Joint predictors of preterm birth and perinatal death among singleton births at a zonal referral hospital in northern Tanzania: A birth registry based study from 2000 to 2017

UNIVERSITY OF
KWAZULU - NATAL

INYUVESI
YAKWAZULU-NATALI

Innocent B. Mboya

October, 2021

Joint predictors of preterm birth and perinatal death among singleton births at a zonal referral hospital in northern Tanzania: A birth registry based study from 2000 to 2017

by

Innocent B. Mboya

A thesis submitted to the

University of KwaZulu-Natal

in fulfilment of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

BIOSTATISTICS

Thesis Supervisors:

Prof. Henry G. Mwambi, Dr. Michael J. Mahande

& Dr. Joseph Obure

UNIVERSITY OF KWAZULU-NATAL

SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE

PIETERMARITZBURG CAMPUS, SOUTH AFRICA

Declaration

I, Innocent B. Mboya, declare that;

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then;
   (a) their words have been re-written but the general information attributed to them has been referenced, or
   (b) where their exact words have been used, then their writing has been placed in italics and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference sections.

[redacted] 12 October 2021

Innocent B. Mboya (Student) Date

[redacted] 12 October 2021

Prof. Henry G. Mwambi (Supervisor) Date

[redacted] 12 October 2021

Dr. Michael J. Mahande (Co-Supervisor) Date

[redacted] 12 October 2021

Dr. Joseph Obure (Co-Supervisor) Date

Disclaimer

This document describes work undertaken as a PhD programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institution.

Contents

List of Tables

List of Figures

Dedication

To God almighty, the Father of my Lord and Saviour Jesus Christ, who created all things, and in Him all things consist, the giver of life to those who love Him and keep His commandments.

To my beloved wife Winfrida, my daughter Gianna, and all women who deserve the right to live birth, and nurturing a healthy child.

Acknowledgements

This work was funded by GSK Africa Non- Communicable Disease Open Lab through the DELTAS Africa Sub- Saharan African Consortium for Advanced Biostatistics (SSACAB) Grant No. 107754/Z/15/Z- training programme. The views expressed in this publication are those of the author(s) and not necessarily those of GSK. It is through the financial support from GSK SSACAB that I was able to pursue my PhD studies. I am grateful to the SSACAB consortium for giving me this opportunity.

I would like to acknowledge my employer, the Kilimanjaro Christian Medical University College (KCMUCo) for the permission and support during my studies. Special appreciations are conveyed to all the staff of the Institute of Public health (IPH) at KCMUCo; Prof. Sia E. Msuya, Prof. Rachel Manongi, Ms. Beatrice John, and Ms. Caroline Amour, to name a few. I am indebted to my mentors and supervisors, Prof. Henry G. Mwambi, Dr. Michael J. Mahande and Dr. Joseph Obure whose comments, contributions and positive criticisms paved a way towards the success of this project.

To the faculty and my colleagues at the University of Kwazulu-Natal, School of Mathematics, Statistics, and Computer Science (Pietermaritzburg campus), Ms. Christel Bernard, Ms. Jesca Batidzirai, Ms. Danielle J. Roberts, and fellow PhD candidates; Alexanda Kasyoki and Mohanad Mohammed. I highly appreciate the spiritual and moral support from my beloved Pastors Rev. Joel M. Senny of the

Infinity Glory Ministries Tanzania and Pastor Shaun Ramroop of the Church on Campus (UKZN). May the Lord God multiply His goodness in all your lives. Amen!

The midwives who participated in data collection and all women and children whose information enabled the availability of data used in this study are highly appreciated. I also thank the staff at the Birth Registry for capturing these data in the electronic system. Lastly but not least, the Centre for International Health at the University of Bergen in Norway and the Department of Obstetrics and Gynecology of the KCMC hospital in Tanzania for establishing the KCMC medical birth registry, which facilitated the availability of data to conduct this study.

List of publications

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

1. **Mboya, I. B.**, Mahande, M. J., Obure, J., & Mwambi, H. G. (2020). Predictors of perinatal death in the presence of missing data: A birth registry-based study in northern Tanzania. PloS one, 15(4), e0231636.

2. **Mboya, I. B.**, Mahande, M. J., Mohammed, M., Obure, J., & Mwambi, H. G. (2020). Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern Tanzania. BMJ open, 10(10), e040132.

3. **Mboya, I. B.**, Mahande M. J., Obure J., & Mwambi H. G. (2021) Predictors of singleton preterm birth using multinomial regression models accounting for missing data: A birth registry-based cohort study in northern Tanzania. PLoS ONE 16(4): e0249411.

4. **Mboya, I. B.**, Mahande, M. J., Mohammed, M., Obure, J., & Mwambi, H. G. (2020). Joint predictors of singleton preterm birth and perinatal death in the presence of missing data: a birth registry-based cohort study in northern Tanzania. *Submitted to Frontiers in Pediatrics*

The author has also contributed to the following publications during the course of his PhD training:

1. Mwimo, J. L., Somoka, S., Leyaro, B. J., Amour, C., Mao, E., & **Mboya, I. B.** (2021). Knowledge, attitude and practice of physical activity among patients with diabetes in Kilimanjaro region, Northern Tanzania: a descriptive cross-sectional study. BMJ open, 11(9), e046841.

2. Mohammed, M., Mwambi, H., **Mboya, I. B.**, Elbashir, M. K., & Omolo, B. (2021). A stacking ensemble deep learning approach to cancer type classification based on TCGA data. Scientific Reports, 11(1), 1-22.

3. Mosha, D., Paulo, H. A., Mwanyika-Sando, M., **Mboya, I. B.**, Madzorera, I., Leyna, G., ... & Fawzi, W. (2021). Risk factors for overweight and obesity among women of reproductive age in Dar es Salaam, Tanzania. BMC Nutr 7, 37 (2021).

4. Philemon, R. N., Mmbaga, B. T., Bartlett, J., Renju, J., Mtuy, T., **Mboya, I. B.**, & Msuya, S. E. (2021). Do Women Enrolled in PMTCT Understand the Recommendations: A Case Study from Kilimanjaro. Patient Preference and Adherence, 15, 1301-1309.

5. Shao, E. R., **Mboya, I. B.**, Lyamuya, F., Temu, E., Nkwama, M., Kilonzo, S., ... & Maro, V. (2021). Uptake of Cost-Free Hepatitis B Vaccination among Healthcare Workers in Northern Tanzania. Tanzania Medical Journal, 32(2), 39-56.

6. Kimondo, F. C., Kajoka, H. D., Mwantake, M. R., Amour, C., & **Mboya, I. B**. Knowledge, attitude, and practice of cervical cancer screening among women living with HIV in the Kilimanjaro region, northern Tanzania. Cancer Reports (2021): e1374.

7. Mahande, M.J., Oneko, O., Amour, C., Pollie, M., Smith, C., **Mboya, I.B.**, & Finkel, M. (2021). Feasibility and acceptability of human papillomavirus self-sampling in a semi-urban area in northern Tanzania. Int. J. Gynecol. Obstet., 154: 113-118.

8. Shao, E. R., **Mboya, I. B.**, Lyamuya, F., Temu, E., Nkwama, M., Kilonzo, S., ... & Maro, V. (2021). Uptake of Cost-Free Hepatitis B Vaccination among Healthcare Workers in Northern Tanzania. Tanzania Medical Journal, 32(2), 39-56.

9. Ali, F., Mgongo, M., Mamseri, R., George, J. M., **Mboya, I. B.**, & Msuya, S. E. Prevalence of and factors associated with early initiation of breastfeeding

among women with children aged <24 months in Kilimanjaro region, northern Tanzania: a community-based cross-sectional study. Int Breastfeed J 15, 80 (2020).

10. **Mboya, I. B.**, Mamseri, R., Leyaro, B. J., George, J., Msuya, S. E., & Mgongo, M. (2020). Prevalence and factors associated with anemia among children under five years of age in Rombo district, Kilimanjaro region, Northern Tanzania. F1000Research, 9(1102), 1102.

11. Kiwango, F., **Mboya, I. B.**, John, B., Hashim, T., Msuya, S. E., & Mgongo, M. (2020). Prevalence and factors associated with timely initiation of breastfeeding in Kilimanjaro region, northern Tanzania: a cross-sectional study. BMC pregnancy and childbirth, 20(1), 1-7.

12. **Mboya, I. B.**, Ngocho, J. S., Mgongo, M., Samu, L. P., Pyuza, J. J., Amour, C., ... & Msuya, S. E. (2020). Community engagement in COVID-19 prevention: experiences from Kilimanjaro region, Northern Tanzania. The Pan African Medical Journal, 35(Suppl 2).

13. **Mboya, I.B.**, Leyaro, B. J., Kongo, A., Mkombe, C., Kyando, E., & George, J. (2020). Internet addiction and associated factors among medical and allied health sciences students in northern Tanzania: a cross-sectional study. BMC psychology, 8(1), 1-8.

14. Yussuf M. H., Elewonibi B. R., Rwabilimbo M. M., **Mboya I. B.**, & Mahande M. J. (2020). Trends and predictors of changes in modern contraceptive use among women aged 15–49 years in Tanzania from 2004–2016: Evidence from Tanzania Demographic and Health Surveys. PLoS ONE 15(6): e0234980.

15. Adams, D. J., Ndanzi, T., Rweyunga, A. P., George, J., Mhando, L., Ngocho, J. S., & **Mboya, I. B**. (2021). Depression and associated factors among geriatric population in Moshi district council, Northern Tanzania. Aging & mental health, 25(6), 1035-1041.

16. **Mboya I. B.**, John B., Kibopile E.S., Mhando L., George J., Ngocho J. S. (2020). Factors associated with mental distress among undergraduate students in northern Tanzania. BMC psychiatry. Dec 1;20(1):28.

17. Mvunta MH, **Mboya I. B.**, Msuya SE, John B., Obure J., Mahande M. J. (2019). Incidence and recurrence risk of low birth weight in Northern Tanzania: A registry-based study. PLoS ONE 14(4): e0215768.

18. Shao, E. R., **Mboya, I. B.**, Gunda, D. W., Ruhangisa, F. G., Temu, E. M., Nkwama, M. L., Pyuza, J. J., Kilonzo, K. G., Lyamuya, F. S., & Maro, V. P. (2018). Seroprevalence of hepatitis B virus infection and associated factors among healthcare workers in northern Tanzania. BMC Infect Dis, 18(1), 474.

19. Mauka, W. I., Mtuy, T. B., Mahande, M. J., Msuya, S. E., **Mboya, I. B.**, Juma, A., & Philemon, R. N. (2018). Risk factors for inappropriate blood requisition among hospitals in Tanzania. PLoS One, 13(5), e0196453.

Abbreviations

aOR	Adjusted Odds Ratio
AIC	Akaike Information Criteria
ANC	Antenatal Care
BMI	Body Mass Index
FIML	Full Information Maximum Likelihood
GEE	Generalized Estimating Equations
GLMM	Generalized Linear Mixed Models
GLM	Generalized Linear Models
ICD-10	ICD, 10th revision
ICD	International Classification of Diseases
IPI	Inter-pregnancy Interval
IUGR	Intrauterine Growth restriction
KCMC	Kilimanjaro Christian Medical Center
KCMUCo	Kilimanjaro Christian Medical University College
LRT	Likelihood Ratio Test
LIML	Limited Information Maximum Likelihood
LIMQL	Limited Information Maximum Quasilikelihood
LBW	Low birth Weight
MoH	Ministry of Health
MoHCD	Ministry of Health and Community Development
MoHCDGEC	Ministry of Health, Community Development, Gender, Elderly and Children
MICE	Multiple imputation by Chained Equations
NBS	National Bureau of Statistics
OR	Odds Ratio
OCGS	Office of the Chief Government Statistician
OLR	Ordered Logistic Regression
PMR	Perinatal Mortality Rate
PROM	Premature Rupture of Membranes

PTB	Preterm Birth
pPROM	Preterm Premature Rupture of Membranes
QIC	Quasi-likelihood Information Criteria
REML	Restricted Maximum Likelihood
RR	Risk Ratio
SSA	Sub-Saharan Africa
TDHS-MIS	Tanzania Demographic and Health Survey and Malaria Indicator Survey
UNDP	United Nations Development Programme
UKZN	University of Kwazulu-Natal
WHO	World Health Organization

Abstract

Background: Globally, preterm birth (births before 37 completed weeks of gestational) contributes to under-five and newborn deaths. Tanzania ranks the tenth country with the highest preterm birth rates globally and shares 2.2% of the global proportion of all preterm births and contributes to perinatal deaths. Perinatal deaths (stillbirths and early neonatal deaths) continue to increase relative to under-five deaths, especially in low- and middle-income countries. Previous exposure to perinatal death increases preterm birth risk. Understanding the independent and joint predictors of these outcomes may inform interventions to accelerate progress towards achieving sustainable development goals. The study aimed to determine the joint predictors of preterm birth and perinatal death among singleton births in northern Tanzania.

Methods: The study utilized birth registry data from Kilimanjaro Christian Medical Center (KCMC) zonal referral hospital from 2000 to 2017, located in Moshi Municipality, Kilimanjaro region, Northern Tanzania. Generalized estimating equations (GEE) estimated the marginal effects of covariates on perinatal death. The predictive capacity of machine learning algorithms was compared with the classical logistic regression model to predict perinatal death. Multinomial logistic regression with cluster adjusted robust standard errors determined predictors of preterm birth. Joint predictors of preterm birth and perinatal death and the co-occurrence were estimated using the random-effects models to account for the correlation between these outcomes.

Results: Perinatal mortality in this cohort slightly declined while preterm birth rates were increasing. Maternal demographic characteristics and pregnancy-related conditions and complications increase the risk of these outcomes. The joint predictors of higher risk of preterm birth and perinatal death were inadequate (<4) ANC visits, referred for delivery, and complications during pregnancy and childbirth, specifically pre-eclampsia/eclampsia, PPH, LBW, abruption placenta, and breech presentation. Younger maternal age (15-24 years), PROM, placenta previa, and male children have higher odds of preterm birth but a lessened likelihood of perinatal death.

Conclusion: ANC is a critical entry point for delivering the recommended interventions to pregnant women, especially those at high risk of experiencing adverse pregnancy outcomes. Improved management of complications during pregnancy and childbirth and the postnatal period may eventually lead to substantially reducing adverse perinatal outcomes towards improving maternal and child health.

Keywords: Preterm birth, perinatal death, GEE, logistic regression, machine learning, missing data mechanisms, MAR, nonmonotone pattern, FCS, joint/multivariate regression, joint modelling, correlated binary outcomes, Tanzania.

Chapter 1

Background

1.1 Epidemiology of preterm birth

Every year, an estimated 15 million babies are born preterm (before 37 completed weeks of gestation) in the world, and this number is rising (World Health Organization, 2020). Preterm birth is a leading cause of deaths in children under five years of age globally and it was responsible for over 1 million deaths in 2015 (Liu et al., 2016; Quinn et al., 2016; World Health Organization, 2020). It increases the risk of babies dying from other causes especially neonatal infections (Blencowe et al., 2013). Despite modern advances in obstetric and neonatal management, the rate of preterm birth are reported to be increasing in the developed world (Georgiou et al., 2015).

Every year, an estimated 15 million babies (11%) are born preterm (before 37 completed weeks of gestation) globally (Chawanpaiboon et al., 2019; Liu et al., 2016; World Health Organization, 2020), majority (81.1%) of these occurs in Asia and sub-Saharan Africa (SSA) (Chawanpaiboon et al., 2019). The rates of preterm birth in SSA are notably high in Nigeria (6.9%), Ethiopia (12.0%), and Tanzania (16.6%) (Chawanpaiboon et al., 2019). Tanzania ranks the tenth country with the highest preterm birth rates in the world, and contributes a 2.2% of the global

proportion of all preterm births (Chawanpaiboon et al., 2019). About seven years ago (2013), Tanzania was not in the top 10 countries with the highest (>15%) preterm birth rates globally. By then, Malawi had the highest preterm birth rate (18%) in SSA and South East Asia (Blencowe et al., 2013; van den Broek et al., 2014). The country specific estimates shows that the proportion of preterm birth ranged between 12-13% in Mwanza region (Watson-Jones et al., 2007; Mahande et al., 2013b; Temu et al., 2016; Rugaimukam et al., 2017) to as high as 24% among HIV infected women in Dar es Salaam (Zack et al., 2014). In northern Tanzania, a 14.2% prevalence of preterm birth was reported (Temu et al., 2016).

Preterm birth is a syndrome with a variety of causes which can be classified into two broad subtypes: (1) spontaneous preterm birth (spontaneous onset of labour or following pre-labour premature rupture of membranes (pPROM)) and (2) provider-initiated preterm birth (defined as induction of labor or elective caesarean birth before 37 completed weeks of gestation for maternal or fetal indications (both "urgent" or "discretionary"), or other non-medical reasons) (Goldenberg et al., 2008; Blencowe et al., 2013; Quinn et al., 2016; Phillips et al., 2017; World Health Organization, 2020). Higher risk of preterm birth is reported among women with a history of preterm delivery, those with low ($\leq$24) or high maternal age ($\geq$40), short inter-pregnancy intervals (<24 months), low maternal body mass index (BMI), multiple pregnancies, maternal infections such as urinary tract infections, malaria, bacterial vaginosis, HIV and syphilis and those with inadequate (<4) ANC visits (Blencowe et al., 2013; van den Broek et al., 2014; Mahande & Mahande, 2016; Mahande & Obure, 2016; Temu et al., 2016; Fuchs et al., 2018). Stress and excessive physical work or long times spent standing, smoking and excessive alcohol consumption, sex of the child (more among males compared to females), hypertensive disorders of pregnancy such as pre-eclampsia or eclampsia, placental abruption, cholestasis, fetal distress, fetal growth restriction, small gestational age (a birth weight below the tenth percentile for the gestational age), and early

induction of labor or cesarean birth (before 39 completed weeks of gestation) whether for medical or non-medical reasons also increases the risk of preterm birth (Blencowe et al., 2013; Temu et al., 2016; van Zijl et al., 2016; Yang et al., 2016; Teoh et al., 2018; World Health Organization, 2020).

The risk of preterm birth is reported to recur in subsequent pregnancies (Mazaki-Tovi et al., 2007; Mahande et al., 2013b; Laughon et al., 2014; Yang et al., 2016; Su et al., 2018). Recurrent preterm birth is defined as two or more deliveries before 37 completed gestational weeks (Mazaki-Tovi et al., 2007). Recurrence risk of preterm delivery occurs both in low- and high-income countries (Mahande et al., 2013b; Laughon et al., 2014). The associated factors include; pre-eclampsia/eclampsia and diabetes, perinatal death, low birthweight, gestational age at delivery in the first pregnancy, short (<24 months) and long (>36 months) inter-pregnancy intervals, depression, a previous stillbirth and maternal infections (e.g. UTI) (Mahande et al., 2013b; Laughon et al., 2014; Grantz et al., 2015; Yang et al., 2016; Malacova et al., 2018).

The world has made substantial progress in improving child survival since 1990. However, SSA remains to be one of the regions with the highest newborn deaths in the world (UNICEF et al., 2020; You et al., 2015). The average under-five mortality rate in SSA was 76 deaths per 1,000 live births in 2019 compared to 180 deaths per 1,000 live births in 1990 (UNICEF et al., 2020). A substantial progress in reducing childhood mortality rates has also been reported in Tanzania. For example, the under-5 mortality rates have declined from 147 to 67 deaths per 1,000 live births in 1999 and 2015-2016, respectively (MoHCDGEC [Tanzania Mainland] et al., 2016). Despite progress in reducing under-five mortality, the world is witnessing an increase in neonatal deaths (i.e. deaths within one month after birth). Forty seven percent of the global under-five deaths in 2019 occurred before the first month of life (neonatal period) (UNICEF et al., 2020). Hence, child survival remains an

urgent concern especially in Sub-Saharan African countries. Research to determine predictors of preterm births is an important stage towards addressing the burden of this problem.

To increase child survival rates and reduce complications due to preterm birth, WHO recommends essential care during childbirth and in the postnatal period for every mother and baby (i.e. routine practice for the safe childbirth before, during and after birth), provision of antenatal steroid injections, kangaroo mother care and antibiotics to treat newborn infections (World Health Organization, 2020). Tanzania has also adopted these strategies (Ministry of Health and Social Welfare [MoHSW], 2015). It is one of the five countries where WHO implements a clinical trial on the immediate kangaroo mother care (KMC) for women at risk of preterm birth (World Health Organization, 2020). This study may therefore contribute to reliable information about preterm birth and associated factors that may be used to improve newborn outcomes in Tanzania.

1.2 Epidemiology of perinatal death

The perinatal period commences at 28 completed weeks of gestation and ends seven completed days after birth. Perinatal mortality refers to the number of stillbirths (pregnancy loss that occurs after 7 months of gestation) and early neonatal deaths (deaths of live births within the first 7 days of life) (MoHCDGEC [Tanzania Mainland] et al., 2016; World Health Organization, 2019). Because perinatal and maternal health are closely linked, perinatal mortality is used as an indicator to measure maternal health status as well as quality of antenatal, intrapartum and newborn care, and is an important health indicator (Mmbaga et al., 2012b; Mpembeni et al., 2014; World Health Organization, 2016a, 2019).

Globally, more than 5 million perinatal deaths occur each year (World Health Organization, 2016a). Children face the highest risk of dying in their first month of

life, at a global rate of 17 deaths per 1,000 live births (UNICEF et al., 2020). Globally 2.6 million children died in the first month of life in 2016 – approximately 7,000 newborn deaths every day – most of which occurred in the first week, with about 1 million dying on the first day and close to 1 million dying within the next six days. Sub-Saharan Africa is one of the regions with the highest rate – 28 per 1000 live births (UNICEF et al., 2020). The patterns of these deaths are similar to the patterns for maternal deaths; the majority occurring in developing countries (World Health Organization, 2019). In Tanzania, between the year 2004-2005 and 2015-16, the under-five mortality rate declined from 112 to 67 deaths per 1,000 live births but the country has witnessed an increase in the number of stillbirths (from 143 to 187), number of early neonatal deaths (from 156 to 214) as well as perinatal mortality rate (from 36 to 39) deaths per 1,000 live births between 2010-11 and 2015-16 respectively (MoHCDGEC [Tanzania Mainland] et al., 2016).

The causes of stillbirths and early neonatal deaths are closely linked, and examining just one or the other is reported to bias the true level of mortality around delivery (Ouyang et al., 2013; MoHCDGEC [Tanzania Mainland] et al., 2016). The risk of perinatal mortality has been associated with preterm birth, shorter birth interval (<24 months), congenital anomaly, previous history of early neonatal death, low birth weight, maternal anemia, placental abruption, ruptured uterus, systemic infections/sepsis, pre-eclampsia, eclampsia, obstetric hemorrhage, having a home delivery, fetal growth restrictions and maternal infections such as syphilis and malaria (Nankabirwa et al., 2011; Bayou & Berhan, 2012; Mmbaga et al., 2012a; Mpembeni et al., 2014; Unterscheider et al., 2014; Vogel et al., 2014; Getiye & Fantahun, 2017). Although these factors may be common across low-, middle- and high-income countries, they are likely to differ depending on the context or country specific conditions such as availability of quality obstetric and newborn care services at different levels of care.

WHO has classified the causes of perinatal deaths according to the timing of death as; antepartum (the time during pregnancy but before child birth), intrapartum (the time between the onset of labor through delivery of the placenta) or neonatal period (time between birth up to 7 days post delivery) as well contributing maternal conditions e.g. hypertension (World Health Organization, 2016a; Getiye & Fantahun, 2017). Classifying death by timing still provides valuable information for analysis and targeting of programmes in these areas.

As with other negative maternal outcomes, perinatal deaths are reported to recur in subsequent pregnancies (Salihu et al., 2011; Mahande et al., 2013a; Ouyang et al., 2013). At the end of their second pregnancies, mothers whose first pregnancy had ended in a stillbirth and whose first infant had died as a neonate were at increased risk of experiencing recurrence of these events in developing countries (Ouyang et al., 2013). Perinatal deaths have also been reported to recur in a hospital-based study in northern Tanzania (Mahande et al., 2013a). Furthermore, recurrence risk of perinatal death has also been associated with characteristics in the first pregnancy including; pre-eclampsia/eclampsia, placental abruption, placenta previa, induced labor, preterm delivery, low birth weight, stillbirth and early age at initiation of pregnancy (Salihu et al., 2011; Mahande et al., 2013a).

WHO recommends a minimum of eight contacts for antenatal care that can reduce perinatal deaths by up to 8 per 1000 births when compared to a minimum of four visits (World Health Organization, 2016b). Early identification and management of women with complications has been recommended to improve maternal and perinatal outcomes (Vogel et al., 2014). Appropriate and informed interventions are therefore necessary to accelerate progress towards achieving the second indicator of sustainable development goal 3, i.e. “by 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at

least as low as 25 per 1,000 live births" (UNDP, 2018). These interventions should consider context specific factors that can better explain the risk of perinatal deaths. As perinatal mortality rate is reported to have increased countrywide, this study will determine current trends and associated factors among deliveries from women recorded at KCMC medical birth registry in northern Tanzania.

1.3 Literature review

1.3.1 Literature search strategy

Literature was searched from Google scholar, Science Direct, PubMed and Scopus databases. The local UKZN library system provided more access to articles that would not be accessible through Google Scholar from some of the publishers. Books (those available at the UKZN library and on-line) particularly those that are specific to statistical models were also searched and used to enrich the literature referred in this study. The search terms used include preterm delivery, preterm birth, recurrence of preterm birth, recurrence risk of preterm birth, predictors of preterm birth/delivery, perinatal death, perinatal mortality, stillbirth and early neonatal death, recurrence of perinatal death, recurrence of adverse perinatal outcomes, predictors of perinatal death/mortality, generalized estimating equations, ordinal logistic regression, ordered logistic regression, joint modeling of categorical outcomes, joint modelling of correlated binary outcomes, multivariate models for binary data, marginal models for categorical data.

1.3.2 Preterm births rates

Every year, an estimated 15 million babies are born preterm and the rates are increasing (Blencowe et al., 2013; World Health Organization, 2020). This estimate is most probably an underestimation, as most countries have an incomplete birth registration (van Zijl et al., 2016). The global estimate of preterm birth rates in 2010 was 11.1% (Blencowe et al., 2013). Majority (60%) of all preterm births occur in

sub-Saharan Africa or South Asian countries, but preterm birth is truly a global problem (Blencowe et al., 2013; World Health Organization, 2020). Low- and middle-income countries (LMIC) account for the majority of the world's preterm births whereby, in the lower-income countries, on average, 12% of babies are born too early compared with 9% in higher-income countries (Quinn et al., 2016; Purisch & Gyamfi-Bannerman, 2017; Vogel et al., 2018). Within countries, poorer families are at higher risk (World Health Organization, 2020). Preterm birth complications are a leading cause of deaths in children under-5 years of age (Blencowe et al., 2013; Liu et al., 2016; UNICEF et al., 2020; World Health Organization, 2020) and according to WHO, in almost all countries with reliable data, preterm birth rates are increasing (World Health Organization, 2020).

There are significant variations in the incidence of preterm birth worldwide (Purisch & Gyamfi-Bannerman, 2017). The rates of preterm birth in 184 countries in 2010 ranged from 5% in several Northern European countries to 18% in Malawi (Blencowe et al., 2013; Purisch & Gyamfi-Bannerman, 2017). The PTB rate in USA was reported to be 9.6% in 2015 though not very different from the rate reported in the 1980s i.e. 9.5% (Purisch & Gyamfi-Bannerman, 2017). Grantz et al. (2015) reported that 7.8% of women in Utah state USA delivered preterm. This suggests that, there are differences in PTB rates across states in this country. In the UK, a proportion of 14.5% was reported among women at high-risk for preterm delivery in view of their pregnancy history, which included previous late miscarriage, PTB or significant cervical surgery (Teoh et al., 2018). Studies among high risk pregnancies are likely to report a large proportion of PTB due to several factors, including the documented recurrence of negative pregnancy outcomes in subsequent pregnancies (Mazaki-Tovi et al., 2007; Mahande et al., 2013b; Laughon et al., 2014; Grantz et al., 2015).

Different hospital based studies on PTB have also been conducted in Asian

countries. The proportion of PTB was reported to be 6.1% in India (Ahankari et al., 2001), 8.6% from eight hospitals in Western China (Lu et al., 2015a) and 5.1% in Iran (Alijahan et al., 2014). This indicates variations in PTB rates across countries. India and China are the top two countries with the highest number of preterm births though their PTB rates are not among the top 10 in the world (World Health Organization, 2020).

The highest documented PTB rates in sub-Saharan Africa for the past seven years (in 2013) was from Malawi, which had the highest PTB rate (18.1 preterm births per 100 births) in the world (van den Broek et al., 2014; World Health Organization, 2020). However, as previously mentioned, a recent systematic review and modelling analysis indicated that, Tanzania is the tenth country with the highest (16·6%) PTB rate in the world (Chawanpaiboon et al., 2019). This systematic review and modelling analysis only focused on estimating the global burden of preterm birth and documented challenges in data quality and gestational age estimations. Reasons for Tanzania being ranked tenth with high preterm birth rates are not discussed in the paper. However, the study utilized data from an observational study among women attending antenatal care in Mwanza city from 1997 to 2000 and followed-up to the delivery assessed occurrence of adverse pregnancy outcomes, including preterm birth (Watson-Jones et al., 2007); the KCMC Medical Birth registry data reported by Mahande et al. (2013b) from 2000-2018 and 1999-2006 data by Habib et al. (2011).

Hospital-based data such as the KCMC Medical Birth registry are most likely to overestimate the actual burden of preterm birth in Tanzania. Although 75% of all deliveries are self-referred for delivery to KCMC hospital, there may still be some concerns about the generalizability of these findings to the Tanzanian population. The current study reported trends and determinants of preterm birth and perinatal deaths and is crucial to informing interventions. Based on the findings presented in

this thesis, the study is an essential milestone to bridge gaps in data availability, analysis, interpretation, and reporting (Gliklich et al., 2020).

There are also variations in PTB rate between and within countries in SSA. Among women who delivered in Gondar town health institutions, Northwest Ethiopia, prevalence of PTB was 4.4% (Gebreslasie, 2016). In Kenya, prevalence ranged between 18.3% in 2013 (Wagura et al., 2018) and 20.2% in 2017 in Kenyatta National Hospital (Okube & Sambu, 2017), an increase of about 2%. The high absolute number of preterm births in African and Asian countries are related, in part, to high fertility and the large number of births in those two regions in comparison to other parts of the world (Blencowe et al., 2013). A meta analysis in East African Countries (Tanzania, Kenya and Uganda) had estimated the prevalence of PTB to range from 2.7% in Mwanza Tanzania to 5.9% in Kabale Uganda (Marchant et al., 2012). A birth cohort study in rural Uganda reported a PTB rate of 19.4% (Bater et al., 2020), which is even higher than global estimates (Chawanpaiboon et al., 2019). Current estimates are crucial to inform interventions and policy decisions.

Furthermore, there are also differences in PTB rates across regions in Tanzania. Prevalence ranged between 12-13% in Mwanza region (Watson-Jones et al., 2007; Rugaimukam et al., 2017) to 24% among HIV infected women in Dar es Salaam (Zack et al., 2014). Prevalence in Kilimanjaro region, northern Tanzania was consistently reported to be about 14% (Mahande et al., 2013b; Temu et al., 2016). Tanzania is one of the countries where WHO is implementing the immediate kangaroo mother care (KMC) multi-country trial (which is a continuous skin-to-skin contact, breastfeeding support, and promotion of early hospital discharge with follow-up) for improving outcomes of preterm births (World Health Organization, 2020). Current estimates on PTB rate in the country, which also document trends over time are crucial to inform public health interventions and programs. The PTB rates are on the rise in both low-, middle- and high-income

countries (Georgiou et al., 2015; Purisch & Gyamfi-Bannerman, 2017; Chawanpaiboon et al., 2019; World Health Organization, 2020). Without context specific interventions, these rates will continue to rise.

1.3.3 Predictors of preterm births

Preterm labour is now thought to be a syndrome initiated by multiple causes though a precise mechanism cannot be established in most cases; therefore, factors associated with PTB, but not obviously in the causal pathway, have been sought to explain preterm labour (Goldenberg et al., 2008; Vogel et al., 2018; World Health Organization, 2020). There are many maternal and fetal characteristics that have been associated with increased risk of preterm birth. These are categorized in maternal demographic characteristics, nutritional status, pregnancy history, present pregnancy characteristics, psychological characteristics, adverse behaviours, infections, uterine contractions and cervical length, and biological and genetic markers (Goldenberg et al., 2008; Vogel et al., 2018). High risk of PTB in high income countries such as USA and UK, has been linked to racial and ethnic disparities especially among blacks compared to women from other racial groups (Goldenberg et al., 2008; Purisch & Gyamfi-Bannerman, 2017).

Generally, maternal factors associated with the risk of PTB include home delivery (Bater et al., 2020), low ($<$20 years) and high ($>$35 years) maternal age, short inter-pregnancy intervals ($\leq$24 months), low maternal body mass index (BMI) (i.e. $<$19 Kg/m^2), multiple pregnancies, low socio-economic and educational status, being single and smoking and alcohol consumption during pregnancy (Blencowe et al., 2013; Goldenberg et al., 2008; Mahande & Obure, 2016; Okube & Sambu, 2017). Teoh et al. (2018) did not find child gender differences in the occurrence of PTB contrary to study by Temu et al. (2016) who reported lower risk among females compared to males though results were not statistically significant. Goldenberg et al. (2008) indicated that, the mechanisms by which the maternal demographic

characteristics are related to preterm birth are unknown.

Other factors associated with increased risk of preterm birth include maternal infections such as urinary track infection, malaria, HIV and other STIs, maternal conditions such as hypertension during pregnancy, pre-eclampsia/ eclampsia, maternal anemia and gestational diabetes, low birth weight (LBW), inadequate ($\leq$4) ANC visits, multiple pregnancies, heavy physical work during pregnancy, placenta previa, placenta abruption, complications during pregnancy, cervical incompetence, polyhydramnios (excess of amniotic fluid in the amniotic sac), preterm premature rupture of membranes (PPROM) and antepartum hemorrhage (Watson-Jones et al., 2007; Alijahan et al., 2014; van den Broek et al., 2014; Lu et al., 2015a; Gebreslasie, 2016; Mahande & Mahande, 2016; Mahapula et al., 2016; Temu et al., 2016; Ahankari et al., 2001; Okube & Sambu, 2017; Rugaimukam et al., 2017; Wagura et al., 2018; Bater et al., 2020). Intimate partner violence (IPV) during pregnancy has also been linked with higher odds of PTB (Hill et al., 2016; Vogel et al., 2018). Other studies did not find a significant association between several of these factors (including socio-demographic characteristics, maternal anemia and HIV positive status) with increased risk of PTB (Watson-Jones et al., 2007; Ahankari et al., 2001; Wagura et al., 2018). This can be explained by the timing of anemia screening (i.e. before or on the day of delivery) (Ahankari et al., 2001), the mediating effect of HIV infection on women with malaria and/ or anemic which may increase the risk of negative pregnancy outcomes (Watson-Jones et al., 2007) and the small number of women who delivered prematurely (Wagura et al., 2018).

A systematic review by Malacova et al. (2018) has linked previous exposure to adverse pregnancy outcomes such as PTB and stillbirth with increased risk of PTB. Likewise, history of preterm birth has been reported by several studies to be a major risk factors for PTB recurrence in subsequent pregnancies (Mazaki-Tovi et al., 2007; Mahande et al., 2013b; Laughon et al., 2014; Grantz et al., 2015; Okube &

Sambu, 2017; Phillips et al., 2017; Malacova et al., 2018; Su et al., 2018). Also, women who deliver a twin pregnancy are at greater risk for delivering prematurely in a subsequent singleton pregnancy (Schaaf et al., 2012). Defining risk factors for prediction of preterm birth is a reasonable goal for identification of at-risk women that allow initiation of risk-specific treatment and also provide important insights into mechanisms leading to PTB (Goldenberg et al., 2008). This study will determine predictors of preterm birth taking into account the natural ordering of this outcome which is crucial in informing clinical, epidemiological and public health decisions and interventions.

1.3.4 Perinatal death rates

Globally, more than 5 million perinatal deaths occur each year (World Health Organization, 2016a; UNICEF et al., 2020). Children face the highest risk of dying in their first month of life, at a global rate of 17 deaths per 1,000 live births. Globally 2.4 million children died in the first month of life in 2019 – approximately 6,700 neonatal deaths every day, with Central and South East Asia and SSA carrying the highest burden (UNICEF et al., 2020). "About a third of all neonatal deaths occur within the first day after birth, and close to three quarters occur within the first week of life" (Lawn et al., 2014; UNICEF et al., 2020).

Significant differences exists in perinatal death rates in developed compared to developing countries. Perinatal mortality rate (PMR) was reported to be 5.4 per 1000 live births in Ireland (Unterscheider et al., 2014) which is slightly lower than that reported in USA (6.0 per 1000 births) (Gregory et al., 2018). These rates are over five times lower compared to the rates in developing nations (Bayou & Berhan, 2012; Getiye & Fantahun, 2017). The estimated perinatal mortality rate in Brazil, Southern America, was about four times higher (20.3 per 1000 births) (Miranda et al., 2017) than that reported in Ireland (5.4 per 1000 live births) (Unterscheider et al., 2014). These inequalities may reflect differences in the availability of quality

obstetric and newborn care services.

There is no significant variation in the reported PMR in Asian countries such as India (49.4 per 1000 births) (Bellad et al., 2010) as compared to estimates in some SSA countries such as Nigeria (49.9 per 1000 live births) (Oyira et al., 2017). A Meta-Analysis of Demographic and Health Surveys from 21 SSA countries revealed a PMR of 34.7 per 1000 births, Eastern African region carrying the highest burden (Akombi & Renzaho, 2019). Tanzania has the highest PMR in the region (Akombi & Renzaho, 2019). Also, country specific estimates in Eastern African ranged from 41 per 1000 live births in Uganda to 75.3 per 1000 live births in Eastern Sudan (Nankabirwa et al., 2011; Ali et al., 2013). The unacceptably higher rates in Sudan were linked to higher rates of home deliveries, low antenatal care coverage and maternal conditions such as anemia and malaria infection (Ali et al., 2013). There could also be some variations in PMR in East African countries depending on the source of data i.e. community/population vs hospital/facility based surveys. For instance, the Rwanda DHS has reported a rate of 29 deaths per 1,000 pregnancies (NISR [Rwanda] et al., 2015), which is nearly 2-3 times lower than that reported in Uganda and Sudan (Nankabirwa et al., 2011; Ali et al., 2013).

As earlier indicated, the perinatal mortality rate has been reported to increase in Tanzania from 36 to 39 deaths per 1,000 live births between 2010-11 and 2015-16 DHS survey rounds, relative to under-5 mortality, respectively (MoHCDGEC [Tanzania Mainland] et al., 2016). Similar trends were also reported in northern Tanzania. PMR ranged between 27 per 1000 births in Manyara region to 57.7 per 1000 births in Kilimanjaro region (Hinderaker et al., 2003; Mmbaga et al., 2012a). Mahande et al. (2013a) reported that, among 3,909 women who delivered singletons for the subsequent (second and above) pregnancy in the KCMC medical birth registry, 7% lost their child in a perinatal death in their first recorded pregnancy. The estimated PMR in Kilimanjaro region is higher than the national

and zonal estimates (MoHCDGEC [Tanzania Mainland] et al., 2016). This increase in PMR in Tanzania could impair efforts to improve child survival and the achievement of SDGs. Registry based cohort studies serve as an important data source to provide current trends and rate of perinatal mortality hence inform interventions and policy decisions to improve maternal and child health.

1.3.5 Predictors of perinatal deaths

The WHO has classified the causes of perinatal deaths according to the timing of death as; antepartum (the time during pregnancy but before child birth), intrapartum (the time between the onset of labor through delivery of the placenta) or neonatal period (time between birth up to 7 days post delivery) (World Health Organization, 2016a; Allanson et al., 2016). This classification also links the contributing maternal conditions with perinatal deaths, given that a maternal condition is frequently found in the context of a perinatal death (Allanson et al., 2016). Worldwide, PTB complications are reported to be the leading cause of deaths among children under-5 years of age, followed by pneumonia and intrapartum related events (Liu et al., 2016; World Health Organization, 2020). Although Liu et al. (2016) reported pneumonia to be a leading cause of under-5 deaths in SSA, PTB complications remains to be one of the biggest threats to child survival, given its increasing contribution to under-5 deaths both in the region and worldwide.

Findings of the WHO multicountry survey on maternal and newborn health have associated perinatal deaths with placental praevia, placental abruption, ruptured uterus, systemic infections/sepsis, pre-eclampsia, eclampsia and severe anemia among other complications or diseases (Vogel et al., 2014). A review by Nijkamp et al. (2017) has provided critical information regarding maternal medical history, obstetric history, current pregnancy (complications), drugs or medications and other risk factors that are associated with perinatal deaths that should be evaluated and recorded to improve management and prevent recurrence in subsequent

pregnancies. Despite variations in the causes of perinatal deaths between and within countries, various studies have reported similar factors associated with this outcome. For instance, in Ireland, perinatal death was associated with low birth weight (LBW), prematurity, pregnancy comorbidities such as hypertension and diabetes, poor obstetric histories such as prior perinatal death, intrauterine growth restriction (IUGR) recurrent pregnancy loss and ethnicity (high among ethnic minorities) (Unterscheider et al., 2014) as also reported by other studies (Bellad et al., 2010; Miranda et al., 2017; Nijkamp et al., 2017).

In African countries, similar and some other factors have been associated with increased risk of perinatal deaths. These include, PTB, pregnancy induced hypertension, obstetric hemorrhage, antepartum hemorrhage, LBW, birth asphyxia (limited blood flow and deprivation of oxygen to a newborn infant before, during, or after the birth process), maternal anemia, hypertensive disorders of pregnancy (including gestational hypertension, pre-eclampsia and eclampsia), congenital anomalies (also known as birth defects, congenital disorders or congenital malformations), recurrent pregnancy outcomes such as LBW, neonatal and perinatal deaths, short ($<$24 months) or long ($>$37 months) inter-pregnancy intervals, low maternal age ($<$20 years), nulliparous (women who never given birth), higher rates of home delivery, parity ($\geq$3), obstructed labour (a condition frequently known as labour dystocia), fetal malpresentation, never use of mosquito nets (for malaria prevention), low uptake of antenatal care (ANC) services and antenatal iron supplementation (Nankabirwa et al., 2011; Bayou & Berhan, 2012; Ali et al., 2014; Habimana-Kabano et al., 2015; Allanson et al., 2016; Getiye & Fantahun, 2017; Oyira et al., 2017).

Likewise, studies in Tanzania reported that, the risk of perinatal deaths was also associated with prematurity (PTB), birth asphyxia, preeclampsia/eclampsia, congenital malformations, infections (such malaria, syphilis and urinary tract

infections), LBW, history of previous pregnancy outcomes such as child death, hypertensive disorders, prolonged obstructed labour, nulliparity and nutritional deficiencies (Hinderaker et al., 2003; Mmbaga et al., 2012b; Mpembeni et al., 2014). High risk of recurrence has been documented among women who experienced adverse pregnancy outcomes such as perinatal death (including stillbirth and/ or early neonatal death) in the previous pregnancy(ies) in both developed and developing countries (Watson-Jones et al., 2007; Salihu et al., 2011; Mahande et al., 2013a; Ouyang et al., 2013). With increasing regional and national rates of perinatal deaths, it is important to determine the associated factors which is necessary for targeted interventions.

1.4 Problem statement

Despite all efforts put forward during the MDG era, perinatal and newborn deaths continues to increase relative to under-five deaths. Child survival remains at the heart of the SDG agenda and an urgent concern (You et al., 2015). The highest number of deaths in children under-five years of age are reported to occur during the perinatal period (UNICEF et al., 2020). An estimated 15 million babies are born too early every year (World Health Organization, 2020). Preterm birth is a leading cause of deaths in children under five years of age (especially perinatal deaths) (Liu et al., 2016; Quinn et al., 2016; World Health Organization, 2020) and a significant cause of long-term loss of human potential among survivors around the world (Blencowe et al., 2013; Malacova et al., 2018; World Health Organization, 2020). There is, therefore, a close relationship between preterm births and perinatal deaths because as it also reflect perinatal care services (Lee et al., 2019a), with provider-initiated preterm birth (often caesarean section) used to prevent potential fetal deaths. According to World Health Organization (2020), in almost all countries with reliable data, preterm birth rates are increasing. This would imply a proportional increase in the number of both stillbirths and early neonatal deaths, especially in SSA and South-Eastern Asia. To accelerate progress towards reducing

preterm births and perinatal death rates, country specific estimates are needed to document current trends and risks using available data.

This study utilized cohort data from KCMC zonal referral hospital birth registry, which documents deliveries prospectively since the year 2000. The study will provide an 17 years (i.e. 2000-2017) trend in preterm births and perinatal deaths and associated factors. Such data are crucial in assessing effectiveness of interventions to improve child survival towards achievement of sustainable development goals (SDG). No study has assessed the joint predictors of preterm birth and perinatal death which underpins the need for this study. We developed a joint model to assess the joint predictors of preterm birth and perinatal death, particularly using the random effects approach (Fitzmaurice et al., 2009; Molenberghs & Verbeke, 2005). While other studies such as those by (Chuwa et al., 2017; Isaksen et al., 2015; Mahande, 2015; Mmbaga et al., 2011, 2012b) ignored the presence of missing values in their analysis (by performing complete case analysis), we applied multiple imputation techniques to account for missing data, hence improved parameter estimates.

Furthermore, most studies on preterm birth analyzed this outcome as a binary variable ignoring the natural ordering or other preterm birth categories. Based on the WHO definition, preterm birth is further grouped according to the levels of severity (World Health Organization, 2020). It is therefore imperative to consider these sub-categories and determine associated risk factors compared to those reported in binary regression analyses that will encourage focused care for women at high risk of experiencing this and related pregnancy outcomes.

1.5 Study justification

A good understanding of predictors for different negative pregnancy outcomes may contribute to designing of preventive strategies especially for high-risk

pregnancies. Existence of a well established medical birth registry better serves this purpose among many others (Bergsjo et al., 2007). A joint model between preterm birth and perinatal death will contribute to better understanding of potential risk factors hence inform preventative and therapeutic measures. This study will also provide current trends of these outcomes in Tanzania which is crucial to reduce the risk and related negative pregnancy outcomes.

Women experiencing preterm birth and perinatal death are mostly faced with psychological, social and economic challenges associated with caring for the babies born preterm (for instance, preterm deliveries are associated with prolonged hospital stays for intensive care) and the loss of an infant. Most of the risk factors for both preterm birth and perinatal deaths can be reduced through cost-effective interventions before, during and after delivery, during antenatal, labor and postnatal care (World Health Organization, 2016b, 2020).

Perinatal death is a devastating obstetric complication (Nijkamp et al., 2017). Determination of causes of death (or associated risk factors) helps in understanding why and how it occurs, and it is an indispensable aid to parents wanting to understand why their baby died and to determine the recurrence risk and management in subsequent pregnancy. Capturing the chain of maternal and fetal events that led to perinatal death can inform preventative and therapeutic measures (Nankabirwa et al., 2011; Allanson et al., 2016). Hence, findings from this study will contribute valuable information that can be used to inform care throughout the course of pregnancy to reduce the devastation that parents are likely to experience due to pregnancy related complications. Also, this study may inform future studies on these and related adverse maternal and fetal outcomes.

1.6 Study objectives

1.6.1 Broad objective

The main objective of this study was to assess the independent and joint predictors of preterm birth and perinatal death among singleton birth in northern Tanzania based on KCMC zonal referral hospital birth registry data between 2000 to 2017.

1.6.2 Specific objectives

1. To determine predictors of perinatal death among singleton births using the KCMC zonal referral hospital birth registry data between 2000 to 2017.

2. To determine predictors of singleton preterm birth using the KCMC zonal referral hospital birth registry data between 2000 to 2017.

3. To determine the joint predictors of preterm birth and perinatal death among singleton birth using the KCMC zonal referral hospital birth registry data between 2000 to 2017.

4. To develop and apply novel statistical methodology to model PTB and perinatal death outcomes, including the methods for handling missing information in the data set.

The rest of the thesis is organized as follows: Chapter 2 describes the data source and exploratory analysis; Chapter 3 describes the statistical and computational analysis methods; Chapter 4-7 presents background, methods, results, and discussion based on each paper/manuscript; Lastly, Chapter 8 focus on the general discussion, conclusion, and recommendations.

Chapter 2

Description of the data source and exploratory analysis

2.1 Study design

We conducted a secondary analysis of birth cohort data from the Kilimanjaro Christian Medical Centre (KCMC) referral hospital, situated in the Moshi Municipality of Kilimanjaro region, Northern Tanzania. The KCMC medical birth registry was established at KCMC to; secure a working system for medical birth registration, provide research data on the women reproductive health and for monitoring perinatal health and quality of care at KCMC among other purposes (Bergsjo et al., 2010). A description of the KCMC medical registry is also available elsewhere (Bergsjo et al., 2007, 2010; Mmbaga et al., 2012a; Mahande, 2015).

2.2 Study area

The Kilimanjaro Christian Medical Centre (KCMC) is the national zonal referral hospital located in Moshi Municipality, Kilimanjaro Region, Northern Tanzania. Based on the 2012 National Population and Housing Census, the region was projected to have a total population of 1,790,113 in the year 2017 of which males were 865,692 (48.4%) and females 924,421 (51.6%) (NBS & OCGS, 2018). The region

has a total of seven districts, including Moshi Municipality and Moshi rural district where approximately 70% of births that occur at KCMC comes from. This means that the main catchment area of the hospital is the local population (Mmbaga et al., 2012b; Mahande, 2015). The hospital also admits referred cases from the rest of the regional districts as well as six other regions; Arusha, Kilimanjaro, Manyara, Tanga, Dodoma and Singida (Mahande, 2015). The KCMC Medical birth registry data is not linked to other health facilities across these regions but captures information for births recorded in the department of obstetrics and gynecology.

