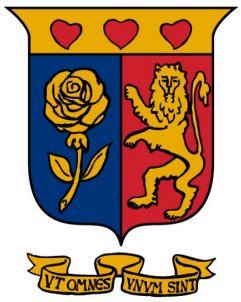
2021

A Joint modelling approach of monthly anthropometry and time to death among hospitalized severe malnourished children in Kenya.

Maronga, Christopher Sianyo
Strathmore Institute of Mathematical Sciences
Strathmore University

A JOINT MODELLING APPROACH OF MONTHLY ANTHROPOMETRY AND TIME TO DEATH AMONG HOSPITALIZED SEVERE MALNOURISHED CHILDREN IN KENYA



Strathmore
UNIVERSITY

Christopher Sianyo Maronga

122458

A thesis submitted to Strathmore University, Institute of Mathematical Sciences in fulfillment for the requirement of Master of Science in Statistical Science.

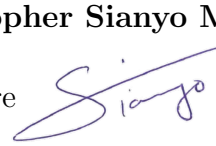
September 2021

Declaration and approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Christopher Sianyo Maronga

Signature 

Date **13th September 2021**

Approval

The thesis of Christopher Sianyo Maronga was reviewed and approved by the following:

Dr. Elphas Okango,
Lecturer, Imstitute of Mathematical Sciences,
Strathmore University.

Dr. Moses Ngari,
Biostatistician, KEMRI/Wellcome Trust Research Programme.

Dr. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

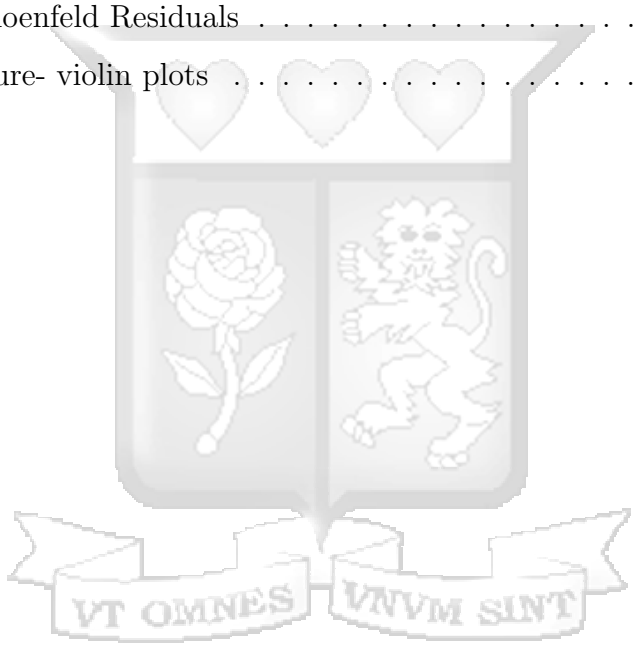
Background: In follow up studies, interest often lies in understanding the association between biomarkers measured over time and a time-to-event outcome. For this, a two-stage separate analysis or the use of time-dependent Cox models are often used. The former approach does not account for shared features between the two processes while the latter ignores the endogeneity in the biomarker, resulting in inefficient and biased estimates. The objective of this project was to fix joint models on longitudinal anthropometry and time to death among children hospitalized with complicated SAM in four hospitals in Kenya.

Methods: Data from a randomised placebo-controlled trial for 1,778 children aged 2 to 59 months admitted to hospital with complicated Severe Acute Malnutrition (SAM) but without HIV was analysed. We used Linear mixed effects models to model longitudinal anthropometry and Cox proportional hazards model to assess the effect of a priori selected baseline covariates on mortality. The two models were linked through current value and slope association to create a joint model used to study the effect of longitudinal anthropometry on risk of death.

Results: The joint model results showed that a unit centimetre gain in monthly mid-upper arm circumference (MUAC) was associated with 46.8% reduction in hazard of death, 0.532(95% CI: 0.476-0.596), while a unit gain in standard deviation (SD) for weight-for-height (WHZ) was associated with 37.1% reduction in the risk of death, 0.629(95% CI:0.579-0.683). A unit gain in SD for monthly weight-for-age (WAZ) and height-for-age (HAZ) was associated with 21.2%, 0.788(95% CI: 0.742-0.837) and 2.5%, 0.227(95% C.I: 0.008 - 6.556) reduction in risk of mortality respectively.

Conclusion: In studying the relationship between survival outcome and covariates, researchers often use baseline values of the covariates which fails to account for the inter-dependencies. Using joint modelling framework, we quantified the association between four longitudinal anthropometry and risk of death. Through current value and slope association, MUAC and WHZ have the strongest association with risk of death respectively hence are better metrics and can be used to screen and identify high-risk children.

Table of Contents



List of Figures

List of Tables

Acknowledgment

I thank the Almighty God for the knowledge, strength and good health during the study and research period.

I would like to express my heartfelt gratitude to my supervisors; Dr. Elphas Okango and Dr. Moses Ngari for their invaluable supervision and guidance that they offered to me during my Masters study. My sincere thanks extends to KEMRI-Wellcome Trust Research Programme and Professor James Berkley for the funding opportunity to undertake my studies at Strathmore University.

I also acknowledge the support of my fiancee Perpetual Njeri for her continued understanding and encouragement, without which, it would have been impossible for me to complete my study. I thank my colleague Céline Bourdon for her enormous input. Her immense knowledge and plentiful experience have helped me shape the ideas for my project.

Lastly, I would like to extremely thank my course mates (Mwikali Lilian, Gakuo Stephen, Joram Andrew, Bore Laban, Kipngeno Victor, Odok Eric) and Strathmore Institute of Mathematical Sciences (SIMS) for offering an enabling environment for learning.

Abbreviations

CGH	Coast General Hospital
CTX	Daily co-trimoxazole prophylaxis to prevent mortality in children with complicated severeacute malnutrition
GEE	Generalized Estimating Equations
GLMMs	Generalized linear mixed models
HAZ	Height-for-age z score
HIV	Human immunodeficiency virus
KCH	Kilifi County Hospital
LMMs	Linear mixed-effects models
MAR	Missing at random
MUAC	mid-upper arm circufrence
SAM	Severe acute malnutrition
UNICEF	United Nations Children's Fund
WAZ	Weight-for-age z score
WHO	World Health Organization
WHZ	Weight-for-height z score

Chapter 1

Introduction

1.1 Background

In most epidemiological studies and clinical trials, patients are followed up over time, during which longitudinal data is collected such as disease biomarker and time to event outcome. These types of outcomes are traditionally analysed separately or by use of time-varying Cox model[35]. However, these approaches have limitations, and can fail to account for the shared random effects between the longitudinal and survival process thus leading to biased parameter estimates[35,57].

More often, the researcher is interested in monitoring the disease progression with an aim to understand how the patient's longitudinal measurement or disease biomarker is associated with the discrete time-to-event outcome. Joint models are widely used in these kinds of scenarios because they analyse the two outcomes simultaneously thereby accounting for their inter- dependencies[57]. Arguably, follow up data from the same individual is most likely related to the event of interest making these longitudinal biomarkers endogenous; a characteristic that is not considered when using the aforementioned methods.

Malnutrition is a global health problem with relatively high prevalence in developing countries, and an important risk factor for mortality in children[6,7]. Studies have shown that mid-upper arm circumference (MUAC) is a significant predictor of mortality in children below the age of five years[39]. Severe acute malnutrition (SAM) in children aged between 6 and 59 months is defined by either weight-for-height (WHZ) z-score <-3 standard deviations from World Health Organization (WHO) defined growth standards or MUAC <11.5 cm or presence of nutritional oedema[44].

The common approaches used for analysing longitudinal and time-to-event data involves separately analysing longitudinal measures using linear mixed effect models with random effects to assess baseline covariates on evolution of longitudinal outcome and the use of

Cox proportional hazard regression to assess effect of covariates of interest on time-to-event outcome. While many studies have used the first approach[19,63], it may not appropriately account for random-effects shared by the linked processes of longitudinal evolution and time-to-event outcome. Using time-varying Cox models is an improved method but can also produce biased estimates for the time-to-event outcome due to endogenous nature of the longitudinal biomarkers[35]. By endogenous we mean that the biomarker is internal to the patient i.e., the measurement can only be realised if the patient is a live at a given time-point. If the patient dies, no further measurements for the outcome can be made.

Due to the mentioned issues, we are propose using joint modelling framework to assess the association of serial longitudinal measurement of anthropometry with risk of dying in children. Joint models simultaneously estimate both the longitudinal and time-to-event components and jointly estimate the hazard of time-to-event outcome conditional on the longitudinal outcome. The framework is based on conditional joint density as opposed to separate marginal independent densities derived from mixed-effects sub-model (for longitudinal outcome data) and Cox regression for time-to-event outcome(survival sub-model). In joint models, the longitudinal and survival models are linked through shared parameters allowing the time-to-event model to depend on features of the longitudinal sub-model, thus sometimes referred to as shared-parameter model[57].

1.2 Problem Statement

Most analysis associating the effect of anthropometry measurements with risk of death have used a single baseline measurement or time-varying cox models, which leads to an underestimation of the association between longitudinal anthropometry and death. Furthermore, there is little data on the relation between monthly anthropometry and risk death despite hospitalization being an important marker of susceptibility[40]. Changing anthropometry as the child recovers is likely related to risk of subsequent readmission or death[42]. By considering only the baseline anthropometry, information on the longitudinal history is not utilized to understand the association between the biomarker and the risk of death.

Daily co-trimoxazole prophylaxis to prevent mortality in children with complicated severe acute malnutrition(CTX) is a two-arm, double-blind, randomised, placebo-controlled trial

looking at survival of children in four hospitals in Kenya (2 in rural areas, i.e. Kilifi and Malindi, and 2 in urban setting i.e. Nairobi and Mombasa). To optimally evaluate the complete history of longitudinal data available for this study, we propose a joint modelling framework. Joint modelling offers a great benefit in linking longitudinal measurement of anthropometry and time-to-event outcome (death).This enables us to assess the effect of longitudinal process (anthropometry measurement) on mortality while allowing for unequally spaced longitudinal measurements, missing covariate data or censoring of survival times[35], which are the common challenges inherent to clinical data collection.

Joint models are alternative approach to using time-varying Cox model and provide higher prediction accuracy because they can account for individual's within and between variability. This characterization of a child's risk of death as their anthropometry changes, is important for the timing and evaluation of interventions.

1.3 Objectives

The main objective is to assess the association between longitudinal anthropometric measures and the risk of death 12 months post discharge using a joint modelling approach.

1.3.1 Specific objective:

i To evaluate the effect of longitudinal changes in each of the four anthropometric measurements MUAC, WHZ, WAZ and HAZ on 12 months mortality in children.

ii To determine the predictors of mortality among hospitalized severe malnourished children.

1.4 Scope

The focus is to apply joint modelling technique to assess the effect of longitudinal measurements (anthropometry) on the risk of death. We used linear mixed effects models (with random slope and intercept) as our longitudinal sub-model and Cox regression model with

selected baseline characteristics as the survival sub-model. However, Joint modelling with multivariate longitudinal data and dynamic predictions are beyond the scope of this work.

Chapter 2

Literature review

2.1 Joint modelling

Most often in observational follow up studies or clinical trials, different types of outcomes are collected mainly multiple longitudinal responses e.g., disease biomarkers such as blood pressure, MUAC measurements, blood biochemistry values, etc. and time-to-event for a particular pre-defined outcome of interest e.g., death, relapse of a disease, etc. Other implicit outcomes are generated such as missing data due to patient drop-out, lost to follow up and patient visit mechanism which is usually random in nature (non-equal visit intervals for subjects). Methods for analysing these separate outcomes are well established in the literature[15,27,28]

More than often in research or medicine, it becomes important to understand the relationship between the changes of a disease biomarker with time and how these changes affect the survival or recovery of a patient or disease progression. This is a perfect description of relating the two outcomes, i.e., one outcome measured repeatedly over time and the other indicating the presence or absence of an event or a condition of interest. For instance, using joint modelling, studies have showed that declining CD4 counts overtime coincide with disease progression and mortality of HIV patients initiated on ART[9,35].

Traditionally, the two kinds of outcomes are analysed separately or by use of time-varying Cox model[19,56]. However, separate analysis of these two outcomes does not provide the association that exists between the hazard of an event and the repeatedly measured longitudinal outcome.[53]. The latter approach assumes that the longitudinal biomarker are exogenous time-varying covariates, and this leads to inefficient or biased estimates for the time-to-event model[35]. To mitigate the short-coming of separate analysis and use of time-varying Cox model, a modelling framework has been proposed known as joint modelling that can adequately associate endogenous time-varying covariates and survival outcome[30,52].

The whole idea of joint modelling is to model simultaneously the two outcomes by combining two models; a survival sub-model together with an appropriate sub-model for repeated measurement that takes into account the special features of the longitudinal outcome. The key advantage of joint modelling is the ability to handle irregularly and imperfectly measured time-varying covariates correctly and produce robust parameter estimates and an estimate of the association between a longitudinal outcome and time-to-event outcome[1]. This formulation has an applicable practical advantage, for instance, you can use joint modelling to associate monthly anthropometry with risk of life-threatening events in SAM children in order to understand the subject-specific effects of the longitudinal biomarkers on the risk of death.

2.2 Longitudinal data

The key feature of longitudinal studies is that data is measured repeatedly from the same subject or observational unit over time. This contrasts with cross-sectional studies in which outcome data is measured at a single time-point across subjects. Longitudinal measurements (i.e., repeated measures obtained in patients' overtime) can enable direct study of change through time within subjects, which is a key benefit of conducting longitudinal studies. However, repeated measurements on the same subject are correlated and thus violate the independence assumptions required by many statistical methods and hence specialized statistical modelling technique that takes into account the within-subject correlation are needed.

The most popular methods for analysing longitudinal data that accounts for correlation structure include the Generalized Estimating Equations (GEE)[2,27,28] and Generalized Linear Mixed Effects Models (GLMMs)[15,22,58]. The use of marginal GEE model is appropriate when the focus of inference is about the population, sometimes referred to as population-average models while GLMMs are useful when subject-specific inference is of interest, as they include random effects in additional to fixed effect predictors[24,38]. GLMMs are powerful when the model and distribution of the outcome are correctly specified while GEE approach is semi parametric and thus more robust to model miss-specification[3], thus GEE is the modelling choice when the distribution of the outcome is uncertain since the model

parameters and standard error estimates will be consistent even if miss-specified [48].

Some of the key advantages of longitudinal studies over cross-sectional studies include; higher statistical power in relation to sample size (i.e., fewer observed units are required within a longitudinal framework), ability to separate aging effects (i.e. changes over time within individuals), from cohort effects (i.e. differences between subjects at baseline) and ability to provide subject-specific information over time, which cannot be achieved using cross-section studies hence making it possible to understand differences in the population and factors that affect growth at an individual level. More benefits of longitudinal studies have been documented in-depth [20,29].

Despite the rich benefits that come with conducting longitudinal studies, challenges arise with this kind of study design such as missing data hurdles. For instance, subjects in an observational or cohort study can drop-out or re-join during follow up (creating unbalanced data). This introduces random or non-random missing data problem and intermittent measurement of outcome variable, which may lead to biased estimates. Repeated measures over-time is highly likely to show intra-subject correlation rendering the use of several traditional models inadequate due to their assumptions of independence, and lastly, the difficulty in incorporating associations of time-varying covariates (e.g., weight, monthly anthropometry or CD4 count, smoking status, family income or structure, seasonality etc) [21].

