.37 (0.906,2.063)	1.41 (0.937,2.126)
Clinician's gender: female	1.51 (1.057, 2.183)	1.53 (1.064, 2.195)	1.44 (1.095, 1.910)	1.45 (1.106, 1.894)
Clinician's cadre: MO[e]	1.02 (0.709, 1.468)	1.04 (0.720, 1.490)	1.18 (0.878, 1.582)	1.20 (0.888, 1.611)
Hospital workload: low	0.93 (0.624, 1.376)	1.12 (1.080, 1.372)	1.42 (0.974, 2.068)	1.40 (1.103, 2.063)
Malaria prevalence: low	0.95 (0.644, 1.401)	0.94 (0.640, 1.389)	1.18 (0.748, 1.865)	1.18 (0.742, 1.87)
Time (months)	1.05 (0.969, 1.145)	1.05 (0.967, 1.141)	0.99 (0.904, 1.094)	0.99 (0.905, 1.103)
Enhanced A&F[f] arm	0.18 (0.095, 0.349)	0.18 (0.093, 0.341)	0.11 (0.054, 0.227)	0.11 (0.053, 0.236)
Time $\times$ Enhanced A&F	1.15 (1.018, 1.307)	1.16 (1.020, 1.308)	1.27 (1.125, 1.484)	1.29 (1.117,1.482)
Variance between random clinician's intercepts	1.328 (1.151)		1.161 (1.073)	

GEE[a]: -Generalized estimating equations, MI[b]: -Multiple imputation, CI[c]:-Confidence Intervals,PAQC[d]:-Paediatric admission quality of care, *MO*[e] : Medical officer, A&F[f]: -Audit and feedback

In Table 3.3, we present proportional odds ratios and the corresponding 95% confidence interval obtained before and after multilevel MI. For the GEE model, we reported robust (empirically corrected) standard errors which were in agreement with model based (naive) standard errors (Appendix Table A.2). Under complete case analysis, only 1619/2127 (76.1%) case records were considered. This loss information led to larger standard errors in comparison to those obtained after MI of missing covariates. These observations were made in both random-effects and GEE model families. Furthermore, the proportional odds ratios were consistently smaller under complete case analyses compared to those obtained after MI (Table 3.3). These results were an indication of bias and inefficiency of parameters estimated under complete case analysis. PAQC score intercepts presented in Table 3.3 denote thresholds (cut points) differentiating adjacent levels of the response variable. For example, intercept 1 in Table 3.3 denote the odds of PAQC score $= 1$ vs PAQC score ≥ 2 for a female patient in age-group 2 to 11 months, with no comorbidities, and admitted by a male medical officer in a high workload hospital located in the high malaria prevalence region. The individual fixed effect parameters are the proportional odds ratios of individual variables on PAQC score holding all other variables in the model constant.

From study results, enhanced audit and feedback led to improved uptake of new pneumonia paediatric guideline over time. For instance, considering a patient admitted in an intervention hospital (enhanced audit and feedback arm), the odds of PAQC score=1 versus PAQC score ≥ 2 were 1.16 (95% CI: 1.02-1.308) times higher the odds of a patients admitted in a control hospital, for a unit increase in follow-

up time and holding other variables constant at reference levels. Likewise, for a patient admitted in an intervention hospital, the odds of PAQC score=1 versus PAQC score $\geq$ 2 were 1.29 (95% CI: 1.17-1.482) times higher than the odds of a patients admitted in a control hospital, for a unit increase in follow-up month (GEE model after MI). These interpretations hold for all other response (PAQC score) levels. The study results also exhibited shifts in statistical significance before and after MI for selected variable. Specifically, adjusting for other variables, complete cases analysis led to insignificant difference between low and high admission workload hospitals on levels of PAQC score in both random-effects model and GEE model. However, after MI, the odds of higher pneumonia PAQC score in low workload hospitals were 1.12 (95% CI: 1.08-1.372) and 1.40 (95% CI: 1.103-2.063) times higher than for high workload hospitals for the random intercepts and GEE model respectively (Table 3.3).

Regarding random-effects model, the variance component between clinicians and the corresponding standard error were inflated under complete cases analysis.

3.4 Discussion

In this chapter, we sought to investigate the effect of enhanced A&F on routine paediatric pneumonia care in 12 Kenyan hospitals during a cluster randomized trial. Among covariates, about 22% of clinicians had missing gender and cadre respectively. In contrast, patient level variables were fully observed except patient's gender which had less than 1% missingness. The sharp contrast in the level of missingness could be due the fact that continued CIN audit and feedback

reports focus on the documentation of patient level variables rather than documentation of clinicians' characteristics. Through preliminary investigations, we established that missing clinicians' characteristics occurred in 6 out of 12 hospitals participating in the trial. The patterns of missingness in the two clinicians level variables was highly correlated. That is, clinicians who did not document their gender were also likely not to document their cadre and vice versa.

To alleviate bias and inefficiency, we used MI within the joint modelling (JM) imputation framework assuming a MAR mechanism (Quartagno and Carpenter, 2018; Grund et al., 2017). Although JM imputation framework does not address the full range of complexities that are typical of multilevel data (Enders et al., 2016; Grund et al., 2018; Quartagno et al., 2019), it was preferred due to its flexibility coupled with recent statistical software developments in handling categorical variables with more than two levels in second and higher levels of hierarchy (Quartagno, 2016). This ensured compatibility between imputation and analysis models of interest thus minimizing bias in parameter estimates (Grund et al., 2018).

From study results multilevel MI led to more precise parameter estimates compared to complete case analyses in both random-effects and GEE models. Adjusting for patients, clinicians and hospital level factors, enhanced A&F improved uptake and adherence to recommended paediatric pneumonia guidelines over time among children aged 2 to 59 months admitted in 6 CIN hospitals during the trial period compared to standard A&F on general inpatient paediatric care. The significant difference in the uptake of pneumonia guidelines between the in-

tervention arms could be due to difference in baseline performance observed in the LOESS curves. Control hospitals exhibited high baseline performance (on average) thus leaving smaller room for improvement compared to low baseline performance in the enhanced A&F arm hence larger room for improvement over time.

From results, the quality of pneumonia care differed between male and female clinicians. It was also evident that junior clinicians (medical officers and clinical officer interns) were responsible for much care during the trial period. However, the quality of care provided did not differ between the cadres. The high number of interns was due to the fact that majority of the study sites were teaching and referral hospitals.

Strengths and Implications of the Study

We evaluated missing data patterns underlying the trial data set. This was useful in revealing trends and gaps in the quality of routine care. Insight into such information is useful when designing cost effective follow-up or new interventions programmes for optimal and efficient utilization of already stretched resources (Bitton et al., 2017). For instance, based on this study results, a follow up intervention programme aimed at improving documentation and reporting of clinician characteristics, should be directed to specific hospitals with low documentation of clinicians' level variables, while directing resources in hospitals with good documentation practices elsewhere.

Our choice of proportion odds models to analyse the ordinal outcome, was ascertained through formal test, further enhancing the validity of these study results. In instances when the proportional odds assumptions are violated, multinomial logistic regression model is recommended (Molenberghs et al., 2014, Chapter 27, p. 493)agresti2002.

In contrast to previous studies reporting quality of inpatient paediatric routine care (Gachau et al., 2017; Thomas et al., 2017; Agweyu et al., 2018a), this study accounted for clinicians who are essential for the delivery of health intervention (Rowe et al., 2005). Ignoring variation at clinician level may lead to biased estimates, overestimation or underestimation of variations in other levels of clustering (Cook et al., 2018).

A limitation of this study is that we relied on data collected after patient discharge. Therefore, we are unable to ascertain if patients received pneumonia care as documented by health workers (Ayieko et al., 2019). In conclusion, adjusting for hospitals, admitting clinicians and patient level factors, enhanced audit and feedback improved uptake of WHO recommended paediatric pneumonia guidelines compared to standard audit and feedback. Additionally, female clinicians and hospitals with low admission workload were associated with higher uptake of the new paediatric pneumonia guidelines during the trial period. In both random-effects and marginal model, parameter estimates were biased and inefficient under complete case analysis. Therefore, MI is recommended.

Chapter 4

Handling Missing data in a Composite Outcome with Partially Observed Subcomponents

4.1 Introduction

Composite measures combine information from multiple measures into a single summary score (Caldis, 2007; Chen et al., 2013; Profit et al., 2014; Shwartz et al., 2015; Ibrahim et al., 2016; Cordoba et al., 2010). In health care settings, composite measures are used as scorecards to measure and benchmark performance and quality of care in neonates (Profit et al., 2010) and cardiovascular care among adults, (Caldis, 2007; Eapen et al., 2011; Chen et al., 2013; EUnetHTA, 2013). Profit et al. (2010) presented a conceptual framework on composite indicator development in paediatrics care. More recently, Opondo et al. (2016) developed and val-

idated the PAQC score; a 7-point composite score aimed at benchmarking processes of care among children admitted with common childhood illnesses in low income settings. In the validation study, PAQC score was shown to be a good proxy for outcome of care (Opondo et al., 2018).

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 (Profit et al., 2014; Shwartz et al., 2015; EUnetHTA, 2013). In addition, the issue of multiple testing is avoided (Proschan and Waclawiw, 2000; Freemantle et al., 2003). Although composite outcomes complement single measures, weak theoretical and statistical assumptions may undermine the overall reliability (Caldis, 2007). 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 (Caldis, 2007; Profit et al., 2014; Ibrahim et al., 2016; EUnetHTA, 2013; Commission et al., 2008). In the literature, MI, proposed by Rubin 1976, offers a good, often best practice, solution in dealing with partially observed outcomes and covariates (Molenberghs and Verbeke, 2005; van Buuren and Groothuis-Oudshoorn, 2011; Carpenter and Kenward, 2013; Enders et al., 2016). 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 (Grund et al., 2017). 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

(Ibrahim et al., 2016; Simons et al., 2015).

In chapter 3, partially observed subcomponents of a composite outcome (PAQC score) were scored with value 0 representing suboptimal care (Opondo et al., 2016). This approach of dealing with missing PAQC score subcomponents 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 underestimated standard errors (Carpenter and Kenward, 2013; Plumpton et al., 2016). Using routine paediatric pneumonia data from Kenyan hospitals, we explored appropriate strategies of dealing with missing data in the PAQC score subcomponents. Through a range of simulation scenarios, that is, three missing data rates under two missing data mechanisms, we assessed 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 subcomponents across the simulation scenarios was obtained and compared between MI and the conventional method.

The remainder of this chapter is structured as follows. Section 4.2 presents multilevel MI methods used to handle missing covariates and PAQC score subcomponents. This is in addition to a simulation scheme as used to assess performance of MI and the conventional approach in handling missing PAQC score subcomponents. Results are presented in Section 4.3 and we conclude with a discussion in Section 4.4.

4.2 Methods

4.2.1 Data

We analyzed pneumonia trial data described in Chapter 1, Section 1.2. The outcome of interest was pneumonia PAQC score introduced in Chapter 3, Section 3.2.1. Missing data occurred in covariates across two levels of the hierarchical structure (Table 1.1) as well as PAQC score subcomponents in the assessment and treatment domains of care (Table 1.2). In Chapter 3, PAQC score was constructed using 12 pneumonia care indicators using the conventional approach presented in Table 3.1. The 12 indicators comprised of nine signs and symptoms in the assessment domain, one indicator in the diagnosis and classification domain and two indicators in the treatment domain). Specifically, six binary indicators were created with one representing adherence to recommended childhood pneumonia guidelines and zero representing inappropriate care. Under the conventional approach, variation on the 7-point scale was due to missing data and/or inappropriate care across the three domains of care. In this study, inappropriate care refers 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 (Opondo et al., 2016). However, considering the controlled study inclusion criteria (i.e., inclusion of patients with pneumonia signs and symptoms), undocumented signs and symptoms in the assessment domain were regarded as inappropriate care. Therefore, we restricted our focus on missing data in treat-

ment domain subcomponents, namely amoxicillin dose prescribed, frequency of administration and weight of patient (Table 1.2). Appendix Figure B.1 provides a graphical representation of missing data pattern underlying pneumonia trial data.

4.2.2 Multilevel Multiple Imputation of Missing Covariates and PAQC Score subcomponents

We imputed incomplete variables of interest using a two-level joint imputation model corresponding to,

$$\begin{aligned} Y_{ijl}^{(1)} &= \boldsymbol{X}_{ijl}^{(1)}\beta^{(1)} + b_{jl}^{(1)} + e_{ijl}^{(1)} \\ \boldsymbol{Y}_{jl}^{(2)} &= \boldsymbol{X}_{jl}^{(2)}\beta^{(2)} + b_{jl}^{(2)}, \end{aligned} \tag{4.1}$$

$$e_{ijl} \sim N(0, \boldsymbol{\Sigma}) \ \ and \ \ \left(b_{jl}^{(1)}, b_{jl}^{(2)}\right) \sim N(0, \boldsymbol{\Omega}_b),$$

where $Y_{ijl}^{(1)}$ denotes partially observed patient's gender (level 1 covariate) and outcome subcomponents in the treatment domain (i.e., missing patient's amoxicillin dose and frequency of administration). Level 1 predictors of interest ($\boldsymbol{X}_{ijl}^{(1)}$) included fully observed covariates (i.e., an interaction term between follow-up time and intervention arm, hospital workload and malaria prevalence status, patient's age and number of comorbid illnesses). We also included outcome subcomponents in the assessment and diagnosis domains as level 1 predictors. The second level of the imputation model targeted missing clinicians' cadre and gender respectively.

A burn-in of 1000 updates and 100 iterations between each of the 30 imputations were considered upon satisfactory convergence. After MI of missing subcomponents in the treatment domain, pneumonia PAQC score was constructed before fitting substantive models of interest. 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 of oral amoxicillin administration (Opondo et al., 2016). Random-effects model (3.7) and generalized estimating equation (GEE) model (3.8) were used to analyse the trial data. Final parameter estimates were pooled according to Rubin's rules (Rubin, 1976). MI results were compared to complete case analysis results (after deletion of case records with missing clinician's cadre, gender and patient's gender) combined with conventional approach of handling missing PAQC score elements. A 5% level of significance was considered in all statistical analyses.

4.2.3 Simulation Study

We sought to simulate data mimicking the observed pneumonia trial data set. However, simulating a standard data set based on model parameters while preserving the correlation structure was a challenge due to the complex multilevel structure of the trial data set. This was in addition to mixed variable types in covariates and outcome subcomponents. To circumvent this challenge, missing data were generated in a complete subset of pneumonia trial data.

Specifically, our simulation study targeted subcomponents in the treatment domain namely patient's weight, oral amoxicillin dose prescribed and frequency of oral amoxicillin administration. These three pneumonia care indicators are required in the calculation of correctness of prescribed oral amoxicillin dose. To create a subset of pneumonia trial data, complete in the treatment domain subcomponents of interest, we excluded 65/2127 (3.1%) case records with missing oral amoxicillin prescription. Out 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 base data set used in the simulation study consisted of 2001(94.1%) pneumonia patients nested within 372 admitting clinicians in 12 hospitals. Although the data set was complete in the outcome subcomponents of interest, one patient and two clinician level covariates still had missing data. Specifically, patients' gender was missing in $<1\%$ of the case records while clinicians' cadre and gender were missing in 22.3% (83/372) and 25.1% (82/372) cases respectively.

4.2.4 Standard Parameter Estimates

The base dataset was used to estimate standard parameter estimates as follows. First, pneumonia PAQC score was constructed for each patient. Thereafter, missing covariates were imputed 10 times using the latent normal approach within multilevel joint model imputation framework presented in section (3.2).

In this case, only partially observed covariates in level 1 and level 2 were imputed. Each imputed data set was analyzed using proportional odds random clinician's intercepts model (equation 3.7). Final parameter estimates were pooled according to Rubin's rules (Rubin, 1976). The pooled estimates henceforth referred to as standard estimates and denoted hereby (β^*_{MI}) were used as reference estimates against which results from different simulation scenarios were benchmarked.

4.2.5 Simulation Scheme

Missing data were induced in the base data set targeting three outcome subcomponents in the treatment domain, that is, patient's weight, oral amoxicillin dose prescribed and frequency of oral amoxicillin administration. Missingness was generated assuming MCAR and MAR mechanisms, respectively. Binary missing data indicators were generated by sampling random numbers from a random binomial distribution with success rates of 3%, 10% and 40%. A 3% missing data rate was selected to evaluate the impact of low proportion of missingness while 10% and 40% were chosen to assess the extent of bias in moderate to high rates of missingness. Under MCAR mechanism, missing values in the target treatment domain subcomponents were induced independent of other variables in the base data set, that is, covariates and outcome subcomponents in the assessment and clinical diagnosis domains. For the MAR condition, probabilities of missing data were conditionally dependent on variables associated with probability of missingness in the three variables of interest (based on the observed trial data set) (Supplementary Table B.1).

In both MAR and MCAR, missing data in the target variables were induced independently of each other, such that either one, two or all three variables were missing for any given patient. Each scenario was simulated 1,000 times. Random number generators (seeds) were chosen and maintained for different scenarios to ensure reproducibility of results.

Thereafter, two approaches were used to handle missing data in each simulated data set. In the first approach, only missing covariates at patient and clinician level were handled using MI. On the other hand, all partially observed outcome subcomponents including patient's weight, amoxicillin dose and frequency of administration (outcome subcomponents within treatment domain) were handled using the conventional approach where they were score with value 0 at PAQC score construction stage described in Table 3.1. In this case, PAQC score was constructed prior to multiple imputation of missing covariates and hence included in the imputation model as a one of the predictor variables.

In the second approach, MI was used to handle partially observed covariates and missing treatment domain subcomponents. Outcome subcomponents in the assessment and diagnosis domains were included in the imputation model as predictor variables. In this approach, PAQC score was constructed after MI of incomplete subcomponents in the treatment domain. Variation on the 7-point scale was attributed to inappropriate pneumonia care which encompassed; lack of documentation of all primary sign and symptom, lack of documentation of all secondary signs and symptoms (assessment domain), misdiagnosis or misclassification of disease severity, failure to prescribe oral amoxicillin, prescription of

oral amoxicillin in wrong dosage or wrong frequency of oral amoxicillin administration.

4.2.6 Performance Measures

A proportional odds random intercepts model (equation 3.7) was fitted to each imputed data set to obtain imputation specific parameter estimates. Imputation-specific estimates were pooled using Rubin's rules (Rubin, 1976) to produce a single estimate $(\hat{\beta}_{i,MI})$ for the i^{th} simulation. This procedure was repeated in all the scenarios. Bias in regression coefficients was calculated as the differences between estimates (log odds) averaged over 1000 simulated data sets $(\bar{\hat{\beta}}_{MI} = \sum_{i=1}^{N} \hat{\beta}_{i,MI}/N)$ and estimated (log odd)(β^*_{MI}) from the base data set. That is,

$$Bias = \bar{\hat{\beta}}_{MI} - \beta^*_{MI}. \tag{4.2}$$

To assess accuracy, model based standard errors calculated as the average of the estimated within simulation standard errors were used. That is,

$$Model\ \ based\ \ SE(\hat{\beta}) = \sum_{i=1}^{N} SE(\hat{\beta}_{i,MI})/N.$$

The model based standard errors were compared with empirical standard errors calculated as the standard deviation of the estimates of interest (Burton et al., 2006) across the 1000 data sets, that is,

$$Empircal\ \ SE(\hat{\beta}) = \sqrt{1/(N-1)\sum_{i=1}^{N}(\hat{\beta}_{i,MI} - \bar{\hat{\beta}}_{i,MI})^2}, \tag{4.3}$$

where N denotes the number of simulations, $\hat{\beta}_{i,MI}$ is the coefficient estimated in the i^{th} simulation and $\bar{\hat{\beta}}_{i,MI}$ is estimator's average over 1000 simulations. The mean square error (MSE) which incorporates both measures of bias and variability (Burton et al., 2006; Enders et al., 2016; Grund et al., 2018) was calculated for the regression coefficients as

$$MSE = (Bias)^2 + (Empircal\ \ SE(\hat{\beta}))^2. \tag{4.4}$$

Bias and accuracy of the corresponding standard errors were assessed in a similar manner. Coverage probability of the 95% confidence intervals were not applicable in this simulation study because missing data were simulated on the same subset of the pneumonia trial data set. Computation time was also used to assess performance of the two strategies employed in handling missing PAQC score subcomponents. To assess variability due to finite number of simulations (Morris et al., 2019), Monte-Carlo standard errors for estimated bias in regression parameters were calculated using

$$Monte-CarloSE(\beta) = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(\hat{\beta}_{i,MI} - \bar{\hat{\beta}}_{i,MI})^2}. \tag{4.5}$$

Simulations were carried out using a server with the following specification: 40 GB memory, Intel Xeon E5-4650 (2.70GHz) processor (12 cores/24 threads), Gnu/Linux Ubuntu 14.04 OS, and R (version 3.4.4) programming language.

4.3 Results

4.3.1 Pneumonia Trial Data Results

Table 4.1: Parameter estimates (standard errors) from random- effects and marginal GEE models: complete case analysis and after multilevel MI.

	Random-effects model				Marginal model: GEE[a]			
	Complete case analysis N=1619 (76.1%)		Multilevel MI[b] N=2127 (100%)		Complete cases analysis N=1619 (76.1%)		Multilevel MI N=2127 (100%)	
Effect	Estimate (SE[c])	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
PAQC[d] score intercept 0	Reference	-	Reference	-	Reference	-	Reference	-
PAQC score intercept 1	-7.77 (1.076)	<0.001	-7.74 (0.829)	<0.001	-7.41 (1.032)	<0.001	-7.19 (0.794)	< 0.00
PAQC score intercept 2	-1.77 (0.383)	<0.001	-2.20 (0.341)	<0.001	-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)	<0.001
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)	<0.001	1.30 (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)	<0.001
Patient's 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
Patient's gender: 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
Clinician's gender: 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
Clinician's cadre: MO[e]	0.02 (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)	<0.001	0.25 (0.060)	<0.001	0.27 (0.052)	<0.001
Variance between random intercepts	1.328 (1.151)		1.161 (1.073)					

GEE[a]: -Generalized estimating equations, MI[b]: -Multiple imputation, SE[c]:- Standard Error, MO[e]:- Medical Officer, A&F[f]:-Audit and feedback

Table 4.1 presents random intercepts and GEE models parameter estimates (in log odds) and the corresponding standard errors obtained under complete case analysis and after MI of missing covariates and missing PAQC score subcomponents in the treatment domain. Overall, MI led to more precise estimates across all variables compared to complete case records methods in both GEE and random intercepts models. We also observed change in the regression coefficients before and after imputing missing covariates and PAQC score subcomponents in the treatment 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 4.1). 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 gender effect (log odds= -0.03) under complete case analysis which changed to a positive effect (log odds= 0.01) after MI (Table 4.1).

4.3.2 Simulation Study Results

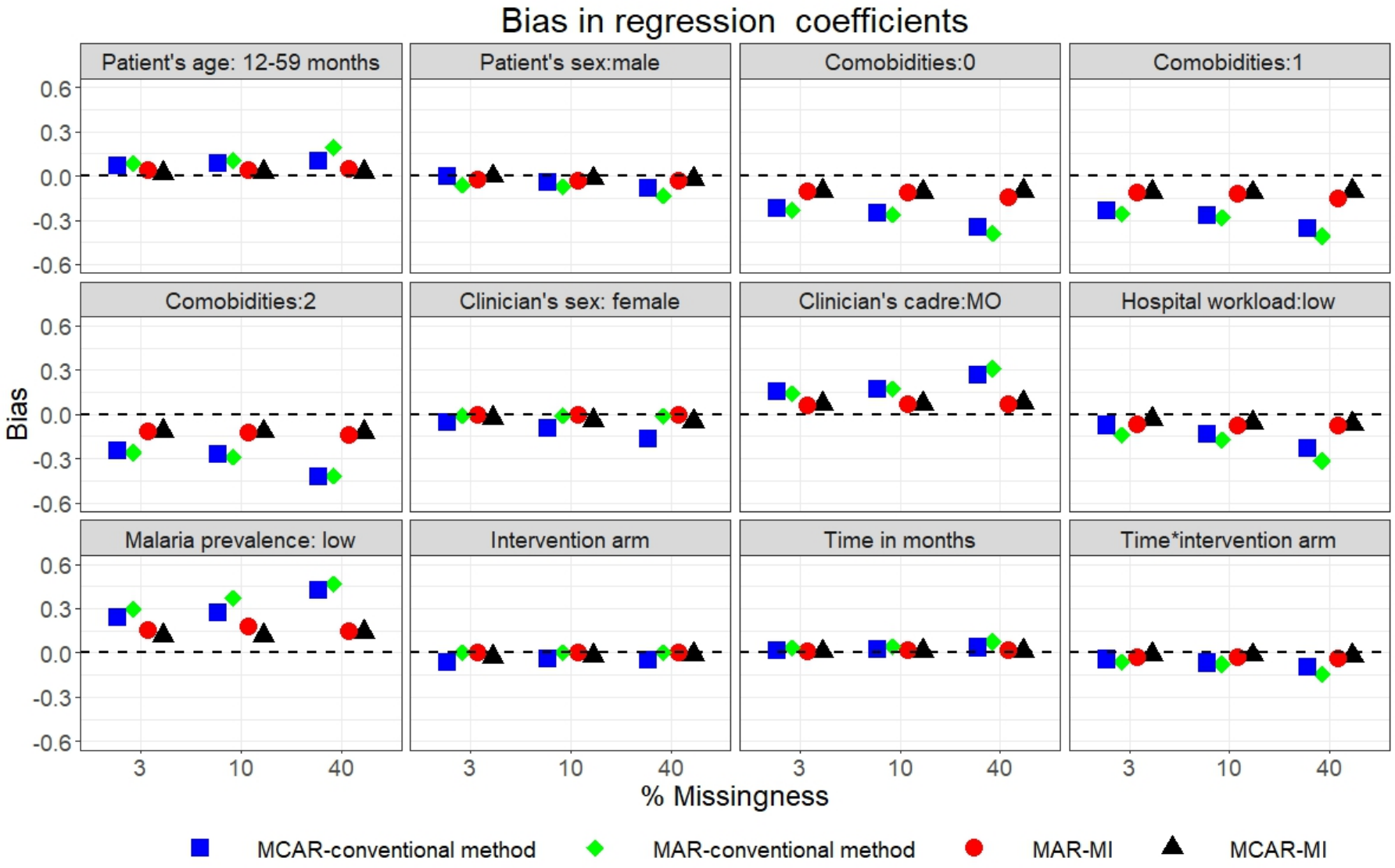

Figure 4.1: Bias in regression coefficients under the conventional approach of handling missing PAQC score subcomponents and after multiple imputation of missing PAQC score subcomponents in the treatment domain and missing covariates imputed across missing data rates and missing data mechanisms

Figure 4.2: Bias in standard errors under the conventional approach of handling missing PAQC score subcomponents 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.

