University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Biometry

Random Forests application in missing data and predictive modelling for hierarchical routine clinical data: A case stud of childhood pneumonia in Kenya

Research Report in Mathematics, Number 16, 2019

Steven Wambua

June 2019

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

Random Forests application in missing data and predictive modelling for hierarchical routine clinical data: A case study of childhood pneumonia in Kenya

Research Report in Mathematics, Number 16, 2019

Steven Wambua

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

Health stakeholders usually need complete, accurate and reliable estimates of various health outcomes to make decisions on improving health care delivery. Missing observations especially in clinical routine data is one of major setbacks in evaluating public health problems efficiently. One of the tools used to measure clinical quality is the Paediatric Admission Quality of Care (PAQC) score. We seek to identify factors that influence clinical quality in this study after dealing with missing values.The main objective of this study is to identify key determinants of Pediatric Admission Quality of Care (PAQC) score using Random Forests.

Data on a total of 2027 children between 2 and 59 months who were admitted in selected county hospitals in Kenya was used. The data contained clinical data from admission to treatment. Random forests missForest package was used to impute missing data. Cumulative logit mixed models were fit with PAQC score as an outcome and age, sex, comorbidity, weight, clinician sex and cadre, hospital workload, malaria prevalence, intervention arm and time of admission as predictors to determine the significant determinants of clinical quality. The models were nested within both hospital and clinician levels. Both Random forests and conditional random forests were used to determine variable importance.

The cumulative logit mixed model nested within both clinician and hospital level was selected based on AIC. Weight of the child, clinician sex, cadre and the time of admission were significant determinants of PAQC score based on the P values at 0.05 level of significance. A unit increase in weight increases the probability of a higher PAQC score by 0.06, while being attended by a medical officer relative to a clinical officer increases the probability by 0.27. The time of admission increases the probability by 0.11. On the other hand, PAQC scores would be lower if the clinician was male. The probability of a reduced PAQC score if a clinician is male is 0.49. Month, weight, intervention arm and hospital workload were the most important variables in predicting the quality of care while age and the number of comorbidities were the least important using Random forests models.

Based on the cumulative logit mixed models, the study concludes that hospital level, weight of the child, clinician sex, cadre and the time of admission are key determinants of PAQC score. On the other hand, age and the number of comorbidities for a given patient may not strongly influence the quality of care provided based on the random forests models. The mechanisms around these associations however need to be studied extensively. Pneumonia

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

Steven Wambua
Reg No.I56/9313/2019

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

Signature Date

Dr Nelson Owuor
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: onyango@uonbi.ac.ke

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

Signature Date

Susan Gachau
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: sgachau@kemri-wellcome.org

Dedication

This research project is dedicated to my family and friends for their efforts, contribution and sacrifice to ensure completion of the work.

Contents

References 35

Acknowledgments

Firstly, I would like to thank SSACAB under the DELTAS Africa Programme for enabling me carry out this research project. Secondly, I thank KEMRI - Wellcome Trust Research Programme for the data towards this study. Thirdly, I heartily thank and appreciate my supervisors Dr. Nelson Owuor and Dr. Sam Aketch for their encouragement and tireless effort in ensuring that my work was successful and being my mentors. God bless you and your families. I would also like to thank my collaborator Susan Gachau. May the Lord bless you all and be with you.....

Steven Wambua

Nairobi, 2019.

1 Introduction

1.1 Introduction

This chapter presents the following: Background on use of random survival forests in imputing missing observations, the importance of Pediatric Admission Quality of Care Score (PAQC) in measuring quality of health care delivery, statement of the problem, objectives and justification for the study.

1.2 Background

Health stakeholders usually need complete, accurate and reliable estimates of various health outcomes to make decisions on improving health care delivery(Kaplan & Frosch, 2005). Missing observations especially in clinical routine data is one of major set - backs in evaluating public health problems efficiently. Being an important aspect for any research to consider, researchers have come up with various methods to handle missing data(De Silva, Moreno-Betancur, De Livera, Lee, & Simpson, 2017; Powney, Williamson, Kirkham, & Kolamunnage-Dona, 2014; Sullivan, White, Salter, Ryan, & Lee, 2018; Y. Zhang et al., 2017) based on the nature of missing observations(Boyko, 2013; Nakagawa & Freckleton, 2008). Unfortunately, there has been no consensus on the best technique to handling missing data(Cheema, 2014).

A recent systematic review (Li et al., 2014) on best standards in handling missing data especially in Patient Centered Outcomes research dealt more on best practices in estimation of missing data. The study involved 1790 guidance documents that had formal recommendations regarding missing data. The study concluded that researchers need to adopt extremely thorough and careful methods in estimation of missing data through techniques that promote good science. Researchers need to prioritize use of existing guidelines on handling missing data in order to provide useful research conclusions.

Understanding good practices and creating standards for the prevention and handling of missing data can help to improve the translation of the research into complete, accurate, and reliable evidence for health care decision-making.

Missing data present various problems. First, the absence of data reduces statistical power, which refers to the probability that the test will reject the null hypothesis when it is false. Second, the lost data can cause bias in the estimation of parameters. Third, it can reduce the representativeness of the samples. Fourth, it may complicate the analysis of the study. Each of these distortions may threaten the validity of the trials and can lead to invalid conclusions(Kang, 2013).

1.2.1 PAQC score

Health systems usually seek to ensure quality of healthcare at all levels is up to set standards of practice. This calls for an evaluation on delivery of healthcare in various health facilities. One of the tools used for this exercise is the Paediatric Admission Quality of Care (PAQC) score(Opondo, Allen, Todd, & English, 2016). This evaluation tool measures compliance to guidelines on patient admissions, assessment, diagnosis and treatment for various illnesses. It measures full compliance for the various set out to – do lists for each illness in the three domains of the score; Assessment, diagnosis and treatment. This score would thus indicate whether healthcare delivery is up to standards or needs urgent interventions to reduce morbidity and mortality that may occur due to inappropriate healthcare at any point on the three domains.
This score is obtained through a systematic process where data is obtained from various health facilities from health records. The set guidelines are used as the gold standards to see whether clinicians complied with them. A score is then obtained based on the adherence to the set standards of practice. Most data from hospital records is sparse. Entries are missing due to ignorance or lack of knowledge of the guidelines. This would thus be an hindrance to an accurate and generalizable score. Researchers therefore need to find ways of imputing the missing data accurately. Due to the hierarchical nature of this data, this study seeks to use a machine learning technique to impute the missing values in a pediatric routine data while taking into account clustering due to the various levels the data was obtained.

1.2.2 Common methods used in handling missing data

Most researchers use two common techniques in handling missing observations. These are removing observations(Langkamp, Lehman, & Lemeshow, 2010) where there are missing values in any variable in order to fit statistical models and imputation methods(Bertsimas, Pawlowski, & Zhuo, 2017; Bertsimas, Orfanoudaki, & Pawlowski, 2018). Although these methods are widely used, there are various shortcomings. For instance removing observations has been associated with producing biased parameters and estimates. In some cases the researcher may opt to delete the entire variable with missing values. This should only be done where data is missing for more than 60 observations but only if that variable is insignificant. To deal with some of these disadvantages associated with dropping observations and variables, imputation methods have been recommended. These include use of mean, median and mode, linear regression and multiple imputation. Although use of measures of tendency is fast, mean imputation for instance has been linked to reduced variance in the dataset. While linear regression(Beyad & Maeder, 2013; Karama, Farouk, & Atiya, 2018) has proved theoretically to provide good estimates of missing values, it relies on the assumption that the variables used in the regression equation are linearly related which may not hold in most cases. Moreover the replaced values are as a result of prediction by other variables hence they tend to fit 'too well'. This deflates the standard

error and thus biased estimates.
Multiple imputation(Royston, 2004) has however been a breakthrough in the field of missing data in the recent past. It's use of the Markov Chain Monte Carlo (MCMC) simulation and pooling of analysis results has been shown to improve efficiency an accuracy(Lin, 2010). It has been shown to provide unbiased estimates which are more valid than other ad hoc methods in estimation of missing values. It's easy to use due to the availability of various algorithms already developed in various standard statistical softwares and it preserves the sample size through use of all the available data. Moreover the results are readily interpreted and preserves the statistical power(McCleary, 2002). However, this method is limited to analytical models without interactions and where the proportion of missing data is not too large(Jakobsen, Gluud, Wetterslev, & Winkel, 2017). These shortcomings could be largely checked through use of machine learning techniques developed recently. These include the use of K Nearest Neighbours(García-Laencina, Sancho-Gómez, Figueiras-Vidal, & Verleysen, 2009), XGBoost(Chen & Guestrin, 2016) and Random Forests(Shah, Bartlett, Carpenter, Nicholas, & Hemingway, 2014; Stekhoven & Bühlmann, 2011) algorithms.