The Tanzania demographic and health survey and malaria indicator survey (DHS-MIS) 2015-16 estimated the total fertility rate in Kilimanjaro region to be 3.4 children per woman for the three years preceding the survey (MoHCDGEC [Tanzania Mainland] et al., 2016). For a 10 years period, Maro et al. (2016) estimated a very high maternal mortality rate at KCMC hospital, i.e., 492.1 per 100,000 live births, though this figure is slightly lower than the national estimate of 556 per 100,000 live births (MoHCDGEC [Tanzania Mainland] et al., 2016).

2.3 Study population and eligibility criteria

Women and their respective siblings were recruited and data prospectively collected since the year 2000 to date. The birth registry was designed to include every birth recorded at the KCMC zonal referral hospital for all consenting mothers. The hospital has an average of 3500-4000 births each year (Chuwa et al., 2017; Mahande & Mahande, 2016; Mitao et al., 2016), which is close to 70,000 deliveries to date. All consenting mothers are interviewed using a standardized questionnaire and data are prospectively recorded in the KCMC medical birth registry. We analyzed data for all women and their respective singleton deliveries recorded at the KCMC medical birth registry between the years 2000-2017. Women with multiple pregnancies were excluded from our analyses to minimize over-representation of high risk pregnancies (Mahande, 2015).

Prospective sample size calculations allow for optimal sample size planning in order to obtain adequate control over the risks of type I and II errors (Columb & Atkinson, 2015). Longitudinal studies such as the birth registries tend to be more powerful than cross-sectional studies (Diggle et al., 2002). Hospital-based studies to estimate the risk of different adverse pregnancy outcomes often require a large sample size in order to have large statistical power and enhance generalizability of results (Mahande, 2015; Manor et al., 2000). Between the year 2000-2017, there was a total of 55907 recorded deliveries in the KCMC medical birth registry from 45324 mothers aged 15-49 years. The large sample size from this registry enhanced statistical power and reliability of the study findings. Large sample sizes also provide room for performing stratified analyses (Katz et al., 2013; Mboya et al., 2020b).

2.4 Data collection methods

Birth data at KCMC have been recorded using a standardized questionnaire (Appendix A) and entered into a computerized database located at the birth registry. Data collection is done by trained Project Midwives, mothers being interviewed within the first 24 hours after birth given a normal delivery (Bergsjo et al., 2010; Mmbaga et al., 2012a). Although the printed questionnaires are in English language, the Project Midwives who conduct the interviews are well vested in English and Swahili languages (Bergsjo et al., 2010). Patient files are used for verification of interview data and to extract supplementary information. Following a cesarean section, mothers are interviewed on the second or third day depending on her condition. A unique identification number was assigned to each woman at first admission and used to trace her medical records at later admissions (Mahande, 2015).

2.5 Study variables

The KCMC birth registry contains data on the basic demographic/background characteristics of parents and the child, mothers health before and during pregnancy, delivery-related information, including complications, and child status. Demographic characteristics included age (years), area of residence, occupation, tribe, current marital status, and highest education level. Information about mother's health before and during pregnancy included body weight and height, smoking and alcohol use during pregnancy, history of serious diseases such as diabetes, hypertension, malaria, and anemia, family planning practices, antenatal care (ANC) during the present pregnancy and if so, the number of visits, HIV status, as well as diseases and complications during present pregnancy such as gestational diabetes, diabetes, hypertension, preeclampsia/eclampsia, bleeding and anemia.

Information concerning the delivery included parity, induction of labor, amount of blood loss, mother's health after delivery, including whether experienced maternal death or not, and complications, particularly premature rapture of the membranes (PROM), bleeding (>500 ml), 3-4 degree tear, abruption placenta, and placenta previa. The registry also captures child-related information such as the date and time of delivery, gestational age at birth, child's sex, birth weight, Apgar scores (1, 5, and 10 minutes), presentation, mode of delivery, and child status, i.e., whether a child is a live-born, stillborn or experienced neonatal death.

2.6 Data management

Data were imported from Access database to STATA version 15.1 (StataCorp LLC, College Station, Texas, USA) for cleaning and analysis. We also used R version 3.4.1 for data analysis (for work in Chapter 5). Unique identification numbers (ID) created to link mothers and her subsequent deliveries was used to merge/link the

mother-child records. Duplicate and unusual observations were identified and resolved prior to analysis in consultation with the KCMC birth registry data managers. We reshaped the data from wide (where all deliveries occurred in columns) to a long format (where a mother could have multiple records) to account for repeated deliveries within a mother. Depending on the distribution of variables in the dataset (both numeric and categorical), some were re-categorized for better interpretation and comparison with other studies. Re-categorization of these variables was informed by previous literature. For instance, maternal age was categorized as 15-19, 20-24, 25-34, 35-39, and 40+. Also, due to small number of deliveries born at $<$28 gestational weeks (extremely preterm), preterm birth was grouped as $<$32 (extremely/very preterm), 32-$<$37 (moderate to late preterm), and $\geq$37 weeks (term), especially for regression analysis of this outcome. The patterns of missing values was also identified to inform imputation.

2.7 Ethical considerations

Ethical approval for the establishment of the birth registry at KCMC were given by the local Institutional Review Board of KCMC hospital, the National Ethics Committee in Norway and by the Tanzania Ministry of Health, Commission for Science and Technology (Bergsjo et al., 2010). For practical reasons, since the interview was administered just after the woman had given birth, consent was given orally. The midwife-nurse gave every woman oral information about the birth registry, the data needed to be collected from them, and the use of the data for research purposes. Women were also informed about the intention to gather new knowledge, which will, in turn, benefit mothers and children in the future. Participation was voluntary and had no implications on the care women would receive.

Following consent, mothers were free to refuse to reply to single questions. For privacy and confidentiality, unique identification numbers were used to both

identity and then link mothers with child records. There was no any person-identifiable information in any electronic database, and instead, unique identification numbers were used. Necessary measures were taken by midwives to ensure privacy during the interview process. Necessary measures were taken by project midwives to ensure privacy during the interview (Bergsjo et al., 2010). This study was approved by the Kilimanjaro Christian Medical College Research Ethics and Review Committee (KCMU-CRERC) with approval number 2424, particularly for data access.

Chapter 3

Statistical and computational analysis methods

3.1 Descriptive analysis

Descriptive statistics were summarized using measures of central tendency (mean/median) with their corresponding measures of dispersion (standard deviation/interquartile range), respectively for numeric variables. Categorical variables were summarized using frequency and percentages. Line graphs were used to display trends over time. We used standard linear regression models to assess linear trends in the proportions of preterm birth and perinatal death over time (years). Chi-square (χ^2) test was used to compare proportion of preterm birth and perinatal death across different levels of explanatory variables. Data analysis involved handling missing data using multiple imputation and followed three stages; the first stage was the complete case analysis. The second stage was the imputation of the missing values in both the outcomes and covariates, followed by the analysis of the imputed dataset. Results are compared before and after imputations.

Furthermore, as we have also stated in (Mboya et al., 2020b), for the analysis of

missing data, we assumed a nonmonotone pattern of missingness in which some subject values were observed again after a missing value occurs (Ibrahim & Molenberghs, 2009; Jakobsen et al., 2017). Multiple imputation is a commonly used method to deal with missing data, which accounts for the uncertainty associated with missing data (Pedersen et al., 2017; Jakobsen et al., 2017; Sterne et al., 2009). Under a nonmonotone pattern of missingness, it is recommended to use the chained equations, also referred to as fully conditional specification (FCS) (Van Buuren et al., 2006a; Azur et al., 2011) or the Markov Chain Monte Carlo (MCMC) method to impute missing values (Jakobsen et al., 2017). We, therefore, used multiple imputation by FCS to handle missing data in this study. This technique is a powerful and statistically valid method for creating imputations in large datasets, which include both categorical and continuous variables (Liu & De, 2015a; Van Buuren et al., 2006a; White et al., 2010; Azur et al., 2011). Additional details about the assumptions, multiple imputation technique implementation, and the FCS algorithm are in Section 3.4.

For objective 1 (addressed in Chapter 4), we used generalized estimating equations (GEE) with binomial family, logit link, exchangeable correlation structure, and robust variance estimator. Results from this model were compared to the GEE log-linear regression model, i.e., Poisson family, log link funcion, an exchangeable correlation structure, and robust variance estimator (Mboya et al., 2020b). GEE is an extension of Generalized Linear Models (GLM), which is a subject-specific or conditional models assuming independence of observations within an individual subject i.e. observation are independent and identically distributed (referred as the iid requirement) (Hardin & Hilbe, 2003). Contrary to this assumption, there are many common data situations for which responses are correlated. For instance, data of patients from different health facilities within the same district may be different. This may be influenced by the level of health facility (e.g. district vs referral hospital) and the type of services offered at each level. Based on such

conditions, the assumption of independence is violated because of the correlation of observations. GEE models were therefore developed to address the dependence of longitudinal and clustered data (Hardin & Hilbe, 2003; Bergsma et al., 2009; Reddy, 2017). A population-averaged model is one which includes the within-panel (cluster) dependence by averaging effects over all panels (Hardin & Hilbe, 2003). Registry based studies include data collected at different time points, successive records being linked using unique identifiers. A mother can have one or more deliveries recorded each time they deliver in the health facility.

In the crude/unadjusted analysis, a 10% significance level was used to select covariates to include in the multivariable models. Variable selection was performed using stepwise regression applying both forward selection and backward elimination methods (Chowdhury & Turin, 2020). However, variables reported in the literature to significantly increase the risk of perinatal death were retained in the models regardless of their level of significance. Relative risk (RR) and Odds Ratios (OR) with their 95%CI were used to determine the strength of association at 5% significant level. Furthermore, a follow-up paper (which is the content of Chapter 5) was written to predict perinatal deaths using machine learning models compared to the logistic regression model (Mboya et al., 2020a). Details about the applied machine learning models, the statistical approach, and results are presented in Chapter 5, which are also published in (Mboya et al., 2020a).

For objective 2 (which is the content of Chapter 6), we used multinomial logistic regression models to determine the predictors of preterm birth as opposed to previous studies. Epidemiologists are often interested in estimating the risk of adverse events originally measured on an interval scale (such as gestational age in weeks). Still, they often choose to divide the outcome into two or more categories to compute an estimate of effect (risk or odds ratio) (Ananth & Kleinbaum, 1997). Such restrictions assume that these outcomes share similar characteristics,

including associated factors that may not always be the case. Previous studies have analyzed preterm birth as a binary variable (Ahankari et al., 2001; Laughon et al., 2014; Lu et al., 2015a; Mahande et al., 2013b; Malacova et al., 2018; Teoh et al., 2018; van den Broek et al., 2014) ignoring other severe forms of preterm birth (World Health Organization, 2020). Cluster adjusted robust variance estimator was used to account for repeated observations/deliveries within mothers. Likelihood ratio test was used to compare nested models while Akaike Information Criteria (AIC) for comparing non-nested models and for model selection. Likewise, in the unadjusted analysis, a 10% significance level was used to select covariates to include in the multivariate models with stepwise regression used for variable selection. Also, variables identified in the literature as clinically important in the association with preterm birth were retained in the model regardless of their level of significance. Odds ratios (OR) and 95%CI were used to determine the strength of association at 5% significance level.

Furthermore, for objective 3 (content of Chapter 7), we jointly modelled the predictors of preterm birth and perinatal death using the random effects approach. Studies on the joint modelling of adverse pregnancy events such as preterm birth and perinatal death are limited. Preterm birth is a known major risk factor for perinatal death (Goldenberg et al., 2008; Bayou & Berhan, 2012; Georgiou et al., 2015; Getiye & Fantahun, 2017; Liu et al., 2016). Traditionally, joint models were developed for the analysis of longitudinal and time-to-event data (Henderson et al., 2000; McCrink et al., 2011; Andrinopoulou, 2014; Asar et al., 2015). Joint models applied to categorical responses are also termed as multivariate models (McCullagh, 1980; Glonek, 1996; Molenberghs & Lesaffre, 1999). By fitting two independent models of preterm birth and perinatal death, we are making a restrictive assumption that the two data generation processes are independent. This may not be a valid assumption because two outcomes from the same individual could be highly correlated. Therefore, a joint model is relevant to account for

dependence nature of the two outcomes. The statistical methodology for joint modelling of correlated binary outcomes applied in this thesis is presented in Section 7.2.5.

3.2 Marginal models for longitudinal data

In this section, we illustrate and describe statistical methodology applied in the analysis of repeated measurements using generalized estimating equations (GEE) in relation to generalized linear models (GLMs). GEE are an extension of GLMs to correlated observations as opposed to the simpler case of independent observations. The methods described in this section were applied to analyse data for Objective 1, presented in Chapter 4 and have been published here (Mboya et al., 2020b).

3.2.1 Full information maximum likelihood (FIML) estimating equation for GLM

Without loss of generality, we focus on the case of binary data as an example. Because GEE is an extension of GLM, we will start by introducing the likelihood based equation for Bernoulli regression, then describe the estimating equation for a population averaged model including the description of relevant correlation structures for repeated measurements. Let Y_i denote our binary outcome of interest for an individual i where $i = 1, ..., n$ such that,

$$Y = \begin{cases} 1 & \text{if the event of interest has occurred} \\ 0 & \text{if the event has not occurred} \end{cases}$$

Hardin & Hilbe (2003) provided a simple description of the estimating equation for Bernoulli distributed data in their book on Generalized Estimating Equations. We will use their description to establish the foundation for modeling binary data in the generalized linear models and later build up to GEE that account for correlation of observations within clusters. According to Hardin & Hilbe (2003), the Bernoulli

distribution, a limiting case of the binomial distribution, is the appropriate choice for the estimation with binary data. Its density function in the case of a single observation or outcome is given by;

$$f(y|p) = p^y(1-p)^{1-y} \tag{3.1}$$

where $p \in (0,1)$ is the probability of success and $E(y) = p \in [0,1]$ are the expected values and $V(y) = p(1-p) \in (0,1)$ representing the variance for y. The exponential family of the distribution has a location parameter θ, a scale parameter $\alpha(\phi)$, and a normalizing term $c(y,\phi)$ with the probability density;

$$f(y;\theta,\phi) = exp\left\{\frac{y\theta - b(\theta)}{\alpha(\phi)} + c(y,\phi)\right\} \tag{3.2}$$

where $E(y) = b'(\theta) = \mu$ and $V(y) = b''(\theta)\alpha(\phi)$ and b'' is the variance of μ. Based on equation (3.1) for Bernoulli distribution and in the case of n independent observations, the joint density is the product of the densities for the individual outcomes.

$$f(y_1,...,y_n) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} \tag{3.3}$$

$$= \prod_{i=1}^{n} \exp\left\{y_i \ln\left(\frac{p}{1-p}\right) + \ln(1-p)\right\} \tag{3.4}$$

The likelihood is obtained by considering the outcome as Bernoulli distributed and the parameter p as unknown. Thus;

$$L(p|y_1,...,y_n) = \prod_{i=1}^{n} \exp\left\{y_i \ln\left(\frac{p}{1-p}\right) + \ln(1-p)\right\} \tag{3.5}$$

In order to introduce covariates that model the outcome, we introduce a subscript to the notation to allow the mean response to reflect a dependence on the linear combination of the covariates and the associated coefficients. The notation p will now change to p_i. We also use $\mu_i = E(y_i)$ instead of the Bernoulli expected value p_i

for consistence with use among various distributions. The likelihood will now be;

$$L(\mu|y_1,...,y_n) = \prod_{i=1}^{n} \exp\left\{ y_i \ln\left(\frac{\mu_i}{1-\mu_i}\right) + \ln(1-\mu_i) \right\} \tag{3.6}$$

Thus, the log-likelihood is given by;

$$L(\mu_i|y_1,...,y_n) = \sum_{i=1}^{n} \{ y_i \ln(\mu_i) - y_i \ln(1-\mu_i) + \ln(1-\mu_i) \} \tag{3.7}$$

where μ_i is the mean specific to i^{th} observation. This is because, μ_i now depends on covariates associated with that observation. We let $\mu = (\mu_1, \ldots, \mu_n)$. However, handling the model as it is leads to what is known as a saturated model i.e. a model with parameters equal to the number of observation. Therefore, we need to introduce a structure to the mean (μ_i).

The covariates are introduced in the model through the expected value of the outcome variable, which are in the range (0, 1) through the linear predictor - a set of independent covariates and associated coefficients to be estimated. Let the linear predictor be defined as

$$\eta_i = X_i'\beta \in \Re \tag{3.8}$$

where $X_i = (1, X_{i1}, \ldots, X_{ip})'$ is the vector of covariates including allowance for a constant term and $\beta = (\beta_0, \ldots, \beta_p)$ are the $p+1$ regression coefficients including the intercept. The variance of the outcome (y_i) is given by;

$$V(y_i) = \mu_i(1-\mu_i) \tag{3.9}$$

where $\mu_i \in (0, 1)$ is the expected value of the outcome. Finally, if we parameterize the model using the natural or canonical link the model becomes;

$$g(\mu_i) = \ln\left(\frac{\mu_i}{1-\mu_i}\right) = X_i'\beta \tag{3.10}$$

This implies;

$$\mu_i = \frac{e^{X_i\beta}}{1+e^{X_i\beta}} = \mu_i(X_i'\beta) \tag{3.11}$$

The general FIML estimating equation $\Psi(\Theta) = 0$ for $\Theta = (\beta)$ i.e. the vector of coefficients for Bernoulli distributed data is given by;

$$\left[\left\{\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^{n}\left(\frac{y_i}{\mu_i} - \frac{1-y_i}{1-\mu_i}\right)\left(\frac{\partial \mu}{\partial \eta}\right)_i x_{ij}\right\}_{j=1,\ldots,p}\right]_{p\times 1} = [0]_{p\times 1} \tag{3.12}$$

The general estimating equation for the exponential family (also referred as limited information maximum likelihood (LIML)) uses the same specification and parameterization of the linear predictor and is given by;

$$\left[\left\{\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^{n}\frac{y_i-\mu_i}{a(\phi)V(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta}\right)_i x_{ij}\right\}_{j=1,\ldots,p}\right]_{p\times 1} = [0]_{p\times 1} \tag{3.13}$$

3.2.2 Limited information maximum quasilikelihood (LIMQL) estimating equation for GEE

The GEE model begins with consideration of the LIMQL (Liang et al., 1992; Hardin & Hilbe, 2003; Ziegler, 2011). This is applicable on functions that are not from an exponential family that is where the log-likelihood implied by the estimating equation is called *quasilikelihood* (Hardin & Hilbe, 2003; Lee et al., 2006). Furthermore, Hardin & Hilbe (2003) indicated that, the estimating equation based on LIMQL is relevant for GEE models due to very restrictive assumptions of the exponential family that assumes; 1) the form of variance function is a known function of the mean and 2) independence of observations. When we have repeated measurements within a cluster, the assumptions of independence and constant variance will not hold.

Suppose we have independent responses $y_1, \ldots, y_n$ with means $E(y_i) = \mu_i$ and variance $V(y_i) = a(\phi)V(\mu_i)$, where μ_i is a function of unknown regression parameters $\beta = (\beta_0, \ldots, \beta_p)$ where β is a vector of coefficients for p covariates. The

estimating equation based on quasilikelihood is given by (Hardin & Hilbe, 2003; Lee et al., 2006);

$$\frac{\partial q(\mu_i; y_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{V(\mu_i)a(\phi)} d\mu_i \tag{3.14}$$

Assuming within panel independence of observations, the LIMQL estimating equation for GLMs is given by;

$$\left[\left\{ \Psi(\beta) = \sum_{i=1}^{n} \sum_{t=1}^{n_i} \frac{y_{it} - \mu_{it}}{a(\phi)V(\mu_{it})} \left(\frac{\partial \mu}{\partial \eta} \right)_{it} x_{ijt} \right\}_{j=1,\ldots,p} \right]_{p\times 1} = [0]_{p\times 1} \tag{3.15}$$

where p is the column dimension of the matrix of covariates $\mathbf{X}$. In matrix form we can re-write this equation as;

$$\left[\left\{ \Psi(\beta) = \sum_{i=1}^{n} x_{ij}^T D \left(\frac{\partial \mu}{\partial \eta} \right) [V(\mu_i)]^{-1} \left(\frac{y_i - \mu_i}{a(\phi)} \right) \right\}_{j=1,\ldots,p} \right]_{p\times 1} = [0]_{p\times 1} \tag{3.16}$$

where D denotes a diagonal matrix. $V(\mu)$ is also a diagonal matrix which can be decomposed into;

$$V(\mu_i) = \left[D(V(\mu_{it}))^{1/2} I_{(n_i \times n_i)} D(V(\mu_{it}))^{1/2} \right]_{(n_i \times n_i)} \tag{3.17}$$

implying the correlation matrix is identity, I.

3.2.3 GEE modification of the LIMQL

The essential idea behind the GEE approach is to generalize and extend the usual likelihood equations for generalized linear model for univariate response by incorporating the covariance matrix of the vector of responses, Y_i (Fitzmaurice et al., 2011). Marginal models for longitudinal data has the following assumptions (Diggle et al., 2002; Fitzmaurice et al., 2011):

1. The marginal expectation of the response, $E(Y_{ij}|X_{ij}) = \mu_{ij}$, depends on covariates through a known link function, $g(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta$. For example, the link function can either be logit for binary response and log for counts.

2. The marginal variance of each Y_{ij} given the covariates depends on the marginal mean according to, $Var(Y_{ij}|X_{ij}) = \phi v(\mu_{ij})$, where $v(\mu_{ij}$ is a known variance function and ϕ is a scale parameter which may be known or may need to be estimated.

3. The correlation between Y_{ij} and Y_{ik} is a function of marginal means and an additional parameters α. That is $Corr(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}, \alpha)$, where $\rho(.)$ is a known function.

In the GEE formulation, Liang et al. (1992) introduces a more general correlation matrix $\mathbf{R}(\alpha)$ that replaces the identity matrix, when we have repeated (correlated) measurements within a subject or clustered observations within a cluster depending on the observation unit. The Correlation matrix is estimated using a parameter vector α. Replacing $I_{(n_i \times n_i)}$ with $R(\alpha)$ in equation (3.17), we now have;

$$V(\mu_i) = \left[D(V(\mu_{it}))^{1/2}\mathbf{R}(\alpha)_{(n_i \times n_i)}D(V(\mu_{it}))^{1/2}\right]_{(n_i \times n_i)} \tag{3.18}$$

According to Hardin & Hilbe (2003), the solution for the estimating equation is obtained using optimization techniques which iterate toward a solution by updating a current estimate until convergence. Using a Taylor series expansion of an estimating equation given by $\Psi(\beta) = 0$, and given a starting estimate of β^0, the solution is iterated using the following relationship;

$$\beta^{(k)} = \beta^{(k-1)} + \left[-\frac{\partial}{\partial\beta}\left(\beta^{(k-1)}\right)\right]^{-1}\Psi\left(\beta^{(k-1)}\right) \tag{3.19}$$

3.2.4 The working correlation structures

There are different correlation matrices used to parameterize the correlation structure. These include independence, exchangeable (which goes with different other names including compound symmetry), stationary, nonstationary, unstructured and autoregressive correlation structures (Liang et al., 1992; Hardin & Hilbe, 2003; Hedeker & Gibbons, 2006; Ziegler, 2011). Our interest is not to compare these correlation structures but to describe those recommended for longitudinal

studies.

3.2.4.1 Exchangeable correlation

The simplest form of the correlation matrix is the identity matrix assumed by the independence model which assumes that observations within a cluster have a common correlation (Hardin & Hilbe, 2003). An exchangeable/compound symmetry correlation is a simple extension of this structure that assumes observations within a panel/cluster have a common correlation. Only one parameter α is estimated in this working correlation matrix and has the following structure;

$$R(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha & \dots & \alpha \\ \alpha & 1 & \alpha & \dots & \alpha \\ \alpha & \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \dots & 1 \end{pmatrix} \tag{3.20}$$

The panels for the data used in this thesis represents mothers and repeated measurements are children within mothers. In an exchangeable correlation structure, Pearson residuals given by (Hardin & Hilbe, 2003; Molenberghs & Verbeke, 2005; Wang, 2014)

$$\hat{r}_{ij} = \frac{(y_{ij} - \hat{\mu}_{ij})}{\sqrt{v(\hat{\mu}_{ij})}} \tag{3.21}$$

from the current fit of the model are used to estimate the common correlation parameter. The estimate of $\hat{\alpha}$ from these residuals is given by (Wang, 2014)

$$\hat{\alpha} = \frac{1}{(N-p)\phi} \sum_{i=1}^{K} \sum_{j \neq k} r_{ij} r_{ik} \tag{3.22}$$

where $N = \sum_{i=1}^{K} n_i(n_i - 1)$ and r_{ij} defines the pearson residuals given in equation 3.21. The estimation of the scale/over-dispersion parameter ($\hat{\phi}$) is described in

Section 3.2.4.4.

3.2.4.2 Autoregressive correlation

The more reasonable working correlation structure for repeated measurements is the autoregressive working correlation of order 1 (commonly abbreviated as AR or AR1) (Ziegler, 2011). The AR1 correlation, $Corr(y_{jk}, y_{jk})$=ρ_{jk}=$\alpha^{|j-k|}$ has the following structure;

$$R(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{n-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{n-3} \\ . & . & . & \dots & . \\ \alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \dots & 1 \end{pmatrix} \tag{3.23}$$

for general n x n matrix.

This working correlation reflects the fact that the correlation decreases exponentially across the lags of the time points (Hedeker & Gibbons, 2006; Ziegler, 2011). We can also estimate the correlation parameters α and scale parameter ϕ using Pearson residuals from the current fit of the model. The AR1 estimate of $\hat{\alpha}$ is given by (Wang, 2014)

$$\hat{\alpha} = \frac{1}{(N-p)\phi} \sum_{i=1}^{K} \sum_{j \leq n_i - 1} r_{ij} r_{i,j+1} \tag{3.24}$$

where $N = \sum_{i=1}^{K} (n_i - 1)$. The correlation matrix is then built from the autoregressive structure implied by the AR correlations. An autoregressive process of order k has nonzero correlations for many more than k lags; the matrix is constant along all major diagonals (Hardin & Hilbe, 2003).

3.2.4.3 Unstructured correlation

The most general of the correlation structures is the unstructured correlation matrix which does not impose any assumption on a specific form to the correlation matrix hence called unstructured working correlation (Hedeker, 2003; Ziegler, 2011). At each time point, a different variance and covariance is estimated with unstructured correlation matrix given as;

$$R(\alpha) = \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & \alpha_{14} & \alpha_{15} \\ \alpha_{12} & 1 & \alpha_{23} & \alpha_{24} & \alpha_{25} \\ \alpha_{13} & \alpha_{23} & 1 & \alpha_{34} & \alpha_{35} \\ \alpha_{14} & \alpha_{24} & \alpha_{34} & 1 & \alpha_{45} \\ \alpha_{15} & \alpha_{25} & \alpha_{35} & \alpha_{45} & 1 \end{pmatrix} \tag{3.25}$$

for a 5 x 5 correlation matrix. It is mostly assumed that the elements of the matrix are time independent. Likewise, using Pearson residuals $\hat{r}_{ij}$ in equation 3.21 from the current fit of the model, the working correlation parameter ($\hat{\alpha}_{jk}$) for the unstructured correlation is given by (Molenberghs & Verbeke, 2005)

$$\hat{\alpha}_{jk} = \frac{1}{(K-p)\phi} \sum_{i=1}^{K} r_{ij} r_{ik} \tag{3.26}$$

3.2.4.4 Estimating the scale variance/dispersion parameter

Using Pearson residuals $\hat{r}_{ij}$, statistical packages such as Stata and SAS have in-built options of estimating two separate equations of the dispersion parameter (ϕ) given as (Hardin & Hilbe, 2003; Fitzmaurice et al., 2011; Wang, 2014)

$$\hat{\phi} = \frac{1}{(N-p)} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \hat{r}_{ij}^2 \tag{3.27}$$

where N=$\sum n_i$, which is the total number of observations, $\hat{r}_{ij}^2$ is the i^{th} Pearson residual and p is the number of covariates in the model. Equation 3.27 above is the default in SAS and model results (for independent correlation) match exactly the GLM results. Alternatively, other software packages uses the following equation

(Hardin & Hilbe, 2003; Molenberghs & Verbeke, 2005; Wang, 2014)

$$\hat{\phi} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i}\hat{r}_{ij}^2 \tag{3.28}$$

which differs with 3.27 in the denominator (the later not subtracting p). This equation is the default in Stata and is said to produce an invariant results (with any correlation structure) to panel-level changes in of the dataset (Hardin & Hilbe, 2003).

3.3 Random effects models for longitudinal data

In longitudinal designs, repeated observations are nested within subjects which results in multilevel or hierachical data in which the level-1 observations (subjects or repeated observations) are nested within the higher level-2 observations (clusters or subjects) (Hedeker, 2005). Generalized linear mixed model (GLMM) is an extension of the generalized linear model that permits random effects as well as fixed effects in the linear predictor (Hedeker, 2005; Agresti, 2010). GLMMs are often referred to as conditional models in contrast to the marginal generalized estimating equations (GEE) models (Hedeker, 2005).

Let Y_{ij} denote the response variable for the i^{th} subject on the j^{th} measurement occasion. Let X_{ij} denote a vector of predictor variables for jth observation in the ith subject. Because subjects may not have the same number of repeated measures and may not be measured at the same set of occasions, we assume there are n_i repeated measurements of the response of the i^{th} subject and that each Y_{ij} is observed at time t_{ij}. For the i^{th} subject, $Y_i = (Y_{i1}, \ldots, Y_{in_i})$ and $i = 1, \ldots, N$. Assume the conditional distribution of each Y_{ij} given a $q \times 1$ vector of random effects b_i belongs to the exponential family of distribution. Then $Var(Y_{ij}|b_i) = \phi v(Y_{ij}|b_i)$, where $v(.)$ is a known variance function. Given the random effects (b_i), the repeated measurements, $Y_{i1}, \ldots, Y_{ini}$ are assumed to be independent of one another; "the conditional independence assumption" (Diggle et al., 2002; Fitzmaurice et al., 2011).

The general form of the GLMM is given as

$$g\{E(Y_{ij}|b_i)\}) = \eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i \tag{3.29}$$

for some known link function, $g(.)$. The random effects b_i are assumed to have a multivariate normal distribution, with zero mean and $q \times q$ covariance matrix, G such that, $b_i \sim N(0, G)$. The conditional variance is assumed to depend on the conditional mean according to Var$(Y_{ij}|b_i) = \phi v\{E(Y_{ij}|b_i)\}$, where $v(.)$ is a known variance function and ϕ is a scale parameter that may be know or may need to be estimated (Fitzmaurice et al., 2009).

3.3.1 GLMM for a binary response

When the response variable is binary, logistic regression models enable assessment of the association between an independent variable(s) and the response variable. When the response variable has more than two categories, generalizations of the logistic model have been suggested (Manor et al., 2000). Estimation of regression coefficients using GLM has been described in Section 3.2.1. In this section, the GLM will be extended for repeated measurements or in other words longitudinal data using random effects models. For a binary response, Y_{ij} takes the values 1 if an event has occurred and 0 if otherwise. Conditional on a single random effect b_i, Y_{ij} have a Bernoulli distribution with $Var(Y_{ij}|b_i) = E(Y_{ij}|b_i)(1-E(Y_{ij}|b_i))$ in which case $\phi = 1$. The conditional mean of Y_{ij} depends on fixed and random effects via the following linear predictor (Fitzmaurice et al., 2011)

$$\eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i = X'_{ij}\beta + b_i, \tag{3.30}$$

where Z_{ij}=1 for all $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$. Using a logit link function, this can be expressed as;

$$logit\left\{\frac{P(Y_{ij} = 1|b_i)}{P(Y_{ij} = 0|b_i)}\right\} = \eta_{ij} = X'_{ij}\beta + b_i \tag{3.31}$$

The single random effect b_i is assumed to have a univariate normal distribution with zero mean and variance g_{11} (Fitzmaurice et al., 2011). Random effects are assumed to have a bivariate normal distribution with zero mean and 2 x 2 covariance matrix, G. No particular computational simplification arises when we focus on the logistic model with Gaussian random effects (Diggle et al., 2002). The likelihood function for β and G for binary data is

$$L(\beta, G; Y_i) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \{\mu_{ij}(\beta, b_i)\}^{y_{ij}} \{1 - \mu_{ij}(\beta, b_i)\}^{1-y_{ij}} f(b_i; G) db_i \tag{3.32}$$

where $\mu_{ij}(\beta, b_i) = E(Y_{ij}|b_i; \beta)$. With the logit link and Gaussian assumption on the b_i, this reduces to (Diggle et al., 2002)

$$\begin{aligned} &\prod_{i=1}^{N} \int exp\left[\beta' \sum_j x_{ij} y_{ij} + b_i' \sum_j d_{ij} y_{ij} - \sum_j \log\{1 + exp(x_{ij}'\beta + d_{ij}' b_i)\}\right] \\ &\times (2\pi)^{-1} |G|^{-q/2} exp(-b_i' G^{-1} b_i / 2) db_i \end{aligned} \tag{3.33}$$

where G is the $q \times q$ variance matrix of each b_i.

3.3.2 GLMM for ordinal response

Suppose Y_{ij} is an ordinal response with C categories where $(c = 1, \ldots, C)$ distinct ordered categories for $C \geq 2$. A logistic mixed effects model for the cumulative response probabilities is given by (Fitzmaurice et al., 2011; Kombo et al., 2017);

$$\eta_{ij} = \alpha_c + X_{ij}'\beta + Z_{ij}' b_i, (c = 1, \ldots, C - 1). \tag{3.34}$$

Now using a logit link, we will have a cummulative logit model conditional on random effects defined as,

$$\log \left\{ \frac{P(Y_{ij} \leq c|b_i)}{P(Y_{ij} > c|b_i)} \right\} = \eta_{ij} = \alpha_c + X_{ij}'\beta + Z_{ij}' b_i \tag{3.35}$$

The random effects are assumed to have a bivariate normal distribution with zero mean and 2×2 covariance matric, G. This is the proportional odds regression

model with randomly varying intercepts and slopes (Fitzmaurice et al., 2011). In equation (3.35), $P(Y_{ij} \leq c|b_i)$ is the probability of being at or below category c, given the random effects and a set of predictors. For C categories of the ordinal response Y_i, α_c is the intercept terms that depend on the j^{th} response category and are needed to compute predicted probabilities, but usually are not of substantive interest by themselves. The regression parameters, β, reflect the association between the predictor variables and the outcome for each of the $C-1$ cumulative logits, implying that $X'_{ij}\beta$ is independent of c (Kombo et al., 2017).

Violation of the assumption of identical log-odds could lead to the formulation of an incorrect or misspecified model (Ananth & Kleinbaum, 1997). To relax the proportional odds assumption, the model could be extended to partial-proportional odds and constrained partial proportional odds models (Ananth & Kleinbaum, 1997; Hedeker, 2008). In these models, covariates are allowed to have differential effects on the $C-1$ cumulative logits (Hedeker, 2008).

3.3.3 Partial Proportional Odds Model

Partial proportional odds model is recommended where there is violation of the proportional odds assumption. In this case covariates are allowed to have differential effects on the $C-1$ cumulative logits (Hedeker, 2008). The random effects model for the $C-1$ cumulative logits can be written as

$$\eta_{ij} = \alpha_c + (X^*_{ij})'\beta_c + X'_{ij}\beta + Z'_{ij}b_i, (c = 1, \ldots, C-1). \tag{3.36}$$

where X^*_{ij} is a $p \times 1$ vector containing the values of observation ij on the set of p covariates for which proportional odds is not assumed. In this model, β_c is a $p \times 1$ vector of regression coefficients associated with these p covariates. Because β_c carries the c subscript, the effects of these p covariates are allowed to vary across the $C-1$ cumulative logits (Hedeker, 2008).

3.3.4 Estimation and inference for random effects models

According to Verbeke & Molenberghs (2009), it follows from equation 3.29 that, conditional on the random effect b_i, Y_i is normally distributed with mean vector $X_i\beta + Z_i b_i$ and with covariance matrix Σ_i. Further, b_i is assumed to be normally distributed with mean vector 0 and covariance matrix G. Let $f(Y_i, b_i)$ and $f(b_i)$ be the corresponding density functions. The marginal density function of Y_i is then given by (Verbeke & Molenberghs, 2009; Fitzmaurice et al., 2011)

$$f(Y_i, b_i) = f(Y_i|b_i)f(b_i) \tag{3.37}$$

where

$$f(Y_i|b_i) = f(Y_{i1}|b_i)f(Y_{i2}|b_i)\dots f(Y_{in_i}|b_i) \tag{3.38}$$

which is assumed to be normally distributed with mean vector $X_i\beta$ and covariance matrix $V_i = Z_i G Z_i' + \Sigma_i$. Fitzmaurice et al. (2011) stated that, $f(Y_{ij}|b_i)$ is assumed to have an exponential family distribution, whereas $f(b_i)$ is assumed to have a multivariate normal distribution with zero mean and covariance matrix G i.e. $f(b_i) \sim MVN(0, G)$. Let ϕ denote the vector of all variance and covariance parameters (usually called variance components) found in $V_i = X_i G X_i'$. The MLE for GLMM is obtained by maximizing the marginal model obtained by integrating out the random effects (Diggle et al., 2002; Fitzmaurice et al., 2011).

$$\begin{aligned} L(\beta, G, \phi) &= \prod_{i=1}^{N} f\left(y_i|b_i, G, \phi\right) \\ &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \left(f(y_{ij}|b_i, G, \phi) f(b_i|G)\, db_i \right. \end{aligned} \tag{3.39}$$

This is just the marginal distribution of Y obtained by integrating the joint distribution of Y and b_i with respect to b_i (Diggle et al., 2002). The MLE estimate of

β with respect to θ is given as;

$$\hat{\beta} = (X_i' V_i^{-1} X_i)^{-1} X' V^{-1} y_i \tag{3.40}$$

Given the ML estimate of β, ϕ and G, the random effects b_i for any particular subject can be predicted as follows

$$\hat{b_i} = E(b_i|Y_i; \hat{\beta}, \hat{\phi}, \hat{G}) = \int b_i h(b_i|Y_i, X_i; \hat{\phi}) db_i \tag{3.41}$$

That is, the predicted random effects for the i^{th} subject are simply estimated as the conditional mean of b_i given Y_i (and $\hat{\beta}$, $\hat{\phi}$,$\hat{G}$) (Fitzmaurice et al., 2011). Here, $h(.)$ is the empirical posterior distribution of b_i (Grilli & Rampichini, 2012).

3.3.5 Polytomous/multinomial regression models

The polytomous logistic model is an extension of the logistic model for binary responses to accommodate multinomial responses which does not have any restrictions on the ordinality of the response (Ananth & Kleinbaum, 1997). Let Y_{ij} denote a nominal response variable for the ith subject and jth measurement occasion. Adding random effects b_i to the fixed-effects multinomial logistic regression model, we get that the probability that $Y_{ij} = c$ (a response occurs in category c) for a given level-2 unit i is given by (Hedeker, 2008)

$$P_{ijc} = Pr(Y_{ij} = c|X_{ij}, b_i) = \frac{\exp(\eta_{ijc})}{1 + \sum_{c=2}^{C} \exp(\eta_{ijc})} \quad \text{for} \quad c = 2, 3, \ldots, C \tag{3.42}$$

$$P_{ij1} = Pr(Y_{ij} = 1|X_{ij}, b_i) = \frac{1}{1 + \sum_{c=2}^{C} \exp(\eta_{ijc})} \tag{3.43}$$

A nominal model to allow for any possible set of $C-1$ response categories is written as

$$P_{ijc} = \frac{\exp(\eta_{ijc})}{\sum_{c=1}^{C} \exp(\eta_{ijc})} \quad \textbf{for} \quad c = 1, 2, \ldots, C \tag{3.44}$$

where the multinomial logit linear predictor $\eta_{ijc} = X'_{ijc}\beta_c + Z'_{ijc}b_i$. Comparing this to the logit for ordered responses, we see that all of the effects β_c vary across categories $(c = 1, 2, \ldots, C)$. An important distinction between the model for ordinal and nominal responses is that the former uses cumulative comparisons of the categories, whereas the latter uses comparisons to a reference category (Hedeker, 2008). Parameter estimation follows the procedure described for ordinal outcomes. Let Y_i denote the vector of nominal responses from level-2 unit i (for the n_j level-1 units nested within). The probability of any Y_i conditional on the random effects b_i is equal to the product of the probabilities of the level-1 responses

$$\ell(Y_i|b_i) = \prod_{j=1}^{n_i} \prod_{c=1}^{C} (P_{ijc})^{Y_{ijc}} \tag{3.45}$$

where $Y_{ijc} = 1$ if $Y_{ij} = c$, and 0 otherwise. The marginal density of the response vector Y_i is given by

$$h(Y_i) = \int_{b_i} \ell(Y_i|b_i) g(b_i) d(b_i) \tag{3.46}$$

3.3.6 Correlation induced by random effects variability

As the variance σ_u^2 of the random effects increases, the correlation $Corr(Y_{ij}, Y_{it})$ between two observations within the same cluster also tends to increase. This correlation is called *intraclass correlation (ICC)* (Agresti, 2010; Fitzmaurice et al., 2011). Consider the cumulative logit link model with a random intercept, b_i. The latent outcome for observation j in cluster i is

$$y^*_{ij} = \alpha + \beta' X_{ij} + b_i + \epsilon_{ij} \tag{3.47}$$

Suppose the random effects are independent $N(0, \sigma_b^2)$ and the errors ϵ_{ij} are also independent of the random effects and have a variance σ^2. This model assumes that observations within a cluster are exchangeable and the positive correlation among observations is accounted for by sharing a common random effect (Fitzmaurice

et al., 2011). This correlation is expressed as,

$$Corr(Y_{ij}^*, Y_{it}^*) = \rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} \tag{3.48}$$

This is the proportion of the total residual variance that is due to the variability σ_b^2 in the random effect. The correlation is positive and increases as σ_b^2 increases, for fixed σ^2 (Agresti, 2010). When the logistic model is used, the residual at level one are assumed to follow the standard logistic distribution with zero mean and variance $\pi^2/3 = 3.29$ (Hedeker, 2008; Agresti, 2010; Grilli & Rampichini, 2012; Arfan & Sherwani, 2017). Although the above model appears to be the same as an ordinary multilevel regression model for continuous outcomes, it is one in which the error variance is fixed and not estimated (Hedeker, 2008). The within groups variation for dichotomous and ordinal outcomes is defined as

$$Corr(Y_{ij}^*, Y_{it}^*) = \rho = \frac{\sigma_b^2}{\sigma_b^2 + \pi^2/3} \tag{3.49}$$

where σ_b^2 is the variance between clusters (level 2) and $\pi^2/3$ is the variance of standard logistic regression (level 1) for observations within clusters (Arfan & Sherwani, 2017). Other correlation structures relevant for longitudinal data can also be incorporated depending on the data structure.

3.4 Analysis of missing data

3.4.1 The missing data model

Longitudinal studies are an important source of information in health sciences and other areas but often have the problem of missing data (Rubin, 1976; Kombo et al., 2017). Missing values in longitudinal studies occur when not all of the planned measurements of a subject outcome vector are actually observed, turning the statistical analysis into the missing data problem (Kombo et al., 2017). In such situations it is advised to avoid simple *ad hoc* methods such as complete case analysis particularly given it is near impossible to justify that data are missing

completely at random. The process that causes missing data are referred to missing data mechanisms. There are three missing data mechanisms i.e. missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). To illustrate differences between these mechanisms, let Y be the complete set of measurements which would have been obtained if no missing values. Y can be partitioned into Y^o denoting the observed values and Y^m denoting the unobserved or missing values such that, $Y = (Y^o, Y^m)$. We also let R denote a set of indicator random variables denoting which elements of Y are observed Y^o or missing Y^m. For observation unit i at the j^{th} measurement, R is given as

$$R_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is observed} \\ 0 & \text{if otherwise} \end{cases} \tag{3.50}$$

As Diggle et al. (2002) indicated, a probability model for the missing value mechanism defines the probability distribution of R conditional on $Y = (Y^o, Y^m)$ where $R_i = (R_{i1}, R_{i2}, \ldots, R_{in})$. Data are therefore MCAR if R is independent of both Y^o and Y^m which is also referred as the non-response process (Reddy, 2017), MAR if R is independent of Y^m i.e. missingness depend on the observed data Y^o only and MNAR if R depend on Y^m which is a non-random process. Molenberghs & Kenward (2007) stated that, the joint distribution of the full data Y and the indicator vector variable R can be factorized based on;

$$f(y_i, r_i|X_i, W_i, \theta, \psi) = f(y_i|X_i, \theta) f(r_i|y_i, W_i, \psi) \tag{3.51}$$

where X_i and W_i denote design matrices for the measurements and missing mechanism while ψ and θ denotes the corresponding parameter vectors, respectively (Molenberghs & Kenward, 2007; Kombo et al., 2017). The first factor in equation 3.51 is the marginal density of the measurement process and the second one is the density of the missingness process, conditional on the outcomes (Molenberghs & Kenward, 2007). The conditional distribution of the missing data can be expressed as $f(r_i|y_i, W_i, \psi)$. The MAR assumption states, conditional on the

observed outcomes;

$$f(r_i|y_i, W_i, \psi) = f(r_i|y_i^0, W_i, \psi) \tag{3.52}$$

Thus the joint distribution of the observed data can be partitioned as,

$$f(y_i, r_i|X_i, W_i, \theta, \psi) = f(y_i|X_i, \theta)f(r_i|y_i^0, W_i, \psi) \tag{3.53}$$

and hence at the observed data level (where R does not depend on Y^m),

$$f(y_i^0, r_i|X_i, W_i, \theta, \psi) = f(y_i^0|X_i, \theta)f(r_i|y_i^0, W_i, \psi) \tag{3.54}$$

Under MAR, equation 3.54 implies that the missingness mechanism is ignorable if the parameters spaces for θ and ϕ are orthogonal.

3.4.2 Methods for handling missing data

Different methods and techniques such as direct likelihood and Bayesian analyses, expectation maximization (EM) algorithm, multiple imputation (MI) and weighted generalized estimating equations are used to handle missing data in longitudinal studies (Molenberghs & Kenward, 2007). MI is the most popular approach (Molenberghs & Kenward, 2007; White et al., 2010; Bartlett et al., 2015; Kombo et al., 2017) which provides a way to capture the uncertainty associated with imputations and can also be used for both continuous, binary and categorical variables (Kombo et al., 2017). Standard MI procedures assume that data are MAR (White et al., 2010; Kombo et al., 2017), an assumption that will also be made in this study. Commonly, the observed values are used as a basis to impute values for the missing observations (Molenberghs & Kenward, 2007; White et al., 2010). White et al. (2010) further stated that, when correctly implemented, MI produces asymptotically unbiased estimates and standard errors and is asymptotically efficient.

MI involves three distinct stages. Firstly, missing values are filled with $m \geq 2$

plausible values to generate m complete datasets from the posterior predictive distribution of the missing data conditional on the observed data. The resulting m imputed datasets are analyzed separately using standard methods for complete data analysis then the results from m analyses are combined into one using the Rubin's rules (White et al., 2010; Bartlett et al., 2015) for inference, where the standard errors of the estimates take account of the variation within and between the m imputations (Molenberghs & Kenward, 2007; White et al., 2010; Azur et al., 2011; Bartlett et al., 2015; Kombo et al., 2017).

Multivariate normal imputation (MVI) and multiple imputation by chain equations (MICE), also referred as fully conditional specification (FCS) or sequential regression multivariate imputation are two of the MI approaches used to impute missing values. The focus of this thesis is on the application of FCS which is a more flexible method that specifies the multivariate model by a series of conditional models for each of the incomplete variables as opposed to MVI which relies on the multivariate normal assumption even for binary and categorical variables (White et al., 2010; Liu & De, 2015a; Kombo et al., 2017).

An important feature of FCS is its ability to handle different variable types (i.e. continuous, binary, unordered categorical and ordered categorical) because each variable is imputed using its own imputation model (according to its distribution), for example, binary variables modeled using logistic regression and continuous variables modeled using linear regression (White et al., 2010; Azur et al., 2011; Liu & De, 2015a; Bartlett et al., 2015). In the FCS procedure a series of regression models are run whereby each variable with missing data is modeled conditional upon the other variables in the data (Azur et al., 2011). To minimize bias, it is widely accepted that the same variables in the model should also be in the imputation model, including any interactions that will be assessed in the analysis model whether or not they are associated with missingness or not (White et al., 2010; Azur

et al., 2011; Kombo et al., 2017). In other words the imputation model aught to be as rich as possible to capture the missingness as accurately as possible.

3.4.3 The FCS algorithm

For the illustration of FCS algorithm, we let Y denote the fully observed outcome in this study i.e., preterm birth, X denote the partially observed covariates $X = X_1, \ldots, X_p$, and W denote the fully observed covariates $W = W_1, \ldots, W_q$. Let X^o and X^m denote the vectors of observed and missing values of X for n subjects. For each partially observed covariate X_j, we posit an imputation model $f(X_j|X_{-j}, W, Y, \theta_j)$ with parameter θ_j where $X_{-j} = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p)$ (Bartlett et al., 2015). This according to (Bartlett et al., 2015) is typically a generalized linear model chosen according to the type of X_j (e.g. continuous, binary, multinomial, and ordinal). Furthermore, a noninformative prior distribution $f(\theta_j)$ for θ_j is specified. We further let x_j^o and x_j^m denote the vectors of observed and missing values in X_j for the n subjects and y and w denote the vector and matrix of fully observed values of Y and W across n subjects.