2.3 Survival analysis and time-to-event data

In traditional logistic regression modelling, the key interest is understanding how risk factors affect the presence or absence of a disease (i.e., binary outcome), however in some instances, a researcher is interested in how these risk factors or treatment affects time to disease onset or time-to some other relevant event (e.g., mortality) [13]. For the latter objective, binary logistic regression cannot be used, but survival analysis, also known as, time-to-event analysis, can evaluate these relationships. The event can be time until first employment after graduating from college, time until tumour recurrence in cancer research studies, time until AIDS for HIV patients, time until death for mortality studies, time until readmission etc. The response is usually continuous and often referred to as failure time, survival time or event time [13,16] and the quantities of interest include survival probability (i.e., probability of an event happening

beyond a given time), hazard rate (instantaneous potential of experiencing an event at a given time, given survival to that time) and cumulative hazard rate (cumulative force of mortality).

The main feature of survival analysis is that the data is often censored, i.e., while we know some information about a subject's event time, we don't know the exact event time or survival time due to lost to follow up[33]. There are several types of censoring mechanisms as discussed in the following texts[16,31]. For any given censoring mechanism, it must be assumed to be non-informative i.e., the censoring mechanism is independent of the event of interest, in other words the censoring is caused by a different process beside the impending failure. Survival times are usually continuous and highly skewed and for this reason models that assume normality cannot be used to analyse this data. Censoring arising from patient drop-out and lost to follow up potentially introduce non-random missing data that needs to be taken into account during statistical modelling.[14]

In estimating the survival probability and how various covariates influence the survival time of individuals, several parametric approaches have been proposed. These models make distributional assumptions about the failure times and parameter estimates are obtained for inference. The most popular examples include exponential, Weibull, Gamma, Log-normal and log-logistic. Details about these models and more can be found in these texts[16,32,36].

The most used model to explore the relationship between explanatory variables and survival is the semi-parametric Cox regression model[17,54]. This model assumes that the relationship between the covariates and the failure time is exponential while the baseline hazard is left unspecified, hence the semi-parametric aspect. In addition, the Cox regression models makes a proportional hazard assumption for the covariate i.e., the multiplicative hazard for each covariate is constant over time.

2.4 Malnutrition and post-discharge mortality

Child malnutrition remains a major global health problem in developing countries and a main contributing factor to child morbidity and mortality, when associated with other infectious diseases. More than half of childhood deaths globally are linked to malnutrition, with high rates of mortality among hospitalized SAM children; especially in sub-Saharan Africa[5,34].

These findings are also concurred in the recent report by United Nations Children's Fund (UNICEF). More than one in five children under the age of 5 had stunted growth in 2019 (21.3%) with nearly two out of five of these children living in sub-Saharan Africa. Globally, 47 million children under 5 were wasted with 14.3 million being severely wasted in 2019, a quarter of these wasted children live in sub-Saharan Africa[45,46]

SAM is defined by WHO and UNICEF as weight-for-height (WHZ) of below -3 standard deviations or mid-upper arm circumference (MUAC) of less than 115mm among children aged between 6 to 60 months[47]. Children that meet this criterion have an elevated risk of dying. Studies have reported high mortality rates following discharge of children hospitalized with SAM and other infections[10,12,41,43]. A recent study conducted in Kilifi Kenya reported that child mortality was seven-fold higher among post-discharge children than those within the community of the same age. Further, the study found that of all the children discharged from the hospital, 4.5% died within one year[37]

Understanding the association between longitudinally measured anthropometry is key to monitoring recovery of children with SAM following hospitalization as well as inform early interventions such as admission to therapeutic programmes or commencement of therapeutic diet/supplementary feeds. It could also inform hospital discharge criteria and continued required medical care for children with impending risk of life-threatening events such as death and/or re-hospitalization. The difficulty of determining a child's nutrition status based on a single index has been reported[25] and thus using several combined anthropometry measurements obtained longitudinally would give a better understanding and screening of the population at risk of death due to SAM and a solid trace of their growth & disease recovery trajectories.

Chapter 3

Methodology

3.1 Introduction

In this chapter, we describe the study design and population. Details of the statistical models used are provided in depth and a discussion of the choices of the parametrization for the joint modelling framework. In the last section, we give an overview of the statistical software choice and preference of the packages used to implement the joint model.

3.2 Data source and Study design

In this analysis, we used data from KEMRI-Wellcome Trust Research Programme, specifically secondary data from a multicentred trial that aimed to assess the efficacy of the antibiotic co-trimoxazole as a prophylaxis treatment to reduce post-discharge mortality and/or hospital readmission. This study was a two-arm, double-blind, randomised, placebo-controlled trial that was conducted in four hospitals in Kenya (2 in rural setting, i.e. Kilifi and Malindi, and 2 urban i.e. Nairobi and Mombasa). This trial recruited 1,778 children aged 2 to 59 months while admitted to hospital with complicated SAM but without HIV. Children were randomized to take either co-trimoxazole or placebo for the first six months and were followed up for 12 months (monthly for the first six months and every 2 months thereafter up to month 12) with a approximately over 14,500 follow-up measurements. Interested readers are directed to this paper[4] for more details about the trial design and reported results.

Anthropometry measurements such as MUAC, height, weight etc. were taken during follow-up visits. Other information collected included study product adherence and household geographical location. The z-scores were derived from the raw anthropometry variables. The recruitment period spanned between November 20, 2009 and March 14, 2013.

3.3 Covariates and study outcomes

The primary outcome was mortality during the 365 days of study duration, which was assessed monthly for a duration of 12 months. Secondary endpoints included frequency of non-fatal illness episodes resulting in readmission to hospital, illness episodes treated as outpatient and nutritional recovery. The longitudinal covariates evaluated include; MUAC, WHZ, WAZ and HAZ.

The anthropometric z-scores were computed using the WHO 2006 child growth reference [18,26]. Child's age, gender, randomization arm, and hospital/site (whether urban or rural), selected admission diagnosis and composite indicator for presence of underlying medical condition were used as baseline covariates for the survival process. Repeated measures comprise of the anthropometric measurements for the same child done at different time points. Measurement times are the follow-up times for each hospital visit or phone follow-up assessment defined by months elapsed between study entry and a hospital visit.

3.4 Statistical Analysis

The joint model is fully defined using two linked sub-models: a longitudinal model for the disease biomarker, in this case monthly anthropometry and a relative risk model for the time-to-event outcome [1,59]. The linkage of the two models is done through shared parameters as detailed in the sections that follow.

The notations that following will be used to formulate the joint model based on the two mentioned sub-models. Let $P_n = \{T_i, \upsilon_i, y_i; i = 1, 2, \cdots, n\}$ represent a sample from our data, where T_i^* denotes the true time to event for the i^{th} child, true time to death or hospital admission: C_i the censoring time such that $T_i = min(T_i^*, C_i)$ is the corresponding observed time for the i^{th} child and $\upsilon_i = I(T_i^* \leq C_i)$ is the indicator for the event of interest such that $I(.)$ takes the value 1 when $T_i^* \leq C_i$, otherwise censored (takes value 0).

3.4.1 Longitudinal sub-model

Longitudinal data is a grouped data with repeated measurements from the same individual over time, in this case monthly anthropometry. The hierarchy or grouping is at the individual level resulting in correlations that should be taken into account. I used linear mixed effects model for analysing continuous, hierarchical data[23] as they account for the correlation and include subject-specific coefficients (random slopes and intercepts) that cannot be explained by measured covariates.

Consider a longitudinal study with a sample size of N. Let y_{ij} denote the measurement of continuous longitudinal outcome y in our case the anthropometry for i^{th} child at the j^{th} time point for $i = 1, 2, \cdots N$ and $j = 1, 2, \cdots n_i$; where as defined above, T_i^* denotes the true event time for the i^{th} child.

The linear mixed effect model for the correlated monthly anthropometry is formulated as below:-

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t_{ij} + \epsilon_{ij} \tag{3.1}$$

where:

- y_{ij} is the jth anthropometry value for the ith child measured with error.
- t_{ij} is the jth observational time point for the ith child.
- β's are fixed effects (intercept β_0 and slope β_1).
- b_i's are random effects (random intercept b_0 and random slope b_1).

ϵ_{ij} is a subject specific random error with mean zero and constant variance.

$i = 1, 2, \cdots N$ i.e number of subjects and $j = 1, 2, \cdots n_i$ number of timepoints for the ith child

The general form of the postulated linear mixed effects model is given below.

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \epsilon_i \\ b_i \sim N(0, D), \quad \epsilon_i \sim N(0, \sigma^2 I_{ni}) \end{cases}$$

where:

X is a design matrix for fixed effects β

Z is a design matrix for random effects b_i

b_i and ϵ_i are independent

3.4.2 Time-to-event sub-model

Event time is the time elapsed up to the occurrence of a clinical event of interest e.g. death, re-occurrence of a disease etc. The semi-parametric Cox proportional hazard model was used for analyzing time-to-event data also reffered to as survival data; in this case time until death of a child.

The survival sub-model is formulated as:

$$h_i(t) = h_0(t) \exp(A^T W_i) \tag{3.2}$$

where:

$h_i(t)$ is the hazard for an event for child i at time t

$h_0(t)$ is the baseline hazard

W_i is a vector baseline covariates with their corresponding regression co-efficients A^T

The quantity h_0 describes the risk of the event occurring, given it has not yet occurred when all covariates W_i have value zero. It is assumed to be invariant across all individuals while the effect parameters $A^T W_i$ describe how the hazard varies as a function of the explanatory covariates W_i.

3.4.3 Mathematical implementation of joint modelling

The longitudinal model and the survival model are linked through shared parameters that induces the dependence between them. This implementation models the hazard for an event of interest as depending on a function of subject-specific linear predictor $\mu_i(t)$, so for the survival process, the model contains extra parameters from equation 3.1 [53].

$$h_i\big(t|H_i(t), W_i(t)\big) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\big[t \leq T_i^* < t + \Delta t | T_i^* \geq t, H_i(t), W_i(t)\big] \quad (3.3)$$

$$= h_0 \exp\big(A^T W_i(t) + f\{H_i(t), b_i(t), \alpha\}\big) \quad (3.4)$$

$$= h_0(t) \exp\{A^T W_i + \alpha \mu_i(t)\} \quad (3.5)$$

where:

- $H_i(t) = \mu_i(k), 0 \leq k < t$ is the history of underlying longitudinal process (monthly anthropometry) up to time t
- $\mu_i(t)$ is assumed to be the true and unobserved anthropemetric measure at time t

The vector parameter α quantifies the strength of association between the longitudinal process i.e. monthly anthropometry and the hazard of the event (death) at the same time point t.

Based on the type of association structure between longitudinal and time-to-event processes, various forms of $f(.)$ can be defined for the joint model to form the basis for inference. Depending on what parametres these two models share, we discuss three most common formulation of joint models: current value association, current value and slope association, and shared random effect association.

Current value association

Under this framework, the assumption is that for any patient i, the true observed value $\mu_i(t)$ of the longitudinal measure at time t is associated with the risk $h_i(t)$ of occurrence or non-occurrence of the event at the same time. Thus the Cox model with this association

structure becomes

$$h_i(t) = h_0(t) \exp\{A^T W_i + \alpha_1 \mu_i(t)\}$$

Hence, $\mu_i(t)$ is associated with $h_i(t)$ at each and every time point. With this association, specifying the correct longitudinal sub-model for $y_i(t)$ accurately is key to good estimates of α_1.

The co-efficient α_1 measures the strength of the association between $\mu_i(t)$ and the time-to-event process at the same time (*a unit increase in* $\mu_i(t)$ *leads to an exp(*α_1*) multiplicative increase in patient's risk of experiencing the event at time* t).

The coefficient α_1 between longitudinal value and risk of an event is the same across all patients but $h_i(t)$ vary across individuals as a function of other covariates w_i

Current value and slope association

This association structure allows the risk of event to be associated additionally with the rate of change of $\mu_i(t)$ i.e. one patient may have an increasing trajectory/profile, whereas another decreasing trajectory/profile even if they have the same longitudinal score $\mu_i(t)$.

This estimate of rate of change is the derivative of $\mu_i(t)$ with respect to time.

$$\mu_i^{'}(t) = \frac{d\mu_i(t)}{dt}$$

Hence, the sub-model with this type of association becomes:-

$$h_i(t) = h_0(t) \exp\{A^T W_i + \alpha_1 \mu_i(t) + \alpha_2 \mu_i^{'}(t)\}$$

Here, the hazard of experiencing the event at time t is assumed to be affected by both the value of $\mu_i(t)$ and it's rate of change or slope $\mu_i^{'}(t)$. The coefficient α_2 measures the strength of association between the true rate of change of trajectory of $\mu_i(t)$ and time-to-event process at time t.

Patients having the same level of $\mu_i(t)$, the hazard rate for a one-unit increase of the current

rate of change $\mu_i'(t)$ is $\exp(\alpha_2)$, while for individuals having the same rate of change $\mu_i'(t)$, the hazard rate for a one-unit increase of the current value $\mu_i(t)$ is $\exp(\alpha_1)$.

Shared random effects association

In this association structure, only random effects parameters from the longitudinal model are shared with time-to-event process as linear predictors of this survival sub-model (both random intercept and slope if specified in the longitudinal sub-model).

In the shared random-effects association structure, we could state that for individuals who have a lower (or higher) measure at baseline (i.e., random intercept) or who show a steeper decrease (or increase) in their longitudinal trajectories (i.e., random slope) are more or less likely to experience the event. The sub-model takes the form

$$h_i(t) = h_0(t) \exp\{A^T W_i + \alpha^T b_i\}$$

Where b_i are the random effects, which implies that for a specific patient i, b_i is unique regardless of the time t as opposed to the current value of $\mu_i(t)$ and current rate of change $\mu_i'(t)$

In this analysis, we used the current value and slope association structure in frequentist approach and inference.

3.5 Parameter estimation in joint models

In order to estimate parameters in the joint modeling framework, the below two methods are widely used[61]:

1. Two-stage approach method
2. Likelihood estimation method

In this study, we used the likelihood method to to estimate the parameters of the joint model. However, I highlight both approaches while putting more emphasis on the implementation of likelihood estimation.

3.5.1 Two-stage methods

The two-stage approach is implemented in two steps:

1. Step 1: The estimates of the parameter and predictions are calculated for longitudinal outcome without taking into account survival outcome.
2. Step 2: The survival model is fitted by utilizing the predicted longitudinal values as true covariates.

As much as the two-stage approach is pretty easy to be implemented with the existing software, it has been previously shown from simulation studies that it often produces biased results[61] because it does not utilize information from the longitudinal and the survival process simultaneously in each model fitting steps.