1.2.3 Machine Learning Techniques in handling missing data

Random Forests algorithm

Also known as random decision forests, Random Forests is a non – parametric supervised learning ensemble method used for classification and regression(Breiman, 2001). It implements these tasks through building predictive models through multiple learning algorithms. To obtain the best possible result, RF constructs a whole forest of random unassociated decision trees to attain at the best possible solution.
This method uses Breiman's algorithm in handling missing data. It has since developed into various different algorithms for imputing missing observations. They are an improvement to Breiman's earlier algorithm, the randomForest package (RColorBrewer & Liaw, 2018) that is provided in R. Advancements from this algorithm include the 'on – the – fly – imputation' method(Tang, 2017) in the randomSurvivalForest R-package. This approach imputes data by concurrently growing a survival tree. A generalization of the above two algorithms is the randomForest SRC package which includes classification and regression(Tang, 2017). The latest RF algorithm for handling missing data is the missForest package(Stekhoven, 2011) which imputes missing data using a prediction framework. It involves regressing each predictor against all the other predictors and the missing values for the response variable are predicted using the fitted forest(Stekhoven, 2015).
Recent studies have sought to compare performance among the different RF algorithms. A study for instance showed that in situations where correlation is high, missForest algorithm performed best(Tang, 2017). However in big data settings, where computational speed is a major factor to consider, mForest is the most efficient. According to the study,

mForest achieved upto a 10 – fold reduction in computed time relative to missForest.
Studies have also been comparison of RF with other machine leraning methods, for instance(Waljee et al., 2013), and another study that looked at performance of a compination of RF and other methods. In this study, imputed estimates had reduced bias when using RF and MICE (multivariate imputation using chaned equation) together, where performance was assessed using computation speed and imputation accuracy(Shah et al., 2014).

K Nearest Neighbors (KNN)

This is another non – parametric supervised learning algorithm for classification and regression. KNN assumes that identical things occur in close proximity and chooses k neighbors to each data point using the Euclidian distance(C. Zhang, Kai, Feng, & Yang, 2013).To obtain the appropriate k, the algorithm is run several times for different values of k. The value of k that reduces the number of errors and maintains accuracy in prediction of the method given new data is chosen.
To impute missing values, k neighbors are selected based on a distance metric, commonly the Euclidian distance, from the missing value. The mean of the selected k nearest neighbors thus becomes the imputed estimate for the missing value. The distance metric is determined by the type of data. For continuous observations; Euclidean, Manhattan and Cosine distance metrics are commonly used. On the other hand, Hamming distance is used for categorical observations.
The KNN algorithm is simple to understand and easy in implementation. It's non – parametric nature also allows it to work very well in different settings. However, this algorithm becomes slow and hence time – consuming especially with big data because of searching identical instances throughout the entire data. Moreover, it's accuracy can be reduced adversely due to the small difference between the nearest and farthest neighbor.

XGBoost

XGBoost is a scalable supervised learning library for tree boosting. It implements machine learning algorithms through the gradient boosting framework(Chen & Guestrin, 2016). It's scalability has made it one of the most sought machine learning system. One of it's notable feature is that it can run more faster than other machine learning methods and scales to billions of examples(Nielsen, 2016). Furthermore, it hosts a novel tree learning algorithm for handling sparse data. To handle missing values, XGBoost contains an algorithm that learns the best direction to impute a missing observation or measurement(Chen & Guestrin, 2016).

1.2.4 Choice of Random Forests for this study

Random Forests algorithm is among the machine learning techniques which have been proven to impute missing values accurately. Compared to other methods like k Nearest Neighbours and XGBoost, it has been shown to perform better(Waljee et al., 2013). RF is more attractive due to the fact that it hosts properties for handling missing value for mixed data, and adapts easily to interactions and non – linear settings.
Due to its wide use in most research projects, different algorithms have been developing from the original package by Breiman(Tang, 2017). Studies have also saught to evaluate the performance of the various algorithms using large diverse types of data. This has added more knowledge on the best Random Forests algorithm to apply for different settings with accuracy and efficiency. This method would thus inform the best approach to use to impute missing values when there is clustering. Random forests can also be used in predictive modelling with multiple responses(Cutler, Cutler, & Stevens, 2012).

1.3 Statement of the Problem

Data analysis phase of any clinical research is one of major determinants for the accuracy and reliability of any study. Failure to take into account or improper methods for replacing missing values has been shown to affect estimation of the study parameters.
Random forests have proven to be fast and more efficient approach to handling missing values while still maintaining a high level of accuracy. The flexibility of random forests in providing various algorithms to deal with missing data in different data settings present an opportunity to explore more on best way to handle missing data especially when the data is clustered. This study seeks to impute missing data for data of hierarchical nature. The main aim of this project is to identify key determinants of clinical quality through the process part of the health care delivery. The Paediatric Quality of Care score was used as the measure of clinical quality. We would use routine clinical data from a clustered randomized trial involving 12 Kenyan county – level hospitals between March and November 2016, which aimed at investigating the effect of enhanced audit and feedback (AF) on adoption of pneumonia guidelines by the World Health Organization (WHO).
Since the PAQC score was developed recently, there hasn't been a study to identify the key features influencing the different levels of the score. The results of this study will thus provide useful information that will provide insights to the different levels of clinical care to both health care practitioners and government agencies involved in providing policies for pneumonias management.

1.4 Objectives

1.4.1 Main Objective

Estimate missing values in pediatric routine data and identify key determinants of Pediatric Admission Quality of Care score using Random Forests.

1.4.2 Specific Objectives

i. Use random forests to impute missing values in routine pediatric data

ii. Identify the key determinants of PAQC score as a multiple outcome using RF

iii. Quantify variable importance of predictors of PAQC score using RF

1.5 Sampling Design

1.5.1 Location of study

The study was conducted at both the University of Nairobi and Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme in Nairobi County, Kenya.

1.5.2 Data source

This study was conducted through a retrospective cross – sectional study design, where data is collected from hospital records. It is based on data from a cluster randomized trial which was conducted in 12 Kenyan hospitals to evaluate the effect of enhancing audit and feedback on childhood pneumonia management(Ayieko et al., 2019). The data was obtained through a formal request to the KEMRI – Wellcome Trust Research Programme through the relevant data committee.

1.5.3 Sampling technique used

The hospitals were the basic unit of randomization. The researchers selected the 12 hospitals based on the following inclusion criteria. The hospital had to be public, government-owned and had at least 1000 admissions of children each year. They were also supposed to be located in either a low or high malaria transmission setting based on the main malaria ecological zones in Kenya. This process was done through consultation with the ministry of health.
A case whose data was obtained had to meet the following criteria. A child was supposed to be aged between 2 and 59 months. If a child had a history of pneumonia clinical

Assessment of Signs	
Primary signs	**Secondary signs**
Cough or difficult breathing	Central cyanosis or (in)ability to drink/breastfeed or AVPU or grunting or acidotic breathing if very severe, or central cyanosis and (in)ability to drink/breastfeed or AVPU, and grunting and acidotic breathing if severe, or central cyanosis and (in)ability to drink/breastfeed or AVPU, and grunting and acidotic breathing and respiratory rate if non-severe.

Table 1. Assessment of pneumonia signs

Diagnosis and Classification	
Present	**Absent**
• Pneumonia A child was expected to be classified has having pneumonia if at least 2 signs were present (either primary or secondary) • Severe Pneumonia A child was expected to be diagnosed with severe pneumonia if the 2 primary signs and at least 5 secondary signs were present.	. No classification

Table 2. Diagnosis and Classification of pneumonia

diagnosis whether signs or symptoms were present or not were also included. Children with a cough exceeding more than two weeks were not selected for the study. If a case had other co – morbities; Meningitis, HIV, severe malnutrition, severe malaria, surgical conditions and sepsis was also excluded. Information on the intervention process and outcomes can be found in the documented clinical trial publication(Ayieko et al., 2019).

Checklist of guidelines used for management of Pneumonia

Assessment

At the assessment phase, children who are admitted were expected to be assessed for 2 primary and 7 secondary signs. These are showed in

Diagnosis and classification

The diagnosis and classification of pneumonia patients criteria is presented in 2. **Treat-**

ment
A child who is between 2 and 59 months diagnosed with pneumonia was expected to receive Amoxicillin of between 32 – 48 mg/kg with a frequency of 2 times a day for the specified duration.

1.5.4 Sample size and power calculation

In total, data for 2,127 children between the age brackets of 2 to 59 months was obtained for this study.

1.5.5 Variables of Interest

Response variable
For the modelling of multilevel data, the outcome variable will be the PAQC score from all the domains as a sum of the scores at the various points of measurements in each domain.

Domains of PAQC Score
The domains of the PAQC score are presented in 3.

Predictor variables
The covariates are presented in 4.

1.6 Justification for the study

Childhood pneumonia is a leading cause of mortality for under – fives in developing countries. Of all the under – fives deaths that occur in the world, a fifth of them are caused by pneumonia(WHO, Fact Sheet: Pneumonia. 2016). It is estimated at 1.9 million each year. In Kenya, pneumonia is the second major cause of under - five deaths accounting for 16 of deaths in those under the age of five (Black et al., 2010). In Kenya, for children under the age of five, pneumonia is diagnosed through the Integrated Management of Childhood Illness (IMCI) criteria in public health facilities(Gera, Shah, Garner, Richardson, & Sachdev, 2016).
Despite stringent measures to curb pneumonia, there are still quality of health care gaps in the diagnosis, assessment and treatment of this leading cause of mortality in children. This is as a result of differences in healthcare settings or lack of knowledge about the set standard guidelines in managing pneumonia. This has resulted to a slow reduction in deaths for under-fives due to the disease. A Paediatric Admission Quality of Care score was recently developed to assess compliance to the set guidelines in management of pneumonia. This score identified possible gaps that could be filled and thus improve quality of health care.