Let $x^{m(t)}$ denote imputations of the missing values x_j^m at iteration t and $x_j^{(t)} = (x_j^o, x_j^{m(t)})$ denote vectors of observed and imputed values at iteration t. Let $x_{-j}^{(t)} = (x_1^{(t)}, \ldots, x_{j-1}^{(t)}, x_{j+1}^{(t-1)}, \ldots, x_p^{(t-1)})$. The tth iteration of the algorithm consists of drawing from the following distributions (up to constants of proportionality)

(Bartlett et al., 2015);

$$\left.\begin{array}{rcl} \theta_1^{(t)} & \sim & f(\theta_1) f(x_1^o | x_{-1}^{(t)}, w, y, \theta_1) \\ x_1^{m(t)} & \sim & f(x_1^m | x_{-1}^{(t)}, w, y, \theta_1^{(t)}) \\ \theta_2^{(t)} & \sim & f(\theta_2) f(x_2^o | x_{-2}^{(t)}, w, y, \theta_2) \\ x_2^{m(t)} & \sim & f(x_2^m | x_{-2}^{(t)}, w, y, \theta_2^{(t)}) \\ \vdots & & \\ \theta_p^{(t)} & \sim & f(\theta_p) f(x_p^o | x_{-p}^{(t)}, w, y, \theta_p) \\ x_p^{m(t)} & \sim & f(x_p^m | x_{-p}^{(t)}, w, y, \theta_p^{(t)}) \end{array}\right\} \tag{3.55}$$

The FCS starts by calculating the posterior distribution $p(\theta|x^o)$ of θ given the observed data. This is followed by drawing a value of θ^* from $p(\theta|x^0)$ given $(x^o, x_{-j}^{(t)}, w, y)$, which is the product of the prior $f(\theta_j)$ and the likelihood corresponding to fitting the imputation model for X_j to subjects for whom X_j is observed, using the observed and most recently imputed values of X_{-j} (Bartlett et al., 2015). Missing values in X_j are then imputed from the imputation model using the parameter value drawn in the preceding step (Bartlett et al., 2015). Finally, a value x^* is drawn from the conditional posterior distribution of x^m given $\theta = \theta^*$. The process is then repeated depending on the desired number of imputations (Van Buuren et al., 2006b; Azur et al., 2011; Bartlett et al., 2015; Kombo et al., 2017). Within each imputation, there is an iterative estimation process until the distribution of the parameters governing the imputations have converged in the sense of becoming stable, although more cycles may be required depending on certain conditions such as the amount of missing observations in the data (Azur et al., 2011; Bartlett et al., 2015). Rubin's rule is then used to provide the final inference for $\hat{\theta}$ by averaging the estimates across M imputations given by (Bartlett et al., 2015);

$$\hat{\theta}_M = \frac{\sum_{m=1}^{M} \hat{\theta}^m}{M} \tag{3.56}$$

while the estimate of the variance of $\hat{\theta}^M$ is given by;

$$\widehat{Var}(\hat{\theta}_M) = \left[\frac{1}{M}\sum_{m=1}^{M}\widehat{Var}(\hat{\theta}^m)\right] + \left[(1+1/M)\frac{1}{M-1}\sum_{m=1}^{M}(\hat{\theta}^m - \hat{\theta}_M)^2\right] \tag{3.57}$$

which is a combination of within and between imputation variances. Detailed descriptions on implementation of the FCS/MICE algorithm in STATA is well-presented elsewhere (StataCorp, 2017; Royston et al., 2011).

Kombo et al. (2017) has well described that, using a Bayesian approach, imputations are done stepwise starting with the variable with the least amount of missing values and progressing like that until the variable with the most missing data is finally handled. He further stated that, the process involves two phases in each imputation: the fill-in stage and the imputation stage. During every stage, draws are randomly done from both the posterior distribution of the parameters and posterior distribution of the missing values. At the fill-in stage, the missing values are filled in sequentially over the variables, one after the other with preceding variables serving as covariates or independent variables and the variable being imputed is the dependent variable (Azur et al., 2011; Kombo et al., 2017). The filled-in values are then used as starting values for the imputation stage. At the imputation stage, the filled-in values are replaced with imputed values for each variable sequentially at each iteration (Kombo et al., 2017).

Different software packages such as R, STATA and SAS are equipped with relevant tools to both implement and assess convergence of the imputation model (Royston et al., 2011; Van Buuren et al., 2006a; Liu & De, 2015a). Kernel density estimate plots are used to visually compare the distributions of the observed, imputed and completed values of each variable (Liu & De, 2015a). Imputation diagnostics are important in identifying potentially problematic variables. These diagnoses were a basis for increasing the number of imputation to 20 and 500 iterations as described in Chapter 4 and 6.

Chapter 4

Predictors of perinatal death in the presence of missing data: a birth registry-based study in northern Tanzania

4.1 Introduction

Perinatal death refers to the number of stillbirths (pregnancy loss that occurs after seven months of gestation and before birth) and early neonatal deaths (deaths of live births within the first seven days of life) (MoHCDGEC [Tanzania Mainland] et al., 2016; World Health Organization, 2019). Perinatal and maternal health are closely linked; hence perinatal mortality is used as an essential indicator to monitor maternal health status and quality of antenatal, intrapartum, and newborn care (Mmbaga et al., 2012b; Mpembeni et al., 2014; World Health Organization, 2016a, 2019). Globally, more than five million perinatal deaths occur each year (World Health Organization, 2016a). Children face the highest risk of dying in their first month of life at a global rate of 17 deaths per 1,000 births (UNICEF et al., 2020). Globally, 2.4 million children died in the first month of life in 2019 – 6,700 deaths

every day (UNICEF et al., 2020). The patterns of these deaths are similar to the patterns for maternal deaths, the majority occurring in developing countries (World Health Organization, 2019). In Tanzania, between the years 2004-2005 and 2015-16, the under-five mortality rate was reported to have declined from 112 to 67 deaths per 1,000 births. The country has, however, witnessed an increase in the number of stillbirths (from 143 to 187), the number of early neonatal deaths (from 156 to 214) as well as perinatal mortality rate (from 36 to 39) deaths per 1,000 births, respectively (MoHCDGEC [Tanzania Mainland] et al., 2016).

The risk factors for stillbirths and early neonatal deaths are closely linked, and examining just one or the other is reported to bias the true level of mortality around delivery (MoHCDGEC [Tanzania Mainland] et al., 2016; Ouyang et al., 2013). The risk of perinatal mortality has been associated with preterm birth, shorter birth interval (<24 months), congenital anomalies, previous history of early neonatal death, low birth weight, maternal anemia, placental abruption, ruptured uterus, systemic infections/sepsis, pre-eclampsia, eclampsia, obstetric hemorrhage, having a home delivery, fetal growth restrictions and maternal infections such as syphilis and malaria (Bayou & Berhan, 2012; Getiye & Fantahun, 2017; Mmbaga et al., 2012a; Mpembeni et al., 2014; Nankabirwa et al., 2011; Unterscheider et al., 2014; Vogel et al., 2014; Liu et al., 2016). As with other adverse maternal outcomes, perinatal deaths recur in subsequent pregnancies (Mahande et al., 2013a; Ouyang et al., 2013; Salihu et al., 2011). Although these factors may be common across low, middle- and high-income countries, they are likely to differ depending on the context or country-specific conditions such as availability of quality obstetric and newborn care services at different levels of care.

Despite challenges in the coverage and content of antenatal care (Benova et al., 2018), the WHO recommends a minimum of eight contacts for antenatal care (ANC) that can reduce perinatal deaths by up to eight per 1000 births when

compared to a minimum of four visits (World Health Organization, 2016b). Early identification and management of women with complications have also been recommended to improve maternal and perinatal outcomes (Vogel et al., 2014). Informed interventions are therefore crucial to accelerate progress towards achieving the second indicator of the third sustainable development goal, i.e., by 2030, end preventable deaths of newborns and children under five years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births (UNDP, 2018). These interventions should consider context-specific factors that can better explain the risk of perinatal deaths. As perinatal mortality rate is increasing in Tanzania, this study was aimed to determine current trends and associated factors from a maternally linked medical birth registry at KCMC referral hospital in northern Tanzania, comparing results before and after imputation of missing values.

Missing data is a common problem that occurs in almost all medical and epidemiological research (Azur et al., 2011; Van Buuren et al., 2006a; Liu & De, 2015a; Kombo et al., 2017; Pedersen et al., 2017). Hospital-based longitudinal studies are also facing the same problem. Individuals with missing data may differ from those with no missing data in terms of the outcome of interest and prognosis in general (Pedersen et al., 2017). Previous studies assessing predictors of perinatal death have adopted simple methods such as complete case analysis or available case analysis hence ignoring important information about missing data. Ignoring missing data in statistical analysis often produces biased and inefficient estimates of association (Liu & De, 2015a; Pedersen et al., 2017), especially when data are missing at random (Sterne et al., 2009). This study aimed to determine predictors of perinatal death accounting for missing values.

4.2 Methods

4.2.1 Study design and participants

This study utilized data from the KCMC Medical Birth Registry, which contains maternally linked cohort data since the year 2000. KCMC is a national zonal referral hospital located in Moshi Municipality, Kilimanjaro Region, Northern Tanzania. The study population was women who delivered singleton babies. The study considered all deliveries recorded from January 2000 to December 2015, a total of 55,003 deliveries from 43,084 mothers aged 15-49 years. We excluded 3,316 multiples gestations to avoid over-representation of high-risk pregnancies (Chuwa et al., 2017). We further excluded 49 records missing hospital numbers (i.e., unique identification number used to link mothers and their subsequent births) and 791 observations with a mismatch between dates of births of children from the same mother or were of unknown sequence (i.e., whether was a singleton or multiple births). We, therefore, analyzed data for 50,847 recorded deliveries born from 41,498 mothers (Figure 4.1).

4.2.2 Data collection methods

A detailed description of the data collection procedure and data collected for the birth registry have been previously published (Mmbaga et al., 2012a,b; Mahande, 2015). Briefly, birth data at KCMC have been recorded using a standardized questionnaire. Specially trained project midwives did data collection, and mothers interviewed within the first 24 hours after birth given a normal delivery or on the second or third day in case of complicated deliveries depending on their condition, following informed consent. Although the printed questionnaires are in the English language, the Project Midwives performing the interviews are well versed in English, Swahili, and one other tribal language. Patient files and antenatal care cards were used for verification of interview data and to extract additional information. Information of neonates admitted in the neonatal intensive care unit

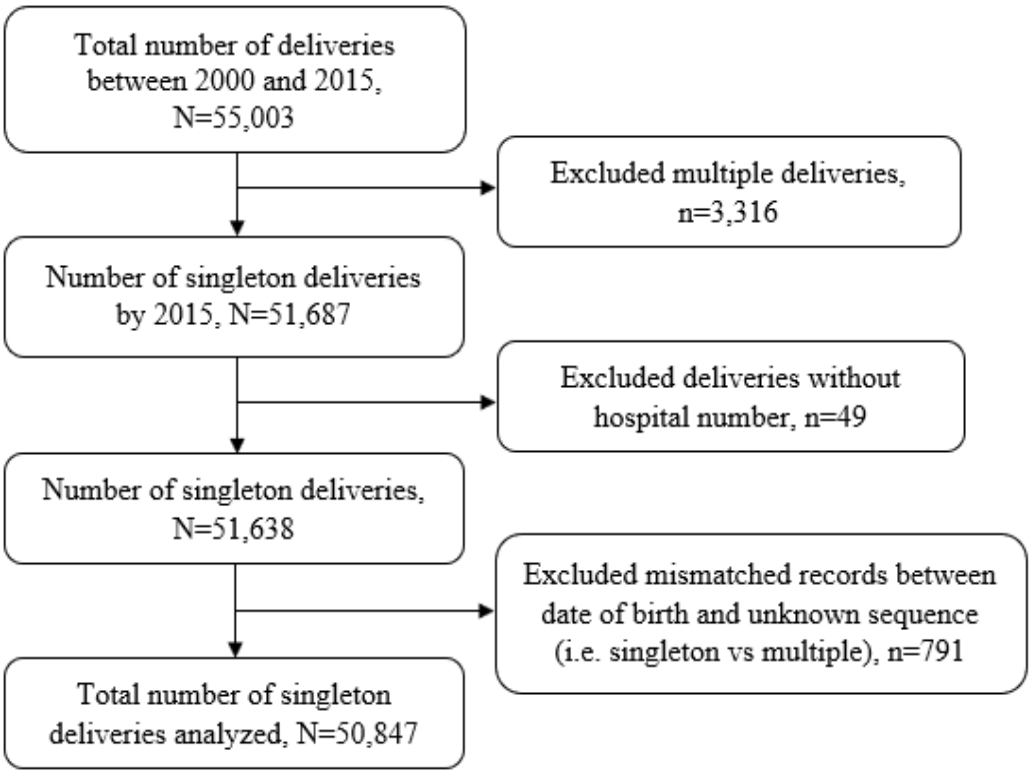


Figure 4.1: Schematic diagram showing the number of participants, KCMC medical birth registry, 2000-2015.

were recorded separately in the neonatal registry form and later linked with mother's information during entry in the birth registry using unique identification numbers. Furthermore, a unique identification number was assigned to each woman at first admission and used to trace her medical records at later admissions.

4.2.3 Study variables and variable measurements

The response variable was perinatal death, which comprises stillbirths (pregnancy loss that occurs after seven months of gestation) and early neonatal death (death of live births within the first seven days of life) (World Health Organization, 2019; MoHCDGEC [Tanzania Mainland] et al., 2016). We coded perinatal death as binary, i.e., 'yes' if death occurred during the perinatal period and 'no' if otherwise. The perinatal mortality rate is calculated as the number of perinatal deaths per 1,000 pregnancies of seven or more months' duration (MoHCDGEC [Tanzania Mainland]

et al., 2016).

Independent variables included maternal background characteristics such as age categories (15-19, 20-34, 35-39 and 40+) in years, area of residence (rural vs. urban), highest education level (none, primary, secondary and higher), marital status (single, married and widow/divorced), occupation (unemployed, employed and others), whether referred for delivery or not and the number of antenatal care visits (<4 and ≥ 4 visits). Maternal health before and during pregnancy; maternal body mass index (BMI) categorized as underweight (<18.5 Kg/m^2), normal weight (18.5–24.9 Kg/m^2), overweight (25–29.9 Kg/m^2) and obese (≥ 30 Kg/m^2). Alcohol consumption during pregnancy, maternal anemia, malaria, systemic infections/sepsis, and pre-eclampsia/eclampsia categorized as yes or no, with yes indicating the occurrence of these outcomes. Maternal HIV status was categorized as positive or negative. Information concerning delivery included complications during delivery, i.e., premature rupture of the membranes (PROM), postpartum hemorrhage, placenta previa, and placenta abruption categorized as yes and no. Gestational age at birth was estimated based on the date of the last menstrual period and was recorded in whole weeks. Preterm birth included any birth before 37 completed weeks of gestation (Quinn et al., 2016; World Health Organization, 2020). Newborn characteristics included sex (whether male or female), and low birth weight, defined as an infant birth weight of less than 2500g (Mitao et al., 2016).

4.2.4 Statistical analysis

We analyzed data using STATA version 15.1 (StataCorp LLC, College Station, Texas, USA). Numeric variables were summarized using means and standard deviations while categorical variables using frequency and percentages. We used the Chi-square test to compare the proportion of perinatal death across different levels of explanatory variables. Data analysis involved three stages; the first stage being complete case analysis, followed by imputation of missing values in both the

outcome and the covariates, and finally, analysis of the imputed dataset. Results are compared before and after imputation to assess the precision of estimates.

Two separate models were fit before and after imputation of missing values. The first was generalized estimating equations (GEE) model with binomial family, logit link, exchangeable correlation structure, and robust variance estimator (Model I). The second was GEE log-linear regression model with the Poisson family, log link, an exchangeable correlation structure, and a robust variance estimator (Model II). The GEE models, often called population average or marginal models describes changes in the population mean given changes in covariates, while accounting for within-cluster correlation of observations (Hubbard et al., 2010). In this thesis, data analysis accounted for the correlation of repeated deliveries within mothers, i.e., mothers having more than one child born at different times. These models are an extension to generalized linear models to longitudinal data, directly modelling the mean response, at each occasion using an appropriate link function (Fitzmaurice et al., 2009). The estimation procedures are not likelihood-based but uses a series of estimating equations. Odds Ratio (OR) and Relative risk (RR) with their corresponding 95% confidence intervals (CIs) were used to determine the strength of association at 5% significant level, respectively. Incomplete data were assumed to be missing at random (MAR) where the probability of data being missing does not depend on the unobserved data, conditional on the observed data (White et al., 2010; Ibrahim & Molenberghs, 2009; Pedersen et al., 2017; Jakobsen et al., 2017); hence the variables in the dataset were used to predict missingness. Stepwise regression was used for variable selection, and it included a consecutive assessment of the effect of adding or removing different variables from the models.

Furthermore, for the analysis of missing data, we assumed a nonmonotone pattern of missingness in which some subject values were observed again after a missing value occurs (Ibrahim & Molenberghs, 2009; Jakobsen et al., 2017). Multiple

imputation is a commonly used method to deal with missing data, which accounts for the uncertainty associated with missing data (Pedersen et al., 2017; Jakobsen et al., 2017; Sterne et al., 2009). Under a nonmonotone pattern of missingness, it is recommended to use the chained equations (also referred to as fully conditional specification (FCS) (Van Buuren et al., 2006a; Azur et al., 2011)) or the Markov Chain Monte Carlo (MCMC) method to impute missing values (Jakobsen et al., 2017). We, therefore, used multiple imputation by FCS to handle missing data in this study. This technique is a powerful and statistically valid method for creating imputations in large datasets, which include both categorical and continuous variables (Liu & De, 2015a; Van Buuren et al., 2006a; White et al., 2010; Azur et al., 2011). After declaring the preferred data structure using *mi set mlong* command, the *mi impute chained* function, part of *mi* package in Stata, implemented this technique, referred to in this software as Multiple Imputation by Chained Equations (MICE) approach (Royston et al., 2011; Aloisio et al., 2014; StataCorp, 2019). Before imputation, *mi register imputed* command registered variables for the imputation model. Interested readers can find more about the *mi* command by typing *help mi* in their Stata command window.

Maternal age and education level were imputed using *ologit* command for ordinal response variables, while maternal occupation, marital status, and BMI (because normal weight (18.5–24.9 Kg/m^2) was a reference category) using *mlogit* for multinomial distribution. The rest of the variables were binary, and so imputed using the *logit* command. Pre-eclampsia/eclampsia, anemia during pregnancy, malaria, sepsis/systemic infections, PROM, PPH, abruption placenta, and placenta previa did not contain any missing values, hence used as auxiliary variables in the imputation model. The imputation model generated 20 imputed datasets after 500 iterations (imputation cycles). A random seed of 5000 was specified for replication of imputation results each time this analysis was performed (Jakobsen et al., 2017; Royston et al., 2011). We repeated similar procedures when imputing data for

stratified analysis (i.e., by referral status), and determining the independent predictors of stillbirth after excluding early neonatal deaths. For the imputation model stratified by referral status, we imputed the data conditional on referral status (i.e., using "if" other than the "by" option in the Stata *mi impute chained* command options). After the imputation of missing values, we performed the analysis preceding with the *mi estimate:* command.

4.3 Results

4.3.1 Maternal background characteristics at first birth

The mean (SD) age of 41,498 mothers was 27 (6) years. More than three quarters (76.9%) were aged between 20-34 years, 55.9% resided in urban areas, 55.7% had primary education level, and 85.8% were married. The prevalence of underweight, overweight, and obese was 5.9%, 26.4%, and 11.1%, respectively. Twenty-eight percent of mothers drank alcohol during pregnancy (Table 4.1).

4.3.2 Obstetric care characteristics and complication

Malaria was the most common (13.5%) disease in this cohort. The proportion of pre-eclampsia/eclampsia was 4.1%, while that of HIV was 5.3%. About 30% of all deliveries had <4 ANC visits, 11.1% were delivered preterm (<37 weeks of gestation), 10.9% with low birth weight, and 34% delivered through Caesarean Section (CS). Almost a quarter (23.5%) were referred for delivery. More than half (51.7%) of these deliveries were males (Table 4.2).

4.3.3 Distribution of missing values in the KCMC Medical Birth registry

Table 4.3 summarizes frequencies and percentages of missing values for the variables with missing information in this study. Maternal BMI and HIV status had the highest proportions of missing values, 31.3%, and 23.7%, respectively. These two variables contributed to over half (55%) of all missing values in the dataset.

Table 4.1: Maternal background characteristics at first birth (N=41,498)

Characteristics	Frequency	Percent (95%CI)
Age groups (years)*		
15-19	4,250	10.3 (10.0, 10.6)
20-34	31,866	76.9 (76.5, 77.3)
35-39	4,201	10.1 (9.9, 10.4)
40+	1,101	2.7 (2.5, 2.8)
Area of residence*		
Rural	18,244	44.1 (43.6, 44.6)
Urban	23,137	55.9 (55.4, 56.4)
Highest education level*		
None	856	2.1 (1.9, 2.2)
Primary	23,081	55.7 (55.3, 56.2)
Secondary	4,895	11.8 (11.5, 12.1)
Higher	12,582	30.4 (29.9, 30.8)
Occupation*		
Unemployed	9,386	22.8 (22.4, 23.2)
Employed	28,973	70.3 (23.8, 24.6)
Others	2,865	6.9 (6.7, 7.2)
Marital Status*		
Single	5,774	14.0 (13.6, 14.3)
Married	35,468	85.8 (85.5, 86.1)
Widowed/Divorced	99	0.2 (0.2, 0.3)
Body mass index categories*		
Underweight (<18.5)	1,671	5.9 (5.6, 6.1)
Normal weight (18.5-24.9)	16,164	56.6 (56.0, 57.2)
Overweight (25-29.9)	7,542	26.4 (25.9, 26.9)
Obese (≥30)	3,184	11.1 (10.8, 11.5)
Drink alcohol during this pregnancy *		
Yes	11,490	28.0 (27.6, 28.5)
No	29,513	71.9 (71.5, 72.4)

*Frequencies do not tally to the total due to missing values in these variables

The proportion of missing values was about 8% in the gestational age at birth variable and 3.7% on referral status. The perinatal status (primary outcome) contributed only 0.2% of missing values in this dataset.

Table 4.2: Diseases and complications during pregnancy and delivery (N=50,847)

Characteristics	Frequency	Percent (95%CI)
Pre-eclampsia/eclampsia		
No	48,779	95.9 (95.8, 96.1)
Yes	2,068	4.1 (3.9, 4.2)
Anaemia		
No	50,054	98.4 (98.3, 98.5)
Yes	793	1.6 (1.5, 1.7)
Malaria		
No	43,961	86.5 (86.2, 86.8)
Yes	6,886	13.5 (13.2, 13.8)
Infections		
No	49,982	98.3 (98.2, 98.4)
Yes	865	1.7 (1.6, 1.8)
HIV Status*		
Negative	36,726	94.7 (94.5, 94.9)
Positive	2,064	5.3 (5.1, 5.5)
Number of ANC visits*		
≥ 4	33,905	67.9 (67.5, 68.3)
<4	16,006	32.1 (31.7, 32.5)
PROM		
No	49,770	97.9 (97.8, 98.0)
Yes	1,077	2.1 (2.0, 2.2)
PPH		
No	50,572	99.5 (99.4, 99.5)
Yes	275	0.5 (0.5, 0.6)
Abruption placenta		
No	50,676	99.7 (99.6, 99.7)
Yes	171	0.3 (0.3, 0.4)
Placenta previa		
No	50,740	99.8 (99.7, 99.9)
Yes	107	0.2 (0.2, 0.3)
Gestational age (weeks) *		
Term birth (≥ 37)	41,646	88.9 (88.6, 89.2)
Preterm birth (<37)	5,184	11.1 (10.8, 11.4)
Delivery mode*		
Vaginal	33,526	66.1 (65.7, 66.5)
CS	17,179	33.9 (33.5, 34.3)
Referred for delivery*		
Yes	11,488	23.5 (23.1, 23.8)
No	37,479	76.5 (76.2, 76.9)
Birth weight*		
NBW	45,269	89.3 (89.0, 89.5)
LBW	5,445	10.7 (10.5, 11.0)
Sex of the baby*		
Male	26,159	51.7 (51.2, 52.1)
Female	24,461	48.3 (47.9, 48.8)

*Frequencies do not tally to the total due to missing values in these variables

Table 4.3: Distribution of missing values in the KCMC Medical Birth Registry, 2000-2015 (N=50,847)

Variables with missing data	Frequency	Percent Missing
Body Mass Index (BMI)	15,911	31.3
HIV status	12,057	23.7
Gestational age at birth categories	4,017	7.9
Referral status	1,880	3.7
Antenatal Care visits	936	1.8
Alcohol use during pregnancy	636	1.3
Occupation	308	0.6
Sex of the child	227	0.5
Marital Status	186	0.4
Mode of delivery	142	0.3
Birth weight of the child	133	0.3
Perinatal status (primary outcome)	123	0.2
Area of residence	120	0.2
Education level	100	0.2
Age categories	85	0.2

4.3.4 Perinatal status by maternal characteristics during pregnancy and delivery

Among 50,724 deliveries with complete records on perinatal status in the KCMC medical birth registry between 2000-2015, 4.2% (95%CI 4.0%, 4.3%) ended in perinatal death (equivalent to a perinatal mortality rate (PMR) of 41.6 (95%CI 39.9, 43.3) deaths per 1,000 births). After the imputation of missing values, the proportion of perinatal death remained relatively the same. The proportion of perinatal death was significantly different ($p<0.05$) across maternal characteristics during pregnancy and delivery except for malaria, placenta previa, and sex of the child. Among deliveries from pre-eclamptic/eclamptic mothers, 13.6% ended up in a perinatal death. About 5% of perinatal deaths were from HIV positive mothers, 3.7% among those who drank alcohol during pregnancy, and 6.4% among those with <4 ANC visits. Deliveries from mothers who experienced postpartum hemorrhage (PPH) had high (22.3%) prevalence of perinatal death compared to those who were not (4.1%), $p<0.001$. Likewise, the proportion was 58.8%, 15.6%, 19.1%, and 8.4% among deliveries from mothers who experienced abruption of

placenta, preterm birth, low birth weight (LBW), and those referred for delivery, respectively (Table 4.4). In addition, the proportion of perinatal death among 41,396 singleton deliveries was 4.6% and was significantly different ($p<0.001$) from 2.3% among 9,328 multiple births (results not shown in the table).

4.3.5 Trends of perinatal death from 2000 to 2015 in northern Tanzania

Between 2000 and 2015, perinatal deaths have been declining slightly in this cohort (Figure 4.2). The proportion of perinatal death decreased significantly by 12.4% (95%CI 8-16.9%, $p<0.001$) for every one-year increase. The proportion of stillbirths is higher than that of early neonatal deaths over the years (Figure 4.3). However, between the years 2013-2015, the proportion of early neonatal deaths increased while that of stillbirth decreased, from 0.2% and 3.7% in 2013 to 1.8% and 2.2% in 2015, respectively. The proportion of stillbirth decreased significantly by 6.2% (95%CI 1.7-10.6%, p=0.0.01) while that of early neonatal death decreased by 12.4% (95%CI 0.4-12.4%, p=0.04) for every one-year increase.

4.3.6 Predictors of perinatal death

Adjusted analysis for the predictors of perinatal death before and after imputation is in Table 4.5. Ignoring missing values in the analysis of pregnancy-related outcomes, especially in registry-based studies, produces biased parameter estimates. The coefficients resulting from complete case analysis are observed to be either higher or lower than should have been if there were no missing values. At the same time, standard errors are all relatively larger. For instance, considering Model II, there was a reduced risk of perinatal death among deliveries of adolescent mothers from (RR=0.582, 95%CI 0.478, 0.708) to (RR=0.674, 95%CI 0.575, 0.789) compared to those aged 20-34 years before and after imputation, respectively. Also, the risk of perinatal death was observed to have reduced from (RR=1.432 1.243, 1.648) to (RR=1.423, 1.263, 1.603) among pre-eclamptic/eclamptic mothers and increased from (RR=2.072, 95%CI 1.847, 2.324) to (RR=2.111, 1.906, 2.338)

Table 4.4: Perinatal status by maternal characteristics during pregnancy and delivery (N=50,847)

Characteristics	Perinatal status		p-value
	Alive	Died	
Pre-eclampsia/eclampsia			<0.001
No	46,836 (96.2)	1,827 (3.8)	
Yes	1,780 (86.4)	281 (13.6)	
Anaemia			0.001
No	47,875 (95.9)	2,057 (4.1)	
Yes	741 (93.6)	51 (6.4)	
Malaria			0.45
No	42,045 (95.9)	1,811 (4.1)	
Yes	6,571 (95.7)	297 (4.3)	
Infections			0.17
No	47,779 (95.8)	2,080 (4.2)	
Yes	837 (96.8)	28 (3.2)	
HIV Status			0.001
Negative	35,369 (96.5)	1,296 (3.5)	
Positive	1,959 (95.1)	101 (4.9)	
Drank alcohol during this pregnancy			0.01
Yes	13,593 (96.3)	525 (3.7)	
No	34,436 (95.7)	1,538 (4.3)	
Number of ANC visits			<0.001
≥4	32,840 (97.1)	976 (2.9)	
<4	14,947 (93.6)	1,029 (6.4)	
PROM			0.004
No	47,569 (95.8)	2,082 (4.2)	
Yes	1,047 (97.6)	26 (2.4)	
PPH			<0.001
No	48,403 (95.9)	2,047 (4.1)	
Yes	213 (77.7)	61 (22.3)	
Abruption placenta			<0.001
No	48,546 (96.0)	2,008 (4.0)	
Yes	70 (41.2)	100 (58.8)	
Placenta previa			0.21
No	48,517 (95.8)	2,101 (4.2)	
Yes	99 (93.4)	7 (6.6)	
Gestational age (weeks)			<0.001
Term birth (≥ 37)	40,483 (97.4)	1,064 (2.6)	
Preterm birth (<37)	4,362 (84.4)	805 (15.6)	
Delivery mode			0.32
Vaginal	32,078 (95.8)	1,396 (4.2)	
CS	16,435 (96.0)	682 (4.0)	
Referred for delivery			<0.001
Yes	10,487 (91.6)	967 (8.4)	
No	36,365 (97.2)	1,033 (2.8)	
Birth weight			<0.001
NBW	44,132 (97.7)	1,037 (2.3)	
LBW	43,92 (80.9)	1,035 (19.1)	
Sex of the baby			0.88
Male	25,018 (95.9)	1,082 (4.1)	
Female	23,408 (95.9)	1,006 (4.1)	
Total	**48,616 (95.8)**	**2108 (4.2)**	

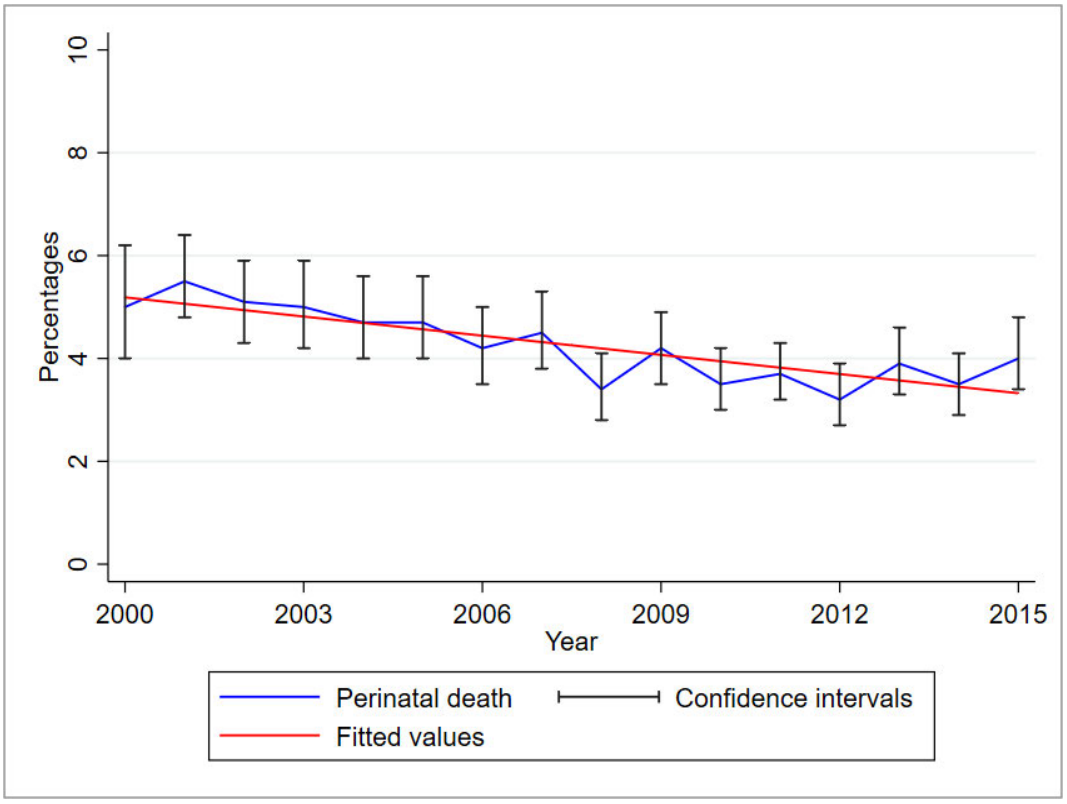


Figure 4.2: Trends of perinatal deaths, KCMC Medical Birth Registry data, 2000-2015 (N=50,724).

among preterm deliveries, before and after imputation, respectively. Consumption of alcohol during pregnancy lowered the risk of perinatal death except in Model II after imputation. While the direction of the association for both Models I and II is almost the same; results from Model II provide more precise estimates of the predictors of perinatal death as it has both considered missing values and accounted for overdispersion of the data. The precision of estimates is not only affected by missing data but also the type of regression model used for analysis. For instance, results for the association between abruption placenta and risk of perinatal death in Model I before and after imputation of missing values shows a nearly 20 times higher risk of perinatal death and too wide confidence intervals. Model II, however, provide more precise estimates both before and after imputation (Table 4.5). We will, therefore, focus on results for Model II for interpretation of our findings.

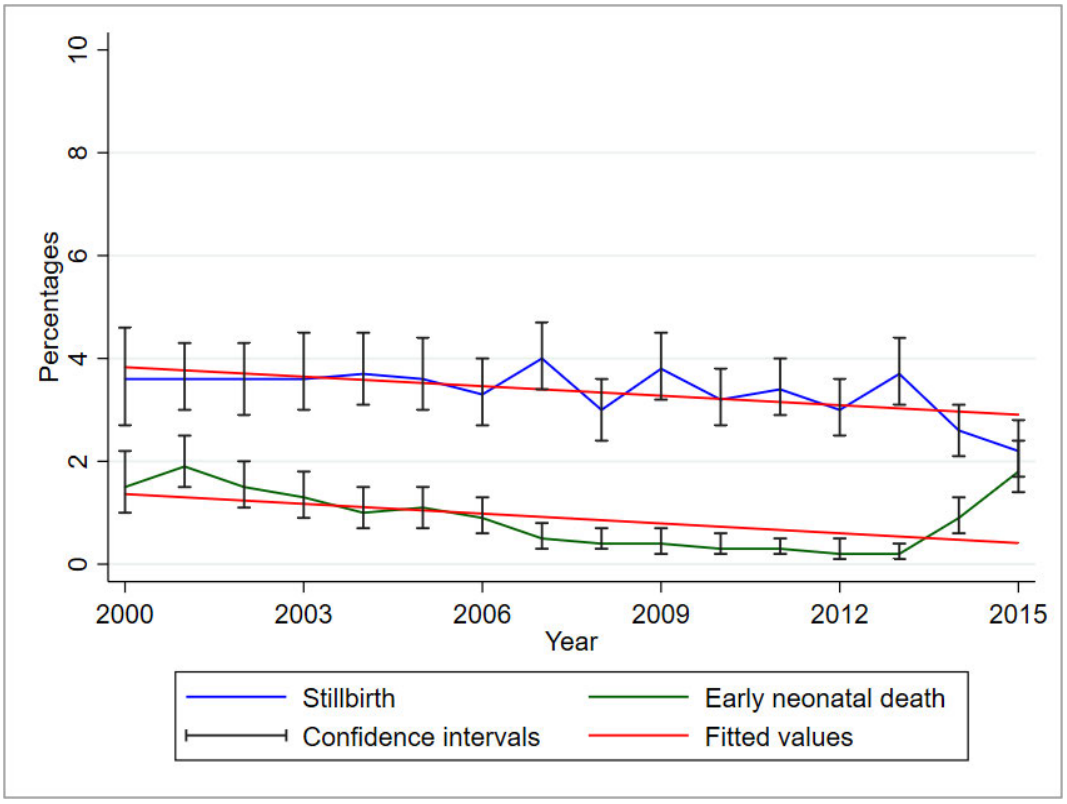


Figure 4.3: Trends of stillbirths and early neonatal deaths, KCMC Medical Birth Registry data, 2000-2015 (N=50,724)

Based on Model II, the risk of perinatal death was higher among deliveries from mothers who resided in rural compared to urban areas (RR=1.241, 95%CI 1.137, 1.355), with primary (RR=1.201, 95%CI 1.083, 1.332) compared to higher education levels, with <4 ANC visits (RR=1.250, 95%CI 1.146, 1.365) compared to ≥4 visits, with PPH (RR=2.638, 95%CI 1.997, 3.486), abruption placenta (RR=4.218, 95%CI 3.438, 5.175), LBW (RR=4.210, 95%CI 3.788, 4.679), male child (RR=1.090, 95%CI 1.007, 1.181), and were referred for delivery (RR=2.108, 95%CI 1.919, 2.317). On the other hand, a lower risk of perinatal death was observed among deliveries from mothers who experienced PROM (RR=0.411, 95%CI 0.283, 0.598) and delivered through CS (RR=0.662, 95%CI 0.604, 0.724).

We further examined the effect of eight or more ANC visits on the risk of perinatal

death as recommended by WHO (World Health Organization, 2016b). We would like to indicate it here that data analyzed in this study are between the years 2000 and 2015, during which the recommendation was at least four ANC visits, but WHO issued the new guidelines in 2016. About 94% of 49,911 recorded deliveries had <8 ANC visits. Results from the complete case analysis of the GEE log-linear model (Poisson family, log link function, exchangeable correlation and robust variance estimator) almost agreed with those presented in Table 4.5 for the four or more ANC visits. The risk of perinatal death among deliveries from women with <8 ANC visits in this model was (RR=1.372, 95%CI 1.031, 1.825) compared to ≥8 ANC visits. The relative risk of perinatal death in the same model (Model II in Table 4.5, in the complete case analysis) for <4 visits is 1.267 (95%CI 1.148, 1.399), though slightly lower.

4.3.7 Stratified analysis by referral status

Deliveries from women referred for delivery in this cohort had a higher risk of experiencing perinatal death (RR=2.577, 95%CI 2.280, 2.912) compared to those who were not (Table 4.5). These women are at risk of having serious pregnancy complications (Mmbaga et al., 2012b) that are likely to increase the risk of perinatal death. We, therefore, performed stratified analysis (of the imputed data) by referral status to better understand the risk factors of perinatal death (Table 4.6). Findings of the GEE log-linear regression model (i.e., Poisson family, log link, exchangeable correlation and robust variance estimator) indicated that, the risk of perinatal death among those referred for delivery was significantly higher among deliveries from women aged 35-39 years (RR=1.250, 95%CI 1.062, 1.471) compared to those aged 15-19 years, with preeclampsia/eclampsia (RR=1.339, 95%CI 1.135, 1.580), <4 ANC visits (RR=1.357, 95%CI 1.194, 1.541), with PPH (RR=3.182, 95%CI 2.308, 4.387), abruption placenta (RR=4.024, 95%CI 3.121, 5.186), preterm birth (RR=1.687, 95%CI 1.451, 1.960) and delivered a LBW baby (RR=2.722, 95%CI 2.351, 3.151) compared to those who did not. PROM and CS delivery were still protective against perinatal

Table 4.5: Adjusted analysis showing predictors of perinatal death

Characteristics	BEFORE IMPUTATION (N=43,198)				AFTER IMPUTATION (N=50,847)			
	Model I[a]		Model II[b]		Model I[a]		Model II[b]	
	AOR[c] (SE)	95%CI	ARR[d] (SE)	95%CI	AOR[c] (SE)	95%CI	ARR[d] (SE)	95%CI
Age groups (years)								
15-19	0 524 (0 059)	0 419, 0 654***	0 582 (0 058)	0 478, 0 708***	0 614 (0 058)	0 510, 0 738***	0 674 (0 054)	0 575, 0 789***
20-34	1		1		1		1	
35-39	1 277 (0 102)	1 092, 1 493**	1 196 (0 081)	1 048, 1 365**	1 281 (0 090)	1 116, 1 471***	1 201 (0 070)	1 072, 1 347**
40+	1 375 (0 197)	1 039, 1 821*	1 320 (0 150)	1 056, 1 650*	1 439 (0 178)	1 129, 1 834**	1 348 (0 131)	1 114, 1 630**
Area of residence (Rural)	1 221 (0 072)	1 087, 1 371***	1 202 (0 061)	1 088, 1 328***	1 274 (0 067)	1 150, 1 412***	1 241 (0 056)	1 137, 1 355***
Highest education level								
None	1 291 (0 254)	0 878, 1 898	1 182 (0 212)	0 832, 1 679	1 548 (0 223)	1 167, 2 054**	1 401 (0 171)	1 103, 1 780**
Primary	1 224 (0 082)	1 073, 1 397**	1 211 (0 071)	1 080, 1 358**	1 216 (0 075)	1 078, 1 372**	1 201 (0 063)	1 083, 1 332***
Secondary	1 013 (0 109)	0 821, 1 251	1 024 (0 094)	0 855, 1 226	0 956 (0 095)	0 787, 1 161	0 975 (0 082)	0 827, 1 150
Higher	1		1		1		1	
Pre-eclampsia/eclampsia (Yes)	1 638 (0 156)	1 359 ,1 975***	1 432 (0 103)	1 243, 1 648***	1 642 (0 136)	1 396, 1 932***	1 423 (0 087)	1 263, 1 603***
Drank alcohol during this pregnancy (Yes)	0 860 (0 058)	0 755, 0 981*	0 886 (0 051)	0 791, 0 991*	0 889 (0 053)	0 791, 0 998*	0 910 (0 046)	0 825, 1 004
Number of ANC visits (<4)	1 313 (0 076)	1 172, 1 470***	1 267 (0 064)	1 148, 1 399***	1 294 (0 067)	1 169, 1 433***	1 250 (0 056)	1 146, 1 365***
PROM (Yes)	0 322 (0 082)	0 195, 0 530***	0 377 (0 088)	0 238, 0 597***	0 346 (0 074)	0 228, 0 525***	0 411 (0 079)	0 283, 0 598***
PPH (Yes)	4 299 (1 014)	2 708, 6 825***	2 648 (0 435)	1 918, 3 654***	4 781 (0 993)	3 183, 7 182***	2 638 (0 375)	1 997, 3 486***
Abruption placenta (Yes)	19 469 (4 515)	12 357, 30 673***	4 472 (0 525)	3 552, 5 629***	18 349 (3 821)	12 200, 27 598***	4 218 (0 440)	3 438, 5 175***
Gestational age (<37 weeks)	2 416 (0 161)	2 121, 2 752***	2 072 (0 122)	1 847, 2 324***	2 495 (0 148)	2 222, 2 802***	2 111 (0 110)	1 906, 2 338***
Delivery mode (CS)	0 569 (0 036)	0 502, 0 645***	0 634 (0 034)	0 571, 0 704***	0 594 (0 034)	0 531, 0 664***	0 662 (0 031)	0 604, 0 724***
Birth weight (LBW)	5 064 (0 336)	4 446, 5 768***	4 243 (0 258)	3 765, 4 781***	5 118 (0 302)	4 560, 5 745***	4 210 (0 227)	3 788, 4 679***
Sex of the baby (Male)	1 119 (0 062)	1 005, 1 247*	1 115 (0 052)	1 017, 1 222*	1 098 (0 054)	0 998, 1 208	1 090 (0 045)	1 007, 1 181*
Referred for delivery (Yes)	2 577 (0 161)	2 280, 2 912***	2 218 (0 120)	1 994, 2 466***	2 465 (0 139)	2 207, 2 754***	2 108 (0 101)	1 919, 2 317***
Year	0 954 (0 007)	0 941, 0 967***	0 962 (0 006)	0 951, 0 973***	0 950 (0 006)	0 939, 0 961***	0 960 (0 005)	0 951, 0 969***

[a]GEE model with binomial family, logit link, exchangeable correlation and robust variance estimator

[b]GEE log-linear regression model, i.e., Poisson family, log link, an exchangeable correlation, and robust variance estimator

[c]Adjusted Odds Ratio

[d]Adjusted Risk Ratio

* p<0.05, ** p<0.01, *** p<0.001

death, (RR=0.426, 95%CI 0.261, 0.697) and (RR=0.628, 95%CI 0.556, 0.708), respectively. Maternal area of residence, education level, alcohol use during pregnancy and sex of the child were not associated with the risk of perinatal death in this analysis. Although not statistically significant, deliveries from women who consumed alcohol during pregnancy had 1.152 (95%CI 0.998, 1.329) times the risk of experiencing a perinatal death.

It is worth noting here that, compared to women referred for delivery, there is a stronger association between most of the covariates and the risk of perinatal death in the group of women not referred for delivery (Table 4.6). Higher risk of perinatal death is still on deliveries from women aged 35-39 years, and 40+ years compared to 20-34 years, rural residents, those with low education level, with pre-eclampsia/eclampsia, <4 ANC visits, PPH, abruption placenter, preterm birth, LBW, and male child. Deliveries from women aged 15-19 years, with PROM and delivered through CS had lower risk of experiencing a perinatal death (Table 4.6). It was noted that under Model II after imputation of missing values, consumption of alcohol during pregnancy among women not referred for delivery had a strong protective effect on the risk of perinatal death (RR=0.757, 95%CI 0.659, 0.869, $p<0.001$).

4.3.8 Predictors of stillbirth

The proportion of stillbirths in this study is higher than that of early neonatal deaths as shown in Figure 4.3. To better understand the risk factors for death among neonates in this cohort, we excluded 415 (0.8%) early neonatal deaths among 50,847 total deliveries. We then reanalyzed the data after performing multiple imputation of missing values. It can be observed (in Table 4.7) that, results from this analysis (the GEE log-linear model – Poisson family, log link, exchangeable correlation, and robust variance estimator) agrees with those for the predictors of perinatal deaths in Table 4.5 except for the sex of the child and

Table 4.6: Adjusted analysis showing predictors of perinatal death stratified by referral status

Characteristics	Referred for delivery (N=11,488)		Not referred for delivery (N=37,479)	
	ARR[a] (SE)	95%CI	ARR[a] (SE)	95%CI
Age groups (years)				
15-19	0.619 (0.065)	0.503, 0.761***	0.667 (0.093)	0.508, 0.875**
20-34	1		1	
35-39	1.250 (0.104)	1.062, 1.471**	1.233 (0.100)	1.051, 1.445*
40+	1.273 (0.172)	0.977, 1.660	1.407 (0.211)	1.049, 1.886*
Area of residence (Rural)	1.131 (0.075)	0.993, 1.288	1.257 (0.076)	1.117, 1.414***
Highest education level				
None	1.249 (0.194)	0.921, 1.694	1.511 (0.297)	1.028, 2.220*
Primary	1.115 (0.967)	0.941, 1.321	1.229 (0.085)	1.072, 1.408**
Secondary	0.999 (0.128)	0.777, 1.285	0.923 (0.110)	0.731, 1.165
Higher	1		1	
Pre-eclampsia/eclampsia (Yes)	1.339 (0.113)	1.135, 1.580**	1.588 (0.141)	1.335, 1.890***
Drank alcohol during this pregnancy (Yes)	1.152 (0.084)	0.998, 1.329	0.757 (0.053)	0.659, 0.869***
Number of ANC visits (<4)	1.357 (0.088)	1.194, 1.541***	1.175 (0.074)	1.039, 1.330*
PROM (Yes)	0.426 (0.107)	0.261, 0.697**	0.243 (0.089)	0.118, 0.500***
PPH (Yes)	3.182 (0.521)	2.308, 4.387***	1.816 (0.470)	1.094, 3.015*
Abruption placenta (Yes)	4.024 (0.521)	3.121, 5.186***	5.023 (0.734)	3.772, 6.689***
Gestational age (<37 weeks)	1.687 (0.129)	1.451, 1.960***	2.523 (0.201)	2.158, 2.949***
Delivery mode (CS)	0.628 (0.039)	0.556, 0.708***	0.651 (0.048)	0.563, 0.752***
Birth weight (LBW)	2.722 (0.203)	2.351, 3.151***	6.220 (0.492)	5.328, 7.263***
Sex of the baby (Male)	0.983 (0.059)	0.875, 1.105	1.156 (0.068)	1.030, 1.297*
Year	0.979 (0.007)	0.965, 0.994**	0.946 (0.007)	0.932, 0.959***

Note: Stratified analysis was performed after imputing the missing data (conditional on each referral status) using the GEE log-linear regression model, i.e., Poisson family, log link, exchangeable correlation, and robust variance estimator.
[a] Adjusted Risk Ratio
* p<0.05, ** p<0.01, *** p<0.001.

consumption of alcohol during pregnancy. The risk of perinatal death is significantly higher among deliveries from women aged 35-39 (RR=1.252, 95%CI 1.103, 1.421) and 40+ (RR=1.366, 95%CI 1.096, 1.701) compared to 20-34 years, rural residents (RR=1.273, 95%CI 1.153, 1.405), with no education (RR=1.428, 95%CI 1.089, 1.873) and primary education level (RR=1.224, 95%CI 1.090, 1.374) compared to higher education. The risk is also high among deliveries from women with pre-eclampsia/eclampsia (RR=1.497, 95%CI 1.311, 1.710), <4 ANC visits (RR=1.286, 95%CI 1.163, 1.423), PPH (RR=3.100, 95%CI 2.282, 4.210), abruption placenta (RR=4.683, 95%CI 3.722, 5.893), preterm birth (RR=2.210, 95%CI 1.961, 2.492), delivered a LBW baby (RR=4.619, 95%CI 4.091, 5.214) and were referred for delivery (RR=2.080, 95%CI 1.870, 2.313). The risk of stillbirth decreased significantly

with every one-year increase in time (RR=0.970, 95%CI 0.960, 0.980). Furthermore, the risk of perinatal death was low among deliveries from women aged 15-19 years (RR=0.574, 95%CI 0.475, 0.695), who experienced PROM (RR=0.302, 95%CI 0.184, 0.497), and delivered through CS (RR=0.555, 95%CI 0.449, 0.617).