3.5.2 Likelihood methods

Semi-parametric maximum likelihood method proposed by[30] is the popular estimation method for joint models. Within the likelihood method, we make an assumption of full conditional independence. This implies that the random effects represented takes into account both the association between the longitudinal and time-to-event outcomes, and the correlation between the repeated measures in the longitudinal process. This framework is formulated as below:

$$p(T_i, \delta_i, y_i|b_i; \theta) = p(T_i, \delta_i|b_i; \theta)p(y_i|b_i; \theta) \tag{3.6}$$

$$p(y_i|b_i; \theta) = \prod_j p\{y_i(t_{ij})|b_i; \theta\} \tag{3.7}$$

Assuming non-informative censoring and visiting processes (censoring, timing, and measurement processes depend only on the observed history and latent random effects and not on the future risk time itself), the log-likelihood contribution for the i^{th} subject is formulated as:

$$\begin{aligned}\log p(T_i, \delta_i, y_i; \theta) &= \log \int p(T_i, \delta_i, y_i, b_i; \theta)\ db_i \\ &= \log \int p(T_i, \delta_i | b_i; \theta) \Big[\prod_j p\big\{ y_i(t_{ij}) | b_i; \theta \big\} \Big] p(b_i; \theta)\ db_i,\end{aligned} \tag{3.8}$$

with θ the parameter space, y_i the longitudinal information of the i^{th} subject, δ_i the event indicator.

$$\begin{aligned}p(T_i, \delta_i | b_i; \theta) &= h_i(T_i | \mathrm{M}_i(T_i); \theta)^{\delta_i} S_i(T_i | \mathrm{M}_i(T_i); \theta) \\ &= \Big[h_0(T_i) \exp\big\{ \boldsymbol{\gamma}^T \mathbf{w}_i + \alpha m_i(T_i) \big\} \Big]^{\delta_i} \exp\left(- \int_0^{T_i} \Big[h_0(t) \exp\big\{ \boldsymbol{\gamma}^T \mathbf{w}_i + \alpha m_i(t) \big\} \Big]\ dt \right).\end{aligned} \tag{3.9}$$

The joint density for the longitudinal outcome together with the random effects will take the form:

$$\begin{aligned}p(y_i | b_i; \theta) p(b_i; \theta) &= \prod_j p\big\{ y_i(t_{ij}) | b_i; \theta \big\} p(b_i; \theta) \\ &= (2\pi\sigma^2)^{n_i/2} \exp\Big\{ - \frac{\| yi - X_i\beta - Z_i b_i \|^2}{2\sigma^2} \Big\} \\ &\times \quad (2\pi)^{q_b/2} \det(\mathrm{D})^{1/2} \exp\left(b_i^T D^1 b_i / 2 \right),\end{aligned}$$

where q_b is the dimensionality of the random-effects vector, and $\|x\| = \big\{ \Sigma_i x_i^2 \big\}^{1/2}$ indicates the Euclidean vector norm.

Maximizing the log-likelihood function $\mathrm{l}(\theta) = \sum_i \log \mathrm{p}(T_i, \delta_i, y_i; \theta)$ with respect to θ can be achieved using either Expectation-Maximization (E-M) algorithm or the Newton-Raphson [52]. Unfortunately, the main weakness of the E-M algorithm is slow convergence at the maximum [53].

Implementation of Expectation-Maximization Algorithm

Two steps exist in this process: the Expectation-step also known as the E-step and the Maximization-step also known as the M-step [51]. In the Expectation-step, we fill the missing

data using the observed data and current parameters estimates by conditional expectation; in the Maximization-step, we maximize the conditional expectation from step one. The observed log-likelihood function can be expressed as:

$$\log(\theta) = \sum_{i=1}^{n} \{\log p(T_i, \delta_i | b_i; \theta, \beta) + \log p(yi | b_i; \theta) + \log p(b_i; \theta)\}$$

In the E-step the expected value of the complete log-likelihood function given the conditional distribution of $\mathbf{b}_i$ is:

$$\mathrm{Q}(\theta|\theta^m) = \sum_{i=1}^{n} \int \{\log p(T_i, \delta_i | b_i; \theta, \beta) + \log p(yi | b_i; \theta) + \log p(b_i; \theta)\} p(b_i | T_i, \delta_i, y_i; \theta^m)\, db_i,$$

The two integrals in $\mathrm{Q}(\theta|\theta^m)$ need to be solved numerically using Gaussian quadrature rules or Monte Carlo sampling.

For the M-step, closed-form solutions can be obtained for the variance and covariance matrix of residuals and random effects respectively. However, the fixed effects for every longitudinal model and parameters in the time-to-event model has to be solved numerically. The main steps for these parameters are as follows:

Step 1: Estimate the variance of residuals of each longitudinal model by closed form expressions:

$$\begin{aligned} \hat{\sigma^2} &= \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} (y_i - X_i \beta^m)^T \left(y_i - X_i^T \beta^m - 2 Z_i^T \mathrm{E}(b_i | T_i, \delta_i, y_i; \theta^m)\right) \\ &\quad + \mathrm{tr}(Z_i^T Z_i \mathrm{Var}(b_i | T_i, \delta_i, y_i; \theta^m)) + \\ &\quad \mathrm{E}(b_i | T_i, \delta_i, y_i; \theta^m)^T Z_i^T Z_i \mathrm{E}(b_i | T_i, \delta_i, y_i; \theta^m) \end{aligned}$$

where tr is the trace function of a matrix, and E, the expectation function.

Step 2: Estimation of variance-covariance matrix of random effects b_i by

$$\begin{aligned} \hat{D} &= \frac{1}{n} \sum_{i}^{n} \mathrm{Var}(b_i | T_i, \delta_i, y_i; \theta^m) \\ &\quad \mathrm{E}(b_i | T_i, \delta_i, y_i; \theta^m) \mathrm{E}(b_i | T_i, \delta_i, y_i; \theta^m)^T \end{aligned}$$

Step 3: Since the parameters of the event time model (θ) and score equations for the fixed effect coefficient (β) lack closed form solutions, we proceed and implement the one-step Newton-Raphson algorithm for these parameters:

$$\begin{aligned}\hat{\beta}^{m+1} &= \hat{\beta}^{m} - \left(\frac{\partial S(\hat{\beta}^{m})}{\partial \beta}\right)^{-1} S(\hat{\beta}^{m}),\\ \hat{\theta}^{m+1} &= \hat{\theta}^{m} - \left(\frac{\partial S(\hat{\theta}^{m})}{\partial \theta}\right)^{-1} S(\hat{\theta}^{m}),\end{aligned}$$

where $\hat{\beta}^m$ and $\hat{\theta}^m$ are values of β and θ at the present iteration, and $\frac{\partial S(\hat{\beta}^m)}{\partial \beta}$ and $\frac{\partial S(\hat{\theta}^m)}{\partial \theta}$ are the corresponding blocks of the Hessian matrix (H-matrix) calculated at $\hat{\beta}^m$ and $\hat{\theta}^m$. The score vector's components then take the form:

$$\begin{aligned}S(\beta) &= \sum_{i=1}^{n} \frac{1}{\sigma^2} X_i^T \left(y_i - X_i^T\beta - Z_i^T \mathrm{E}(b_i|T_i,\delta_i,y_i;\theta^m)\right) + \delta_i \alpha x_i(T_i)\\ &\quad - \exp(\gamma^T w_i) \int\int_0^{T_i} h_0(u)\alpha x_i(s)\exp\left[\alpha\{x_i^T(s)\beta + z_i^T(s)b_i\}\right]\\ &\quad \times p(b_i|T_i,\delta_i,y_i;\theta)\, du\, db_i,\end{aligned}$$

Step 4: In this step, the survival model parameters could be updated too using the Newton-Raphson algorithm. Similarly, the baseline hazard function is estimated non-parametrically using piecewise-constant function. Score equations used in the Newton-Raphson algorithm are:

$$
\begin{aligned}
S(\gamma) &= \sum_{i=1}^{n} w_i \Big[\delta_i - \exp(\gamma^T w_i) \int \int_0^{T_i} h_0(u) \exp\big[\alpha\{x_i^T(s)\beta + z_i^T(s) b_i\}\big] \\
&\quad \times p(b_i | T_i, \delta_i, y_i; \theta)\, du\, db_i \Big],
\end{aligned}
$$

$$
\begin{aligned}
S(\alpha) &= \sum_{i=1}^{n} \delta_i \{(x_i^T(T_i)\beta + z_i^T(T_i) \mathrm{E}(b_i | T_i, \delta_i, y_i; \theta^m)\} \\
&\quad - \exp(\gamma^T w_i) \int \int_0^{T_i} h_0(u) \exp\big[\alpha\{x_i^T(s)\beta + z_i^T(s) b_i\}\big] \\
&\quad \times p(b_i | T_i, \delta_i, y_i; \theta)\, du\, db_i,
\end{aligned}
$$

$$
\begin{aligned}
S(\theta_{h_0}) &= \sum_{i=1}^{n} \delta_i \frac{\partial h_0(s; \theta_{h_0})}{\partial \theta_{h_0}^T} \\
&\quad - \exp(\gamma^T w_i) \int \int_0^{T_i} \frac{\partial h_0(T_i; \theta_{h_0})}{\partial \theta_{h_0}^T} \exp\big[\alpha\{x_i^T(s)\beta + z_i^T(s) b_i\}\big] \\
&\quad \times p(b_i | T_i, \delta_i, y_i; \theta)\, du\, db_i.
\end{aligned}
$$

The corresponding blocks of the H-matrix for $\frac{\partial S(\hat{\beta})}{\partial \beta}$ and $\frac{\partial S(\hat{\theta})}{\partial \theta}$ can be calculated by central difference approximation.

To get a solution of the expected likelihood function, a pseudo-adaptive Gaussian-Hermit quadrature rule can be utilized in approximating the integrals rizopoulos2012joint. The Expectation-step and the Maximization-step iterate to a pre-specified convergence criterion.

3.6 Data preparation and summary

Data wrangling and summary output was done in R[50] version 4.0.3. Data preparation involved creating two analogous flat files (i.e., long and wide) appropriate for each sub model. The long format (multiple rows for each individual) is suitable for longitudinal data analysis whereas the wide format (a single row for each individual) is used for survival analysis. All data munging and summary were done using tidyverse[60] group of libraries.

For exploratory data analysis, we created individual spaghetti plots to visualize the longitu-

dinal profiles as well average profile plots stratified by age group to compare the evolution of each anthropometric measurement over time. We also used a correlation matrix plot to visualize and compare the correlation structure across the four different longitudinal outcomes.

3.7 Software Implementation

All statistical analyses were conducted in R[50] version 4.0.3. Linear mixed models for longitudinal outcome were fitted using the {nlme} package[49] and Cox proportional hazard sub-model (for time-to-event outcome) was performed using the package {survival} from base R[55]. Refer to appendix 1 for a complete print out of R session containing all loaded packages during the analysis.

We used {JM}[51] developed by Dimtris Rizopoulos for the full implementation of the joint modelling framework. There exists other R packages that can be used to performing joint modelling as documented by Sezen Cekic in his tutorial on joint modelling[11].

Chapter 4

Results

4.1 Introduction

Data from CTX clinical trial conducted in four public hospitals in Kenya by the KEMRI-Wellcome Trust Research Programme was used. This chapter documents the analysis results. We explore the subject specific longitudinal profile, the mean and the correlation structure of the repeated anthropometric measurements.

We then present the model results for both longitudinal data analysis (mixed effects models) and time-to-event a.k.a survival analysis (Cox proportional hazard model). The joint models are discussed in the last part of the chapter.

4.2 Patients' demographic and baseline characteristics

The trial recruited a total of 1778 children across all the 4 sites, 887 (49.9%) children were randomised to the intervention arm while 891 (50.1%) received placebo. The mean age in months at index admission was 12.8 (SD=8.91), approximately 17.2% of the children were less than six months at enrolment. The screening criteria was MUAC<11.5cm and children had an overall mean of 10.6 (SD=1.07). Approximately 50% of the children were female and less than 15% were pre-mature births. Three hundred children had nutrional oedema at admission.

Over half of the of participants were diagnosed with diarrhoea and 656 (36.9%) had a pneumonia co-infections. About 7% of children were unconscious while 10.3% were treated for shock before enrolment.Haemoglobin levels at admission indicated a moderately anemic population with less than 8% being severely anemic. Up to 20% of the children had an underlying condition, such as tuberculosis, sickle cell disease, rickets, cerebral palsy and congenital

anomaly. Table 4.1

Table 4.1: Demographics and patient characteristics

	Survived (N=1522)	**Died (N=256)**	**Total (N=1778)**
Age in months-Median(IQR)	11.0 (53)	8.0 (47)	11.0 (53)
MUAC (cm)	10.6 (1.1)	10.2 (1.1)	10.6 (1.1)
Weight-for-age z-score(SD)	-3.9 (1.1)	-4.3 (1.2)	-3.9 (1.2)
Weight-for-height z-score(SD)	-3.2 (1.3)	-3.0 (6.5)	-3.2 (2.8)
Height-for-age z-score(SD)	-2.8 (1.6)	-3.1 (1.8)	-2.9 (1.7)
Younger than 6 months - Yes(%)	231 (15.2%)	75 (29.3%)	306 (17.2%)
Sex - Female (%)	745 (48.9%)	130 (50.8%)	875 (49.2%)
Nutrional oedema - Yes(%)	270 (17.7%)	30 (11.7%)	300 (16.9%)
Underlying medical condition - Yes(%)	293 (19.3%)	75.0 (29.3%)	368 (20.7%)
Haemoglobin (g/L)	9.83 (2.21)	9.66 (2.09)	9.80 (2.19)
Anemic status			
Mild	287 (18.9%)	44 (17.2%)	331 (18.6%)
Moderate	693 (45.5%)	123 (48.0%)	816 (45.9%)
Normal	382 (25.1%)	65 (25.4%)	447 (25.1%)
Severe	113 (7.4%)	19 (7.4%)	132 (7.4%)
Missing	47.0 (3.1%)	5 (2.0%)	52.0 (2.9%)
Premature birth - Yes(%)	177 (11.6%)	44 (17.2%)	221 (12.4%)
Admission for pneumonia - Yes(%)	533 (35.0%)	123 (48.0%)	656 (36.9%)
Admission for diarrhoea- Yes(%)	894 (58.7%)	127 (49.6%)	1021 (57.4%)
Treated for shock - Yes(%)	164 (10.8%)	20 (7.8%)	184 (10.3%)
Impaired consciousness - Yes(%)	90 (5.9%)	23 (9.0%)	113 (6.4%)

Numeric data are in Mean (SD). Underlying medical condition is a composite variable indicating presence of either tuberculosis, sickle cell disease, rickets, cerebral palsy or congenital anomaly

The primary outcome was death assessed monthly and then bi-monthly after 6 months for the period of 12 months. During the study, 256 (14.4%) of the children died;56 (3%) were lost to follow up while 36 (2%) voluntarily withdrew. A total 308 (17.3%) children had at least a single hospital readmission over this period.

4.3 Exploring subject specific profiles

The longitudinal measures (anthropometry) were evaluated monthly for the first six months and bi-monthly up to month 12. Values missing during follow up were considered missing at random (MAR), thus, we are able to track longitudinal trend by visualising the repeated

measurements.

4.3.1 Mid-Upper Arm Circumference (MUAC)

MUAC was used to screen for malnutrition at enrolment. Figure 6.3 shows spaghetti plot or subject-specific profile plot of MUAC index for all children. The figure shows presence of heterogeneity at baseline (each subject having differing intercepts) and a substantial variability within and between subject over time (differing average slope). In general, we observe a linear increasing trend for MUAC between baseline and month 12.