No.	Domain	Sub - domain	Binary Classification
1	Assessment	Primary signs	1 – The two symptoms were assessed 0 – At least one or both of the symptoms were not assessed
2	Assessment	Secondary signs	1 – All 7 symptoms were assessed 0 – At least one of the symptoms was not assessed
3	Assessment	Both primary and secondary signs	1 – All 9 symptoms were assessed 0 – At least one of the 9 is not documented
4	Diagnosis	Diagnosis	1 – Right diagnosis & classification of pneumonia 0 – Otherwise
5	Treatment	Treatment	1 – Those diagnosed with pneumonia got Amoxi 0 – Those diagnosed with pneumonia did not get Amoxil
6	Treatment	Dosage and Frequency	1 – Pneumonia patient received correct dosage and in right frequency 0 – Patient did not get right dosage and in right frequency

Table 3. Domains of PAQC Score

No.	Predictor	Variable type	Predictor classification
1	Malaria prevalence	Binary	0 – Low 1 – High
2	Hospital Workload	Binary	0 – Low 1 – High
3	Intervention	Binary	0 – Control arm 1 – Intervention arm
4	Child gender	Binary	0 – Female 1 – Male
5	Age group	Binary	0 – 2 to 11 months 1 – 12 to 59 months
6	Comobidity	Categorical	0 – None 1 – One condition 2 – Two conditions 3 – Greater or equal to 3 conditions
7	Weight	Continuous	Weight of the child in kg
8	Clinician sex	Binary	0 – Female 1 – Male
9	Clinician cadre	Binary	1 – Clinical Officer (CO) 2 – Medical Officer (MO)
10	Month	Continuous	The time a child presented to the hospital

Table 4. Covariates

Clinical routine data used to calculate this score was sparse. Most fields were missing due to various reasons. The accuracy and validity of the score could be affected by the significant number of missing values. To overcome this possible shortcoming, this study sought to use random forests to estimate the missing values and in – turn calculate a new PAQC score having handled the missing values.

The study will also identify the key determinants of the score. This would thus inform health care providers on ways to improve delivery of health care services. It will also add more knowledge on determining PAQC score by overcoming the major problem of missing observations in hospital records. This would inform other researchers interested in determining PAQC scores for other chronic diseases and hence improve health care.

2 Literature Review

2.1 Introduction

This chapter outlines previous studies on PAQC score in measuring quality of health care. It will also involve discussion on modelling data of hierarchical nature with multiple outcomes. An in-depth explanation on nature and causes of missing data in clinical routine data will also be discussed. Application of random forests in handling missing data and in predictive modelling with multiple outcomes will also be discussed.

2.2 History of clinical quality

Quality of healthcare in clinical practice is an interaction between patients and the clinicians and how inputs within the health system are transferred into health outcomes. Clinical quality focuses mostly on the process of healthcare rather than inputs such as drugs, facilities and equipment(Donabedian, 1988). Therefore, for an up – to standard healthcare, it should be effective and evidence based(Baker, 2001). Although the availability of inputs makes it easy to measure, they can't be solely be used to determine whether there is an improvement in health care provided(Peabody, Taguiwalo, Robalino, Frenk, et al., 2006). Therefore, to evaluate clinical processes, the behavior of clinicians and their associated measurements and assessments to patients provide critical information in the development of tools and methods towards improving the patients' health care services. The healthcare delivery could thus be evaluated based on the current guidelines to identify gaps and hence improve care. However, there are quite a number of challenges in evaluating clinical quality. There is need for strong and credible evidence to provide useful estimates of quality of care for policy purposes by healthcare providers and as a standard to evaluate interventions(Boren & Balas, 1999).
For many developing countries, clinical quality guidelines either exist and are poorly enforced or are unavailable in totality. Moreover, where the guidelines exist there are no clear standards and cut – offs for distinguishing high and low quality of care. This thus calls for careful judgement through scientific research.

2.3 Quality of healthcare in management and control of childhood pneumonia.

Globally, Pneumonia is the leading cause of deaths among under – fives. According to WHO, 920136 children under the age of five years died from pneumonia in 2016(WHO,

Fact Sheet: Pneumonia. 2016). More deaths were reported in Sub – Saharan Africa. To reduce this prevalence, a better understanding of childhood pneumonia clinical practices in assessment, diagnosis and treatment required to improve both outpatient and inpatient pediatric care especially in settings with resources limitations.

Guidelines to improve general population health have in most cases focused on expanding health care delivery through provision of quick and easily accessible services to the population. However the quality of the health care has not been explored and taken into account. Health care stakeholders have always assumed the expansion of health care services implies natural improvement in quality of care(Ensor & Cooper, 2004). This hypothesis is not correct. People have developed rational ways to seek healthcare services based on the quality of services delivered previously. This is due to the common belief that poor quality of health care is a hindrance to coverage of universal health which is independent of access to healthcare services(Berendes, Heywood, Oliver, & Garner, 2011). To improve the health care services strategies have been developed to ensure essential inputs in-line with the developing modern healthcare landscape are in place and easily accessible at all levels. These strategies are only tailored towards strengthening provision of services in line with the standard guidelines(Heiby, 2014). The need to go a step further to evaluate healthcare services using patient – centered models has however been explored recently.

To advance pneumonia control and management in protection, prevention and treatment of childhood pneumonia, a Global Action Plan for Pneumonia and Diarrhoea (GAPPD) has been launched. It is an initiative by the WHO and UNICEF. This intervention outlined three ways. Firstly, protection through exclusive breastfeeding and sufficient complimentary feeding. Secondly, prevention by use of vaccinations, handwashing with soap, reducing air pollution in the household and HIV prevention practices for HIV-infected and exposed children. Thirdly, treatment through access to the recommended kind of care and drugs through all levels of health care delivery given the diversity due to differences in resources and severity.

These guidelines have been widely shared but most countries have not yet implemented them. Kenya has however taken steps towards the implementation through the Integrated Management of Childhood Illness (IMCI) criteria in public health facilities. There has however been need to measure compliance to the guidelines. This is particularly important due to the differences in health care settings and resources in healthcare delivery units. There are also common challenges towards evaluating quality of healthcare. In most developing countries for instance, medical records are not well maintained and in most cases are sparse with missing information both at random and systematically. This shortcomings may therefore not provide a reflection of the actual practice. In some situations quality of health care can be evaluated through use of covert research where under-cover patients are used to monitor quality of healthcare. This has however in the past raised ethical concerns(Leonard & Masatu, 2010). In light of these challenges and many more, researchers have sought to find ways on evaluating quality of healthcare.

2.4 Pediatric admission quality of care (PAQC) score as a measure of clinical quality

PAQC score is a logical summative patient – level measure of quality of healthcare from admission to treatment of a disease(Opondo et al., 2016). This tool measures compliance by clinicians to the recommended clinical guidelines on the various stages of healthcare delivery. According to the designers of this important tool in clinical quality, process metrics are evaluated across three domains of healthcare delivery. These domains are assessment, diagnosis and treatment stages for a health condition. Medical records are usually mined for data on the clinician's compliance to the checklist of guidelines on the administration of healthcare services at each of these domains for a particular health condition. This data can be obtained for all admissions for the particular ailment at all health units and compiled.

At the assessment level, evaluation is done on whether clinicians assessed all symptoms recommended in the clinical guidelines for diagnosis of a particular disease. Once the data on assessment has been obtained, the next stage is to evaluate whether clinicians recorded correct diagnosis for the symptoms identified at the assessment level. The last domain focuses on if the patient received the correct treatment based on the diagnosis. Recommended medicines, correct dosage and frequency of treatment prescription for the correct diagnosis are evaluated at this stage. At each stage Boolean indicators are used to show whether compliance to guidelines was followed or not. These indicators at each of the domains are summed up to provide a cumulative score that is used as a process metric of quality of healthcare. The higher the score the better the quality of care provided and the lower the score the poorer the process of health care delivery.

Recently, a validation study was carried out on the effectiveness of this score. In this study to evaluate the association of the PAQC score with mortality in Kenyan hospitals found out that a unit increase in the PAQC score significantly reduced the odds of a mortality case in a hospital, which is a pooled estimate across the 27 hospitals whose data was obtained(Opondo, Allen, Todd, & English, 2018). The data was obtained for children who were admitted for treatment of malaria, pneumonia, diarrhea, or dehydration. This findings bring to light for the first time a tool that could be used across the whole spectrum of health conditions using the WHO guidelines on the management of chronic diseases to improve healthcare delivery and reduce major causes of mortality and morbidity. This could inform policies to increase awareness on adherence to the guidelines which has been shown to improve the score. This was evident in a study to determine how processes of care improve for children hospitalized with diarrhoea(Akech et al., 2019). The study found that participating in a clinical network by health stakeholders could improve sensitization on use of the clinical guidelines in healthcare delivery and significantly increase the PAQC score translating to better clinical quality.