Table 4.7: Adjusted analysis showing predictors of stillbirth

Characteristics	Before Imputation (N=43,198)		After imputation (N=50,432)	
	ARR[a] (SE)	**95%CI**	**ARR[a] (SE)**	**95%CI**
Age groups (years)				
15-19	0.486 (0.059)	0.383,0.616***	0.574 (0.056)	0.475, 0.695***
20-34	1		1	
35-39	1.206 (0.091)	1.040,1.397*	1.252 (0.081)	1.103, 1.421***
40+	1.324 (0.171)	1.028,1.705*	1.366 (0.153)	1.096, 1.701**
Area of residence (Rural)	1.217 (0.069)	1.089,1.360***	1.273 (0.064)	1.153, 1.405***
Highest education level				
None	1.192 (0.242)	0.801,1.774	1.428 (0.198)	1.089, 1.873*
Primary	1.225 (0.080)	1.078,1.392**	1.224 (0.072)	1.090, 1.374**
Secondary	0.999 (0.102)	0.818,1.219	0.957 (0.091)	0.795, 1.152
Higher	1		1	
Pre-eclampsia/eclampsia (Yes)	1.506 (0.119)	1.290,1.759***	1.497 (0.101)	1.311, 1.710***
Drank alcohol during this pregnancy (Yes)	0.888 (0.057)	0.784,1.007	0.909 (0.051)	0.815, 1.015
Number of ANC visits (<4)	1.276 (0.072)	1.143,1.425***	1.286 (0.066)	1.163, 1.423***
PROM (Yes)	0.288 (0.087)	0.160,0.519***	0.302 (0.077)	0.184, 0.497***
PPH (Yes)	3.080 (0.546)	2.176,4.359***	3.100 (0.484)	2.282, 4.210***
Abruption placenta (Yes)	5.039 (0.650)	3.913,6.489***	4.683 (0.549)	3.722, 5.893***
Gestational age (<37 weeks)	2.188 (0.145)	1.922,2.490***	2.210 (0.135)	1.961, 2.492***
Delivery mode (CS)	0.534 (0.033)	0.474,0.602***	0.555 (0.030)	0.499, 0.617***
Birth weight (LBW)	4.581 (0.314)	4.005,5.241***	4.619 (0.286)	4.091, 5.214***
Sex of the baby (Male)	1.121 (0.058)	1.012,1.242*	1.081 (0.050)	0.988, 1.183
Referred for delivery (Yes)	2.185 (0.132)	1.940,2.460***	2.080 (0.113)	1.870, 2.313***
Year	0.972 (0.006)	0.960,0.985***	0.970 (0.005)	0.960, 0.980***

Note: Analysis performed using the GEE log-linear regression model, i.e., Poisson family, log link, exchangeable correlation, and robust variance estimator.
[a] Adjusted Risk Ratio
* p<0.05, ** p<0.01, *** p<0.001.

4.4 Discussion

The proportion of perinatal death in this study was 4.2%, an estimated PMR of 41.6 per 1000 births. This study found that, under MAR assumption, ignoring missing values leads to biased parameter estimates. Higher risk of perinatal death was associated with maternal demographic characteristics (i.e., age, area of residence

and education level), pregnancy and delivery-related characteristics (i.e., pre-eclampsia/eclampsia, number of ANC visits, PPH, abruption placenta, preterm birth, low birth weight, sex of the child and referral status). PROM and CS delivery were associated with a lower risk of perinatal death in this population. There were no differences in the predictors of perinatal death and stillbirth in this study.

The PMR in this study 41.6 per 1000 births is higher than 27.0 per 1,000 births reported in Manyara region, northern Tanzania (Hinderaker et al., 2003), and slightly higher than the national estimate, 39 per 1000 births (MoHCDGEC [Tanzania Mainland] et al., 2016) but significantly higher than those reported from high-income countries such as Ireland (5.4/1,000 births)(Unterscheider et al., 2014) and USA (6.0/1,000 births)(Gregory et al., 2018). Higher PMR than the national level could be because this study was conducted at the consultant referral hospital and therefore attending women with most complications or higher risk pregnancy compared to estimates at lower-level facilities and population surveys. PMR in this study is lower than those reported in India (49.4/1,000 births) (Bellad et al., 2010), Nigeria (49.9/1,000 births)(Oyira et al., 2017), and in Eastern Sudan (75.3 per 1,000 births)(Ali et al., 2014). Higher rate in Sudan was linked to the higher proportion of home deliveries, low ANC coverage, and maternal infections such as malaria and anemia (Ali et al., 2014). These inequalities reflect differences in the availability and quality of obstetric and newborn care services within and between countries. The proportion of stillbirth in this study was consistently larger than the early neonatal death, possibly because women and their newborns were discharged soon after birth; hence deaths occurring at home could not be recorded (Bailey et al., 2017). However, between the years 2013-2015, the proportion of early neonatal deaths increased while that of stillbirth decreased. A previous study on the recording of maternal deaths reported a considerable under-reporting of these deaths in the medical birth registry (Bergsjo et al., 2010), which might also be the case in the reporting of stillbirth and early neonatal deaths in this study. Concerted efforts are

needed to articulate and address the health system and provided related challenges in the delivery of quality reproductive, maternal, and newborn care services to avert the rising trends of PMR in Tanzania.

Ignoring missing values in the analysis of clinical events such as adverse pregnancy outcomes like perinatal death produces biased parameter estimates (Kombo et al., 2017; Liu & De, 2015a; Pedersen et al., 2017; Sterne et al., 2009; Van Buuren et al., 2006a), as also shown in this study. After the imputation of missing values, the effect of covariates on perinatal death appeared to be stronger than before imputation. This emphasizes the need for considering missing values in modeling adverse pregnancy outcomes, particularly in longitudinal studies where such issues are common (Pedersen et al., 2017; Sterne et al., 2009; Ibrahim & Molenberghs, 2009). We further observed that, the GEE log-binomial regression model slightly over-estimated parameter estimates compared to the log-linear model. For more precise parameter estimates, data analysts should carefully consider the choice of the regression models that better fit their data.

Unexpectedly, alcohol use during pregnancy was observed to reduce the risk of perinatal death before imputation of missing values and after imputation using the GEE binomial regression model. We did not find a significant association in the GEE with a log-linear regression model, which agrees with other studies (Baptista et al., 2017). However, maternal alcohol consumption during pregnancy has been associated with a decreased risk of being small for gestational age (below 10^th percentile related to gestational age and sex) and that of preterm birth (Isaksen et al., 2015). Although the stratified analysis in our study agrees with these findings only among women not referred for delivery, this finding should be interpreted with caution due to the low proportion (3.5%) of women who drank alcohol during pregnancy in this category. Moreover, the study did not collect data on the dose-effect relationship for the association between alcohol consumption and

perinatal death. More should be done to investigate the effect of alcohol use during pregnancy on the adverse pregnancy outcomes because it is unlikely that alcohol consumption in itself can explain these findings (Isaksen et al., 2015). Furthermore, Isaksen et al. (2015) recommended further insights on diet (and nutrition), lifestyle factors, and maternal health that could explain this paradoxical association. Despite the observed association, the WHO recommends screening and counseling of pregnant women on the harmful effects of substance use during pregnancy (including alcohol) at every ANC visit (World Health Organization, 2016b). We found the risk of perinatal death to be high among deliveries from mothers in higher age categories (35-39 and 40+ years) compared to those aged 20-34 years.

Advanced maternal age (>35 years) has been linked with a higher risk of stillbirth (Nijkamp et al., 2017), while in Uganda, there was no significant age effect (Nankabirwa et al., 2011). However, the risk of perinatal death was high among pregnant women aged 30+ years (Nankabirwa et al., 2011), which can be explained by small number of perinatal deaths in the 30+ age category. Rural residents had a higher risk of experiencing perinatal death in this study, contrary to findings from Uganda, who found the risk to be high in urban areas (Nankabirwa et al., 2011). Women from urban areas in Uganda were living in slums (characterized by low socio-economic status (SES) and poor access to care), which could explain the observed discrepancies. Furthermore, compared to our findings, there was no significant association between perinatal death and maternal education level in this study (Nankabirwa et al., 2011). It is worth noting here that, having no education was not associated with the risk of perinatal death before but not after imputation of missing values. Pregnant women residing in rural areas and with low education levels could be in a disadvantaged position in accessing quality health care during pregnancy and childbirth. Low socioeconomic status has also been linked to adverse pregnancy outcomes such as perinatal death (Bellad et al., 2010; Nijkamp et al., 2017; Blumenshine et al., 2010), but this could not be measured in our study.

Area of residence and education level were not significant predictors of perinatal death among those referred for delivery probably because most of those referred were coming from rural areas of residence, had low SES and low education. Deliveries from women who had <4 ANC visits had a higher risk of perinatal death. Antenatal care provides a critical opportunity for women and babies to benefit from good quality maternal care (Benova et al., 2018; World Health Organization, 2016b). Furthermore, the risk of perinatal death was higher among males compared to female children in this study, which also agrees with findings from Brazil (Miranda et al., 2017). This may be linked to early pulmonary maturation among females that lowers the risk of respiratory complications (Miranda et al., 2017), one of the leading causes of under-five deaths in low-income countries (Liu et al., 2016). However, sex of the child was not associated with the risk of perinatal death and stillbirth among women referred for delivery. Interventions to increase coverage and uptake of recommended routine ANC services for pregnant women are crucial for early identification and management of pregnancy-related complications.

The risk of perinatal death was higher among deliveries from women with hypertensive disorders of pregnancy particularly pre-eclampsia/eclampsia, those with postpartum hemorrhage, abruption placenta, delivered preterm and low birth weight baby, which is similar to other studies (Vogel et al., 2014; Bellad et al., 2010; Mpembeni et al., 2014; Nijkamp et al., 2017). We did not find a significant association between maternal anemia, malaria infection, positive HIV status, systemic infections/sepsis, and placenta previa with an increased risk of perinatal death. PROM and delivery by CS reduced the risk of perinatal death in this cohort. The latter is known to reduce the risk of obstetric complications when medically indicated (Vogel et al., 2014; MoHCDGEC [Tanzania Mainland] et al., 2016). The protective effect of PROM could reflect timely management of such pregnancies

considering that these deliveries were attended at a tertiary care facility where comprehensive emergency obstetric and newborn care (CEmONC) services are available. Despite that, a study in Uganda has shown that CS delivery, particularly among women with PROM, increases the risk of perinatal mortality among other adverse pregnancy outcomes (Kayiga et al., 2018).

Women referred for delivery are potentially at risk of experiencing adverse maternal and perinatal outcomes (Mahande et al., 2013a; Mmbaga et al., 2012b; Vogel et al., 2014). In this study, we included referral cases in our analyses and later performed a stratified analysis. As one would expect, the risk of perinatal death was almost twice higher among deliveries from women referred for delivery compared to those who were not. This is because these women are likely to have experienced pregnancy-related complications that required specialized care (Mmbaga et al., 2012b; Vogel et al., 2014). Findings from the stratified analysis by referral status were almost comparable with non-referral results except for area of residence, education level, alcohol use during pregnancy, and sex of the child, which were not statistically significant, in those referred. Yet, we cannot ignore the fact that women referred for delivery are most at risk of experiencing adverse pregnancy outcomes. It is essential to strengthen the referral system to ensure timely and proper referral mechanisms and to promote appropriate health care seeking behavior to reduce the risk of perinatal deaths (Mmbaga et al., 2012b). At the same time, we observed a stronger association between covariates and the risk of perinatal death in the group of women not referred for delivery compared to those referred. This group of women should also be given due attention in order to prevent the rise in perinatal death cases.

Being a registry-based study from a zonal referral hospital, findings (on predictors of perinatal death) may not be generalized to a larger population. However, population-based estimates agree with these results (Ali et al., 2014; Hinderaker

et al., 2003; Nankabirwa et al., 2011). Furthermore, the fact that this birth registry only captured deaths occurring in the health facility lowers the number of recorded neonatal deaths. It could also underestimate the proportion and rates of perinatal death. Hence, stillbirth contributed to a higher percentage of the perinatal death numbers. This could explain the unobserved differences between predictors of perinatal death and stillbirth in this study. To our knowledge, no study has assessed the effect of ignoring missing values on determining predictors of adverse pregnancy outcomes, particularly perinatal death. This study has provided evidence for the need to consider missing values in the analysis of pregnancy outcomes. Our findings also emphasize the necessity of proper choice of statistical models for more precise parameter estimates. Increased surveillance and management of different maternal risk factors for different perinatal outcomes should be at the heart of improving child survival.

4.5 Conclusion

Perinatal mortality in this cohort is higher than the national estimate. Higher risk of perinatal death was associated with low maternal education level, rural residence, <4 ANC visits, PPH, abruption placenta, LBW delivery, child's sex, and being referred for delivery. Ignoring missing values in the analysis of adverse pregnancy outcomes produces biased covariate coefficients and standard errors. Close clinical follow-up of women at high risk of experiencing perinatal death, particularly during ANC visits and delivery, is of high importance to increase perinatal survival.

Chapter 5

Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern Tanzania

5.1 Introduction

Neonatal survival is at the heart of sustainable development goals (SDG) agenda (UNDP, 2018; World Health Organization, 2016a). The Every Newborn Action Plan to end Preventable Deaths set a goal for all countries to reach the target of ten or less newborn deaths per 1000 live births and ten or less stillbirths per 1000 total births by the year 2035 (World Health Organization, 2014a). Furthermore, the United Nations set the target of reducing neonatal mortality to 12 deaths per 1000 live births or fewer by 2030 (UNDP, 2018). Globally, neonatal deaths declined by 51% from 5 million in 1990 to 2.5 million in 2017. But this decline has not been realized in low- and middle-income countries (LMICs), which carries the highest burden of neonatal deaths, with south Asia and sub-Saharan Africa accounting for 79% of the total burden of neonatal deaths in 2017 (Hug et al., 2019). Furthermore,

the under-five mortality rate has decreased almost across the world, but the proportions of neonatal deaths remained high in this group (Burstein et al., 2019; UNICEF et al., 2020). Neonatal deaths accounted for 47% of all under-five deaths in 2019, and it has increased from 40% in 1990, with sub-Saharan Africa bearing the highest burden (UNICEF et al., 2020). Globally 2.4 million children died in the first month of life in 2018, with approximately 6,700 newborn deaths every day (UNICEF et al., 2020). Nearly three-quarters of these deaths occur during the first week, with about one million dying on the first day and close to one million dying within the next six days (UNICEF et al., 2020).

Globally, more than five million perinatal deaths occur each year (World Health Organization, 2016a). The majority (95%) of these deaths occur in sub-Saharan Africa and Southern Asia (Akombi & Renzaho, 2019). According to the Tanzania Demographic and Health Survey, the perinatal mortality rate has slightly increased from 36 to 39 deaths per 1,000 live births between 2010-11 and 2015-16 survey rounds, respectively relative to under-five mortality (MoHCDGEC [Tanzania Mainland] et al., 2016). In addition, perinatal mortality rate in Tanzania is the highest in East Africa (Akombi & Renzaho, 2019).

Early identification of pregnant women at risk for adverse maternal and perinatal outcomes during the prenatal period and timely provision of high-quality health care services have been reported to improve maternal and newborn survival (Kuhle et al., 2018). Machine learning (hereafter denoted as 'ML') models are methodologies for developing algorithms that learn from existing data to make predictions on new data (Kuhle et al., 2018). ML models have shown better predictive performance over the classical or conventional regression models (Raita et al., 2019), and they can better handle a significant number of potential predictors. However, there is conflicting evidence of the performance of these models. Previous investigators have demonstrated that, compared with the classical

regression models, ML models have superior performance for early differentiation of sepsis and non-infectious systemic inflammatory response syndrome in critically ill children (Lamping et al., 2018), in predicting neonatal and under-five mortality (Nasejje & Mwambi, 2017; Houweling et al., 2019; Hoodbhoy et al., 2019; Muktan et al., 2019; Lee et al., 2019b), and critical care and hospitalization outcomes (Raita et al., 2019; Goto et al., 2019; Vellido et al., 2018). In contrast, other studies have shown no predictive performance benefit of the ML models in prediction of clinical outcomes (Christodoulou et al., 2019; Kuhle et al., 2018).

The first step in addressing high perinatal mortality is the accurate capture and classification of the causes of those deaths across all settings (Allanson et al., 2016). The World Health Organization (WHO) International Classification of Diseases (ICD-10) is a standardized tool used for the classification of deaths occurring during the perinatal period: ICD-PM (World Health Organization, 2016a; Wojcieszek et al., 2016; Allanson et al., 2016). ML models may be an essential tool in the assessment of risk factors for deaths during the perinatal period and triage pregnant women at high risk of experiencing adverse perinatal outcomes, especially in low resourced settings where the majority of perinatal deaths occur at home (Pitt et al., 2016; World Health Organization, 2019; Khan et al., 2020; Chaibva et al., 2019). Capturing the chain of events that led to the perinatal mortality, from both the maternal and the perinatal side, informs the design and development of preventative and therapeutic measures (World Health Organization, 2016a).

Using data from the medical birth registry at Kilimanjaro Christian Medical Center (KCMC) referral hospital in northern Tanzania, we aimed to determine the key predictors of perinatal death using machine learning models. Previous studies using the same data (Mmbaga et al., 2011; Mitao et al., 2016; Mahande et al., 2013a,b; Chuwa et al., 2017; Isaksen et al., 2015) applied standard regression models to assess risk factors for adverse perinatal outcomes. A major weakness of

conventional regression analysis, as opposed to ML models, is that many covariates are excluded based on specific model assumptions. In contrast, ML techniques which are non-parametric in nature find the most predictive groupings of factors based on their frequency and strength of association, with no particular model assumptions (Hamilton et al., 2020). In this study, we compared the predictive performance of the ML models with the conventional regression analysis, particularly logistic regression.

5.2 Methods

5.2.1 Data source

Data for this study comes from the KCMC referral hospital medical birth registry between the years 2000-2015, which were collected among mothers who delivered at the department of obstetrics and gynecology. More description of the KCMC medical birth registry is also available in Chapter 2 and (Bergsjo et al., 2007; Mahande, 2015; Mmbaga et al., 2011, 2012a,b; Mahande et al., 2013a,b; Chuwa et al., 2017). Briefly, the KCMC medical birth registry is within hospital grounds at the Reproductive and Child Health Centre. The birth registry has been in operation since the year 2000, established to serve both clinical, administrative, and research purposes (Bergsjo et al., 2007; Mahande, 2015). Trained midwives collected data using a standardized questionnaire (within 24 hours after delivery or later in case a mother had recovered from complications), after which data is entered into a computerized database located at the birth registry. Also, additional data were abstracted from the antenatal (ANC) cards and the hospital medical records of the mother (Mahande et al., 2013b).

A unique hospital identification number was assigned to each woman at first admission and used to trace her medical records at later admissions, and further to link records of successive births of the same woman (Mahande, 2015). Data

captured information on the background characteristics of mother and father, mother's health before and during present pregnancy, information about delivery including complications, and child characteristics including their status (i.e., whether dead or alive).

5.2.2 Study variables and variable measurements

The main outcome variable in this study was perinatal death which was defined as the number of stillbirths (pregnancy loss that occurs after seven months of gestation) and early neonatal deaths (deaths of live births within the first seven days of life)(MoHCDGEC [Tanzania Mainland] et al., 2016; World Health Organization, 2019). The perinatal death was coded as binary, i.e., 'yes' if death occurred during the perinatal period and 'no' if otherwise. This outcome only captured deaths that occurred within the hospital before the discharge of mothers. There are no follow-up mechanisms for deaths that occur outside the health facility (KCMC hospital).

We included a total of 32 predictor variables for the machine learning models. Previous literature informed the selection of these variables (Allanson et al., 2016; Getiye & Fantahun, 2017; Hug et al., 2019; Nijkamp et al., 2017; Mitao et al., 2016; Mpembeni et al., 2014; Mutsaerts et al., 2014; Ouyang et al., 2013; Unterscheider et al., 2014; Vogel et al., 2014), most of which are available in the birth registry. These included maternal and paternal background characteristics; age in years, area of residence (rural vs. urban), highest education level (none, primary, secondary and higher), marital status (single, married and widow/divorced), and occupation (unemployed, employed and others). Further, specific characteristics of the mother included referral status (whether referred for delivery or not), and the number of antenatal care visits (<4 and ≥ 4 visits).

We excluded maternal body mass index (BMI) and HIV status because they

contributed to nearly 47% of all missing values in the dataset. Maternal health during pregnancy included; alcohol consumption, smoking, gestational diabetes, diabetes, hypertension, pre-eclampsia/eclampsia, bleeding (i.e., the woman observed blood from the vagina at any time during the pregnancy), anemia, malaria, and systemic infections/sepsis. Variables with information concerning delivery included; induction of labor (yes or no), mode of delivery (vaginal vs. caesarean section), presentation (breech vs. cephalic), complications during birth, particularly premature rupture of the membranes (PROM), postpartum hemorrhage (PPH), placenta previa, and placenta abruption, all categorized as yes and no. Gestational age at birth was estimated based on the date of the last menstrual period and recorded in full weeks. Preterm was defined as babies born alive before 37 weeks of pregnancy are completed (World Health Organization, 2019). Child characteristics included sex (male or female), low birth weight defined as an infant birth weight of less than 2500g (Mitao et al., 2016; World Health Organization, 2014b), and year of birth.

5.2.3 Statistical and computational analysis

Data were cleaned and then analyzed using Stata version 15.1 (StataCorp, 2019). Categorical variables were summarized using frequencies and proportions. The Chi-square (χ^2) statistic was used to test the relationships between a set of independent variables and perinatal death. For the ML models (i.e., from feature selection, training, testing, and comparison of the predictive performance of the machines), we used R version 3.6.3 (R Core Team, 2020). The training dataset contained 70% of randomly selected samples used to develop six different ML models to predict perinatal death. These are artificial neural networks (ANN), random forests (RF), Naïve Bayes (NB), bagged trees, boosting, and the logistic regression (Lreg) model. We used the caret package to implement these models in R.

Briefly, ANN is a method constructed from three layers of connected nodes: input, hidden, and output (Dwivedi, 2018). The input where each input variable appears as a node; the hidden layer contains several nodes determined during the model tuning phase. In contrast, the output layer contains several nodes equal to the number of classes to be predicted (Stephens & Diesing, 2014). Between these layers there are weighted links (Kuhle et al., 2018; Stephens & Diesing, 2014; Dwivedi, 2018), the hidden layer receives a sum of the multiplication of the input variables with associated weights values plus the bias (Dwivedi, 2018; Stephens & Diesing, 2014). This value is entered into an activation function, such as a logistic or sigmoid function, to decide the class prediction. Outputs of the network are interpreted as class probabilities and sum to one (Stephens & Diesing, 2014). We used nnet package to construct the ANN model.

RF is an extension of classification and regression trees (CART) (Stephens & Diesing, 2014; Raita et al., 2019; Kuhle et al., 2018; Breiman, 2001). RF performance is better compared to bagged trees because it decorrelates the trees (Hastie et al., 2009), hence improves accuracy (Breiman, 2001). Several forests of decision trees are grown using a random bootstrapped training sample. Also, instead of using all the variables/features in each tree, a random sample of variables are selected and tested at each split in each tree (Stephens & Diesing, 2014; Raita et al., 2019; Breiman, 2001). The prediction is made for unobserved data by taking a majority vote of the individual trees (Breiman, 2001; Stephens & Diesing, 2014). We used *randomForest* package to construct the RF model. NB is an effective classifier (Dwivedi, 2018) due to its simplicity, exhibiting a surprisingly competitive predictive accuracy (De Campos et al., 2011). NB uses probability theory to find the most possible sample class in a classification problem. NB has two assumptions; (1) each attribute is conditionally independent of the other attributes given the class, and (2) all the attributes have an impact on the class (De Campos et al., 2011; Stephens & Diesing, 2014). We used *naivebayes* package to construct the NB model.

Logistic regression (Lreg) is a standard multivariate classification method. It arises from the desire to model the posterior probabilities via linear functions in covariates, such that besides predicting class labels, it provides a probabilistic interpretation of this labeling (Dreiseitl & Ohno-Machado, 2002; Hastie et al., 2009; Musa, 2014). Lreg uses a sigmoid function instead of a linear function to map predictions to probabilities between 0 and 1 (Hastie et al., 2009). We used glm method to construct the Lreg model. Bagging, or bootstrap aggregation and boosting are general techniques for improving prediction rules and accuracy of the resulting predictions, by reducing the associated variance of prediction (Hastie et al., 2009; Sutton, 2005). Bagging divides the available data into many bootstrap samples and then train a separate model for each bootstrap, and then make a final prediction by averaging and voting for regression and classification, respectively (Sutton, 2005). Boosting, on the other hand, is a committee-based approach that uses a weighted average of prediction from various samples. The incorrectly predicted cases from a given step are given a higher weight during the next step. Thus, it is an iterative procedure, incorporating weights, as opposed to simple averaging of predictions (Sutton, 2005). We used *treebag* method and *gbm* package to construct the bagging and boosting models, respectively.

In the training set, parameter tuning and cross-validation aim to find a balance between building a model that can classify the training data effectively without overfitting to the random fluctuations (Stephens & Diesing, 2014). For each ML model, we used 10-fold cross-validation as a resampling method, where the training set is divided equally into ten parts (folds). Therefore, every nine folds are used together for training the model and the remaining one-fold for testing. This training-testing process is repeated ten times. Feature selection was performed using the RF algorithm. After selecting the most important features, they were retained in the dataset and used them for analysis in both the training and testing

data for all models. The Synthetic Minority Over-sampling Technique (SMOTE) method (Chawla, 2009; Johnson & Khoshgoftaar, 2019) to address the class imbalance in the outcome (i.e., the low proportion of perinatal deaths), by specifying the additional sampling to be "smote" on train control parameter specifications. SMOTE is a method that produces artificial minority samples by interpolating between existing minority samples and their nearest minority neighbors (Chawla, 2009; Johnson & Khoshgoftaar, 2019).

Using the testing set (30% of the remaining randomly selected sample), we computed the predictive performance of the six models (including logistic regression model) from the training set using the area under the receiver-operating-characteristics curve (AUC). We used the ROCR package for plotting ROC curves, obtaining the AUC values, and comparison of models using AUC values. We also used measures from the confusion matrix results (i.e., accuracy, sensitivity, specificity, positive and negative predictive values), and the net benefit through decision curve analysis (Vickers & Elkin, 2006; Zhang et al., 2018b) – which quantifies whether a machine provides a relevant improvement in the prediction. We used *epiR* package to obtain confidence intervals for the performance measures and DCA package (`http://www.decisioncurveanalysis.org`) for decision curve analysis. We further used *ggplot2* package to plot the decision curves. A good model will have a higher net benefit (Vickers & Elkin, 2006). We used Delong's test to compare the receiver-operating-characteristics curve (ROC) between models, where, a p-value of <0.05 was considered statistically significant. The variable importance is a scaled measure with a maximum value of 100 (Goto et al., 2019).

5.3 Results

5.3.1 Characteristics of study participants

A total of 55,003 total deliveries were recorded at the KCMC medical birth registry from 2000 to 2015. Of these, we excluded 3,316 (6%) multiple gestations (to avoid over-representation of high-risk pregnancies) (Chuwa et al., 2017), 49 (0.1%) records missing maternal identification numbers (hence could not be linked to child records), 791 (1.4%) records with a mismatch between the date of birth and unknown sequence (i.e., singleton vs. multiple births). We further excluded a total of 8,528 (15.5%) observations with missing values in both the outcome (perinatal status) and covariates. We, therefore, analyzed data for a total of 42,319 singleton deliveries with complete records (Figure 5.1). The characteristics of the participants

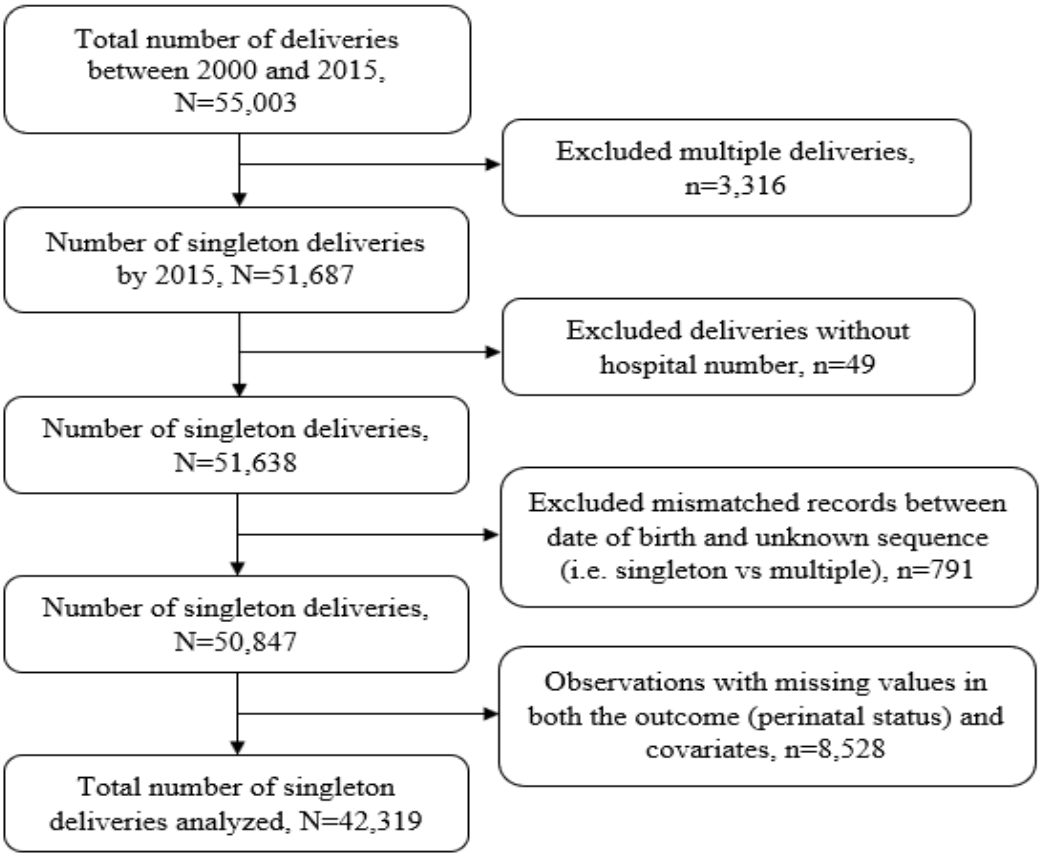


Figure 5.1: Schematic diagram showing the number of singleton deliveries analyzed, KCMC medical birth registry data, 2000-2015

are shown in Table 5.1. The overall proportion of perinatal death among 42,319 singleton deliveries in this study was 3.7%. The proportion of perinatal deaths among mothers aged 20-34, 35-39, and 40+ years was 3.5%, 4.7%, and 5.8%, respectively. Mothers with no education (5.6%) and those with primary education level (4.4%), who resided in rural areas (4.8%), had less than four antenatal care (ANC) visits (5.9%), and those referred for delivery (7.8%) had a higher proportion of perinatal death. Among fathers, a higher proportion of perinatal death is among those aged 35+ years (4.1%), with no (9.6%) or with primary education level (4.6%) as well as those who were unemployed (5.7%).

Furthermore, the most common obstetric care and complications in this birth cohort included induction of labor (22.7%), malaria (13.2%), preterm birth (10.8%), and LBW (10.2%). About 4% of mothers in this cohort experienced pre-eclampsia/eclampsia during pregnancy. Less than half of all children were females. The proportion of perinatal death among women who experienced induction of labor, with malaria, delivered preterm, delivered LBW baby, and experienced pre-eclampsia/eclampsia during pregnancy was 4.8%, 3.8%, 14.2%, 17.6%, and 12.5%, respectively. The proportion of perinatal death is almost similar among males (3.8%) compared to females (3.6%) children in this cohort (Table 5.2). The trends in the proportion of perinatal deaths that occurred at KCMC between the years 2000-2015 are shown in Figure 5.2. Overall, the proportion of perinatal deaths has slightly declined over the years by 6% (95%CI, 0.3%, 12.3%), though this decline was not statistically significant (p=0.06).

5.3.2 Variable importance

We used the random forest algorithm for feature/ variable selection. This model selected a total of 20 important predictors (Figure 5.3) based on its threshold measure of importance out of the 32 variables. We used these 20 variables in all the subsequent analysis for all models in both training and testing sets.

Table 5.1: Background characteristics of study participants (N=42,319)

Characteristics	Total	Perinatal death	P-value*
	n (%)	n (%)	
Maternal			
Age (years)			<0.001
15-19	3470 (8.2)	99 (2.9)	
20-34	32675 (77.2)	1158 (3.5)	
35-39	4984 (11.8)	235 (4.7)	
40+	1190 (2.8)	69 (5.8)	
Education level			<0.001
None	567 (1.3)	32 (5.6)	
Primary	23010 (54.4)	1019 (4.4)	
Secondary	5275 (12.5)	159 (3.0)	
Higher	13467 (31.8)	351 (2.6)	
Occupation			0.37
Unemployed	9316 (22.0)	365 (3.9)	
Employed	30061 (71.0)	1085 (3.6)	
Others	2942 (7.0)	111 (3.8)	
Marital status			0.89
Single	4954 (11.7)	186 (3.8)	
Married	37300 (88.1)	1372 (3.7)	
Widowed/Divorced	65 (0.2)	3 (4.6)	
Area of residence			<0.001
Urban	25056 (59.2)	725 (2.9)	
Rural	17263 (40.8)	836 (4.8)	
Alcohol consumption during pregnancy			0.001
No	30759 (72.7)	1191 (3.9)	
Yes	11560 (27.3)	370 (3.2)	
Smoking during pregnancy			0.97
Yes	53 (0.1)	2 (3.8)	
No	42266 (99.9)	1559 (3.7)	
Number of ANC visits			<0.001
≥4	28742 (67.9)	760 (2.6)	
<4	13577 (32.1)	801 (5.9)	
Referred for delivery			<0.001
No	32762 (77.4)	819 (2.5)	
Yes	9557 (22.6)	742 (7.8)	
Paternal characteristics			
Age (years)			0.001
<25	3938 (9.3)	122 (3.1)	
25-29	10593 (25.0)	346 (3.3)	
30-34	12303 (29.1)	457 (3.7)	
35+	15485 (36.6)	636 (4.1)	
Education level			<0.001
None	281 (0.7)	27 (9.6)	
Primary	18987 (44.9)	868 (4.6)	
Secondary	4565 (10.8)	154 (3.4)	
Higher	18486 (43.7)	512 (2.8)	
Occupation			<0.001
Unemployed	5710 (13.5)	323 (5.7)	
Employed	36102 (85.3)	1218 (3.4)	
Others	507 (1.2)	20 (3.9)	
Total	**42319**	**1561 (3.7%)**	

Table 5.2: Obstetric care characteristics and complications (N=42,319)

Characteristics	Total	Perinatal death	P-value*
	n (%)	n (%)	
Gestational diabetes			0.89
No	42288 (99.9)	1560 (3.7)	
Yes	31 (0.1)	1 (3.2)	
Diabetes			0.002
No	42240 (99.8)	1553 (3.7)	
Yes	79 (0.2)	8 (10.1)	
Hypertension			<0.001
No	42241 (99.8)	1550 (3.7)	
Yes	78 (0.2)	11 (14.1)	
Bleeding			<0.001
No	41897 (99.0)	1528 (3.6)	
Yes	422 (1.0)	33 (7.8)	
Anemia			0.004
No	41661 (98.4)	1523 (3.7)	
Yes	658 (1.6)	38 (5.8)	
Malaria			0.79
No	36746 (86.8)	1352 (3.7)	
Yes	5573 (13.2)	209 (3.8)	
Sepsis/ infections			0.43
No	41588 (98.3)	1538 (3.7)	
Yes	731 (1.7)	23 (3.1)	
Complications			
Pre-eclampsia/eclampsia			<0.001
No	40668 (96.1)	1355 (3.3)	
Yes	1651 (3.9)	206 (12.5)	
Induction of labor			<0.001
No	32732 (77.3)	1105 (3.4)	
Yes	9587 (22.7)	456 (4.8)	
PROM			0.006
No	41416 (97.9)	1543 (3.7)	
Yes	903 (2.1)	18 (2.0)	
PPH			<0.001
No	42091 (99.5)	1516 (3.6)	
Yes	228 (0.5)	45 (19.7)	
3-4-degree tear			0.49
No	42305 (99.9)	1560 (3.7)	
Yes	14 (0.1)	1 (7.1)	
Abruption placenta			<0.001
No	42193 (99.7)	1490 (3.5)	
Yes	126 (0.3)	71 (56.3)	
Placenta previa			0.04
No	42245 (99.8)	1555 (3.7)	
Yes	74 (0.2)	6 (8.1)	
Presentation			<0.001
Cephalic	41833 (98.9)	1459 (3.5)	
Breach/Transverse	486 (1.1)	102 (21.0)	
Gestational age at birth			<0.001
Term birth (≥37 weeks)	37764 (89.2)	914 (2.4)	
Preterm birth (<37 weeks)	4555 (10.8)	647 (14.2)	
Birth weight			<0.001
NBW	37991 (89.8)	801 (2.1)	
LBW	4328 (10.2)	760 (17.6)	
Child's sex			0.42
Female	20430 (48.3)	738 (3.6)	
Male	21889 (51.7)	823 (3.8)	

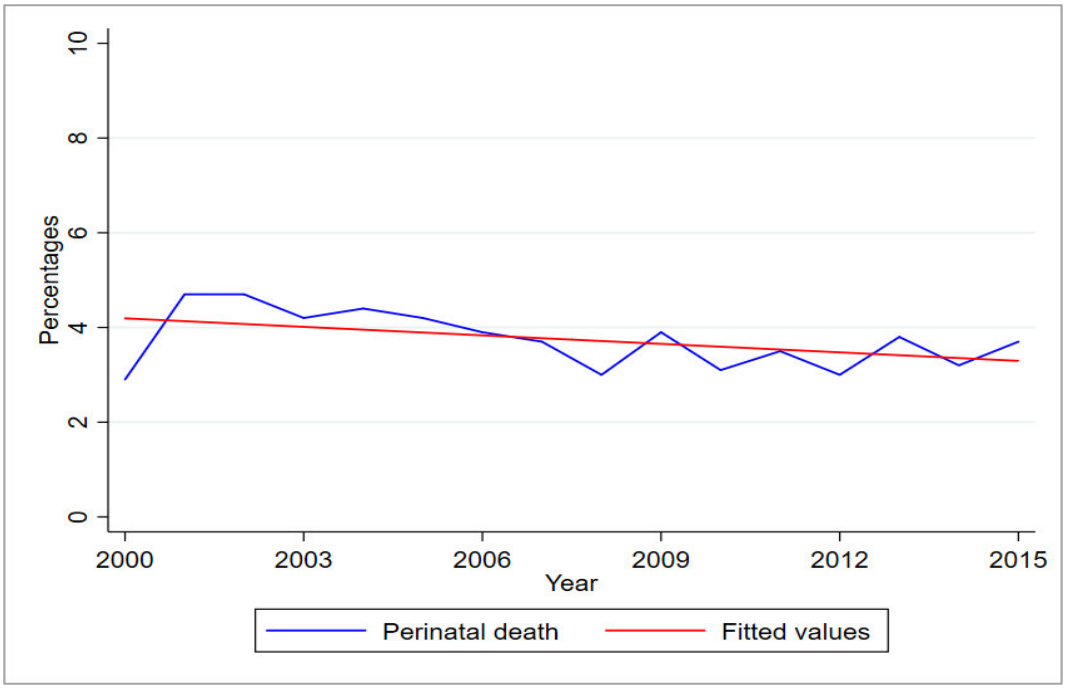


Figure 5.2: Trends of perinatal death, KCMC Medical Birth Registry data, 2000-2015

5.3.3 Predicting perinatal deaths

The discriminatory abilities of all models for the prediction of perinatal death are in Figure 5.4a and Table 5.3. There were no significant differences ($p>0.05$) in the area under the receiver-operating characteristics curve (AUC) between logistic regression with random forest (RF), artificial neural networks (ANN), boosting, and naïve Bayes (NB). However, bagging had significantly lower predictive performance (AUC: 0.76, 95%CI 0.74, 0.79, p=0.006) compared to the logistic regression model (AUC: 0.78, 95%CI 0.76, 0.81). Furthermore, the ANN model (Sensitivity: 0.60, 95% CI 0.55, 0.64) and NB model (Sensitivity: 0.57, 95% CI 0.52, 0.62) had slightly higher sensitivity compared to logistic regression (Sensitivity: 0.56, 95%CI 0.51, 0.60) while boosting (Specificity: 0.89 95%CI 0.88, 0.89) and RF (Specificity: 0.88, 95%CI 0.88, 0.89) had slightly higher specificity compared to logistic regression (Specificity: 0.87, 95%CI 0.86, 0.88). Due to the low prevalence of perinatal deaths (3.7%), all models had high negative predictive values (NPV: 0.98, 95%CI 0.98, 0.98).

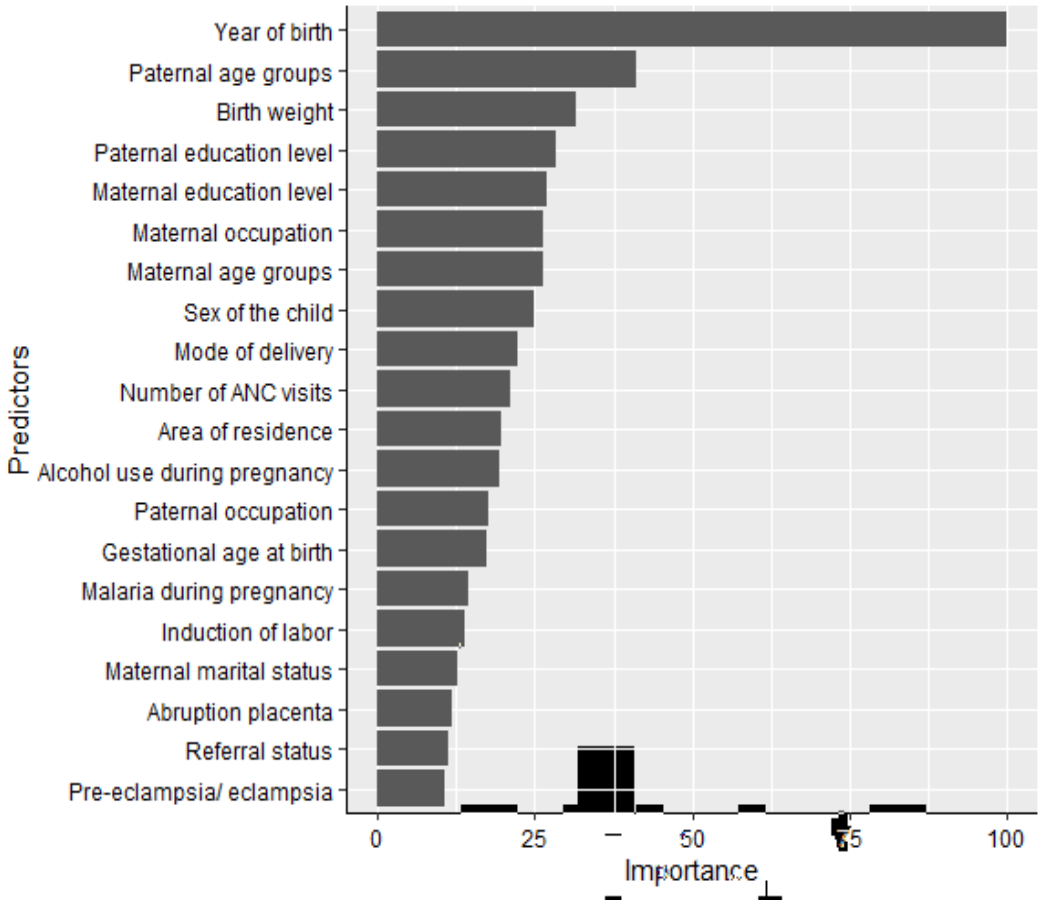


Figure 5.3: Variable importance of predictors for perinatal death in the random forest model scaled to have a maximum value of 100

Table 5.3: Prediction performance of the reference and machine learning models in the test set

Model	Lreg	ANN	RF	NB	Bagging	Boosting
ACC	0.86 (0.85, 0.86)	0.83 (0.82, 0.83)	0.87 (0.86, 0.87)	0.84 (0.83, 0.85)	0.82 (0.81, 0.83)	0.87 (0.87, 0.88)
AUC	0.78 (0.76, 0.81)	0.78 (0.76, 0.80)	0.79 (0.76, 0.81)	0.79 (0.76, 0.81)	0.76 (0.74, 0.79)	0.79 (0.76, 0.81)
P-value*	Reference	0.59	0.37	0.65	0.006	0.2
Sensitivity	0.56 (0.51, 0.60)	0.60 (0.55, 0.64)	0.54 (0.49, 0.58)	0.57 (0.52, 0.62)	0.55 (0.50, 0.59)	0.54 (0.49, 0.58)
Specificity	0.87 (0.86, 0.88)	0.84 (0.83, 0.84)	0.88 (0.88, 0.89)	0.85 (0.84, 0.86)	0.83 (0.82, 0.84)	0.89 (0.88, 0.89)
PPV	0.14 (0.12, 0.16)	0.12 (0.11, 0.14)	0.15 (0.13, 0.17)	0.13 (0.11, 0.14)	0.11 (0.10, 0.12)	0.15 (0.14, 0.17)
NPV	0.98 (0.98, 0.98)	0.98 (0.98, 0.98)	0.98 (0.98, 0.98)	0.98 (0.98, 0.98)	0.98 (0.98, 0.98)	0.98 (0.98, 0.98)

Abbreviations: ACC, Accuracy; AUC, Area Under the Curve; PPV, Positive Predictive Value; NPV, Negative Predictive Value

*We calculated p-values to compare the area under the receiver-operating-characteristics curve (AUC) of logistic with each machine learning model

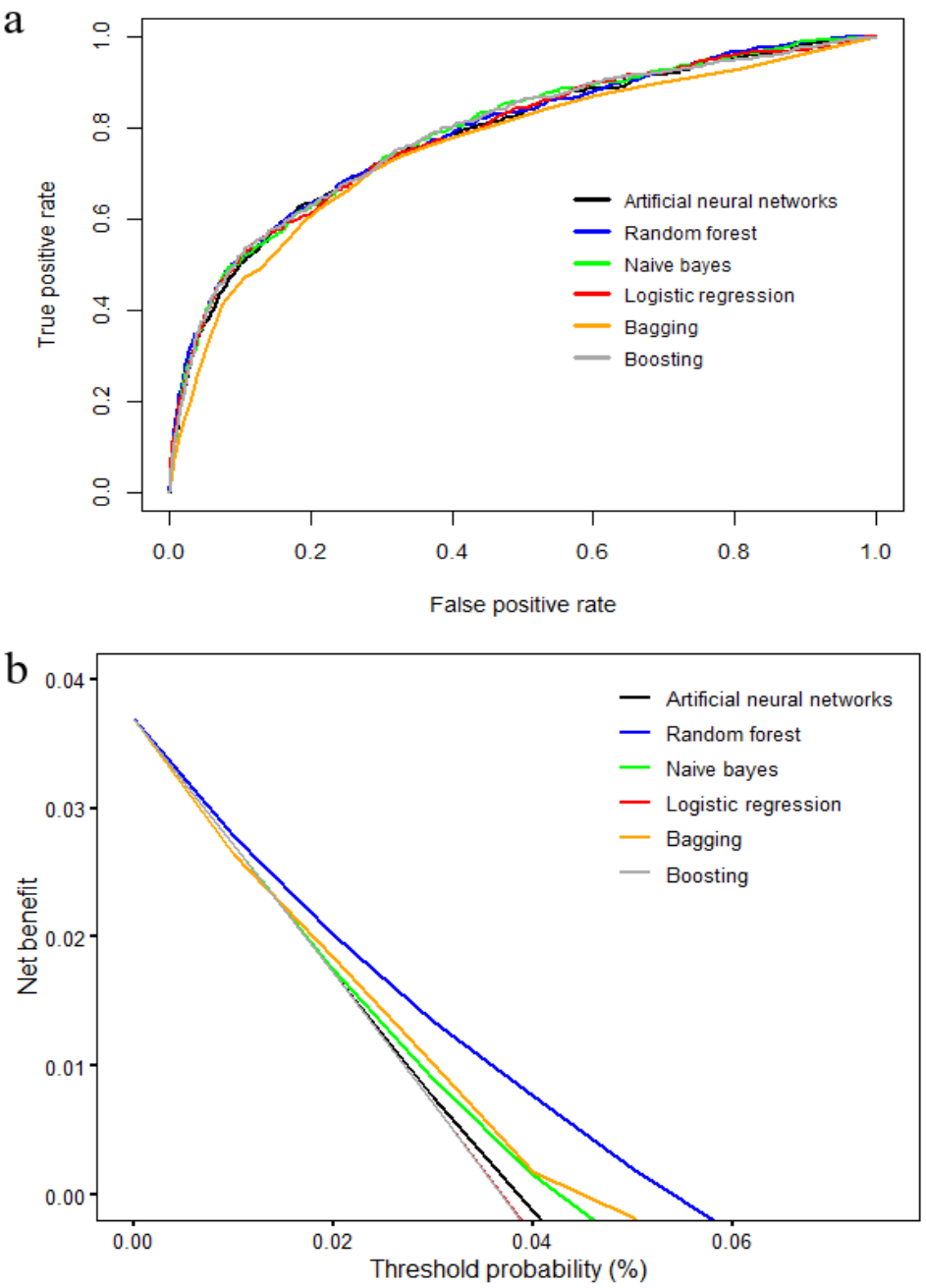


Figure 5.4: Prediction ability of perinatal deaths comparing different machine learning models in the test set: [a] Receiver-operating-characteristics (ROC) curves. The corresponding values of the area under the receiver-operating-characteristics curve (AUC) for each model are in Table 5.3. [b] Decision curve analysis. The net benefit of the machine learning models (except for boosting) is larger over a range of threshold probability values compared to that of the logistic regression model

With regard to the number of actual and predicted outcomes (Table 5.4), all models correctly predicted perinatal deaths by more than half of 468 deaths in the testing set. The numbers of correct classification were higher in the ANN 280 (59.8%) and NB 267 (57.1%), followed by the logistic regression model 261 (55.8%) and Bagging 260 (55.6%). The decision curve analysis (Figure 5.4b) demonstrated that the net benefit of the RF model surpassed that of other machine learning models, including logistic regression for all threshold values, indicating that the RF model is more superior in predicting the risk of perinatal deaths in this cohort. The accuracy of the RF model was 0.87, 95%CI (0.86, 0.87), compared to 0.87, 95%CI (0.87, 0.88) for boosting and 0.86, 95%CI (0.85, 0.86) for the logistic regression model (Table 5.3). Furthermore, other ML models (except for boosting) demonstrated high net benefit over a range of threshold probability values relative to that of the logistic regression model. Also, the random forest model had a superior net benefit over all models (Figure 5.4b).