Figure 4.1 and Figure 4.2 respectively, present bias estimated in regression coefficients and standard errors under the conventional and MI approaches across 6-simulation scenarios (i.e., 3 missing data rates of 3%, 10% and 40% and 2 missing data mechanism namely MAR and MCAR). Results for specific scenarios with regard to estimated bias, empirical standard errors, model based standard errors and MSE for regression coefficients and corresponding standard errors are presented in Tables 4.1, Table 4.2, and supplementary Tables B.2-B.7. Monte-Carlo standard errors and confidence interval around bias for regression parameters across simulation scenarios are presented in Supplementary Tables B.8-B.9.

Across 6-simulation scenarios, the regression coefficients either underestimated (negative bias) or overestimated (positive bias) the standard estimates (Figure 4.1). Moreover, the magnitude of bias varied across variables and tended to increase with an increase in the proportion of missingness. However, the bias was much smaller when MI was used to handle incomplete treatment subcomponents compared to the conventional approach (Figures 4.1). On the other hand, the standard errors tended to overestimate the base data set resulting to positive bias across simulation scenarios (Figure 4.2). For individual variables, it was further observed that the standard errors were less prone to bias compared to regression coefficients. These observations were made within and across simulation series. Moreover, simulation results exhibited larger bias when missingness in treatment domain subcomponents were generated under MAR mechanism compared to MCAR mechanism.

Across simulation scenarios, the estimated empirical standard errors were close

to the estimated model based standard errors (Table 4.1, Table 4.2 and Supplementary Tables B.2-B.7). In addition, the magnitude of both measures of accuracy tended to increase with an increase in the proportion of missing data in PAQC score components. The results further showed that MSEs were slightly larger under the conventional approach compared to MI approach and were somewhat larger under MAR mechanism compared with MCAR mechanism (Table 4.2, Table 4.3 and Supplementary Tables B.2-B.7).

Across simulation scenarios, that is, missing data mechanisms and rates of missingness, Monte-Carlo standard errors of estimated bias in regression parameters ranged between 0.001 and 0.04 (Supplementary Tables B.8-B.9). The corresponding 95% confidence intervals around bias in the parameters of interest were narrow across simulation settings. Finally, the simulation process was on average more time intensive under MI strategy compared to the conventional approach. Furthermore, the computational time increased with an increase in the proportion of missing data irrespective of the mechanism used to generate missing data.

Table 4.2: Performance measures of regression coefficients after multiple imputation of covariates and outcome elements: MAR[a] mechanism.

		Proportion Missing											
		3%				10%				40%			
Effect	True est[b]	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
Patient's 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
Patient's gender: 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
Clinician's gender: 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
Clinician's cadre: MO	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	-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

MAR[a]:-Missing at Random, True est[b]:-True estimate, Emp SE[c]:-Empirical standard error, MSE[d]:- Mean Square Error, MO[e]:- Medical Officer, A&F[f]:-Audit and feedback

Table 4.3: Performance measures of regression coefficients after multiple imputation of covariates and outcome elements: MAR[a] mechanism.

		Proportion Missing											
		3%				10%				40%			
Effect	True est[b]	Bias	Model-based SE	Emp SE[c]	MSEd	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's gender: 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' gender: 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	-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[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

MAR[a]:-Missing at Random, True est[b]:-True estimate, Emp SE[c]:-Empirical standard error, MSE[d]:- Mean Square Error, MO[e]:- Medical Officer, A&F[f]:-Audit and feedback

4.4 Discussion

In this chapter, we sought to explore and propose appropriate strategy for handling missing data in PAQC score (Opondo et al., 2016), an ordinal composite outcome. In composite measures development guidelines, a required step is a strategy for handling missing data to minimize bias and enhance reliability of a composite score (Profit et al., 2010; Commission et al., 2008). However, in the literature, most studies reporting composite measures avoid missing items in composite scores by conducting complete case analysis (Simons et al., 2015; Cordoba et al., 2010). In Chapter 3, we imputed missing covariate while all missing PAQC score subcomponents were handled using the conventional method. In the present chapter, we used MI to handle missing PAQC score subcomponents in the treatment domain in addition to partially observed covariates. Although the proportion of missingness in PAQC score subcomponents of interest (treatment domain) was small, we observed notable differences in parameter estimates.

Through a range of simulation conditions, we explored bias in regression coefficients and standard errors associated with missing data in PAQC score treatment domain subcomponents. The study results demonstrated superiority of MI as a strategy for dealing with partially observed PAQC score domain subcomponents over the conventional method. Nevertheless, MI approach led to some level of bias in the regression coefficients. These observations could be due to lack of compatibility between the imputation model and the analysis model considering

that PAQC score was not included in the imputation model. To be specific, incomplete subcomponents in the treatment domain were included in the imputation model as target while subcomponents in the assessment and diagnosis domains were included as predictors variables. In this case, the composite outcome was constructed after MI step. Therefore, further research is needed to compare the performance of proposed MI method with that of MI including the composite outcome, possibly adapting substantive model compatible MI approaches (Bartlett et al., 2015) to this setting, in order to guarantee that the relation between subcomponent and composite outcome is preserved (Quartagno and Carpenter, 2018).

Simulation results further showed that regression coefficients were more prone to bias compared to standard errors across simulation scenarios. A possible explanation for these results is that case records with missing PAQC score subcomponents in the treatment domain were not discarded in both conventional approach and MI approach and hence no major impact on estimation of standard errors.

Previously, MI has been used to address missing data at component level in composite scores assessing quality of patient's care (Blough et al., 2009; Plumpton et al., 2016). Elsewhere, Simons et al. (2015) proposed MI at index level particularly for smaller samples. In the case of PAQC score, there are no possibilities of missing PAQC score at aggregate level (the only possibilities are values between zero and six). Therefore, multiple imputation can only be implemented at subcomponents level. In this study, MI was used to handle missing treatment domain subcomponents while undocumented pneumonia signs and symptoms

in the assessment domain were regarded as inappropriate care and hence scored zero in the binary indicators at PAQC score construction stage. This was in consideration of the trial's inclusion criterion which required recruitment of patients with syndromic pneumonia. That is, patients with at least one of the two primary pneumonia signs and symptoms (i.e., presence of cough or difficult breathing) in addition to at least one secondary sign and symptom necessary for pneumonia severity classification (Organization, 2013; Ayieko et al., 2017). As such, imputation of undocumented signs and symptoms was not expected to have any meaningful impact on the simulation study. However, it should be noted that MI can be extended to handle missing PAQC score subcomponents in other domains of routine care in studies without such restrictive inclusion criteria. Moreover, analyses and MI procedure proposed in this study can be extended to other MI techniques (Jenghara et al., 2018; Mostafa, 2019) in order to examine performance in terms of computational cost, bias and measures of accuracy as appropriate.

In conclusion, MI produce minimally biased estimates in comparison with conventional method. However, the regression coefficients are more prone to bias compared to standards errors more so when the underlying mechanism is MAR. Besides, bias tended to increase with an increase in proportion of missing variable in the outcome subcomponents. Therefore, missing data in subcomponents composite measures should be addressed carefully to alleviate potential for biased estimate and misleading inferences.

Chapter 5

Sensitivity analysis of departure from missing at random (MAR) assumption

5.1 Introduction

In practice, MAR and MNAR mechanisms cannot be distinguished using observed data only, hence the need for sensitivity analyses (Mackinnon, 2010). Sensitivity analyses entail scrutinizing plausible models assuming MNAR mechanisms to assess departures from the MAR assumption. Alternatively, the primary analysis model is changed through a number of alterations and the stability of inferences across the alternative settings assessed (Héraud-Bousquet et al., 2012; Liublinska and Rubin, 2014).

As already mentioned, sensitivity analyses following MI can be conducted within

three generic frameworks namely; pattern-mixture models, selection models and shared parameter models (Héraud-Bousquet et al., 2012; Little et al., 2012; Liublinska and Rubin, 2014). Nonetheless, sensitivity analysis within 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 (Héraud-Bousquet et al., 2012; Smuk et al., 2017). Besides, sensitivity analyses methods are underdeveloped in standard statistical software thus limiting their application in practice (Héraud-Bousquet et al., 2012).

In health care settings, completeness of routine data depends on an interplay of factors that operate at the patient, clinician and healthcare facility levels (Yelin et al., 2015). 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 (Rowe et al., 2005). 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 (Thomas et al., 2017; Tuti et al., 2017).

This problem of missing data at the clinician level is compounded when missing data are handled using inappropriate methods, that increase the risk for obtaining biased and inefficient estimates hence misleading inference (Mackinnon, 2010;

Tompsett et al., 2018). 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 (Mackinnon, 2010). However, similarities between CCA and MI results could lead to false reassurances that data are either MCAR or missing at random with a mechanism not involving the outcome (i.e., covariate-dependent MAR (Molenberghs et al., 2014, Chapter 1) whereas a MNAR mechanism could be in operation (Mackinnon, 2010).

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 also conducted sensitivity analyses aimed at assessing robustness of inference under MAR mechanism using two approaches within the pattern-mixture models framework.

The rest of this Chapter is structured as follows. Section 5.2 provides details of multilevel MI under MAR mechanism and MNAR mechanism respectively followed by results in Section 5.3. The chapter concludes with a discussion in Section 5.4.

5.2 Methods

5.2.1 Multiple Imputations under MNAR assumption: Sensitivity analyses

In chapters 3 and 4, we imputed missing covariates and PAQC score components in the treatment domain within the joint model framework while assuming a MAR mechanism. More details on multilevel MI assuming a MAR mechanism are presented in Chapter 4 Section 4.2.3. As mentioned above, MAR assumption cannot be verified using observed data alone. Therefore, we imputed missing data assuming a MNAR mechanism to assess possible departures from MAR mechanism. Our sensitivity 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. Specifically, we considered MNAR imputation in level two variables (i.e., clinician's gender and cadre) while retaining the MAR imputation models for patient-level (level one) variables for two reasons. First, we aimed to minimize complexities at analysis stage considering that three out of four patient-level variables (i.e., patient's weight, amoxicillin dose prescribed and frequency of amoxicillin administration) were subcomponents 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 (Smuk et al., 2017). In the second approach, we modified MI model assuming MAR mechanism through a range of sensitivity parameters (delta adjustment approach) (Carpenter and Kenward, 2013). These changes can be informed by opinions elicited from experts in the subject matter or contextual knowledge (Molenberghs et al., 2014, Chapter 20).

Pattern mixture models

The PMM assumes that observations are stratified based on patterns of missing data, and distinct models formulated to estimate parameters within each pattern. 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 (Molenberghs et al., 2014, Chapter 19, p. 439)

5.2.2 Elicitation of Experts' Opinion

In this study, we elicited clinical experts' opinions to enable us to define suitable MNAR assumption about the differences in the distribution of clinicians with observed cadre/gender and clinicians with missing cadre/gender. Preliminary investigations into underlying missing data pattern showed that nearly all clinicians with missing cadre had missing gender (Figure 3.1). Further assessment revealed that intervention arm and paediatric admission workload were predictor variables for missing cadre and gender respectively. Preliminary analysis sug-

gested that both trial arm and admission workload were strongly associated with the observed clinician's cadre and gender respectively. Therefore, we defined 4 patterns labelled k based on combinations of admission workload and trial arm as

$$k = \begin{cases} 1 & \text{if hospital is in control arm and has high paediatric admission workload} \\ 2 & \text{if hospital is in control arm and has low paediatric admission workload} \\ 3 & \text{if hospital is in intervention arm and has high paediatric admission workload} \\ 4 & \text{if hospital is in intervention arm and has low paediatric admission workload.} \end{cases} \tag{5.1}$$

For each k, we estimated data predicted probabilities of a clinician belonging to a cadre (i.e., clinical officers or clinical officer interns or medical officers or medical officer interns) under the MAR assumption (Smuk et al., 2017). Specifically, we imputed missing clinicians' cadre and gender jointly assuming a MAR mechanism with trial arm and admission workload as predictor variables. Thereafter, we separately regressed clinicians' cadre on trial arm and admission workload using a multinomial logistic model on each imputed data set. The final estimates (log odds) were pooled according to Rubin's rule (Rubin, 1976) were then used to determine data predicted probabilities of clinicians belonging to either of the four cadre categories for each k.

Similarly, we fitted a logistic regression model to each imputed data set with clinician's gender as the outcome and 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 (Figure 5.1) and clinician's gender (Appendix Figure C.1) were then presented to experts in 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. 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 for each $k = 1, 2, 3, 4$. Here, we use $j = 1$ to denote 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. Similarly, they filled in their belief about clinicians with missing gender being females $(j = 1)$ or males $(j = 2)$ in each k. Experts' predicted probability for clinician's gender/cadre are denoted by θ_{jk}.

After the elicitation exercise, we pooled experts predicted probabilities and calculated 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.

Control arm and high paediatric admission workload hospital				
	Clinician cadre			
	CO	CO intern	MO	MO intern
Data prediction (%)	1	38	1	60
Your prediction (%)				
Minimum	0	0	0	0
Maximum	100	100	100	100
Control arm and low paediatric admission workload hospital				
	Clinician cadre			
	CO	CO intern	MO	MO intern
Data prediction (%)	3	45	2	50
Your prediction (%)				
Minimum	0	0	0	0
Maximum	100	100	100	100
Intervention arm and high paediatric admission workload hospital				
	Clinician cadre			
	CO	CO intern	MO	MO intern
Data prediction (%)	1	42	2	55
Your prediction (%)				
Minimum	0	0	0	0
Maximum	100	100	100	100
Intervention arm and low paediatric admission workload hospital				
	Clinician cadre			
	CO	CO intern	MO	MO intern
Data prediction (%)	1	50	2	47
Your prediction (%)				
Minimum	0	0	0	0
Maximum	100	100	100	100

Figure 5.1: Questionnaire tables used to elicit experts' opinions about missing clinician's cadre

Dirichlet Conjugate Prior for Multinomial Distribution

For clinicians' cadre with four categories, we chose a Dirichlet distribution as an appropriate conjugate prior distribution (Smuk et al., 2017). A Dirichlet distribution with four parameters is formulated as

$$f(x_{jk}, x_{jk}, x_{jk}, x_{jk}, \alpha_{jk}, \alpha_{jk}, \alpha_{jk}, \alpha_{jk}) = \frac{\Gamma \sum_{j=1}^{J} \alpha_{jk}}{\prod_{j=1}^{J} \Gamma(\alpha_{jk})} \prod_{i=1}^{k} x_{jk}^{\alpha_{jk}-1}, \tag{5.2}$$

where the vector x_{jk} denotes probabilities for different categories in the variable of interest, $\sum_{j=1}^{J} x_{jk} = 1$ and α_{jk} are concentration parameters. The mean and variance of Dirichlet distribution are denoted by

$$E(x_{jk}) = \frac{\alpha_{jk}}{L_k}, \tag{5.3}$$

and

$$Var(X_{jk}) = \frac{\alpha_{jk}(L_k - \alpha_{jk})}{L_k^2(L_k + 1)}, \tag{5.4}$$

where $L_k = \sum_{j=1}^{J} \alpha_{jk}$. Using the means and variances of experts predicted probabilities $(E[\theta_{jk}])$ and $(Var[\theta_{jk}])$ for j^{th} cadre $(j = 1, 2, 3, 4)$ in each k, $(k = 1, 2, 3, 4)$, we estimated Dirichlet distribution concentration parameters using the methods of moments (Smuk et al., 2017) as follows:

1. Using a sequence of values between 1 and 40 (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 dis-

tribution in equation (5.3) using

$$\alpha_{jk} = L_k * E(\theta_{jk}). \tag{5.5}$$

2. We substituted α_{jk} values obtained in step 1 in the variance formulae (equation 5.5) to estimate Dirichlet distribution variances $Var(x_{jk})$ for each value in the sequence L_k.

3. We plotted Dirichlet distribution variance 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 category (i.e., clinical officers, clinical officer interns, medical officers and medical officer interns) (Figure 5.2). The step was repeated for the other patterns of missingness, that is, $k = 2, 3, 4$. The corresponding figures are presented in the Supplementary file (Appendix Figures C.2-C.4).

4. We determined the value in the sequence L_k for which estimated Dirichlet estimated 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 category. We summed L_k values across the four cadre categories and divided the total by four. The mean was denoted by $E(L_k)$.

5. We determined Dirichlet distribution parameters for the j^{th} cadre in each k

by multiplying expert predicted mean probabilities $E(\theta_{jk})$ by $E(L_k)$, that is,

$$\hat{\alpha}_{jk} = E(L_k) * E(\theta_{jk}). \tag{5.6}$$

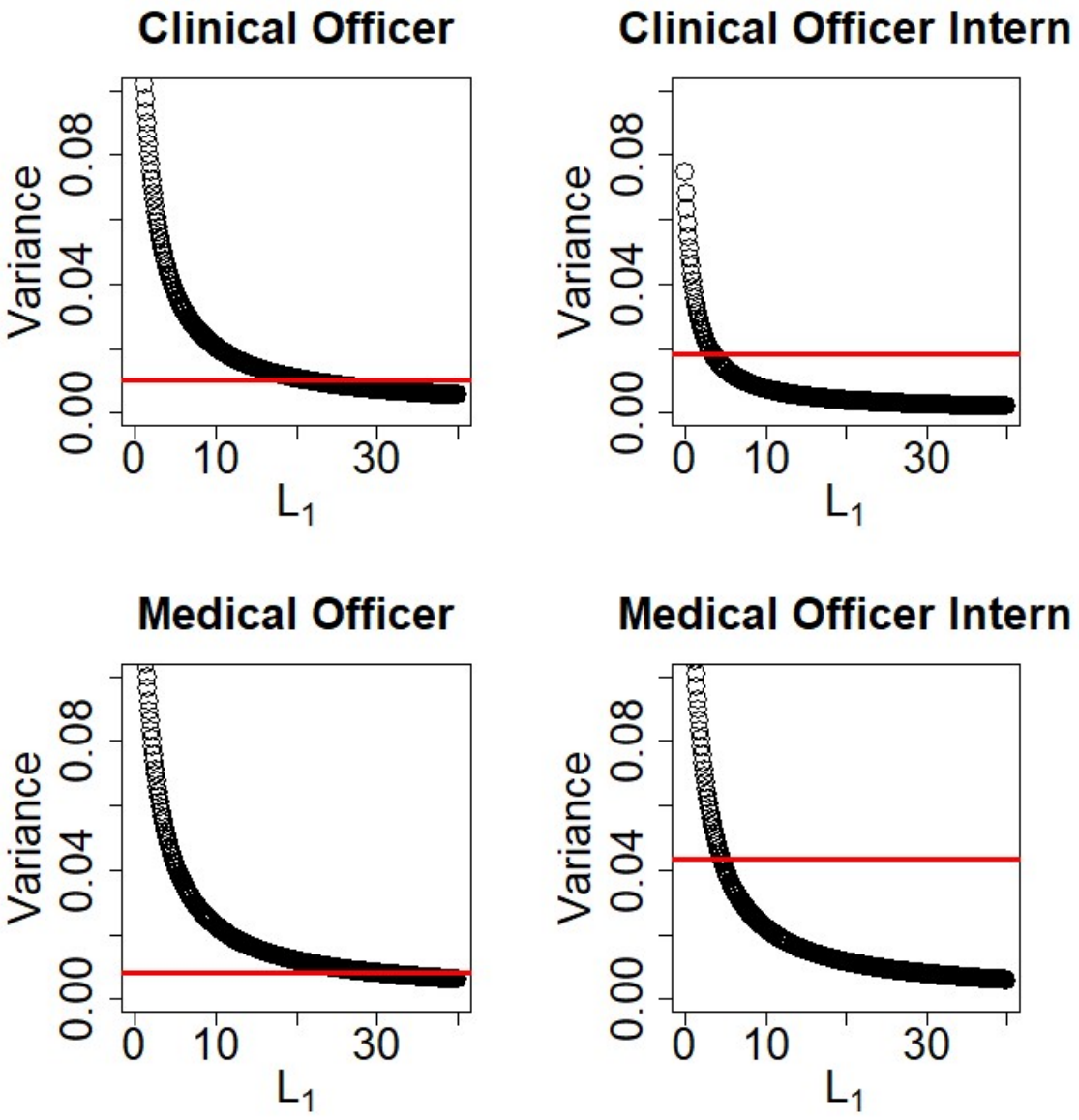

Figure 5.2: Estimated Dirichlet variances (black curves) and experts' variances (horizontal red lines) in a control hospital with high admission workload $(k = 1)$

Estimated concentration parameters for Dirichlet distribution for a given k, $(k, 1, 2, 3, 4)$

are presented in Table 5.1. The parameter vectors were used to generate random vectors of probabilities of j^{th} cadre probabilities in each k.

Table 5.1: Estimated concertation parameters (α) and precision (L) for the Dirichlet distribution estimated using moments method.

	CO	CO intern	MO	MO intern	
k	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$E[L_k]$
1	1.19	5.08	1.34	7.63	14.96
2	1.96	11.78	2.62	16.36	32.73
3	0.58	3.01	1.59	3.09	8.35
4	1.06	4.38	2.12	5.58	13.28

Beta Conjugate Prior for Binomial Distribution

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

$$f(x) = \frac{\Gamma(\alpha_{jk} + \beta_{jk})}{\Gamma(\alpha_{jk})\Gamma(\beta_{jk})} x^{\alpha_{jk}-1}(1-x)^{\beta_{jk}-1}, \tag{5.7}$$

where $\alpha_{jk} > 0$ and $\beta_{jk} > 0$. Using the mean $(E[\theta_{jk}])$ and variances $(Var[\theta_{jk}])$ of experts predicted probabilities for j^{th} $(j = 1, 2)$ clinician's gender category in the k^{th} strata $(k = 1, 2, 3, 4)$, we estimated α_{jk} and β_{jk} using the moments method (Lunn et al., 2012) as shown below

$$\hat{\beta}_{jk} = \frac{E[\theta_{jk}](1 - E[\theta_{jk}])^2}{Var[\theta_{jk}]} + E[\theta_{jk}] - 1 \tag{5.8}$$

$$\hat{\alpha}_{jk} = \frac{E[\theta_{jk}] * \hat{\beta}_{jk}}{1 - E[\theta_{jk}]} + E[\theta_{jk}] - 1. \tag{5.9}$$

The approximated $\hat{\alpha}_{jk}$ and $\hat{\beta}_{jk}$ parameters for each k, $k = 1, 2, 3, 4$ are presented in Table 5.2.

Table 5.2: Beta distribution parameters approximated using moments method

	j=female clinician	
k	$\hat{\alpha}$	$\hat{\beta}$
1: Control arm and high paediatric admission workload	6.5	4.5
2: Control arm and low paediatric admission workload	4.3	3.6
3: Intervention arm and high paediatric admission workload	4.6	3.9
4: Intervention arm and low paediatric admission workload	7.3	6.1

5.2.3 Multiple Imputations from MNAR Prior Distributions

Using Dirichlet and beta prior parameter estimate vectors (Tables 5.1 and 5.2), we generated 20 random probability vectors for each k, $(k = 1, 2, 3, 4)$. The number of random draws 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 $(i = 1, 2, \ldots, 20)$ was then used to determine the proportion of occurrence of clinicians' cadre/gender category in the k^{th} stratum. For clinician's gender, 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. 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.

5.2.4 Multiple Imputation with Delta Adjustment Method)

MI with delta adjustment involves adding a fixed quantity δ to the linear predictor of the imputation model (Yuan, 2014; Tompsett et al., 2018). For continuous target variables, δ represents the difference in mean between non-respondents and respondents (Little et al., 2012). When the variable of interest is categorical, addition of shift parameter in the imputation model modifies the predicted probabilities for the classification levels thus producing MNAR imputed values (Little et al., 2012; Tompsett et al., 2018).

In this study, we conducted separate MI-MNAR analyses for clinicians' gender and clinicians' cadre rather than two-dimensional sensitivity analysis. In first imputation model, we modified the probability of classification among clinicians with missing gender while missing clinicians' cadre was imputed without any modifications. In second imputation model, 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 functions from the *jomo* package in R (version 3.5.4) (Quartagno et al., 2019). These functions are not yet available in the version of the package available in CRAN but will be included in the near future. The modified multilevel joint imputation model is formulated as follows

$$\begin{aligned} \boldsymbol{Y}_{ijl}^{(1)} &= \boldsymbol{X}_{ijl}^{(1)}\boldsymbol{\beta}^{(1)} + b_{jl}^{(1)} + e_{ijl}^{(1)} \\ \boldsymbol{Y}_{jl}^{(2)} &= \boldsymbol{X}_{jl}^{(2)}\boldsymbol{\beta}^{(2)} + \delta(1 - R_{jl}) + b_{jl}^{(2)} \end{aligned} \tag{5.10}$$

$$e_{ijl} \sim N(\mathbf{0}, \mathbf{\Sigma}) \ \ and \ \ \left(b_{jl}^{(1)}, b_{jl}^{(2)}\right) \sim N(\mathbf{0}, \mathbf{\Omega}_b),$$

where $\mathbf{Y}_{ijl}^{(1)}$ is a vector of partially observed level one 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 $\mathbf{Y}_{jl}^{(2)}$ while R_{jl} is a binary indicator with value 1 if clinicians' gender/cadre is observed and 0 if clinicians' gender/cadre is missing. When δ is zero, a MAR mechanism is implied (Carpenter and Kenward, 2013).