Most medical records are usually sparse. This may affect significantly the cumulative PAQC score. Methods to address the missing fields is therefore needed to ensure estimates are more accurate. This study will use Random Forests algorithm to impute missing values

in routine pediatric data and determine key determinants of the PAQC score with multiple outcomes.

2.4.1 Statistical methods used in analysis of PAQC score

Having been developed recently, PAQC score hasn't been researched on widely. However, few studies have sought to find the association between PAQC score and outcome measures particularly mortality. One such study was used for validation of the tool. The main aim of this study was to determine the association of adhering to guidelines making up the PAQC score to mortality among under – fives who were admitted to 8 hospitals across Kenya for acute illnesses (Opondo et al., 2018). Hierarchical logistic regression models were used to assess the association of the score with mortality. Mortality was the outcome variable. The models were adjusted for comorbidities, illness severity, age, gender, the duration they were admitted in the hospital, trial arm and the survey number. The study found that for hospitals where the guidelines were adhered to the risk of mortality among the patients reduced as compared to where they were not followed. Another study that explored the same objective was carried out a year later to determine the association between the score among severe pneumonia children
(Opoka et al., 2019). This study involved a review of hospital records for over 1300 children. Logistic regression models were fit to determine the association where mortality was used as an outcome variable. The study found that inpatient deaths were significantly reduced for children who were managed according to the guidelines. To determine how the PAQC score fared overtime for different settings or interventions, a study to study the pattern of improvement in clinical quality for hospitals that participate in a clinical network was carried out(Akech et al., 2019). In this study, the PAQC score was classified into a binary variable where a score greater or equal to 5 was a "good score". Mixed effects fractional polynomial regression was fit to adjust for hospital level. The study found that after joining a clinical network the scores increased gradually.
Clearly, there is study that has sought to find the key determinants of the score as most of the studies investigated the association between the score and outcome measures particularly mortality with the score as a predictor variable. This study presents for the first time a model to determine the predictors of the score and determining the most important predictors that influence clinical quality in management of pneumonia especially in children under the age of five years.

2.5 Analysis of hierarchical data for an ordered outcome

Most clinical research problems have shown to have patients or samples nested within clusters such as clinicians and hospitals. To carry out statistical models to understand various relationships and associations between features of interest, we need to adjust for the effects of these clusters. There has been various methods in use. Linear fixed models have been widely used for outcomes that are continuous.For instance, in a study in Uganda

to identify quality cattle breeds using yield as outcome variable, a linear mixed model was fit with cattle – associated characteristics were nested within herds group(Onyango, 2009). Multilevel exploratory data analysis using Random Forests have also been proposed in the recent past. The first implementation of this technique was in the development of the MLEDA package that used scores of a mathematics test for children as outcome variable(Martin, von Oertzen, & Rimm-Kaufman, 2015). In this study, students were nested within schools. Children – related characteristics were used as predictors. Random intercept logit models have also been used for binary outcomes(Rodriguez & Goldman, 2001). Methods for multi-level analysis for ordered outcomes have been proposed in the recent past. An earlier study to explore this technique involved 2408 children in 169 schools to determine proficiency in reading at the end of the first grade(O'Connell, 2010). Proficiency in reading was used as an ordered outcome with 6 levels. Children were nested within schools. The study used the hierarchical proportion odds model approach. Boys in public schools and from low social economic status homes had greater probabilities of being at or below a given proficiency level relative to their peers. A year later, another study to explore when a linear mixed model would be ideal for ordinal outcomes used simulated data with various levels and clusters(Bauer & Sterba, 2011). Five hundred samples were used for this study. Penalized quasi-likelihood and maximum likelihood using adaptive quadrature methods were used to evaluate the deviance of the estimates from a cumulative logit model which is a standard model for ordinal outcomes. This study found that when the marginal distribution of the category responses was almost normal and the categories were seven, the linear model bias of the fixed effects is in the acceptable levels. Subsequent studies that have explored use of ordinal outcome using cumulative – logit mixed models include a study to determine the effect of bio – energy treatments on breast cancer using mice (Schmidt, 2012) and another to determine student satisfaction on courses taught at the university(Grilli & Rampichini, 2012).
Generally the standard methods for analyzing hierarchical data are the cumulative – logit models and the hierarchical proportions odds methods. We will use the cumulative – logit mixed models under the assumption of proportional odds and multilevel exploratory data analysis using Random Forests to evaluate the objectives of the study.

2.6 Cumulative logit models approach

One of the widely used method for analysis of data with an ordinal outcome is the proportional odds cumulative logit model.This technique transforms ordinal levels into binary at a given threshold that is based on cumulative probabilities.
Given $Y = 1, 2, ..., J$ are levels of an ordinal outcome with probabilities $\pi_1, \pi_2, ..., \pi_j$, then the cumulative probability for a level less or equal to j is given by

$$P(Y \leq j) = \pi_1, \pi_2, ..., \pi_j \tag{1}$$

Therefore,

$$log(\frac{P(Y \leq j}{P(Y > j}) = log(\frac{P(Y \leq j}{1 - P(Y \leq j}) = log(\frac{\pi_1 + \pi_2 + ... + \pi_j}{\pi_{j+1} + ... + \pi_J}) \quad \textbf{(2)}$$

is the cumulative logit.

This is an estimate of the likelihood of a response being in level j or below to a level higher than j.
A sequence of these cumulative logits is given by;

$$\begin{aligned} L_1 &= log(\frac{\pi_1}{\pi_2 + \pi_3 + ... + \pi_r}) \\ L_2 &= log(\frac{\pi_1 + \pi_2}{\pi_3 + \pi_4 + ... + \pi_r}) \\ &\vdots \\ L_{r-1} &= log(\frac{\pi_1 + \pi_2 + ... + \pi_{r-1}}{\pi_2 + \pi_3, ..., \pi_r}) \end{aligned}$$

L_j is the log - odds for a certain level being less or equal to level j

Adding covariates to the model we have;

$$\begin{aligned} L_1 &= \alpha_1 + \beta_1 X_1 + ... + \beta_p X_p \\ L_2 &= \alpha_2 + \beta_1 X_1 + ... + \beta_p X_p \\ &\vdots \\ L_{r-1} &= \alpha_{r-1} + \beta_1 X_1 + ... + \beta_p X_p \end{aligned}$$

This is the proportional - odds cumulative logit model where;

α_j is the log odds of being less or equal to level j

β_k is the effect of variable X_k on the outcome

For instance considerig only one covariate we have;

$$logit[(P(Y \leq j)] = \alpha_j + \beta_x \tag{3}$$

The cumulative probabilities will therefore be given by $P(Y \leq j) = \frac{\alpha_j + \beta_x}{1 + exp(\alpha_j + \beta_x)}$

Since β is constant the odds - ratio is proportional to the difference between x_1 and x_2 where β is the constant of proportionality. Hence the name 'proportional odds model'.

3 Methods

3.1 Introduction

This chapter outlines derivations and model specifications for the models fit. Imputation of missing values and calculation of the PAQC score will also be discussed.

3.2 Derivations and model specification

This study will use four steps to evaluate the objectives of the study. These include imputation of the missing values using RF, calculating the PAQC score after accounting for missing values,cumulative - logit mixed models and multilevel exploratory data analysis models using the overall PAQC score as an outcome and the other variables as predictors.

3.2.1 Imputing missing values using RF

The package missForest in R will be used to impute missing values in variables of interest. All predictor variables will be passed onto the algorithm for imputation. This method is best in situations with mixed type of data and it provides OOB imputation error estimate.

3.2.2 Calculation of PAQC score

The binary indicators for the specific points of measurement at each domain in table (3) will be summed up to obtain an overall score for each observation. This will form the PAQC score.

3.2.3 Cumulative - logit Mixed models

Model Specification

$$logit(Y_{cij}) = \alpha_c + X_{ij}\beta + \mu; c = 0, 1, \ldots, c-1 \quad \textbf{(4)}$$

where;

- Y_{cij} - cumulative probability upto the c-th category for patient i in hospital/clinician j

- α_c - Cumulative intercept for a certain PAQC score level
- i - Patient ID
- j - clustering variable
- X_{ij} - Covariate vector with patient characteristics
- β - vector of fixed parameters
- μ_j - random effect (unobserved factors)

Model Estimation

Let;

- Y_j - Vector of n_j ordinal responses of the j^th cluster
- X_j -Covariate matrix of cluster j with rows X_{ij}
- θ - (α,β,δ_μ) be the vector of the model parameters where $\alpha^{'}$= α_1,…,c-1

The likelihood of Y_j conditional on μ_j is equal to the product of the conditional probabilities of the responses.

$$L_j(Y_j|\mu_j;X_j:\alpha,\beta) = \prod_{i=1}^{n_i}\prod_{c=1}^{c}(Y_{cij}-Y_{c-1ij})^{d_{cij}} \quad \textbf{(5)}$$

Where; d_{cij} is the indicator of (Y_{ij}=y_c) The likelihood of Y_j is obtained by integrating out the random effects μ_j

$$L_j(Y_j|X_j;\theta) = \int_{\mu_j} L_j(Y_j|\mu_j;X_j:\alpha,\beta)f(\mu_j:\delta_{\mu_j}) \quad \textbf{(6)}$$

Where; $f(\mu_j;\delta_{mu_j})$ is the density of μ_j

$$\mu_j \sim N(0,\delta_\mu)$$

Given independence across clusters, the log - likelihood for the j cluster is

$$logL = \sum_{j=1}^{j} logL_j(Y_j|X_j;\theta) \quad \textbf{(7)}$$

3.2.4 Multi - Level Random Forests model

RF is a predictor consisting of a collection of randomized base regression trees.