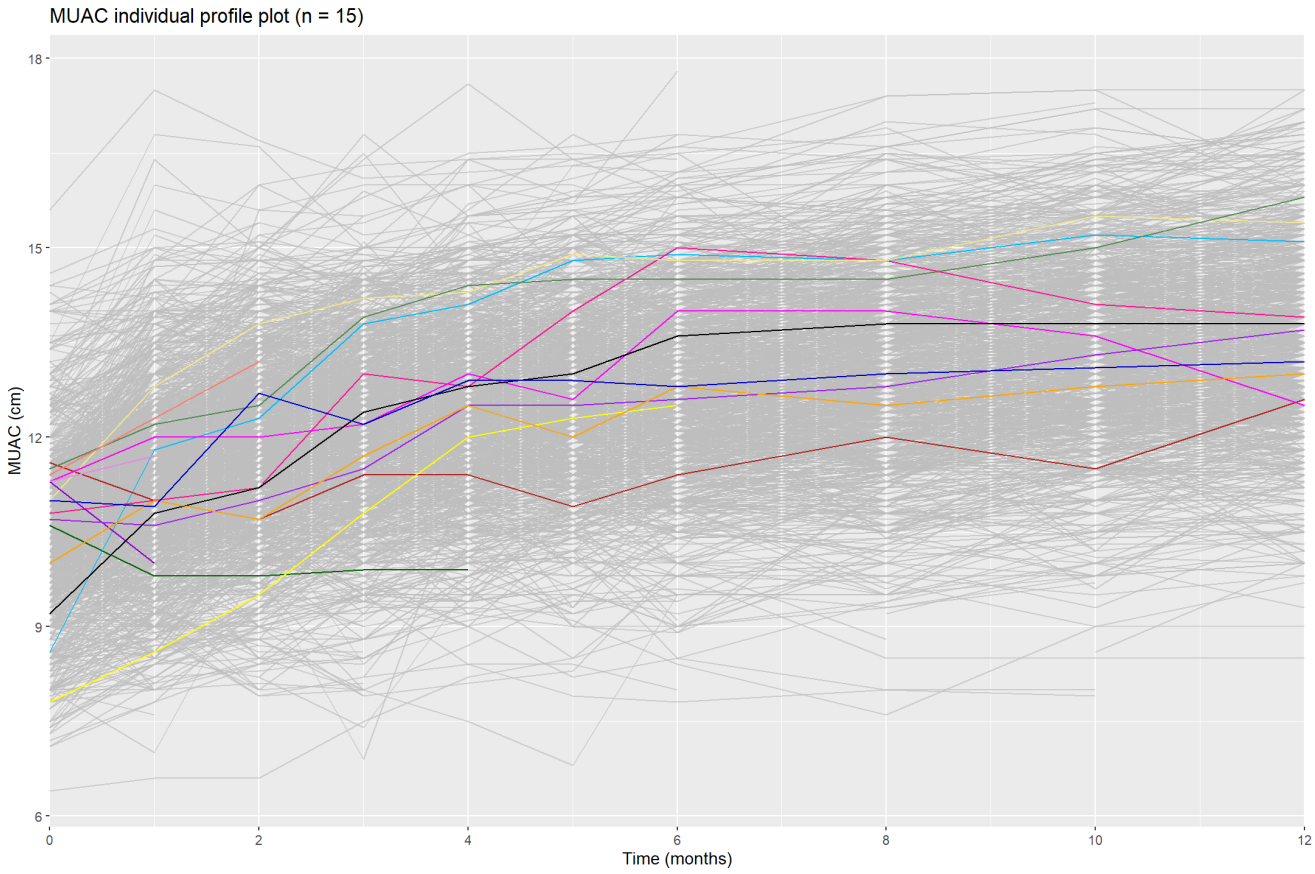


Figure 4.1: Spaghetti plot representing individual profiles of MUAC over time. Each grey line depicts the profile of a child, with colored highlights for 15 random subjects

4.3.2 Height-for-age Z-score (HAZ)

HAZ is a measure that indicates the deviance of a child's height from the median value of children his/her and measured as standard deviation (SD). Low values of HAZ is indicative of stunting in children. Figure 4.2 shows the profile plot for HAZ.

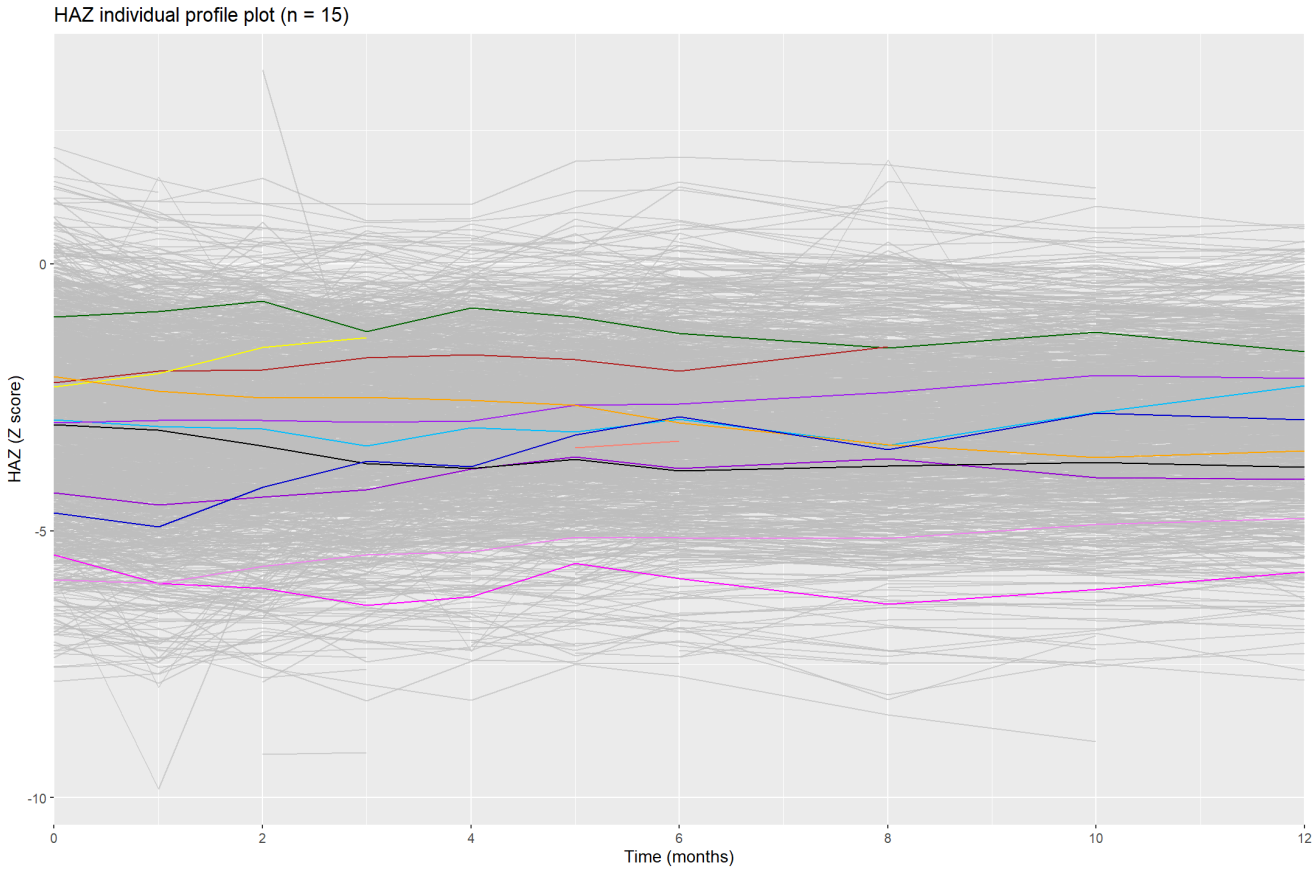


Figure 4.2: Spaghetti plot representing individual profiles of HAZ over time. Each grey line depicts the profile of a child, with colored highlights for 15 random subjects

We observe heterogeneous pattern for baseline values but less within and between variability over time for HAZ. This indicates that stunting status does not improve over the 12 months period (slow/ no vertical growth over time).

4.3.3 Weight-for-age Z-score (WAZ)

Low values of HAZ are classified as "underweight" and is indicative of malnutrition in children. Figure 4.3 each individual profile plots of WAZ over time.

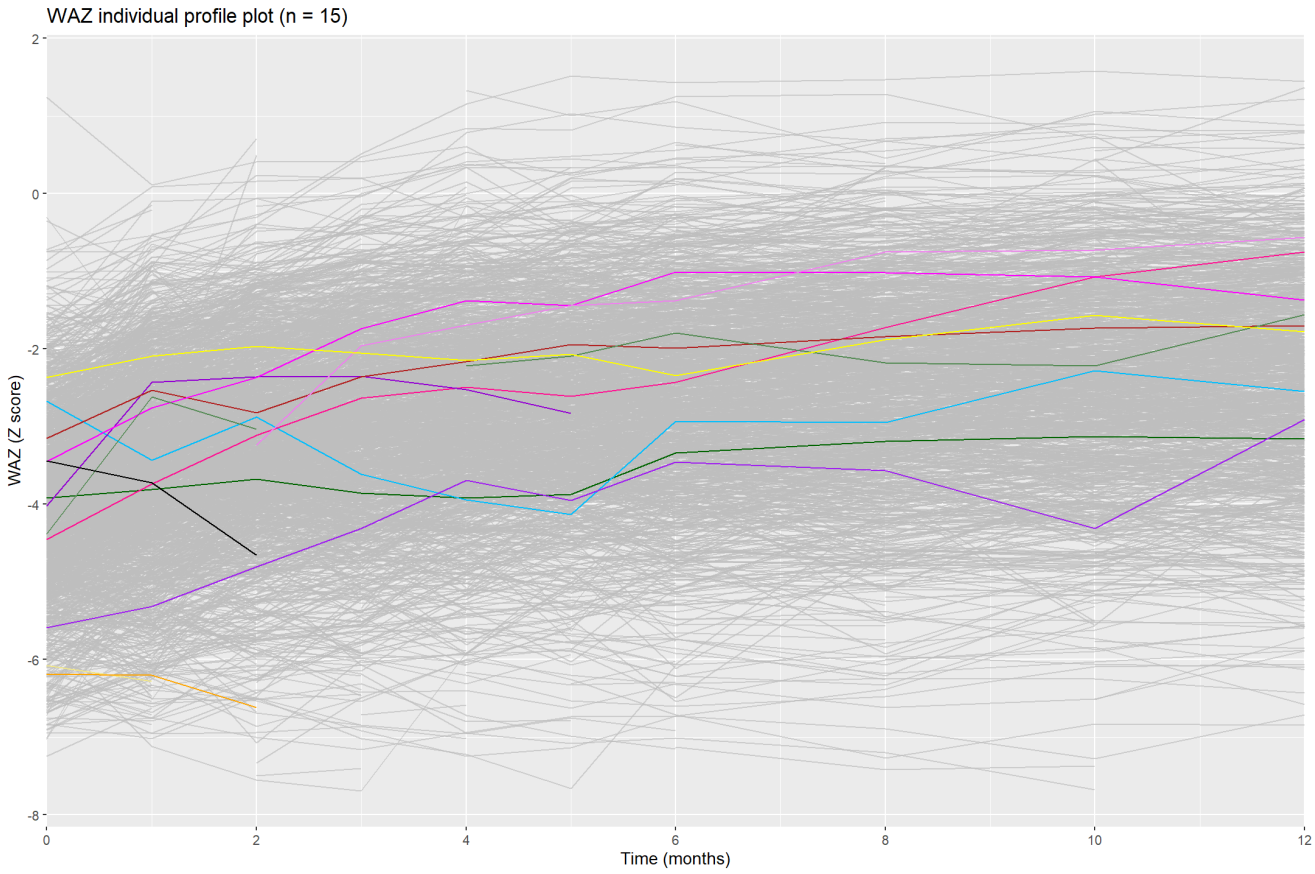


Figure 4.3: Spaghetti plot representing individual profiles of WAZ over time. Each grey line depicts the profile of a child, with colored highlights for 15 random subjects

We observe a linearly increasing trend with fair heterogeneity at baseline and substantial within and between subject variability over the follow up period. From this plot, we can see that children in this study gained weight over the follow up period.

4.3.4 Weight-for-height Z-score (WHZ)

The final z-score that we investigated is WHZ whose low values are indicative of wasting in children. Figure 4.4 depicts the subject-specific evolution of WHZ over the follow up period.

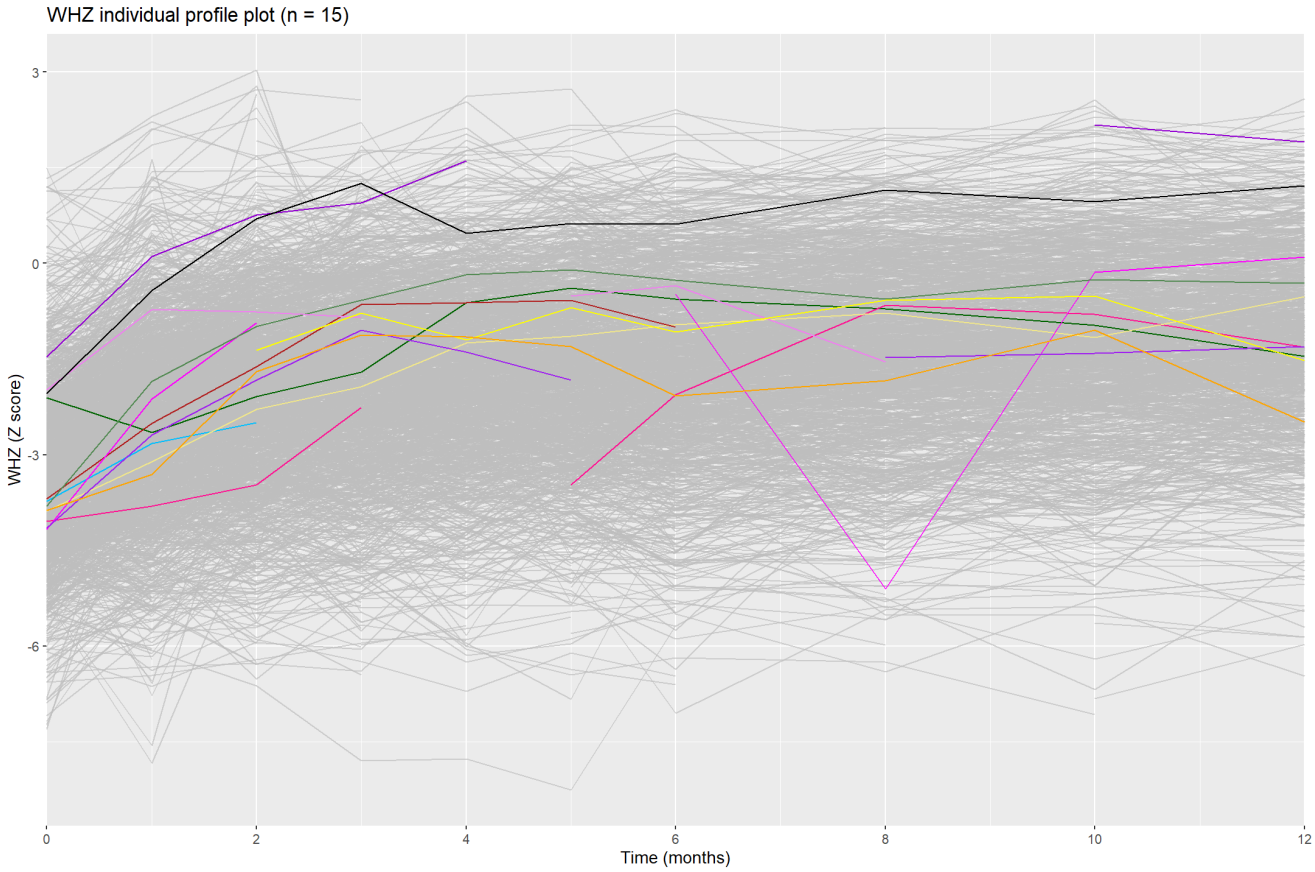


Figure 4.4: Spaghetti plot representing individual profiles of WHZ over time. Each grey line depicts the profile of a child, with colored highlights for 15 random subjects

Just like the previous anthropometry measurements, most children show an increasing trend in WHZ presence of heterogeneity at baseline with a fair amount within subject variability. Similar to WAZ profile plot (figure 4.3), WHZ profiles also shows a gain in weight in children in this study over the follow up period.

4.4 Exploring the mean structure

The mean structure for the various anthropometry measures was explored using average profile plots stratified by age category with an inclusion of 95% C.I error bars for each time point as shown in Figure 4.5.