Table 5.4: The number of actual and predicted outcomes of prediction models in the test set

Prediction Model	**Classification**	**Perinatal status**	
		Alive	**Died**
	Actual number of events	12227	468
Logistic Regression	Correctly predicted outcome	10627	261
	Incorrectly predicted outcome	1600	207
Artificial Neural Network	Correctly predicted outcome	10225	280
	Incorrectly predicted outcome	2002	188
Random Forests	Correctly predicted outcome	10774	251
	Incorrectly predicted outcome	1453	217
Naïve Bayes	Correctly predicted outcome	10386	267
	Incorrectly predicted outcome	1841	201
Bagging	Correctly predicted outcome	10175	260
	Incorrectly predicted outcome	2052	208
Boosting	Correctly predicted outcome	10852	252
	Incorrectly predicted outcome	1375	216

5.4 Discussion

In this study, the perinatal death was predicted using five ML models (Artificial Neural Networks, Random Forests, Naïve Bayes, Bagging, and Boosting). There were no differences in the predictive performance between ML models except for bagging, which had a lower predictive performance. The artificial neural networks and naïve Bayes had higher sensitivity compared to the logistic regression, and other ML models. Specificity for all models was high, mainly due to the low prevalence of perinatal deaths in this cohort. Additionally, results from the decision curve analysis revealed that the ML models (except for boosting) had a higher net benefit over a range of threshold probability values compared to the logistic regression model, indicating high accuracy. The random forest model demonstrated a superior net benefit over other models.

In the present study, maternal characteristics before and during pregnancy, pregnancy history, and paternal characteristics identified pregnancies at high risk of experiencing adverse perinatal outcomes that might need close clinical follow-ups. It is worth noting here that paternal age and education level were highly predictive of perinatal death more than the known pregnancy-related conditions or complications such as prematurity. Previous literature shows that paternal characteristics, particularly advanced paternal age, increase the risk of adverse perinatal outcomes, such as low birth weight, prematurity, small for gestational age, and low Apgar scores (Khandwala et al., 2018; Meng & Groth, 2018; Tough et al., 2003), despite conflicting evidence from other studies (Hurley & DeFranco, 2017). Furthermore, studies using data from the KCMC Medical Birth Registry (same data source to the current study) focused on modelling the association between maternal and pregnancy-related characteristics and complications during pregnancy and childbirth with the risk of adverse perinatal outcomes (Chuwa et al., 2017; Isaksen et al., 2015; Mitao et al., 2016; Mahande et al., 2013a,b; Mmbaga et al., 2011) but ignored paternal characteristics. Despite

challenges in male involvement in pregnancy and childbirth in Tanzania (Gibore & Bali, 2020; Peneza & Maluka, 2018), their participation is critical to improving maternal and child health outcomes.

On top of clinicians' judgment, previous investigators applied standard regression models in prediction of risk for adverse perinatal outcomes, particularly perinatal death (Allanson et al., 2016; Ali et al., 2014; Carvalho et al., 2020; Getiye & Fantahun, 2017; Habimana-Kabano et al., 2015; Hinderaker et al., 2003; Mahande et al., 2013a; Mutsaerts et al., 2014; Mpembeni et al., 2014; Nankabirwa et al., 2011; Ouyang et al., 2013; Unterscheider et al., 2014; Vogel et al., 2014). We found no differences in the predictive performance of the ML models, except for bagging, which had lower predictive capacity. The sensitivity of the machine learning models was also almost comparable to that of logistic regression, which indicates that both models correctly classified perinatal deaths. Our finding is consistent with a recent systematic review that showed no performance benefit of ML models over logistic regression for the prediction of clinical outcomes (Christodoulou et al., 2019). The possible explanation for lack of differences in the performance between the compared models could be attributed to the low proportion of outcome and exposures in this study, as well as data quality and recording challenges inherent in registry-based studies.

In contrast, some previous investigators have demonstrated that ML models offer better predictions of clinical or adverse pregnancy outcomes compared to classical regression models (Hoodbhoy et al., 2019; Houweling et al., 2019; Lee et al., 2019a; Muktan et al., 2019). The application of ML models may improve the classification of adverse events occurring during the perinatal period and, therefore, assist in triaging and provision of close clinical follow-up for women at high-risk. Other studies also provide evidence of improved prediction of under-five and neonatal mortality (Houweling et al., 2019; Hoodbhoy et al., 2019; Muktan et al., 2019;

Nasejje & Mwambi, 2017) using ML models. The utility of these models may, therefore, improve the prediction of adverse pregnancy outcomes as opposed to standard regression models.

In this study, the decision curve analysis that accounts for the impact of false-negative and false-positive misclassification errors showed superior predictive performance of the ML approaches over the logistic regression model. This demonstrates a higher net benefit for the prediction of perinatal deaths. The higher net benefit in the prediction ability of the ML approaches has also been documented elsewhere (Goto et al., 2019; Raita et al., 2019). This is because ML approaches can incorporate the high order nonlinear interactions between predictors, which cannot be addressed by traditional modeling approaches, including the logistic regression model. Furthermore, the use of cross-validation is also known to reduce potential overfitting in ML models. It is important to note that ML approaches are, to a large extent, non-parametric as opposed to the logistic regression model that relies on strong distributional assumptions.

The strength of this study is that it is the first to apply modern ML approaches to predict perinatal deaths, particularly in Tanzania and to a large extent sub-Saharan Africa, compared with the classical logistic regression model. Our study demonstrated that ML models might be used to improve the prediction of perinatal deaths and triage of women at risk. We also used the SMOTE balancing technique to avoid the bias of the model toward skewed data (reduce over-fitting), hence improving the prediction accuracy of the ML algorithm (Chawla, 2009; Hoodbhoy et al., 2019; Johnson & Khoshgoftaar, 2019). However, SMOTE is not very effective for high dimensional data (Blagus & Lusa, 2013b,a). Our study also had some limitations that are worth considering when interpreting the results. Firstly, we excluded observations with missing values in both the outcome and exposures from the analysis, a problem inherent in cohort studies, including birth registries,

which may lead to under-estimation of the proportion of perinatal death. Two excluded variables (maternal BMI and HIV status) have been associated with perinatal and under-five deaths (Hug et al., 2019; Nijkamp et al., 2017; UNICEF et al., 2020); hence their exclusion might increase the risk of residual confounding bias. The effect of exclusion of these two variables and missing values to predict perinatal deaths remains unquantified.

Secondly, selection/referral bias is a common problem in hospital-based studies, which affects the generalization of findings to the general population. This might also be the case in the present study. However, our findings might reflect a similar setting in Tanzania and probably in other sub-Saharan African countries. Thirdly, the KCMC Medical Birth Registry cohort only captures perinatal deaths occurring in the health facility (KCMC hospital), which may underestimate the observed perinatal deaths in the wider population. Currently, the hospital has no mechanisms to follow-up the birth outcomes from deliveries that occur at home and post-discharge outcomes of the babies after mothers are discharged from the hospital within the first week, especially within the KCMC hospital catchment area. Future extensions include ways of handling missing values before applying the machine learning algorithms to predict perinatal death and other adverse pregnancy outcomes.

5.5 Conclusion

The ML models (except for bagging) performed equally with the logistic regression model to predict perinatal deaths using maternal, paternal, and obstetric factors in this cohort. The ML models, however, have a higher net benefit, demonstrating superiority in the prediction of perinatal death. Furthermore, the random forest model also demonstrated superior performance over other ML models. These models are a useful and alternative strategy over the standard logistic regression model to predict perinatal deaths, considering the richness of the medical birth

registries. Moreover, the ML models are capable of handling many predictors at the same time, which is crucial in capturing multiple risk factors for adverse perinatal outcomes such as perinatal deaths. The application of ML models may, therefore, increase the prediction ability of adverse perinatal outcomes and thereby helping in triage women most at risk.

Chapter 6

Predictors of singleton preterm birth using multinomial regression models accounting for missing data: a birth registry-based cohort study in northern Tanzania

6.1 Introduction

Every year, an estimated 15 million babies (11%) are born preterm (before 37 completed weeks of gestation) globally (Chawanpaiboon et al., 2019; World Health Organization, 2020), majority (81.1%) of these occurs in Asia and sub-Saharan Africa (SSA) (Chawanpaiboon et al., 2019). The rates of preterm birth in SSA are notably high in Nigeria (6.9%), Ethiopia (12.0%), and Tanzania (16.6%) (Chawanpaiboon et al., 2019). Tanzania ranks the tenth country with the highest preterm birth rates in the world, and shares a 2.2% of the global proportion of all preterm births (Chawanpaiboon et al., 2019). The country specific estimates shows that the proportion of preterm birth ranged between 12-13% in Mwanza region

(Watson-Jones et al., 2007; Mahande et al., 2013b; Temu et al., 2016; Rugaimukam et al., 2017) to as high as 24% among HIV infected women in Dar es Salaam (Zack et al., 2014).

Preterm birth is a syndrome with a variety of causes, which can be classified into two broad clinical sub-types: spontaneous preterm birth (spontaneous onset of labour or following prelabour premature rupture of membranes) and provider-initiated preterm birth (induction of labor or elective caesarean birth before 37 completed weeks of gestation for maternal or fetal indications, both "urgent" or "discretionary", or other non-medical reasons) (Goldenberg et al., 2008; Blencowe et al., 2013; Quinn et al., 2016; Phillips et al., 2017; World Health Organization, 2020).

A higher risk of preterm birth is reported among women with a history of preterm delivery, those with low ($\leq$24) or high maternal age ($\geq$40), short inter-pregnancy intervals ($<$24 months), low maternal body mass index (BMI), multiple pregnancies, maternal infections such as urinary tract infections, malaria, bacterial vaginosis, HIV and syphilis and those with inadequate ($<$4) ANC visits (Blencowe et al., 2013; van den Broek et al., 2014; Mahande & Mahande, 2016; Mahande & Obure, 2016; Temu et al., 2016; Fuchs et al., 2018). Stress and excessive physical work or long times spent standing, drug abuse such as smoking and excessive alcohol consumption, sex of the child (more among males compared to females), hypertensive disorders of pregnancy such as pre-eclampsia or eclampsia, placental abruption, cholestasis, fetal distress, fetal growth restriction, small for gestational age (a birth weight below the 10th percentile for the gestational age), and early induction of labor or cesarean birth (before 39 completed weeks of gestation) whether for medical or non-medical reasons also increases the risk of preterm birth (Blencowe et al., 2013; Temu et al., 2016; van Zijl et al., 2016; Yang et al., 2016; Teoh et al., 2018; World Health Organization, 2020).

Globally, preterm birth is a leading cause of deaths among children under five years of age (Liu et al., 2016; Quinn et al., 2016; Chawanpaiboon et al., 2019; World Health Organization, 2020). SSA is one of the regions with the highest under five deaths in the world (Liu et al., 2016; UNICEF et al., 2020). In 2018, preterm birth complications accounted for 18% of death of children under the age of five and 35% of all newborn deaths globally (UNICEF et al., 2019). Preterm birth also increases the risk of babies dying from other causes, especially neonatal infections (Blencowe et al., 2013). Despite modern advances in obstetric and neonatal management, the rate of preterm birth are on the rise in both low-, middle- and high-income countries (Georgiou et al., 2015; Purisch & Gyamfi-Bannerman, 2017; Chawanpaiboon et al., 2019; World Health Organization, 2020), while in many low- and middle-income countries, preterm newborns are reported to die because of a lack of adequate newborn care (Chawanpaiboon et al., 2019).

Despite a substantial progress in improving child survival since 1990 (You et al., 2015; Chawanpaiboon et al., 2019), preterm birth remains a crucial issue in child mortality and improving quality of maternal and newborn care (Chawanpaiboon et al., 2019). To increase child survival and reduce preterm birth complications, the World Health Organization (WHO) recommends essential care during childbirth and postnatal period for every mother and baby (i.e. routine practice for the safe childbirth before, during and after birth), provision of antenatal steroid injections, magnesium sulfate for prevention of cerebral palsy in the infant and child, kangaroo mother care, and antibiotics to treat newborn infections (World Health Organization, 2015, 2020). Tanzania has also adopted these strategies (Ministry of Health and Social Welfare [MoHSW], 2015; MoHCDGEC, 2016) and is one of the five countries where WHO implements a clinical trial on the immediate kangaroo mother care (KMC) for preterm and babies weighing <2000 grams (MoHCDGEC, 2016; World Health Organization, 2020).

Epidemiologists are often interested in estimating the risk of adverse events originally measured on an interval scale (such as gestational age in weeks), but they often choose to divide the outcome into two or more categories in order to compute an estimate of effect (risk or odds ratio) (Ananth & Kleinbaum, 1997). In this study, we applied the multinomial logistic regression models, to show the effect of covariates on several preterm birth categories (Purisch & Gyamfi-Bannerman, 2017; World Health Organization, 2020) to avoid the bias that might be introduced by performing a binary analysis. A number of previous studies to assess predictors of preterm birth collapsed all preterm birth categories and performed a binary regression analysis (Ahankari et al., 2001; Grantz et al., 2015; Laughon et al., 2014; Mahande et al., 2013a; Malacova et al., 2018; Rugaimukam et al., 2017; Su et al., 2018; Teoh et al., 2018; van den Broek et al., 2014; Zack et al., 2014). This may introduce potential bias in estimating the effect of covariates on the risk of preterm birth due to a loss of information resulting from collapsing these categories. For a more focused care in the high-risk pregnancies, it is essential to estimate the risk factors for preterm birth, which may differ by the gestational age at birth.

Furthermore, missing data are common in epidemiological and clinical research (Sterne et al., 2009). Ignoring missing values in the analysis of such data potentially produces biased parameter estimates (Ibrahim & Molenberghs, 2009; Sterne et al., 2009; Kombo et al., 2017; Pedersen et al., 2017). Sterne et al. (2009), further indicated that "missing data in several variables often leads to exclusion of a substantial proportion of the original sample, which in turn causes a substantial loss of precision and power". Therefore, data analysis in this study accounted for missing data, for more precise parameter estimates.

6.2 Methods

6.2.1 Data source

We utilized secondary birth registry data from a prospective cohort of women who delivered singletons in the Kilimanjaro Christian Medical Center (KCMC) between the years 2000-2015. A detailed description of the KCMC Medical birth registry is also available elsewhere (Bergsjo et al., 2007; Mmbaga et al., 2012a; Mahande, 2015; Chuwa et al., 2017; Mvunta et al., 2019; Mboya et al., 2020b). Briefly, KCMC is one of the four zonal referral hospitals in the country and is located in the Moshi municipality, Kilimanjaro region, northern Tanzania. The centre primarily receives deliveries of women from the nearby communities, but also referral cases from within and outside the region. On average, the hospital has approximately 4000 deliveries per year (Mitao et al., 2016; Chuwa et al., 2017; Mvunta et al., 2019).

The study population in this study was singleton deliveries for women of reproductive age (15-49 years) recorded in the KCMC birth registry between 2000-2015, a total of 55,003 deliveries from 43,084 mothers. We excluded 3,316 multiple deliveries, 49 records missing hospital numbers (i.e. unique identification number used to link mothers and their subsequent births), 791 observations with a mismatch between dates of births of children from the same mother or were of unknown sequence (i.e. whether a singleton or multiple births), and 6,730 deliveries with gestational age $<$20 weeks and $>$42 weeks. Data was, therefore, analyzed for 44,117 deliveries born from 35,871 mothers (Figure 6.1).

6.2.2 Study variables and variable definitions

The response variable was preterm birth, defined as any birth before 37 completed weeks of gestation and further categorized based on gestational age as $<$28 weeks (extremely preterm), $[28, 32)$ weeks (very preterm), $[32, 37)$ weeks (moderate to late preterm), and $\geq$37 weeks (term) for a full-term pregnancy (World Health

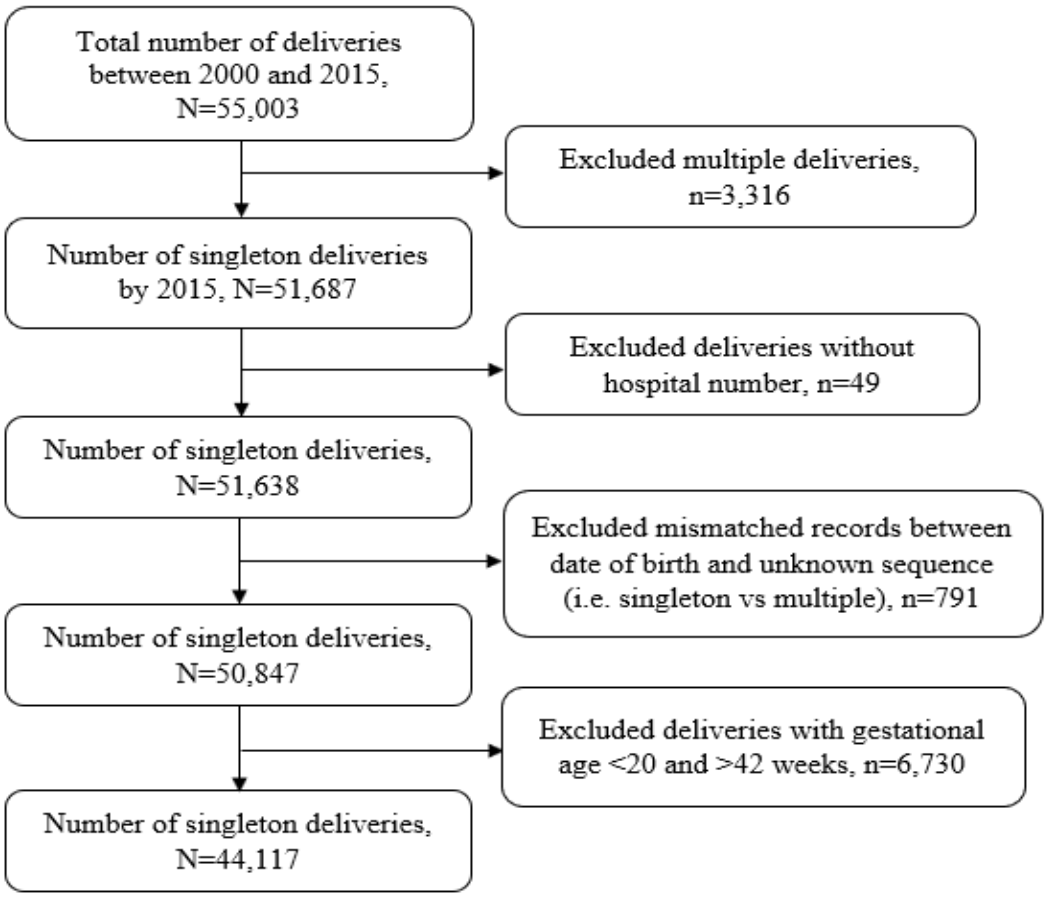


Figure 6.1: Flow chart showing the number of deliveries analyzed in this study. Data from the KCMC Medical Birth registry, 2000-2015.

Organization, 2020). Gestational age was estimated from the date of last menstrual period of the mother and recorded in completed weeks (Mahande et al., 2013b).

Independent variables included maternal background characteristics, particularly age categories (15-19, 20-24, 25-34, 35-39 and 40+) in years, area of residence (rural vs urban), education level (none, primary, secondary and higher), marital status (single, married and widow/ divorced), occupation (unemployed, employed and others), parity (primipara vs multipara (para 2-6)), referral status (referred for delivery or not), number of antenatal care visits (<4 and ≥ 4 visits), and body mass index (underweight [<18.5 Kg/m^2], normal weight [18.5–24.9 Kg/m^2], overweight [25–29.9 Kg/m^2], and obese [≥ 30 Kg/m^2]). Maternal health before and during

pregnancy included, alcohol consumption during pregnancy, maternal anemia, malaria, systemic infections/sepsis and pre-eclampsia/eclampsia (all categorized as binary, yes/no). Maternal HIV status was categorized as positive or negative. Complications during pregnancy and delivery included premature rapture of the membranes (PROM), postpartum hemorrhage (PPH), placenta previa and placenta abruption also categorized as binary, yes/no, with "yes" indicating the occurrence of these outcomes. Newborn characteristics included sex (male vs female), perinatal status (dead if experienced stillbirth/early neonatal death vs alive) (Mboya et al., 2020b), and low birth weight (LBW) defined as an absolute infant birth weight of $<$2500g regardless of gestational age at birth (World Health Organization, 2014b; Cutland et al., 2017).

6.2.3 Statistical and computational analysis

Data were analyzed using STATA version 15.1 (StataCorp LLC, College Station, Texas, USA). The primary unity of analysis was singleton deliveries for women recorded in the KCMC Medical Birth Registry between the years 2000 and 2015. We summarized numeric variables using means and standard deviations, and categorical variables using frequencies and percentages. The Chi-square test was used to compare the proportion of preterm birth by participants characteristics. We used multinomial logistic regression models to determine the predictors of preterm birth as opposed to previous studies (Mahande et al., 2013b; Laughon et al., 2014; van den Broek et al., 2014; Zack et al., 2014; Grantz et al., 2015; Lu et al., 2015a; Ahankari et al., 2001; Rugaimukam et al., 2017; Malacova et al., 2018; Su et al., 2018; Teoh et al., 2018) that performed a binary regression analysis.

The multinomial/polytomous regression model is an extension of the logistic model for binary responses to accommodate multinomial responses which does not have any restrictions on the ordinality of the response (Ananth & Kleinbaum, 1997). Let Y_i denote a nominal response variable for the ith subject, and $Y_i = c$ (the

response variable occuring in category c), while $Pr(Y_i)$ defines the probability that $Y_i = c$. The multinomial logit model can be written as

$$P_{ic} = Pr(Y_i = c|X_{ij}) = \frac{\exp(\eta_{ic})}{1 + \sum_{c=2}^{C} \exp(\eta_{ic})} \quad \text{for} \quad c = 2, 3, \ldots, C \tag{6.1}$$

$$P_{i1} = Pr(Y_i = 1|X_i) = \frac{1}{1 + \sum_{c=2}^{C} \exp(\eta_{ic})} \tag{6.2}$$

A nominal model to allow for any possible set of $c - 1$ response categories is written as

$$P_{ic} = \frac{\exp(\eta_{ic})}{\sum_{c=1}^{C} \exp(\eta_{ic})} \quad \text{for} \quad c = 1, 2, \ldots, C \tag{6.3}$$

where the multinomial logit $\eta_{ic} = X'_{ic}\beta_c$. In this model, all of the effects β_c vary across categories $(c = 1, 2, \ldots, C)$ and makes comparisons to a reference category compared to the ordinal regression model that uses cumulative comparisons of the categories (Hedeker, 2008). We used robust standard errors adjusted for clusters to account for nested observations/deliveries within mothers.

We would like to indicate here that we performed preliminary analysis using the binary and ordinal logistic regression models. There were a couple of variables that did not satisfy the proportional odds (PO) assumption, hence the ordinal logistic regression model could not be used. The close alternative model that relaxes the PO assumption are the generalized ordered logistic regression models. However, we encountered a non-convergence problem, especially with four preterm birth categories and appropriate interpretation of results. For instance, the order of gestational age categories is <28 weeks (extremely preterm), $[28, 32)$ weeks (very preterm), $[32, 37)$ weeks (moderate to late preterm), and 37+ weeks (term/normal). Assuming the variable is coded as 0 to 3 (with 0 being term birth), the first panel of coefficients will be interpreted as; 0 vs. 1+2+3, then 0+1 vs 2+3 etc (Williams, 2016). This will imply modeling the probability of delivering at a normal gestational age

(category 0) compared to preterm (categories 1-3), probability of delivering term and very preterm vs other preterm categories, etc. Similar interpretations will apply even if preterm birth is coded from extremely preterm (0) to term (3). Such interpretation could be somehow misleading given the nature of this outcome and may not be appealing to clinicians or public health practitioners. Nevertheless, the choice of regression models often depends on the research question one would like to address. In this study, the choice of multinomial regression model was relevant to determine preterm birth predictors across different preterm birth categories, other than performing a binary or an ordinal regression analysis.

As previously indicated, data analysis in this study considered missing values in the covariates. A description of how missing data were imputed is also reported in (Mboya et al., 2020b). Data were imputed using a multiple imputation technique, which is a commonly used method to deal with missing data, which accounts for the uncertainty associated with missing data (Sterne et al., 2009; Jakobsen et al., 2017; Pedersen et al., 2017). We assumed the data were missing at random (MAR) where the probability of data being missing does not depend on the unobserved data, conditional on the observed data (Ibrahim & Molenberghs, 2009; Sterne et al., 2009; Kombo et al., 2017; Pedersen et al., 2017); hence the variables in the dataset were used to predict missingness (Mboya et al., 2020b). We also assumed a nonmonotone pattern of missingness in which some subject values were observed again after a missing value occurs (Ibrahim & Molenberghs, 2009; Mboya et al., 2020b; Jakobsen et al., 2017). Under a nonmonotone pattern of missingness, it is recommended to use chained equations, which goes with several names such as the Markov chain Monte Carlo (MCMC), and the fully conditional specification (FCS), to impute missing values (Liu & De, 2015b; Pedersen et al., 2017; Jakobsen et al., 2017; Van Buuren et al., 2006a; StataCorp, 2017; Azur et al., 2011).

Furthermore, the FCS method allows imputation of all types of variables

simultaneously, namely some continuous and other categorical. The FCS algorithm is described in Section 3.4.3. Detailed descriptions on implementation of the FCS/MICE algorithm in STATA is well-presented elsewhere (StataCorp, 2017; Royston et al., 2011). Maternal age and education level were imputed as ordinal variables, while maternal occupation, marital status, and BMI (because normal weight (18.5–24.9 Kg/m^2) was a reference category) as multinomial variable (Mboya et al., 2020b). The rest of the variables were binary, and so imputed using the binomial distribution. Preterm birth (the outcome in this study), parity, pre-eclampsia/eclampsia, anemia, malaria, systemic infections/sepsis, PROM, PPH, abruption placenta, placenta previa, and year of birth did not contain any missing values, hence used as auxiliary variables in the imputation model. The imputation model generated 20 imputed datasets after 500 iterations (imputation cycles). A random seed of 5000 was specified for replication of imputation results each time a multiple imputation analysis is performed (Jakobsen et al., 2017).

We developed a multivariable analysis model by including all covariates in the multinomial logit analysis model (StataCorp, 2017), with standard errors adjusted for clusters (i.e., deliveries nested within mothers). We then performed stepwise regression, in which variables with $p<0.1$ or $p<10\%$ were retained in the model. The next steps entailed performing a series of adjusted analysis to test the effect of retaining and dropping variables in the multivariable model. Variables in the final model were evaluated at p-value<0.05 level of statistical significance. We used AIC to compare model performance and non-nested models (Vrieze, 2012), and Likelihood ratio test to compare nested models. After the imputation of missing values, we estimated parameter estimates adjusting for the variability between imputations (StataCorp, 2017; Royston et al., 2011). Before the analysis of imputed data, we firstly performed complete case analysis using multivariable multinomial regression model. The final model from this analysis was then compared to those from the multiply imputed dataset. We followed the recommendations suggested

by Sterne et al. (2009) for reporting and analysis of missing data.

6.3 Results

6.3.1 Maternal background characteristics by gestational age categories

The overall proportion of preterm birth in this study was 12.8%, of which 9.8% children were born at $[32, 37)$ weeks (moderate to late preterm), 1.6% at $[28, 32)$ weeks (very preterm), and 0.4% at <28 weeks (extremely preterm) of gestation. The proportions of preterm birth differed significantly by maternal background and obstetric care characteristics (Tables 6.1 and 6.2, respectively). Among adolescent mothers (15-19 years), 12.3% delivered at $[32, 37)$ weeks and 1.8% at $[28, 32)$ weeks of gestation, which is almost similar to that among older mothers (40+ years). The proportion of women who delivered at $[32, 37)$ weeks of gestation was 10.8% among rural residents, 11.0% among those with primary education level, 9.6% among those employed, and 9.6% among mothers who were married (Table 6.1).

6.3.2 Diseases and complications during pregnancy and delivery by gestational age categories

The diseases and complications during pregnancy and delivery by gestational age categories are shown in (Table 6.2). There were statistically significant differences in the proportion of preterm birth categories by diseases and complications during pregnancy and delivery except for anaemia, infections/sepsis and child's sex. Significantly higher proportion of deliveries born at $[32, 37)$ weeks of gestation was among mothers who experienced placenta previa (39.6%), abruption placenta (37.3%), delivered LBW baby (37.1%), perinatal death (28.1%), pre-eclampsia/eclampsia mothers (24.3%), PROM (18.9%) with <4 ANC visits (17.0%), and postpartum hemorrhage (14.8%). Also, the proportion of deliveries born at $[28, 32)$ weeks of gestation was significantly higher among mothers with pre-eclampsia/eclampsia (6.2%), abruption placenta (10.0%), placenta previa

Table 6.1: Maternal background characteristics by gestational age categories (N=44,117)

Characteristics	Total (%)	Gestational age at birth				p-value
		≥37	32-<37	28-<32	<28	
Mother's age groups (years)*						<0.001
15-19	3637 (8.3)	3101 (85.3)	447 (12.3)	67 (1.8)	22 (0.6)	
20-24	11113 (25.2)	9797 (88.2)	1108 (10.0)	171 (1.5)	37 (0.3)	
25-34	22767 (51.7)	20321 (89.3)	2031 (8.9)	342 (1.5)	73 (0.3)	
35-39	5262 (12.0)	4576 (87.0)	556 (10.6)	110 (2.1)	20 (0.4)	
40+	1267 (2.9)	1080 (85.2)	158 (12.5)	21 (1.7)	8 (0.6)	
Current area of residence*						<0.001
Rural	18083 (41.1)	15690 (86.8)	1951 (10.8)	360 (2.0)	82 (0.5)	
Urban	25935 (58.9)	23155 (89.3)	2349 (9.1)	352 (1.4)	79 (0.3)	
Mother's highest education level*						<0.001
None	640 (1.5)	544 (85.0)	74 (11.6)	19 (3.0)	3 (0.5)	
Primary	24038 (54.6)	20857 (86.8)	2654 (11.0)	426 (1.8)	101 (0.4)	
Secondary	5406 (12.3)	4752 (87.9)	540 (10.0)	102 (1.9)	12 (0.2)	
Higher	13967 (31.7)	12730 (91.1)	1028 (7.4)	164 (1.2)	45 (0.3)	
Occupation*						0.04
Unemployed	9617 (21.9)	8397 (87.3)	1020 (10.6)	161 (1.7)	39 (0.4)	
Employed	31233 (71.2)	27618 (88.4)	2999 (9.6)	502 (1.6)	114 (0.4)	
Others	3023 (6.9)	2701 (89.3)	269 (8.9)	45 (1.5)	8 (0.3)	
Marital Status*						<0.001
Single	5202 (11.8)	4490 (86.3)	572 (11.0)	112 (2.2)	28 (0.5)	
Married	38697 (88.0)	34279 (88.6)	3698 (9.6)	589 (1.5)	131 (0.3)	
Widowed/Divorced	87 (0.2)	62 (71.3)	18 (20.7)	5 (5.7)	2 (2.3)	
Body mass index categories*						<0.001
Underweight (<18.5)	1582 (5.2)	1382 (87.4)	167 (10.6)	30 (1.9)	3 (0.2)	
Normal weight (18.5-24.9)	16417 (53.9)	14735 (89.8)	1439 (8.8)	201 (1.2)	42 (0.3)	
Overweight (25-29.9)	8510 (27.9)	7763 (91.2)	633 (7.4)	94 (1.1)	20 (0.2)	
Obese (≥30)	3947 (13.0)	3581 (90.7)	307 (7.8)	48 (1.2)	11 (0.3)	
Total (row %)		38933 (88.2%)	4309 (9.8%)	714 (1.6%)	161 (0.4%)	

* Variables with missing values.

(16.7%), experienced perinatal death (14.0%), and those who delivered a LBW baby (12.2%).

6.3.3 Distribution of missing values

Percentage distribution of missing values in this study are summarized in Table 6.3. Maternal BMI (31.0%) and HIV status (23.5%) accounted for more than half (54.5%) of all missing values. The proportion of missing values was 3.7%, 1.7% and 1.2% for referral status, number of ANC visits and alcohol consumption during pregnancy, respectively. The rest of the variables had less than 1% of missing values.

Table 6.2: Diseases and complications during pregnancy and delivery by gestational age categories (N=44,117)

Characteristics	Total (%)	Gestational age at birth				p-value
		≥37	32-<37	28-<32	<28	
Pre-eclampsia/eclampsia						<0.001
No	42282 (95.8)	37674 (89.1)	3864 (9.1)	600 (1.4)	144 (0.3)	
Yes	1835 (4.2)	1259 (68.6)	445 (24.3)	114 (6.2)	17 (0.9)	
Anaemia						0.63
No	43427 (98.4)	38331 (88.3)	4238 (9.8)	699 (1.6)	159 (0.4)	
Yes	690 (1.6)	602 (87.2)	71 (10.3)	15 (2.2)	2 (0.3)	
Malaria						0.002
No	38145 (86.5)	33579 (88.0)	3785 (9.9)	637 (1.7)	144 (0.4)	
Yes	5972 (13.5)	5354 (89.7)	524 (8.8)	77 (1.3)	17 (0.3)	
Infections						0.37
No	43352 (98.3)	38244 (88.2)	4243 (9.8)	706 (1.6)	159 (0.4)	
Yes	765 (1.7)	689 (90.1)	66 (8.6)	8 (1.0)	2 (0.3)	
HIV Status*						0.003
Negative	32000 (94.8)	28367 (88.6)	3047 (9.5)	472 (1.5)	114 (0.4)	
Positive	1769 (5.2)	1521 (86.0)	213 (12.0)	31 (1.8)	4 (0.2)	
Consumed alcohol during pregnancy*						<0.001
No	31287 (71.8)	27472 (87.8)	3150 (10.1)	543 (1.7)	122 (0.4)	
Yes	12292 (28.2)	10998 (89.5)	1099 (8.9)	158 (1.3)	37 (0.3)	
Number of ANC visits*						<0.001
≥4	29490 (68.0)	27489 (93.2)	1830 (6.2)	125 (0.4)	46 (0.2)	
<4	13884 (32.0)	10879 (78.4)	2366 (17.0)	540 (3.9)	99 (0.7)	
Parity						0.001
Primipara	35871 (81.3)	31599 (88.1)	3519 (9.8)	606 (1.7)	147 (0.4)	
Multipara	8246 (18.7)	7334 (88.9)	790 (9.6)	108 (1.3)	14 (0.2)	
PROM						<0.001
No	43157 (97.8)	38187 (88.5)	4128 (9.6)	681 (1.6)	161 (0.4)	
Yes	960 (2.2)	746 (77.7)	181 (18.9)	33 (3.4)	0 (0.0)	
PPH						<0.001
No	43874 (99.4)	38739 (88.3)	4273 (9.7)	702 (1.6)	160 (0.4)	
Yes	243 (0.6)	194 (79.8)	36 (14.8)	12 (4.9)	1 (0.4)	
Abruption placenta						<0.001
No	43967 (99.7)	38857 (88.4)	4253 (9.7)	699 (1.6)	158 (0.4)	
Yes	150 (0.3)	76 (50.7)	56 (37.3)	15 (10.0)	3 (2.0)	
Placenta previa						<0.001
No	44021 (99.8)	38891 (88.3)	4271 (9.7)	698 (1.6)	161 (0.4)	
Yes	96 (0.2)	42 (43.8)	38 (39.6)	16 (16.7)	0 (0.0)	
Perinatal status*						<0.001
Alive	42230 (96.0)	37868 (89.7)	3796 (9.0)	462 (1.1)	104 (0.2)	
Died	1780 (4.0)	975 (54.8)	500 (28.1)	250 (14.0)	55 (3.1)	
Birth weight*						<0.001
NBW	39202 (89.1)	36543 (93.2)	2500 (6.4)	107 (0.3)	52 (0.1)	
LBW	4801 (10.9)	2334 (48.6)	1779 (37.1)	585 (12.2)	103 (2.1)	
Sex of the baby*						0.48
Male	22684 (51.6)	20032 (88.3)	2216 (9.8)	349 (1.5)	87 (0.4)	
Female	21242 (48.4)	18743 (88.2)	2070 (9.7)	359 (1.7)	70 (0.3)	
Refereed for delivery*						<0.001
Yes	9610 (22.6)	7883 (82.0)	1382 (14.4)	278 (2.9)	67 (0.7)	
No	32878 (77.4)	29575 (90.0)	2807 (8.5)	409 (1.2)	87 (0.3)	
Total		38933 (88.2%)	4309 (9.8%)	714 (1.6%)	161 (0.4%)	

* Variables with missing values.

Table 6.3: Distribution of missing values, KCMC medical birth registry, 2000–2015 (N=44,117)

Variable	Frequency	Percent Missing
Body Mass Index (BMI)	13,661	31.0
HIV status	10,348	23.5
Referral status	1,629	3.7
Number of antenatal care visits	743	1.7
Consumed alcohol during pregnancy	538	1.2
Occupation	244	0.6
Sex of the child	191	0.4
Marital status	131	0.3
Birth weight of the child	114	0.3
Area of residence	107	0.2
Perinatal status	99	0.2
Maternal education level	71	0.2
Maternal age categories	66	0.2

6.3.4 Trends of preterm birth from 2000-2015

The proportion of moderate to late preterm (32 to <37) and very preterm (28 to <32) increased significantly over the years between 2000-2015 (Figure 6.2). The annual increase of PTB at [32,37) weeks of gestation was 22.2% (95%CI 12.2%, 32.1%, p<0.001) while for [28,32) weeks of gestation was 4.2% (95%CI 1.9%, 6.6%, p=0.002). Despite a slight increasing trend of extremely preterm birth (<28 weeks) deliveries, this increase was not statistically significant (p=0.37). However, further analysis of the trends in the very/extremely preterm birth (i.e., all deliveries at <32 weeks of gestation) revealed a significant annual increase of 4.6% (95%CI 2.2%, 7.0%, p=0.001). Regression analysis both before and after imputation of missing values, considered two preterm birth categories, i.e., <32 weeks (combined <28 and [28,32) weeks and [32,37] weeks) compared to ≥37 weeks, due to small sample size in the <28 category and increased statistical power to detect the observed effect.

6.3.5 Predictors of preterm birth

Due to a small number of deliveries 161 (0.4%) at <28 weeks of gestation recorded at the KCMC Medical birth registry between 2000 and 2015, we combined this

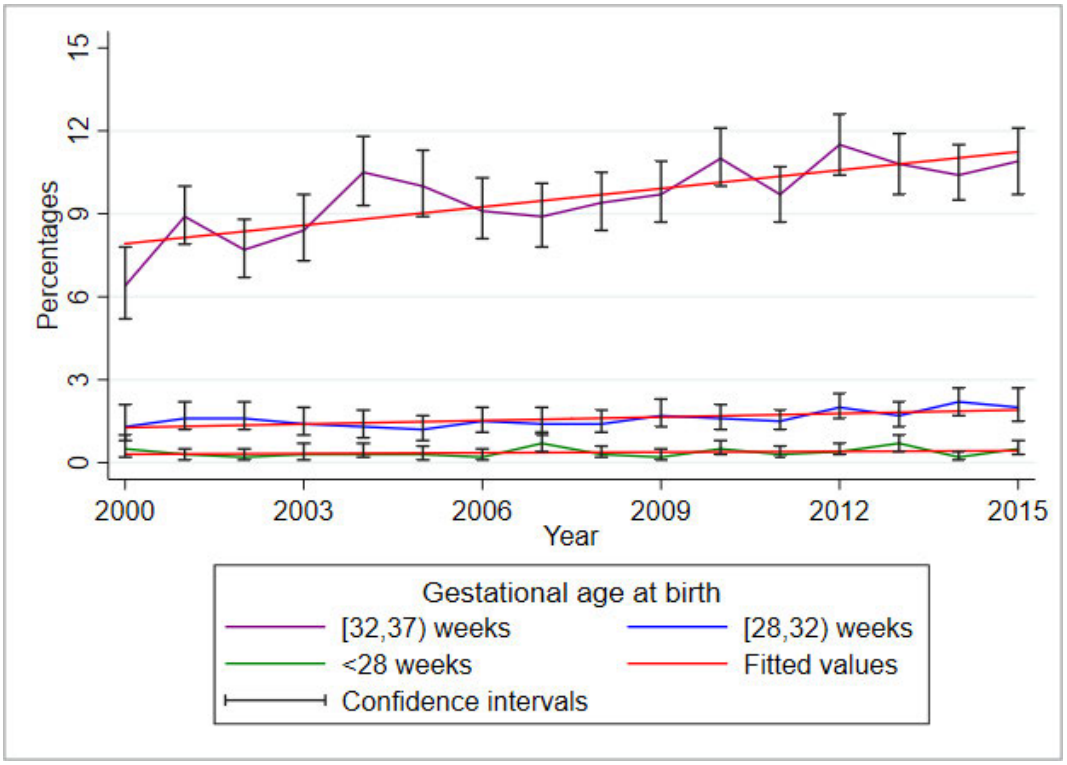


Figure 6.2: Trends of preterm birth from 2000-2015 in the KCMC Medical Birth Registry (N=44,117)

category with deliveries at [28,32) weeks of gestation, 714 (1.6%). This gives a total of 875 (2.0%) in the new <32 (very/extremely preterm) category. The collapsed categories increased statistical power and improved model performance, given a non-convergence problem of models with all three preterm birth categories.

6.3.5.1 Results before imputation of missing values

Findings from the adjusted analysis of the multinomial regression model before imputation of missing values are shown in Table 6.4. The standard errors are robust (adjusted) to clustering of deliveries within mothers. Higher odds of delivering at [32,37) weeks of gestation (moderate to late preterm) were among adolescent (15-19) mothers (OR=1.29, 95% 1.13, 1.48) and those aged 20-24 years (OR=1.17, 95%CI 1.07, 1.28) compared to those aged 25-34 years and those with primary education level (OR=1.28, 95%CI 1.17, 1.39) compared to higher education level. Also, mothers referred for delivery (OR=1.20, 95%CI 1.10, 1.31), with

pre-eclampsia/eclampsia (OR=1.88, 95%CI 1.63, 2.15), with inadequate (<4) ANC visits (OR=2.56, 95%CI 2.38, 2.75), experienced PROM (OR=1.83, 95%CI 1.51, 2.22), abruption placenta (OR=2.01, 95%CI 1.24 3.24), placenta previa (OR=4.90, 95%CI 2.73, 8.77), delivered through cesarean section (OR=1.16, 95%CI 1.07, 1.25), delivered a LBW baby (OR=8.05, 95%CI 7.41, 8.75), experienced perinatal death (OR=2.06, 95%CI 1.78, 2.37), and delivered a male child (OR=1.11, 95%CI 1.03, 1.19), compared to their respective reference levels had higher odds of delivering moderate to late preterm birth. Primiparous women were less likely to deliver moderate to late preterm (OR=0.89, 95%CI 0.80, 0.98). For every year increase, the odds of delivering at [32,37) weeks of gestation increased significantly by 2% (OR=1.02, 95%CI 1.01, 1.03).

Moreover, in the adjusted analysis, maternal age, referral status, pre-eclampsia/eclampsia, number of ANC visits, placenta previa, LBW, perinatal status, child's sex, and year of birth remained significantly associated with delivering at <32 weeks of gestation (very/extremely preterm). Notably, the odds of delivering at <32 of gestation were nearly forty times (OR=36.23, 95%CI 29.91, 43.89) among deliveries born with LBW compared to normal weight at birth. This is more than four times higher odds compared to the effect in the gestational age of [32,37) weeks. Mothers aged 15-19 years (OR=1.37, 95%CI 1.03, 1.81), referred for delivery (OR=1.30, 95%CI 1.08, 1.55), with pre-eclampsia/eclampsia (OR=1.51, 95%CI 1.18, 1.92), with inadequate (<4) ANC visits (OR=5.55, 95%CI 4.61, 6.69), experienced placenta previa (OR=8.68, 95%CI 3.75, 20.10), experienced perinatal death (OR=5.38, 95%CI 4.41, 6.56), and delivered male children (OR=1.22, 95%CI 1.04, 1.43) had higher odds of delivering very/extremely preterm birth (<32 weeks of gestation) as compared to their counterparts. Furthermore, for every year increase, the odds of delivering at <32 weeks of gestation increased significantly by 6% (OR=1.06, 95%CI 1.04, 1.09), which is three-times higher than the effect in the [32,37) weeks of gestation. These results demonstrate the advantage of the

Table 6.4: Adjusted analysis for predictors of preterm birth using multinomial regression model before imputation of missing values (N=41,271)

Characteristics	32-<37 vs. ≥37 weeks		<32 vs. ≥37 weeks	
	AOR† (SE‡)	95%CI	AOR† (SE‡)	95%CI
Mother's age groups (years)				
15-19	1.29 (0.09)	1.13,1.48***	1.37 (0.20)	1.03,1.81*
20-24	1.17 (0.05)	1.07,1.28***	1.11 (0.12)	0.91,1.37
25-34	1		1	
35-39	1.03 (0.06)	0.92,1.15	1.04 (0.13)	0.82,1.33
40+	1.13 (0.11)	0.93,1.38	0.84 (0.20)	0.53,1.33
Maternal highest education level				
None	1.15 (0.17)	0.85,1.54	1.43 (0.37)	0.86,2.38
Primary	1.28 (0.06)	1.17,1.39***	1.11 (0.11)	0.91,1.35
Secondary	1.11 (0.07)	0.98,1.26	0.97 (0.14)	0.74,1.28
Higher	1		1	
Referred for delivery (Yes)	1.20 (0.05)	1.10,1.31***	1.30 (0.12)	1.08,1.55**
Pre-eclampsia/eclampsia (Yes)	1.88 (0.13)	1.63,2.15***	1.51 (0.19)	1.18,1.92***
Number of ANC visits (<4)	2.56 (0.10)	2.38,2.75***	5.55 (0.53)	4.61,6.69***
Parity (Primipara)	0.89 (0.04)	0.80,0.98*	0.96 (0.11)	0.76,1.21
PROM (Yes)	1.83 (0.18)	1.51,2.22***	1.51 (0.35)	0.96,2.39
Abruption placenta (Yes)	2.01 (0.49)	1.24,3.24**	1.60 (0.57)	0.80,3.20
Placenta previa (Yes)	4.90 (1.46)	2.73,8.77***	8.68 (3.72)	3.75,20.10***
Delivery mode (CS)	1.16 (0.04)	1.07,1.25***	0.93 (0.08)	0.78,1.11
Birth weight (LBW)	8.05 (0.34)	7.41,8.75***	36.23 (3.55)	29.91,43.89***
Sex of the baby (Male)	1.11 (0.04)	1.03,1.19**	1.22 (0.10)	1.04,1.43*
Perinatal death (Yes)	2.06 (0.15)	1.78,2.37***	5.38 (0.55)	4.41,6.56***
Year	1.02 (0.00)	1.01,1.03***	1.06 (0.01)	1.04,1.09***

†**AOR**: Adjusted Odds Ratio, adjusted for maternal age groups (years), highest level of education, referral status, pre-eclampsia/eclampsia, number of ANC visits, parity, PROM, abruption placenta, placenta previa, delivery mode, child's birth weight, perinatal status and year of birth.
‡**SE**: Standard errors adjusted for clustering of deliveries within mothers.
* p<0.05, ** p<0.01, *** p<0.001.

multinomial regression as opposed to the simple binary regression models. We see that the effect of some covariates (LBW, inadequate ANC visits, placenta previa, and perinatal death) are more pronounced for the extreme preterm birth category than the moderately to late preterm birth category (Table 6.4).