Estimation of Intra – cluster correlations

PAQC score is assumed to be numeric in this analysis
The first step was to estimate Intra – cluster correlations to assess significance of the clinician and hospital level variables.
This was executed by fitting random intercept models only, without predictors. One model using clinician level as cluster and the other using hospital level. If the estimate of the ICC is above 0.15, the clustering of the outcome variable in that cluster is significant and hence needs to be accounted for in the multilevel modelling.
In cases where the ICC is small, this suggests that clustering may not strongly influence the outcome and thus we may just fit the normal model without nesting the data into the cluster variable.

RF Multilevel Exploratory Data Analysis (MLEDA) models

These models will be fit to determine most important variables in predicting the PAQC score.

Determination as to which variables were most important in predicting PAQC score.

Two RF models will fit, one is the normal RF model and the other the conditional random forest model. The implementation the conditional random forest algorithm differs from the normal RF model with respect to the base learners used and the aggregation scheme applied. Particularly, the conditional random forest aggregation scheme works by averaging observation weights extracted from each of the trees, this is contrary to the RF scheme that averages predictions directly Hothorn et al Meinshausen (2006).
The third model is a linear mixed effects model adjusting for clustering at both the clinician and hospital level.
Variable importance plots for the three models will be plotted to illustrate the relative importance of each predictor in the three models.

Model specification and estimation

$$r_n(X, \theta_m, D_n), m \geq 1 \tag{8}$$

Where $\theta_1, \theta_2, \ldots$ are $i.i.d$ outputs of a randomized variable θ

These random trees are combined to form the aggregated regression estimate.

$$\bar{r_n}(X, D_n) = E_\theta[r_n(X, \theta, D_n)] \tag{9}$$

E_θ is the expectation wrt random parameters, conditionally on X and the data set D_n

4 Data Analysis

4.1 Introduction

This chapter presents the descriptive statistics for the PAQC score on association with predictors used in this study. The descriptive statistics will be presented as counts, percentages, means and medians.Associations will be explored using box-plots. In this chapter, results of multilevel mixed models and random forests models findings will be presented. The various regression coefficients and estimates will be interpreted based on 5 level of significance.

4.2 Data cleaning procedures

The raw data obtained from the KEMRI - Wellcome Trust Research programme was imported into R. Variables that were in string format were transformed into factor format. These variables are; Hospital ID, Malaria prevalence, Hospital workload, Intervention arm, Child sex, Age group, Comobidity, Amoxl frequency, Clinician Sex, Cadre, Correct dose and Correct treatment.
A new variable named comobidities was created based on the other ailments recorded by clinicians at the time a child presented to the hospital. This variable had categories for whether a child had one, two, three or more than 3 comorbidities.Variales that were used to create other variables were removed from the dataset. These included variables used to define the primary, secondary, diagnosis and treatment scores.
Once the variables of interest were identified, missing values for each covariate were visualized and a test to determine the nature of missingness evaluated. Below is a visualization of the missing values.

From the missing values patterns in the figure 1, missing values on clinician sex and cadre were high and were associated. The values were thus not missing at random. The same trend is observed for Amoxl frequency and dosage. The Amoxl frequency and weight of the child also seem correlated. Child sex and age did not have any significant association with any other variable. Data on these two would thus have occurred at random.
There was a significant trend on missing values in recording Amoxl dose by hospital ID as shown in the figure 2. H9 and H10 had significantly higher missing values in this variable. Data was imputed using the random forests *missForest* package in R. The out of bag error (00B) showed precision and accuracy in the imputation. The Normalised root mean squared error (NRMSE) for continuous variables was 0.00008 while the proportion of false

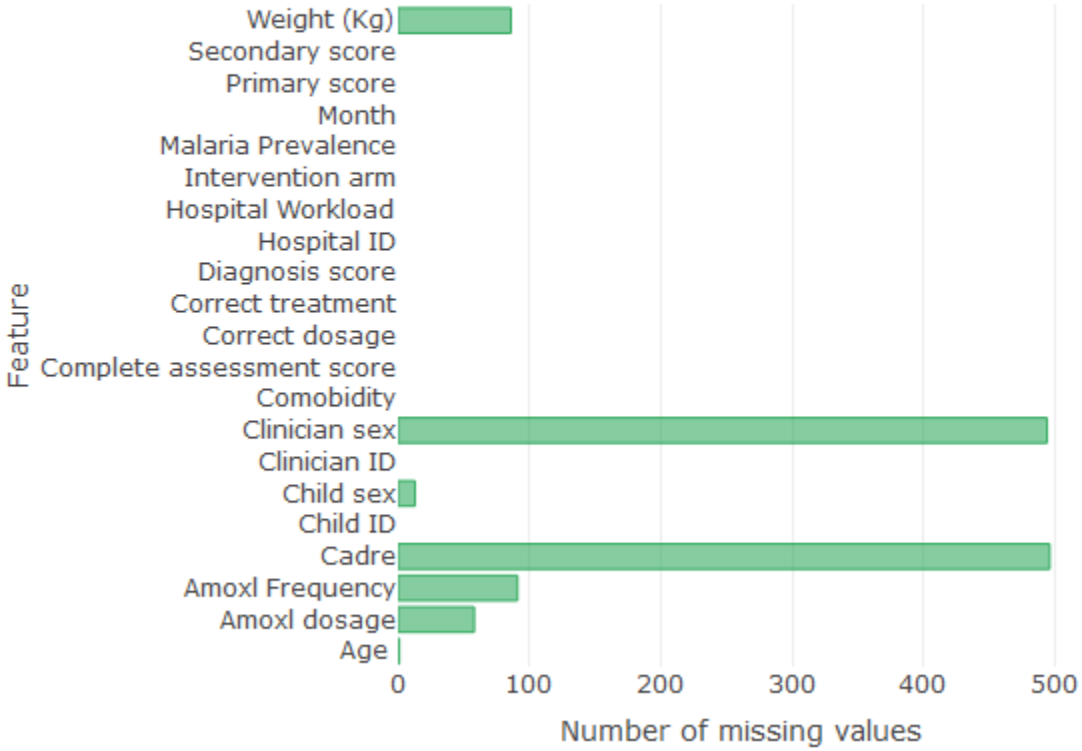

Figure 1. Missing values pattern

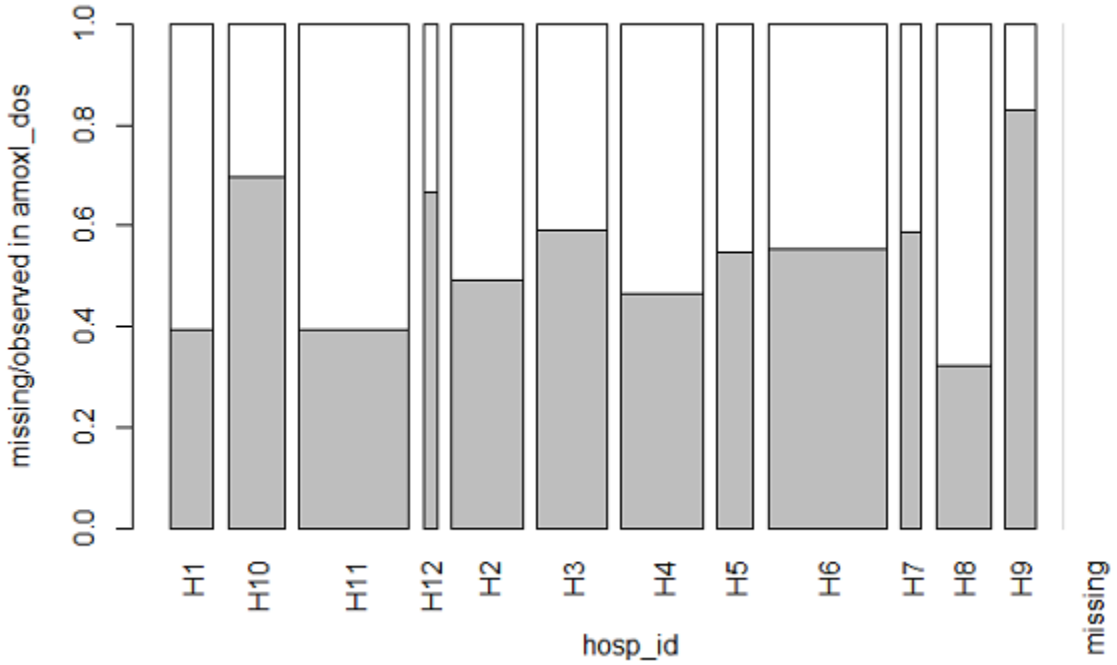

Figure 2. Missing values by hospital

classifications for categorical variables (PFC) was 0.09. This shows accuracy in imputation.

4.3 Descriptive data analysis

Descriptive statistics for each covariate and domains is presented in table 5.