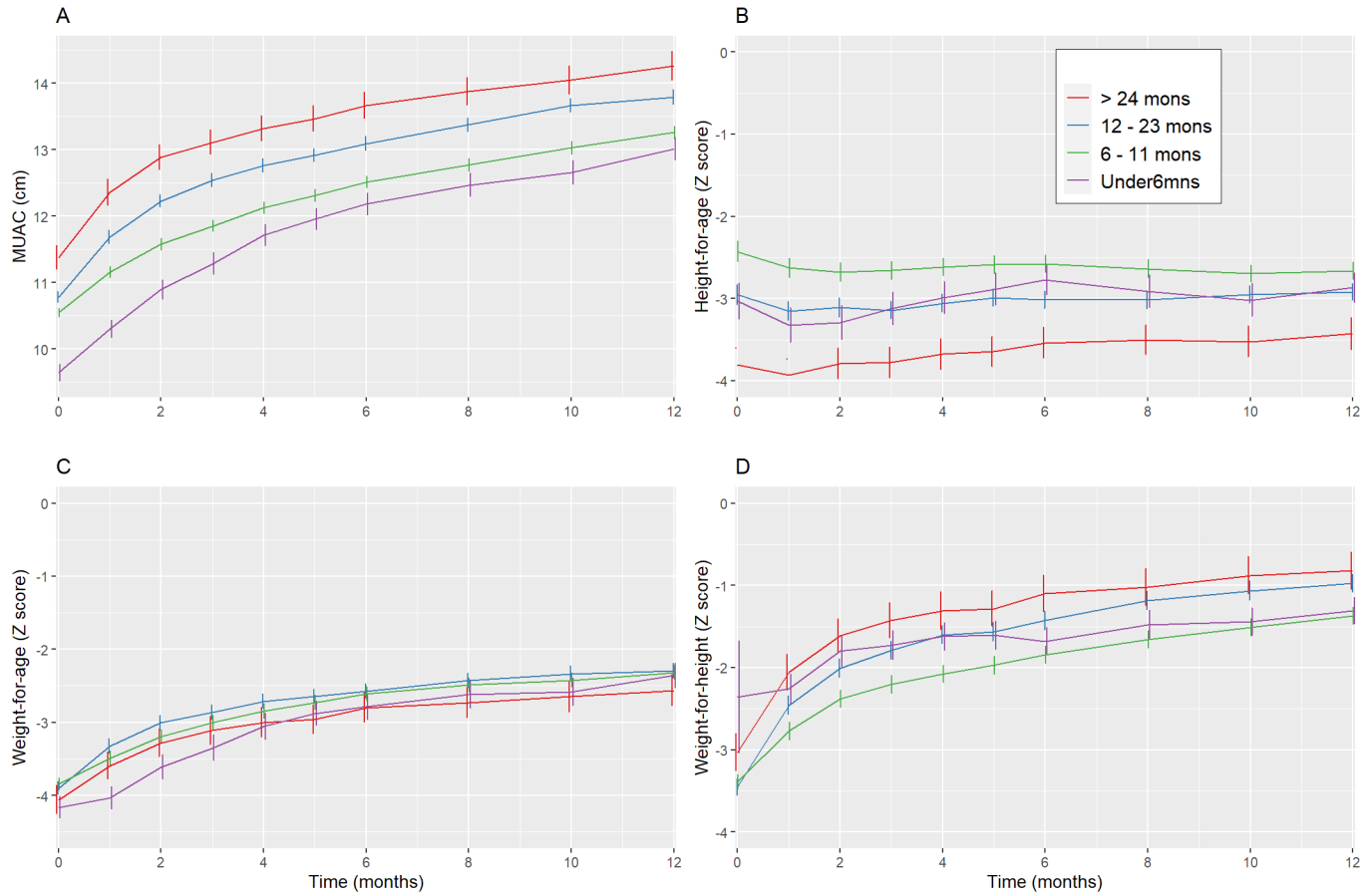


Figure 4.5: Profiles plots depicting the cohort average split by age group (as per coloured legend) for: A) MUAC, B)HAZ, C)WAZ, D)WHZ. Vertical bars indicate 95% confidence intervals at each time point.

All the four longitudinal measurements depicted a linear trend over time. Young infants (under six months) showed a low gain in MUAC over time compared to children older than twenty four months. A significant improvement/gain was observed in all age groups for MUAC and weight (both WHZ and WAZ) with a sharp increase in the first two months and then slowing down thereafter.

4.5 Exploring the correlation structure

Anthropometric measures repeated within the same subject are correlated and requires analysis methods that can account for the intra-subject correlation of response measurements;

which if ignored, can grossely invalidate inferences. Figure 4.6 shows the correlation structure/matrices of the four response measurement.

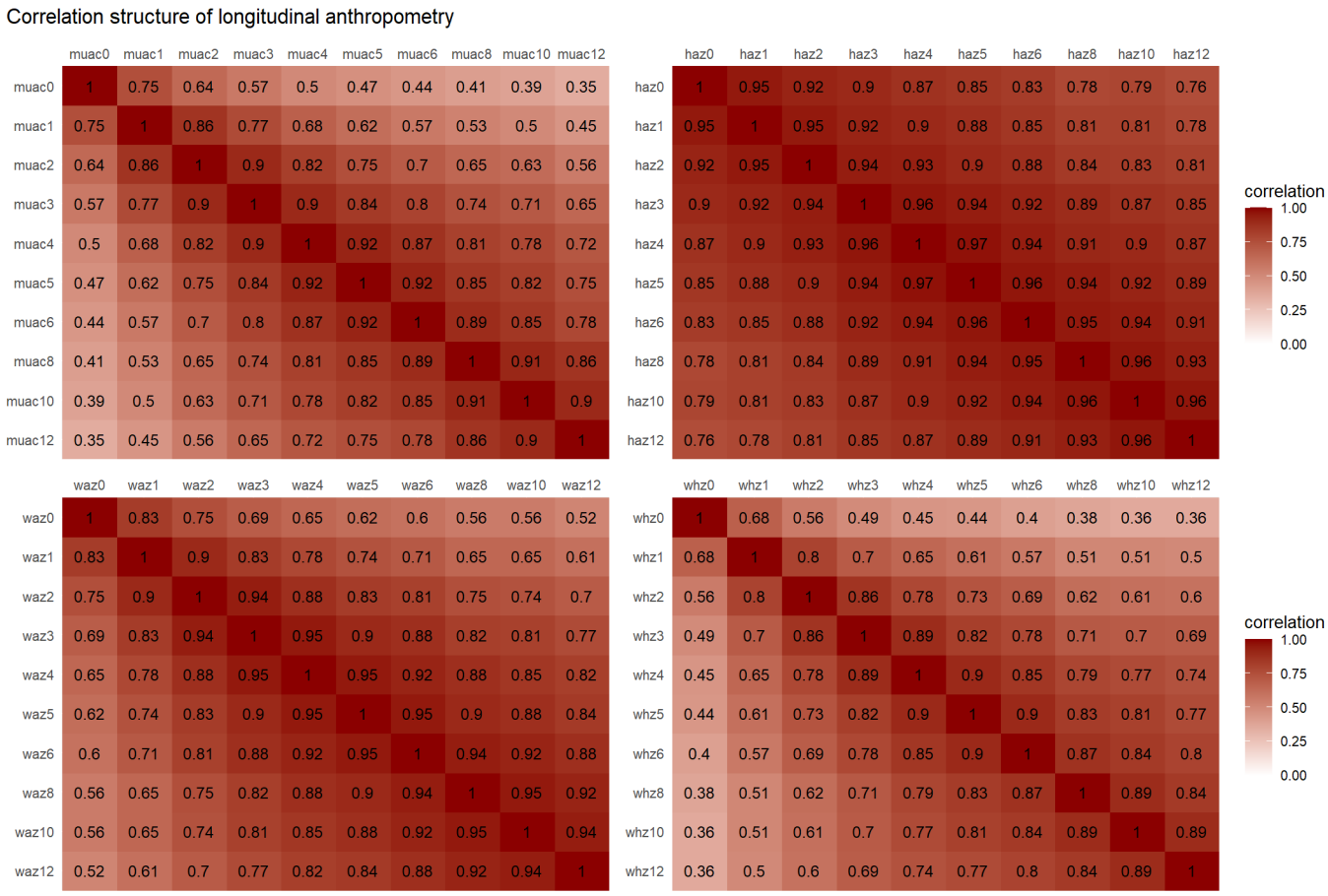


Figure 4.6: Correlation structure

From the matrices above, HAZ has the strongest intra individual correlation, and WHZ & MUAC shows the weakest intra-subject correlation over time.

4.6 Longitudinal data analysis

Linear mixed-effect models are an extension of the traditional least squares linear models with additional random-effect terms that can account for clustering of dependent data (also known as hierarchial or multilevel data).

We fitted four LME models for each of the anthropometry measurements while adjusting for child's age in months at admission and the treatment arm. Both random intercept and slope were included in the models. The tables in section 4.6.1 to section 4.6.4 displays the results (both of random and fixed effects estimates) from fitting the LMMs.

4.6.1 MUAC LMM

Table 4.2 and Table 4.3 for random and fixed effects respectively for longitudinal MUAC.

Table 4.2: MUAC: Random effects

Random Effects	**StdDev**	**Corr**
Intercept	0.969	Intr
time	0.112	-0.155
Residual	0.650	

Table 4.3: MUAC: Fixed effects

Fixed Effects	**Value**	**Std.Error**	**DF**	**t-value**	**p-value**
Intercept	10.409	0.049	12169.000	211.490	0.000
time	0.223	0.003	12169.000	66.2221	0.000
Arm - placebo	-0.061	0.048	1775.000	-1.269	0.204
Age (months)	0.058	0.002	1775.000	21.618	0.000

4.6.2 HAZ LMM

Table 4.4 and Table 4.5 for random and fixed effects estimates respectively for longitudinal Height-for-age Z-score.

Table 4.4: HAZ: Random effects

Random Effects	**StdDev**	**Corr**
Intercept	1.539	Intr
time	0.087	-0.415
Residual	0.376	

Table 4.5: HAZ:Fixed effects

Fixed Effects	Value	Std.Error	DF	t-value	p-value
Intercept	-2.554	0.071	11756.000	-36.2448	0.0000
time	0.006	0.003	11756.000	2.354	0.019
Arm - placebo	-0.069	0.068	1775.000	-1.022	0.307
Age (months)	-0.0304	0.0038	1775.0000	-7.9775	0.0000

4.6.3 WAZ LMM

Table 4.6 and Table 4.7 for random and fixed effects estimates respectively for longitudinal Weight-for-age Z-score.

Table 4.6: WAZ:Random effects

Random effects	StdDev	Corr
Intercept	1.149	Intr
time	0.095	-0.113
Residual	0.495	

Table 4.7: WAZ: Fixed effects

Fixed effects	Value	Std.Error	DF	t-value	p-value
Intercept	-3.633	0.057	11737.000	-64.1605	0.0000
time	0.1177	0.003	11737.000	41.575	0.000
Arm - placebo	-0.074	0.056	1775.000	-1.336	0.182
Age (months)	0.005	0.003	1775.000	1.547	0.122

4.6.4 WHZ LMM

Table 4.8 and Table 4.9 for random and fixed effects estimates respectively for longitudinal Weight-for-height Z-score.

Table 4.8: WHZ:Random effects

Random effects	StdDev	Corr
Intercept	1.208	Intr
time	0.108	-0.230
Residual	0.751	

Table 4.9: WHZ: Fixed effects

Fixed effects	Value	Std.Error	DF	t-value	p-value
Intercept	-2.868	0.060	11699.000	-47.761	0.000
time	0.144	0.003	11699.000	41.636	0.000
Arm - placebo	-0.066	0.057	1775.000	-1.123	0.261
Age (months)	0.017	0.003	1775.000	5.125	0.000

4.7 Predictors of mortality

4.7.1 The Kaplan Meier survival function

The primary endpoint was mortality which was assessed over a period of twelve months. Secondary endpoints included hospital readmission, illness episodes, treated as outpatient and achieving nutritional recovery. Non-informative right censoring mechanism was considered with censoring constituting of drop-outs due to lost-to-follow up, voluntary withdrawals and administrative censoring. Figure 4.7 shows the Kaplan Meier survival function for all cause mortality.

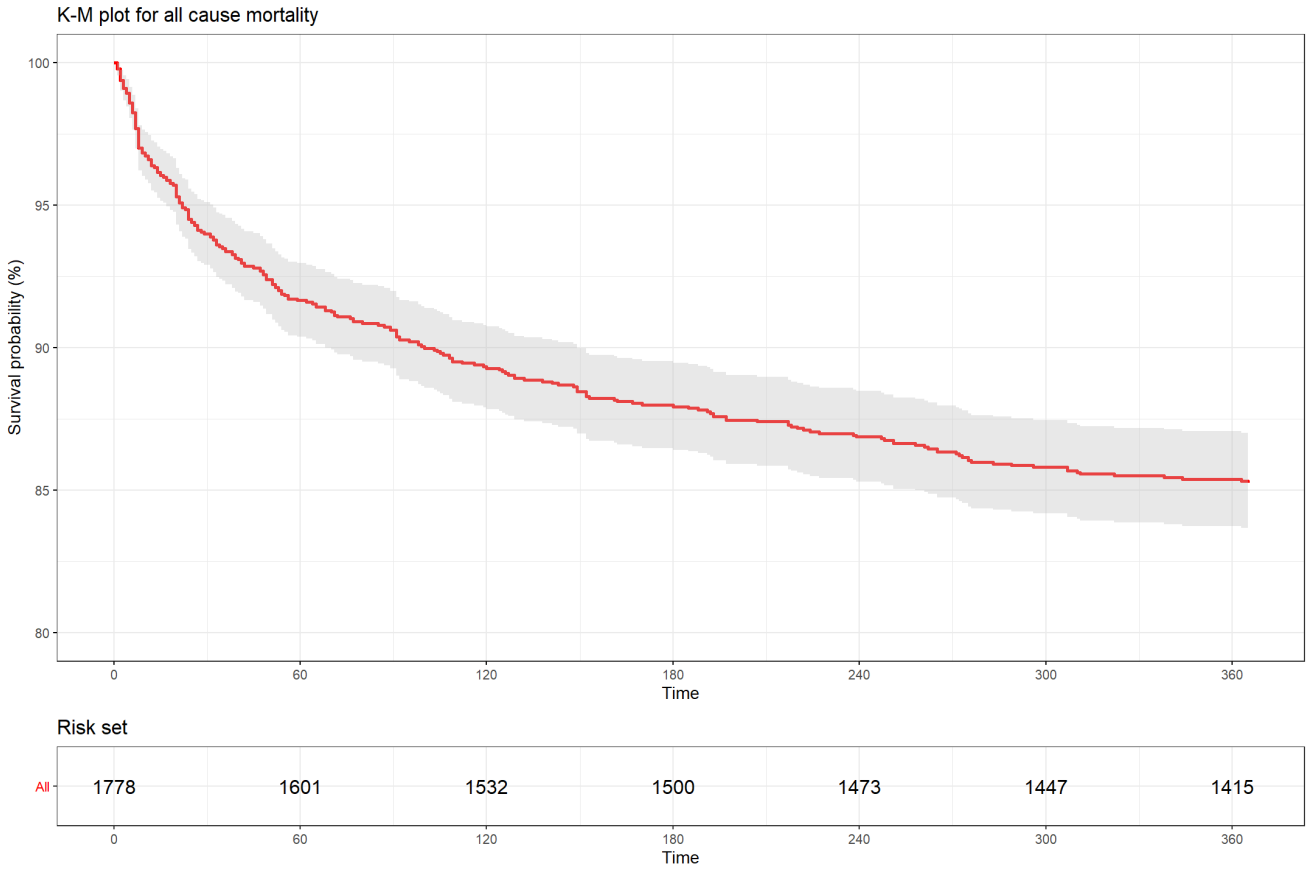


Figure 4.7: The Kaplan Meier curve for time to death

4.7.2 Cox proportional hazards model

In order to understand the effect of pre-selected baseline covariates on the risk of death, Cox proportional hazards was fitted and results presented in Table 4.10. The fundamental assumption of proportional hazards (i.e. testing whether the hazard of each covariate differs with time) was evaluated using Schonenfeld residuals by evaluating an interaction effect between a covariate and log(time) Table 4.11.