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 quartiles values. We chose three shift parameters (i.e., $\delta = -0.2, -0.3, -0.5$) to alter probability of classification in the imputation of clinicians' gender. These parameters corresponded to the three quartiles of the prior distributions elicited from experts. Specifically, 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 shift values used to alter classification probabilities for clinicians' gender were also used to alter classification probabilities among clinicians with missing cadre. This was in consideration of the similarities in missing data patterns underlying clinicians' gender and clinicians' cadre (Appendix, Figure A.1). In this case, negative shift parameters increased the probability of being medical of-

ficers 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 differences in proportion of classification increased with an increase in the magnitude of shift parameters. The MI-MNAR analysis under the delta-adjusted approach was repeated for different shift parameters.

5.2.5 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 procedure outlined in Section 3.2.1. For each imputed data set, we fitted a proportional odds random intercepts model (3.7) in Section 3.2.5. Thereafter, we combined MI estimates using Rubin's rules (Rubin, 1976) and compared inference between 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 patients and clinicians level variables.

5.3 Results

Table 5.3 presents a summary of data predicted probabilities and experts' predicted probabilities (mean and variance) for four cadre categories in each combination of trial arm and admission workload respectively. Among clinicians with missing cadre, experts believed that medical officers and clinical officers

were more than clinical officer interns and medical officer interns respectively. These observations were in contrast 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 5.3). 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 5.3). In both clinicians' gender and cadre, experts' responses did not vary widely across cadre/gender categories and across stratification groups.

Table 5.3: Data predicted and expert predicted probabilities (mean and variance) for clinicians' cadre.

K	Data probabilities (p_{jk})	Elicited probabilities $E(\theta_{jk}), Var(\theta_{jk})$
Clinician's cadre		
1: Control arm and high workload		
CO intern	0.38	0.12 (0.08)
CO	0.01	0.14 (0.10)
MO intern	0.60	0.49 (0.12)
MO	0.01	0.25 (0.09)
2: Control arm and low workload		
CO intern	0.45	0.17 (0.12)
CO	0.03	0.39 (0.11)
MO intern	0.50	0.29 (0.10)
MO	0.02	0.15 (0.05)
3: Intervention arm and high workload		
CO intern	0.42	0.23 (0.05)
CO	0.01	0.23 (0.09)
MO intern	0.55	0.22 (0.06)
MO	0.02	0.31 (0.08)
4: Intervention arm and low workload		
CO intern	0.50	0.25 (0.04)
CO	0.01	0.25 (0.12)
MO intern	0.47	0.31 (0.06)
MO	0.02	0.19 (0.05)
Clinician's 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)

Table 5.4 shows the distribution of clinicians' cadre and gender under complete case analysis and under MAR and MNAR mechanisms. When clinician's cadre was the sensitivity variable of interest, we observed a systematic increase in the proportion of clinicians imputed as medical officer and medical officer interns respectively. Similarly, when clinician gender was the sensitivity variable of interest, more clinicians were imputed as males compared to females. For clinician's cadre, the proportions of medical officer and medical officer interns tended to increase with an increase in magnitude of sensitivity parameter (delta values).

Furthermore, we observed similarities in the proportions of clinicians' gender and clinicians' cadre after MI from prior distributions and delta adjustment with a sensitivity parameter equal to -0.2 (Table 5.4). Considering the few numbers of clinical officer and medical officers , we grouped clinician into two categories as done in previous chapters.

Table 5.4: Proportion of clinicians' cadre and gender in complete records and under MAR and MNAR: 20 imputations used

			Sensitivity analysis variable: Clinician's cadre				Sensitivity analysis variable: Clinician's gender			
			MI-MNAR				MI-MNAR			
	Complete Records	MI-MAR	$\delta = -0.2$	$\delta = -0.3$	$\delta = -0.5$	Dirichlet prior	$\delta = -0.2$	$\delta = -0.3$	$\delta = -0.5$	Beta Prior
Clinician's cadre										
CO	0.52	1.05	0.55	0.60	0.69	1.58	0.69	0.68	0.88	0.64
COI	39.80	43.58	40.31	39.59	36.59	39.19	44.47	43.53	44.38	45.34
MO	2.62	2.82	3.62	4.17	4.51	4.71	2.87	2.88	2.62	2.63
MOI	57.05	52.70	55.53	55.64	58.33	54.53	51.97	52.91	52.11	51.38
Clinician's gender										
Males	58.61	57.34	58.31	56.44	55.79	57.33	60.21	61.26	63.70	60.34
Females	41.39	42.66	41.69	43.56	44.21	42.67	39.79	38.74	36.30	39.66

MI-MAR: -Multiple imputation assuming Missing at Random, MI-MNAR: -Multiple imputation assuming Missing at Not Random, MO: -Medical officers

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 5.5 and Table 5.6 respectively. After MI assuming MAR mechanism, enhanced audit and feedback led to improved 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 $\geq$ 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 5.5/Table 5.6). 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 $\geq$ 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 5.5/Table 5.6). 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 $\geq$ 2.

To assess stability of parameter estimates under MI assuming MAR mechanism, we imputed missing clinicians' cadre (Table 5.5) and clinicians' gender (Table 5.6)

assuming MNAR mechanism. This 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 MI with the delta adjustment method and MI 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 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 5.5). 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 5.5). Nevertheless, the observed changes in magnitude did not change the conclusion.

Table 5.5: 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	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 (0.010, 0.076)	0.03 (0.010, 0.079)	0.03 (0.010, 0.079)	0.03 (0.010, 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 (0.270, 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)
Patient's age group: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)
patient's 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, 2.220)	1.51 (1.029, 2.220)	1.51 (1.029, 2.220)	1.50 (1.016, 2.226)
Comorbidities :1	1.59 (1.005, 2.498)	1.34 (0.910, 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 (0.930, 2.078)	1.39 (0.930, 2.078)	1.39 (0.930, 2.078)	1.35 (0.897, 2.033)
Clinician's gender: female	1.52 (1.057, 2.183)	1.37 (0.977, 1.912)	1.37 (0.981, 1.931)	1.39 (0.985, 2.110)	1.35 (0.892, 1.951)	1.37 (0.973, 1.937)
Clinician's 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, 1.020)	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 (0.610, 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 $\times$ 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)

MI-MAR: -Multiple imputation assuming Missing at Random, MI-MNAR: -Multiple imputation assuming Missing at Not Random, MO: -Medical officers, A&F: -Audit and feedback

Table 5.6: 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	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 (0.010, 0.076)	0.03 (0.010, 0.076)	0.03 (0.010, 0.077)	0.03 (0.010, 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)
Patient's age group: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)
Patient's 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, 2.220)	1.51 (1.030, 2.222)
Comorbidities :1	1.59 (1.005, 2.498)	1.34 (0.910, 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 (0.930, 2.079)	1.39 (0.930, 2.078)	1.38 (0.929, 2.076)
Clinician's 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's 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, 1.020)	0.73 (0.530, 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, 1.400)	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 (0.940, 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 $\times$ 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)

MI-MAR: -Multiple imputation assuming Missing at Random, MI-MNAR: -Multiple imputation assuming Missing at Not Random, MO: -Medical officers, A&F: -Audit and feedback

5.4 Discussion

In this study we sought to conduct sensitivity analyses to assess stability and robustness of inference under assumed MAR mechanism. Missing data occurred in patient and clinician-level covariates, as well as pneumonia care indicators used to construct PAQC score; a composite measure for quality of care.

The focus of these sensitivity analyses was clinician-level variables in pneumonia trial data set. In order to define suitable assumptions reflecting MNAR missing data mechanism in the two variables of interest, we elicited and incorporated experts' opinions into the analysis. Specifically, we interviewed 15 clinical experts in paediatrics wards in two study hospitals and incorporated their opinions into our sensitivity analyses using two approaches. In the first approach, we incorporated uncertainty about the missing data mechanism in the form of conjugate prior distributions. In the second approach, we incorporated experts' opinion in the form of shift parameters within the delta adjustment method. Although this approach is a transparent and flexible means by which to impute data under MNAR mechanisms, the choice of appropriate sensitivity parameters is less straightforward.

In this study, we utilized elicited probabilities combined with additional information probed from experts during interview sessions in the choice of sensible shift parameters. According to experts' contextual knowledge, hospitals with high workload were more likely to be teaching and referral hospitals, hence more medical officers and medical officer interns. Furthermore, experts' opinions indicated that there are more male medical officers/interns than female medical

officers/interns, compared to the observed data. Therefore, clinicians with missing information in high workload hospitals were more likely to be male medical officers/interns than female medical officers/interns. In this analysis, we implemented experts' opinion over a range of three shift parameters (i.e., δ = -0.2, -0.3 and -0.5). The shift parameters altered the probabilities with which the multilevel joint imputation model imputed missing clinicians' cadre and gender. 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) under MI assuming MNAR scenarios were close to those estimated under MAR assumption. The similarities were an indication of robust inferences under MAR assumptions. For delta adjusted over a range of parameters, 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. If 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 (White and

Carlin, 2010).

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 (Héraud-Bousquet et al., 2012; Smuk et al., 2017; Tompsett et al., 2018). In this case, we applied 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 MI with delta adjustment method while the other provides a single inference based on informative prior distributions (i.e., MI from prior distribution). Despite 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 PMM 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 analysis, we elicited experts' opinions in face-to-face interviews, which allowed us to probe for additional information and clarifications not captured in the questionnaires. In instances where face-to-face interviews are impractical, telephone discussions or electronic questionnaires can be considered (Molen-

berghs et al., 2014, Chapter 20). 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 (Leacy et al., 2017; Tompsett et al., 2018). 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 (Yuan, 2014).

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 rules (Rubin, 1976). A recent study by Tang (2017) evaluated the extent of bias associated with Rubin's variance estimator under the delta-adjusted PMM 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. The study results further showed that inference based on Rubin's rules in the delta-adjusted PMM was approximately valid. For this reason, we only reported estimates based on Rubin's rules. The alternative asymptotic sampling variance estimator suggested by Tang (2017) can be considered in future studies.

This analysis 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 sep-

arate MI-MNAR analysis for"y analysis. This because eliciting experts' opinions for the two variables jointly would have been complicated and more difficult to implement. Thirdly, although pneumonia data had clustering at hospital (n=12) and clinician level (n=378), we only accounted for clinicians' random effect in the analysis model while hospital characteristics were included as fixed effects. This was to ensure compatibility between analysis and imputation models. Moreover, statistical software used could only accommodate random-effects only at the second level of hierarchy.

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 two sensitivity analyses 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.

Chapter 6

Pairwise Joint Modelling of Clustered and High-dimensional Vector of Outcomes with Covariate Missingness

6.1 Introduction

Multiple responses reflecting different aspects of patient care is a common phenomenon in routine care studies, investigating research questions such as the level of adherence to standard quality of care guidelines by clinicians in different health care facilities. Despite measuring, for each patient, a correlated vector of response variables, inferences in most routine care studies are based on one primary outcome or multiple separate analyses (Gachau et al., 2017; Ogero

et al., 2018; Ayieko et al., 2019). Alternatively, the outcomes are combined into a single composite score (Profit et al., 2014; Opondo et al., 2016; Ogero et al., 2020), to provide global trends and insight into the quality of patient care. While these approaches are relatively straightforward, some research questions require joint modelling of all outcomes simultaneously (Molenberghs and Verbeke, 2005, Chapter25-25), for instance, when the association among outcomes and joint effects of covariates on all outcomes are of primary research interest (Fieuws and Verbeke, 2006; Fieuws et al., 2006; Verbeke et al., 2014).

In principle, a joint model links two or more models, using random-effects that capture association among outcomes of interest. Statistically, joint modelling has advantages over separate analyses of multiple outcomes. This includes efficiency gain and bias reduction, especially when data are MAR in some of the outcomes (Gueorguieva, 2001; Fieuws and Verbeke, 2004, 2006; McCulloch, 2008). In addition, joint modelling allows for different types of models for the different outcomes (e.g. linear, non-linear, and generalized linear mixed models) (Faes et al., 2008), while the interpretation of parameter estimates is the same as interpretation from the separate univariate models (Fieuws and Verbeke, 2004).

Although joint models have been extended from the common bivariate to the multivariate cases (Fieuws and Verbeke, 2006), standard fitting procedures are difficult to implement with high-dimensional outcomes (Fieuws and Verbeke, 2006; McCulloch, 2008; Jaffa et al., 2014; Hickey et al., 2016; Nassiri et al., 2017). The computational complexities stem from an increase in the number of parameters to be estimated, for every new outcome added into the joint model (Molen-

berghs and Verbeke, 2005, Chapter 25, p .470), and relatedly the increasing dimension of the random-effects vector.

To overcome these challenges, the shared random-effects model, which assumes that all outcomes share the same set of random-effects, can be considered. In this case, the dimension of the random-effects does not increase with an increase with the number of outcomes (Jaffa et al., 2014). The price to pay is a sometimes restrictive, less realistic model (McCulloch, 2008). For instance, when dealing with discrete outcomes (e.g., binomial and Poisson), that have a natural link between the mean and variance.

A plausible alternative is the pairwise joint modelling approach, which allows fitting of the correlated random-effects joint model, while circumventing the computational complexity associated with a full joint multivariate model (Fieuws and Verbeke, 2006; Fieuws et al., 2006).

As mentioned earlier, missing data in either outcomes or covariates is a common problem in routine data. Although joint modelling can be used to mitigate the effect of missing data among outcomes, appropriate strategies of handling missing covariates in high-dimensional joint modelling is hardly addressed in the literature. For instance, a previous joint modelling study reported deletion of case records with missing covariates to alleviate computational challenges (Long and Mills, 2018). The repercussion of suboptimal missing data handling techniques includes risk of biased and inefficient estimates, hence misleading inferences (Carpenter and Kenward, 2013).

In this chapter, we sought to jointly analyse 9 binary outcomes, at the same time

accounting for covariate missingness in a paediatric pneumonia trial data set. Specifically, we used MI, based on the joint modelling framework, to address missing covariates across two levels of the hierarchy. Thereafter, we used the pairwise approach to estimate the joint effects of covariates on outcomes. This was in addition to estimating the strength of association among pneumonia outcomes.

The remainder of this chapter is organized as follows. Section 6.2 presents methods on joint modelling approach using mixed models and the pairwise fitting approach followed by application to pneumonia trial data. Results are presented in Section 6.3 and we conclude with a discussion in Section 6.4.

6.2 Methods

6.2.1 Correlated Random-effects Joint Model

Let Y_{rij} denote the r^{th} $(r = 1, 2, \ldots, p)$ outcome for the i^{th} $(i = 1, 2, \ldots, N)$ subject in cluster j $(j = 1, 2, \ldots, n_i)$. The corresponding univariate random-effects model for the r^{th} outcome can be defined as

$$h^{-1}\left(E(Y_{rij}|b_{ri}, \boldsymbol{X}_{rij}, \boldsymbol{Z}_{rij})\right) = \boldsymbol{X}'_{rij}\beta_r + \boldsymbol{Z}'_{rij}b_{ri}, \tag{6.1}$$

where $h^{-1}(\cdot)$ is an appropriate link function depending on the of type variable (i.e. continuous, binary, count, etc.) (Fitzmaurice et al., 2009a), $\boldsymbol{X}_{rij}$ is a vector of known covariates with fixed effects β_r and $\boldsymbol{Z}_{rij}$ is a vector of covariates with random-effects b_{ri}. The univariate random-effects model can be extended to jointly model all the outcomes simultaneously, by imposing a joint multivari-

ate distribution on the random-effects (Gueorguieva and Agresti, 2001; Fieuws and Verbeke, 2006; Faes et al., 2008). Moreover, the number of random-effects can vary among the outcomes of interest. Conditional on the vector of random-effects (b_{ri}), the outcomes are assumed to be independent (Molenberghs and Verbeke, 2005) and the corresponding log-likelihood contribution for subject i equals

$$l_i(y_{1i}, y_{2i}, \ldots, y_{pi}|\boldsymbol{\Theta}^*) = ln \int \prod_{r=1}^{p} f_{ri}(y_{ri}|b_{ri}, \boldsymbol{\theta}_r) f(b_{ri}|\boldsymbol{D}) db_{ri}. \tag{6.2}$$

The vector Θ^* contains all parameters of the full joint model (i.e. fixed parameters denoted by $\boldsymbol{\beta}^*$ and covariance parameters denoted by Σ^*), while $f_{ri}(y_{ri}|b_{ri}, \boldsymbol{\Theta}_r)$ is the density of y_{ri} conditional on the random- effects for the r^{th} outcome on subject i. The vector of random-effects b_i is assumed to follow a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{D}$ (Molenberghs and Verbeke, 2005), that is,

$$b_i = \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{pi} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{D}_{11} & \boldsymbol{D}_{12} & \ldots & \boldsymbol{D}_{1p} \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} & \ldots & \boldsymbol{D}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{D}_{p1} & \boldsymbol{D}_{p2} & \ldots & \boldsymbol{D}_{pp} \end{pmatrix} \right].$$

The elements $\boldsymbol{D}_{rs}$ in $\boldsymbol{D}$ correspond to blocks of random-effects variance-covariance between the r^{th} and the s^{th} outcomes $(r, s = 1, 2, \ldots, p)$. For example, if each outcome has a random intercept (b_0) and a random slope (b_1), then $\boldsymbol{D}_{rs}$ is given by

$$\boldsymbol{D}_{rs} = \begin{bmatrix} \sigma^2_{b_{0r}} & \sigma_{b_{0r}b_{1r}} & \sigma_{b_{0r}b_{0s}} & \sigma_{b_{0r}b_{1s}} \\ & \sigma^2_{b_{1r}} & \sigma_{b_{1r}b_{0s}} & \sigma_{b_{1r}b_{1s}} \\ & & \sigma^2_{b_{0s}} & \sigma_{b_{0s}b_{1s}} \\ & & & \sigma^2_{b_{1s}} \end{bmatrix}.$$

The elements of the variance covariance matrix $\boldsymbol{D}$ can be used to measure the strength of association between any two outcomes of interest. As mentioned earlier, the dimension of the random-effects vector b_i in the full joint model, increases with an in increase in the number of outcomes. This leads to computational challenges for high dimensional vector of outcomes (Molenberghs and Verbeke, 2005; Fieuws and Verbeke, 2006).

6.2.2 Pairwise Modelling Approach

In light of computational challenges highlighted above, Fieuws and Verbeke (2006) proposed a pairwise approach within the pseudo-likelihood framework to handle a high-dimensional vector of outcomes. With a vector of p outcomes, the pairwise approach maximizes the likelihood for all $Q = p(p-1)/2$ pairwise models separately, instead of maximizing the full joint multivariate likelihood (Fieuws and Verbeke, 2006; Fieuws et al., 2006; Kundu, 2011).

Precisely, this produces a so-called pseudo-likelihood $(pl)$of the following form:

$$pl(\boldsymbol{\Theta}) = l(Y_1Y_2|\boldsymbol{\Theta}_{1,2})l(Y_1Y_3|\boldsymbol{\Theta}_{1,3}),\ldots,l(Y_{p-1}Y_p|\boldsymbol{\Theta}_{p-1,p}) = \prod_{r=1}^{p-1}\prod_{s=1}^{p} l(Y_rY_s|\boldsymbol{\Theta}_{r,s}). \tag{6.3}$$

For a given pair of responses $(r,s = 1,2,\ldots,p)$, $l(Y_r,Y_s|\boldsymbol{\Theta}_{rs})$ denotes the likelihood, while $\boldsymbol{\Theta}_{rs}$ is the vector of all parameters encountered in a pairwise joint model (Fieuws and Verbeke, 2006; Kundu, 2011). The corresponding pseudo-log likelihood (pll) function is given by

$$\begin{aligned} pll(\Theta) &= \sum_{r=1}^{p-1}\sum_{s=r+1}^{p} ll(Y_rY_s|\boldsymbol{\Theta}_{rs}) \\ &= \sum_{q=1}^{Q} ll(Y_q|\boldsymbol{\Theta}_q), \end{aligned} \tag{6.4}$$

where $\boldsymbol{Y}_q$ and $\boldsymbol{\Theta}_q$ contains all the observations and parameters, respectively, in the q^{th} response pair $(q = 1,2,\ldots,Q)$. All Q pair-specific parameter vectors $\boldsymbol{\Theta}_q$ $(q = 1,2,\ldots,Q)$ are stacked together into $\boldsymbol{\Theta}$. It is clear that if $\hat{\boldsymbol{\Theta}}_q$ maximizes $l(Y_q|\boldsymbol{\Theta}_q)$, then $\hat{\boldsymbol{\Theta}}$, the stacked vector combining all $\boldsymbol{\Theta}_q$, maximizes $pll(\boldsymbol{\Theta})$ (Kundu, 2011). The asymptotic distribution of $\hat{\boldsymbol{\Theta}}$ is multivariate normal given by

$$\sqrt{N}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) \sim MVN(\mathbf{0}, \boldsymbol{H^{-1}GH^{-1}}), \tag{6.5}$$

where $\boldsymbol{H^{-1}GH^{-1}}$ is a sandwich estimator and $\boldsymbol{H}$ and $\boldsymbol{G}$ are based on cluster-wise Hessians and gradients of the log-pseudo-likelihood function, respectively (Fieuws et al., 2006; Fieuws and Verbeke, 2006; Kundu, 2011). The vector of all

parameters in the full joint model ($\boldsymbol{\Theta}^*$) and stacked vector from pairwise models ($\boldsymbol{\Theta}$) are not equivalent. Specifically, some parameters in $\boldsymbol{\Theta}^*$ have a single counterpart in $\boldsymbol{\Theta}$, while other elements in $\boldsymbol{\Theta}^*$ have multiple counterparts in $\boldsymbol{\Theta}$. A set of fixed effects ($\boldsymbol{\beta}^*$), for the full joint model, are obtained by averaging duplicate parameter estimates from the pairwise joint models (Fieuws and Verbeke, 2006). This can be achieved by multiplying the stacked vector of regression parameters ($\boldsymbol{\beta}$) with an appropriate weight matrix $\boldsymbol{A}$ as below

$$\boldsymbol{\beta}^* = \boldsymbol{A}\boldsymbol{\beta}. \tag{6.6}$$

The standard errors follow as the square root of diagonal elements of variance-covariance estimator:

$$\boldsymbol{\Sigma}^* = \boldsymbol{A}(\boldsymbol{H}^{-1}\boldsymbol{G}\boldsymbol{H}^{-1})\boldsymbol{A}^T. \tag{6.7}$$

Further details on estimation of fixed effects and corresponding standard errors are presented in the application section.

6.2.3 Pneumonia Trial Data

We analyzed routine paediatric data introduced in Chapter 1, Section 1.2. There were 12 pneumonia care indicators of interest in the trial data (i.e., nine signs and symptoms in the assessment domain, 1 indicator in the diagnosis and classification domain and two indicators in the treatment domain (Table 1.2)). However, 3 signs and symptoms, namely central cyanosis, grunting and ability to drink

were documented for all the case records. Therefore, we considered 9 binary outcomes in a subsequent analysis. Table 6.1 describes how we created each binary outcome. The value one for the binary indicators represents documentation/adherence to recommended paediatric pneumonia guidelines, while zero represents lack of documentation or documentation of inappropriate care.

Table 6.1: Quality of care indicators among children admitted with pneumonia during the trial period.

Quality of care domain	Indicator	Scores in binary indicators
1. Assessment		
Primary signs and symptoms	Cough	1: if cough is documented, 0: if it is not documented.
	Difficult breathing	1: if difficult breathing is documented, 0: if it is not documented.
Secondary sign and symptoms	Respiratory rate	1: if respiratory rate is documented, 0: if it is not documented.
	Oxygen saturation	1: if oxygen saturation is documented, 0: if it is not documented.
	AVPU[a]	1: if AVPU is documented, 0: if it is not documented.
	Lower chest wall indrawing	1: if indrawing is documented, 0: if it is not documented.
2. Diagnosis and classification	Correct diagnosis	1: if the admitting clinician documented pneumonia as the clinical diagnosis
		0: if documented clinical diagnosis is severe pneumonia or missing classification.
3. Treatment	Correct prescription	1: if oral amoxicillin was prescribed and documented in the medical record. 0: if amoxicillin was not prescribed
	Correct oral amoxicillin dose	1: if oral amoxicillin was prescribed in correct dose and frequencies, i.e., 32-48 international units/Kilogram (IU/Kg) every 12 hours.
		0: if oral amoxicillin prescription is an under dose ($<$32 IU/Kg) or over dose ($>$48 IU/Kg) or missing amoxicillin dose or wrong frequency or missing frequency or missing patient's weight.

AVPU[a] :-Alert, Verbal response, Pain response, Unresponsive, $*$Pneumonia diagnosis for patients with history of cough and/or difficult breathing (primary signs) in combination with signs of lower chest wall indrawing and/or respiratory rate (RR) $\geq$50 ($\geq$40) for patients aged 2-11 (12-59 months), in the absence of danger any sign (inability to drink/breastfeed, cyanosis, grunting or oxygen saturation $<$ 90% or AVPU= 'V', 'P' or 'U').

Covariates

The covariates of interest are as described in Chapter 3, Section 3.2.1. Missingness occurred in patients' gender at 7% while 21.7% and 21.9% of the clinicians had missing cadre and gender, respectively. Before fitting a pairwise model, we imputed partially observed covariates using the latent normal joint modelling MI approach described in Chapter 3, Section 3.2.3.