PAQC Score by various attributes

From the boxplots in the figure 3, hospital is an important factor in predicting quality of healthcare using the PAQC score. H1 and H11 had the highest median score of 5, while H12, H2, H4, H6, H7 and H9 had the least median score of 3.

From figure 4, hospitals with high hospital workload had a significantly higher scores as compared to those with low workloads.

4.4 Random effects model analysis

Proportion of PAQC Scores

The proportion of PAQC scores is presented in table 6.

Random Effects models

Table 7 shows the first random intercept models without covariates to assess whether clustering is significant at either hospital or clinician level. Two models are fit, one with

Name of variable	**Frequencies and descriptive statistics (N = 2127)**	
	n	%
Child sex: Male	1174	55.2
Clinician sex: Male	1290	60.7
Malaria prevalence :Low	1576	74.1
Hospital workload: High	1067	50.2
Intervention arm: Intervention	953	44.8
Age group : 12 – 59 months	1224	57.6
Comobidity		
0	995	46.8
1	633	29.8
2	381	17.9
>=3	118	5.6
Cadre : Medical Officers	1335	62.8
PAQC Sub - domains		
Primary score	2111	99.3
Secondary score	1183	55.6
Assessment score	1173	55.2
Diagnosis score	1473	69.2
Correct treatment score	1036	48.7
Both correct dosage & frequency	1093	51.4
Numeric Variables		
	Mean Median Range IQR	
Weight	9.2 8.9 (3 – 24) (6.8 – 11)	
PAQC Score	3.8 4 (0 – 6) (3 – 5)	

Table 5. Covariates

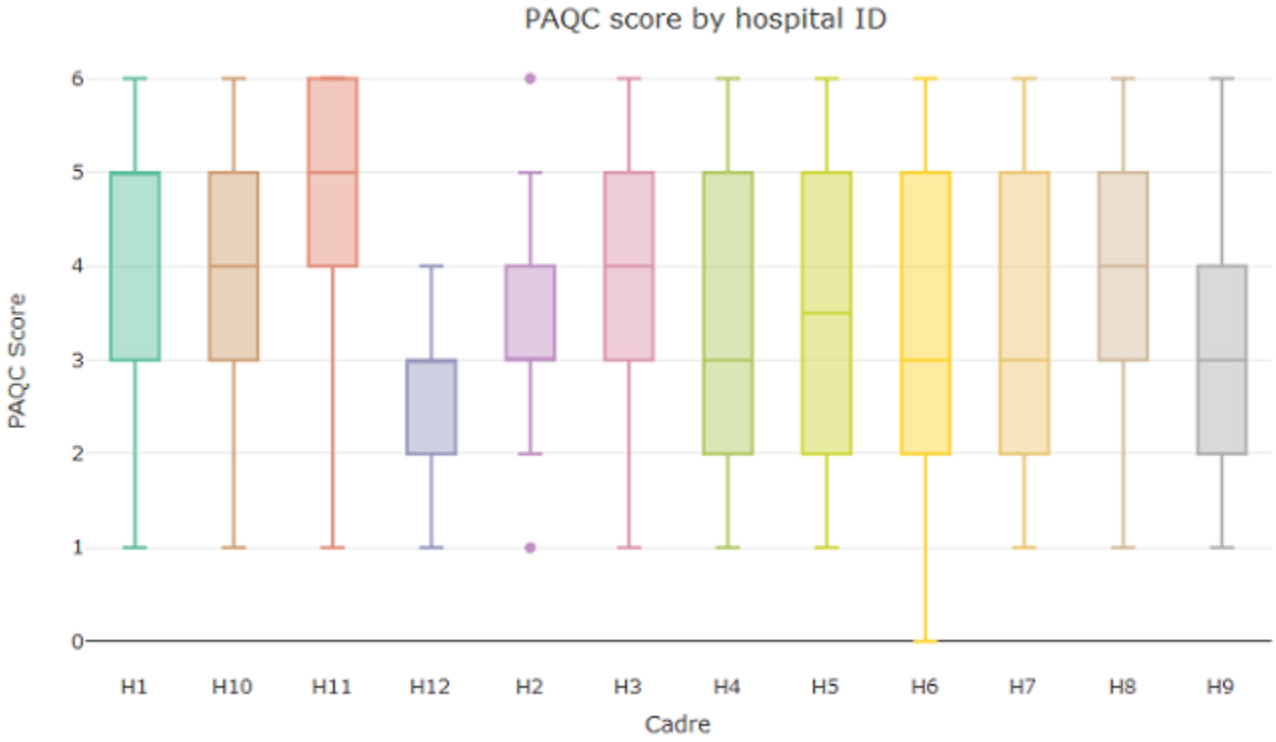

Figure 3. Hospital vs PAQC score

Level	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**Totals**
Frequency	2	164	303	431	472	412	343	2127
Percentage	0.09	7.71	14.25	20.26	22.19	19.37	16.13	100

Table 6. Proportion of PAQC scores

Random effects: Hospital Level		
Groups Name	Variance	Standard Deviation
Hospital Level (Intercept)	0.491	0.7007
Number of hospitals: 12		
Random effects: Clinician Level		
Groups Name	Variance	Standard Deviation
Clinician Level (Intercept)	1.444	1.202
Number of clinicians: 378		

Table 7. Random Effects model results

Figure 4. Hospital workload vs PAQC score

Fit	No.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
Null Model	6	7462.9	-3725.5			
Random Effects Model (Hospital Level)	7	7154.8	-3570.4	310.14	1	<2.2e-16
Random Effects Model (Clinician Level)	7	6988.3	-3487.2	476.6	1	<2.2e-16

Table 8. Likelihood Ratio Test

hospital level and another with clinician level as random factors.
At hospital level, a cluster variance of 0.491 (with a standard deviation of 0.7007) is obtained. At clinician level, a cluster variance of 1.444 (with a standard deviation of 1.202) is obtained. Clearly there is more variation at clinician level relative to the hospital level on the PAQC score. The study goes on to assess whether the variance is statistically significant.

The table 8 shows the results of the LRT to assess whether the cluster variances are statistically significant, the two models were compared with the model without random effects, the null model.
The LRT statistics for hospital and clinician levels are 310.14 and 476.6 respectively. Each with 1 degrees of freedom associated with p-values less than 0.05 (2.2e-16). The null hypothesis of no significant clustering is rejected. There is therefore evidence of unobserved heterogeneity at hospital and at clinician level. These random factors therefore need to be accounted for in the model to be fit.

4.5 Cumulative Link Mixed Models Results

The table 9 presents the results of the cumulative logit mixed models.

4.5.1 Discussion of the results

Model Nested within Hospital

Weight of the child, clinician sex, cadre and the time of admission were significant key determinants of PAQC score based on the P-values at 0.05 level of significance.
The probability of having a higher PAQC score is higher as the weight of the child increases if the child was attended to by a medical officer rather than a clinical officer and as the months increase. Specifically, the probability of increasing the PAQC score by one level as weight increases is 6%, while being attended by a medical officer relative to a clinical officer increases the probability by 19% and as the time of admission increase the probability increases by 12%.

	Nested within Hospital			Nested within Clinician			Nested within both Hospital and Clinician		
Feature	**Estimate**	**Std. Error**	**Pr(>|z|)**	**Estimate**	**Std. Error**	**Pr(>|z|)**	**Estimate**	**Std. Error**	**Pr(>|z|)**
Age group: 12 – 59 months	-0.10	0.10	0.34	-0.10	0.11	0.37	-0.09	0.11	0.37
Child sex: Male	-0.04	0.08	0.62	-0.01	0.08	0.92	-0.02	0.08	0.80
Commodities:									
0	0.14	0.17	0.43	0.06	0.19	0.74	0.10	0.19	0.60
1	0.07	0.18	0.68	0.03	0.19	0.88	0.06	0.19	0.76
2	0.10	0.18	0.57	0.00	0.20	0.98	0.03	0.20	0.87
Child weight	0.06	0.02	**0.00**	0.06	0.02	**0.00**	0.06	0.02	**0.00**
Clinician sex: Male	-0.47	0.09	**0.00**	-0.53	0.16	**0.00**	-0.49	0.15	**0.00**
Cadre: Medical Officer	0.19	0.09	**0.03**	0.33	0.15	**0.02**	0.28	0.14	**0.05**
Hospital Workload: Low	-0.52	0.38	0.18	-0.38	0.17	**0.03**	-0.45	0.41	0.26
Malaria Prevalence: Low	0.06	0.36	0.88	-0.17	0.18	0.33	-0.07	0.39	0.85
Intervention Arm	-0.54	0.36	0.13	-0.59	0.17	**0.00**	-0.59	0.39	0.12
Time of admission	0.12	0.02	**0.00**	0.11	0.03	**0.00**	0.11	0.03	**0.00**

Table 9. Cumulative logit mixed models results

On the other hand, PAQC scores would be lower if the clinician was male. For a male clinician, the probability of reduced PAQC score by one level is 47% relative to a female clinician.