A significant p value ($p < 0.05$) indicates that the proportional hazards assumption has been violated. This is important because, for any single covariate that violates this assumption, model results could be invalid[62]. None of the variables used in the Cox model violated the proportional hazard assumption (see Table 4.11)

Table 4.10: Cox proportional hazard model results

Covariate	Estimate	aHR	aHR (95% CI)	Pr(> |z|)
Age (under 6 months) - Yes	0.686	1.985	1.985 (1.504 - 2.621)	0.000
Arm - placebo	0.075	1.078	1.078 (0.843 - 1.378)	0.549
Gender-Male	-0.106	0.900	0.899 (0.703 - 1.151)	0.401
Urban site - Yes	0.059	1.060	1.060 (0.781 - 1.439)	0.706
Unconscious - Yes	0.370	1.447	1.447 (0.937 - 2.234)	0.095
Medical condition -Yes	0.421	1.523	1.523 (1.158- 2.004)	0.003
Pneumonia - Yes	0.257	1.293	1.293 (0.991 - 1.685)	0.058
Diarrhoea - Yes	-0.183	0.833	0.833 (0.644 - 1.077)	0.163

Younger age and presence of any underlying medical condition were significant predictors of mortality. Young infants (< six months) were at elevated risk of dying with an adjusted hazard ratio of 1.985 (95% CI: 1.504 - 2.621) compared to older children (> six months). Children presenting with an underlying medical condition at admission had 52.3% increased risk of dying compared to children without an underlying medical condition.

Pneumonia diagnosis was a risk factor for mortality while children with diarrhoea were at a lower risk of dying with adjusted hazard ratios of 1.293 (95% CI:0.991 - 1.685) and 0.833 (95% CI:0.644 - 1.077) respectively. In general, other than age, underlying medical conditions and gender, all other factors were not associated with risk of mortality (p >= 0.05).

Table 4.11: Testing for the proportional hazards assumption

Covariate	chisq	df	p-value
Age (under 6 months) - Yes	2.637	1	0.104
Arm - placebo	0.549	1	0.459
Gender-Male	0.167	1	0.683
Urban site - Yes	0.28	1	0.596
iconsc	0.401	1	0.527
Medical condition -Yes	2.223	1	0.136
Pneumonia - Yes	3.053	1	0.081
Diarrhoea - Yes	3.827	1	0.05
GLOBAL	9.806	8	0.279

4.7.3 Time-varying Cox regression models

Data for performing time-varying Cox model needs reshaping so that multiple columns for each subject along with covariate values apply across intervals. This format is referred to as counting process style or (start, stop) [63].

We fitted four time-varying Cox regression models to evaluate the effect of monthly anthropometry on risk of death while adjusting for a priori selected baseline covariates (see tables 4.12, 4.13, 4.14 and 4.15)

MUAC

Table 4.12: Time-varying cox model for monthly MUAC

Covariate	Estimate	aHR	aHR (95% C.I)	p-value
Age (under 6 months) - Yes	0.005	1.005	1.005(0.728-1.387)	0.976
Arm - placebo	0.024	1.024	1.024(0.796-1.319)	0.851
Gender-Male	-0.099	0.906	0.906(0.707-1.160)	0.432
Urban site - Yes	0.059	1.061	1.061(0.778-1.447)	0.709
Unconscious - Yes	0.454	1.575	1.5751.030-2.409)	0.036
Medical condition -Yes	0.279	1.322	1.321(0.999-1.748)	0.051
Pneumonia - Yes	0.238	1.268	1.268(0.973-1.653)	0.079
Diarrhoea - Yes	-0.138	0.871	0.871(0.667-1.138)	0.311
MUAC	-0.618	0.539	0.539(0.484-0.599)	0.000

A one centimetre gain in monthly MUAC is associated with 46.1% reduction in hazard of death.

HAZ

Table 4.13: Time-varying Cox model for monthly HAZ

Covariate	Estimate	aHR	aHR(95% CI)	p-value
Age (under 6 months) - Yes	0.658	1.930	1.930(1.465-2.543)	0.000
Arm - placebo	0.072	1.075	1.075(0.838-1.377)	0.570
Gender-Male	-0.189	0.828	0.828(0.644-1.065)	0.142
Urban site - Yes	0.159	1.172	1.1722(0.858-1.601)	0.318
Unconscious - Yes	0.433	1.542	1.542(0.996-2.386)	0.052
Medical condition -Yes	0.376	1.456	1.456(1.107-1.915)	0.007
Pneumonia - Yes	0.288	1.333	1.333(1.025-1.733)	0.031
Diarrhoea - Yes	-0.149	0.862	0.862(0.663-1.120)	0.266
HAZ	-0.141	0.869	0.869(0.793-0.951)	0.002

A unit SD gain in monthly HAZ reduces the hazard of deaths by 13.1%, Table4.13

WAZ

Table 4.14: Time-varying Cox model for monthly WAZ

Covariate	Estimate	aHR	aHR(95% CI)	p-value
Age (under 6 months) - Yes	0.469	1.599	1.599(1.196-2.137)	0.002
Arm - placebo	0.057	1.058	1.058(0.822-1.362)	0.660
Gender-Male	-0.346	0.707	0.7070.548-0.912)	0.008
Urban site - Yes	0.112	1.119	1.119(0.817-1.531)	0.484
Unconscious - Yes	0.484	1.622	1.622(1.049-2.507)	0.030
Medical condition -Yes	0.203	1.225	1.225(0.926-1.621)	0.155
Pneumonia - Yes	0.313	1.368	1.368(1.054-1.775)	0.018
Diarrhoea - Yes	-0.083	0.921	0.921(0.704-1.204)	0.546
WAZ	-0.519	0.595	0.595(0.534-0.663)	0.000

A unit SD gain in monthly WAZ is associated with 40.5% reduction in risk of death.

WHZ

Table 4.15: Time-varying Cox model for monthly WHZ

Covariate	Estimate	aHR	aHR(95% CI)	p-value
Age (under 6 months) - Yes	0.839	2.314	2.314(1.759-3.0423)	0.000
Arm - placebo	0.061	1.063	1.063(0.828-1.365)	0.631
Gender-Male	-0.292	0.747	0.747(0.582-0.958)	0.022
Urban site - Yes	-0.208	0.812	0.812(0.593-1.114)	0.197
Unconscious - Yes	0.320	1.377	1.377(0.895-2.118)	0.145
Medical condition -Yes	0.222	1.248	1.248(0.947-1.645)	0.115
Pneumonia - Yes	0.243	1.275	1.275(0.979-1.660)	0.071
Diarrhoea - Yes	-0.142	0.868	0.868(0.665-1.1334)	0.298
WHZ	-0.465	0.628	0.628(0.572-0.689)	0.000

For the monthly WHZ, a unit gain in SD is associated with 37.2% reduction in risk of mortality.

4.8 Joint models for longitudinal and time-to-event data

The time-varying Cox model should only be applied to exogenous time-dependent covariates and thus cannot handle longitudinal bio-markers which are by definition endogenous[52]. Therefore, when the primary interest lies in the association between endogenous time-dependent variables and time-to-event outcome, joint modelling framework is applied.

In the analysis and in fitting the four joint models, we hypothesized that a child's hazard of dying at any given time is associated with both the current value of the longitudinal measurement and it's associated rate of change or slope, also known as current value and slope association.

4.8.1 MUAC joint model

In determining the association between MUAC and hazard of death of a child, the current value of MUAC, was statistically significant ($p < 0.0001$) but it's rate/slope, MUAC.slope was not ($p = 0.8763$). In particular, for children having the same rate of change of MUAC, a unit gain in one centimetre in the current value of MUAC is associated with 46.8% reduction in hazard of death 0.532(95% CI: 0.476-0.596). Table 4.16.

Table 4.16: Current value and slope MUAC joint model

Effect	**aHR (95% CI)**	**Std.Err**	**z-value**	**p-value**
Age (under 6 months) - Yes	1.109(0.761 -1.615)	0.192	0.539	0.590
Arm - placebo	1.096(0.814-1.477)	0.152	0.605	0.545
Gender-Male	0.964(0.720-1.290)	0.149	-0.246	0.806
Urban site - Yes	1.489(1.028-2.156)	0.189	2.105	0.035
Unconscious - Yes	1.468(0.810-2.661)	0.303	1.266	0.206
Medical condition -Yes	1.2080(0.867-1.684)	0.169	1.116	0.265
Pneumonia - Yes	0.817(0.585-1.141)	0.170	-1.187	0.235
Diarrhoea - Yes	0.868(0.641-1.175)	0.154	-0.917	0.359
MUAC	0.532(0.476-0.596)	0.058	-10.961	<0.0001
MUAC.slope	1.436(0.015-136.196)	2.323	0.156	0.876

4.8.2 HAZ joint model

For Children having the same current value of HAZ, the slope or rate of change was not statistically significant ($p = 0.3879$). A unit increase in the trajectory is associated with a

hazard ratio of 0.227 (95% C.I: 0.008 - 6.556). A unit SD increase in the value of HAZ is associated with 2.5% reduction in risk of death in children with same HAZ trajectory. Table 4.17

Table 4.17: Current value and slope HAZ joint model

Effect	**aHR (95% CI)**	**Std.Err**	**z-value**	**p-value**
Age (under 6 months) - Yes	2.038(1.541-2.694)	0.142	5.002	<0.0001
Arm - placebo	1.058(0.826-1.354)	0.126	0.447	0.655
Gender-Male	0.919(0.716-1.179)	0.127	-0.658	0.511
Urban site - Yes	1.036(0.759-1.411)	0.158	0.223	0.824
Unconscious - Yes	1.426(0.922-2.205)	0.222	1.596	0.111
Medical condition -Yes	1.547(1.174-2.039)	0.141	3.104	0.002
Pneumonia - Yes	1.283(0.985-1.671)	0.135	1.851	0.064
Diarrhoea - Yes	0.815(0.629-1.055)	0.132	-1.551	0.121
HAZ	0.975(0.955-0.995)	0.010	-2.388	0.017
HAZ.slope	0.227(0.008-6.557)	1.715	-0.863	0.388

4.8.3 WAZ joint model

The trajectory of WAZ was not significantly associated with the risk of death (p = 0.9557). However, children with same rate of change in WAZ, a unit gain in one SD in the current value of WAZ is associated with 21.2% lower risk of death 0.788(95% CI: 0.742-0.837). Table 4.18

Table 4.18: Current value and slope WAZ joint model

Effect	**aHR(95% CI)**	**Std.Err**	**z-value**	**p-value**
Age (under 6 months) - Yes	2.012(1.474-2.747)	0.159	4.406	<0.0001
Arm - placebo	0.906(0.690-1.190)	0.139	-0.706	0.481
Gender-Male	0.749(0.566-0.991)	0.143	-2.022	0.043
Urban site - Yes	1.131(0.805-1.589)	0.174	0.708	0.479
Unconscious - Yes	1.294(0.814-2.057)	0.237	1.088	0.277
Medical condition -Yes	1.408(1.039-1.907)	0.155	2.212	0.027
Pneumonia - Yes	1.279(0.945-1.731)	0.154	1.596	0.111
Diarrhoea - Yes	1.101(0.824-1.472)	0.148	0.651	0.515
WAZ	0.788(0.742-0.837)	0.031	-7.818	<0.0001
WAZ.slope	1.169(0.005-285.838)	2.806	0.056	0.956

4.8.4 WHZ joint model

Both current value and rate of change of WHZ was significantly associated with the risk of dying in children. For same level of current values of WHZ, the adjusted log hazard associated with a unit increase in the trajectory was 5.72 (95% CI: 0.29 - 11.14), however, this was not statistically significant as the confidence interval is too wide and contains one.

Children having the same rate of change in WHZ, a unit increase in one SD in the current value of WHZ was associated with 37.1% reduction in the hazard of death 0.629(95% CI: 0.579-0.683). Table 4.19

Table 4.19: Current value and slope WHZ joint model

Effect	**aHR(95% CI)**	**Std.Err**	**z-value**	**p-value**
Age (under 6 months) - Yes	2.036(1.465-2.831)	0.168	4.232	<0.0001
Arm - placebo	1.054(0.786-1.414)	0.150	0.353	0.724
Gender-Male	0.745(0.563-0.987)	0.143	-2.053	0.040
Urban site - Yes	0.891(0.633-1.253)	0.174	-0.662	0.508
Unconscious - Yes	1.622(0.988-2.662)	0.253	1.913	0.056
Medical condition -Yes	1.199(0.875-1.641)	0.160	1.130	0.258
Pneumonia - Yes	1.059(0.781-1.435)	0.155	0.370	0.711
Diarrhoea - Yes	1.076(0.801-1.444)	0.150	0.490	0.625
WHZ	0.629(0.579-0.683)	0.042	-10.955	<0.0001
WHZ.slope	303.584(1.335-69006.22)	2.769	2.065	0.039

4.9 Time-varying Cox models vs joint models

We compared models for monthly anthropometry from the two modelling framework i.e. joint modeling and use of time-varying cox model. The results for hazard ratios and 95% confidence intervals are displayed in Table 4.20.

Table 4.20: Comparing results from joint models and time-varying Cox models

Joint Model	**aHR (95% CI)**	**p-value**	**Cox model**	**aHR (95% CI)**	**p-value**
MUAC	0.532(0.476-0.596)	0.000	MUAC	0.539(0.484-0.599)	0.000
HAZ	0.9750.955-0.995)	0.017	HAZ	0.869(0.793-0.951)	0.002
WAZ	0.788(0.742-0.837)	0.000	WAZ	0.595(0.534-0.663)	0.000
WHZ	0.629(0.579-0.683)	0.000	WHZ	0.628(0.572-0.689)	0.000

Using time-varying cox model on endogenous time-dependent covariates leads to in-efficient and biased estimates[51,53,63]. These modeling framework ignores the data generating process of the bio-marker i.e. an internal covariate that is as result of a stochastic process genereted by an individual on follow up and only observed as long as the patient survives[63]

Chapter 5

Discussion and Conclusion

5.1 Discussion

In this study, we applied a joint model for longitudinal and time-to-event data to quantify the association strength between the four common anthropometry measurements and the risk of dying in children with SAM. Results from the longitudinal models (LMMs) show no statistical difference between children who were randomised to co-trimoxazole and those in placebo arm with regards to improvement of the anthropometry measurements over time. These results are in concordance to those found by [4] that daily co-trimoxazole given for 6 months did not reduce mortality nor improve growth. The age of a child at admission was a significant determinant of the anthropometry trajectory. Notably, older children had a slow trajectory of HAZ over time, thus delayed recovery from stunting.

Cox proportional hazards regression analysis results show no statistical difference in survival between co-trimoxazole group and placebo. A child younger than the age of six months was at a remarkably high risk of dying (1.98-fold increase in hazard of dying) than older children. Presence of underlying medical conditions such as tuberculosis, sickle cell disease, rickets, cerebral palsy or congenital anomaly at admission puts children at elevated risk of death.

Children presenting with diarrhoea had a high survival rate while comorbidities such as pneumonia and experiencing bouts of unconsciousness during admission tend to show an increased risk of mortality, however these results were not statistically significant at 5% α-level. Patients recruited from urban sites (Mombasa and Nairobi) were shown to have increased risk of experiencing the event, again this difference is not statistically significant at 5% α-level. Compared to female, male children were shown to have better survival rates.

Results from time-varying cox model showed that the monthly anthropometry is associated with risk of death. For a one-unit gain in monthly MUAC, HAZ, WAZ and WHZ; was associated with reduction of hazard of death by 46.1%, 13.1%, 40.5% and 37.2% respectively

(see Tables 4.12, 4.13, 4.14 and 4.15). However, it should be noted that these estimates are grossly biased because monthly anthropometry is an internal covariate and thus time-varying Cox models are inappropriate approach to modelling their association with risk of death.