Application: Model Fitting and Inference

We applied the pairwise approach to jointly model the probability of documentation among 9 pneumonia outcomes (1=cough, 2=difficult breathing, 3=respiratory rate, 4=oxygen saturation, 5=AVPU, 6=lower chest wall indrawing, 7=correct pneumonia diagnosis, 8=correct pneumonia treatment, and 9=correct treatment dose).

For each imputed data set we fitted 36 pairwise models defined by

$$
\begin{aligned}
&logit[P(Y_{PAQCScore;rijl}=1)] = \beta_{r0} + \beta_1 X_{agegroup;rijl} + \beta_2 X_{patientgender;rijl} + \\
&\beta_3 X_{comorbidity=0;rijl} + \beta_4 X_{comorbidity=1;rijl} + \beta_5 X_{comorbidity=2;rijl} + \\
&\beta_6 X_{cliniciancadre;rjl} + \beta_7 X_{cliniciangender;rjl} + \beta_8 X_{admissionworkload;rl} + \\
&\beta_9 X_{malariaprevalence;rl} + \beta_{10} X_{timeinmonths;rl} + \beta_{11} X_{trialarm;rl} + \\
&beta_{12} X_{timeinmonths;rl} * X_{trialarm;rl} + b_{rjl} \\
&logit[P(Y_{PAQCScore;sijl}=1)] = \beta_{s0} + \beta_1 X_{agegroup;sijl} + \beta_2 X_{patientgender;sijl} + \\
&\beta_3 X_{comorbidity=0;sijl} + \beta_4 X_{comorbidity=1;sijl} + \beta_5 X_{comorbidity=2;sijl} + \\
&\beta_6 X_{cliniciancadre;sjl} + \beta_7 X_{cliniciangender;sjl} + \beta_8 X_{admissionworkload;sl} + \\
&\beta_9 X_{malariaprevalence;sl} + \beta_{10} X_{timeinmonths;sl} + \beta_{11} X_{trialarm;sl} + \\
&beta_{12} X_{timeinmonths;sl} * X_{trialarm;sl} + b_{sjl}
\end{aligned}
\tag{6.8}
$$

where Y_{rijl} and Y_{sijl} denote the r^{th} and the s^{th} outcomes, $r \neq s$ for $(r, s = 1, 2, \ldots, 9)$ for patient i admitted by clinician j in hospital l. Each outcome occurred in 8 specific pairs and we included a random clinician intercept in each model. Due to relatively low numbers of clinical and medical officers, we grouped clinicians into two cadres from the initial four. That is, clinical officers (CO) combine clinical officers and clinical officer interns and medical officers (MO) combine medical officers and medical officer interns, respectively. We fitted all pairwise joint models using the JMbayes package (Rizopoulos et al., 2020) in R version 3.5.4. Compete case analysis was also conducted using a SAS macro provided by Fieuws et al.

(2006) for verification purposes.

Under complete case analysis, regression estimates and standard errors were averaged across 36 pairwise models using the pseudo-likelihood approach presented in subsection 6.2.2. Likewise, regression parameters were averaged across the various pairwise models for each imputed data set. Variance-covariance estimators (6.7) were also obtained for each imputed data set. This step resulted in 20 sets of averaged regression parameters and variance-covariance estimators respectively. Thereafter, Rubin's rules (Rubin, 1976) were applied to obtain final estimates while accounting for within and between imputation variability. More details on the two-step procedure are as follows. Each bivariate model in the m^{th} imputed data set had a vector of 26 regression coefficients (i.e., 13 regression coefficients for each outcome) denoted by $\hat{\boldsymbol{\beta}}_{qm}$, $q = 1, 2, \ldots, 36$, $m = 1, 2, \ldots, 20$. We stacked the 36 pairwise parameter estimate vectors resulting into a column vector with 936 rows, that is,

$$\hat{\boldsymbol{\beta}}_m = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{1m} \\ \hat{\boldsymbol{\beta}}_{2m} \\ \vdots \\ \hat{\boldsymbol{\beta}}_{36m} \end{bmatrix}_{936\times 1}, \quad m = 1, 2, \ldots, 20.$$

Any two pairwise joint models with a common outcome (e.g., $l(Y_r, Y_s)$ and $l(Y_r, Y_{s'})$ $s \neq s'$), shared the parameters for the r^{th} outcome (Fieuws and Verbeke, 2006; Fieuws et al., 2007; Kundu, 2011). To account for duplicate parameter estimates,

we pre-multiplied $\boldsymbol{\beta}_m$ with an appropriate weight matrix $\boldsymbol{A}$ as follows,

$$\hat{\boldsymbol{\beta}}^*_m = \boldsymbol{A}\hat{\boldsymbol{\beta}}_m \quad \text{m=1,2,...,20.} \tag{6.9}$$

The weight matrix $\boldsymbol{A}$ had 117 rows and 936 columns, and it was constructed such that, it averaged all duplicate parameter estimates of an outcome across the 8 pairwise models in which it occurred. The resulting vector, $\hat{\boldsymbol{\beta}}^*_m$ was a stacked column vector of regression parameters for all nine outcomes. Each outcome had 13 regression parameters denoted by $\hat{\boldsymbol{\beta}}^*_{mr}$.

Inference for Standard Errors

The corresponding standard errors were obtained using the pseudo-likelihood approach introduced above. For each bivariate pair, $\boldsymbol{Y}_{mq}$, $q = 1, 2, \ldots, 36$, in the m^{th} imputed data set, we estimated the variance-covariance matrix, $\boldsymbol{H^{-1}GH^{-1}}$. Since $\boldsymbol{H}$ and $\boldsymbol{G}$ depend on the unknown parameters in $\boldsymbol{\Theta}$ (Molenberghs and Verbeke, 2005; Kundu, 2011), estimation proceeded as follows.

1. We obtained $\hat{\boldsymbol{J}}_{mq}$ and $\hat{\boldsymbol{K}}_{mq}$ for each pairwise model using

$$\hat{\boldsymbol{J}}_{mq} = \sum_{i=1}^{N} X^T_{imq}\hat{T}_{imq}X_{imq}$$

and

$$\hat{\boldsymbol{K}}_{mq} = \left[X^T_{1mq}\hat{T}_{1mq}, X^T_{2mq}\hat{T}_{2mq}, \ldots, X^T_{Nmq}\hat{T}_{Nmq}\right],$$

where X_{imq} corresponds to the i^{th} subject's contribution in the design matrix for the fixed effects, $T_{imq} = (Z_{imq}\hat{D}_{mq}Z^T_{imq})^{-1}$ where Z_{imq} is the i^{th} subject's

contribution in the design matrix for random-effects (Kundu, 2011) and D_{mq} is the variance-covariance matrix for the random-effects for the q^{th} pair in the m^{th} imputed data set. N indicates the number of subjects.

2. We combined $\hat{\boldsymbol{J}}_{mq}$ and $\hat{\boldsymbol{K}}_{mq}$ estimated across all the 36 pairs, (i.e., ($\hat{\boldsymbol{J}}_{m1}$,$\hat{\boldsymbol{K}}_{m1}$),($\hat{\boldsymbol{J}}_{m2}$,$\hat{\boldsymbol{K}}_{m2}$), ..., ($\hat{\boldsymbol{J}}_{m36}$,$\hat{\boldsymbol{K}}_{m36}$)) as follows:

$$\hat{\boldsymbol{J}}_m = \begin{bmatrix} \hat{\boldsymbol{J}}_{m1} & 0 & \dots & 0 \\ 0 & \hat{\boldsymbol{J}}_{m2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \hat{\boldsymbol{J}}_{m36} \end{bmatrix}_{936\times 936} \quad \text{and} \quad \hat{\boldsymbol{K}}_m = \begin{bmatrix} \hat{\boldsymbol{K}}_{m1} \\ \hat{\boldsymbol{K}}_{m2} \\ \vdots \\ \hat{\boldsymbol{K}}_{m36} \end{bmatrix}_{936\times N.}$$

3. We estimated $\boldsymbol{H}_m$ and $\boldsymbol{G}_m$ as follows

$$\hat{\boldsymbol{H}}_m = \frac{1}{N}\hat{\boldsymbol{J}}_m \quad \text{and} \quad \hat{\boldsymbol{G}}_m = \frac{1}{N}\hat{\boldsymbol{K}}_m\hat{\boldsymbol{K}}_m^T,$$

where N is the number of subjects.

4. We obtained a variance-covariance matrix, $\hat{\boldsymbol{\Sigma}}_m^*$ for each imputed data set using

$$\hat{\boldsymbol{\Sigma}}_m^* = \boldsymbol{A}\boldsymbol{\Omega}_m\boldsymbol{A}^T, \quad m = 1, 2, \dots, 20, \tag{6.10}$$

where $\hat{\boldsymbol{\Omega}}_m = \hat{\boldsymbol{H}}_m^{-1}\hat{\boldsymbol{G}}_m\hat{\boldsymbol{H}}_m^{-1}$ and $\boldsymbol{A}$ is the weight matrix defined above. Each $\hat{\boldsymbol{\Sigma}}_m^*$ was a 117×117 covariance matrix and the diagonal elements corresponded to variances of fixed regression parameters in $\hat{\boldsymbol{\beta}}_m^*$.

Pooling Final Estimates

In the final step, we pooled the final estimates using Rubin's rules (Rubin, 1976) for each of the nine outcomes. This was based on the set of pairwise regression parameters and the estimated variance covariance matrices $\hat{\boldsymbol{\Sigma}}_m^*$. The pooled MI estimator for β is given by

$$\bar{\beta}_r^* = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}_{mr}^*, \tag{6.11}$$

with variance estimator

$$V_r = W_r + \left(\frac{M+1}{M}\right) \times B_r,$$

where

$$W_r = \frac{1}{M}\sum_{m=1}^{M}\hat{\sigma}_{mr}^2$$

is the average imputation variance, $\hat{\sigma}_{mr}^2$ are the diagonal elements of $\hat{\Sigma}_m^*$ and

$$B_r = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\beta}_{mr}^* - \bar{\beta}_r^*)^2$$

is the between imputation variance. Final MI estimates were compared to those obtained under complete case analysis.

6.2.4 Statistical Tests for Multiple Parameters

To test for the joint effect of covariates on the outcomes, we used the Wald-type test. The full (saturated) model contained all the covariates while the reduced

(null) models dropped one covariate at a time. The general linear hypothesis under complete case analysis corresponded to

$$H_0 : \boldsymbol{L\beta} = 0 \text{ vs } H_A : \boldsymbol{L\beta} \neq 0 \tag{6.12}$$

where L is an appropriate matrix (Molenberghs and Verbeke, 2005; Fieuws and Verbeke, 2006). For example, to test for the joint interaction effect between intervention arm and follow-up time, the null hypothesis corresponded to

$$H_0 : \beta_{1,13} = \beta_{2,13} = \beta_{3,13} = \beta_{4,13} = \beta_{5,13} = \beta_{6,13} = \beta_{7,13} = \beta_{8,13} = \beta_{9,13} = 0.$$

We compared the test statistic with a chi-square with nine degrees of freedom. A 5% level of significance was considered.

6.2.5 Association Among Pneumonia Outcomes

The strength of association among documentation of pneumonia care indicators was evaluated using the covariances of the random-effects. Since the covariance matrix $\boldsymbol{D}$ was not estimated directly at analysis stage, we constructed it using blocks of random-effects variance-covariance matrices estimated in the pairwise joint models. Under MI, we first averaged duplicate variances across 36 pairwise random intercept models for each of the 20 imputed data set.

Specifically, for each imputed data set, we extracted the random intercepts variance-covariance matrix for all 36 pairwise joint models, that is,

$$\boldsymbol{D}_{m1} = \begin{bmatrix} \sigma^2_{b_{m01}} & \sigma_{b_{m01}b_{m02}} \\ & \sigma^2_{b_{m02}} \end{bmatrix}, \boldsymbol{D}_{m2} = \begin{bmatrix} \sigma^2_{b_{m01}} & \sigma_{b_{m01}b_{m03}} \\ & \sigma^2_{b_{m03}} \end{bmatrix}, \ldots, \boldsymbol{D}_{m36} = \begin{bmatrix} \sigma^2_{b_{m08}} & \sigma_{b_{m08}b_{m09}} \\ & \sigma^2_{b_{m09}} \end{bmatrix}.$$

We then created an overall variance-covariance matrix $\boldsymbol{D}_m$ for each imputed data set accounting for overlapping information. For example, in each imputed data set, $(m = 1, 2, \ldots, 20)$, documentation of cough occurred in the variance-covariance matrices of the first 8 pairs, that is,

$$\boldsymbol{D}_{m1} = \begin{bmatrix} \sigma^2_{b_{m01}} & \sigma_{b_{m01}b_{m02}} \\ & \sigma^2_{b_{m02}} \end{bmatrix}, \boldsymbol{D}_{m2} = \begin{bmatrix} \sigma^2_{b_{m01}} & \sigma_{b_{m01}b_{m03}} \\ & \sigma^2_{b_{m03}} \end{bmatrix}, \ldots, \boldsymbol{D}_{m8} = \begin{bmatrix} \sigma^2_{b_{m01}} & \sigma_{b_{m01}b_{m09}} \\ & \sigma^2_{b_{m09}} \end{bmatrix}.$$

We extracted the random intercept variances of each outcome from the pairs it occurred in and averaged them into a single random intercept variance estimate of Y_r (e.g., $\sigma^2_{b_{01}}$ denoting the random intercept variance for cough). On the other hand, unique off-diagonal elements corresponding to covariance between any two outcomes were also mapped into $\boldsymbol{D}_m$. Thereafter, we averaged all the 20 $\boldsymbol{D}_m$ matrices, $m = 1, 2, \ldots, 20$ to obtain the overall 9×9 variance covariance matrix $\boldsymbol{D}$ for all the nine outcomes. We used the same procedure to construct the random-intercepts variance-covariance under complete case analysis where we averaged duplicate variances across 36 pairwise random intercept models. The strength of

association between any two outcomes, say Y_r and Y_s, was calculated using

$$Corr(b_{0r}, b_{0s}) = \frac{Cov(b_{0r}b_{0s})}{\sqrt{Var(b_{0r}) \times Var(b_{0s})}} = \frac{\sigma_{b_{0r}b_{0s}}}{\sqrt{\sigma^2_{b_{0r}} \times \sigma^2_{b_{0s}}}}. \tag{6.13}$$

We performed principal component analysis (PCA) on random clinicians' intercepts variance-covariance matrices obtained under complete case analysis and after MI in order to visualize the factor loadings amongst pneumonia outcomes of interest.

6.3 Results

Regarding to outcomes of interest, the level of documentation and adherence to recommended pneumonia care varied within and across domains of care. Most of the signs and symptoms in the assessment domain were well documented except for oxygen saturation and respiratory rate, which had documentation rates of 60.9% (1297/2127) and 88.8% (1889/2127), respectively. On the other hand, we observed poorer performance in diagnosis and classification domain as well as treatment domain. Specifically, of all 2127 syndromic pneumonia cases, only 1473 (69.3%) had correct clinical diagnosis documented in the medical record. In the treatment domain, about 48.7% (1036/2127) were prescribed with oral amoxicillin as per the guidelines. However, only 25% (523/2127) of all pneumonia patients got the right oral amoxicillin dosage, that is, 32-48 international units/Kilogram (IU/Kg) every 12 hours.

Wald-type Tests for Joint Covariates Effect

Under compete case analysis, Wald test results revealed a significant joint interaction effect between intervention arm and follow-up time on documentation of 9 paediatric pneumonia quality of care indicators (Table 6.2). At hospital level, paediatric admission workload and malaria prevalence status exhibited significant joint effects on documentation of all the 9 paediatric pneumonia care outcomes (Table 6.2). At patients' level, age and comorbidity showed significant joint effect on documentation of paediatric pneumonia care outcomes. Likewise, clinicians' gender and cadre had significant joint effect on documentation and adherence to recommended paediatric pneumonia care guidelines (Table 6.2).

Table 6.2: Wald test for the joint effect of covariates on 9 pneumonia outcomes under complete case analysis.

Effect	Test statistics	P-Value
Patient's age	19.62	0.02
Patient's gender	12.20	0.21
Comorbidity	20.54	0.01
Clinician's gender	20.91	0.01
Clinician's cadre	19.94	0.02
Admission workload	25.56	0.002
Malaria prevalence	17.89	0.04
Follow-up time (months)	19.26	0.02
Enhanced A&F [a] arm	17.98	0.04
Enhanced A&F arm $\times$ follow-up time	18.13	0.03

Table 6.3 and 6.4 present the odds ratios and the corresponding 95% confidence intervals for the nine paediatric pneumonia outcomes under complete case analysis and after MI of missing covariates. In general, the magnitude and direction of covariate effects varied among outcomes. Holding other covariates at reference levels, enhanced audit and feedback improved adherence to recommended guidelines in 6 out of 9 paediatric pneumonia care indicators over time. That is, for a unit increase in follow-up time, the change in the odds of oxygen saturation, respiratory rate, lower wall indrawing documentation (in the assessment domain), correct pneumonia diagnosis, oral amoxicillin prescription and correct dosage among patients admitted in an intervention hospital (enhanced audit and feedback (A&F) arm) was significantly more positive in comparison to the change among patients admitted in a control hospital. These observations were made under complete case analysis (Table 6.3) and after MI of missing covariates (Table 6.4). However, the 95% confidence intervals estimated after multiple imputation were consistently narrower compared to those estimated under complete case analysis. On the other hand, there was no significant difference between enhanced A&F and standard A&F on the documentation of cough, difficult breathing and AVPU over time.

Although results were largely consistent before and after MI, we observed a few instances of contrasting results. For example, under complete case analysis, the odds of cough documentation were significantly lower among patients admitted in hospitals with low paediatric admission workload (Table 6.3). After MI, however, the difference was no longer significant (Table 6.4). Similarly, the odds

of documentation of difficult breathing and lower wall chest indrawing among patients in low malaria prevalence hospital were lower compared to the odds of patients in high malaria hospital (Table 6.3). However, after MI, the difference was no longer statistically significant (Table 6.4).

Table 6.3: Odds ratios and 95% confidence intervals for 9 pneumonia care indicators under complete case analysis.

	Cough	Difficult breathing	Respiratory rate	Oxygen saturation	AVPU	Indrawing	Correct diagnosis	Correct treatment	Correct dose
Effect	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Intercept	8.45 (6.72, 10.18)	4.76 (4.17, 5.35)	5.75 (5.56, 5.94)	3.40 (3.23, 3.57)	3.94 (3.64, 4.24)	1.01 (0.93, 1.09)	2.31 (2.02, 2.61)	0.60 (0.5, 0.71)	0.14 (0.11, 0.17)
Patients'age group:12-59 months	2.16 (1.09, 3.23)	1.35 (1.11, 1.59)	0.91 (0.85, 0.97)	0.99 (0.92, 1.06)	1.09 (0.98, 1.21)	1.92 (1.67, 2.17)	1.33 (1.12, 1.54)	1.13 (0.92, 1.34)	1.18 (0.94, 1.42)
Patient's gender: males	1.02 (0.89, 1.15)	0.94 (0.83, 1.05)	0.89 (0.79, 0.99)	0.89 (0.83, 0.96)	0.86 (0.76, 0.97)	2.44 (2.31, 2.57)	0.89 (0.86, 0.92)	0.98 (0.73, 1.23)	0.96 (0.51, 1.41)
Comorbidities: 0	6.05 (5.08, 7.02)	2.76 (2.32, 3.21)	0.82 (0.69, 0.95)	1.34 (1.13, 1.55)	1.49 (1.26, 1.72)	1.78 (1.37, 2.19)	0.87 (0.73, 1.01)	2.09 (1.76, 2.42)	2.12 (1.48, 2.76)
Comorbidities :1	2.92 (1.90, 3.94)	3.40 (2.16, 4.64)	0.83 (0.75, 0.92)	1.30 (1.19, 1.41)	1.66 (1.53, 1.79)	1.24 (1.16, 1.32)	1.06 (0.97, 1.15)	2.13 (1.93, 2.33)	2.28 (2.04, 2.52)
Comorbidities :2	2.71 (1.35, 4.07)	3.37 (2.48, 4.26)	0.96 (0.86, 1.06)	1.22 (1.10, 1.34)	4.20 (3.82, 4.58)	1.12 (1.03,1.21)	1.08 (0.97, 1.19)	1.83 (1.62, 2.04)	1.98 (1.73, 2.23)
Clinician's gender: female	0.92 (0.73, 1.11)	1.25 (1.02, 1.48)	1.08 (0.91, 1.25)	2.24 (1.96, 2.52)	1.45 (1.31, 1.60)	1.23 (1.10, 1.36)	1.12 (1.07, 1.17)	1.24 (1.01, 1.47)	1.29 (1.16, 1.42)
Clinician's cadre: MO[b]	2.11 (2.01, 2.21)	0.52 (0.19, 0.84)	1.64 (1.55, 1.73)	1.45 (1.37, 1.53)	1.13 (1.07, 1.20)	1.83 (1.60, 2.06)	0.94 (0.73, 1.17)	0.91 (0.80, 1.03)	0.98 (0.86, 1.10)
Hospital workload: low	0.57 (0.37, 0.79)	2.54 (1.72, 3.36)	1.01 (0.71, 1.31)	0.64 (0.47, 0.81)	0.87 (0.67, 1.07)	0.25 (0.20, 0.31)	1.17 (1.01, 1.33)	1.40 (1.22, 1.58)	0.44 (0.24, 0.64)
Malaria prevalence: low	0.15 (0.12, 0.19)	0.7 (0.55, 0.85)	3.61 (2.72, 4.50)	0.22 (0.19, 0.25)	0.8 (0.69, 0.91)	0.89 (0.79, 0.99)	1.19 (1.03, 1.36)	0.85 (0.78, 0.92)	0.81 (0.75, 0.87)
Time (months)	1.24 (1.04, 1.44)	1.17 (1.02, 1.32)	1.06 (0.94, 1.18)	1.27 (1.14, 1.41)	5.10 (4.64, 5.56)	1.52 (1.27, 1.77)	1.02 (0.95, 1.09)	0.93 (0.88, 0.98)	1.04 (0.99, 1.09)
Enhanced A&Fc arm	1.01 (0.92, 1.11)	0.86 (0.67, 1.05)	0.85 (0.64, 1.06)	0.16 (0.12, 0.20)	0.79 (0.63, 0.95)	0.89 (0.70, 1.09)	0.79 (0.65, 0.93)	0.30 (0.26, 0.35)	0.94 (0.82, 1.06)
Time× Enhanced A&F arm	1.04 (0.85, 1.23)	0.96 (0.83, 1.09)	1.13 (1.03, 1.23)	1.10 (1.01, 1.20)	1.17 (0.99, 1.35)	1.24 (1.08, 1.40)	1.36 (1.18, 1.54)	1.34 (1.17, 1.51)	1.15 (1.01, 1.29)

AVPU[a] :-Alert, Verbal response, Pain response, Unresponsive, MO:- Medical officers, A&F:-Audit and feedback

Table 6.4: Odds ratios and 95% confidence intervals for 9 pneumonia care indicators after multiple imputation.