Model Nested within Clinician

Weight of the child, clinician sex, cadre, hospital workload, intervention arm and the time of admission were significant key determinants of PAQC score based on the P-values at 0.05 level of significance. It's evident hospital workload and intervention arm factors come into play as compared to the model nested within hospital level.
Weight of the child increases the probability of a higher PAQC score level by 6%, being a medical officer increases this probability by 33% relative to a clinical officer and as time of admission increases this probability increases by 11%.
The probability that the score decreases by one level on the other hand is 53% if the clinician is male relative to a female counterpart, 38% if the hospital workload is lower than 1000, and 59% if the hospital received an intervention relative to the control group of hospitals.

Model nested within both hospital and clinician level

Typical to the model nested within the hospital level, weight of the child, clinician sex, cadre and the time of dmission were significant determinants of PAQC score based on the P-values at 0.05 level of significance.
The probability the PAQC score increases by one level as weight increases is 6%, while being attended by a medical officer relative to a clinical officer increases the probability by 27% and as the time of admisssion increase the probability increases by 11%.
On the other hand, PAQC scores would be lower if the clinician was male. For a male clinician, the probability of reduced PAQC score by one level is 49% relative to a female clinician.

4.6 Random Forests Models Results

4.6.1 Estimation of Intra – cluster correlations

The intra - cluster correlations are presented in table 10. The amount of variance in PAQC scores associated with hospital level and clinician level is 0.34 and 0.73 respectively.
This implies that just under 34% and 73% of the variation in PAQC scores is associated with the hospital a child was admitted and the clinician that attended to the child respectively. In other words, a third of the variability in PAQC scores across children admitted is associated with the hospital attended while close to three – quarters of the variability in the score is associated with the clinician that attended the child. The presence of such a large ICC suggests that clustering is important at both clinician and hospital level, and that multilevel appropriate methods for fitting recursive partitioning models should be strongly considered. Therefore as we move forward, we have to adjust for clinician and hospital level.

Random effects:		
Groups Name	Variance	Std.Dev.
Hospital (Intercept)	0.3427	0.5854
Clinician (Intercept)	0.7339	0.8567
Number of observations: 2127, Groups: Hospital Levels, 12 Clinicians, 378		

Table 10. Random intercept models

4.6.2 RF Multilevel Exploratory Data Analysis (MLEDA) models results

The figure 5 shows the variable importance for RF models based on a continuous outcome. We fit a RF MLEDA model using PAQC score as outcome and the other variables as predictors. The proportion of variance in PAQC score explained by the RF model based on the cross – validation sample is 0.236.
In other words, a model was fit to a training set of patients made up of randomly selected 50% of the total sample, and then the model was applied to the other 50%, who made up the cross-validation sample. For this second sample, the RF model accounted for 23.6% of the variance in PAQC scores.
Month and weight were the most important variables in predicting the quality of care while age and the number of comorbidities were the least important.
The figure 6 shows the variable importance for RF models based on an ordinal outcome. Weight, clinician and month were the most important variables in predicting the quality of care while the intervention arm and malaria prevalence were the least important.

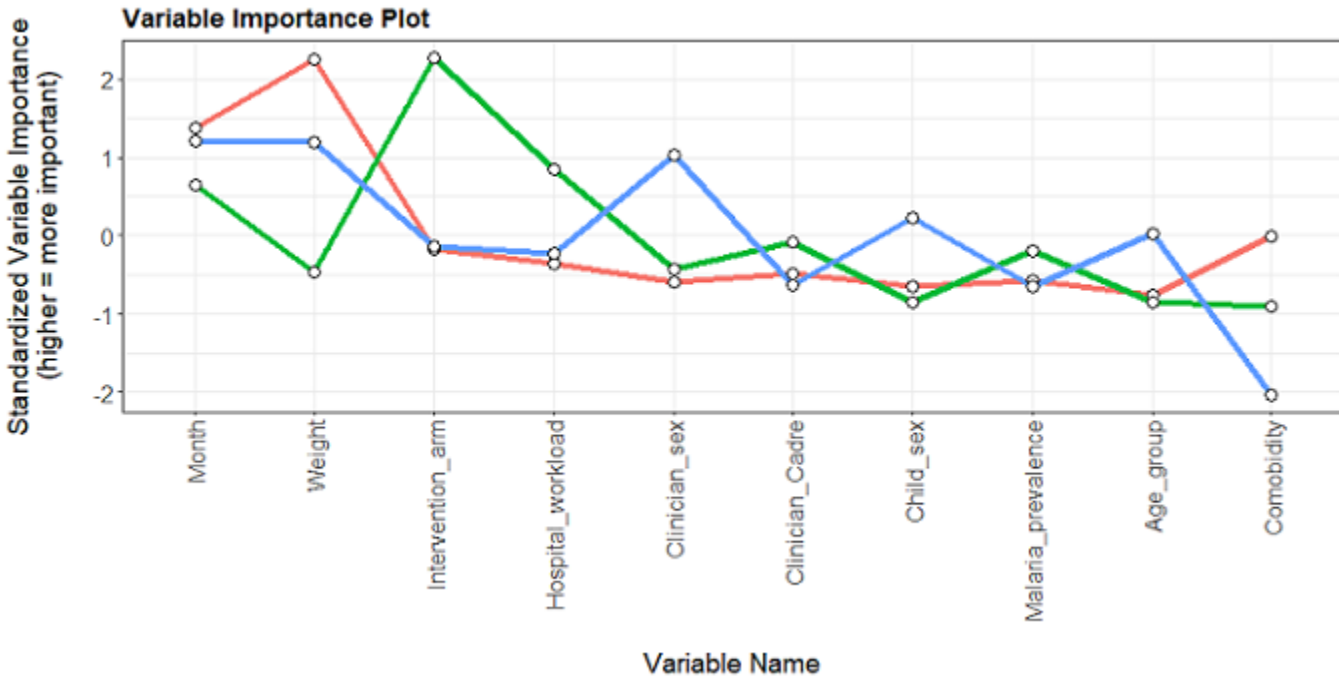

Figure 5. RF using PAQC score as a numeric outcome

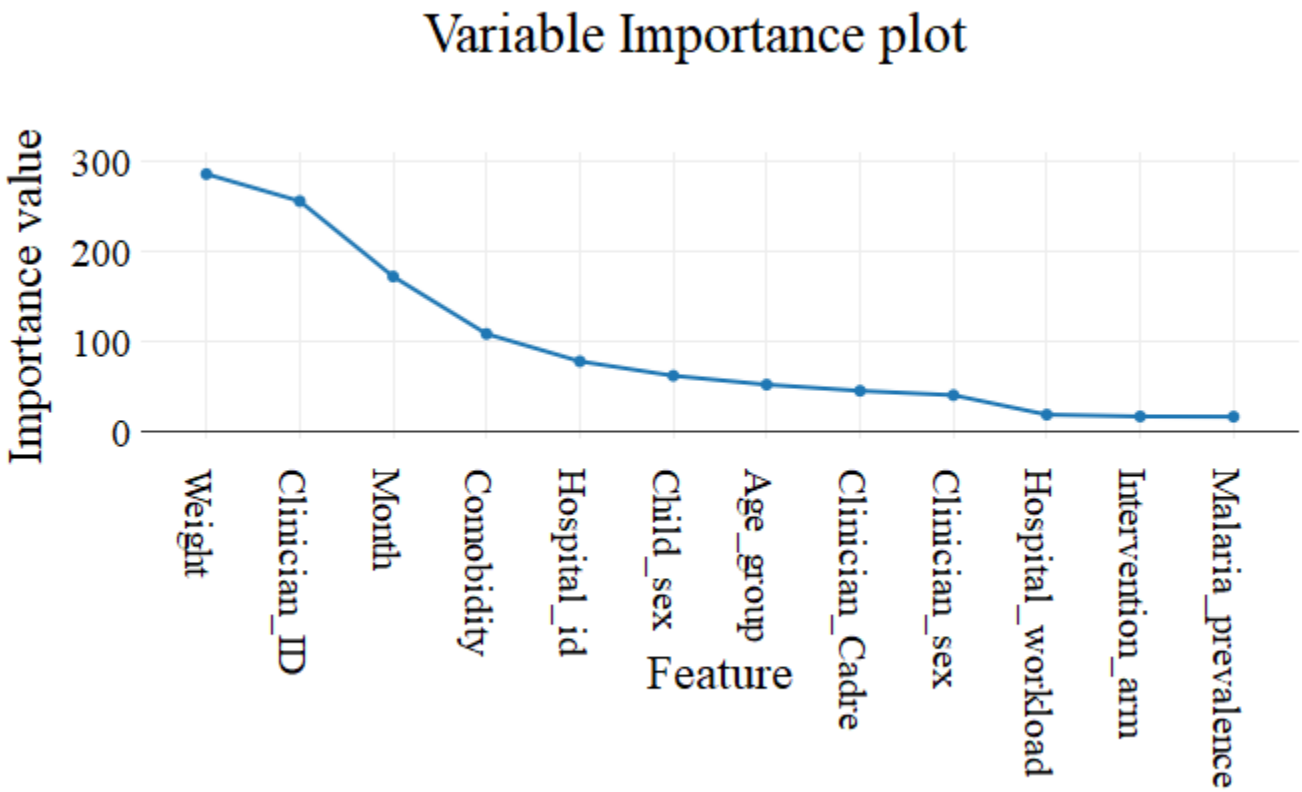

Figure 6. RF using PAQC score as an ordinal outcome

5 Conclusion

5.1 Summary

Although based on the AIC, the model nested within hospital and at clinician level provides more information, the other two models bring out a few insights worth noting.
Child weight is a significant determinant of PAQC score across all the models with same effect size of 0.06.
The model nested within the hospital level provides the same significant determinants of PAQC score as the model nested within both hospital and clinician level. The direction of the effect sizes is same and the magnitude of the estimates are not very different.
The model nested within clinician level is however different with additional factors influencing the score. These are the hospital workload and intervention arm. These are possibly clinician-level specific covariates and they could have a direct effect on the clinicians. Both have a negative effect on the PAQC score. This can also be confirmed by the higher cluster variance in the random effects model at clinician level relative to hospital level.
Since the model nested within both hospital level and clinician level provides more information based on the AIC. The study concludes that the weight of the child, clinician sex, clinician cadre and time of admission are key determinants of PAQC score.
The standard cumulative logit model with no clustering was not interpreted however we note that the results are similar in terms of the direction of effect to the PAQC score as the model nested within the clinician level.
Based on the RF models; age, the number of comorbidities and malaria prevalence for a given patient may not strongly influence the quality of care provided.