The joint model simultaneously evaluates the longitudinal evolution of anthropometry and time-to-event outcome (death). The key feature of joint modelling is the ability to incorporate a time-dependent covariate measured with error into a survival sub-model by shared parameters with the risk factor variable, in this case mortality[51]. In this study, we used the current value and slope association parameter, i.e., at any given time, the hazard of experiencing an event is influenced by both the current value of a longitudinal biomarker and the rate of change/trajectory.

Results from our joint model showed a significant association between the four monthly anthropometry and the risk of death. MUAC had the highest impact on risk of death (reducing risk of death by 46.8% per unit gain). This means that for children admitted with SAM, a unit gain in MUAC lowers their risk of dying significantly implying than MUAC is a good predictor of mortality Table 4.16. Our findings are similar to previous research studies[8,39] that shows a benefit of using MUAC and/or WHZ to identify and screen high-risk children. WHZ was the second highest significant predictor of mortality, followed by WAZ and finally HAZ which had the least effect on risk of dying in this cohort. Tables 4.19, 4.18 and 4.17 respectively.

The joint model was also used to evaluate the effect of the rate of change/trajectory of the repeated monthly anthropometry on mortality. There was no statistically significant effect of slope for MUAC, HAZ and WAZ on risk of death. However, results showed that the slope for WHZ for same current values of WHZ is a significant predictor of mortality Table 4.19.

5.2 Conclusion

With joint modelling framework, we were able to measure the strength of association between four longitudinal anthropometry and risk of death in children aged 2 to 59 months, by utilizing both the values of the biomarkers over time and rate of change/trajectory (current value and slope shared parameters). In studying the relationship between a survival outcome and a biomarker, researchers have used the naive Cox regression model with only baseline

values, thus failing to use the longitudinal history of bi-omarkers and time-varying Cox model which ignores the endogeneity of the biomarker, producing biased model estimates.

Through applying joint modelling, this study shows that both MUAC and WHZ are strongly associated with mortality in children with SAM. However, MUAC has a better ability to assess and identify children at risk of dying while HAZ is a less effective predictor of mortality. The rate of change overtime also contributes to evaluating mortality risk (tables 4.16, 4.17, 4.18 and 4.19)

5.3 Recommendations and further research

In follow up observational studies and clinical trials where various disease biomarkers are measured repeatedly over time with the aim of understanding patients' recovery path, health status or growth profile for children and how this is associated with the risk of a life-threatening event occurring, the joint modelling framework is perfectly suited for these kinds of analyses. This approach is beneficial because you get the change of utilizing the entire longitudinal history of the biomarker to inform risk of experiencing an event of interest.

Traditional methods for analysing these two outcomes have fallen short in the sense that they do not recognize the dependencies that exist during the data generation process. In this analysis, the interest was to quantify the association between four monthly anthropometry (MUAC, HAZ, WAZ and WHZ) indices and risk of dying in children aged between 2 to 59 months diagnosed with SAM.

We used univariate joint modelling in this analysis. We therefore recommend an improvement on this separate analysis by proposing an extension to multivariate joint modelling to help understand how the four combined anthropometry predict mortality in children with SAM. For instance, a multivariate joint modelling for WHZ and HAZ will be beneficial to understand how combined wasting and stunting predict mortality. A full grown multivariate joint model for the four anthropometry measurements can also give insights into the underlying interaction in the trajectories and how this affects growth and risk of dying in children.

Chapter 6

Appendix

6.1 R session information

```
> sessionInfo()

R version 4.0.3 (2020-10-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 18363)

Matrix products: default

locale:
[1] LC_COLLATE=English_Kenya.1252  LC_CTYPE=English_Kenya.1252
[3] LC_MONETARY=English_Kenya.1252 LC_NUMERIC=C
[5] LC_TIME=English_Kenya.1252

attached base packages:
[1] splines   stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
 [1] rjags_4-10         coda_0.19-4        JM_1.4-8           nlme_3.1-152
 [5] MASS_7.3-53.1      lubridate_1.7.9.2  DT_0.17            haven_2.3.1
 [9] readxl_1.3.1       here_1.0.1         table1_1.4         scales_1.1.1
[13] gghighlight_0.3.1  brolgar_0.1.0      survminer_0.4.8    ggpubr_0.4.0
[17] survival_3.2-7     patchwork_1.1.1    lcsm_0.1.2         ggthemes_4.2.4
```

```
[21] qwraps2_0.5.0     forcats_0.5.1     stringr_1.4.0     dplyr_1.0.5
[25] purrr_0.3.4       readr_1.4.0       tidyr_1.1.2       tibble_3.0.6
[29] ggplot2_3.3.3     tidyverse_1.3.0

loaded via a namespace (and not attached):
 [1] fs_1.5.0             httr_1.4.2           rprojroot_2.0.2
 [4] tools_4.0.3          backports_1.2.1      R6_2.5.0
 [7] DBI_1.1.1            colorspace_2.0-0     withr_2.4.1
[10] tidyselect_1.1.0     gridExtra_2.3        mnormt_2.0.2
[13] curl_4.3             compiler_4.0.3       cli_2.4.0
[16] rvest_0.3.6          xml2_1.3.2           survMisc_0.5.5
[19] digest_0.6.27        pbivnorm_0.6.0       foreign_0.8-80
[22] tsibble_1.0.1        rmarkdown_2.7        rio_0.5.16
[25] pkgconfig_2.0.3      htmltools_0.5.1.1    dbplyr_2.1.0
[28] htmlwidgets_1.5.3    rlang_0.4.10         rstudioapi_0.13
[31] farver_2.0.3         generics_0.1.0       zoo_1.8-8
[34] jsonlite_1.7.2       distributional_0.2.2 zip_2.1.1
[37] car_3.0-10           magrittr_2.0.1       Formula_1.2-4
[40] Matrix_1.3-2         Rcpp_1.0.6           munsell_0.5.0
[43] abind_1.4-5          lifecycle_1.0.0      stringi_1.5.3
[46] yaml_2.2.1           snakecase_0.11.0     carData_3.0-4
[49] lavaan_0.6-8         grid_4.0.3           crayon_1.4.1
[52] lattice_0.20-41      hms_1.0.0            tmvnsim_1.0-2
[55] anytime_0.3.9        knitr_1.32           pillar_1.4.7
[58] ggsignif_0.6.0       stats4_4.0.3         reprex_1.0.0
[61] glue_1.4.2           evaluate_0.14        fabletools_0.3.1
[64] data.table_1.13.6    modelr_0.1.8         vctrs_0.3.6
[67] cellranger_1.1.0     gtable_0.3.0         km.ci_0.5-2
[70] assertthat_0.2.1     xfun_0.22            openxlsx_4.2.3
[73] janitor_2.1.0        xtable_1.8-4         broom_0.7.6
[76] rstatix_0.7.0        KMsurv_0.1-5         ellipsis_0.3.1
```

6.2 R codes for actual model fitting

6.2.1 Cox proportional model

```
# cox-proportional model
    Cox.Fit <- coxph(Surv(study_time, outcome) ~ u6months + arm + sex + urban_site + i
                 data = ctx_wide_data,
                 model = TRUE,
                 x = TRUE) # include design matrix
```

Time-varying Cox models

```
   # MUAC
    muac.CoxFit <- coxph(
   Surv(
      time = tstart,
      time2 = tstop,
      event = end_point
   ) ~ u6months + arm + sex + urban_site + iconsc + med_condition + sev_pneum + diarrh
   data = muac_tvCox
)
# HAZ
haz.CoxFit <- coxph(
   Surv(
      time = tstart,
      time2 = tstop,
      event = end_point
   ) ~ u6months + arm + sex + urban_site + iconsc + med_condition + sev_pneum + diarrh
   data = haz_tvCox
)
# WAZ
waz.CoxFit <- coxph(
```

```
   Surv(
      time = tstart,
      time2 = tstop,
      event = end_point
   ) ~ u6months + arm + sex + urban_site + iconsc + med_condition + sev_pneum + diarrh
   data = waz_tvCox
)
# WHZ
whz.CoxFit <- coxph(
   Surv(
      time = tstart,
      time2 = tstop,
      event = end_point
   ) ~ u6months + arm + sex + urban_site + iconsc + med_condition + sev_pneum + diarrh
   data = whz_tvCox
)
```

6.2.2 Linear mixed effects models

```
# MUAC
LmeFit.muac <- lme(muac ~ time + arm + agemonth,
                   random = ~ time | subjid,
                   data = ctx_long_data,
                   na.action = na.omit)
# HAZ
LmeFit.haz <- lme(haz ~ time + arm + agemonth,
                  random = ~ time | subjid,
                  data = ctx_long_data,
                  na.action = na.omit)
 # WAZ
LmeFit.waz <- lme(waz ~ time + arm + agemonth,
                  random = ~ time | subjid,
```

```
                    data = ctx_long_data,
                    na.action = na.omit)
# WHZ
LmeFit.whz <- lme(whz ~ time + arm + agemonth,
                    random = ~ time | subjid,
                    data = ctx_long_data,
                    na.action = na.omit)
```

6.2.3 Joint models

```
# MUAC
muac_jointFit <- jointModel(LmeFit.muac, # longitudinal model
                            Cox.Fit,  # relative risk model
                            timeVar = "time",
                            parameterization = "both",
                            derivForm = derivation_form,
                            iter.EM = 1000,
                            method = "spline-PH-GH")
# HAZ
haz_jointFit <- jointModel(LmeFit.haz, # longitudinal model
                           Cox.Fit,  # relative risk model
                           timeVar = "time",
                           parameterization = "both",
                           derivForm = derivation_form,
                           iter.EM = 1000,
                           method = "spline-PH-GH")
# WAZ
waz_jointFit <- jointModel(LmeFit.waz, # longitudinal model
                           Cox.Fit,  # relative risk model
                           timeVar = "time",
                           parameterization = "both",
                           derivForm = derivation_form,
```

```
                              iter.EM = 1000,
                              method = "spline-PH-GH")
# WHZ
whz_jointFit <- jointModel(LmeFit.whz, # longitudinal model
                              Cox.Fit,  # relative risk model
                              timeVar = "time",
                              parameterization = "both",
                              derivForm = derivation_form,
                              iter.EM = 1000,
                              method = "spline-PH-GH")
```