6.3.5.2 Results after imputation of missing values

After imputation of missing values (in the covariates), the standard errors were relatively lower while the coefficients (odds ratios) (Table 6.5) were either lower or higher compared to those in the complete case analysis (Table 6.4). Results from the imputed data indicated significantly higher odds of moderately to late preterm

delivery (32 to <37 weeks) were among adolescent mothers aged 15-19 years (OR=1.29, 95%CI 1.13, 1.3479), aged 20-24 years (OR=1.15, 95%CI1.06, 1.26), with primary education level (OR=1.27, 95%CI 1.17, 1.39), and referred for delivery (OR=1.20, 95%CI 1.10, 1.30). Also, significantly higher odds of moderately to late preterm delivery were among mothers with pre-eclampsia/eclampsia (OR=1.86, 95%CI 1.62, 2.13), inadequate (<4) ANC visits (OR=2.56, 95%CI 2.38, 2.75), experienced PROM (OR=1.87, 95%CI 1.55, 2.26), abruption placenta (OR=1.98, 95%CI 1.23, 3.19), and placenta previa (OR=4.76, 95%CI 2.73, 8.28). Likewise, delivery through CS (OR=1.16, 95%CI 1.08, 1.25), delivering LBW baby (OR=8.09, 95%CI 7.45, 8.78), experiencing perinatal death (OR=2.10, 95%CI 1.83, 2.42), and delivering male children (OR=1.11, 95%CI 1.04, 1.20) were associated with higher odds of delivering moderately to late preterm. Primiparous women were less likely to deliver moderately to late preterm (OR=0.90, 95%CI 0.82, 0.99) compared to multiparous. For every one year increase, the odds of delivering moderately to late preterm increased significantly by 2% (OR=1.02, 95%CI 1.01, 1.03), Table 6.5.

Furthermore, after imputation of missing values the positive effect of PROM on very/extremely preterm birth (<32 weeks of gestation) is observed to be statistically significant (OR=1.63, 95%CI 1.06, 2.50) compared to results before imputation of missing values (Table 6.4). Significantly higher odds of very/extreme preterm birth was among mothers referred for delivery (OR=1.28, 95%CI 1.08, 1.52), with pre-eclampsia/eclampsia (OR=1.61, 95%CI 1.27, 1.03), inadequate (<4) ANC visits (OR=5.64, 95%CI 4.67, 6.80), experienced placenta previa (OR=8.07, 95%CI 3.61, 18.07), delivered LBW baby (OR=38.21, 95%CI 31.65, 46.14), experienced perinatal death (OR=5.29, 95%CI 4.37, 6.40), and delivered male children (OR=1.22, 95%CI 1.05, 1.43). Also, the odds of delivering very/ extreme preterm birth increased significantly by 4% (OR=1.04, 95%CI 1.02, 1.06) for every calendar year.

Table 6.5: Adjusted analysis for predictors of preterm birth using multinomial regression model after imputation of missing values (N=42,089)

Characteristics	32-<37 vs. ≥37 weeks		<32 vs. ≥37 weeks	
	AOR† (SE‡)	95%CI	AOR† (SE‡)	95%CI
Mother's age groups (years)				
15-19	1.29 (0.09)	1.13,1.47***	1.30 (0.18)	0.99,1.71
20-24	1.15 (0.05)	1.06,1.26**	1.13 (0.11)	0.93,1.38
25-34	1		1	
35-39	1.03 (0.06)	0.92,1.15	1.06 (0.13)	0.84,1.34
40+	1.11 (0.11)	0.91,1.34	0.92 (0.20)	0.60,1.40
Maternal highest education level				
None	1.11 (0.17)	0.82,1.49	1.44 (0.36)	0.88,2.34
Primary	1.27 (0.06)	1.17,1.39***	1.09 (0.10)	0.91,1.32
Secondary	1.10 (0.07)	0.98,1.25	1.00 (0.14)	0.77,1.32
Higher	1		1	
Referred for delivery (Yes)	1.20 (0.05)	1.10,1.30***	1.28 (0.11)	1.08,1.52**
Pre-eclampsia/eclampsia (Yes)	1.86 (0.13)	1.62,2.13***	1.61 (0.19)	1.27,2.03***
Number of ANC visits (<4)	2.56 (0.10)	2.38,2.75***	5.64 (0.54)	4.67,6.80***
Parity (Primipara)	0.90 (0.04)	0.82,0.99*	0.98 (0.11)	0.78,1.23
PROM (Yes)	1.87 (0.18)	1.55,2.26***	1.63 (0.36)	1.06,2.50*
Abruption placenta (Yes)	1.98 (0.48)	1.23,3.19**	1.46 (0.52)	0.73,2.93
Placenta previa (Yes)	4.76 (1.35)	2.73,8.28***	8.07 (3.32)	3.61,18.07***
Delivery mode (CS)	1.16 (0.04)	1.08,1.25***	0.91 (0.08)	0.77,1.08
Birth weight (LBW)	8.09 (0.34)	7.45,8.78***	38.21 (3.67)	31.65,46.14***
Perinatal death (Yes)	2.10 (0.15)	1.83,2.42***	5.29 (0.52)	4.37,6.40***
Sex of the baby (Male)	1.11 (0.04)	1.04,1.20**	1.22 (0.10)	1.05,1.43*
Year	1.02 (0.00)	1.01,1.03***	1.04 (0.01)	1.02,1.06***

†**AOR**: Adjusted Odds Ratio, adjusted for maternal age groups (years), highest level of education, referral status, pre-eclampsia/eclampsia, number of ANC visits, PROM, abruption placenta, placenta previa, delivery mode, child's birth weight, perinatal status and year of birth.
‡**SE**: Standard errors adjusted for clustering of deliveries within mothers.
* p<0.05, ** p<0.01, *** p<0.001.

6.4 Discussion

Globally, the trends of preterm birth rate has been increasing over time (Blencowe et al., 2013; Lu et al., 2015a; Chawanpaiboon et al., 2019; World Health Organization, 2020). Findings in the current study also revealed the rising trends of both moderate to late preterm (32 to <37 weeks of gestation) and very/extremely preterm birth (<32 weeks of gestation) between the years 2000-2015. A recent systematic review and modelling analysis revealed that Tanzania is among the top 10 countries (tenth position) with the highest preterm birth rate (16·6%) and contributed to 2.2% of the global preterm birth estimates (Chawanpaiboon et al.,

2019). Based on the estimates released seven years ago (2013) by Blencowe et al. (2013), Tanzania was not in the top 10 countries with the highest (>15%) preterm birth rates globally. By then, Malawi had the highest preterm birth rate (18%) in SSA and South East Asia (Blencowe et al., 2013; van den Broek et al., 2014).

Previous studies at the KCMC zonal referral hospital (Mahande et al., 2013b; Temu et al., 2016) and Bugando Medical Center in Mwanza region (Rugaimukam et al., 2017) reported the preterm birth rate of 14%; where (Mahande et al., 2013b) utilized cohort data between the years 2000-2008 while (Temu et al., 2016) and (Rugaimukam et al., 2017) conducted case-control studies. The rising trends and relatively high preterm birth rates in Tanzania are alarming, given the documented short- and long-term consequences, particularly an increased risk of recurrence in subsequent pregnancies, stillbirths, and neonatal mortality (Marchant et al., 2012; Mahande et al., 2013b; van den Broek et al., 2014; Liu et al., 2016; Phillips et al., 2017; Malacova et al., 2018; Vogel et al., 2018). In fact, mothers who experienced perinatal death in this study were more likely to deliver preterm. The effect of perinatal death almost doubled in the very/extremely preterm category.

Multiple imputation was performed to increase precision of parameter estimates, as it accounts for the uncertainty associated with missing data (Ibrahim & Molenberghs, 2009; Sterne et al., 2009; Jakobsen et al., 2017; Pedersen et al., 2017). After the imputation of missing values, the standard errors are relatively lower and coefficients (odds ratios) were either lower or higher than those in the complete case analysis. Although the direction of associations remained the same, precision of parameters estimates is increased after imputation of missing data. It has been reported that "multiple imputation provides unbiased and valid estimates of associations based on information from the available data – ie, yielding estimates similar to those calculated from full data" (Pedersen et al., 2017). Data analysts should consider accounting for missing data in their analysis using proper

techniques to reduce the bias associated with simple analysis (such as analyzing available or complete cases) that ignore missing values (Liu & De, 2015b; Jakobsen et al., 2017; Pedersen et al., 2017).

Results from the imputed data revealed that adolescent (15-19 years) mothers and mothers aged 20-24 years had higher odds of delivering moderately to late preterm births (32 to <37 weeks) as well as very/extremely preterm (<32 weeks though this association was not statistically significant) compared to mothers aged 25-34 years. Our findings are consistent with previous studies (van den Broek et al., 2014; Zack et al., 2014; Grantz et al., 2015; Lu et al., 2015a; Fuchs et al., 2018). Authors in these studies revealed that younger (<24 years) mothers are at increased risk of delivering preterm. A previous study in Canada indicated that women aged 20-24 years were more at risk of delivering spontaneous preterm birth (Fuchs et al., 2018). However, authors in this study did not include adolescent mothers. Data from the Tanzania Demographic and Health Survey 2015/16 revealed the rising trends of teenage childbearing (15-19 years) from 23% in 2010 to 27% in 2015/16 (MoHCDGEC [Tanzania Mainland] et al., 2016).

Younger age at first pregnancy is a public health concern due to an increased risk of complications during pregnancy and child birth as well as maternal and neonatal mortality (MoHCDGEC [Tanzania Mainland] et al., 2016; Fuchs et al., 2018). A systematic review and meta-analysis in SSA documented an association between adolescent child-bearing and an increased risk of low birth weight, pre-eclampsia/eclampsia, preterm birth and maternal and perinatal mortality (Grønvik & Fossgard Sandøy, 2018). Findings in this study suggests that interventions in Tanzania should emphasize on delayed age at first pregnancy and provision of adolescent and youth friendly sexual and reproductive health services (Kozuki et al., 2013; MoHCDGEC, 2016; Wado et al., 2019), for positive pregnancy experiences.

Mothers referred for delivery at the KCMC zonal referral hospital were more likely to deliver preterm compared to those who had self-referred (normal clinic attendance). Similar findings has been reported elsewhere(Grønvik & Fossgard Sandøy, 2018; Mboya et al., 2020b), where women referred for delivery are more likely to have more pregnancy-related complications such as pre-eclampsia, which increases the risk of preterm birth. Close clinical follow-up is recommended to this group of women during prenatal care to minimize pregnancy-related complications, such as preterm birth and associated consequences. Mothers with primary education compared to higher (college/university) education level had significantly higher odds of delivering moderately to late, but not very/extremely preterm. These findings were consistent to a meta-analysis of 12 European Cohorts, where poor health at birth was higher among babies born from mothers with low education levels (Ruiz et al., 2015). Policies and programs to improve maternal and child care in Tanzania should address health inequalities and prioritize the marginalized groups taking a multi-sectoral approach.

Furthermore, male children were more likely to be delivered preterm compared to females. This might be associated with shorter gestational duration for male compared to female fetuses (Zhang et al., 2018a). A study in the UK found no significant relationship between fetal gender and the risk of preterm birth among women at high risk of delivering preterm (ie, with a history of miscarriage, preterm birth or cervical surgery) (Teoh et al., 2018). We also found that primiparous women were less likely to deliver preterm compared to multiparous. Findings from a meta-analysis using data from cohort studies in LMIC indicated that nulliparous, aged $<$18 years and parity $\geq$3 aged $\geq$35 years women were more likely to experience adverse neonatal outcomes, including preterm birth (Kozuki et al., 2013). Other studies found no significant association between parity and the risk of preterm birth (van den Broek et al., 2014; Lu et al., 2015b; Temu et al., 2016;

Malacova et al., 2018). Despite that, interventions to improve maternal and child care should be delivered through out the course of woman's reproductive period.

Among the factors associated with the rise in trends of preterm birth is the iatrogenic early delivery (i.e. following labour induction and/or caesarean delivery) carried out for fetal or maternal indications (Lisonkova et al., 2012). In this study, women who delivered moderately to late preterm were more likely to deliver through caesarean section (CS). It is possible that these women had other obstetric complications such as a previous CS, severe pre-eclampsia/eclampsia, placenta praevia, preterm premature rupture of membranes, and high birthweight that contributed highly to CS delivery and hence preterm birth (Worjoloh et al., 2012; Tarimo et al., 2020). The odds of delivering both moderately to late and very/extremely preterm was high among mothers with pre-eclampsia/eclampsia, experienced placenta previa, and abruption placenta, as also reported elsewhere (Georgiou et al., 2015; Temu et al., 2016; Vogel et al., 2018). The effect of placenta previa on delivering very/extremely preterm were almost twice compared to the moderately to late preterm birth category. These conditions are both the risk factors as well as common indications for preterm birth (Lu et al., 2015a; Vogel et al., 2018). PROM increases the risk of preterm birth (Blencowe et al., 2013; Requejo et al., 2013; Georgiou et al., 2015; Lu et al., 2015a), which is consistent to the findings in this study. Previous studies have shown that PROM is among the common indications of spontaneous preterm birth (Requejo et al., 2013; Georgiou et al., 2015; Purisch & Gyamfi-Bannerman, 2017).

LBW was associated with eight-fold higher odds of moderately to late preterm ([32,37) weeks of gestation) and nearly 40 times higher odds of very/extremely preterm (<32 weeks of gestation). In fact, the proportions of moderately to late and very/extremely preterm birth were significantly higher among deliveries born with LBW than in the normal birth weight deliveries (37.1% and 14.3%, vs 6.4% and

0.4%, respectively) (results before imputation). Our findings agree with a previous case-control study in northern Tanzania, where LBW was associated with over 34-folds risk of preterm delivery (Temu et al., 2016). The observed increase in preterm birth due to LBW could be attributed to two factors; the fact that preterm birth is also a risk factor for LBW (low birth weight but appropriate for gestation age) and intrauterine growth retardation or small for gestational age. Literature shows that extremely preterm babies are more likely to be born with LBW, while newborns small for gestational age are at a higher risk of experiencing morbidity and mortality (Katz et al., 2013; Unterscheider et al., 2014). In this study, 81.2% (688/847) of very/extreme preterm newborns were born with both LBW and preterm compared to 41.6% (1779/4279) among moderately to late preterm (results before imputation). On the other hand, babies born preterm are at an increased risk of being born with LBW (Mitao et al., 2016) and experiencing perinatal and neonatal morbidity and mortality (UNICEF et al., 2020; Mboya et al., 2020b). Care for the LBW and preterm babies is a critical intervention for improving child survival. Special attention should be given to babies born with LBW at <32 weeks of gestation.

According to the WHO recommendations, antenatal care visit remains to be a critical entry point where high-risk pregnancies can be identified and managed (Requejo et al., 2013; World Health Organization, 2015, 2016b). We found that women with inadequate (<4) ANC visits are more likely to deliver moderately to late and very/extremely preterm. Similar findings were also reported in other studies (Lu et al., 2015a; Mahapula et al., 2016; Temu et al., 2016; Rugaimukam et al., 2017). However, these studies estimated the association between the number of ANC visits in the overall preterm birth categories (<37 weeks of gestation) compared to our study that showed different risk patters in two sub-categories of preterm birth (<32 and [32,37) weeks of gestation). In Tanzania, over half (51%) of pregnant women had at least four ANC visits during their last pregnancy

(MoHCDGEC [Tanzania Mainland] et al., 2016). Considering the current WHO recommendations of eight or more visits (World Health Organization, 2016b), different strategies are needed to promote health care seeking behaviors for pregnant women, and provision of quality ANC services at all levels of care. The timing and number of ANC visits is as important as the content and quality of care (Benova et al., 2018).

In this study, we applied the multinomial regression models with two categories of preterm birth (<32 and [32,37) weeks of gestation) due to rarity of cases in the <28 gestational weeks category. Eventually, the collapsed categories increased statistical power. Nevertheless, it is also possible that there may be under-reporting of extreme premature deliveries in the KCMC Medical birth registry. Despite the low accuracy of gestational age estimation based on the date of last menstrual period (Blencowe et al., 2013; Quinn et al., 2016; Vogel et al., 2018), it remains the widely used method in resource-limited settings like Tanzania. Even where ultrasound is available, this method "requires skilled technicians, equipment and for maximum accuracy, first-trimester antenatal clinic attendance" Blencowe et al. (2013), which is still a challenge in Tanzania (MoHCDGEC [Tanzania Mainland] et al., 2016). There are alternative gestational age estimation methods, such as a combination of ultrasound and LMP (Blencowe et al., 2013; Quinn et al., 2016; Vogel et al., 2018), but the question remains on the feasibility and applicability of these options in resource-limited settings.

Another limitation of this study is that it was hospital-based, utilizing the KCMC Medical Birth Registry data from the KCMC zonal referral hospital in northern Tanzania, hence suffers from referral bias. Nearly a quarter of all women were referred for delivery during the study period. This may affect the generalization of the results. Nevertheless, this is the only birth-registry in the country (and potentially one of the few in SSA) providing critical information for pregnancy

monitoring, administrative, and research purposes. Such registries allows for routine and inter-generational linkage and analysis of mother-child records. The KCMC hospital and its partners should promote routine data quality checks, resolve data quality and reporting challenges to ensure a sustainable operation of the birth registry, for current and future use.

6.5 Conclusion

The findings from this study support other studies showing improved precision of parameter estimates after imputation of missing values and the rising trends of preterm birth rates. The multinomial regression models allowed for the simultaneous assessment of predictors of different preterm birth categories as opposed to binary regression analysis. Policy decisions should intensify efforts on improved maternal and child care throughout the course of pregnancy and childbirth, towards prevention of preterm birth. Interventions to increase the uptake and quality of ANC services should also be strengthened in Tanzania at all levels of care, where several interventions can easily be delivered to pregnant women (World Health Organization, 2016b), especially those at high-risk of experiencing adverse pregnancy outcomes. The number of ANC visits is as important as the content of care.

Chapter 7

Joint modelling of singleton preterm birth and perinatal death using birth registry cohort data in northern Tanzania

7.1 Introduction

Globally, there is a notable decline of under five mortality rates since the year 1990 (UNICEF et al., 2020). Despite this decline, the share of mortality burden increased in the group of children in younger ages, especially in the first 28 days of life (neonatal period) (Burstein et al., 2019; Hug et al., 2019; UNICEF et al., 2020). The UN Inter-agency Group for Child Mortality Estimation report indicated that at a global rate of 17 deaths per 1000 live births, and approximately 6700 neonatal deaths everyday in 2019, neonatal period is the most vulnerable time for children under five years of age (UNICEF et al., 2020). The share of neonatal mortality to under five deaths has increased from 40% in 1990 to 47% in 2019 (UNICEF et al., 2020). In addition, sub-Saharan Africa (SSA) caries the highest burden of neonatal mortality rates in the world (Burstein et al., 2019; UNICEF et al., 2020). Most of the

neonatal deaths occurs during the perinatal period (Baqui et al., 2016; Mmbaga et al., 2012b; UNICEF et al., 2020). A recent meta-analysis in 21 SSA countries estimated a perinatal mortality rate of 34.7 per 1000 liver births. The Eastern Africa region had a rate of 34.5 per 1000 live births, and was highest (39.5 per 1000 live births) in Tanzania (Akombi & Renzaho, 2019; MoHCDGEC [Tanzania Mainland] et al., 2016).

Preterm birth complications are among the leading causes of perinatal and neonatal deaths (UNICEF et al., 2020). In 2018 alone, preterm birth complications accounted for 35% of all neonatal deaths, followed by intrapartum-related complications (24%) (UNICEF et al., 2019). Globally, preterm birth rate was 10.6%, equivalent to nearly 15 million live preterm births in 2014, 81% occuring in Asia and SSA (Chawanpaiboon et al., 2019). If these estimates are left unchecked within and between countries, there may be a proportional increase in perinatal deaths. Currently, Tanzania ranks the tenth country with the highest preterm birth rate in the world (16.6%) and shares a 2.2% of the global preterm birth proportions (Chawanpaiboon et al., 2019). Timely, quality, and skilled newborn care at birth and treatment immediately after birth and first days of life is essential to increase child survival (Chawanpaiboon et al., 2019; UNICEF et al., 2020).

Previous studies assessed the independent predictors of preterm birth and perinatal deaths or as the determinants of each other (Jena et al., 2020; Mahande et al., 2013a,b; Mpembeni et al., 2014; Mboya et al., 2020b,a; Rugaimukam et al., 2017; Temu et al., 2016). Maternal characteristics and conditions and complications in the current pregnancy increase preterm birth and perinatal death risk (Bailey et al., 2017; Nijkamp et al., 2017; van Zijl et al., 2016; Vogel et al., 2014, 2018). Also, previous exposure to these outcomes increases the recurrence risk (Malacova et al., 2018; Mahande et al., 2013a,b; Ouyang et al., 2013; van Zijl et al., 2016). These demonstrate the association between preterm birth and perinatal deaths. In other

words, two outcomes within the same individual are highly correlated. Birth registries are examples of such data where several outcomes are highly correlated. Joint modelling is relevant to reveal more about their relationship, hence inform clinical and public health decisions.

Joint modelling, particularly using the random effects approach, have been previously applied to clinical outcomes such as HIV and HCV (Del Fava et al., 2011; Ghebremichael, 2015), hearing thresholds (Fieuws & Verbeke, 2006; Fieuws et al., 2007), and body mass index with other clinical targets among diabetic patients (Ivanova et al., 2016). The application of these methods to pregnancy-related adverse outcomes is limited. This study aimed to jointly model preterm birth and perinatal death using the KCMC zonal referral hospital medical birth registry data in northern Tanzania. To our knowledge, no studies have jointly modelled preterm birth and perinatal death in Tanzania. A joint model of the two outcomes will help better understand potential risk factors for early diagnosis and management of high-risk pregnancies.

7.2 Methods

7.2.1 Data source

Data used in this study comes from a prospective hospital-based maternally linked cohort data from the KCMC zonal referral hospital in Moshi Municipality, Northern Tanzania. Details about this birth registry are presented in Chapter 2 and also published elsewhere (Bergsjo et al., 2007; Mahande, 2015; Mboya et al., 2020b,a; Mmbaga et al., 2012a; Temu et al., 2016). Briefly, the KCMC medical birth records information for women and their subsequent deliveries from 2000 to date. The hospital has an average of 3500-4000 births every year, close to 70,000 recorded deliveries to date. All consenting mothers are interviewed using a specially designed questionnaire by the project midwives 24 hours after normal delivery.

Mothers undergoing cesarean delivery or who experienced a complicated birth are interviewed on the second or third day, depending on their condition.

7.2.2 Study variables and variable definitions

The primary outcomes were preterm birth and perinatal death. Perinatal death comprises stillbirths (pregnancy loss that occurs after seven months of gestation) and early neonatal death (death of live births within the first seven days of life) (World Health Organization, 2019; MoHCDGEC [Tanzania Mainland] et al., 2016). We coded perinatal death was as binary, i.e., 'Yes' if the child died and 'No' if otherwise. Preterm birth is any birth before 37 completed weeks of gestation or fewer than 259 days from the first date of a woman's last menstrual period (Chawanpaiboon et al., 2019; Mahande et al., 2013b; World Health Organization, 2020) and was also analyzed as a binary variable ($<$37 vs $\geq$37 weeks of gestation).

The secondary outcome was the co-occurrence of preterm birth and perinatal death. We generated a categorical variable from the two outcomes with the following categories; '0' if none of the events occurred, '1' if both occurred, '2' if perinatal death only, and '3' if preterm birth only occurred. We then used a multinomial random-effects regression model to predict the independent and co-occurrence of preterm birth and perinatal death.

The independent variables included maternal and paternal background characteristics and maternal conditions and complications during pregnancy and delivery. Previous literature (Bailey et al., 2017; Malacova et al., 2018; Nijkamp et al., 2017; van Zijl et al., 2016; Vogel et al., 2014, 2018) and analyses of this cohort data informed selection of these variables (Mboya et al., 2020a,b). The background characteristics were maternal age (15-19, 20-24, 25-34, 35-39, and 40+), paternal age (15-24, 25-29, 30-34, and 35+), maternal and paternal highest level of education (none, primary, secondary, and higher), paternal and maternal occupation

(employed, unemployed, farmer, and others), marital status (married, single, and widowed/divorced), the current area of residence (rural, urban), body mass index (BMI) in Kg/m^2 (normal [18.5-24.9], underweight [<18.5], overweight [25-29.9], and obese [30+]), and paternal age [15-24, 25-29, 30-34, 35+].

Maternal conditions and complications during pregnancy and delivery were number of antenatal care visits (4+, <4), parity (primipara, multipara), HIV status (positive, negative), and referral status (Yes, No). Maternal anemia and malaria during pregnancy, infections, pre-eclampsia/eclampsia, premature rupture of the membranes (PROM), postpartum haemorrhage (PPH), abruption placenta, and placenta previa were all binary (Yes, No). Other information included sex of the child (male, female), birth weight (normal [≥2500g], low birth weight (LBW) [<2500g]) (World Health Organization, 2014b), presentation at birth (cephalic, breech, and transverse), mode of delivery (vaginal, cesarean section (CS)), and Apgar score at 5 minutes (high [7+], low [<7]).

7.2.3 Study population and eligibility criteria

The study population for this study was women who delivered singleton babies from January 2000 to December 2017. For this period, there were 60,840 deliveries from 45,324 mothers aged 15-49 years. We excluded 52 records missing unique identification numbers (used to link mothers and their subsequent births) and 3,669 multiple gestations (i.e., twins and triplets) to avoid over-representing high-risk pregnancies. We further excluded 1,212 deliveries of unknown sequence (i.e., whether singleton or multiple births). We, therefore, analyzed data for 55,907 recorded deliveries, of which 49,113 had complete information on gestational age and 55,736 on perinatal status (Figure 7.1).

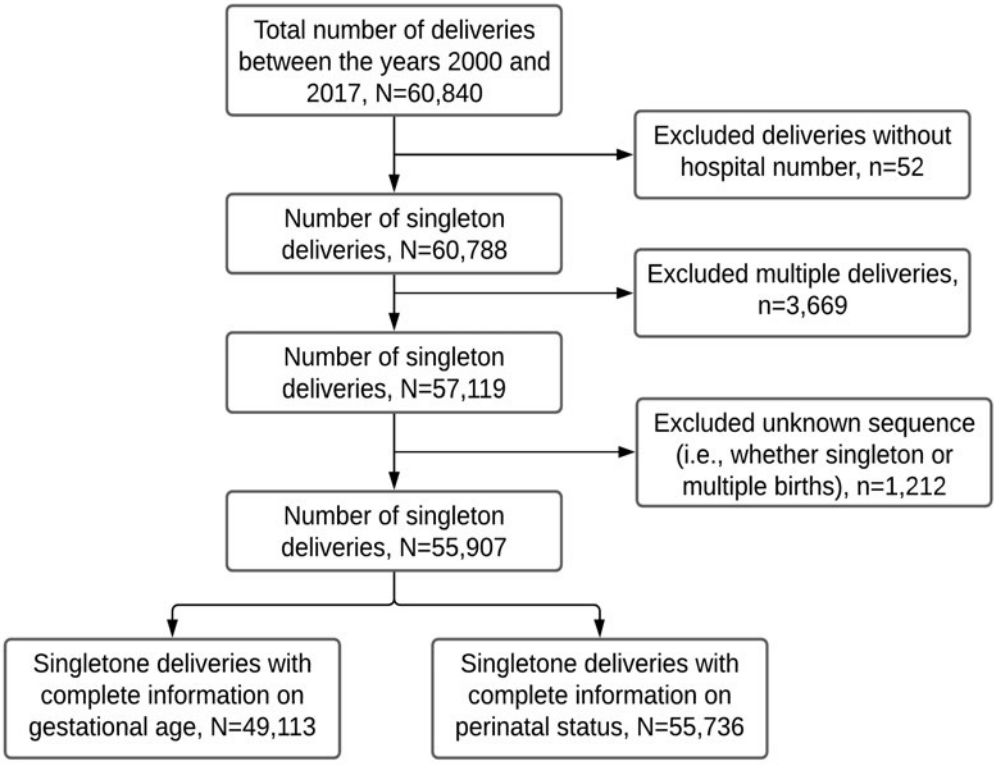


Figure 7.1: Flow chart showing the number of singleton deliveries analyzed in this study. Data from the KCMC Medical birth registry, 2000-2017.

7.2.4 Descriptive analysis

Data were analyzed using STATA version 15.1 (StataCorp LLC, College Station, Texas, USA)(StataCorp, 2019). The primary unity of analysis was singleton deliveries for women recorded in the KCMC Medical Birth Registry between 2000-2017. We summarized numeric variables using means and standard deviations and categorical variables using frequencies and percentages. The Chi-square test compared the proportion of preterm births and perinatal deaths by maternal and paternal background characteristics and maternal conditions and complications during pregnancy and childbirth. Ordinary least-squares linear regression assessed linear trends of proportions of the two outcomes for every year increase. Findings from previous analyses for the predictors of preterm birth (Mboya et al., 2021) and perinatal death (Mboya et al., 2020a,b) informed selection of variables to include in the initial steps of multivariable analysis. The next step

was a separate stepwise manual reduction of variables not significantly associated with preterm birth and perinatal death ($p<0.05$) using the mixed-effects generalized linear models with exchangeable correlation structure. This step was essential given additional variables, such as paternal characteristics, which were significant predictors of perinatal death in the previous analysis using machine learning models (Mboya et al., 2020a). Of importance, we tested the effect of including paternal characteristics in this step, which were not significant predictors of any of the two outcomes.

7.2.5 Joint modeling of correlated binary outcomes

7.2.5.1 The joint model of two binary responses

We developed a joint model of preterm birth and perinatal death using random effects models with an exchangeable correlation structure. Both outcomes were binary, hence used the binomial family and logit link function. We assumed that a set of latent, unobserved random effects of the same mother's two outcomes are correlated. Therefore, we used shared random intercepts to determine the correlation between the same mother's two outcomes, i.e., preterm birth and perinatal death. The random intercept captures the unobserved factors specific to each individual, which may influence the responses (Ghebremichael, 2015). Let Y_{ij} denote the jth response ($j = 1, 2$) of the ith ($i = 1, 2, \ldots, n$) subject, with $j = 1$ for preterm birth and $j = 2$ for perinatal death. A binary response Y_{ij} takes the values 1 if an event has occurred and 0 if otherwise. Thus, for the ith subject, we have a bivariate binary response vector (Y_{1i}, Y_{2i}). We also let X_{1i} and X_{2i}, represent the vectors of covariates associated with preterm birth and perinatl death, and $\beta_1(\hat{\beta}_1)$ and $\beta_2(\hat{\beta}_2)$ be their corresponding regression coefficients, and estimates in brackets, respectively. Random effects models are used to jointly model two longitudinal outcomes of different nature (Ghebremichael, 2015; Faes et al., 2008; Fieuws & Verbeke, 2006; Ivanova et al., 2016), also referred as multivariate longitudinal models (Fitzmaurice et al., 2009). The association between the covariates and each

outcome (preterm birth and perinatal death) can be examined using separate regression models for each outcome given as (Ghebremichael, 2015)

$$\begin{aligned} logit\left\{E(Y_{1i})\right\} &= logit\left\{Pr(Y_{1i}=1|X_{1i},\beta_1)\right\} \\ &= \beta_1^T X_{1i} \end{aligned} \tag{7.1}$$

and

$$\begin{aligned} logit\left\{E(Y_{2i})\right\} &= logit\left\{Pr(Y_{2i}=1|X_{2i},\beta_2)\right\} \\ &= \beta_2^T X_{2i} \end{aligned} \tag{7.2}$$

Ghebremichael Ghebremichael (2015) correctly indicated that these traditional logistic regression models ignore the correlation between the two outcomes. The random effects capture the unobserved factors specific to each individual, which may influence the responses (Ghebremichael, 2015). The joint models using random effects can be developed following two approaches; shared parameter and the multivariate random-effects models.

7.2.5.2 Shared-parameter models

The joint model is built by describing the joint density $f(y_{1i}, y_{2i})$ of the binary response vectors Y_{1i} and Y_{2i}. Let b_i denote the random effects shared by the two responses of the ith individual. We further let d_{1j} and d_{2j} define the dummy variables, with $d_{1j} = 1$ for $j = 1$ and $d_{2j} = 1$ for $j = 2$. A popular approach is to postulate a so-called shared-parameter model (Fitzmaurice et al., 2009), where the joint density for (Y_{1i}, Y_{2i}) is obtained from

$$f(y_{1i}, y_{2i}) = \int f(y_{1i}, y_{2i}|b_i)f(b_i)db = \int f(y_{1i}|b_i)f(y_{2i}|b_i)f(b_i)db_i \tag{7.3}$$

in which $f(b_i)$ denotes the random-effects density. The joint response model using logit link for binary responses can be given by (Ghebremichael, 2015)

$$\begin{aligned} logit\{E(Y_{ij}|b_i)\} &= logit\{Pr(Y_{ij}=1|X_{ij},\beta_j,b_i\}) \\ &= d_{1i}(\beta_1^T X_{1i}+b_i)+d_{2i}(\beta_2^T X_{2i}+b_i) \end{aligned} \tag{7.4}$$

Alternatively, equation 7.4 can be expressed in a vector form as

$$logit\left\{E\begin{pmatrix}Y_{i1}\\Y_{i2}\end{pmatrix}\right\}=\begin{pmatrix}\beta_1^T X_{1i}+b_i\\ \beta_2^T X_{2i}+b_i\end{pmatrix} \tag{7.5}$$

where the bivariate responses (Y_{1i}, Y_{2i}) of all individuals are stacked into a single response vector (Y_{ij}). The random effect b_i is a "shared parameter" inducing correlation between the two binary responses Y_{1i} and Y_{2i} through the joint dependence on b_i. The conditional independence of Y_{1i} and Y_{2i} given b_i may reflect the belief that a common set of underlying characteristics of the individual governs both outcomes (Fitzmaurice et al., 2009).

The random intercept b_i in 7.5 shared by both outcomes dictates that correlations between parts of measurements from different outcomes must be equal to the product of the correlation between measurements of the two outcomes. "A key disadvantage of shared-parameter models is that they can imply very strong assumptions about the association between the outcomes modeled, given longitudinal measurements of both Y_{1i} and Y_{2i}" Fitzmaurice et al. (2009). Therefore, it is evident that the shared-parameter models provide a limited representation of the dependence structure of multiple longitudinal outcomes.

7.2.5.3 Multivariate random-effects models

The rigid assumption in the shared-parameter models may be relaxed by allowing the models for Y_{1i} and Y_{2i} to depend on separate random effects b_{1i} and b_{2i}, which are themselves correlated (Faes et al., 2008; Fitzmaurice et al., 2009). For two

longitudinal binary responses, a GLMM with correlated random effects is given as Faes et al. (2008)

$$\begin{pmatrix} Y_{i1j} \\ Y_{i2j} \end{pmatrix} = \begin{pmatrix} \dfrac{exp(\alpha_0 + \alpha_1 X_{ij} + b_{1i})}{1 + exp(\alpha_0 + \alpha_1 X_{ij} + b_{1i})} \\ \dfrac{exp(\beta_0 + \beta_1 X_{ij} + b_{2i})}{1 + exp(\beta_0 + \beta_1 X_{ij} + b_{2i})} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1j} \\ \varepsilon_{i2j} \end{pmatrix} \tag{7.6}$$

Rather than being linked by shared dependence on a common random effects as in 7.5, the models for the two binary outcomes are joined by assuming the random-effect vector $\mathbf{b} = (b_1, b_2)'$ are independent of ε_{i1j} and ε_{i2j}, which has a multivariate normal distribution with zero mean vector and covariance matrix D. The random effects b_{i1} and b_{i2} are normally distributed as

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix} \right\} \tag{7.7}$$

The random effects b_{i1} and b_{i2} are used to accommodate the longitudinal structure in the data. It is further assumed that $\text{var}(\varepsilon_{i1j}) = \upsilon_{i1j} = \pi_{i1j}(b_{1i} = 0)[1 - \pi_{1ij}(b_{1i} = 0)]$ and $\text{var}(\varepsilon_{2ij}) = \upsilon_{2ij} = \pi_{2ij}(b_{2i} = 0)[1 - \pi_{2ij}(b_{2i} = 0)]$. The approximate variance covariance matrix of the two responses for subject i at time point j is equal to (Faes et al., 2008)

$$V_{ij} = \begin{pmatrix} \upsilon_{1ij}^2\tau_1^2 + \upsilon_{1ij} & \rho\tau_1\tau_2\upsilon_{1ij}\upsilon_{2ij} \\ \rho\tau_1\tau_2\upsilon_{1ij}\upsilon_{2ij} & \upsilon_{2ij}^2\tau_2^2 + \upsilon_{2ij} \end{pmatrix} \tag{7.8}$$

and the correlation between the two outcomes given as (Faes et al., 2008)

$$\rho_{Y_1,Y_2} = \frac{\rho\tau_1\tau_2\upsilon_{1ij}\upsilon_{2ij}}{\sqrt{\upsilon_{1ij}^2\tau_1^2 + \upsilon_{1ij}}\sqrt{\upsilon_{2ij}^2\tau_2^2 + \upsilon_{2ij}}} \tag{7.9}$$

Expression 7.8 demonstrates explicitly the role of the correlation between the outcome-specific random effects in dictating the between-process outcome correlation at any two time points. The model no longer assumes that the product of the within-process correlations equals the between-process correlation, thus

allowing a more general dependence structure (Fitzmaurice et al., 2009). The shared parameter model in equation 7.5 can be obtained as a special case of equation 7.6 by restricting the correlation between b_1 and b_2 to be equal to 1.

7.2.5.4 Estimation and inference

The joint responses of Y_{1i} and Y_{2i} are assumed to be independent given the shared random effects (b_i). Assume the b_i are normally distributed with zero mean and variance covariance matrix D. Given this assumption, we can write the likelihood function of the joint response model as follows (Ghebremichael, 2015)

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n}\prod_{j=1}^{2} Pr(Y_{ij}=1|X_{ij},\beta_j,b_i) \\ &= \prod_{i=1}^{n}\prod_{j=1}^{2}\left\{\int Pr(Y_{ij}=1|X_{ij},\beta_j,b_i)dG(b_i)\right\} \\ &= \prod_{i=1}^{n}\left\{\int \prod_{j=1}^{2} Pr(Y_{ij}=1|X_{ij},\beta_j,b_i)dG(b_i)\right\} \\ &= \prod_{i=1}^{n}\left\{\int \prod_{j=1}^{2} \frac{e^{b_i+\beta_j^{\mathrm{T}}X_{ij}}}{1+e^{b_i+\beta_j^{\mathrm{T}}X_{ij}}}dG(b_i)\right\} \end{aligned} \tag{7.10}$$

where θ is the vector of all parameters in the conditional distribution and the multivariate normal distribution for b_i. The integrals involved in equation 7.10 cannot be calculated analytically and numerical approaches are needed (Ivanova et al., 2016; Ghebremichael, 2015). Numeric approximations, such as adaptive Gaussian quadrature are recommended to estimate the model parameters (Fieuws & Verbeke, 2006; Ivanova et al., 2016; Molenberghs & Verbeke, 2005; Rabe-Hesketh et al., 2002). The higher the order of the quadrature, the better the approximation will be of the N subjects integrals in the likelihood (Molenberghs & Verbeke, 2005). Once the model has been fitted, inferences for all elements in θ become available using standard likelihood theory (e.g., likelihood ratio tests, Wald tests, score tests) (Ivanova et al., 2016).

We used maximum likelihood estimation using adaptive Gaussian quadrature method based on 10 quadrature points to obtain parameter estimates of the joint models (Ivanova et al., 2016; Ghebremichael, 2015; Molenberghs & Verbeke, 2005). This method gives precise parameter estimates at the price of being computationally intensive (Molenberghs & Verbeke, 2005).

7.2.5.5 A random-effects multinomial regression model for co-occurence

We further developed two additional multinomial random-effects models to assess predictors of co-occurrence of both preterm birth and perinatal death. These models provided additional information to understand the dependence between the two outcomes conditional on the random effects. The first model was random effects, multinomial regression model, with robust standard errors. As previously explained in section 7.2.2, we assessed predictors of both outcomes occurring, the occurrence of preterm birth only and perinatal death only, in a single multinomial variable. This model estimated a single random effects variance to account for mother-to-mother variability of the two responses. Let Y_{ij} denote a nominal response variable for the ith subject and jth measurement occasion. Given the shared random effects (b_i), the probability that a response Y_{ij} occurs in category c for a given level-2 unit (i) allowing for any possible set of $C-1$ response categories is written as

$$P_{ijc} = \frac{\exp(\eta_{ijc})}{\sum_{c=1}^{C}\exp(\eta_{ijc})} \quad \text{for} \quad c = 1, 2, \ldots, C \tag{7.11}$$

where the multinomial logit linear predictor, $\eta_{ijc} = X'_{ijc}\beta_c + Z'_{ijc}b_i$. The random effects b_i are shared across the $C-1$ binary comparisons in the multinomial logit model. The second model was developed similar to in 7.11, but allowing for separate but correlated random effects of the multinomial logits. The random effects b_i in the linear predictor, $\eta_{ijc} = X'_{ijc}\beta_c + Z'_{ijc}b_{ic}$ are now different for each binary comparison in the multinomial logit. A model with separate random effects estimated covariance parameters for each pair of the multinomial outcomes. Assessment of the best-fitting

model can be done using AIC or the the model standard errors. Several other post estimation commands can be used after the gsem estimation in STATA (StataCorp, 2017).

7.3 Results

7.3.1 Preterm birth and perinatal death proportions by maternal and paternal characteristics

The overall proportions of preterm birth and perinatal death between 2000-2017 recorded in the KCMC medical birth registry was 12.8% and 4.3%, respectively and perinatal mortality rate (PMR) of 42.6 per 1000 births. The proportions of preterm birth and perinatal death differed significantly ($p<0.05$) by maternal and paternal background characteristics and obstetric care characteristics (Tables 7.1 and 7.2). The preterm birth proportion was significantly higher among mothers aged 15-19 (15.7%) and 40+ years (17%), those with no education (16.3%), farmers (16.6%), and rural residents (14.3%). The highest proportions of preterm birth were among younger fathers, i.e., 15-24 years (16.2%), with no education (20.5%), and farmers (17.9%). Furthermore, the perinatal death proportions were significantly higher among mothers aged 40+ years (6.4%), with no education (9%), farmers (6.5%), and rural residents (5.6%). Among fathers, perinatal death proportions were high among those aged 30-34 (4.1%) and 35+ years (4.8%), with no education (12.7%), and farmers (7.7%) (Table 7.1).

7.3.2 Preterm birth and perinatal death proportions by maternal conditions and complications during pregnancy and delivery

The preterm birth proportions were highest among mothers with inadequate (<4) ANC visits (27.4%), those referred for delivery (20.1%), experienced pre-eclampsia/eclampsia (33%), PROM (23.6%), PPH (22.4%), abruption placenta (50.9%), and placenta previa (55.9%), delivered LBW baby (53.8%), experienced

breech presentation at birth (27.1%), had <7 five minutes Apgar score (42.0%), and experienced perinatal death (47%). Also, the perinatal death proportions are high among mothers with inadequate ANC visits (6.8%), referred for delivery (8.5%), experienced pre-eclampsia/eclampsia (13.1%), PPH (18.6%), delivered LBW baby (19.1%), breech presentation at birth (20.2%), low (<7) five minutes Apgar score (60.4%), and delivered preterm (15.3%). Notably, the highest proportions of perinatal deaths are among those that experienced abruption placenta (55.4%) and with low (<7) five minutes Apgar score (60.4%) (Table 7.2). The proportion of preterm birth was 47% among 2024 perinatal deaths compared to 11% among 46938 live births. On the other hand, among 6238 preterm deliveries (<37 gestational weeks), 15.3% experienced perinatal death compared to 2.5% among 42724 term (37+ gestational weeks) deliveries. These differences were statistically significantly different, p<0.001 (results not shown in the table).

7.3.3 Trends of preterm birth and perinatal death between 2000-2017.

Between 2000 and 2017, there was a rising trend of preterm birth while perinatal death proportions decline slightly in this cohort. The proportion of preterm birth (<37 gestational weeks) increased significantly by 0.33 (95%CI 0.23, 0.43, p<0.001) while that of perinatal death decreased significantly by 0.11 (95%CI 0.08-0.15, p<0.001) for every one-year increase (Figure 7.2).

7.3.4 Joint predictors of preterm birth and perinatal death

7.3.4.1 Joint model with separate but correlated random effects

Findings of the joint model with separate but correlated random effects are in Table 7.3. The random-effects variance is observed to be equal for both outcomes (Var=0.18, 95%CI 0.004, 9.09) and is significantly greater than zero. The covariance parameter capturing dependence between the two outcomes is not statistically significant (Cov=-0.11, 95%CI -0.42, 0.20). Therefore, the two outcomes are independent conditional on accounting for mother to mother

Table 7.1: Distribution of preterm birth and perinatal death by maternal and paternal characteristics (N=55,907)

Characteristics	Preterm birth			Perinatal death		
	Total (%)	n (%)	P-value	Total (%)	n (%)	P-value
Maternal age in years			**<0.001**			**<0.001**
15-19	3749 (6.7)	589 (15.7)		4416 (7.9)	173 (3.9)	
20-24	11930 (21.4)	1547 (13.0)		13648 (24.5)	520 (3.8)	
25-34	25687 (46.0)	2973 (11.6)		28920 (51.8)	1199 (4.1)	
35-39	6045 (10.8)	862 (14.3)		6792 (12.2)	359 (5.3)	
40+	1609 (2.9)	274 (17.0)		1851 (3.3)	119 (6.4)	
Maternal highest education level			<0.001			<0.001
None	718 (1.3)	117 (16.3)		1031 (1.8)	93 (9.0)	
Primary	25669 (46.0)	3713 (14.5)		29479 (52.8)	1500 (5.1)	
Secondary	6922 (12.4)	914 (13.2)		7768 (13.9)	244 (3.1)	
Higher	15727 (28.2)	1499 (9.5)		17343 (31.1)	511 (2.9)	
Maternal occupation			<0.001			<0.001
Employed	26226 (47.2)	2921 (11.1)		29303 (52.7)	971 (3.3)	
Unemployed	10445 (18.8)	1451 (13.9)		11852 (21.3)	517 (4.4)	
Farmer	9067 (16.3)	1501 (16.6)		10699 (19.3)	700 (6.5)	
Others	3114 (5.6)	360 (11.6)		3554 (6.4)	151 (4.2)	
Marital Status			<0.001			0.08
Married	42385 (76.0)	5232 (12.3)		48037 (86.1)	2036 (4.2)	
Single	6569 (11.8)	988 (15.0)		7477 (13.4)	304 (4.1)	
Widowed/Divorced	89 (0.2)	25 (28.1)		107 (0.2)	9 (8.4)	
Current area of residence			<0.001			<0.001
Urban	29417 (52.8)	3448 (11.7)		32915 (59.0)	1086 (3.3)	
Rural	19576 (35.1)	2801 (14.3)		22673 (40.7)	1276 (5.6)	
Body mass index categories (Kg/m^2)			<0.001			0.64
Normal (18.5-24.9)	18021 (46.8)	2029 (11.3)		20427 (53.0)	696 (3.4)	
Underweight (<18.5)	1766 (4.6)	232 (13.1)		2029 (5.3)	68 (3.4)	
Overweight (25-29.9)	9596 (24.9)	944 (9.8)		10770 (28.0)	395 (3.7)	
Obese (30+)	4601 (11.9)	505 (11.0)		5171 (13.4)	186 (3.6)	
Paternal age (years)			<0.001			<0.001
15-24	4460 (8.0)	721 (16.2)		5149 (9.3)	189 (3.7)	
25-29	11979 (21.6)	1466 (12.2)		13595 (24.5)	486 (3.6)	
30-34	14199 (25.6)	1662 (11.7)		15995 (28.8)	656 (4.1)	
35+	18179 (32.8)	2363 (13.0)		20582 (37.1)	996 (4.8)	
Paternal education level			<0.001			<0.001
None	365 (0.7)	75 (20.5)		529 (1.0)	67 (12.7)	
Primary	21163 (38.0)	3154 (14.9)		24440 (43.9)	1302 (5.3)	
Secondary	6083 (10.9)	851 (14.0)		6776 (12.2)	233 (3.4)	
Higher	21358 (38.4)	2152 (10.1)		23765 (42.7)	741 (3.1)	
Paternal occupation			<0.001			<0.001
Employed	41695 (74.9)	4964 (11.9)		46932 (84.3)	1756 (3.7)	
Unemployed	878 (1.6)	127 (14.5)		1005 (1.8)	23 (2.3)	
Farmer	5637 (10.1)	1009 (17.9)		6671 (12.0)	515 (7.7)	
Others	764 (1.4)	131 (17.1)		915 (1.6)	50 (5.5)	
Total n (%)	**49113**	**6263 (12.8)**		**55736**	**2377 (4.3)**	

Note: Variables may not tally to the total frequencies due to missing values in either the exposure or the outcome of interest.