	Cough	Difficult breathing	Respiratory rate	Oxygen saturation	AVPU	Indrawing	Correct diagnosis	Correct treatment	Correct dose
Effect	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Intercept	7.01 (4.74, 9.24)	5.1 (4.69, 5.51)	6.76 (5.31, 8.21)	3.43 (3.32, 3.54)	3.95 (3.67, 4.23)	1.04 (0.94, 1.14)	2.33 (2.05, 2.61)	0.68 (0.53, 0.83)	0.23 (0.07, 0.39)
Patient's age group: 12-59 months	1.86 (1.11, 2.61)	1.37 (1.17, 1.57)	0.95 (0.82, 1.08)	1.02 (0.95, 1.10)	1.11 (0.94, 1.28)	1.89 (1.71, 2.07)	1.30 (1.10, 1.51)	1.15 (0.80, 1.50)	1.13 (0.75, 1.51)
Patient's gender: males	1.04 (0.95, 1.13)	0.96 (0.89, 1.03)	0.90 (0.85, 0.95)	0.90 (0.84, 0.96)	0.87 (0.84, 0.90)	2.45 (2.33, 2.57)	0.91 (0.80, 1.01)	0.98 (0.77, 1.19)	0.99 (0.61, 1.37)
Comorbidities: 0	6.17 (5.76, 6.58)	2.65 (2.42, 2.88)	0.83 (0.76, 0.95)	1.40 (1.26, 1.55)	1.35 (1.26, 1.44)	1.76 (1.38, 2.14)	0.97 (0.66, 1.28)	2.15 (1.81, 2.49)	1.96 (1.23, 2.69)
Comorbidities :1	2.34 (2.15, 2.53)	3.51 (2.43, 4.59)	0.81 (0.73, 0.89)	1.25 (1.11, 1.39)	1.68 (1.56, 1.80)	1.25(1.18,1.32)	1.03(0.95, 1.12)	1.89(1.75, 2.04)	1.90 (1.78, 2.02)
Comorbidities :2	2.22 (1.75, 2.69)	3.21 (2.55, 3.87)	0.90 (0.72, 1.08)	1.24 (1.14, 1.34)	4.55 (3.83, 5.27)	1.14 (1.08,1.20)	1.10 (0.96, 1.24)	1.66 (1.36, 1.96)	1.54 (1.29, 1.79)
Clinician's gender: female	1.18 (1.02,1.34)	0.99 (0.48, 1.50)	1.05 (0.96, 1.14)	2.17 (1.90, 2.44)	1.47 (1.39, 1.55)	1.24 (1.13, 1.35)	1.17 (1.07, 1.27)	1.14 (1.05, 1.23)	1.32 (1.15, 1.49)
Clinician's cadre: MO[b]	2.18 (2.04, 2.32)	0.65 (0.25, 1.05)	1.58 (1.38, 1.78)	1.47 (1.31, 1.63)	1.17 (1.09, 1.25)	1.83 (1.62, 2.04)	0.96 (0.84, 1.08)	0.95 (0.78, 1.01)	0.91 (0.79,1.04)
Hospital workload: low	0.81 (0.43, 1.19)	2.40 (1.72, 3.08)	0.93 (0.76, 1.1)	0.57 (0.48, 0.68)	0.79 (0.69, 0.89)	0.24 (0.18, 0.30)	1.15 (1.04, 1.26)	1.30 (1.25, 1.35)	0.48 (0.28, 0.68)
Malaria prevalence: low	0.15 (0.07, 0.23)	0.73 (0.34, 1.12)	3.62 (2.81, 4.43)	0.26 (0.23, 0.30)	0.80 (0.68, 0.92)	0.89 (0.78, 1.01)	1.18 (1.05, 1.31)	0.64 (0.31, 0.97)	0.78 (0.58, 0.98)
Time (months)	1.22 (1.09, 1.35)	1.21 (1.07, 1.35)	1.04 (0.95, 1.13)	1.23 (1.15, 1.32)	4.99 (4.71, 5.27)	1.53 (1.28, 1.78)	0.99 (0.96, 1.02)	0.90 (0.81, 0.99)	0.98 (0.93, 1.03)
Enhanced A&F[c] arm	0.95 (0.85, 1.05)	0.76 (0.46, 1.03)	0.82 (0.75, 0.89)	0.14 (0.07, 0.21)	0.76 (0.67, 0.85)	0.86 (0.72, 0.99)	0.74 (0.67, 0.81)	0.42 (0.11, 0.73)	0.87 (0.75, 0.99)
Time $\times$ Enhanced A&F arm	0.98 (0.87, 1.09)	1.05 (0.88, 1.22)	1.11 (1.04, 1.18)	1.08 (1.02, 1.17)	1.16 (0.94, 1.38)	2.56 (2.19, 2.93)	1.37 (1.15, 1.59)	1.32 (1.13, 1.51)	1.20 (1.06, 1.34)

AVPU[a] :-Alert, Verbal response, Pain response, Unresponsive, MO:- Medical officers, A&F:-Audit and feedback

Table 6.5 and 6.6 present variance-correlation matrices of random clinicians' intercepts among nine pneumonia outcomes under complete case analysis and after MI, respectively. Generally, the magnitude of correlation estimated among outcomes was consistently larger under MI compared to complete case analysis. Moreover, the strength and direction of association amongst pneumonia care outcome varied within and across domains of care. For instance, the strength of association between documentation of oxygen saturation and respiratory rate in the assessment domain was somewhat high, compared to association with other indicators in the assessment domain. To be specific, correlation between oxygen saturation and respiratory rate documentation increased from 0.69 (Table 6.5) under complete case analysis to 0.89 (Table 6.6) after MI of missing covariates. In the treatment domain, prescription of oral amoxicillin and correct dosage, exhibited a strong positive association with a correlation coefficient of 0.73 under complete case analysis (Table 6.5) and 0.80 after MI of missing covariates (Table 6.6).

Across domains of care, correct pneumonia diagnosis was strongly associated with prescription of oral amoxicillin and correct dosage both in the treatment domain. We also observed that documentation of three secondary signs and symptoms (S&S), namely: oxygen saturation, respiratory rate and lower wall chest wall indrawing, in the assessment domain were positively associated with correct pneumonia diagnosis, amoxicillin prescription and correctness of the dose. On the other hand, documentation of cough and difficult breathing (primary S&S) and AVPU (a secondary S&S) in the assessment domain were negatively associated with documentation of other pneumonia care indicators.

Table 6.5: Variance-correlation matrix for the random intercepts under complete case analysis.

	Cough	Difficult breathing	Respiratory rate	Oxygen saturation	AVPU[a]	Indrawing	Correct diagnosis	Correct treatment	Correct dose
Cough	1.49								
Difficult breathing	0.07	1.92							
Respiratory rate	-0.29	-0.43	2.71						
Oxygen saturation	-0.17	-0.47	0.63	7.38					
AVPU	-0.14	-0.19	-0.20	0.09	2.26				
Indrawing	-0.22	-0.11	-0.54	-0.39	-0.19	2.33			
Diagnosis	-0.49	-0.53	0.48	0.29	-0.06	0.04	2.64		
Treatment	-0.48	-0.42	0.07	0.16	-0.38	0.66	0.64	1.81	
Dose	-0.54	-0.64	0.57	0.69	-0.21	0.19	0.62	0.73	1.30

Table 6.6: Variance-correlation matrix for the random intercepts after multiple imputation.

	Cough	Difficult breathing	Respiratory rate	Oxygen saturation	AVPU [a]	Indrawing	Correct diagnosis	Correct treatment	Correct dose
Cough	1.05								
Difficult breathing	0.17	0.71							
Respiratory rate	-0.29	-0.60	2.47						
Oxygen saturation	-0.30	-0.78	0.89	2.23					
AVPU	-0.12	-0.24	-0.12	0.22	1.76				
Indrawing	-0.30	0.06	-0.52	-0.50	-0.26	1.82			
Diagnosis	-0.54	-0.65	0.40	0.24	-0.07	0.35	2.14		
Treatment	-0.45	-0.55	0.23	0.26	-0.22	0.52	0.77	0.56	
Dose	-0.47	-0.76	0.63	0.64	-0.18	0.15	0.74	0.80	0.67

AVPU[a] :-Alert, Verbal response, Pain response, Unresponsive.

Under complete case analysis, a principal component analysis (PCA) on the correlation matrix of the random intercepts showed that the first and second principal components explained 57.6% and 24.6% of the variation respectively (Figure 6.1, panel a). After multiple imputation, the first and second principal components explained 60.3% and 26.2% of the variation respectively (Figure 6.1, panel b). Vectors of two positively correlated outcomes in the loading plots were close, forming a small angle between them (e.g. oxygen saturation and respiratory rate). On the other hand, vectors of negatively correlated outcomes (e.g. cough and treatment) were diverging forming a large angle between them. The direction of vectors for all the outcomes was consistent under complete case analysis and after MI.

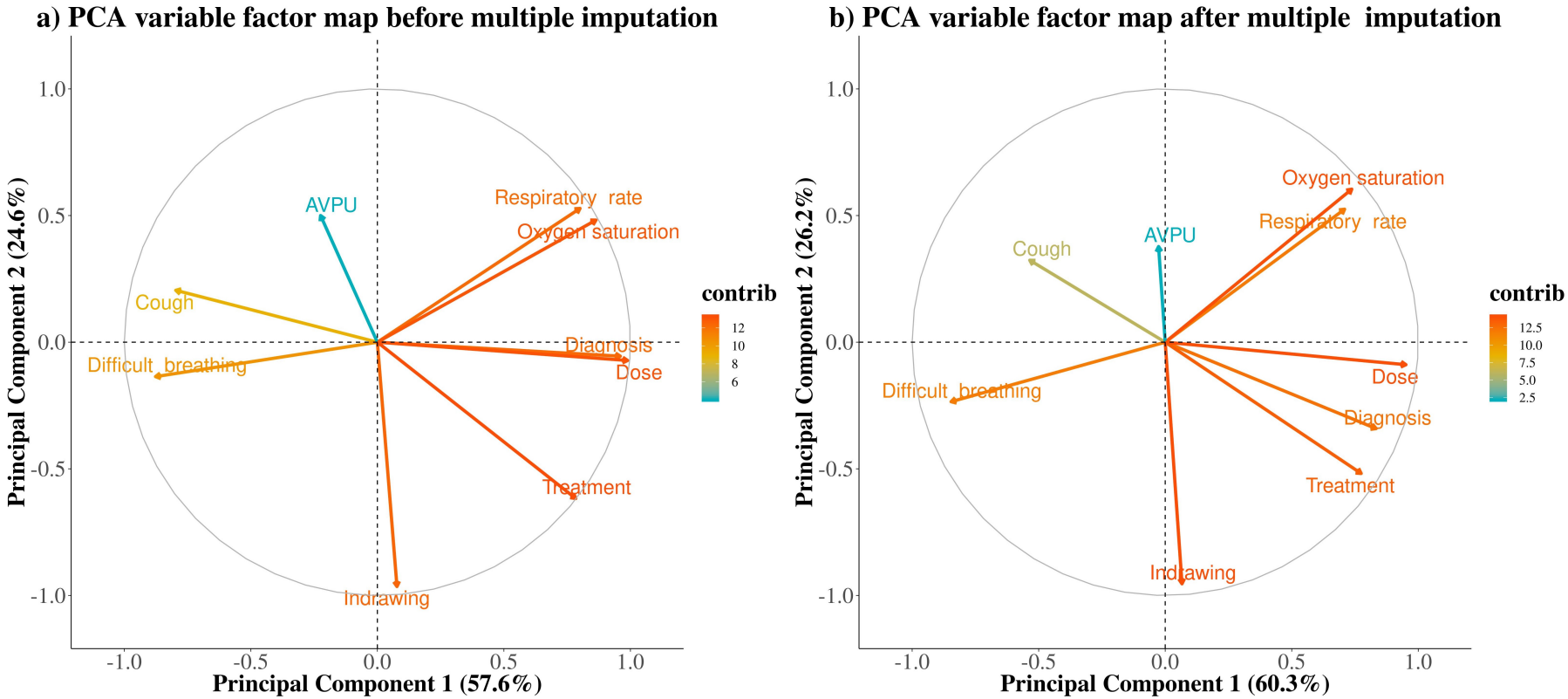

Figure 6.1: Results (component loadings for the first and second principal components) of a principal components analysis on correlation matrix of the random intercepts of model under complete case analysis (panel a) and after multiple imputation (panel b).

6.4 Discussion

In this chapter we sought to estimate the joint effects of covariates on nine paediatric pneumonia quality of care indicators. We also estimated the strength of association among the outcomes using a correlated random-effects joint model (Fieuws and Verbeke, 2006). From study results, there was a significant joint effect of covariates on nine pneumonia outcomes under complete case analysis. The strength and direction of association among pneumonia outcomes varied within and across domains of care. Thus, an assumption of common random-effects amongst all outcomes would be too restrictive and unrealistic for pneumonia trial data analysed in this study.

Further results showed that enhanced audit and feedback improved documentation and adherence to recommended clinical guidelines in six out of nine paediatric pneumonia care indicators over time before and after MI. However, fitting pairwise models after MI led to more precise estimates compared to estimates from pairwise models under complete case analysis. These observations were attributed to loss of information under complete case analysis resulting to larger standard errors hence wider 95% confidence intervals. During the trial period, documentation and adherence to recommended paediatric pneumonia guidelines by clinicians depended on individual quality of care indicators. For instance, documentation of care indicators, that did not require a lot of cognitive effort, were highly documented (e.g. cough, difficult breathing) compared to indicators that required more cognitive effort on the part of the clinician (e.g.

prescribing the right treatment in the right dosage). These variations in delivery of recommended care could also be due to hospital level factors, such as lack of or broken medical devices, impeding delivery of recommended care (e.g. pulse oximeter to measure oxygen saturation).

In the pairwise modelling approach, estimates obtained by averaging over a number of auxiliary estimates (from the various pairs) do not maximize the full multivariate likelihood. However, Fieuws and Verbeke (2005) demonstrated with simulations that the loss of efficiency is small in the pairwise approach relative to a full maximum-likelihood based approach. Moreover, the averaged estimates are consistent and asymptotically normal (Molenberghs and Verbeke, 2005, Chapter 25 ,p. 473), a property which holds for imputed data sets as well thus, ensuring valid within imputation estimates. Validity of within imputation estimates is a prerequisite for the application of Rubin's rules which then account for between imputation variability (Rubin, 1976).

Although we did not evaluate computational complexity explicitly, combining pairwise joint model fitting and MI comes with its computational expense as demonstrated in this study. At imputation stage, the level of complexity is compounded when missing data occur in more than one level of clustering. In such occurrences, it is paramount to account for the hierarchical structure present in the analysis model of interest in the imputation model as well. This is because incompatibility between imputation and analysis model may lead to biased estimates, underestimated cluster level variances and overestimated individual level variances (Grund et al., 2017). In the current study, missing covariates were

imputed using the latent normal approach within the joint model imputation framework while accounting for clustering at clinician level. Additionally, the outcomes of interest, all fully observed were included in the imputation model as auxiliary variables. Nonetheless, there is need for further research possibly through a simulation study to evaluate compatibility between imputation and substantive model or the lack thereof, in the high dimensional joint modelling context.

At analysis stage, complexity stems from calculating parameters of interest (e.g. obtaining variance-covariances matrices for each imputed data set using the pseudo-likelihood approach before applying Rubin's rules). Besides, constructing the overall variance covariance matrix for the random effects is not straight forward, hence the need for greater care to avoid incorrect inferences due to miscalculations. Therefore, future studies can consider developing and incorporating generic functions and packages into standard statistical software to handle missing data and other computational aspects (e.g. Wald-type tests to test for joint covariate effects after MI) more efficiently when the substantive model of interest entails joint modelling of clustered and high-dimensional vectors of outcomes.

The correlated random-effects joint model fitted using the pairwise approach has been previously used to jointly analyse clustered binary data (Fieuws et al., 2006) as well as continuous longitudinal outcomes (Fieuws and Verbeke, 2006). However, there is essentially no example in the literature on how to also account for missing covariates in a high-dimensional joint modelling context, hence the novelty of this study. We estimated joint effects of covariates in addition to quantify-

ing the strength of association among quality of care outcomes, aspects that are largely ignored in routine paediatric care studies.

In previous analysis of the trial data, for instance, diagnosis and classification of pneumonia cases was the primary outcome of interest (Ayieko et al., 2019). In the previous chapters, pneumonia quality of care indicators were combined into a single ordinal composite outcome known as the paediatric quality of care indicator (PAQC) score. Therefore, when there is need for joint inference, we recommend this study as a practical example for handling high-dimensional vector of outcomes using a pairwise fitting approach and at the same time performing MI to account for missing covariates. However, if the research question does not necessitate joint inference, then univariate mixed models as tools for analysis suffice (Fieuws and Verbeke, 2006).

Evidently, this study has a number of limitations. First, our tests for the joint effects were based on complete case analysis only. This was due to lack of functionalities in standard software to perform Wald-type tests for joint covariate effects after MI. Secondly, we imputed missing covariates assuming a MAR mechanism, an assumption that cannot be verified using the observed data alone (Molenberghs et al., 2008; Verbeke and Molenberghs, 2010; Carpenter and Kenward, 2013). Therefore, sensitivity analysis is recommended to explore the robustness of the inferences to the MAR assumptions.

As already noted, fitting pairwise joint models on multiply imputed data sets was time intensive. Future studies may consider MI, an approach suggested by Nassiri et al. (2017) as alternative to the pairwise joint modelling using a sandwich-

type robust variance estimator.

In conclusion, there was a significant joint interaction effect between intervention arm and follow-up time on pneumonia care under complete case analysis. This was in addition to significant joint effects of patients age, clinician's cadre and gender and hospital level factors on pneumonia care indicators. Enhanced audit and feedback improved documentation and adherence to recommended clinical guidelines in six out of nine paediatric pneumonia care indicators over time. Multiple imputation of missing covariates improved precision of parameter estimate compared to complete case analysis. The strength and direction of association estimated using clinicians' random intercepts varied amongst pneumonia outcomes within and across the three domains of pneumonia care. Across domains of care, pneumonia diagnosis was strongly correlated with oral amoxicillin prescription and dosage.

Chapter 7

Conclusion and Recommendations

Research studies that involve analysis of routine heath care data are often subject to incompleteness, thus necessitating application of missing data techniques at analysis stage. In the past, analyses of incomplete data revolved around simple methods such as complete case analysis. However, there has been significant methodological and computational development of principled missing data handling methods, thus broadening the options available to researchers (Sotto, 2009). Despite the methodological advancements, analysis of incomplete data remains less than straightforward. This is because inferential validity of missing data techniques rests on untestable assumptions regarding the underlying missing data mechanism (Molenberghs et al., 2014, Chapter 1).

This report demonstrates practical utility of advanced biostatistical analyses methods of partially observed multivariate hierarchical data, often encountered by researchers in health research. The choice regarding the type of analysis depends on several practical considerations. These include the scientific research question

of interest, computational complexities, as well as the level of incompleteness present in the data at hand. Uptake and utilization of sound statistical methods can improve analysis and reporting of health data used to inform policies and in the long run enhance optimal utilization of limited resources while promoting better patients' outcomes. A summary of methods, the resulting conclusions and areas of further research for the pertinent chapters are as follows:

In chapter 3, we considered the latent normal joint MI approach to handle missing covariates while accounting for multilevel structures of the data at hand. We then used random-effects models and GEE to analyze an ordinal composite outcome. The study results reinforced the strengths of MI over complete case analysis in terms of parameter estimates efficiency and precision. Thus, while complete case analysis remains the default missing data handling technique in most statistical software, it should be used as a preliminary step or supplement to more appropriate methods.

In Chapter 4 we explored handling missing components in a composite outcome of interest. Through a range of simulation scenarios which entailed 2 missing data mechanisms: missing MCAR and MAR, and 3 rates of missingness: 3%, 10% and 40%. We compared performance of MI at item level and the conventional method where missing composite components were scored with value 0. Results indicated that the precision of the estimates in both methods was largely dependent on the amount of missingness as well as the underlying missing data mechanism. Nonetheless, MI produced less biased estimates compared to the conventional approach across simulation scenarios. However, 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.

In Chapter 3 and 4, missing data were imputed assuming a MAR mechanism. To assess robustness of these inferences, sensitivity analysis within the PMM framework was the subject of chapter 5. The sensitivity analyses focused on two categorical covariates in the second level of pneumonia trial data analyzed in this study. The two variables were imputed assuming a MNAR mechanism. To achieve this, we elicited and incorporated 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. In both approaches, inferences were insensitive to departures from the MAR mechanism thus increasing the level confidence in the resulting conclusions. Furthermore, both procedures led to estimates and standard errors that were comparable in magnitude. Due to computational complexity reasons, sensitivity analyses for the two variables of interest were performed separately. Therefore, future studies may consider strategies of performing multi-dimensional sensitivity analyses of two or more target variables. Although sensitivity analyses methods are well established in theory, implementation in statistical software is still lagging especially for multilevel data contexts, thus limiting their utility in practice.

In Chapter 6, nine binary outcomes were jointly modelled under the correlated random-effects approach. Specifically, we applied the pairwise fitting and pseudo-likelihood methods before and after multilevel MI of missing covariates. Even

though the analysis was time intensive under MI, the parameter estimates were more precise compared to those under complete case analysis. However, due to software and computational complexities, Wald-type test for joint covariate effects were restricted to complete case analysis. Consequently, more research work is needed with regards to testing for joint covariates effects in pairwise fitting and pseudo-likelihood methods after MI of missing variables.

When using MI to handle partially observed data, it is vital to ensure compatibility between the imputation model and the analysis model of interest. That is, the imputation model should be compatible with or richer than the analysis model interest. Nonetheless, there is need for further research to evaluate compatibility in greater details in complex settings such as the high dimensional joint modelling and analysis of composite outcomes when some of the subcomponents are partially observed.

References

Aerts, M., Geys, H., Molenberghs, G., and Ryan, L. (2002). *Topics in modelling of clustered data*. CRC Press.

Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, second edition.

Agweyu, A., Lilford, R. J., English, M., and Network, C. I. (2018a). Appropriateness of clinical severity classification of new who childhood pneumonia guidance: a multi-hospital, retrospective, cohort study. *The Lancet Global Health*, 6(1):e74–e83.

Agweyu, A., Oliwa, J., Gathara, D., Muinga, N., Allen, E., Lilford, R. J., and English, M. (2018b). Comparable outcomes among trial and nontrial participants in a clinical trial of antibiotics for childhood pneumonia: a retrospective cohort study. *Journal of clinical epidemiology*, 94:1–7.

Akech, S., Ayieko, P., Gathara, D., Agweyu, A., Irimu, G., Stepniewska, K., English, M., Ngarngar, S., Aduro, N., Mutai, L., et al. (2018). Risk factors for mortality and effect of correct fluid prescription in children with diarrhoea and dehydration without severe acute malnutrition admitted to kenyan hospitals: an

observational, association study. *The Lancet Child & Adolescent Health*, 2(7):516–524.

Akech, S., Chepkirui, M., Ogero, M., Agweyu, A., Irimu, G., English, M., and Snow, R. W. (2019). The clinical profile of severe paediatric malaria in an area targeted for routine rts, s/as01 malaria vaccination in western kenya. *Clinical Infectious Diseases*.

Allison, P. D. (2012). Handling missing data by maximum likelihood. In *SAS global forum*, volume 2012. Statistical Horizons, Havenford, PA.

Amboko, B. I., Ayieko, P., Ogero, M., Julius, T., Irimu, G., and English, M. (2016). Malaria investigation and treatment of children admitted to county hospitals in western kenya. *Malaria journal*, 15(1):506.

Amouzou, A., Velez, L. C., Tarekegn, H., and Young, M. (2016). One is too many: ending child deaths from pneumonia and diarrhoea.

Andridge, R. R. and Little, R. J. (2011). Proxy pattern-mixture analysis for survey nonresponse. *Journal of Official Statistics*, 27(2):153.

Ayieko, P., Irimu, G., and English, M. (2017). 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(1):416.

Ayieko, P., Irimu, G., Ogero, M., Mwaniki, P., Malla, L., Julius, T., Chepkirui, M., Mbevi, G., Oliwa, J., Agweyu, A., et al. (2019). 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*, 14(1):20.

Ayieko, P., Ogero, M., Makone, B., Julius, T., Mbevi, G., Nyachiro, W., Nyamai, R., Were, F., Githanga, D., Irimu, G., et al. (2015). Characteristics of admissions and variations in the use of basic investigations, treatments and outcomes in kenyan hospitals within a new clinical information network. *Archives of disease in childhood*, pages archdischild–2015.

Ayieko, P., Okiro, E. A., Edwards, T., Nyamai, R., and English, M. (2012). Variations in mortality in children admitted with pneumonia to kenyan hospitals. *PLoS One*, 7(11):e47622.

Bartlett, J., Morris, T., et al. (2015). Multiple imputation of covariates by substantive-model compatible fully conditional specification. *The Stata Journal*, 15(2):437–456.

Bebu, I. and Lachin, J. M. (2018). Properties of composite time to first event versus joint marginal analyses of multiple outcomes. *Statistics in medicine*.

Bitton, A., Ratcliffe, H. L., Veillard, J. H., Kress, D. H., Barkley, S., Kimball, M., Secci, F., Wong, E., Basu, L., Taylor, C., et al. (2017). Primary health care as a foundation for strengthening health systems in low-and middle-income countries. *Journal of general internal medicine*, 32(5):566–571.

Blough, D. K., Ramsey, S., Sullivan, S. D., and Yusen, R. (2009). 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*, 18(1):91–101.

Bohlius, J., Maxwell, N., Spoerri, A., Wainwright, R., Sawry, S., Poole, J., Eley, B., Prozesky, H., Rabie, H., Garone, D., et al. (2016). Incidence of aids-defining and other cancers in hiv-positive children in south africa: record linkage study. *The Pediatric infectious disease journal*, 35(6):e164.

Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in medicine*, 25(24):4279–4292.

Caldis, T. (2007). Composite health plan quality scales. *Health care financing review*, 28(3):95.

Carpenter, J. R. and Kenward, M. G. (2013). *Multiple imputation and its applications*. Chichester: John Wiley & Sons.

Chen, L. M., Staiger, D. O., Birkmeyer, J. D., Ryan, A. M., Zhang, W., and Dimick, J. B. (2013). Composite quality measures for common inpatient medical conditions. *Medical care*, 51(9):832.

Christensen, R. H. B. (2015). Package 'ordinal'—regression models for ordinal data. *R package version*, 28:2015.

Commission, J. R. C.-E. et al. (2008). *Handbook on constructing composite indicators: methodology and user guide*. OECD publishing.

Cook, D. A., Pencille, L. J., Dupras, D. M., Linderbaum, J. A., Pankratz, V. S., and Wilkinson, J. M. (2018). Practice variation and practice guidelines: Attitudes of

generalist and specialist physicians, nurse practitioners, and physician assistants. *PloS one*, 13(1).

Cordoba, G., Schwartz, L., Woloshin, S., Bae, H., and Gøtzsche, P. C. (2010). Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *Bmj*, 341:c3920.

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M. G. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, 52(1):111–125.

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M. G. (2011). Generalized shared-parameter models and missingness at random. *Statistical modelling*, 11(4):279–310.

Daniels, M. J. and Hogan, J. W. (2008). Missing data in longitudinal studies. *Monographs on Statistics and Applied Probability*, 109.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Drechsler, J. (2015). Multiple imputation of multilevel missing data—rigor versus simplicity. *Journal of Educational and Behavioral Statistics*, 40(1):69–95.

Eapen, Z. J., Fonarow, G. C., Dai, D., O'brien, S. M., Schwamm, L. H., Cannon, C. P., Heidenreich, P. A., Bhatt, D. L., Peterson, E. D., and Hernandez, A. F. (2011). Comparison of composite measure methodologies for rewarding qual-

ity of care: an analysis from the american heart association's get with the guidelines program. *Circulation: Cardiovascular Quality and Outcomes*, 4(6):610–618.

Eekhout, I. (2015). *Don't Miss Out!: Incomplete data can contain valuable information*. PhD thesis, Vrije Universiteit Amsterdam. Geneeskunde Naam instelling promotie: VU University Naam instelling onderzoek: VU University medical center.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.

Enders, C. K., Mistler, S. A., and Keller, B. T. (2016). Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, 21(2):222.

English, M. (2013). Designing a theory-informed, contextually appropriate intervention strategy to improve delivery of paediatric services in kenyan hospitals. *Implementation Science*, 8(1):39.