6.3 Events by site and treatment arm

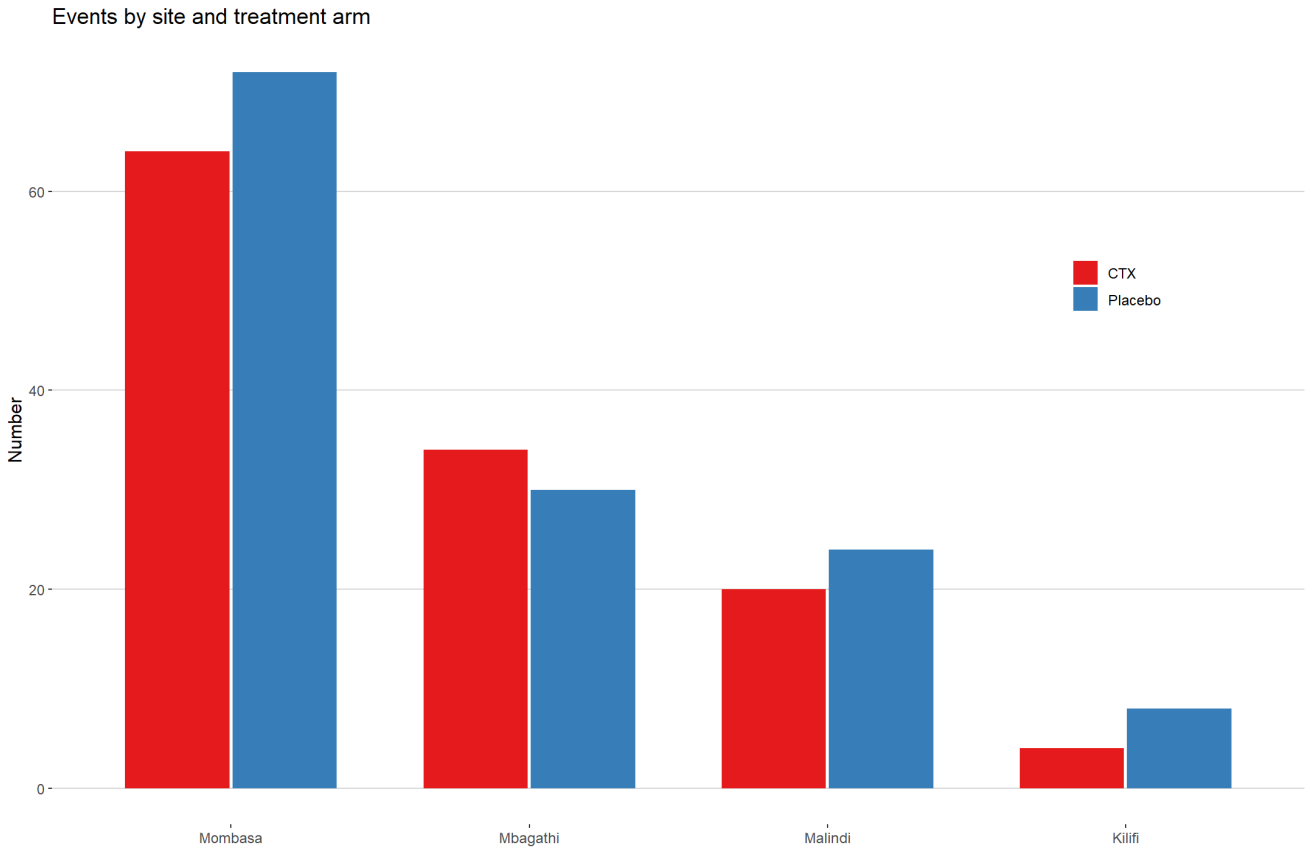


Figure 6.1: Events by site and treatment arm

6.4 Plots for Schoenfeld Residuals

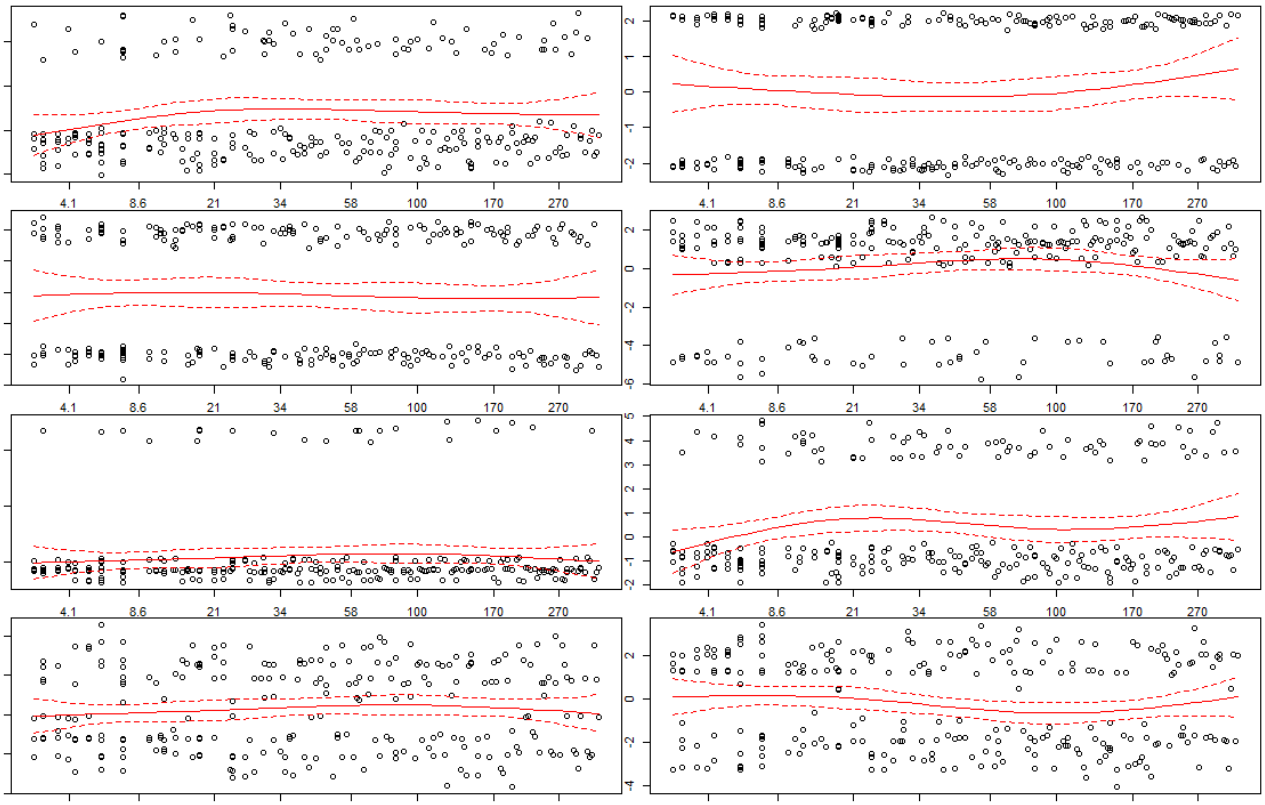


Figure 6.2: Schoenfeld Residuals for baseline covariates

6.5 Mean structure- violin plots

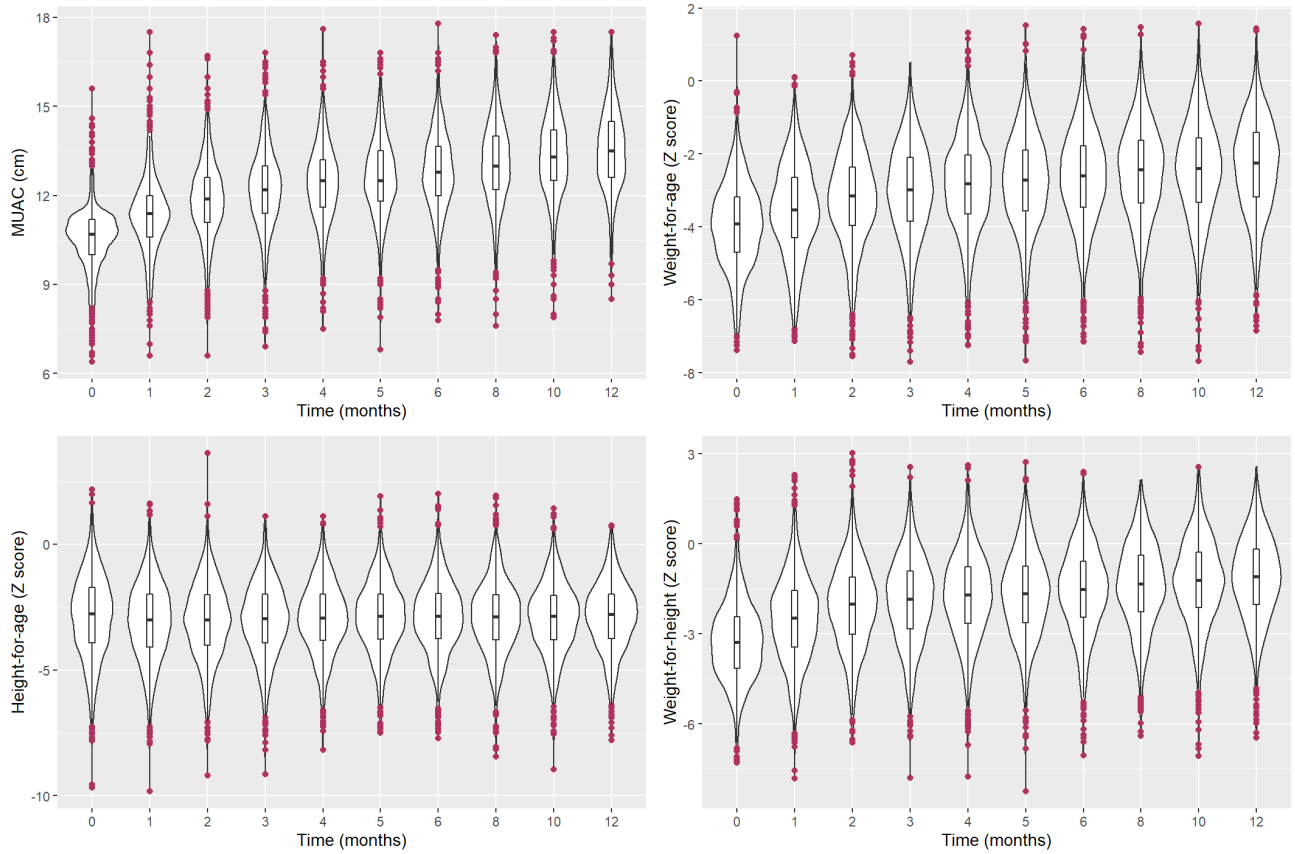


Figure 6.3: Violin plots

References

[1] Asar, Ö., Ritchie, J., Kalra, P. A., and Diggle, P. J. (2015). Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International journal of epidemiology*, 44(1):334–344.

[2] Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational research methods*, 7(2):127–150.

[3] Barnett, I., Torous, J., Staples, P., Keshavan, M., and Onnela, J.-P. (2018). Beyond smartphones and sensors: choosing appropriate statistical methods for the analysis of longitudinal data. *Journal of the American Medical Informatics Association*, 25(12):1669–1674.

[4] Berkley, J. A., Ngari, M., Thitiri, J., Mwalekwa, L., Timbwa, M., Hamid, F., Ali, R., Shangala, J., Mturi, N., Jones, K. D., et al. (2016). Daily co-trimoxazole prophylaxis to prevent mortality in children with complicated severe acute malnutrition: a multicentre, double-blind, randomised placebo-controlled trial. *The Lancet Global Health*, 4(7):e464–e473.

[5] Bhan, M. K., Bhandari, N., and Bahl, R. (2003). Management of the severely malnourished child: perspective from developing countries. *Bmj*, 326(7381):146–151.

[6] Bhattacharya, A., Pal, B., Mukherjee, S., and Roy, S. K. (2019). Assessment of nutritional status using anthropometric variables by multivariate analysis. *BMC public health*, 19(1):1045.

[7] Black, R. E., Morris, S. S., and Bryce, J. (2003). Where and why are 10 million children dying every year? *The lancet*, 361(9376):2226–2234.

[8] Briend, A., Maire, B., Fontaine, O., and Garenne, M. (2012). Mid-upper arm circumference and weight-for-height to identify high-risk malnourished under-five children. *Maternal & child nutrition*, 8(1):130–133.

[9] Buta, G. B., Goshu, A. T., and Worku, H. M. (2015). Bayesian joint modelling of disease progression marker and time to death event of hiv/aids patients under art follow-up. *Journal of Advances in Medicine and Medical Research*, pages 1034–1043.

[10] Bwakura-Dangarembizi, M., Dumbura, C., Amadi, B., Ngosa, D., Majo, F. D., Nathoo, K. J., Mwakamui, S., Mutasa, K., Chasekwa, B., Ntozini, R., et al. (2021). Risk factors for postdischarge mortality following hospitalization for severe acute malnutrition in zimbabwe and zambia. *The American Journal of Clinical Nutrition*.

[11] Cekic, S., Aichele, S., Brandmaier, A. M., Köhncke, Y., and Ghisletta, P. (2019). A tutorial for joint modeling of longitudinal and time-to-event data in r. *arXiv preprint arXiv:1909.05661*.

[12] Chisti, M. J., Graham, S. M., Duke, T., Ahmed, T., Faruque, A. S. G., Ashraf, H., Bardhan, P. K., Shahid, A. S., Shahunja, K., and Salam, M. A. (2014). Post-discharge mortality in children with severe malnutrition and pneumonia in bangladesh. *PloS one*, 9(9):e107663.

[13] Clark, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003a). Survival analysis part i: basic concepts and first analyses. *British journal of cancer*, 89(2):232–238.

[14] Clark, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003b). Survival analysis part iv: further concepts and methods in survival analysis. *British journal of cancer*, 89(5):781–786.

[15] Cnaan, A., Laird, N. M., and Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in medicine*, 16(20):2349–2380.

[16] Collett, D. (2015). *Modelling survival data in medical research*. CRC press.

[17] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

[18] De Onis, M., Garza, C., Onyango, A., and Rolland-Cachera, M. (2008). Who growth standards for infants and young children. *Archives de pediatrie: organe officiel de la Societe francaise de pediatrie*, 16(1):47–53.

[19] Dekker, F. W., De Mutsert, R., Van Dijk, P. C., Zoccali, C., and Jager, K. J. (2008). Survival analysis: time-dependent effects and time-varying risk factors. *Kidney international*, 74(8):994–997.

[20] Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S., et al. (2002). *Analysis of longitudinal data*. Oxford University Press.

[21] Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008). *Longitudinal data analysis*. CRC press.

[22] Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). Applied longitudinal analysis. volume 998, pages 291–486. John Wiley & Sons.

[23] Gałecki, A. and Burzykowski, T. (2013). Linear mixed-effects model. In *Linear Mixed-Effects Models Using R*, pages 245–273. Springer.

[24] Gardiner, J. C., Luo, Z., and Roman, L. A. (2009). Fixed effects, random effects and gee: What are the differences? *Statistics in medicine*, 28(2):221–239.

[25] Ge, K. and Chang, S. (2001). Definition and measurement of child malnutrition. *Biomedical and environmental sciences: BES*, 14(4):283–291.

[26] Group, W. M. G. R. S. et al. (2006). Who child growth standards based on length/height, weight and age. *Acta paediatrica (Oslo, Norway: 1992). Supplement*, 450:76.

[27] Hanley, J. A., Negassa, A., Edwardes, M. D. d., and Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. *American journal of epidemiology*, 157(4):364–375.

[28] Hardin, J. W. (2005). Generalized estimating equations (gee). *Encyclopedia of statistics in behavioral science*.

[29] Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal data analysis*, volume 451. John Wiley & Sons.

[30] Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.

[31] Klein, J. P. and Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

[32] Lawless, J. (2014). Parametric models in survival analysis. *Wiley StatsRef: Statistics Reference Online*.

[33] Leung, K.-M., Elashoff, R. M., and Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual review of public health*, 18(1):83–104.

[34] Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., Lawn, J. E., Cousens, S., Mathers, C., and Black, R. E. (2016). Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals. *The Lancet*, 388(10063):3027–3035.

[35] Mchunu, N. N., Mwambi, H. G., Reddy, T., Yende-Zuma, N., and Naidoo, K. (2020). Joint modelling of longitudinal and time-to-event data: an illustration using cd4 count and mortality in a cohort of patients initiated on antiretroviral therapy. *BMC infectious diseases*, 20:1–9.

[36] Miller Jr, R. G. (2011). *Survival analysis*, volume 66. John Wiley & Sons.

[37] Moïsi, J. C., Gatakaa, H., Berkley, J. A., Maitland, K., Mturi, N., Newton, C. R., Njuguna, P., Nokes, J., Ojal, J., Bauni, E., et al. (2011). Excess child mortality after discharge from hospital in kilifi, kenya: a retrospective cohort analysis. *Bulletin of the World Health Organization*, 89:725–732.

[38] Molenberghs, G. and Verbeke, G. (2006). *Models for discrete longitudinal data*. Springer Science & Business Media.

[39] Mwangome, M. K., Fegan, G., Fulford, T., Prentice, A. M., and Berkley, J. A. (2012). Mid-upper arm circumference at age of routine infant vaccination to identify infants at elevated risk of death: a retrospective cohort study in the gambia. *Bulletin of the World Health Organization*, 90:887–894.

[40] Nemetchek, B., English, L., Kissoon, N., Ansermino, J. M., Moschovis, P. P., Kabakyenga, J., Fowler-Kerry, S., Kumbakumba, E., and Wiens, M. O. (2018). Paediatric postdischarge mortality in developing countries: a systematic review. *BMJ open*, 8(12).

[41] Ngari, M. M., Fegan, G., Mwangome, M. K., Ngama, M. J., Mturi, N., Scott, J. A. G., Bauni, E., Nokes, D. J., and Berkley, J. A. (2017). Mortality after inpatient treatment for severe pneumonia in children: a cohort study. *Paediatric and perinatal epidemiology*, 31(3):233–242.

[42] Ngari, M. M., Mwalekwa, L., Timbwa, M., Hamid, F., Ali, R., Iversen, P. O., Fegan, G. W., and Berkley, J. A. (2018). Changes in susceptibility to life-threatening infections after treatment for complicated severe malnutrition in kenya. *The American journal of clinical nutrition*, 107(4):626–634.

[43] Noble, C., S. J. P. B.-D. M. K. P. A. B. . P. A. J. (2021). Postdischarge interventions for children hospitalized with severe acute malnutrition: a systematic review and meta-analysis. *The American journal of clinical nutrition*.

[44] Organization, W. H. et al. (2009a). Who child growth standards and the identification of severe acute malnutrition in infants and children: joint statement by the world health organization and the united nations children's fund.

[45] Organization, W. H. et al. (2019). Levels and trends in child malnutrition: key findings of the 2019 edition. Technical report, World Health Organization.

[46] Organization, W. H. et al. (2020). Unicef/who/the world bank group joint child malnutrition estimates: levels and trends in child malnutrition: key findings of the 2020 edition.

[47] Organization, W. H., UNICEF, et al. (2009b). Who child growth standars and the identification of severe acute malnutrition in infans and children. *Genever, Switzerland*.

[48] Pan, W. and Wall, M. M. (2002). Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. *Statistics in medicine*, 21(10):1429–1441.

[49] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2021). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-152.

[50] R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[51] Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33.

[52] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.

[53] Rizopoulos, D. (2014). The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *arXiv preprint arXiv:1404.7625*.

[54] Therneau, T. M. and Grambsch, P. M. (2000). The cox model. In *Modeling survival data: extending the Cox model*, pages 39–77. Springer.

[55] Therneau, T. M. and Lumley, T. (2015). Package 'survival'. *R Top Doc*, 128(10):28–33.

[56] Tian, L., Zucker, D., and Wei, L. (2005). On the cox model with time-varying regression coefficients. *Journal of the American statistical Association*, 100(469):172–183.

[57] van Oudenhoven, F. M., Swinkels, S. H., Ibrahim, J. G., and Rizopoulos, D. (2020). A marginal estimate for the overall treatment effect on a survival outcome within the joint modeling framework. *Statistics in Medicine*.

[58] Verbeke, G. (1997). Linear mixed models for longitudinal data. In *Linear mixed models in practice*, pages 63–153. Springer.

[59] Wang, Y. and Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455):895–905.

[60] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

[61] Wu, L., Liu, W., Yi, G. Y., and Huang, Y. (2012). Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012.

[62] Xue, X., Xie, X., Gunter, M., Rohan, T. E., Wassertheil-Smoller, S., Ho, G. Y., Cirillo, D., Yu, H., and Strickler, H. D. (2013). Testing the proportional hazards assumption in case-cohort analysis. *BMC medical research methodology*, 13(1):1–10.

[63] Zhang, Z., Reinikainen, J., Adeleke, K. A., Pieterse, M. E., and Groothuis-Oudshoorn, C. G. (2018). Time-varying covariates and coefficients in cox regression models. *Annals of translational medicine*, 6(7).