5.2 Future Research

A method or algorithm to analyze multi - level data for an ordinal outcome using random forests could be explored.

References

Akech, S., Ayieko, P., Irimu, G., Stepniewska, K., English, M., & authors, C. I. N. (2019). Magnitude and pattern of improvement in processes of care for hospitalised children with diarrhoea and dehydration in kenyan hospitals participating in a clinical network. *Tropical Medicine & International Health*, *24*(1), 73–80.

Ayieko, P., Irimu, G., Ogero, M., Mwaniki, P., Malla, L., Julius, T., … others (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.

Baker, A. (2001). *Crossing the quality chasm: a new health system for the 21st century* (Vol. 323) (No. 7322). British Medical Journal Publishing Group.

Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychological methods*, *16*(4), 373.

Berendes, S., Heywood, P., Oliver, S., & Garner, P. (2011). Quality of private and public ambulatory health care in low and middle income countries: systematic review of comparative studies. *PLoS medicine*, *8*(4), e1000433.

Bertsimas, D., Orfanoudaki, A., & Pawlowski, C. (2018). Imputation of clinical covariates in time series. *arXiv preprint arXiv:1812.00418*.

Bertsimas, D., Pawlowski, C., & Zhuo, Y. D. (2017). From predictive methods to missing data imputation: an optimization approach. *The Journal of Machine Learning Research*, *18*(1), 7133–7171.

Beyad, Y., & Maeder, M. (2013). Multivariate linear regression with missing values. *Analytica chimica acta*, *796*, 38–41.

Black, R. E., Cousens, S., Johnson, H. L., Lawn, J. E., Rudan, I., Bassani, D. G., … others (2010). Global, regional, and national causes of child mortality in 2008: a systematic analysis. *The lancet*, *375*(9730), 1969–1987.

Boren, S. A., & Balas, A. E. (1999). Evidence-based quality measurement. *The Journal of ambulatory care management*, *22*(3), 17–23.

Boyko, J. (2013). Handling data with three types of missing values.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research*, *84*(4), 487–508.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In *Ensemble machine*

learning (pp. 157–175). Springer.

De Silva, A. P., Moreno-Betancur, M., De Livera, A. M., Lee, K. J., & Simpson, J. A. (2017). A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: a simulation study. *BMC medical research methodology*, *17*(1), 114.

Donabedian, A. (1988). The quality of care: how can it be assessed? *Jama*, *260*(12), 1743–1748.

Ensor, T., & Cooper, S. (2004). Overcoming barriers to health service access: influencing the demand side. *Health policy and planning*, *19*(2), 69–79.

García-Laencina, P. J., Sancho-Gómez, J.-L., Figueiras-Vidal, A. R., & Verleysen, M. (2009). K nearest neighbours with mutual information for simultaneous classification and missing data imputation. *Neurocomputing*, *72*(7-9), 1483–1493.

Gera, T., Shah, D., Garner, P., Richardson, M., & Sachdev, H. S. (2016). Integrated management of childhood illness (imci) strategy for children under five. *Cochrane Database of Systematic Reviews*(6).

Grilli, L., & Rampichini, C. (2012). Multilevel models for ordinal data. *Modern analysis of customer surveys: with applications using R*, 391–408.

Heiby, J. (2014). The use of modern quality improvement approaches to strengthen african health systems: a 5-year agenda. *International journal for quality in health care*, *26*(2), 117–123.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, *17*(1), 162.

Kang, H. (2013). The prevention and handling of the missing data. *Korean journal of anesthesiology*, *64*(5), 402.

Kaplan, R. M., & Frosch, D. L. (2005). Decision making in medicine and health care. *Annu. Rev. Clin. Psychol.*, *1*, 525–556.

Karama, A., Farouk, M., & Atiya, A. (2018). A multi linear regression approach for handling missing values with unknown dependent variable (mlrmud). In *2018 14th international computer engineering conference (icenco)* (pp. 195–201).

Langkamp, D. L., Lehman, A., & Lemeshow, S. (2010). Techniques for handling missing data in secondary analyses of large surveys. *Academic pediatrics*, *10*(3), 205–210.

Leonard, K. L., & Masatu, M. C. (2010). Professionalism and the know-do gap: Exploring intrinsic motivation among health workers in tanzania. *Health economics*, *19*(12), 1461–1477.

Li, T., Hutfless, S., Scharfstein, D. O., Daniels, M. J., Hogan, J. W., Little, R. J., … Dickersin, K. (2014). Standards should be applied in the prevention and handling of missing data for patient-centered outcomes research: a systematic review and expert consensus. *Journal of clinical epidemiology*, *67*(1), 15–32.

Lin, T. H. (2010). A comparison of multiple imputation with em algorithm and mcmc method for quality of life missing data. *Quality & quantity*, *44*(2), 277–287.

Martin, D. P., von Oertzen, T., & Rimm-Kaufman, S. E. (2015). Efficiently exploring

multilevel data with recursive partitioning. *Society for Research on Educational Effectiveness*.

McCleary, L. (2002). Using multiple imputation for analysis of incomplete data in clinical research. *Nursing Research*, *51*(5), 339–343.

Nakagawa, S., & Freckleton, R. P. (2008). Missing inaction: the dangers of ignoring missing data. *Trends in Ecology & Evolution*, *23*(11), 592–596.

Nielsen, D. (2016). *Tree boosting with xgboost-why does xgboost win" every" machine learning competition?* (Unpublished master's thesis). NTNU.

O'Connell, A. A. (2010). An illustration of multilevel models for ordinal response data. In *Data and context in statistics education: Towards an evidence-based society. proceedings of the eighth international conference on teaching statistics (icots8).*

Onyango, N. O. (2009). On the linear mixed effects regression (lmer) r function for nested animal breeding data.

Opoka, R. O., Ssemata, A. S., Oyang, W., Nambuya, H., John, C. C., Karamagi, C., & Tumwine, J. K. (2019). Adherence to clinical guidelines is associated with reduced inpatient mortality among children with severe anemia in ugandan hospitals. *PloS one*, *14*(1), e0210982.

Opondo, C., Allen, E., Todd, J., & 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., & 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.

Peabody, J. W., Taguiwalo, M. M., Robalino, D. A., Frenk, J., et al. (2006). *Improving the quality of care in developing countries.*

Powney, M., Williamson, P., Kirkham, J., & Kolamunnage-Dona, R. (2014). A review of the handling of missing longitudinal outcome data in clinical trials. *Trials*, *15*(1), 237.

RColorBrewer, S., & Liaw, M. A. (2018). Package 'randomforest'.

Rodriguez, G., & Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: a case-study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(2), 339–355.

Royston, P. (2004). Multiple imputation of missing values. *The Stata Journal*, *4*(3), 227–241.

Schmidt, J. (2012). *Ordinal response mixed models: A case study* (Unpublished doctoral dissertation). Montana State University.

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using mice: a caliber study. *American journal of epidemiology*, *179*(6), 764–774.

Stekhoven, D. J. (2011). Using the missforest package. *R package*, 1–11.

Stekhoven, D. J. (2015). missforest: Nonparametric missing value imputation using random forest. *Astrophysics Source Code Library*.

Stekhoven, D. J., & Bühlmann, P. (2011). Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118.

Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., & Lee, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials? *Statistical methods in medical research*, *27*(9), 2610–2626.

Tang, F. (2017). Random forest missing data approaches.

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., … Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ open*, *3*(8), e002847.

Zhang, C., Kai, J., Feng, H. C., & Yang, T. (2013). The nearest neighbor algorithm of filling missing data based on cluster analysis. In *Applied mechanics and materials* (Vol. 347, pp. 2324–2328).

Zhang, Y., Alyass, A., Vanniyasingam, T., Sadeghirad, B., Flórez, I. D., Pichika, S. C., … others (2017). A systematic survey of the methods literature on the reporting quality and optimal methods of handling participants with missing outcome data for continuous outcomes in randomized controlled trials. *Journal of clinical epidemiology*, *88*, 67–80.