EUnetHTA, J. (2013). Endpoints used for relative effectiveness assessment of pharmaceuticals healthrelated quality of life and utility measures.

Faes, C., Aerts, M., Molenberghs, G., Geys, H., Teuns, G., and Bijnens, L. (2008). A high-dimensional joint model for longitudinal outcomes of different nature. *Statistics in medicine*, 27(22):4408–4427.

Fieuws, S. and Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statistics in medicine*, 23(20):3093–3104.

Fieuws, S. and Verbeke, G. (2005). Evaluation of the pairwise approach for fitting joint linear mixed models: A simulation study. Technical report, Technical Report TR0527, Biostatistical Centre, Katholieke Universiteit ....

Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62(2):424–431.

Fieuws, S., Verbeke, G., Boen, F., and Delecluse, C. (2006). High dimensional multivariate mixed models for binary questionnaire data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(4):449–460.

Fieuws, S., Verbeke, G., and Molenberghs, G. (2007). Random-effects models for multivariate repeated measures. *Statistical methods in medical research*, 16(5):387–397.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2009a). *Handbooks of modern statistical methods: Longitudinal data analysis*. New York: Taylor & Francis Group.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2009b). *Longitudinal Data Analysis*. Boca Raton: Chapman & Hall.

Freemantle, N., Calvert, M., Wood, J., Eastaugh, J., and Griffin, C. (2003). Composite outcomes in randomized trials: greater precision but with greater uncertainty? *Jama*, 289(19):2554–2559.

Gachau, S., Ayieko, P., Gathara, D., Mwaniki, P., Ogero, M., Akech, S., Maina, M., Agweyu, A., Oliwa, J., Julius, T., et al. (2017). 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(4):e000468.

Gachau, S., Irimu, G., Ayieko, P., Akech, S., Agweyu, A., English, M., et al. (2018). Prevalence, outcome and quality of care among children hospitalized with severe acute malnutrition in kenyan hospitals: A multi-site observational study. *PloS one*, 13(5):e0197607.

Galimard, J.-E., Chevret, S., Curis, E., and Resche-Rigon, M. (2018). Heckman imputation models for binary or continuous mnar outcomes and mar predictors. *BMC medical research methodology*, 18(1):90.

Gathara, D., Malla, L., Ayieko, P., Karuri, S., Nyamai, R., Irimu, G., Hensbroek, M. B., Allen, E., and English, M. (2017). Variation in and risk factors for paediatric inpatient all-cause mortality in a low income setting: data from an emerging clinical information network. *BMC pediatrics*, 17(1):99.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Gordon, D. M., Frenning, S., Draper, H. R., and Kokeb, M. (2013). Prevalence and burden of diseases presenting to a general pediatrics ward in gondar, ethiopia. *Journal of tropical pediatrics*, 59(5):350–357.

Grund, S., Lüdtke, O., and Robitzsch, A. (2017). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, page 1094428117703686.

Grund, S., Lüdtke, O., and Robitzsch, A. (2018). Multiple imputation of missing data at level 2: A comparison of fully conditional and joint modeling in multilevel designs. *Journal of Educational and Behavioral Statistics*, 43(3):316–353.

Grund, S., Robitzsch, A., Luedtke, O., and Grund, M. S. (2019). Package 'mitml'.

Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*, 1(3):177–193.

Gueorguieva, R. V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association*, 96(455):1102–1112.

Harries, A. D., Zachariah, R., and Maher, D. (2013). The power of data: using routinely collected data to improve public health programmes and patient outcomes in low-and middle-income countries. *Tropical medicine & international health*, 18(9):1154–1156.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (redcap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics*, 42(2):377–381.

Hau, D. K., Chami, N., Duncan, A., Smart, L. R., Hokororo, A., Kayange, N. M., and Peck, R. N. (2018). Post-hospital mortality in children aged 2-12 years in tanzania: A prospective cohort study. *PloS one*, 13(8):e0202334.

Héraud-Bousquet, V., Larsen, C., Carpenter, J., Desenclos, J.-C., and Le Strat, Y. (2012). Practical considerations for sensitivity analysis after multiple imputation applied to epidemiological studies with incomplete data. *BMC medical research methodology*, 12(1):73.

Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC medical research methodology*, 16(1):117.

Hooli, S., Colbourn, T., Lufesi, N., Costello, A., Nambiar, B., Thammasitboon, S., Makwenda, C., Mwansambo, C., McCollum, E. D., and King, C. (2016). Predicting hospitalised paediatric pneumonia mortality risk: an external validation of risc and mrisc, and local tool development (risc-malawi) from malawi. *PLoS One*, 11(12):e0168126.

Houngbo, P. T., Buning, T. D. C., Bunders, J., Coleman, H. L., Medenou, D., Dakpanon, L., and Zweekhorst, M. (2017). Ineffective healthcare technology management in benin's public health sector: the perceptions of key actors and their ability to address the main problems. *International journal of health policy and management*, 6(10):587.

Huberty, C. J. and Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological bulletin*, 105(2):302.

Ibrahim, F., Tom, B. D., Scott, D. L., and Prevost, A. T. (2016). A systematic review of randomised controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes. *Trials*, 17(1):272.

Jaffa, M. A., Gebregziabher, M., and Jaffa, A. A. (2014). A joint modeling approach for right censored high dimensional multivariate longitudinal data. *Journal of biometrics & biostatistics*, 5(4).

Jenghara, M. M., Ebrahimpour-Komleh, H., Rezaie, V., Nejatian, S., Parvin, H., and Yusof, S. K. S. (2018). Imputing missing value through ensemble concept based on statistical measures. *Knowledge and Information Systems*, 56(1):123–139.

Karuri, S. W., Murithi, M. K., Irimu, G., English, M., et al. (2017). Using data from a multi-hospital clinical network to explore prevalence of pediatric rickets in kenya. *Wellcome open research*, 2.

Kundu, M. G. (2011). Implementation of pairwise fitting technique for analyzing multivariate longitudinal data in sas. In *PharmaSUG 2011 Conference Proceedings*, pages 1–12.

Leacy, F. P., Floyd, S., Yates, T. A., and White, I. R. (2017). 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*, 185(4):304–315.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134.

Little, R. J., D'agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T.,

Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., et al. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, 367(14):1355–1360.

Little, R. J. and Rubin, D. B. (2002). Bayes and multiple imputation. *Statistical Analysis with Missing Data, Second Edition*, pages 200–220.

Liublinska, V. and Rubin, D. B. (2014). Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. *Statistics in medicine*, 33(24):4170–4185.

Lloyd, J. E., Obradović, J., Carpiano, R. M., and Motti-Stefanidi, F. (2013). Jmasm 32: Multiple imputation of missing multilevel, longitudinal data: A case when practical considerations trump best practices? *Journal of Modern Applied Statistical Methods*, 12(1):29.

Long, J. D. and Mills, J. A. (2018). Joint modeling of multivariate longitudinal data and survival data in several observational studies of huntington's disease. *BMC medical research methodology*, 18(1):1–15.

Lopez, A. D. (2014). *Levels and Trends in Child Mortality: Report 2014, Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. United Nations Inter-agency Group for Child Mortality Estimation (UN IGME).

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 226–233.

Lunn, D., Jackson, C., Best, N., Spiegelhalter, D., and Thomas, A. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman and Hall/CRC.

Mackinnon, A. (2010). The use and reporting of multiple imputation in medical research–a review. *Journal of internal medicine*, 268(6):586–593.

Malla, L., Perera-Salazar, R., McFadden, E., and English, M. (2017). Comparative effectiveness of injectable penicillin versus a combination of penicillin and gentamicin in children with pneumonia characterised by indrawing in kenya: a retrospective observational study. *BMJ open*, 7(11):e019478.

Mbevi, G., Ayieko, P., Irimu, G., Akech, S., and English, M. (2016). Prevalence, aetiology, treatment and outcomes of shock in children admitted to kenyan hospitals. *BMC medicine*, 14(1):184.

Mc Cord, K. A., Salman, R. A.-S., Treweek, S., Gardner, H., Strech, D., Whiteley, W., Ioannidis, J. P., and Hemkens, L. G. (2018). Routinely collected data for randomized trials: promises, barriers, and implications. *Trials*, 19(1):29.

McCulloch, C. (2008). Joint modelling of mixed outcome types using latent variables. *Statistical methods in medical research*, 17(1):53–73.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, pages 538–558.

Meng, X.-L. and Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79(1):103–111.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):371–388.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., and Verbeke, G. (2014). *Handbook of missing data methodology*. CRC Press.

Molenberghs, G. and Kenward, M. (2007). *Missing data in clinical studies*, volume 61. John Wiley & Sons.

Molenberghs, G. and Verbeke, G. (2005). *Models for discrete longitudinal data*. Springer.

Morris, T. P., White, I. R., and Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in medicine*, 38(11):2074–2102.

Mostafa, S. M. (2019). Imputing missing values using cumulative linear regression. *CAAI Transactions on Intelligence Technology*, 4(3):182–200.

Muthee, V., Bochner, A. F., Osterman, A., Liku, N., Akhwale, W., Kwach, J., Prachi, M., Wamicwe, J., Odhiambo, J., Onyango, F., et al. (2018). The impact of routine data quality assessments on electronic medical record data quality in kenya. *PloS one*, 13(4):e0195362.

Mwaniki, P., Ayieko, P., Todd, J., and English, M. (2014). Assessment of paediatric inpatient care during a multifaceted quality improvement intervention in

kenyan district hospitals–use of prospectively collected case record data. *BMC health services research*, 14(1):312.

Nakai, M. and Ke, W. (2011). Review of the methods for handling missing data in longitudinal data analysis. *International Journal of Mathematical Analysis*, 5(1):1–13.

Nassiri, V., Ivanova, A., Molenberghs, G., and Verbeke, G. (2017). Fast precision estimation in high-dimensional multivariate joint models. *Biometrical Journal*, 59(6):1221–1231.

Njagi, E. N., Molenberghs, G., Kenward, M. G., Verbeke, G., and Rizopoulos, D. (2014). A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis. *Biometrical Journal*, 56(6):1001–1015.

Ogero, M., Akech, S., Malla, L., Agweyu, A., Irimu, G., and English, M. (2020). Examining which clinicians provide admission hospital care in a high mortality setting and their adherence to guidelines: an observational study in 13 hospitals. *Archives of Disease in Childhood*.

Ogero, M., Ayieko, P., Boniface Makone, T. J., Malla, L., Oliwa, J., Irimu, G., English, M., et al. (2018). An observational study of monitoring of vital signs in children admitted to kenyan hospitals: an insight into the quality of nursing care? *Journal of global health*, 8(1).

Oliwa, J. N., Gathara, D., Ogero, M., van Hensbroek, M. B., English, M., van't Hoog, A., and Network, C. I. (2019). 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*, 14(9):e0221145.

Omore, R., Tate, J. E., O'Reilly, C. E., Ayers, T., Williamson, J., Moke, F., Schilling, K. A., Awuor, A. O., Jaron, P., Ochieng, J. B., et al. (2016). Epidemiology, seasonality and factors associated with rotavirus infection among children with moderate-to-severe diarrhea in rural western kenya, 2008–2012: the global enteric multicenter study (gems). *PloS one*, 11(8):e0160060.

Opondo, C., Allen, E., Todd, J., and English, M. (2016). 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*, 21(10):1334–1345.

Opondo, C., Allen, E., Todd, J., and English, M. (2018). Association of the paediatric admission quality of care score with mortality in kenyan hospitals: a validation study. *The Lancet Global Health*, 6(2):e203–e210.

Organization, W. H. (2013). *Pocket book of hospital care for children: guidelines for the management of common childhood illnesses*. World Health Organization.

Organization, W. H. et al. (2003). Improving data quality: a guide for developing countries.

Organization, W. H. et al. (2014). Revised who classification and treatment of childhood pneumonia at health facilities. *Geneva: World Health Organization*, pages 6–14.

Plumpton, C. O., Morris, T., Hughes, D. A., and White, I. R. (2016). Multiple imputation of multiple multi-item scales when a full imputation model is infeasible. *BMC research notes*, 9(1):45.

Pocock, S. J., Geller, N. L., and Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics*, pages 487–498.

Profit, J., Kowalkowski, M. A., Zupancic, J. A., Pietz, K., Richardson, P., Draper, D., Hysong, S. J., Thomas, E. J., Petersen, L. A., and Gould, J. B. (2014). Babymonitor: a composite indicator of nicu quality. *Pediatrics*, pages peds–2013.

Profit, J., Typpo, K. V., Hysong, S. J., Woodard, L. D., Kallen, M. A., and Petersen, L. A. (2010). Improving benchmarking by using an explicit framework for the development of composite indicators: an example using pediatric quality of care. *Implementation Science*, 5(1):13.

Proschan, M. A. and Waclawiw, M. A. (2000). Practical guidelines for multiplicity adjustment in clinical trials. *Controlled clinical trials*, 21(6):527–539.

Quartagno, M. (2016). *Multiple Imputation for Individual Patient Data Meta-Analyses*. PhD thesis, London School of Hygiene & Tropical Medicine.

Quartagno, M., Carpenter, J., and Quartagno, M. M. (2019). Package 'jomo'.

Quartagno, M. and Carpenter, J. R. (2018). Multilevel multiple imputation in presence of interactions, non-linearities and random slopes. In *Studies in Theoretical and Applied Statistics-SIS2018-49th Meeting of the Italian Statistical Society, Palermo*, pages 20–22.

Quartagno, M. and Carpenter, J. R. (2019). Multiple imputation for discrete data: Evaluation of the joint latent normal model. *Biometrical Journal*, 61(4):1003–1019.

Rizopoulos, D., Rizopoulos, M. D., Imports, M., SystemRequirements, J., and Rcpp, L. (2020). Package 'jmbayes'.

Rombach, I., Gray, A. M., Jenkinson, C., Murray, D. W., and Rivero-Arias, O. (2018). Multiple imputation for patient reported outcome measures in randomised controlled trials: advantages and disadvantages of imputing at the item, subscale or composite score level. *BMC medical research methodology*, 18(1):87.

Rowe, A. K., De Savigny, D., Lanata, C. F., and Victora, C. G. (2005). How can we achieve and maintain high-quality performance of health workers in low-resource settings? *The Lancet*, 366(9490):1026–1035.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Rubin, D. B. (1987). *Multiple imputation for survey nonresponse*. New York: Wiley.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall/CRC.

Shwartz, M., Restuccia, J. D., and Rosen, A. K. (2015). Composite measures of

health care provider performance: a description of approaches. *The Milbank Quarterly*, 93(4):788–825.

Simons, C. L., Rivero-Arias, O., Yu, L.-M., and Simon, J. (2015). Multiple imputation to deal with missing eq-5d-3l data: Should we impute individual domains or the actual index? *Quality of life research*, 24(4):805–815.

Smuk, M., Carpenter, J., and Morris, T. (2017). What impact do assumptions about missing data have on conclusions? a practical sensitivity analysis for a cancer survival registry. *BMC medical research methodology*, 17(1):21.

Sotto, C. (2009). *Topics in Analysis and Sensitivity Analysis of Incomplete Longitudinal data*. PhD thesis, UHasselt Diepenbeek.

Su, Y.-S., Gelman, A. E., Hill, J., and Yajima, M. (2011). Multiple imputation with diagnostics (mi) in r: Opening windows into the black box.

Tang, Y. (2017). On the multiple imputation variance estimator for control-based and delta-adjusted pattern mixture models. *Biometrics*, 73(4):1379–1387.

Thomas, J., Ayieko, P., Ogero, M., Gachau, S., Makone, B., Nyachiro, W., Mbevi, G., Chepkirui, M., Malla, L., Oliwa, J., et al. (2017). Blood transfusion delay and outcome in county hospitals in kenya. *The American journal of tropical medicine and hygiene*, 96(2):511–517.

Tompsett, D. M., Leacy, F., Moreno-Betancur, M., Heron, J., and White, I. R. (2018). On the use of the not-at-random fully conditional specification (narfcs) procedure in practice. *Statistics in medicine*, 37(15):2338–2353.

Touloumis, A. (2014). R package multgee: A generalized estimating equations solver for multinomial responses. *arXiv preprint arXiv:1410.5232*.

Tsiatis, A. A., Kenward, M. G., Fitzmaurice, G., Verbeke, G., and Molenberghs, G. (2014). *Handbook of missing data methodology*. Chapman and Hall/CRC.

Tuti, T., Agweyu, A., Mwaniki, P., Peek, N., and English, M. (2017). An exploration of mortality risk factors in non-severe pneumonia in children using clinical data from kenya. *BMC medicine*, 15(1):201.

Tuti, T., Bitok, M., Paton, C., Makone, B., Malla, L., Muinga, N., Gathara, D., and English, M. (2015). 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(1):184–192.

UNICEF et al. (2016). Pneumonia kills half a million children under five in subsaharan africa.

Van Buuren, S. (2018). *Flexible imputation of missing data*. Chapman and Hall/CRC.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in r. jstat softw 45 (3): 1–67.

Verbeke, G., Fieuws, S., Molenberghs, G., and Davidian, M. (2014). The analysis of multivariate longitudinal data: A review. *Statistical methods in medical research*, 23(1):42–59.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York:Springer.

Verbeke, G. and Molenberghs, G. (2010). Arbitrariness of models for augmented and coarse data, with emphasis on incomplete data and random effects models. *Statistical Modelling*, 10(4):391–419.

White, I. R. and Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in medicine*, 29(28):2920–2931.

Yelin, E., Yazdany, J., Tonner, C., Trupin, L., Criswell, L. A., Katz, P., and Schmajuk, G. (2015). Interactions between patients, providers, and health systems and technical quality of care. *Arthritis care & research*, 67(3):417–424.

Yuan, Y. (2014). Sensitivity analysis in multiple imputation for missing data. In *Proceedings of the SAS Global Forum 2014 Conference:[http://support. sas. com/resources/papers/proceedings14/SAS270-2014. pdf]*.

Chapter A

Appendices

A Chapter 3 Appendices

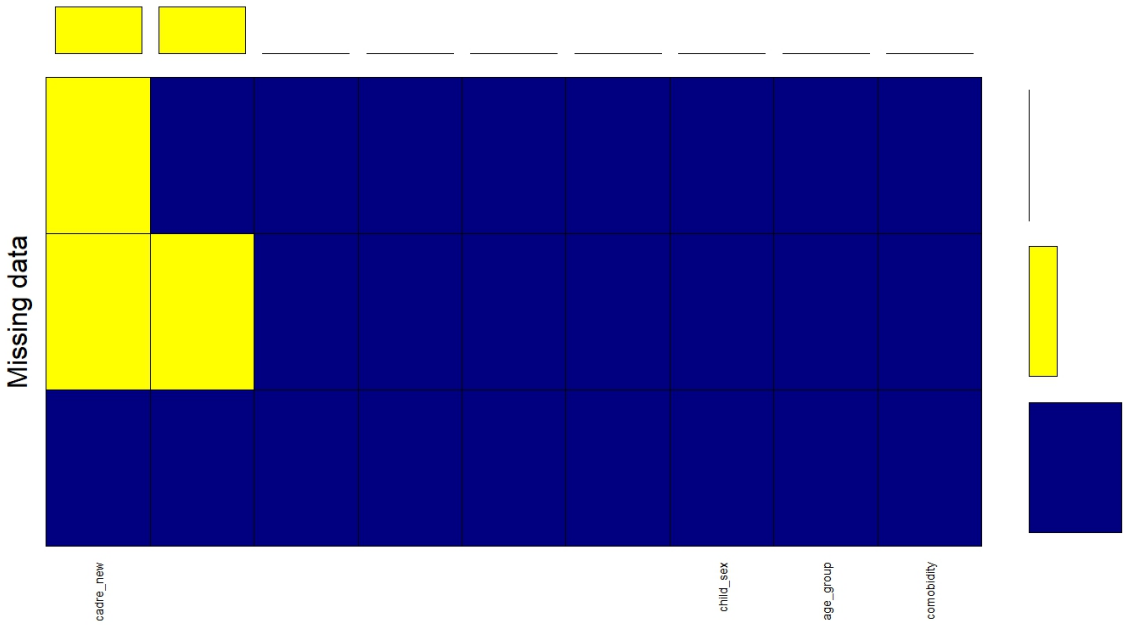

Figure A.1: Missing data pattern underlying pneumonia trial data.

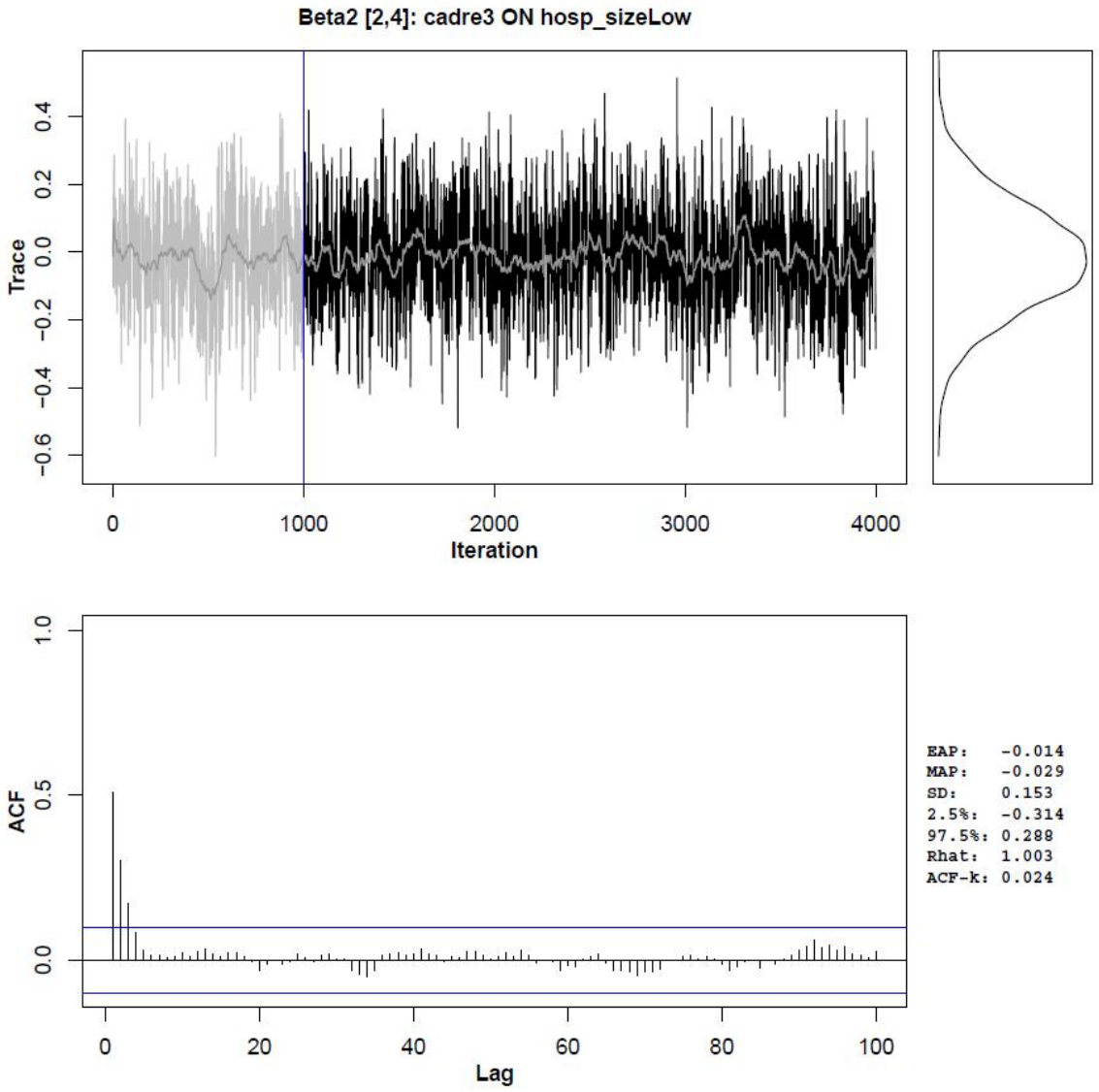

Figure A.2: Example of a chain portraying satisfactory multilevel multiple imputation convergence

Table A.1: Multiple logistic regression model parameter estimates (standard errors) for the probabilities of missing patients' sex, clinician's sex and cadre.

	Patient's sex		Clinician's sex		Clinician's cadre	
Variable	Estimate (s.e)	P value	Estimate (s.e)	P value	Estimate (s.e)	P value
Intercept: PAQC score1	0.66 (0.86)	0.99	9.36 (0.001)	<0.001	9.35 (0.01)	<0.001
Intercept: PAQC score 2	-2.02 (0.83)	0.97	-8.42 (0.001)	<0.001	-8.42 (0.01)	<0.001
Intercept: PAQC score 3	-2.16 (0.62)	0.85	6.44 (0.002)	0.002	6.44 (0.01)	<0.001
Intercept: PAQC score 4	0.33 (0.36)	0.78	-3.98 (0.001)	0.003	-3.98 (0.01)	0.002
Intercept: PAQC score 5	0.43 (0.16)	0.67	1.81 (0.001)	0.002	1.81 (0.01)	0.001
Intercept: PAQC score 6	1.12 (0.50)	0.69	-0.26 (0.001)	0.03	-0.26 (0.01)	0.02
Age-group:12-59 months	-0.56 (0.63)	0.38	0.21 (0.96)	0.03	-0.16 (0.63)	0.04
Comorbidity 0	0.91 (0.82)	0.12	-1.96 (0.97)	0.04	-1.96 (0.97)	0.02
Comorbidity 1	2.04 (6.40)	0.99	-2.09 (1.15)	0.06	-2.09 (1.14)	0.21
Comorbidity 2	2.04 (3.70)	0.98	-2.03 (0.002)	0.10	-2.03 (0.01)	0.11
Malaria prevalence: low	-9.5 (6.24)	0.96	-10.13 (0.001)	0.01	-10.13 (0.01)	0.01
Hospital workload: low	1.63 (0.39)	0.68	-1.06 (1.08)	0.03	-1.06 (1.08)	0.02
Enhanced A&F arm	1.47 (2.34)	0.43	0.21 (0.001)	0.03	0.21 (0.01)	<0.001
Time (months)	-0.34 (0.13)	0.07	-0.48 (0.002)	0.03	-0.47 (0.01)	0.04
Time×Enhanced A&F arm	0.26 (0.40)	0.68	-0.05 (0.001)	<0.001	-0.06 (0.001)	<0.001

Table A.2: Standard errors estimated in random effects model and GEE model under complete case analysis and after multilevel multiple imputation.