Table 7.2: Distribution of preterm birth and perinatal death by maternal conditions and complications during pregnancy and delivery (N=55,907)

Characteristics	Preterm birth			Perinatal death		
	Total (%)	n (%)	P-value	Total (%)	n (%)	P-value
Number of ANC visits			<0.001			<0.001
4+	33291 (60.5)	2488 (7.5)		37619 (68.4)	1111 (3.0)	
<4	15087 (27.4)	3581 (23.7)		17198 (31.3)	1161 (6.8)	
Parity			<0.001			0.23
Multipara	9456 (16.9)	1063 (11.2)		10552 (18.9)	449 (4.3)	
Primipara	39657 (70.9)	5200 (13.1)		45184 (80.8)	1928 (4.3)	
Drank alcohol during this pregnancy			<0.001			0.004
No	35922 (64.4)	4778 (13.3)		40745 (73.0)	1782 (4.4)	
Yes	13123 (23.5)	1474 (11.2)		14874 (26.7)	568 (3.8)	
Referred for delivery			<0.001			<0.001
No	36498 (67.6)	3907 (10.7)		40988 (76.0)	1169 (2.9)	
Yes	10910 (20.2)	2189 (20.1)		12817 (23.7)	1092 (8.5)	
HIV status			<0.001			0.001
Negative	36764 (83.7)	4574 (12.4)		41568 (94.6)	1541 (3.7)	
Positive	1972 (4.5)	304 (15.4)		2265 (5.2)	114 (5.0)	
Anemia			0.54			<0.001
No	48349 (86.5)	6160 (12.7)		54872 (98.1)	2317 (4.2)	
Yes	764 (1.4)	103 (13.5)		864 (1.5)	60 (6.9)	
Malaria			<0.001			0.43
No	42992 (76.9)	5600 (13.0)		48760 (87.2)	2067 (4.2)	
Yes	6121 (10.9)	663 (10.8)		6976 (12.5)	310 (4.4)	
Any infections condition			0.07			0.18
No	48340 (86.5)	6181 (12.8)		54869 (98.1)	2348 (4.3)	
Yes	773 (1.4)	82 (10.6)		867 (1.6)	29 (3.3)	
Pre-eclampsia/eclampsia			<0.001			<0.001
No	47008 (84.1)	5569 (11.8)		53389 (95.5)	2069 (3.9)	
Yes	2105 (3.8)	694 (33.0)		2347 (4.2)	308 (13.1)	
PROM			<0.001			0.002
No	48123 (86.1)	6029 (12.5)		54635 (97.7)	2351 (4.3)	
Yes	990 (1.8)	234 (23.6)		1101 (2.0)	26 (2.4)	
PPH			<0.001			<0.001
No	48760 (87.2)	6184 (12.7)		55343 (99.0)	2304 (4.2)	
Yes	353 (0.6)	79 (22.4)		393 (0.7)	73 (18.6)	
Abruption placenta			<0.001			<0.001
No	48950 (87.6)	6180 (12.6)		55552 (99.4)	2275 (4.1)	
Yes	163 (0.3)	83 (50.9)		184 (0.3)	102 (55.4)	
Placenta previa			<0.001			0.08
No	49002 (87.6)	6201 (12.7)		55616 (99.5)	2368 (4.3)	
Yes	111 (0.2)	62 (55.9)		120 (0.2)	9 (7.5)	
Sex of the baby			0.86			0.65
Female	23664 (42.5)	3005 (12.7)		26831 (48.2)	1128 (4.2)	
Male	25245 (45.3)	3219 (12.8)		28686 (51.5)	1228 (4.3)	
Birth weight			<0.001			<0.001
NBW	43619 (78.2)	3313 (7.6)		49596 (88.9)	1190 (2.4)	
LBW	5373 (9.6)	2889 (53.8)		6008 (10.8)	1148 (19.1)	
Presentation			<0.001			<0.001
Cephalic	48160 (86.6)	6030 (12.5)		54686 (98.3)	2183 (4.0)	
Breech	638 (1.1)	173 (27.1)		729 (1.3)	147 (20.2)	
Transverse	75 (0.1)	11 (14.7)		83 (0.1)	17 (20.5)	
Delivery mode			<0.001			0.03
Vaginal	32085 (57.6)	3744 (11.7)		36426 (65.4)	1586 (4.4)	
CS	16855 (30.3)	2487 (14.8)		19116 (34.3)	757 (4.0)	
Apgar score at 5 minutes			<0.001			<0.001
High (7+)	46015 (83.2)	4981 (10.8)		52117 (94.3)	161 (0.3)	
Low (<7)	2543 (4.6)	1068 (42.0)		3006 (5.4)	1817 (60.4)	
Induced labour			<0.001			<0.001
No	37537 (67.5)	5088 (13.6)		42648 (76.6)	1695 (4.0)	
Yes	11352 (20.4)	1135 (10.0)		12831 (23.1)	667 (5.2)	
Total n (%)	**49113**	**6263 (12.8)**		**55,736**	**2377 (4.3)**	

Note: Variables may not tally to the total frequencies due to missing values in either the exposure or the outcome of interest.

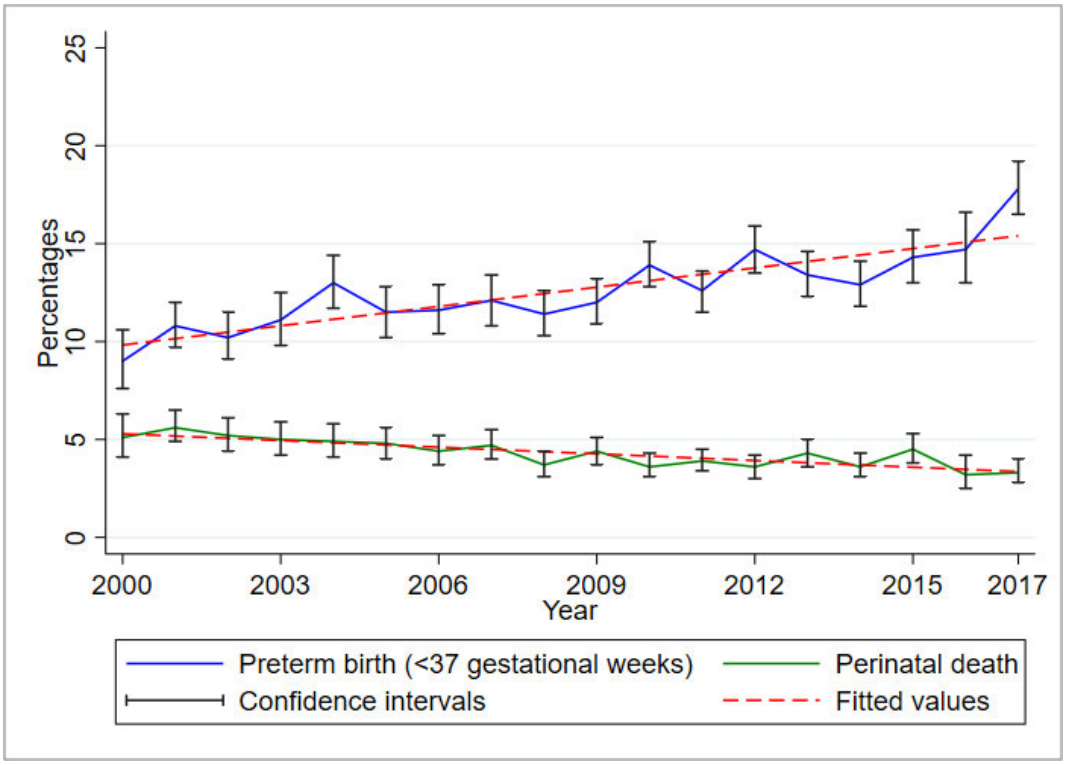


Figure 7.2: Trends of preterm birth and perinatal death. Data from the KCMC Medical Birth Registry, 2000-2017.

variability/heterogeneity.

Conditional on the random-effects, inadequate (<4) ANC visits (OR=2.92, 95%CI 2.71, 3.15 and OR=1.26, 95%CI 1.05, 1.51), being referred for delivery (OR=1.32, 95%CI 1.21, 1.43, and OR=1.37, 95%CI 1.13, 1.66), abruption placenta (OR=1.73, 95%CI 1.03, 2.92 and OR=2.43, 95%CI 1.35, 4.40), and breech presentation (OR=1.54, 95%CI 1.19, 1.99 and OR=4.01, 95%CI 1.96, 8.19) increased the odds of both preterm birth and perinatal death, respectively. For every one year increase, the odds of preterm birth increased significantly by 1.04 (95%CI 1.03, 1.05) while that of perinatal death decreased by 0.97 (95%CI 0.95, 0.99).

Also, conditional on the random effects, adolescent mothers (15-19 years) were significantly more likely to deliver preterm (OR=1.24, 95%CI 1.09, 1.42) but had lower odds of experiencing perinatal death (OR=0.43, 95%CI 0.30, 0.62). Similar

results were among mothers aged 20-24 years though this association was not statistically significant. Likewise, higher odds of preterm birth were among mothers who experienced PROM (OR=1.92, 95%CI 1.59, 2.33), experienced placenta previa (OR=4.71, 95%CI 2.66, 8.35), and among male children (OR=1.12, 95%CI 1.04, 1.19). On the contrary, experiencing PROM (OR=0.35, 95%CI 0.17, 0.72) and placenta previa (OR=0.21, 95%CI 0.06, 0.79), and male children (OR=0.83, 95%CI 0.71, 0.98) were less likely to experience perinatal death. Induction of labour was protective of preterm birth of preterm birth (OR=0.82, 95%CI 0.75, 0.89) but increased the odds of perinatal death (OR=1.43, 95%CI 1.18, 1.73).

7.3.4.2 Predictors of independent and co-occurrence of preterm birth and perinatal death using random effect multinomial regression model

Findings from the random-effect multinomial regression model are in Table 7.4. This model's random-effects variance is not significantly from zero (Var=0.04, 95%CI 0.00, 99.43). The observed results are not surprising. The reason is that we generated a multinomial variable from preterm birth and perinatal death, allowing for modelling the dependence between the two outcomes directly other than through separate and correlated random effects. Significantly higher odds of co-occurrence of preterm birth and perinatal death were among mothers with inadequate (<4) ANC visits (OR=3.46, 95%CI 2.77, 4.32), experienced pre-eclampsia/eclampsia (OR=1.38, 95%CI 1.01, 1.89), PPH (OR=2.24, 95%CI 1.05, 4.78), and abruption placenta (OR=3.98, 95%CI 2.02, 7.82), delivered LBW baby (OR=12.81, 95%CI 9.84, 16.67), and had a breech presentation (OR=3.79, 95%CI 2.03, 7.08). Adolescent mothers (15-19 years) (OR=0.46, 95%CI 0.29, 0.73), with no education (OR=0.41, 95%CI 0.20, 0.84), primipara (OR=0.63, 95%CI 0.48, 0.84), and delivered through CS (OR=0.50, 95%CI 0.40, 0.64) had lower odds of co-occurrence of preterm birth and perinatal death.

The factors independently associated with a higher odds of perinatal death were

Table 7.3: Joint predictors of preterm birth and perinatal death with separate but correlated random effects

Variables	Preterm birth†	Perinatal death4‡
	OR¶ (95%CI)	OR¶ (95%CI)
Maternal age groups		
15-19	1.24 (1.09, 1.42)***	0.43 (0.30, 0.62)***
20-24	1.16 (1.07, 1.26)***	0.95 (0.77, 1.16)
25-29	1.00	1.00
35-39	1.07 (0.97, 1.19)	1.21 (0.93, 1.57)
40+	1.22 (1.03, 1.46)*	1.30 (0.83, 2.02)
Maternal education		
None	1.06 (0.81, 1.39)	-
Primary	1.29 (1.19, 1.40)***	-
Secondary	1.12 (1.00, 1.25)*	-
Higher	1.00	-
Area of residence (Rural)	0.91 (0.84, 0.97)**	-
ANC visits (<4)	2.92 (2.71, 3.15)***	1.26 (1.05, 1.51)*
Referred for delivery (Yes)	1.32 (1.21, 1.43)***	1.37 (1.13, 1.66)**
Parity (Primipara)	-	0.69 (0.53, 0.89)**
Pre-eclampsia/eclampsia (Yes)	1.79 (1.56, 2.05)***	1.05 (0.79, 1.40)
PROM (Yes)	1.92 (1.59, 2.33)***	0.35 (0.17, 0.72)**
PPH (Yes)	-	3.46 (2.02, 5.93)***
Abruption placenta (Yes)	1.73 (1.03, 2.92)*	2.43 (1.35, 4.40)**
Placenta previa (Yes)	4.71 (2.66, 8.35)***	0.21 (0.06, 0.79)*
Sex (Male)	1.12 (1.04, 1.19)**	0.83 (0.71, 0.98)*
LBW (Yes)	10.36 (9.11, 11.77)***	1.32 (0.97, 1.78)
Presentation at birth		
Cephalic	1.00	1.00
Breech	1.54 (1.19, 1.99)**	4.01 (1.96, 8.19)***
Transverse	0.83 (0.31, 2.26)	20.46 (2.79, 149.75)**
Delivery mode (CS)	1.06 (0.98, 1.14)	0.58 (0.46, 0.72)***
Five minutes Apgar score (<7)¶¶	2.29 (2.04, 2.57)***	496.61 (240.07, 1027.28)***
Induced labour (Yes)	0.82 (0.75, 0.89)***	1.43 (1.18, 1.73)***
Year	1.04 (1.03, 1.05)***	0.97 (0.95, 0.99)*
Variance of the random effects	0.18 (0.004, 9.09)	0.18 (0.004, 9.09)
Covariance	-0.11 (-0.42, 0.20)	

† N=45,320; ‡ N=45,378

¶ OR: Odds ratios adjusted for maternal age, education level, area of residence, number of ANC visits, referral status, parity, pre-eclampsia/eclampsia, PROM, PPH, abruption placenta, placenta previa, sex of the child, LBW, presentation at birth, delivery mode, five minutes apgar score, labour induction, and year of birth. Parity and PPH were not included in preterm birth model while education level not included in perinatal death prediction.

¶¶ Odds ratio not estimable due to very small number of perinatal deaths among mothers who delivered children with 5-minutes Apgar score of seven and above. Too wide confidence intervals demonstrates low precision of parameter estimates, except for preterm birth only.

* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

being referred for delivery (OR=1.23, 95%CI 1.00, 1.52), PPH (OR=3.16, 95%CI 1.87, 5.33), abruption placenta (OR=2.21, 95%CI 1.15, 4.24), breech presentation (OR=2.58, 95%CI 1.39, 4.77), and labour induction (OR=1.23, 95%CI 1.10, 1.64). Significantly lower odds of perinatal death were among adolescent mothers (OR=0.56, 95%CI 0.40, 0.79), primipara (OR=0.75, 95%CI 0.60, 0.95), experienced PROM (OR=0.39, 95%CI 0.18, 0.88), and delivered through CS (OR=0.72, 95%CI 0.59, 0.87).

Higher odds of preterm birth were among mothers aged 15-19 (OR=1.24, 95%CI 1.09, 1.42), 20-24 (OR=1.16, 95%CI 1.07, 1.26) and 40+ years (OR=1.20, 95%CI 1.00, 1.43), with primary (OR=1.27, 95%CI 1.16, 1.38) and secondary education (OR=1.14, 95%CI 1.02, 1.27), and with inadequate ANC visits (OR=2.79, 95%CI 2.58, 3.00). Likewise, mothers referred for delivery (OR=1.28, 95%CI 1.18, 1.40), experienced pre-eclampsia/eclampsia (OR=1.76, 95%CI 1.53, 2.02), PROM (OR=2.01, 95%CI 1.66, 2.43), and placenta previa (OR=5.32, 95%CI 3.09, 9.16), male children (OR=1.12, 95%CI 1.04, 1.19), delivered LBW baby (OR=9.57, 95%CI 8.48, 10.80), and delivered through CS (OR=1.11, 95%CI 1.03, 1.19). The odds of delivering preterm increased significantly by 1.04 (95%CI 1.04, 1.05) for every one year increase. Lower odds of delivering preterm were among mothers with no education (OR=0.68, 95%CI 0.51, 0.91), resided in rural areas (OR=0.89, 95%CI 0.82, 0.96), and induced labour (OR=0.81, 95%CI 0.74, 0.89).

7.3.4.3 Predictors of independent and co-occurrence of preterm birth and perinatal death using multinomial regression model with separate but correlated random effects

The proportion of co-occurence of preterm birth and perinatal death was 1.7% (N=55,887) while 1.9% experienced perinatal death only, and 9.5% preterm birth only. Results of the joint model presented in Table 7.4 have a single variance component for the three multinomial outcomes. Table 7.5 contains findings of a

Table 7.4: Predictors of independent and co-occurrence of preterm birth and perinatal death using random effect multinomial regression model (N=51,493)

Variable	Co-occurrence†	Perinatal death only	Preterm birth only
	OR‡ (95%CI)	OR‡ (95%CI)	OR‡ (95%CI)
Maternal age groups			
15-19	0.46 (0.29, 0.73)**	0.56 (0.40, 0.79)***	1.24 (1.09, 1.42)***
20-24	1.10 (0.85, 1.43)	1.00 (0.81, 1.24)	1.16 (1.07, 1.26)***
25-29	1.00	1.00	1.00
35-39	1.35 (0.99, 1.84)	1.20 (0.92, 1.56)	1.07 (0.96, 1.19)
40+	1.60 (0.95, 2.68)	1.02 (0.65, 1.61)	1.20 (1.00, 1.43)*
Maternal education			
None	0.41 (0.20, 0.84)*	0.67 (0.39, 1.16)	0.68 (0.51, 0.91)**
Primary	0.93 (0.72, 1.21)	1.09 (0.88, 1.35)	1.27 (1.16, 1.38)***
Secondary	0.77 (0.53, 1.14)	1.22 (0.90, 1.67)	1.14 (1.02, 1.27)*
Higher	1.00	1.00	1.00
Area of residence (Rural)	0.86 (0.69, 1.07)	0.93 (0.77, 1.11)	0.89 (0.82, 0.96)**
ANC visits (<4)	3.46 (2.77, 4.32)***	1.11 (0.92, 1.33)	2.79 (2.58, 3.00)***
Referred for delivery (Yes)	1.27 (1.00, 1.61)	1.23 (1.00, 1.52)*	1.28 (1.18, 1.40)***
Parity (Primipara)	0.63 (0.48, 0.84)**	0.75 (0.60, 0.95)*	0.99 (0.91, 1.09)
Pre-eclampsia/eclampsia (Yes)	1.38 (1.01, 1.89)*	1.24 (0.90, 1.72)	1.76 (1.53, 2.02)***
PROM (Yes)	0.85 (0.34, 2.13)	0.39 (0.18, 0.88)*	2.01 (1.66, 2.43)***
PPH (Yes)	2.24 (1.05, 4.78)*	3.16 (1.87, 5.33)***	0.96 (0.65, 1.42)
Abruption placenta (Yes)	3.98 (2.02, 7.82)***	2.21 (1.15, 4.24)*	1.75 (0.94, 3.27)
Placenta previa (Yes)	0.60 (0.12, 2.95)	1.39 (0.27, 7.17)	5.32 (3.09, 9.16)***
Sex (Male)	0.88 (0.71, 1.08)	0.95 (0.80, 1.13)	1.12 (1.04, 1.19)**
LBW (Yes)	12.81 (9.84, 16.67)***	1.00 (0.79, 1.26)	9.57 (8.48, 10.80)***
Presentation at birth			
Cephalic	1.00	1.00	1.00
Breech	3.79 (2.03, 7.08)***	2.58 (1.39, 4.77)**	1.25 (0.93, 1.68)
Transverse	10.01 (0.49, 204.86)	25.63 (4.76, 137.90)***	1.07 (0.39, 2.90)
Delivery mode (CS)	0.50 (0.40, 0.64)***	0.72 (0.59, 0.87)***	1.11 (1.03, 1.19)**
Five minutes Apgar score (<7)§	466.88 (294.54, 740.05)***	351.03 (261.03, 472.04)***	1.07 (0.81, 1.40)
Induced labour (Yes)	1.23 (0.96, 1.58)	1.34 (1.10, 1.64)**	0.81 (0.74, 0.89)***
Year	1.02 (1.00, 1.05)	0.98 (0.96, 1.00)	1.04 (1.04, 1.05)***
Variance of the random effect	0.04 (0.00, 99.43)		

† Co-occurrence means the occurrence of both preterm birth and perinatal death.

‡ OR: Odds ratios adjusted for maternal age, education level, area of residence, number of ANC visits, referral status, parity, pre-eclampsia/eclampsia, PROM, PPH, abruption placenta, placenta previa, sex of the child, LBW, presentation at birth, delivery mode, five minutes apgar score, labour induction, and year of birth.

§ Odds ratio not estimable due to very small number of perinatal deaths among mothers who delivered children with 5-minutes Apgar score of seven and above. Too wide confidence intervals demonstrates low precision of parameter estimates, except for preterm birth only (last column).

* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

multinomial regression model with separate but correlated random effects. The variance components indicate high variability for the co-occurrence of both outcomes (Var=1.23, 95%CI 0.20, 7.60) than the outcomes occurring independently (Var=0.70, 95%CI 0.04, 11.5 and Var=0.50, 95%CI 0.28, 0.90, for perinatal death and preterm birth, respectively). The covariance between a pair of these outcomes gives no evidence of dependence between perinatal death and preterm birth, conditional on accounting for mother to mother variability. Furthermore, we also observed relatively larger standard errors (especially for the co-occurrence and perinatal death only) for this model (standard errors not shown) than the model with a single variance component. The confidence intervals for the predictors of co-occurrence and perinatal death in Table 7.4 are relatively narrow compared to those in Table 7.5. Also, model comparison using AIC agreed with the results mentioned above. Specifically, the model corresponding to results presented in Table 7.4 had an AIC of 33108.28, which is smaller than 33111.54 for the more complex model corresponding to Table 7.5. Hence, the best model is the random effect multinomial regression model than the one with separate but correlated random effects (more complex, i.e., has additional parameters).

Similar to the descriptions of results in Section 7.3.4.2, conditional on separate random effects for each outcome in the multinomial logits, significantly higher odds of co-occurrence of preterm birth and perinatal death were among mothers with inadequate (<4) ANC visits (OR=3.97, 95%CI 2.79, 5.66), experienced pre-eclampsia/eclampsia (OR=1.44, 95%CI 1.00, 2.06), PPH (OR=2.44, 95%CI 1.02, 5.87), and abruption placenta (OR=4.80, 95%CI 2.16, 10.67), delivered LBW baby (OR=17.00, 95%CI 9.51, 30.36), and had a breech presentation (OR=4.60, 95%CI 2.01, 10.53). Adolescent mothers (15-19 years) (OR=0.40, 95%CI 0.23, 0.72), with no education (OR=0.36, 95%CI 0.16, 0.83), primipara (OR=0.60, 95%CI 0.43, 0.85), and CS delivery (OR=0.45, 95%CI 0.34, 0.60) had lower odds of co-occurrence of preterm birth and perinatal death.

Referral status was no longer significantly associated with perinatal death in this model. The factors independently associated with a higher odds of perinatal death were PPH (OR=3.40, 95%CI 1.92, 6.00), abruption placenta (OR=2.46, 95%CI 1.11, 5.43), breech presentation (OR=2.86, 95%CI 1.30, 6.28), and labour induction (OR=1.38, 95%CI 1.10, 1.72). Significantly lower odds of perinatal death were among adolescent mothers (OR=0.53, 95%CI 0.35, 0.80), primipara (OR=0.72, 95%CI 0.54, 0.97), experienced PROM (OR=0.36, 95%CI 0.14, 0.90), and CS delivery (OR=0.69, 95%CI 0.52, 0.90).

Higher odds of preterm birth were among mothers aged 15-19 (OR=1.24, 95%CI 1.09, 1.42), and 20-24 (OR=1.16, 95%CI 1.07, 1.26), with primary (OR=1.27, 95%CI 1.16, 1.38) and secondary education (OR=1.14, 95%CI 1.02, 1.27), and with inadequate ANC visits (OR=2.79, 95%CI 2.58, 3.00). Likewise, mothers referred for delivery (OR=1.28, 95%CI 1.18, 1.40), experienced pre-eclampsia/eclampsia (OR=1.76, 95%CI 1.53, 2.02), PROM (OR=2.01, 95%CI 1.66, 2.43), and placenta previa (OR=5.36, 95%CI 3.11, 9.25), male children (OR=1.12, 95%CI 1.04, 1.20), delivered LBW baby (OR=9.60, 95%CI 8.51, 10.83), and delivered through CS (OR=1.11, 95%CI 1.03, 1.19). The odds of delivering preterm increased significantly by 1.04 (95%CI 1.04, 1.05) for every one year increase. Lower odds of delivering preterm were among mothers with no education (OR=0.68, 95%CI 0.51, 0.91), resided in rural areas (OR=0.89, 95%CI 0.82, 0.96), and induced labour (OR=0.81, 95%CI 0.74, 0.89).

7.4 Discussion

The study aimed to determine the joint predictors of preterm birth and perinatal death based on the birth cohort data from the KCMC zonal referral hospital in Northern Tanzania between 2000-2017. Conditional on the random effects, higher odds of both preterm birth and perinatal death were among mothers with

Table 7.5: Predictors of independent and co-occurrence of preterm birth and perinatal death using multinomial regression model with separate but correlated random effects (N=51,493)

Variable	Co-occurrence†	Perinatal death only	Preterm birth only
	OR‡ (95%CI)	OR‡ (95%CI)	OR‡ (95%CI)
Maternal age groups			
15-19	0.40 (0.23, 0.72)**	0.53 (0.35, 0.80)**	1.24 (1.09, 1.42)***
20-24	1.11 (0.82, 1.50)	1.00 (0.79, 1.26)	1.16 (1.07, 1.26)***
25-29	1.00	1.00	1.00
35-39	1.42 (0.99, 2.05)	1.23 (0.91, 1.65)	1.07 (0.96, 1.19)
40+	1.75 (0.94, 3.26)	1.04 (0.63, 1.71)	1.19 (1.00, 1.43)
Maternal education			
None	0.36 (0.16, 0.83)*	0.64 (0.34, 1.19)	0.68 (0.51, 0.91)**
Primary	0.91 (0.67, 1.23)	1.09 (0.86, 1.38)	1.27 (1.16, 1.38)***
Secondary	0.74 (0.47, 1.16)	1.23 (0.88, 1.72)	1.14 (1.02, 1.27)*
Higher	1.00	1.00	1.00
Area of residence (Rural)	0.85 (0.65, 1.09)	0.92 (0.75, 1.12)	0.89 (0.82, 0.96)**
ANC visits (<4)	3.97 (2.79, 5.66)***	1.11 (0.88, 1.40)	2.79 (2.58, 3.00)***
Referred for delivery (Yes)	1.30 (0.98, 1.71)	1.26 (0.99, 1.59)	1.28 (1.18, 1.40)***
Parity (Primipara)	0.60 (0.43, 0.85)**	0.72 (0.54, 0.97)*	0.99 (0.91, 1.09)
Pre-eclampsia/eclampsia (Yes)	1.44 (1.00, 2.06)*	1.26 (0.89, 1.79)	1.76 (1.53, 2.02)***
PROM (Yes)	0.87 (0.29, 2.64)	0.36 (0.14, 0.90)*	2.01 (1.66, 2.43)***
PPH (Yes)	2.44 (1.02, 5.87)*	3.40 (1.92, 6.00)***	0.96 (0.65, 1.41)
Abruption placenta (Yes)	4.80 (2.16, 10.67)***	2.46 (1.11, 5.43)*	1.77 (0.93, 3.34)
Placenta previa (Yes)	0.45 (0.07, 2.96)	1.38 (0.24, 7.80)	5.36 (3.11, 9.25)***
Sex (Male)	0.86 (0.68, 1.10)	0.95 (0.79, 1.15)	1.12 (1.04, 1.20)**
LBW (Yes)	17.00 (9.51, 30.36)***	1.01 (0.60, 1.70)	9.60 (8.51, 10.83)***
Presentation at birth			
Cephalic			
Breech	4.60 (2.01, 10.53)***	2.86 (1.30, 6.28)**	1.25 (0.93, 1.69)
Transverse	12.14 (0.50, 296.70)	28.07 (5.64, 139.69)***	1.06 (0.39, 2.86)
Delivery mode (CS)	0.45 (0.34, 0.60)***	0.69 (0.52, 0.90)**	1.11 (1.03, 1.19)**
Five minutes Apgar score (<7)§	750.47 (228.10, 2469.12)***	457.69 (184.43, 1135.84)***	1.05 (0.80, 1.38)
Induced labour (Yes)	1.26 (0.95, 1.66)	1.38 (1.10, 1.72)**	0.81 (0.74, 0.89)***
Year	1.02 (0.99, 1.06)	0.98 (0.95, 1.00)	1.04 (1.04, 1.05)***
Variance of the random effects	1.23 (0.20, 7.60)	0.70 (0.04, 11.5)	0.50 (0.28, 0.90)
Covariances			
Cov(1,2)	0.54 (-0.78, 2.86)		
Cov(1,3)	0.12 (-0.51, 0.74)		
Cov(2,3)	0.20 (-0.34, 0.73)		

† Co-occurrence means the occurrence of both preterm birth and perinatal death.
‡ OR: Odds ratios adjusted for maternal age, education level, area of residence, number of ANC visits, referral status, parity, pre-eclampsia/eclampsia, PROM, PPH, abruption placenta, placenta previa, sex of the child, LBW, presentation at birth, delivery mode, five minutes apgar score, labour induction, and year of birth.
§ Odds ratio not estimable due to very small number of perinatal deaths among mothers who delivered children with 5-minutes Apgar score of seven and above. Too wide confidence intervals demonstrates low precision of parameter estimates, except for preterm birth only (last column).
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

inadequate (<4) ANC visits, referred for delivery, experienced abruption placenta, and breech presentation. Mothers with inadequate ANC visits, who experienced pre-eclampsia/eclampsia, PPH, and abruption placenta, delivered LBW, and experienced breech presentation had a higher likelihood of co-occurring both preterm birth and perinatal death. Lower odds of co-occurrence were among adolescent mothers (15-19), with no education, primipara, and those delivered through CS. Mothers aged 15-19, experienced PROM and placenta previa, and delivered male children had higher odds of preterm birth but were less likely to experience perinatal death.

Inadequate ANC visits increased the risk of both preterm birth and perinatal death. Previous studies on independent predictors of these outcomes support this finding (Mahapula et al., 2016; Mboya et al., 2020b, 2021; Lu et al., 2015a; Rugaimukam et al., 2017; Temu et al., 2016). According to WHO, "within the continuum of reproductive health care, ANC provides a platform for important health-care functions, including health promotion, screening and diagnosis, and disease prevention" (World Health Organization, 2016b). Tanzania's local and national efforts should promote good healthcare-seeking behaviours during pregnancy and improved coverage and quality of antenatal care services at all levels of care (Benova et al., 2018; Darmstadt et al., 2009). It is also essential to improve intrapartum and postnatal care quality, particularly for women who experienced pregnancy and delivery-related complications (Baqui et al., 2016; Balkus et al., 2021; Iams et al., 2008; UNICEF et al., 2020).

Women referred for delivery had higher odds of preterm birth and perinatal death. Pregnant women referred for delivery are more likely to experience delivery-related complications, where adolescent mothers have elevated risk (Grønvik & Fossgard Sandøy, 2018). In this study, adolescent mothers (15-19 years), primipara, and those with no education were less likely to experience co-occurring

preterm birth and perinatal death. However, the joint random effects model (conditional on the mother-to-mother variability) revealed that those aged 15-19 and 20-24 years were more likely to deliver preterm but had a lower odds of perinatal death. However, the protective effect of 20-24 years of age on the risk of perinatal death was not statistically significant. CS delivery lowered the odds of co-occurrence, which may reflect timely care of these high-risk pregnancies to save both the mother and child's life.

Conditional on the random effects, significantly higher odds of preterm birth and perinatal death, and co-occurrence were among mothers who experienced abruption placenta and breech presentation. Additionally, pre-eclampsia/ eclampsia, PPH, and LBW increased the likelihood of co-occurrence. On top of these complications being among the common risk factors of preterm birth (Mboya et al., 2021; Purisch & Gyamfi-Bannerman, 2017; Rugaimukam et al., 2017; Vogel et al., 2018) and perinatal death (Chaibva et al., 2019; Mboya et al., 2020b,a; Nijkamp et al., 2017; Vogel et al., 2014), they also increase the risk of newborns transfer to intensive care units (Mmbaga et al., 2011). Given their history, women at risk of these adverse pregnancy events should be given due public health and clinical attention and care during antenatal, intrapartum, and postnatal periods. Although we did not assess health system performance regarding pregnancy and childcare, efforts are needed to strengthen health facilities providing delivery services in Tanzania for improved pregnancy outcomes (Akombi & Renzaho, 2019; Koffi et al., 2020; UNICEF et al., 2020).

The study had several strengths compared to previous studies. Firstly, this is the first study in Tanzania and potentially in SSA to assess the joint predictors of preterm birth and perinatal death, to the best of our knowledge. The vast majority of previous studies focused on determining the independent predictors of preterm birth and perinatal death or the determinant of each other. Secondly, joint

modelling using random effects approach accounted for the relationship between the two outcomes for improved precision of parameter estimates. Nevertheless, conditional on the random effects, we observed no statistically significant covariance between preterm birth and perinatal death. In other words, the two outcomes are independent conditional on accounting for mother-to-mother variability.

As we explained elsewhere (Mboya et al., 2020a,b, 2021), the study has several limitations. Data for this study comes from the KCMC zonal referral hospital in northern Tanzania, affecting the generalization of findings. However, less than a quarter (23.8%) of all recorded deliveries were referrals. Hence the study findings may reflect prenatal and intrapartum care practices and adverse events among deliveries from women in the hospital's catchment area, similar settings in Tanzania and SSA. Also, the fact that KCMC medical birth registry cohort only captures perinatal deaths occurring in the health facility (KCMC hospital), may underestimate the reported perinatal death proportions/rates (Mboya et al., 2020a).

7.5 Conclusion

The joint predictors of preterm birth and perinatal death were inadequate (<4) ANC visits, referred for delivery, and complications during pregnancy and childbirth, specifically pre-eclampsia/ eclampsia, PPH, LBW, abruption placenta, and breech presentation. Younger maternal age (15-24 years), PROM, placenta previa, and male children have higher odds of preterm birth but a lessened likelihood of perinatal death. ANC is a critical entry point for delivering the recommended interventions to pregnant women (World Health Organization, 2016b), especially those at high risk of experiencing adverse pregnancy outcomes. Improved management of complications during pregnancy and childbirth and the postnatal period may eventually lead to a substantial reduction of adverse perinatal outcomes and improving maternal and child health.

Chapter 8

General discussion, conclusion, and recommendations

8.1 Discussion of the main findings

The main objective for this study was to determine the joint predictors of preterm birth and perinatal death among singleton birth in northern Tanzania based on medical birth registry data from KCMC zonal referral hospital between 2000 and 2017. We also applied the novel statistical methodology to handle missing data, which reduce the bias introduced by ignoring missing values in the data set. Detailed discussions of the main findings are presented separately in Chapters 4 to 7. The focus of this chapter is to provide a summary discussion of these findings, followed by the study strengths and limitations, conclusions, and recommendations.

8.1.1 Proportions/rates of preterm birth and its predictors

The overall proportion of preterm birth among singleton births recorded in the KCMC medical birth registry between 2000-2017 years was 12.8%. Previous studies in northern Tanzania reported similar estimates (Mahande et al., 2013a; Rugaimukam et al., 2017; Temu et al., 2016), which is slightly higher than the

estimate in SSA (12.3%) reported about ten years ago (March of Dimes et al., 2012) and the recent global estimate of 10.6% (Chawanpaiboon et al., 2019). We observed significant rise in trends of preterm birth over the years, with the highest increase among the moderate to late preterm (32 to <37) (Mboya et al., 2021). Our finding is consistent to the global and regional estimates (March of Dimes et al., 2012; Blencowe et al., 2013; Chawanpaiboon et al., 2019; Lu et al., 2015a; World Health Organization, 2020). Current estimates positioned Tanzania as the tenth country with the highest preterm birth rates globally (Chawanpaiboon et al., 2019).

As reported in previous studies, the factors associated with higher risk of preterm birth include younger maternal age, particularly adolescents (15-19 years) and youths (20-24 years) (Fuchs et al., 2018; Grantz et al., 2015; Lu et al., 2015a; Mboya et al., 2021; van den Broek et al., 2014; Zack et al., 2014), inadequate (<4) ANC visits (Lu et al., 2015a; Mahapula et al., 2016; Rugaimukam et al., 2017; Temu et al., 2016), women referred for delivery (Grønvik & Fossgard Sandøy, 2018), primiparous (Kozuki et al., 2013), maternal low education level (Kim et al., 2018; Ruiz et al., 2015; Vogel et al., 2018), delivery of male children (Teoh et al., 2018; Zhang et al., 2018a), and delivery through CS (Tarimo et al., 2020; Lisonkova et al., 2012; Worjoloh et al., 2012). The association between low education level and increased risk of preterm birth has been linked to low socioeconomic status, which increases the risk of adverse pregnancy outcomes and complications (Kim et al., 2018; Ruiz et al., 2015).

The association between preterm birth and CS delivery in this cohort may be due to pregnancy complications such as pre-eclampsia/eclampsia, PROM, placenta previa, and abruption placenta, which necessitate medical intervention such as CS to save the life of the mother, child, or both. Similar to other studies (Blencowe et al., 2013; Georgiou et al., 2015; Lu et al., 2015a; Requejo et al., 2013; Vogel et al., 2018), these complications were also associated with a higher preterm birth risk in this study and are also common indications of preterm delivery (Georgiou et al.,

2015; Lu et al., 2015a; Purisch & Gyamfi-Bannerman, 2017; Requejo et al., 2013; Vogel et al., 2018). LBW babies were more likely to be born preterm, i.e., nearly 40-times higher risk of very/extremely preterm (<32 gestational weeks) and eight-times higher risk of moderately to late preterm (32-<37 gestational weeks). A case-control study in the KCMC hospital reported similar results (Temu et al., 2016). Preterm babies are likely to be born with LBW (Katz et al., 2013; Mitao et al., 2016), hence increasing the risk of perinatal and neonatal morbidity and mortality (Katz et al., 2013; Mboya et al., 2020b; UNICEF et al., 2020; Unterscheider et al., 2014). Therefore, clinicians should give special attention to the LBW babies born at <32 weeks of gestation towards increasing child survival.

8.1.2 Proportions/rates of perinatal death and its predictors

The perinatal death rate (PMR) between 2000-2017 years in the KCMC medical birth registry was 42.6 per 1000 births. We also observed the decline in trends of perinatal deaths over the years during the study period. The PMR in this study is higher than the national estimate of 39 per 1000 births (MoHCDGEC [Tanzania Mainland] et al., 2016), and 27 per 1000 births in Manyara region, northern Tanzania (Hinderaker et al., 2003). The high proportion of PMR in our study could be attributed to the referral nature of the study population, where the majority of high-risk mothers are delivered compared to lower-level and population-based surveys. A previous study using a similar dataset in 2000-2010 reported a slightly higher PMR of 57.7 per 1000 births (Mmbaga et al., 2012b), demonstrating a decline compared to our estimate. The observed differences from this study may be because Mmbaga et al. (2012b) restricted the analysis to neonatal deaths above 500 grams and included linked records from the neonatal registry. The proper linkage between the neonatal registry and the KCMC medical birth registry data in the reproductive and child health unit is essential for a reliable perinatal mortality estimation. But still, more efforts are needed to accelerate PMR reduction both at the KCMC hospital and other similar settings in Tanzania to improve newborn

survival.

The PMR estimated in our study is lower than those reported in India, Nigeria, and Sudan (Ali et al., 2014; Bellad et al., 2010; Oyira et al., 2017). The differences in estimates between our study and previous studies could be due to differences in sample size and study population such as that from a community-based cross-sectional study in eastern Sudan (Ali et al., 2014). High-income countries documented a nearly six-times lower PMR (Gregory et al., 2018; Unterscheider et al., 2014), which can be linked to the availability and quality of obstetric, newborn, and specialized care services. Furthermore, we found no significant differences in perinatal death and stillbirth predictors (i.e., stillbirths and perinatal deaths had shared risk factors). Women and their newborns were essentially discharged soon after birth; hence, deaths occurring at home could not be recorded, consistent with a previous study (Bailey et al., 2017).

Similar to other studies, the higher risk of perinatal death was observed among mothers with advanced maternal age (>35 years) (Nijkamp et al., 2017), from rural areas of residence, with low education level (Bellad et al., 2010; Blumenshine et al., 2010; Nijkamp et al., 2017), inadequate (<4) ANC visits, male compared to female children (Miranda et al., 2017), and those referred for delivery (Mahande et al., 2013a; Mmbaga et al., 2012b; Vogel et al., 2014). We also stratified our analysis based on referral status and found similar results except for the area of residence, education level, and sex of the child, which were not statistically significant among those referred for delivery. Compared to women who were referred for delivery in the study setting, there was a stronger covariates effect on perinatal death risk in the group of women not referred for delivery (Mboya et al., 2020b). Nevertheless, PMR is higher in women referred for delivery because they are at increased risk of experiencing adverse pregnancy outcomes and could reflect the delay in seeking care or referral. Therefore, it is imperative to strengthen the referral system and

improve care for women referred for delivery at KCMC hospital (Mmbaga et al., 2012b). In addition, women not referred for delivery also needs special attention to improve maternal and newborns survival.

In the present study, the risk of perinatal death was also high among mothers who experienced pre-eclampsia/eclampsia, PPH, abruption placenta, delivered preterm, and LBW baby (Bellad et al., 2010; Mpembeni et al., 2014; Nijkamp et al., 2017; Vogel et al., 2014). PROM and CS delivery offered protection against the risk of perinatal death. CS is known to reduce the risk of complications where medically indicated (Vogel et al., 2014). At the same time, the protective effect of PROM could reflect timely management of high-risk pregnancies at the KCMC hospital. Routine monitoring of women at high risk of experiencing adverse pregnancy outcomes and complications during prenatal period is essential to prevent avoidable complications.

We also applied machine learning approaches to predict perinatal deaths. Although there were no differences in the predictive capacity between machine learning algorithms and the logistic regression model, the former approach had a higher net benefit. In addition to maternal characteristics discussed above, paternal characteristics, such as age, education level, and occupation were also important predictors of perinatal death (Mboya et al., 2020a). The association between paternal characteristics and adverse perinatal outcomes has been well documented (Hurley & DeFranco, 2017; Khandwala et al., 2018; Meng & Groth, 2018; Tough et al., 2003). The previous investigators have emphasized on the need for paternal involvement in pregnancy and child care in Tanzania and similar settings (Gibore & Bali, 2020; Peneza & Maluka, 2018). Therefore, "machine learning algorithms may improve the prediction ability of perinatal deaths, and enable triage of women who are at high risk of experiencing adverse perinatal outcomes" (Mboya et al., 2020a).

8.1.3 Joint predictors of preterm birth and perinatal death

We aimed to determine the joint predictors of preterm birth and perinatal death using birth cohort data from the KCMC zonal referral hospital in northern Tanzania between 2000-2017. We found that conditional on the random effects, the joint predictors of preterm birth and perinatal death were inadequate (<4) ANC visits, referred for delivery, and complications during pregnancy and childbirth, particularly pre-eclampsia/ eclampsia, PPH, LBW, abruption placenta, and breech presentation. Younger maternal age (15-24 years), PROM, placenta previa, and sex of the child (male) have higher odds of preterm birth but a lessened likelihood of perinatal death.

The previous studies assessed the predictors of preterm birth and perinatal death independently. There were no studies reported on the joint predictors of preterm birth and perinatal death to compare with our findings. Nevertheless, existing literature supports these findings as independent predictors of preterm birth (Mboya et al., 2021; Purisch & Gyamfi-Bannerman, 2017; Rugaimukam et al., 2017; Temu et al., 2016; Vogel et al., 2018) and perinatal death (Chaibva et al., 2019; Mboya et al., 2020b,a; Nijkamp et al., 2017; Vogel et al., 2014). Therefore, our findings suggest a need to have focused care among women most at risk of experiencing adverse pregnancy and delivery complications in Tanzania and similar settings in SSA to reduce the risk of preterm birth and perinatal death. "Early identification and management of women with complications could improve maternal and perinatal outcomes" (Vogel et al., 2014).

The random-effects models for the joint predictors of preterm birth and perinatal death and co-occurring using multinomial models showed that there is no evidence of dependence between the two outcomes assessed in this study. In other words, the two outcomes were independent after accounting for mother-to-mother variability. However, there were few women who experienced both preterm birth

and perinatal death that might have affected estimation of within-mother correlation. For instance, among all recorded deliveries between 2000-2015 (N=50847), there were 805 (1.6%) deliveries which experienced both preterm birth and perinatal death and 1.7% (N=55,887) between 2000-2017. This might have been a reason for difficulties in estimating the within-mother correlation.

Furthermore, it is worth noting that convergence of the joint random effects model (results described in section 7.3.4.1) was only achieved upon inclusion of preterm birth as an independent predictor of perinatal death. Indeed, literature also demonstrated that preterm birth complications are a leading cause of perinatal and neonatal deaths globally (Chawanpaiboon et al., 2019; Hug et al., 2019; Koffi et al., 2020; Khan et al., 2020; UNICEF et al., 2020). Based on this fact, a joint random-effects model including preterm birth as a predictor of perinatal death provided additional information about the existing relationship between the two variables. Therefore, our findings have indicated that preterm births had nearly two-fold the risk of perinatal death (OR=1.69, 95%CI 1.04, 2.75) compared to term births, conditional on the random effects and adjusted for other factors. In contrast, the model with perinatal death as a predictor of preterm birth did not converge.

The choice and application of a statistical model depend on the research question one would like to address. This thesis focused on determining joint predictors of preterm birth and perinatal death. Other joint modelling frameworks, such as the copula models (Nikoloulopoulos & Karlis, 2008; Klein et al., 2019) are complex in theory but feasible in the application were explored but not applied in this work, hence remains open for future extensions.

8.2 Study strengths and limitations

8.2.1 Study strengths

The study utilized data from a large birth cohort of women with their subsequent deliveries at KCMC zonal referral hospital in Northern Tanzania. The KCMC medical birth registry is one among a few birth/pregnancy cohorts in SSA (Bone et al., 2020; Campbell & Rudan, 2011) providing essential data for administrative, clinical, public health, and potential policy decisions (Bergsjo et al., 2007; Mulder & Spicer, 2019). These registries provides opportunities for inter-generational record linkages (Lamont et al., 2021). The large sample sizes and wider population coverage enhances statistical power and the generalizability of our findings, especially when our findings are consistent with the population/community-based studies (Ali et al., 2014; Hinderaker et al., 2003; Nankabirwa et al., 2011; van den Broek et al., 2014).

Missing data is a common problem in clinical, longitudinal, and registry-based studies, including those conducted in hospital settings (Pedersen et al., 2017; Sterne et al., 2009). To our knowledge, no study in Tanzania and possibly in SSA has assessed the effect of ignoring missing values on determining predictors of adverse pregnancy outcomes, such as preterm birth and perinatal death. Data analysis in this study accounted for missing data in prediction of these outcomes (Mboya et al., 2020b,a), hence improved precision of parameter estimates (Jakobsen et al., 2017; Pedersen et al., 2017; Sterne et al., 2009).

Likewise, this is the first study to apply modern machine learning approaches to predict perinatal deaths in Tanzania and to a large extent SSA. Previous studies using the KCMC medical birth registry data (Chuwa et al., 2017; Isaksen et al., 2015; Mahande et al., 2013a,b; Mitao et al., 2016; Mmbaga et al., 2011, 2012a) applied standard regression models to assess risk factors for adverse perinatal outcomes. This study extended the analysis by applying the machine learning algorithms

compared to the classical logistic regression model to predict perinatal deaths (Mboya et al., 2020a). The machine learning models can, therefore, be used to improve the prediction of perinatal deaths and triage for women at risk.

8.2.2 Study limitations

First, the birth registry data are often faced with a missing data problem and unmeasured factors/variables/covariates, which increase the potential for residual confounding (Mahande, 2015). Maternal conditions such as periodontitis (Ren & Du, 2017) and genetic components of gestational duration (Zhang et al., 2018a) plays a critical role on the risk of preterm birth and potentially perinatal death. However, these factors were not captured in the KCMC medical birth registry, and therefore, their effects on these outcomes were not evaluated.

Secondly, the KCMC medical birth registry contains information for women and their siblings who were delivered in the obstetrics and gynaecology department at the KCMC hospital. However, the follow-up is limited to the first week after delivery. This implies that any adverse event occurring outside the hospital environment after a woman is discharged, such as early neonatal deaths (before seven days after birth), are not captured in the registry. Stillbirths accounted for the largest proportion (about 3.4%) of all perinatal deaths compared to early neonatal deaths (about 0.5%). Thus, there is potential for underestimating the burden of adverse events such as perinatal death. For instance, stillbirth contributed to a higher percentage of perinatal death numbers in this study, which accounted for the unobserved differences between perinatal death and stillbirth predictors (Mboya et al., 2020b). Findings from birth cohorts should, therefore, be complemented with population or community-based surveys, such as the demographic and health surveys, especially in settings with large proportions of home deliveries (MoHCDGEC [Tanzania Mainland] et al., 2016).

Thirdly, our study is a hospital-based study from a tertiary care health facility in northern Tanzania. Therefore, our findings may not be generalized to lower-level health facilities and settings in Tanzania, mainly because there is potential for selection or referral bias. Despite this limitation, our findings on the distribution and determinants of preterm birth and perinatal death agree with previous hospital (Ahankari et al., 2001; Mahapula et al., 2016; Mpembeni et al., 2014) and population/community-based studies (Ali et al., 2014; Hinderaker et al., 2003; Katz et al., 2013; Nankabirwa et al., 2011; van den Broek et al., 2014). These demonstrate that triangulation of findings from multiple data sources within and across countries is critical to informing context-specific and cost-effective interventions and policy decisions to reduce the burden of preterm birth, perinatal deaths, and other avoidable adverse pregnancy outcomes.

8.3 Conclusion

The proportion of preterm birth between 2000-2017 in our birth cohort was 12.8% which continue to rise over time. At the same time, the proportion of perinatal death proportion was 4.3%. This corresponds to a PMR of 42.6 per 1000 births. Over the years, we observed a slight decline in perinatal death estimates. The joint predictors of higher risk of preterm birth and perinatal death were inadequate (<4) ANC visits, referred for delivery, and complications during pregnancy and childbirth, specifically pre-eclampsia/eclampsia, PPH, LBW, abruption placenta, and breech presentation. Younger maternal age (15-24 years), PROM, placenta previa, and male children have higher odds of preterm birth but a lessened likelihood of perinatal death.

In addition, maternal characteristics, pregnancy-related conditions, complications, and delivery characteristics were significantly associated with an increased risk of preterm birth and perinatal death (Mboya et al., 2020a,b, 2021). Paternal characteristics such as advanced paternal age and education level were also

potential predictors of perinatal death (Mboya et al., 2020a). It is essential to account for missing data in birth/pregnancy cohorts and longitudinal studies in predicting adverse pregnancy outcomes for improved precision of parameter estimates. Furthermore, the choice of data analysis model is also as essential as the research problem/question of interest for improved precision of parameter estimates.