	Random effects model		GEE Model	
	Complete case analysis	Multilevel MI	Complete case analysis	Multilevel MI
Effect	Standard error	Standard error	Standard error	Standard error
Intercept: PAQC score 0	ref	ref	ref	ref
Intercept: PAQC score 1	1.231	1.074	1.031	1.010
Intercept: PAQC score 2	1.075	0.381	0.332	0.332
Intercept: PAQC score 3	0.383	0.378	0.329	0.330
Intercept: PAQC score 4	0.380	0.378	0.336	0.336
Intercept: PAQC score 5	0.380	0.382	0.334	0.334
Intercept: PAQC score 6	0.384	0.387	0.342	0.341
Age-group:12-59	0.389	0.093	0.090	0.086
Child sex: Males	0.100	0.096	0.084	0.084
Comorbidities: 0	0.121	0.120	0.122	0.120
Comorbidities :1	0.141	0.140	0.131	0.130
Comorbidities: 2	0.231	0.230	0.209	0.208
Clinician sex: female	0.186	0.185	0.183	0.182
Clinician Cadre: MO	0.186	0.185	0.166	0.163
Hospital workload: low	0.202	0.201	0.178	0.160
Malaria prevalence: low	0.198	0.198	0.190	0.191
Time (months)	0.042	0.042	0.038	0.038
Enhanced A&F arm	0.332	0.332	0.335	0.338
Time$\times$ Enhanced A&F	0.064	0.063	0.060	0.060

B Chapter 4 Appendices

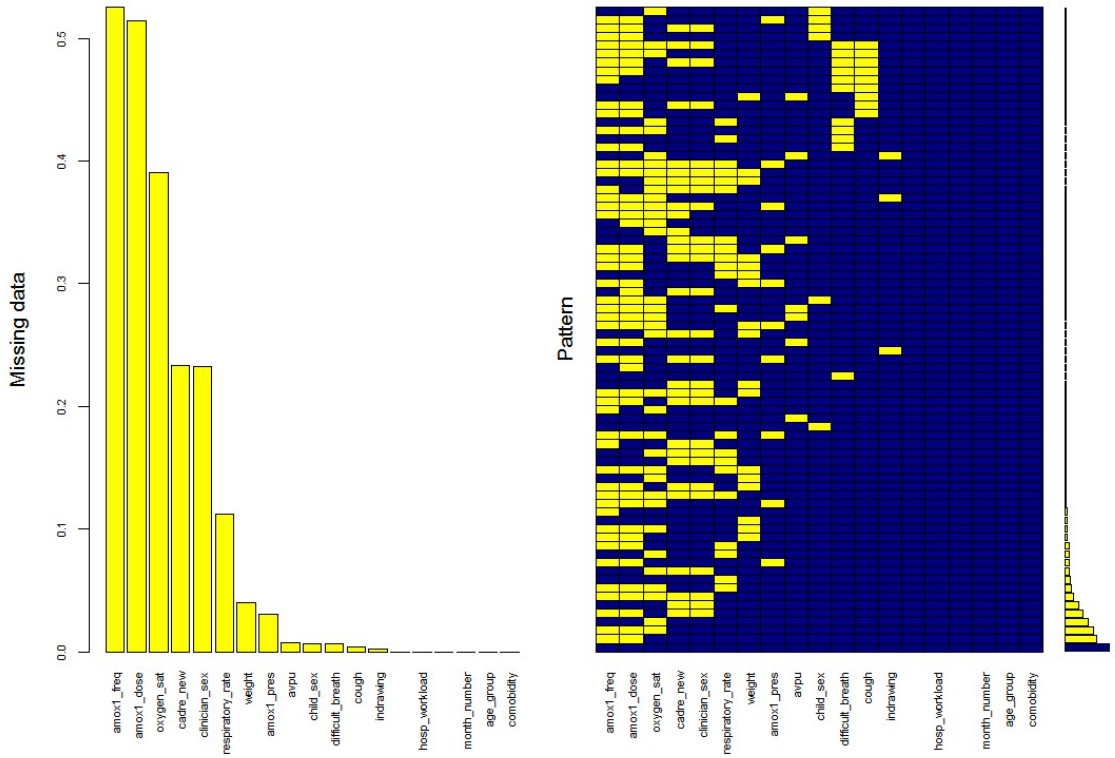

Figure B.1: Missing data pattern underlying covariates and pneumonia outcomes in the trial data.

Table B.1: Coefficients (standard errors) from logistic regression models for the probability missing PAQC score components in the treatment domain and independent variables (patient's sex and clinician's cadre and sex variables).

	Patient's sex		Amoxicillin dose		Patient's weight		Amoxicillin frequency		Clinician's sex		Clinician's cadre	
Variable	Estimate (s.e)	P value	Estimate (s.e)	P value	Estimate (s.e)	P value	Estimate (s.e)	P value	Estimate (s.e)	P value	Estimate (s.e)	P value
Intercept: PAQC score 1	0.66 (0.86)	0.99	7.12 (8.70)	0.86	-0.47 (0.61)	0.46	1.20 (3.50)	0.95	9.36 (0.001)	<0.001	9.35 (0.01)	<0.001
Intercept: PAQC score 2	-2.02 (0.83)	0.97	1.14 (7.56)	0.96	0.50 (2.66)	0.34	1.70 (5.30)	0.86	-8.42 (0.001)	<0.001	-8.42 (0.01)	<0.001
Intercept: PAQC score 3	-2.16 (0.62)	0.85	1.10 (3.56)	0.95	0.31 (2.65)	0.47	0.13 (2.10)	0.97	6.44 (0.002)	0.002	6.44 (0.01)	<0.001
Intercept: PAQC score 4	0.33 (0.36)	0.78	0.59 (3.29)	0.85	-0.76 (0.80)	0.34	1.20 (2.30)	0.90	-3.98 (0.001)	0.003	-3.98 (0.01)	0.002
Intercept: PAQC score 5	0.43 (0.16)	0.67	0.17 (2.36)	0.84	0.15 (0.09)	0.11	0.14 (3.10)	0.89	1.81 (0.001)	0.002	1.81 (0.01)	<0.001
Intercept: PAQC score 6	1.12 (0.50)	0.69	-1.20 (4.50)	0.86	-0.13 (0.11)	0.25	0.60 (5.30)	0.56	-0.26 (0.001)	0.03	-0.26 (0.01)	0.02
Age-group:12-59 months	-0.56 (0.63)	0.38	0.14 (0.09)	0.15	0.43 (0.23)	0.07	0.15 (0.09)	0.11	0.21 (0.96)	0.03	-0.56 (0.63)	0.04
Comorbidity 0	0.91 (0.82)	0.12	-0.13 (0.12)	0.27	0.23 (0.28)	0.41	-0.13 (0.11)	0.25	-1.96 (0.97)	0.04	-1.96 (0.97)	0.02
Comorbidity 1	2.04 (6.40)	0.99	-0.24 (0.14)	0.07	0.51 (0.35)	0.14	-0.20 (0.14)	0.14	-2.09 (1.15)	0.06	-2.09 (1.14)	0.21
Comorbidity 2	2.04 (3.70)	0.98	-0.69 (0.23)	0.002	1.73 (1.03)	0.09	-0.63 (0.23)	0.002	-2.03 (0.002)	0.10	-2.03 (0.01)	0.11
Malaria prevalence: low	-9.5 (6.24)	0.96	-0.39 (0.16)	0.01	0.85 (0.33)	0.01	-0.45 (0.15)	0.003	-10.13 (0.001)	0.01	-10.13 (0.01)	0.01
Hospital workload: low	1.63 (0.39)	0.68	0.16 (0.14)	0.01	0.97 (0.32)	0.02	0.17 (0.14)	0.023	-1.06 (1.08)	0.03	-1.06 (1.08)	0.02
Enhanced A&F arm	1.47 (2.34)	0.43	-1.19 (0.28)	<0.01	-1.09 (0.63)	<0.01	-0.48 (0.63)	0.44	0.21 (0.001)	0.03	0.21 (0.01)	<0.001
Time (months)	-0.34 (0.13)	0.07	-0.11 (0.04)	0.001	-0.09 (0.08)	0.28	-0.07 (0.08)	0.33	-0.48 (0.002)	0.03	-0.47 (0.01)	0.04
Time×Enhanced A&F arm	0.26 (0.40)	0.68	0.28 (0.05)	<0.01	0.11 (0.12)	0.32	0.02 (0.11)	0.79	-0.05 (0.001)	<0.001	-0.06 (0.001)	<0.001

Table B.2: Simulation results for random intercepts models under MCAR mechanism: Estimated bias in regression coefficients after multiple imputation of missing covariates and missing PAQC[a] score treatment domain subcomponents.

		Proportion Missing											
		3%				10%				40%			
Effect	True est[b]	Bias	Model-based SE	Emp SE[c]	MSE[d]	Bias	Model-based SE	Emp SE	MSE	Bias	Model-based SE	Emp SE	MSE
Intercept: PAQC score 1	-7.825	-0.191	0.047	0.047	0.039	-0.054	0.120	0.121	0.017	-0.031	0.152	0.153	0.024
Intercept: PAQC score 2	-2.253	-0.484	0.060	0.062	0.238	-0.580	0.378	0.377	0.479	-0.630	0.588	0.587	0.742
Intercept: PAQC score 3	-1.189	-0.269	0.009	0.010	0.072	-0.313	0.264	0.266	0.167	-0.341	0.379	0.380	0.260
Intercept: PAQC score 4	0.083	-0.266	0.102	0.101	0.081	-0.266	0.155	0.154	0.095	-0.281	0.131	0.132	0.096
Intercept: PAQC score 5	1.371	-0.240	0.447	0.447	0.257	-0.268	0.459	0.459	0.282	-0.292	0.258	0.259	0.152
Intercept: PAQC score 6	2.246	-0.152	0.362	0.363	0.154	-0.086	0.241	0.242	0.065	-0.082	0.244	0.245	0.066
Age-group:12-59	0.154	0.018	0.035	0.035	0.002	0.028	0.061	0.063	0.004	0.026	0.061	0.062	0.004
Child sex: males	-0.046	-0.002	0.159	0.160	0.025	-0.019	0.199	0.200	0.040	-0.025	0.216	0.217	0.047
Comorbidities: 0	0.474	-0.052	0.413	0.415	0.173	-0.084	0.329	0.331	0.115	-0.101	0.275	0.276	0.086
Comorbidities: 1	0.309	-0.043	0.122	0.123	0.017	-0.083	0.223	0.225	0.056	-0.103	0.284	0.285	0.091
Comorbidities: 2	0.335	-0.053	0.249	0.249	0.065	-0.077	0.315	0.316	0.105	-0.093	0.366	0.367	0.142
Clinicians' sex: female	0.337	-0.025	0.023	0.024	0.001	-0.041	0.065	0.065	0.006	-0.048	0.089	0.090	0.010
Clinicians' cadre: MO[e]	0.038	0.072	0.144	0.145	0.026	0.073	0.127	0.128	0.021	0.078	0.104	0.106	0.017
Hospital workload: low	-0.367	-0.035	0.079	0.080	0.007	-0.057	0.136	0.137	0.022	-0.066	0.167	0.168	0.032
Malaria prevalence: low	-0.189	0.113	0.785	0.786	0.628	0.105	0.790	0.791	0.635	0.108	0.813	0.814	0.672
Enhanced A&F[f]	-0.002	-0.030	0.130	0.131	0.018	-0.019	0.149	0.150	0.022	-0.014	0.159	0.160	0.025
Time (months)	-1.754	0.007	0.739	0.740	0.546	0.011	0.729	0.730	0.531	0.010	0.730	0.731	0.533
Time× Enhanced A&F	0.226	-0.012	0.022	0.023	0.001	-0.017	0.036	0.037	0.002	-0.019	0.044	0.045	0.002

[a]PAQC:- Paediatric admission quality of care, [b]True estimate, [c]Empirical standard error, [d]MSE:- Mean Square Error, [e]MO:- Medical Officer, [f]A&F:- Audit and feedback

Table B.3: Simulation results for random intercepts models under MCAR mechanism: Estimated bias in regression coefficients after multiple imputation of missing covariates and conventional methods in handling missing PAQC[a] score treatment domain subcomponents

		Proportion Missing											
		3%				10%				40%			
Effect	True est[b]	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	0.101	0.048	0.049	0.012	0.124	0.121	0.122	0.032	0.376	0.153	0.154	0.165
PAQC score intercept 2	-2.253	-0.316	0.061	0.062	0.104	-0.434	0.379	0.380	0.332	-0.544	0.589	0.590	0.643
PAQC score intercept 3	-1.189	-0.565	0.010	0.011	0.319	-0.72	0.265	0.266	0.589	-0.835	0.38	0.381	0.842
PAQC score intercept 4	0.083	-0.509	0.103	0.103	0.271	-0.612	0.156	0.157	0.399	-0.688	0.132	0.133	0.491
PAQC score intercept 5	1.371	-0.504	0.448	0.449	0.455	-0.616	0.460	0.461	0.591	-0.715	0.259	0.260	0.578
PAQC score intercept 6	2.246	-0.109	0.363	0.364	0.144	-0.158	0.242	0.243	0.084	-0.201	0.245	0.246	0.101
Age-group:12-59	0.154	0.038	0.036	0.037	0.003	0.124	0.062	0.063	0.019	0.164	0.062	0.063	0.031
Child sex: males	-0.046	-0.004	0.160	0.161	0.026	-0.044	0.201	0.202	0.042	-0.061	0.217	0.218	0.051
Comorbidities: 0	0.474	-0.109	0.414	0.414	0.183	-0.193	0.330	0.330	0.146	-0.247	0.276	0.277	0.137
Comorbidities: 1	0.309	-0.090	0.123	0.125	0.023	-0.191	0.224	0.224	0.087	-0.252	0.285	0.286	0.145
Comorbidities: 2	0.335	-0.111	0.250	0.251	0.075	-0.177	0.316	0.316	0.131	-0.228	0.367	0.368	0.186
Clinicians' sex: female	0.337	-0.053	0.024	0.025	0.003	-0.094	0.066	0.067	0.013	-0.118	0.09	0.10	0.022
Clinicians' cadre: MO[e]	0.038	0.151	0.145	0.146	0.044	0.168	0.128	0.129	0.045	0.191	0.105	0.106	0.048
Hospital workload: low	-0.367	-0.074	0.080	0.081	0.012	-0.131	0.137	0.138	0.036	-0.162	0.168	0.167	0.055
Malaria prevalence: low	-0.189	0.237	0.786	0.787	0.675	0.242	0.791	0.792	0.685	0.265	0.814	0.815	0.733
Enhanced A&F[d]	-0.002	-0.063	0.131	0.132	0.021	-0.044	0.151	0.152	0.025	-0.034	0.16	0.161	0.027
Time (months)	-1.754	0.015	0.740	0.741	0.548	0.025	0.732	0.734	0.534	0.025	0.731	0.732	0.535
Time× Enhanced A&F	0.226	-0.025	0.023	0.024	0.001	-0.039	0.037	0.038	0.003	-0.047	0.045	0.046	0.004

[a]PAQC:- Paediatric admission quality of care, [b]True estimate, [c]Empirical standard error, [d]MSE:- Mean Square Error, [e]MO:- Medical Officer, [f]A&F:- Audit and feedback

Table B.4: Simulation results for random intercepts models under MAR mechanism: Estimated bias in standard errors after multiple imputation of missing covariates and missing PAQC[a] score treatment domain subcomponents.

		Proportion Missing											
		3%				10%				40%			
Effect	True est[b]	Bias	Model based SE	Emp SE[c]	MS[d]d	Bias	Model based SE	Emp SE	MSE	Bias	Model based SE	Emp SE	MSE
PAQC score intercept 1	0.796	-0.007	0.223	0.224	0.050	-0.013	0.142	0.142	0.020	-0.013	0.093	0.094	0.009
PAQC score intercept 2	0.352	0.079	0.079	0.079	0.012	0.158	0.121	0.122	0.040	0.162	0.162	0.163	0.052
PAQC score intercept 3	0.349	0.056	0.056	0.057	0.006	0.112	0.106	0.107	0.024	0.116	0.116	0.116	0.027
PAQC score intercept 4	0.348	0.053	0.053	0.053	0.006	0.106	0.115	0.115	0.024	0.110	0.110	0.111	0.024
PAQC score intercept 5	0.350	0.043	0.143	0.144	0.022	0.086	0.109	0.110	0.019	0.089	0.118	0.119	0.022
PAQC score intercept 6	0.353	0.043	0.243	0.244	0.061	0.086	0.026	0.027	0.008	0.089	0.129	0.130	0.025
Age-group:12-59	0.089	0.007	0.270	0.271	0.073	0.015	0.085	0.086	0.007	0.015	0.085	0.086	0.007
Child sex: males	0.087	0.007	0.107	0.107	0.011	0.014	0.101	0.102	0.010	0.014	0.101	0.102	0.010
Comorbidities: 0	0.202	-0.007	0.12	0.121	0.014	-0.013	0.013	0.014	0.000	-0.014	0.114	0.114	0.013
Comorbidities: 1	0.203	-0.004	0.270	0.270	0.073	-0.009	0.068	0.069	0.005	-0.009	0.027	0.028	0.001
Comorbidities: 2	0.211	-0.011	0.099	0.099	0.010	-0.022	0.234	0.235	0.055	-0.023	0.243	0.243	0.060
Clinicians' sex: female	0.176	0.000	0.250	0.250	0.063	0.000	0.060	0.061	0.004	0.000	0.17	0.170	0.029
Clinicians' cadre: MO[e]	0.174	0.036	0.132	0.133	0.019	0.071	0.075	0.076	0.011	0.080	0.075	0.076	0.012
Hospital workload: low	0.176	0.022	0.024	0.025	0.001	0.044	0.026	0.027	0.003	0.045	0.047	0.047	0.004
Malaria prevalence: low	0.178	0.011	0.148	0.149	0.022	0.022	0.166	0.167	0.028	0.023	0.167	0.168	0.028
Enhanced A&F[f]	0.307	0.019	0.123	0.124	0.015	0.037	0.073	0.074	0.007	0.038	0.043	0.043	0.003
Time (months)	0.038	0.002	0.118	0.119	0.014	0.003	0.107	0.108	0.011	0.003	0.106	0.107	0.011
Time×Enhanced A&F	0.055	0.008	0.108	0.109	0.012	0.016	0.155	0.156	0.024	0.016	0.056	0.056	0.003

[a]PAQC:- Paediatric admission quality of care, [b]True estimate, [c]Empirical standard error, [d]MSE:- Mean Square Error, [e]MO:- Medical Officer, [f]A&F:- Audit and feedback

Table B.5: Simulation results for random intercepts models under MCAR mechanism: Estimated bias in standard errors after multiple imputation of missing covariates and missing PAQC[a] score treatment domain subcomponents.

		Proportion Missing											
		3%				10%				40%			
Effect	True est[b]	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	0.796	0.050	0.015	0.015	0.003	0.039	0.016	0.017	0.002	0.075	0.150	0.152	0.028
PAQC score intercept 2	0.352	0.013	0.173	0.174	0.030	0.014	0.054	0.056	0.003	0.015	0.073	0.074	0.006
PAQC score intercept 3	0.349	0.070	0.070	0.070	0.010	0.083	0.081	0.081	0.013	0.103	0.083	0.084	0.017
PAQC score intercept 4	0.348	0.060	0.180	0.181	0.036	0.066	0.059	0.060	0.008	0.096	0.068	0.069	0.014
PAQC score intercept 5	0.350	0.050	0.249	0.250	0.065	0.075	0.091	0.092	0.014	0.010	0.090	0.091	0.008
PAQC score intercept 6	0.353	0.050	0.375	0.375	0.143	0.064	0.059	0.059	0.008	0.012	0.058	0.059	0.004
Age-group:12-59	0.089	0.020	0.056	0.057	0.004	0.035	0.059	0.060	0.005	0.051	0.049	0.049	0.005
Child sex: males	0.087	0.020	0.115	0.117	0.014	0.033	0.116	0.117	0.015	0.037	0.118	0.118	0.015
Comorbidities: 0	0.202	0.010	0.256	0.257	0.066	0.029	0.062	0.063	0.005	0.032	0.140	0.140	0.021
Comorbidities: 1	0.203	0.020	0.176	0.177	0.031	0.027	0.025	0.025	0.001	0.031	0.027	0.028	0.002
Comorbidities: 2	0.211	0.010	0.320	0.321	0.103	0.024	0.032	0.033	0.002	0.028	0.033	0.033	0.002
Clinicians' sex: female	0.176	0.040	0.031	0.032	0.003	0.034	0.031	0.031	0.002	0.042	0.037	0.038	0.003
Clinicians' cadre: MO[e]	0.174	0.050	0.087	0.088	0.010	0.062	0.088	0.089	0.012	0.067	0.086	0.086	0.012
Hospital workload: low	0.176	0.010	0.280	0.280	0.079	0.022	0.061	0.061	0.004	0.028	0.080	0.080	0.007
Malaria prevalence: low	0.178	0.030	0.131	0.131	0.018	0.041	0.132	0.132	0.019	0.054	0.130	0.130	0.020
Enhanced A&F[f]	0.307	0.02	0.027	0.028	0.001	0.043	0.074	0.075	0.007	0.047	0.060	0.060	0.006
Time (months)	0.038	0.021	0.126	0.126	0.016	0.033	0.127	0.127	0.017	0.041	0.123	0.124	0.018
Time× Enhanced A&F	0.055	0.010	0.210	0.211	0.044	0.023	0.052	0.053	0.003	0.026	0.091	0.091	0.009

[a]PAQC:- Paediatric admission quality of care, [b]True estimate, [c]Empirical standard error, [d]MSE:- Mean Square Error, [e]MO:- Medical Officer, [f]A&F:- Audit and feedback

Table B.6: Simulation results for random intercepts models under MAR mechanism: Estimated bias in standard errors after multiple imputation of missing covariates and conventional methods in handling missing PAQC[a] score treatment domain subcomponents

		Proportion Missing											
		3%				10%				40%			
Effect	True est[b]	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	0.796	0.184	0.184	0.184	0.068	0.230	0.230	0.230	0.106	0.249	0.049	0.049	0.064
PAQC score intercept 2	0.352	0.876	0.028	0.028	0.768	0.092	0.092	0.092	0.017	0.186	0.106	0.106	0.046
PAQC score intercept 3	0.349	0.738	0.094	0.094	0.553	0.920	0.220	0.220	0.895	0.199	0.045	0.045	0.042
PAQC score intercept 4	0.348	0.786	0.049	0.049	0.620	0.280	0.280	0.280	0.157	0.165	0.065	0.065	0.031
PAQC score intercept 5	0.35	0.708	0.011	0.011	0.501	0.682	0.182	0.182	0.498	0.238	0.058	0.058	0.060
PAQC score intercept 6	0.353	0.728	0.013	0.013	0.530	0.208	0.108	0.108	0.055	0.186	0.216	0.216	0.081
Age-group:12-59	0.089	0.280	0.028	0.028	0.079	0.362	0.162	0.162	0.157	0.421	0.421	0.421	0.354
Child sex: males	0.087	0.036	0.015	0.015	0.002	0.251	0.159	0.159	0.088	0.278	0.163	0.163	0.104
Comorbidities: 0	0.202	0.122	0.013	0.013	0.015	0.045	0.161	0.161	0.028	0.049	0.174	0.174	0.033
Comorbidities: 1	0.203	0.005	0.022	0.022	0.001	0.128	0.017	0.017	0.017	0.139	0.014	0.014	0.020
Comorbidities: 2	0.211	0.096	0.059	0.059	0.013	0.152	0.087	0.087	0.031	0.165	0.103	0.103	0.038
Clinicians' sex: female	0.176	0.103	0.013	0.013	0.011	0.010	0.157	0.157	0.025	0.013	0.168	0.168	0.028
Clinicians' cadre: MO[e]	0.174	0.208	0.012	0.012	0.043	0.124	0.162	0.162	0.042	0.140	0.189	0.189	0.055
Hospital workload: low	0.176	0.044	0.046	0.046	0.004	0.068	0.070	0.070	0.010	0.102	0.104	0.104	0.021
Malaria prevalence: low	0.178	0.038	0.102	0.102	0.012	0.103	0.091	0.091	0.019	0.140	0.061	0.061	0.023
Enhanced A&F[f]	0.307	0.163	0.043	0.043	0.028	0.049	0.288	0.288	0.085	0.079	0.519	0.519	0.276
Time (months)	0.038	0.083	0.046	0.046	0.009	0.218	0.026	0.026	0.048	0.249	0.011	0.011	0.062
Time× Enhanced A&F	0.055	0.081	0.081	0.081	0.013	0.101	0.101	0.101	0.020	0.110	0.210	0.210	0.056

[a]PAQC:- Paediatric admission quality of care, [b]True estimate, [c]Empirical standard error, [d]MSE:- Mean Square Error, [e]MO:- Medical Officer, [f]A&F:- Audit and feedback

Table B.7: Simulation results for random intercepts models under MCAR mechanism: Estimated bias in standard errors after multiple imputation of missing covariates and conventional methods in handling missing PAQC[a] score treatment domain subcomponents

		Proportion Missing											
		3%				10%				40%			
Effect	True est[b]	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	0.796	0.177	0.112	0.112	0.051	0.201	0.102	0.102	0.051	0.223	0.123	0.123	0.065
PAQC score intercept 2	0.352	0.342	0.232	0.232	0.196	0.126	0.958	0.958	0.934	0.059	0.019	0.019	0.004
PAQC score intercept 3	0.349	0.509	0.209	0.209	0.365	0.211	0.807	0.807	0.696	0.219	0.112	0.112	0.061
PAQC score intercept 4	0.348	0.356	0.221	0.221	0.205	0.256	0.860	0.860	0.805	0.245	0.151	0.151	0.083
PAQC score intercept 5	0.350	0.480	0.248	0.248	0.349	0.124	0.127	0.127	0.032	0.236	0.116	0.116	0.069
PAQC score intercept 6	0.353	0.528	0.210	0.210	0.390	0.190	0.236	0.236	0.092	0.480	0.118	0.118	0.244
Age-group:12-59	0.089	0.269	0.139	0.139	0.110	0.317	0.317	0.317	0.201	0.376	0.176	0.176	0.172
Child sex: males	0.087	0.200	0.149	0.149	0.070	0.220	0.153	0.153	0.072	0.248	0.158	0.158	0.086
Comorbidities: 0	0.202	0.035	0.126	0.126	0.006	0.039	0.142	0.142	0.022	0.044	0.157	0.157	0.027
Comorbidities: 1	0.203	0.099	0.022	0.022	0.012	0.113	0.018	0.018	0.013	0.124	0.016	0.016	0.016
Comorbidities: 2	0.211	0.117	0.055	0.055	0.020	0.133	0.072	0.072	0.023	0.148	0.088	0.088	0.030
Clinicians' sex: female	0.176	0.005	0.128	0.128	0.001	0.009	0.142	0.142	0.020	0.011	0.153	0.153	0.024
Clinicians' cadre: MO[e]	0.174	0.092	0.111	0.111	0.019	0.109	0.131	0.131	0.029	0.125	0.159	0.159	0.041
Hospital workload: low	0.176	0.042	0.024	0.024	0.003	0.060	0.052	0.052	0.006	0.091	0.093	0.093	0.017
Malaria prevalence: low	0.178	0.080	0.104	0.104	0.015	0.091	0.097	0.097	0.018	0.125	0.069	0.069	0.020
Enhanced A&F[f]	0.307	0.036	0.426	0.426	0.017	0.043	0.461	0.461	0.214	0.071	0.493	0.493	0.248
Time (months)	0.038	0.157	0.049	0.049	0.032	0.191	0.038	0.038	0.038	0.223	0.004	0.004	0.050
Time× Enhanced A&F	0.055	0.078	0.018	0.018	0.007	0.089	0.122	0.122	0.023	0.098	0.098	0.098	0.019

[a]PAQC:- Paediatric admission quality of care, [b]True estimate, [c]Empirical standard error, [d]MSE:- Mean Square Error, [e]MO:- Medical Officer, [f]A&F:- Audit and feedback

Table B.8: Monte-Carlo standard errors and confidence intervals for estimated bias in regression parameters across simulation scenarios: MAR mechanism.

	Multiple imputation of covariates and outcome subcomponents in the treatment domain			Multiple imputation of covariates and conventional method in handling outcome subcomponents in the treatment domain		
	Proportion missing			Proportion missing		
	3%	10%	40%	3%	10%	40%
Effect	Monte-Carlo SE (MC[a] 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)
PAQC score intercept 1	0.008 (0.01, 0.03)	0.007 (0.01,0.03)	0.025 (0.03,0.07)	0.006 (0.12, 0.15)	0.005 (0.16, 0.18)	0.006 (0.47,0.50)
PAQC score intercept 2	0.016 (-0.65, -0.58)	0.017 (-0.74, -0.67)	0.035 (-0.81, -0.67)	0.005 (-0.4, -0.36)	0.001 (-0.71, -0.66)	0.02 (-0.85, -0.77)
PAQC score intercept 3	0.01 (-0.35, -0.31)	0.01 (-0.39, -0.36)	0.01 (-0.41, -0.37)	0.004 (-0.74, -0.72)	0.002 (-0.81, -0.79)	0.019 (-0.94, -0.86)
PAQC score intercept 4	0.001 (-0.25, -0.22)	0.01 (-0.28, -0.26)	0.02 (-0.31, -0.26)	0.02 (-0.56, -0.52)	0.001 (-0.67, -0.66)	0.003 (-0.8, -0.77)
PAQC score intercept 5	0.011 (-0.29, -0.24)	0.01 (-0.32, -0.28)	0.01 (-0.34, -0.30)	0.003 (-0.62, -0.57)	0.002 (-0.73.-0.72)	0.004 (-0.84, -0.79)
PAQC score intercept 6	0.02 (-0.08, -0.01)	0.02 (-0.09, -0.01)	0.02 (-0.09, -0.01)	0.037 (-0.16, -0.02)	0.036 (-0.18, -0.04)	0.035 (-0.25, -0.12)
Age-group:12-59	0.001 (0.02,0.04)	0.001 (0.04,0.05)	0.001 (0.04,0.05)	0.002 (0.07,0.10)	0.002 (0.10, 0.11)	0.003 (0.17,0.18)
Child sex: males	0.001 (-0.03, -0.02)	0.004 (-0.04, -0.02)	0.006 (-0.04, -0.02)	0.003 (-0.07, -0.05)	0.002 (-0.08, -0.07)	0.003 (-0.13, -0.12)
Comorbidities: 0	0.003 (-0.14, -0.12)	0.003 (-0.15, -0.14)	0.003 (-0.16, -0.15)	0.001 (-0.31, -0.29)	0.001 (-0.36, -0.34)	0.001 (-0.60, -0.58)
Comorbidities: 1	0.001 (-0.14, -0.13)	0.001 (-0.16, -0.15)	0.006 (-0.18, -0.15)	0.001 (-0.32, -0.29)	0.001 (-0.42, -0.32)	0.002 (-0.62, -0.59)
Comorbidities: 2	0.001 (-0.13, -0.10)	0.001 (-0.13, -0.11)	0.005 (-0.15, -0.11)	0.002 (-0.28, -0.22)	0.001 (-0.31, 0.29)	0.002 (-0.51, -0.49)
Clinicians' sex: female	0.002 (-0.05, -0.02)	0.005 (-0.06, -0.04)	0.002 (-0.09, -0.07)	0.007 (-0.01, -0.004)	0.001 (-0.01, -0.001)	0.002 (-0.02, -0.01)
Clinicians' cadre: MO[c]	0.001 (0.06,0.08)	0.01 (0.07,0.09)	0.001 (0.07,0.08)	0.004 (0.11, 0.17)	0.002 (0.16,0.18)	0.002 (0.27,0.29)
Hospital workload: low	0.004 (-0.07, -0.06)	0.004 (-0.08, -0.06)	0.004 (-0.08, -0.07)	0.003 (-0.16, -0.12)	0.002 (-0.18, -0.17)	0.002 (-0.29, -0.27)
Malaria prevalence: low	0.004 (0.16,0.14)	0.004 (0.17,0.19)	0.002 (0.17,0.21)	0.004 (0.33, 0.37)	0.001 (0.41,0.44)	0.002 (0.72, -0.73)
Enhanced A&F[d]	0.001 (-0.07, -0.06)	0.002 (-0.06, -0.05)	0.002 (-0.07, -0.05)	0.001 (-0.007, -0.002)	0.003 (-0.01, -0.002)	0.002 (-0.019, -0.01)
Time (months)	0.003 (0.01,0.02)	0.003 (0.01,0.02)	0.008 (0.01,0.02)	0.01 (0.029, 0.04)	0.002 (0.039, 0.042)	0.003 (0.06,0.07)
Time× Enhanced A&F	0.001 (-0.03, -0.01)	0.002 (-0.04, -0.03)	0.001 (-0.04, -0.03)	0.004 (-0.069, -0.058)	0.001(-0.08, -0.06)	0.001 (-0.13, -0.11)

[a] MC:-Monte-Carlo, [b]PAQC:- Paediatric admission quality of care, [c]MO:- Medical Officer, [d]A&F:-Audit and feedback

Table B.9: Monte-Carlo standard errors and confidence intervals for estimated bias in regression parameters across simulation scenarios: MCAR mechanism.

	Multiple imputation of covariates and outcome subcomponents in the treatment domain			Multiple imputation of covariates and conventional method in handling outcome subcomponents in the treatment domain		
	Proportion missing			Proportion missing		
	3%	10%	40%	3%	10%	40%
Effect	Monte-Carlo SE (MC[a] 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)	Monte-Carlo SE (MC 95% CI for bias)
PAQC[b] score intercept 1	0.018 (-0.23, -0.16)	0.007 (-0.07, -0.04)	0.025 (-0.06,0.01)	0.013 (0.08,0.13)	0.018 (0.09,0.16)	0.018 (0.34,0.41)
PAQC score intercept 2	0.014 (-0.51, -0.46)	0.016 (-0.61, -0.55)	0.034 (-0.7, -0.56)	0.012 (-0.34, -0.29)	0.016 (-0.47, -0.41)	0.032 (-0.61, -0.48)
PAQC score intercept 3	0.01 (-0.29, -0.25)	0.009 (-0.33, -0.27)	0.009 (-0.36, -0.32)	0.014 (-0.59, -0.54)	0.01 (-0.74, -0.67)	0.007 (-0.85, -0.82)
PAQC score intercept 4	0.001 (-0.27, -0.25)	0.001 (-0.27, -0.25)	0.001 (-0.30, -0.26)	0.003 (-0.51, -0.49)	0.006 (-0.62, -0.58)	0.006 (-0.7, -0.66)
PAQC score intercept 5	0.011 (-0.26, -0.22)	0.008 (-0.28, -0.25)	0.01 (-0.31, -0.27)	0.006 (-0.52, -0.49)	0.006 (-0.63, -0.6)	0.006 (-0.73, -0.69)
PAQC score intercept 6	0.018 (-0.19, -0.12)	0.019 (-0.12, -0.05)	0.02 (-0.12, -0.04)	0.019 (-0.15, -0.07)	0.018 (-0.19, -0.12)	0.018 (-0.24, -0.17)
Age-group:12-59	0.001 (0.01,0.03)	0.001 (0.03,0.025)	0.001 (0.02,0.03)	0.001 (0.04,0.036)	0.002 (0.11,0.13)	0.003 (0.15,0.17)
Child sex: males	0.001 (-0.004,-0.01)	0.001 (-0.02, -0.01)	0.001 (-0.03, -0.01)	0.001 (-0.005, -0.02)	0.001 (-0.05, -0.03)	0.001 (-0.07, -0.05)
Comorbidities: 0	0.005 (-0.06, -0.04)	0.004 (-0.09, -0.07)	0.004 (-0.12, -0.09)	0.003 (-0.11, -0.09)	0.002 (-0.2, -0.18)	0.002 (-0.25, -0.23)
Comorbidities: 1	0.002 (-0.05, -0.04)	0.002 (-0.09, -0.08)	0.001 (-0.13, -0.09)	0.002 (-0.11, -0.07)	0.001 (-0.21, -0.18)	0.001 (-0.26, -0.24)
Comorbidities: 2	0.001 (-0.06, -0.05)	0.001(-0.10, -0.08)	0.001 (-0.11, -0.08)	0.002 (-0.13, -0.10)	0.001 (-0.19, -0.16)	0.001 (-0.24, -0.21)
Clinicians' sex: female	0.002 (-0.03, -0.02)	0.002 (-0.06, -0.04)	0.002 (-0.05, -0.03)	0.001 (-0.06, -0.047)	0.001 (-0.1, -0.09)	0.001 (-0.13, -0.10)
Clinicians' cadre: MO[c]	0.001 (0.06,0.08)	0.001 (0.06,0.08)	0.001 (0.06,0.10)	0.002 (0.14,0.17)	0.001 (0.17,0.15)	0.002 (0.18,0.21)
Hospital workload: low	0.004 (-0.04, -0.03)	0.003 (-0.06, -0.04)	0.004 (-0.07, -0.05)	0.004 (-0.08, -0.067)	0.004 (-0.14, -0.12)	0.004 (-0.17, -0.15)
Malaria prevalence: low	0.001 (0.10,0.12)	0.001 (0.09,0.12)	0.001 (0.09,0.13)	0.001 (0.22,0.26)	0.001 (0.23,0.27)	0.001 (0.25,0.28)
Enhanced A&F[d]	0.002 (-0.04, -0.02)	0.002 (-0.02, -0.01)	0.003 (-0.02, -0.01)	0.001 (-0.07, -0.05)	0.001 (-0.05, -0.03)	0.002(-0.05, -0.02)
Time (months)	0.003 (0.001,0.01)	0.003 (0.01,0.02)	0.003 (0.008,0.08)	0.004 (0.01,0.02)	0.003 (0.01,0.03)	0.003 (0.02,0.03)
Time× Enhanced A&F	0.002 (-0.02, -0.01)	0.001 (-0.024,-0.01)	0.001 (-0.04, -0.01)	0.002 (-0.03, -0.02)	0.001 (-0.05, -0.03)	0.001 (-0.06, -0.04)

[a] MC:-Monte-Carlo, [b]PAQC:- Paediatric admission quality of care, [c]MO:- Medical Officer, [d]A&F:-Audit and feedback

C Chapter 5 Appendices

Control arm and high paediatric admission workload hospital

	Clinician's sex	
	Females	Males
Data prediction (%)	47	53
Your prediction (%)		
Minimum	0	0
Maximum	100	100

Control arm and low paediatric admission workload hospital

	Clinician's sex	
	Females	Males
Data prediction (%)	36	64
Your prediction (%)		
Minimum	0	0
Maximum	100	100

Intervention arm and high paediatric admission workload hospital

	Clinician's sex	
	Females	Males
Data prediction (%)	57	46
Your prediction (%)		
Minimum	0	0
Maximum	100	100

Intervention arm and low paediatric admission workload hospital

	Clinician's sex	
	Females	Males
Data prediction (%)	42	58
Your prediction (%)		
Minimum	0	0
Maximum	100	100

Figure C.1: Questionnaire tables used to elicit experts' opinions about missing clinician's sex

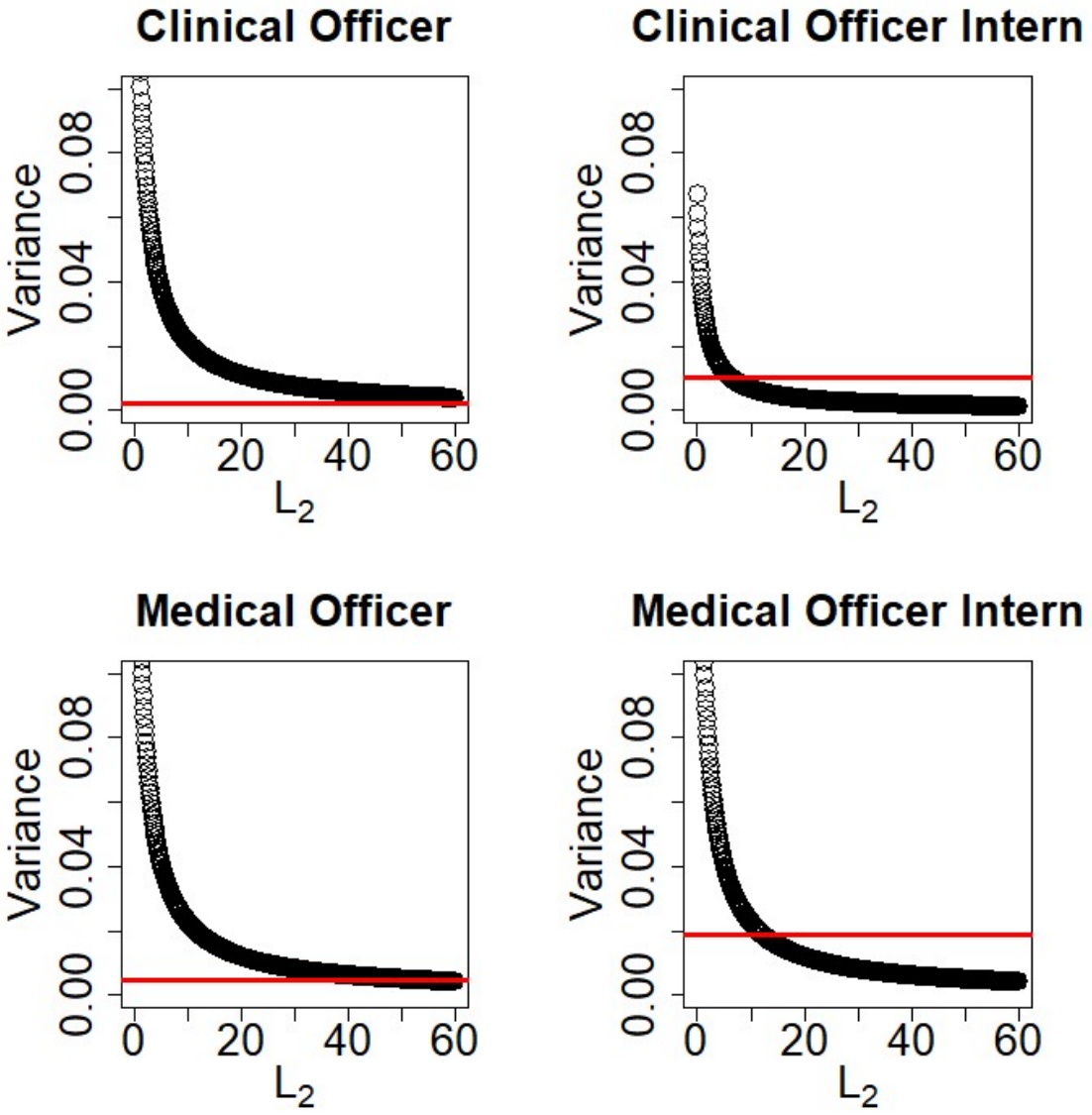

Figure C.2: Clinician's cadre Dirichlet variances (curves) and experts' variances (horizontal red lines) for k=2

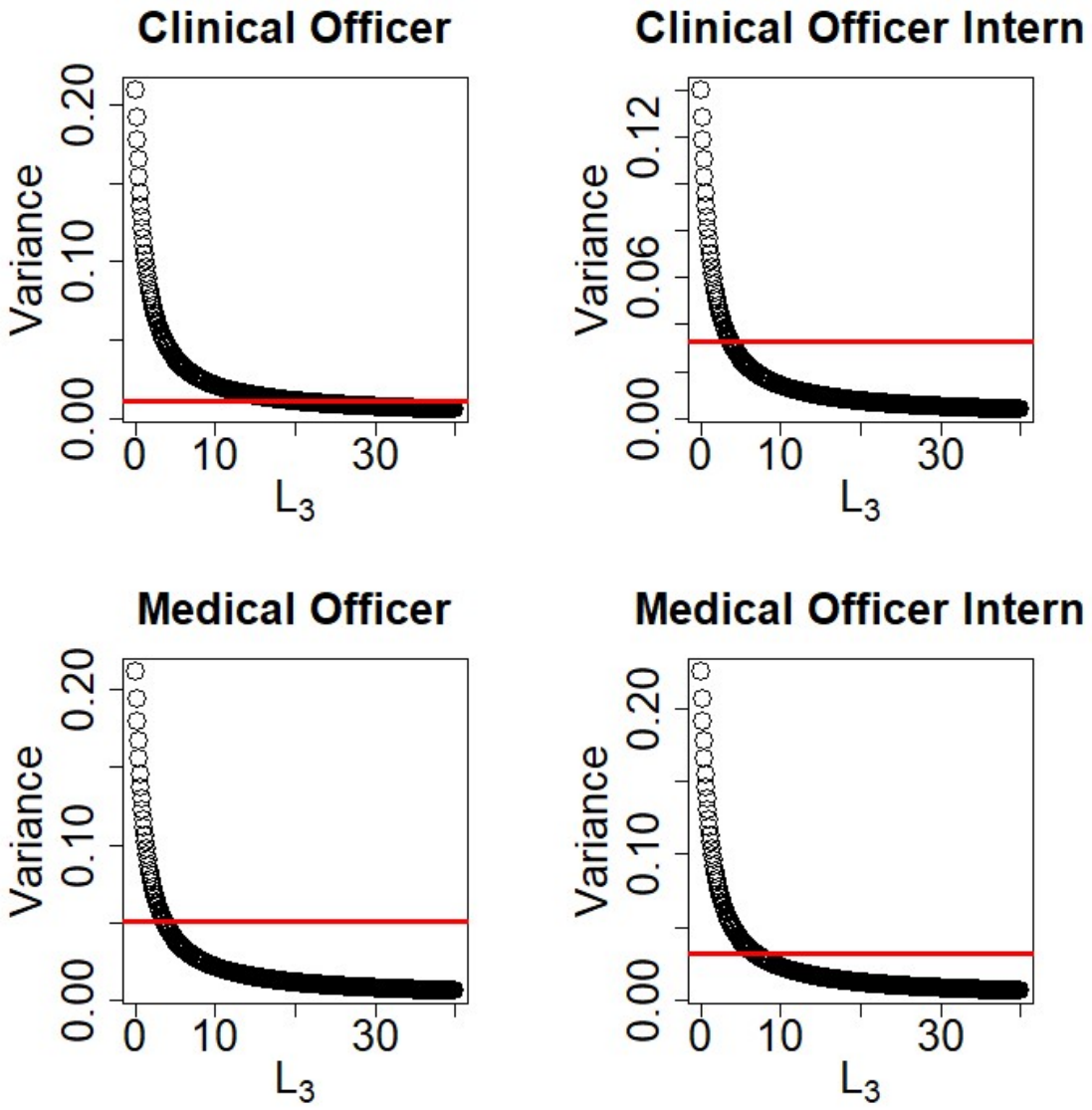

Figure C.3: Clinician's cadre Dirichlet variances (curves) and experts' variances (horizontal red lines) for k=3

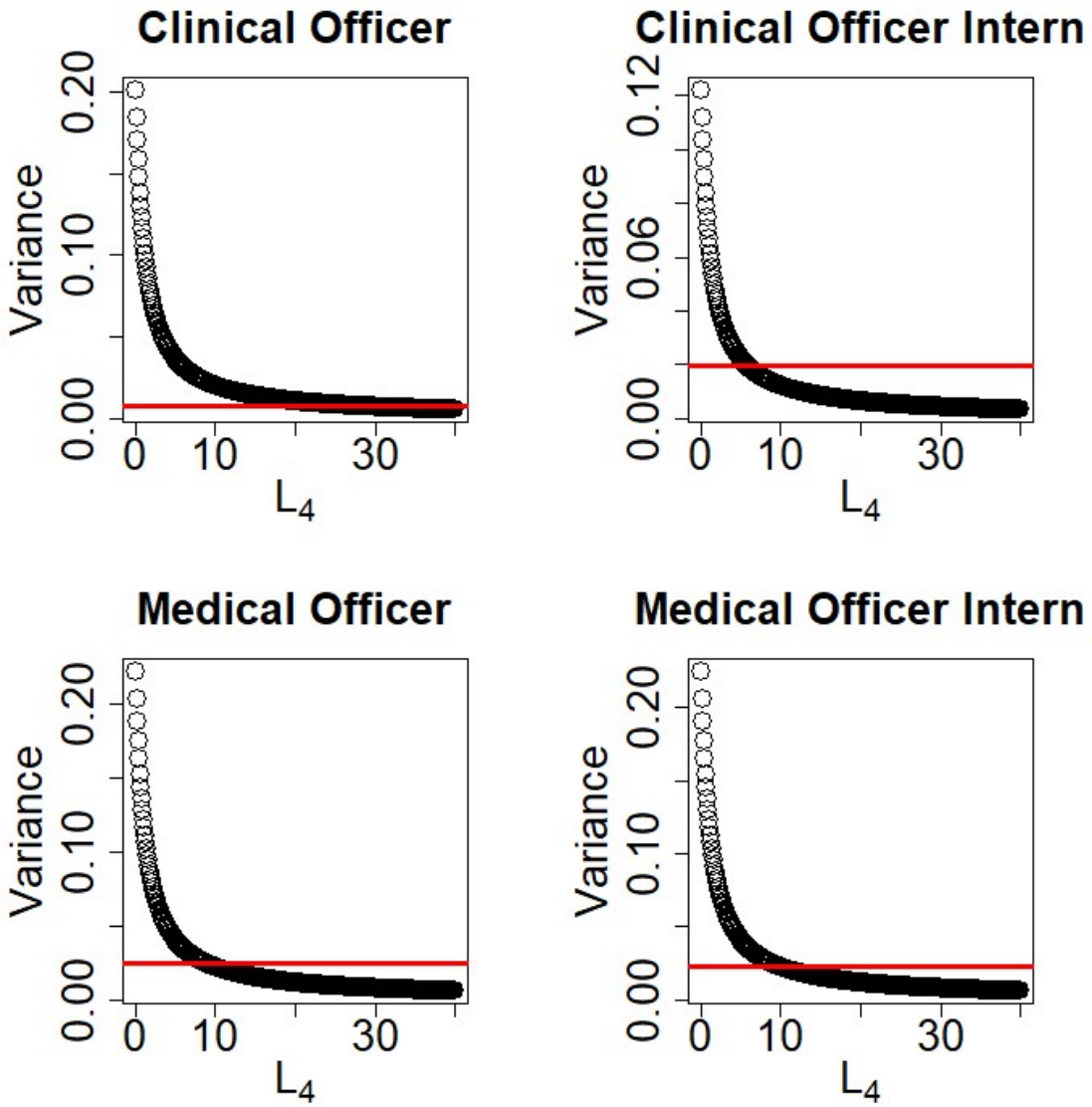

Figure C.4: Clinician's cadre Dirichlet variances (curves) and experts' variances (horizontal red lines) for k=4