8.4 Recommendations

Ensuring child survival is at the heart of the sustainable development goals agenda (UNDP, 2018). Perinatal mortality is an indicator of the health system performance during pregnancy, delivery, and postnatal period (Akombi & Renzaho, 2019). Concerted efforts are needed to articulate and address the health system and provided related challenges in delivering quality reproductive, maternal, and newborn care services to avert the rising trends of preterm birth and further reduce perinatal deaths in Tanzania. The study recommends the following interventions in the delivery of quality antenatal, intrapartum, and postnatal care services in Tanzania and elsewhere towards improving child survival.

8.4.1 Clinical implications

Special attention should be given to first-time parents, especially mothers aged 15-24 years, referred for delivery, and those who experienced complications during pregnancy and childbirth such as pre-eclampsia/eclampsia, PPH, LBW delivery, abruption placenta, placenta previa, PROM, and breech presentation. Child bearing in a younger age such as during the adolescence period increases the risk of pregnancy complications and adverse events (Grønvik & Fossgard Sandøy, 2018).

Regardless of maternal age, first-time parents may not be aware of the critical need for the first ANC visit, which could be among the causes for poor ANC attendance. Therefore, focused care of women who experienced or are at high risk of

experiencing pregnancy complications and adverse events is necessary to reduce the risk of preterm birth and perinatal death. In addition, there is also a need for strengthening antenatal and postnatal education at the facility levels to both first-time parents, early and practical information about parenting skills, and the opportunity to seek support and help from health professionals when needed (Entsieh & Hallström, 2016).

8.4.2 Policy implications

At the national level, policy decisions should intensify and sustain efforts to increase the uptake and quality of ANC at all levels of care in Tanzania. ANC is a critical entry point where several interventions can easily be delivered to pregnant women (World Health Organization, 2016a), especially those at high-risk of experiencing adverse pregnancy outcomes such as preterm birth and perinatal death. "ANC is also an opportunity to promote the use of skilled attendance at birth and healthy behaviours such as breastfeeding, early postnatal care, and planning for optimal pregnancy spacing" (Lincetto et al., 2006). Benova et al. (2018) emphasized on the significance of the content/components of ANC care on top of the timing and the number of visits. In 2016, WHO issued new recommendations on antenatal care, which included a minimum of eight ANC contacts with the first contact scheduled to take place in the first trimester (up to 12 weeks of gestation) (World Health Organization, 2016a). In Tanzania, over half (51%) of pregnant women had at least four ANC visits during their last pregnancy, but only 24% started ANC before the fourth month of pregnancy (MoHCDGEC [Tanzania Mainland] et al., 2016). Therefore, it is evident that more efforts are needed to increase the coverage of ANC services in the country. Such interventions should continue emphasizing continued male involvement in pregnancy and child birth towards improving maternal and child health outcomes.

Although not measured in our study, routine and sustainable educational

campaigns at the health facility setting and the general public, especially using mass media and skilled providers, should be strengthened and implemented in Tanzania. These should include information on when to attend ANC, the danger signs, the relevance of essential recommended interventions, the significance of attending scheduled visits, and the benefit of skilled health professionals (doctor, nurse or midwives) during delivery (World Health Organization, 2016a).

We recommend establishing birth registries in other zonal referral hospitals and, if possible, regional hospitals in Tanzania and across countries in low- and middle-income countries where such registries are scarce. We want to reiterate what Bergsjo et al. (2010) said ten years ago that; "extending the birth registry monitoring system to all health institutions with obstetrical services in a region will give more reliable estimates to be followed over time and serve as a basis for regular auditing, to the benefit of mothers and their children". Lessons learnt from the KCMC medical birth registry can be used to inform sustainable implementation in other health facilities.

A previous study at KCMC referral hospital documented data quality and reporting challenges in the medical birth registry (Bergsjo et al., 2010), as also observed in our study (Mboya et al., 2020b). Therefore, a need for high-quality data from birth registries, population/community-based surveys, longitudinal studies, and randomized controlled clinical trials (Hug et al., 2019; Lisonkova et al., 2012; UNICEF et al., 2020) locally and at the national level is not over-emphasized. High-quality and reliable data are essential to inform administrative and clinical decisions and epidemiological monitoring and surveillance of adverse pregnancy outcomes. These data should go together with standardisation of definitions, measurement, and reporting of adverse outcomes such as preterm birth (Chawanpaiboon et al., 2019; Lisonkova et al., 2012), stillbirths, and early neonatal deaths (Bailey et al., 2017; Blencowe et al., 2016; Hug et al., 2019; Lawn et al., 2011).

8.4.3 Recommendations to the KCMC hospital management

We encourage the KCMC zonal referral hospital management and partners to establish sustainable routine data quality checks and monitoring mechanisms for the medical birth registry. For the past 20 years (since its establishment in the year 2000), the KCMC medical birth registry has become a critical source of information for administrative, clinical, research, and public health decisions. It is, therefore, high time to update the registry based on the current international classification for diseases frameworks, such as the ICD-PM (World Health Organization, 2016a) and determinants of adverse pregnancy outcomes. Electronic data collection systems are available and should be explored and implemented to ensure efficient and effective quality data collection, management, and monitoring (Lamont et al., 2021).

We also found record linkage challenges in the KCMC medical birth registry. For instance, we restricted our analysis for objectives 1, 2, and 4 for cohort data between 2000-2015 because about half of maternal records could not be linked to their siblings. Also, despite analysing data for up to 2017 for objective 3, the same record linkage issues persisted. In addition, we also noted that there are currently no linkages between the medical birth registry in the reproductive and child health centre and the neonatal registry data located in the paediatric department. For this reason, there is potential for underestimation of adverse pregnancy outcomes such as perinatal deaths occurring at the KCMC referral hospital. We, therefore, encourage the hospital management to create mechanisms for improving record linkages.

8.4.4 Future research

Future studies should consider longitudinal follow-up of births occurring both at health facility and the community environment. Such studies are essential to document the burden and identify the health system challenges and barriers to providing timely and quality maternal and child health services.

The application of machine learning and deep learning algorithms to pregnancy-related adverse events is an area for future extensions. In addition, future extensions may include applying other statistical methodologies, including structural equation modelling, to a particular data problem or research question. Such analyses should consider available methods to handle missing data. Standardization of data collection platforms and pooled analysis of birth cohort data from SSA may also highlight the quality of maternal and childcare services and inform interventions to improve birth outcomes (Bone et al., 2020).

There are alternative methods for dealing with missing data in the statistical literature, but they were not the focus of this thesis. These include probability weighting, maximum likelihood estimations, and multiple correspondence analysis (Sterne et al., 2009; Pedersen et al., 2017). While these methods are attractive and novel to handle missing data, they will be compared to multiple imputation as a future extension of the work.

References

Agresti, A. (2010). *Analysis of ordinal categorical data*, vol. 656. John Wiley & Sons.

Ahankari, A., Bapat, S., Myles, P., Fogarty, A., & Tata, L. (2001). Factors associated with preterm delivery and low birth weight: a study from rural maharashtra, india. *F1000Research*, *6*(72), 1–11.

Akombi, B. J., & Renzaho, A. M. (2019). Perinatal mortality in sub-saharan africa: a meta-analysis of demographic and health surveys. *Annals of global health*, *85*(1).

Ali, A. A. A., Elgessim, M. E., Taha, E., & Adam, G. K. (2013). Factors associated with perinatal mortality in kassala, eastern sudan: a community-based study 2010–2011. *Journal of tropical pediatrics*, *60*(1), 79–82.

Ali, A. A. A., Elgessim, M. E., Taha, E., & Adam, G. K. (2014). Factors associated with perinatal mortality in kassala, eastern sudan: a community-based study 2010–2011. *Journal of tropical pediatrics*, *60*(1), 79–82.

Alijahan, R., Hazrati, S., Mirzarahimi, M., Pourfarzi, F., & Hadi, P. A. (2014). Prevalence and risk factors associated with preterm birth in ardabil, iran. *Iranian journal of reproductive medicine*, *12*(1), 47.

Allanson, E., Tuncalp, O., Gardosi, J., Pattinson, R. C., Erwich, J. J., et al. (2016). Classifying the causes of perinatal death. *Bulletin of World Health Organization*, *94*(2), 79–79A.

Aloisio, K. M., Micali, N., Swanson, S. A., Field, A., & Horton, N. J. (2014). Analysis of partially observed clustered data using generalized estimating equations and multiple imputation. *The Stata Journal*, *14*(4), 863–883.

Ananth, C. V., & Kleinbaum, D. G. (1997). Regression models for ordinal responses: a review of methods and applications. *International journal of epidemiology*, *26*(6), 1323–1333.

Andrinopoulou, E.-R. (2014). *Joint Modelling of Longitudinal and Survival Data with Applications in Heart Valve Data*. Ph.D. thesis, Erasmus Universiteit Rotterdam.

Arfan, M., & Sherwani, R. A. K. (2017). Ordinal logit and multilevel ordinal logit models: An application on wealth index mics-survey data. *Pakistan Journal of Statistics and Operation Research*, *13*(1), 211–226.

Asar, , Ritchie, J., Kalra, P. A., & 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.

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*, *20*(1), 40–49.

Bailey, P. E., Andualem, W., Brun, M., Freedman, L., Gbangbade, S., Kante, M., Keyes, E., Libamba, E., Moran, A. C., Mouniri, H., et al. (2017). Institutional maternal and perinatal deaths: a review of 40 low and middle income countries. *BMC pregnancy and childbirth*, *17*(1), 295.

Balkus, J. E., Neradilek, M., Fairlie, L., Makanani, B., Mgodi, N., Mhlanga, F., Nakabiito, C., Mayo, A., Harrell, T., Piper, J., et al. (2021). Assessing pregnancy and neonatal outcomes in malawi, south africa, uganda, and zimbabwe: Results from a systematic chart review. *PloS one*, *16*(3), e0248423.

Baptista, F. H., Rocha, K. B. B., Martinelli, J. L., Avó, L. R. d. S. d., Ferreira, R. A., Germano, C. M. R., & Melo, D. G. (2017). Prevalence and factors associated with

alcohol consumption during pregnancy. *Revista Brasileira de Saúde Materno Infantil*, *17*(2), 271–279.

Baqui, A. H., Mitra, D. K., Begum, N., Hurt, L., Soremekun, S., Edmond, K., Kirkwood, B., Bhandari, N., Taneja, S., Mazumder, S., et al. (2016). Neonatal mortality within 24 hours of birth in six low-and lower-middle-income countries. *Bulletin of the World Health Organization*, *94*(10), 752.

Bartlett, J. W., Seaman, S. R., White, I. R., & Carpenter, J. R. (2015). Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Statistical methods in medical research*, *24*(4), 462–487.

Bater, J., Lauer, J. M., Ghosh, S., Webb, P., Agaba, E., Bashaasha, B., Turyashemererwa, F. M., Shrestha, R., & Duggan, C. P. (2020). Predictors of low birth weight and preterm birth in rural uganda: Findings from a birth cohort study. *PloS one*, *15*(7), e0235626.

Bayou, G., & Berhan, Y. (2012). Perinatal mortality and associated risk factors: a case control study. *Ethiopian Journal of Health Sciences*, *22*(3), 153–162.

Bellad, M., Srividhya, K., Ranjit, K., & Dhaded, S. (2010). Factors associated with perinatal mortality: a descriptive observational study. *Journal of South Asian Federation of Obstetrics and Gynecology*, *2*(1), 49–51.

Benova, L., Tunçalp, Ö., Moran, A. C., & Campbell, O. M. R. (2018). Not just a number: examining coverage and content of antenatal care in low-income and middle-income countries. *BMJ global health*, *3*(2), e000779.

Bergsjo, P., Mlay, J., Lie, R. T., Lie-Nielsen, E., & Shao, J. F. (2007). A medical birth registry at kilimanjaro christian medical centre. *East Afr J Public Health*, *4*(1), 1–4.

Bergsjo, P., Mlay, J., Oneko, O., & Shao, J. (2010). *Instructions manual for registration of deliveries at the KCMC Birth Registry*. Kilimanjaro Christian Medical Center, Moshi, Tanzania and Department of Public health and Primary Health Care of the University of Bergen, Bergen, Norway.

Bergsma, W., Croon, M., & Hagenaars, J. A. (2009). *Marginal Models for Dependent, Clustered and Longitudinal Categorical Data*. Springer, New York, USA.

Blagus, R., & Lusa, L. (2013a). Improved shrunken centroid classifiers for high-dimensional class-imbalanced data. *BMC bioinformatics*, *14*(64).

Blagus, R., & Lusa, L. (2013b). Smote for high-dimensional class-imbalanced data. *BMC bioinformatics*, *14*(106).

Blencowe, H., Cousens, S., Chou, D., Oestergaard, M., Say, L., et al. (2013). Born too soon: the global epidemiology of 15 million preterm births. *Reproductive health*, *10*(1), S2.

Blencowe, H., Cousens, S., Jassir, F. B., Say, L., Chou, D., Mathers, C., Hogan, D., Shiekh, S., Qureshi, Z. U., You, D., et al. (2016). National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. *The Lancet Global Health*, *4*(2), e98–e108.

Blumenshine, P., Egerter, S., Barclay, C. J., Cubbin, C., & Braveman, P. A. (2010). Socioeconomic disparities in adverse birth outcomes: a systematic review. *American journal of preventive medicine*, *39*(3), 263–272.

Bone, J. N., Pickerill, K., Kinshella, M.-L. W., Vidler, M., Craik, R., Poston, L., Stones, W., Sevene, E., Temmerman, M., Etyang, A. K., et al. (2020). Pregnancy cohorts and biobanking in sub-saharan africa: a systematic review. *BMJ global health*, *5*(11), e003716.

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

Burstein, R., Henry, N. J., Collison, M. L., Marczak, L. B., Sligar, A., Watson, S., Marquez, N., Abbasalizad-Farhangi, M., Abbasi, M., Abd-Allah, F., et al. (2019). Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. *Nature*, *574*(7778), 353–358.

Campbell, A., & Rudan, I. (2011). Systematic review of birth cohort studies in africa. *Journal of global health*, *1*(1), 46.

Carvalho, C. A., da Silva, A. A., Victora, C., Goldani, M., Bettiol, H., Thomaz, E. B. A. F., Barros, F., Horta, B. L., Menezes, A., Cardoso, V., et al. (2020). Changes in infant and neonatal mortality and associated factors in eight cohorts from three brazilian cities. *Scientific reports*, *10*(1), 1–8.

Chaibva, B. V., Olorunju, S., Nyadundu, S., & Beke, A. (2019). Adverse pregnancy outcomes,'stillbirths and early neonatal deaths' in mutare district, zimbabwe (2014): a descriptive study. *BMC pregnancy and childbirth*, *19*(1), 1–7.

Chawanpaiboon, S., Vogel, J. P., Moller, A.-B., Lumbiganon, P., Petzold, M., Hogan, D., Landoulsi, S., Jampathong, N., Kongwattanakul, K., Laopaiboon, M., et al. (2019). Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *The Lancet Global Health*, *7*(1), e37–e46.

Chawla, N. V. (2009). *Data mining for imbalanced datasets: An overview*. Springer.

Chowdhury, M. Z. I., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. *Family medicine and community health*, *8*(1).

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, *110*, 12–22.

Chuwa, F. S., Mwanamsangu, A. H., Brown, B. G., Msuya, S. E., Senkoro, E. E., Mnali, O. P., Mazuguni, F., & Mahande, M. J. (2017). Maternal and fetal risk factors for stillbirth in northern tanzania: A registry-based retrospective cohort study. *PloS one*, *12*(8), e0182250.

Columb, M., & Atkinson, M. (2015). Statistical analysis: sample size and power estimations. *Bja Education*, *16*(5), 159–161.

Cutland, C. L., Lackritz, E. M., Mallett-Moore, T., Bardají, A., Chandrasekaran, R., Lahariya, C., Nisar, M. I., Tapia, M. D., Pathirana, J., Kochhar, S., et al. (2017).

Low birth weight: Case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data. *Vaccine*, *35*(48Part A), 6492.

Darmstadt, G. L., Lee, A. C., Cousens, S., Sibley, L., Bhutta, Z. A., Donnay, F., Osrin, D., Bang, A., Kumar, V., Wall, S. N., et al. (2009). 60 million non-facility births: who can deliver in community settings to reduce intrapartum-related deaths? *International Journal of Gynecology & Obstetrics*, *107*, S89–S112.

De Campos, L. M., Cano, A., Castellano, J. G., & Moral, S. (2011). Bayesian networks classifiers for gene-expression data. In *2011 11th International Conference on Intelligent Systems Design and Applications*, (pp. 1200–1206). IEEE.

Del Fava, E., Kasim, A., Usman, M., Shkedy, Z., Hens, N., Aerts, M., Bollaerts, K., Tomba, G. S., Vickerman, P., Sutton, A. J., et al. (2011). Joint modeling of hcv and hiv infections among injecting drug users in italy using repeated cross-sectional prevalence data. *Statistical Communications in Infectious Diseases*, *3*(1).

Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, *35*(5-6), 352–359.

Dwivedi, A. K. (2018). Artificial neural network model for effective cancer classification using microarray gene expression data. *Neural Computing and Applications*, *29*(12), 1545–1554.

Entsieh, A. A., & Hallström, I. K. (2016). First-time parents' prenatal needs for early parenthood preparation-a systematic review and meta-synthesis of qualitative literature. *Midwifery*, *39*, 1–11.

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

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

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

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (2009). *Longitudinal data analysis*. Chapman Hall/CRC, Taylor Francis Group.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis*, vol. 998. John Wiley & Sons.

Fuchs, F., Monet, B., Ducruet, T., Chaillet, N., & Audibert, F. (2018). Effect of maternal age on the risk of preterm birth: A large cohort study. *PloS one*, *13*(1), e0191002.

Gebreslasie, K. (2016). Preterm birth and associated factors among mothers who gave birth in gondar town health institutions. *Advances in Nursing*, *2016*, 1–5.

Georgiou, H. M., Di Quinzio, M. K., Permezel, M., & Brennecke, S. P. (2015). Predicting preterm labour: current status and future prospects. *Disease markers*, *2015*.

Getiye, Y., & Fantahun, M. (2017). Factors associated with perinatal mortality among public health deliveries in addis ababa, ethiopia, an unmatched case control study. *BMC Pregnancy Childbirth*, *17*(1), 245.

Ghebremichael, M. (2015). Joint modeling of correlated binary outcomes: Hiv-1 and hsv-2 co-infection. *Journal of Applied Statistics*, *42*(10), 2180–2191.

Gibore, N. S., & Bali, T. A. (2020). Community perspectives: An exploration of potential barriers to men's involvement in maternity care in a central tanzanian community. *PloS one*, *15*(5), e0232939.

Gliklich, R. E., Leavy, M. B., & Dreyer, N. A. (2020). Analysis, interpretation, and reporting of registry data to evaluate outcomes. In *Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 4th edition*. Agency for Healthcare Research and Quality (US).

Glonek, G. F. (1996). A class of regression models for multivariate categorical responses. *Biometrika*, *83*(1), 15–28.

Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *The lancet*, *371*(9606), 75–84.

Goto, T., Camargo, C. A., Faridi, M. K., Freishtat, R. J., & Hasegawa, K. (2019). Machine learning–based prediction of clinical outcomes for children during emergency department triage. *JAMA network open*, *2*(1), e186937–e186937.

Grantz, K. L., Hinkle, S. N., Mendola, P., Sjaarda, L. A., Leishear, K., et al. (2015). Differences in risk factors for recurrent versus incident preterm delivery. *American Journal of Epidemiology*, *182*(2), 157–167.

Gregory, E., Drake, P., & Martin, J. (2018). Lack of change in perinatal mortality in the united states, 2014-2016. nchs data brief, no 316. *Hyattsville, MD: National Center for Health Statistics*.

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

Grønvik, T., & Fossgard Sandøy, I. (2018). Complications associated with adolescent childbearing in sub-saharan africa: A systematic literature review and meta-analysis. *PloS one*, *13*(9), e0204327.

Habib, N. A., Wilcox, A. J., Daltveit, A. K., Basso, O., Shao, J., Oneko, O., & Lie, R. T. (2011). Birthweight, preterm birth and perinatal mortality: a comparison of black babies in tanzania and the usa. *Acta obstetricia et gynecologica Scandinavica*, *90*(10), 1100–1106.

Habimana-Kabano, I., Broekhuis, A., & Hooimeijer, P. (2015). The effects of interpregnancy intervals and previous pregnancy outcome on fetal loss in rwanda (1996–2010). *International journal of reproductive medicine*, *2015*.

Hamilton, E. F., Dyachenko, A., Ciampi, A., Maurel, K., Warrick, P. A., & Garite, T. J. (2020). Estimating risk of severe neonatal morbidity in preterm births under 32 weeks of gestation. *The Journal of Maternal-Fetal & Neonatal Medicine*, *33*(1), 73–80.

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized Estimating Equations*. Chapman & Hall/CRC, Florida, USA.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, *22*(9), 1433–1446.

Hedeker, D. (2005). Generalized linear mixed models. *Encyclopedia of statistics in behavioral science*.

Hedeker, D. (2008). Multilevel models for ordinal and nominal variables. In *Handbook of multilevel analysis*, (pp. 237–274). Springer.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis: Covariance pattern models*, vol. 451. John Wiley & Sons.

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

Hill, A., Pallitto, C., McCleary-Sills, J., & Garcia-Moreno, C. (2016). A systematic review and meta-analysis of intimate partner violence during pregnancy and selected birth outcomes. *International Journal of Gynecology & Obstetrics*, *133*(3), 269 – 276.

Hinderaker, S. G., Olsen, B. E., Bergsjø, P. B., Gasheka, P., Lie, R. T., & Kvåle, G. (2003). Perinatal mortality in northern rural tanzania. *Journal of Health, Population and Nutrition*, *21*(1), 8–17.

Hoodbhoy, Z., Noman, M., Shafique, A., Nasim, A., Chowdhury, D., & Hasan, B. (2019). Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data. *International Journal of Applied and Basic Medical Research*, *9*(4), 226.

Houweling, T. A., van Klaveren, D., Das, S., Azad, K., Tripathy, P., Manandhar, D., Neuman, M., de Jonge, E., Been, J. V., Steyerberg, E., et al. (2019). A prediction model for neonatal mortality in low-and middle-income countries: an analysis of data from population surveillance sites in india, nepal and bangladesh. *International journal of epidemiology*, *48*(1), 186–198.

Hubbard, A. E., Ahern, J., Fleischer, N. L., [Van der Laan], M., Satariano, S. A., Jewell, N., Bruckner, T., & Satariano, W. A. (2010). To gee or not to gee: Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*, *21*(4), 467–474.

Hug, L., Alexander, M., You, D., Alkema, L., & for Child, U. I.-a. G. (2019). National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: a systematic analysis. *The Lancet Global Health*, *7*(6), e710–e720.

Hurley, E. G., & DeFranco, E. A. (2017). Influence of paternal age on perinatal outcomes. *American journal of obstetrics and gynecology*, *217*(5), 566–e1.

Iams, J. D., Romero, R., Culhane, J. F., & Goldenberg, R. L. (2008). Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. *The lancet*, *371*(9607), 164–175.

Ibrahim, J. G., & Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. *Test*, *18*(1), 1–43.

Isaksen, A. B., Østbye, T., Mmbaga, B. T., & Daltveit, A. K. (2015). Alcohol consumption among pregnant women in northern tanzania 2000–2010: a registry-based study. *BMC pregnancy and childbirth*, *15*(1), 1–10.

Ivanova, A., Molenberghs, G., & Verbeke, G. (2016). Mixed models approaches for joint modeling of different types of responses. *Journal of Biopharmaceutical Statistics*, *26*(4), 601–618.

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

Jena, B. H., Biks, G. A., Gelaye, K. A., & Gete, Y. K. (2020). Magnitude and trend of perinatal mortality and its relationship with inter-pregnancy interval in ethiopia: a systematic review and meta-analysis. *BMC pregnancy and childbirth*, *20*(1), 1–13.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1), 1–54.

Katz, J., Lee, A. C., Kozuki, N., Lawn, J. E., Cousens, S., Blencowe, H., Ezzati, M., Bhutta, Z. A., Marchant, T., Willey, B. A., et al. (2013). Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. *The Lancet*, *382*(9890), 417–425.

Kayiga, H., Lester, F., Amuge, P. M., Byamugisha, J., & Autry, A. M. (2018). Impact of mode of delivery on pregnancy outcomes in women with premature rupture of membranes after 28 weeks of gestation in a low-resource setting: A prospective cohort study. *PloS one*, *13*(1), e0190388.

Khan, F. A., Mullany, L. C., Wu, L. F., Ali, H., Shaikh, S., Alland, K., West Jr, K. P., & Labrique, A. B. (2020). Predictors of neonatal mortality: development and validation of prognostic models using prospective data from rural bangladesh. *BMJ global health*, *5*(1).

Khandwala, Y. S., Baker, V. L., Shaw, G. M., Stevenson, D. K., Lu, Y., & Eisenberg, M. L. (2018). Association of paternal age with perinatal outcomes between 2007 and 2016 in the united states: population based cohort study. *bmj*, *363*.

Kim, M. K., Lee, S. M., Bae, S.-H., Kim, H. J., Lim, N. G., Yoon, S.-J., Lee, J. Y., & Jo, M.-W. (2018). Socioeconomic status can affect pregnancy outcomes and complications, even with a universal healthcare system. *International Journal for Equity in Health*, *17*(1), 2.

Klein, N., Kneib, T., Marra, G., Radice, R., Rokicki, S., & McGovern, M. E. (2019). Mixed binary-continuous copula regression models with application to adverse birth outcomes. *Statistics in medicine*, *38*(3), 413–436.

Koffi, A. K., Kalter, H. D., Kamwe, M. A., & Black, R. E. (2020). Verbal/social autopsy analysis of causes and determinants of under-5 mortality in tanzania from 2010 to 2016. *Journal of global health*, *10*(2).

Kombo, A. Y., Mwambi, H., & Molenberghs, G. (2017). Multiple imputation for ordinal longitudinal data with monotone missing data patterns. *Journal of Applied Statistics*, *44*(2), 270–287.

Kozuki, N., Lee, A. C., Silveira, M. F., Sania, A., Vogel, J. P., Adair, L., Barros, F., Caulfield, L. E., Christian, P., Fawzi, W., et al. (2013). The associations of parity and maternal age with small-for-gestational-age, preterm, and neonatal and infant mortality: a meta-analysis. *BMC public health*, *13*(S3), S2.

Kuhle, S., Maguire, B., Zhang, H., Hamilton, D., Allen, A. C., Joseph, K., & Allen, V. M. (2018). Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. *BMC pregnancy and childbirth*, *18*(1), 1–9.

Lamont, K., Scott, N., & Bhattacharya, S. (2021). The data we have: Pregnancy and birth related data collection in australia, canada, europe and the usa–a web-based survey of practice. *International Journal of Population Data Science*, *6*(1), 1378.

Lamping, F., Jack, T., Rübsamen, N., Sasse, M., Beerbaum, P., Mikolajczyk, R. T., Boehne, M., & Karch, A. (2018). Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious sirs in critically ill children-a data-driven approach using machine-learning algorithms. *BMC pediatrics*, *18*(1), 1–11.

Laughon, S. K., Albert, P. S., Leishear, K., & Mendola, P. (2014). The nichd consecutive pregnancies study: recurrent preterm delivery by subtype. *American Journal of Obstetrics and Gynecology*, *210*(2), e131–138.

Lawn, J. E., Blencowe, H., Oza, S., You, D., Lee, A. C., Waiswa, P., Lalli, M., Bhutta, Z., Barros, A. J., Christian, P., et al. (2014). Every newborn: progress, priorities, and potential beyond survival. *The Lancet*, *384*(9938), 189–205.

Lawn, J. E., Blencowe, H., Pattinson, R., Cousens, S., Kumar, R., Ibiebele, I., Gardosi, J., Day, L. T., Stanton, C., Committee, L. S. S. S., et al. (2011). Stillbirths: Where? when? why? how to make the data count? *The Lancet*, *377*(9775), 1448–1463.

Lee, A. C., Blencowe, H., & Lawn, J. E. (2019a). Small babies, big numbers: global estimates of preterm birth. *The Lancet Global Health*, *7*(1), e2–e3.

Lee, S. M., Lee, M. H., Chang, Y. S., et al. (2019b). The clinical risk index for babies ii for prediction of time-dependent mortality and short-term morbidities in very low birth weight infants. *Neonatology*, *116*(3), 244–251.

Lee, Y., Nelder, J. A., & Pawitan, Y. (2006). *Generalized linear models with random effects: unified analysis via H-likelihood*. Chapman and Hall/CRC.

Liang, K., Zeger, S. L., & Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *54*(1), 3–40.

Lincetto, O., Mothebesoane-Anoh, S., Gomez, P., & Munjanja, S. (2006). Antenatal care. In *Opportunities for Africa's newborns: Practical data, policy and programmatic*

support for newborn care in Africa, (pp. 51–62). Partnership for Maternal, Newborn & Child Health.

Lisonkova, S., Sabr, Y., Butler, B., & Joseph, K. (2012). International comparisons of preterm birth: higher rates of late preterm birth are associated with lower rates of stillbirth and neonatal death. *BJOG: An International Journal of Obstetrics & Gynaecology*, *119*(13), 1630–1639.

Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., et al. (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.

Liu, Y., & De, A. (2015a). Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *International journal of statistics in medical research*, *4*(3), 287.

Liu, Y., & De, A. (2015b). Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *International journal of statistics in medical research*, *4*(3), 287.

Lu, L., Qu, Y., Tang, J., Chen, D., & Mu, D. (2015a). Risk factors associated with late preterm births in the underdeveloped region of china: A cohort study and systematic review. *Taiwan Journal of Obstetrics and Gynecology*, *54*(6), 647–653.

Lu, L., Qu, Y., Tang, J., Chen, D., & Mu, D. (2015b). Risk factors associated with late preterm births in the underdeveloped region of china: A cohort study and systematic review. *Taiwanese Journal of Obstetrics and Gynecology*, *54*(6), 647–653.

Mahande, A. M., & Mahande, M. J. (2016). Prevalence of parasitic infections and associations with pregnancy complications and outcomes in northern tanzania: a registry-based cross-sectional study. *BMC infectious diseases*, *16*(1), 1–9.

Mahande, M. J. (2015). *Recurrence of Perinatal Death, Preterm Birth and Preeclampsia in Northern Tanzania: A Registry Based Study*. Ph.D. thesis, The University of Bergen.

Mahande, M. J., Daltveit, A. K., Mmbaga, B. T., Obure, J., Masenga, G., et al. (2013a). Recurrence of perinatal death in northern tanzania: a registry based cohort study. *BMC pregnancy and childbirth*, *13*(1), 166.

Mahande, M. J., Daltveit, A. K., Mmbaga, B. T., Obure, J., Masenga, G., et al. (2013b). Recurrence of preterm birth and perinatal mortality in northern tanzania: registry-based cohort study. *Tropical Medicine and International Health*, *18*(8), 962–967.

Mahande, M. J., & Obure, J. (2016). Effect of interpregnancy interval on adverse pregnancy outcomes in northern tanzania: a registry-based retrospective cohort study. *BMC Pregnancy Childbirth*, *16*(1), 140.

Mahapula, F. A., Kumpuni, K., Mlay, J. P., & Mrema, T. F. (2016). Risk factors associated with pre-term birth in dar es salaam, tanzania: a case-control study. *Tanzania Journal of Health Research*, *18*(1).

Malacova, E., Regan, A., Nassar, R.-G. C., N., Leonard, H., et al. (2018). Risk of stillbirth, preterm delivery, and fetal growth restriction following exposure in a previous birth: systematic review and meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*, *125*(2), 183–192.

Manor, O., Matthews, S., & Power, C. (2000). Dichotomous or categorical response? analysing self-rated health and lifetime social class. *International journal of epidemiology*, *29*(1), 149–157.

March of Dimes, PMNCH, Save the Children, & WHO (2012). *Born Too Soon: The Global Action Report on Preterm Birth*. Eds CP Howson, MV Kinney, JE Lawn. World Health Organization, Geneva.

Marchant, T., Willey, B., Katz, J., Clarke, S., Kariuki, S., Ter Kuile, F., Lusingu, J., Ndyomugyenyi, R., Schmiegelow, C., Watson-Jones, D., et al. (2012). Neonatal mortality risk associated with preterm birth in east africa, adjusted by weight for gestational age: individual participant level meta-analysis. *PLoS medicine*, *9*(8), e1001292.

Maro, E. W., Mosha, N. R., Mahande, M. J., Obure, J., & Masenga, G. (2016). Ten years trend in maternal mortality at kilimanjaro christian medical center tanzania, 2003–2012: A descriptive retrospective tertiary hospital based study. *Asian Pacific Journal of Reproduction*, *5*(3), 214 – 220.

Mazaki-Tovi, S., Romero, R., Kusanovic, J. P., Erez, P.-B. L., O., et al. (2007). Recurrent preterm birth. *Seminars in Perinatology*, *31*(3), 142–158.

Mboya, I. B., Mahande, M. J., Mohammed, M., Obure, J., & Mwambi, H. G. (2020a). Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern tanzania. *BMJ open*, *10*(10), e040132.

Mboya, I. B., Mahande, M. J., Obure, J., & Mwambi, H. G. (2020b). Predictors of perinatal death in the presence of missing data: A birth registry-based study in northern tanzania. *Plos one*, *15*(4), e0231636.

Mboya, I. B., Mahande, M. J., Obure, J., & Mwambi, H. G. (2021). Predictors of singleton preterm birth using multinomial regression models accounting for missing data: A birth registry-based cohort study in northern tanzania. *Plos one*, *16*(4), e0249411.

McCrink, L., Marshall, A. H., Cairns, K., Fogarty, D., & Casula, A. (2011). Joint modelling of longitudinal and survival data: A comparison of joint and independent models. In *Proc 58th World Statistical Congress*.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, (pp. 109–142).

Meng, Y., & Groth, S. W. (2018). Fathers count: the impact of paternal risk factors on birth outcomes. *Maternal and child health journal*, *22*(3), 401–408.

Ministry of Health and Social Welfare [MoHSW] (2015). *The National Road Map Strategic Plan to Accelerate Reduction of Maternal, Newborn, and Child Deaths in Tanzania,2008–2015: One Plan II*. Ministry of Health and Social Welfare, Dar es Salaam, Tanzania.

Miranda, M. H. H., Fernandes, F. E. C. V., & de Lacerda Campos, M. E. A. (2017). Determinants associated to perinatal mortality and associated factors. *Journal of Nursing UFPE on line*, *11*(3), 1171–1178.

Mitao, M., Philemon, R., Obure, J., Mmbaga, B. T., Msuya, S., & Mahande, M. J. (2016). Risk factors and adverse perinatal outcome associated with low birth weight in northern tanzania: a registry-based retrospective cohort study. *Asian pacific journal of Reproduction*, *5*(1), 75–79.

Mmbaga, B. T., Lie, R. T., Kibiki, G. S., Olomi, R., Kvåle, G., & Daltveit, A. K. (2011). Transfer of newborns to neonatal care unit: a registry based study in northern tanzania. *BMC pregnancy and childbirth*, *11*(1), 1–12.

Mmbaga, B. T., Lie, R. T., Olomi, R., Mahande, M. J., Kvale, G., et al. (2012a). Cause-specific neonatal mortality in a neonatal care unit in northern tanzania: a registry based cohort study. *BMC Pediatrics*, *12*(116).

Mmbaga, B. T., Lie, R. T., Olomi, R., Mahande, M. J., Olola, O., & Daltveit, A. K. (2012b). Causes of perinatal death at a tertiary care hospital in northern tanzania 2000–2010: a registry based study. *BMC pregnancy and childbirth*, *12*(1), 139.

MoHCDGEC (2016). *The national road map strategic plan to improve Reproductive, Maternal, Newborn, Child & Adolescent health in Tanzania (2016-2020): One Plan II*. Ministry of Health, Community Development, Gender, Elderly and Children. Dar Es Salaam, Tanzania.

MoHCDGEC [Tanzania Mainland], MoH [Zanzibar], NBS, OCGS, & ICF (2016). *Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS) 2015-16*. Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and ICF. Dar es Salaam, Tanzania, and Rockville, Maryland, USA.

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

Molenberghs, G., & Lesaffre, E. (1999). Marginal modelling of multivariate categorical data. *Statistics in Medicine*, *18*, 2237–2255.

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. Springer Science & Business Media.

Mpembeni, R., Jonathan, R., & Mughamba, J. (2014). Perinatal mortality and associated factors among deliveries in three municipal hospitals of dar es salaam, tanzania. *Journal of Pediatrics & Neonatal Care*, *1*(4).

Muktan, D., Singh, R. R., Bhatta, N. K., & Shah, D. (2019). Neonatal mortality risk assessment using snappe-ii score in a neonatal intensive care unit. *BMC pediatrics*, *19*(1), 1–4.

Mulder, D. S., & Spicer, J. (2019). Registry-based medical research: data dredging or value building to quality of care? *The Annals of thoracic surgery*, *108*(1), 274–282.

Musa, A. B. (2014). Gene expression data classification with kernel independent component analysis. *Research Journal of Mathematical and Statistical Sciences*, *2320*, 6047.

Mutsaerts, M., Groen, H., Buiter-Van der Meer, A., Sijtsma, A., Sauer, P., Land, J., Mol, B., Corpeleijn, E., & Hoek, A. (2014). Effects of paternal and maternal lifestyle factors on pregnancy complications and perinatal outcome. a population-based birth-cohort study: the gecko drenthe cohort. *Human Reproduction*, *29*(4), 824–834.

Mvunta, M. H., Mboya, I. B., Msuya, S. E., John, B., Obure, J., & Mahande, M. J. (2019). Incidence and recurrence risk of low birth weight in northern tanzania: A registry based study. *PloS one*, *14*(4), e0215768.

Nankabirwa, V., Tumwine, J. K., Tylleskar, T., Nankunda, J., Sommerfelt, H., et al. (2011). Perinatal mortality in eastern uganda: a community based prospective cohort study. *Plos One*, *6*(5), e19674.

Nasejje, J. B., & Mwambi, H. (2017). Application of random survival forests in understanding the determinants of under-five child mortality in uganda in the presence of covariates that satisfy the proportional and non-proportional hazards assumption. *BMC research notes*, *10*(1), 1–18.

NBS, & OCGS (2018). *National Population Projections*. National Bureau of Statistics, Ministry of Finance and Planning, Dar es Salaam and Office of the Chief Government Statistician Ministry of Finance and Planning, Zanzibar.

Nijkamp, J., Sebire, N., Bouman, K., Korteweg, F., Erwich, J., & Gordijn, S. (2017). Perinatal death investigations: What is current practice? In *Seminars in Fetal and Neonatal Medicine*, vol. 22, (pp. 167–175). Elsevier.

Nikoloulopoulos, A. K., & Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, *27*(30), 6393–6406.

NISR [Rwanda], MOH [Rwanda], & International, I. (2015). *Rwanda Demographic and Health Survey 2014-15*. National Institute of Statistics of Rwanda (NISR) [Rwanda], Ministry of Health (MOH) [Rwanda], and ICF International. Rockville, Maryland, USA: NISR, MOH and ICF International.

Okube, O. T., & Sambu, L. M. (2017). Determinants of preterm birth at the postnatal ward of kenyatta national hospital, nairobi, kenya. *Open Journal of Obstetrics and Gynecology*, *7*(09), 973.

Ouyang, F., Zhang, J., Betran, A. P., Yang, Z., Souza, J. P., et al. (2013). Recurrence of adverse perinatal outcomes in developing countries. *Bulletin of World Health Organization*, *91*(5), 357–367.

Oyira, E. J., Ndiok, A., & A., B. P. (2017). Prevalence and factors associated with perinatal mortality between january 2005 to december 2009 in university of calabar teaching hospital, calabar. *International Journal of Science and Research Methodology.*, *6*(4), 1–13.

Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical epidemiology*, *9*, 157.

Peneza, A. K., & Maluka, S. O. (2018). 'unless you come with your partner you will be sent back home': strategies used to promote male involvement in antenatal care in southern tanzania. *Global health action*, *11*(1), 1449724.

Phillips, C., Velji, Z., Hanly, C., & Metcalfe, A. (2017). Risk of recurrent spontaneous preterm birth: a systematic review and meta-analysis. *BMJ Open*, *7*(7), e015402.

Pitt, C., Tawiah, T., Soremekun, S., ten Asbroek, A. H., Manu, A., Tawiah-Agyemang, C., Hill, Z., Owusu-Agyei, S., Kirkwood, B. R., & Hanson, K. (2016). Cost and cost-effectiveness of newborn home visits: findings from the newhints cluster-randomised controlled trial in rural ghana. *The Lancet Global Health*, *4*(1), e45–e56.

Purisch, S. E., & Gyamfi-Bannerman, C. (2017). Epidemiology of preterm birth. In *Seminars in perinatology*. Elsevier.

Quinn, J., Munoz, F. M., Gonik, B., Frau, L., Cutland, C., et al. (2016). Preterm birth: Case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. *Vaccine*, *34*(49), 6047–6056.

R Core Team (2020). R: A language and environment for statistical computing. *Vienna, Austria, R Foundation for Statistical Computing*.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, *2*(1), 1–21.

Raita, Y., Goto, T., Faridi, M. K., Brown, D. F., Camargo, C. A., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical care*, *23*(1), 1–13.

Reddy, T. (2017). *Prediction of time to threshold from a repeatedly measured biomarker*. Ph.D. thesis, Hasselt University.

Ren, H., & Du, M. (2017). Role of maternal periodontitis in preterm birth. *Frontiers in immunology*, *8*, 139.

Requejo, J., Merialdi, M., Althabe, F., Keller, M., Katz, J., & Menon, R. (2013). Born too soon: Care during pregnancy and childbirth to reduce preterm deliveries and improve health outcomes of the preterm baby. *Reproductive health*, *10*(S1), S4.

Royston, P., White, I. R., et al. (2011). Multiple imputation by chained equations (mice): implementation in stata. *J Stat Softw*, *45*(4), 1–20.

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

Rugaimukam, J. J., Mahande, M. J., Msuya, S. E., & Philemon, R. N. (2017). Risk factors for preterm birth among women who delivered preterm babies at bugando medical centre, tanzania. *SOJ Gynecol Obstet Womens Health*, *3*(2), 1–7.

Ruiz, M., Goldblatt, P., Morrison, J., Kukla, L., Švancara, J., Riitta-Järvelin, M., Taanila, A., Saurel-Cubizolles, M.-J., Lioret, S., Bakoula, C., et al. (2015). Mother's education and the risk of preterm and small for gestational age birth: a drivers meta-analysis of 12 european cohorts. *J Epidemiol Community Health*, *69*(9), 826–833.

Salihu, H. M., Duan, J., Nabukera, S. K., Mbah, A. K., & Alio, A. P. (2011). Younger maternal age (at initiation of childbearing) and recurrent perinatal mortality. *European Journal of Obstetrics and Gynecology*, *154*(1), 31–36.

Schaaf, J. M., Hof, M. H., Mol, B. W. J., Abu-Hanna, A., & Ravelli, A. C. (2012). Recurrence risk of preterm birth in subsequent singleton pregnancy after preterm twin delivery. *American journal of obstetrics and gynecology*, *207*(4), 279–e1.

StataCorp, L. (2017). Stata statistical software: Release 15 (2017). *College Station, TX: StataCorp LP*.

StataCorp, L. (2019). Stata multilevel mixed-effects reference manual release 16. *College Station, TX: StataCorp LLC*.

Stephens, D., & Diesing, M. (2014). A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. *PloS one*, *9*(4), e93950.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, *338*.

Su, D., Samson, K., Garg, A., Hanson, C., Berry, A. L. A., Lin, G., & Qu, M. (2018). Birth history as a predictor of adverse birth outcomes: Evidence from state vital statistics data. *Preventive Medicine Reports*, *11*, 63–68.

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of statistics*, *24*, 303–329.

Tarimo, C. S., Mahande, M. J., & Obure, J. (2020). Prevalence and risk factors for caesarean delivery following labor induction at a tertiary hospital in north tanzania: a retrospective cohort study (2000–2015). *BMC pregnancy and childbirth*, *20*(1), 1–8.

Temu, T. B., Masenga, G., Obure, J., Mosha, D., & Mahande, M. J. (2016). Maternal and obstetric risk factors associated with preterm delivery at a referral hospital in northern-eastern tanzania. *Asian Pacific Journal of Reproduction*, *5*(5), 365–370.

Teoh, P. J., Ridout, A., Seed, P., Tribe, R. M., & Shennan, A. H. (2018). Gender and preterm birth: Is male fetal gender a clinically important risk factor for preterm birth in high-risk women? *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *225*, 155–159.

Tough, S. C., Faber, A. J., Svenson, L. W., & Johnston, D. W. (2003). Is paternal age associated with an increased risk of low birthweight, preterm delivery, and multiple birth? *Canadian journal of public health*, *94*(2), 88–92.

UNDP (2018). Sustainable development goals. `http://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-3-good-health-and-well-being.html`. "[Online; accessed September-2018]".

UNICEF, WHO, World Bank Group, & United Nations (2019). Levels & Trends in Child Mortality: Report 2019. Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation.

UNICEF, WHO, World Bank Group, & United Nations (2020). *Levels and Trends in Child Mortality: Report 2020. Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. United Nations Children's Fund.

Unterscheider, J., O'Donoghue, K., Daly, S., Geary, M. P., Kennelly, M. M., et al. (2014). Fetal growth restriction and the risk of perinatal mortality–case studies from the multicentre porto study. *BMC pregnancy and childbirth*, *14*(1), 63.

Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006a). Fully conditional specification in multivariate imputation. *Journal of statistical computation and simulation*, *76*(12), 1049–1064.

Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006b). Fully conditional specification in multivariate imputation. *Journal of statistical computation and simulation*, *76*(12), 1049–1064.

van den Broek, N. R., Jean-Baptiste, R., & Neilson, J. P. (2014). Factors associated with preterm, early preterm and late preterm birth in malawi. *Plos One*, *9*(3), e90128.

van Zijl, M. D., Koullali, B., Mol, B. W., Pajkrt, E., & Oudijk, M. A. (2016). Prevention of preterm delivery: current challenges and future prospects. *International Journal of Womens Health*, *8*, 633–645.

Vellido, A., Ribas, V., Morales, C., Sanmartín, A. R., & Rodríguez, J. C. R. (2018). Machine learning in critical care: state-of-the-art and a sepsis case study. *Biomedical engineering online*, *17*(1), 1–18.

Verbeke, G., & Molenberghs, G. (2009). *Linear mixed models for longitudinal data*. Springer.

Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, *26*(6), 565–574.

Vogel, J. P., Chawanpaiboon, S., Moller, A.-B., Watananirun, K., Bonet, M., & Lumbiganon, P. (2018). The global epidemiology of preterm birth. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *52*, 3–12.

Vogel, J. P., Souza, J. P., Mori, R., Morisaki, N., & Lumbiganon, P. (2014). Maternal complications and perinatal mortality: findings of the world health organization multicountry survey on maternal and newborn health. *BJOG: An International Journal of Obstetrics & Gynaecology*, *121*(Suppl. 1), 76–88.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, *17*(2), 228.

Wado, Y. D., Sully, E. A., & Mumah, J. N. (2019). Pregnancy and early motherhood among adolescents in five east african countries: a multi-level analysis of risk and protective factors. *BMC pregnancy and childbirth*, *19*(1), 59.

Wagura, P., Wasunna, A., Laving, A., Wamalwa, D., et al. (2018). Prevalence and factors associated with preterm birth at kenyatta national hospital. *BMC pregnancy and childbirth*, *18*(1), 107.

Wang, M. (2014). Generalized estimating equations in longitudinal data analysis: a review and recent developments. *Advances in Statistics*, *2014*.

Watson-Jones, D., Weiss, H. A., Changalucha, J. M., Todd, J., Gumodoka, B., Bulmer, J., Balira, R., Ross, D., Mugeye, K., Hayes, R., et al. (2007). Adverse birth outcomes in united republic of tanzania: impact and prevention of maternal risk factors. *Bulletin of the World Health Organization*, *85*, 9–18.

White, I. R., Royston, P., & Wood, A. M. (2010). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, *40*(1), 7–20.

Wojcieszek, A. M., Reinebrant, H. E., Leisher, S. H., Allanson, E., Coory, M., Erwich, J. J., Frøen, J. F., Gardosi, J., Gordijn, S., Gulmezoglu, M., et al. (2016). Characteristics of a global classification system for perinatal deaths: a delphi consensus study. *BMC pregnancy and childbirth*, *16*(1), 1–11.

Worjoloh, A., Manongi, R., Oneko, O., Hoyo, C., Daltveit, A. K., Westreich, D., et al. (2012). Trends in cesarean section rates at a large east african referral hospital from 2005-2010. *Open Journal of Obstetrics and Gynecology*, *2*(03), 255.

World Health Organization (2014a). *Every newborn: an action plan to end preventable deaths*. World Health Organization, Geneva.

World Health Organization (2014b). Wha global nutrition targets 2025: Low birth weight policy brief. Tech. rep., World Health Organization.

World Health Organization (2015). *WHO recommendations on interventions to improve preterm birth outcomes*. World Health Organization, Geneva.

World Health Organization (2016a). *The WHO application of ICD-10 to deaths during the perinatal period: ICD-PM*. World Health Organization, Geneva.

World Health Organization (2016b). *WHO recommendations on antenatal care for a positive pregnancy experience*. World Health Organization, Geneva.

World Health Organization (2019). Maternal and perinatal health. `http://www.who.int/maternal_child_adolescent/topics/maternal/maternal_perinatal/en/`. "[Online; accessed September-2019]".

World Health Organization (2020). Preterm birth. `http://www.who.int/news-room/fact-sheets/detail/preterm-birth`. "[Online; accessed July-2020]".

Yang, J., Baer, R. J., Berghella, V., Chambers, C., & Chung, P. (2016). Recurrence of preterm birth and early term birth. *Obstetrics and Gynecology*, *128*(2), 364–372.

You, D., Hug, L., Ejdemyr, S., Idele, P., Hogan, D., Mathers, C., Gerland, P., New, J. R., Alkema, L., et al. (2015). Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the un inter-agency group for child mortality estimation. *The Lancet*, *386*(10010), 2275–2286.

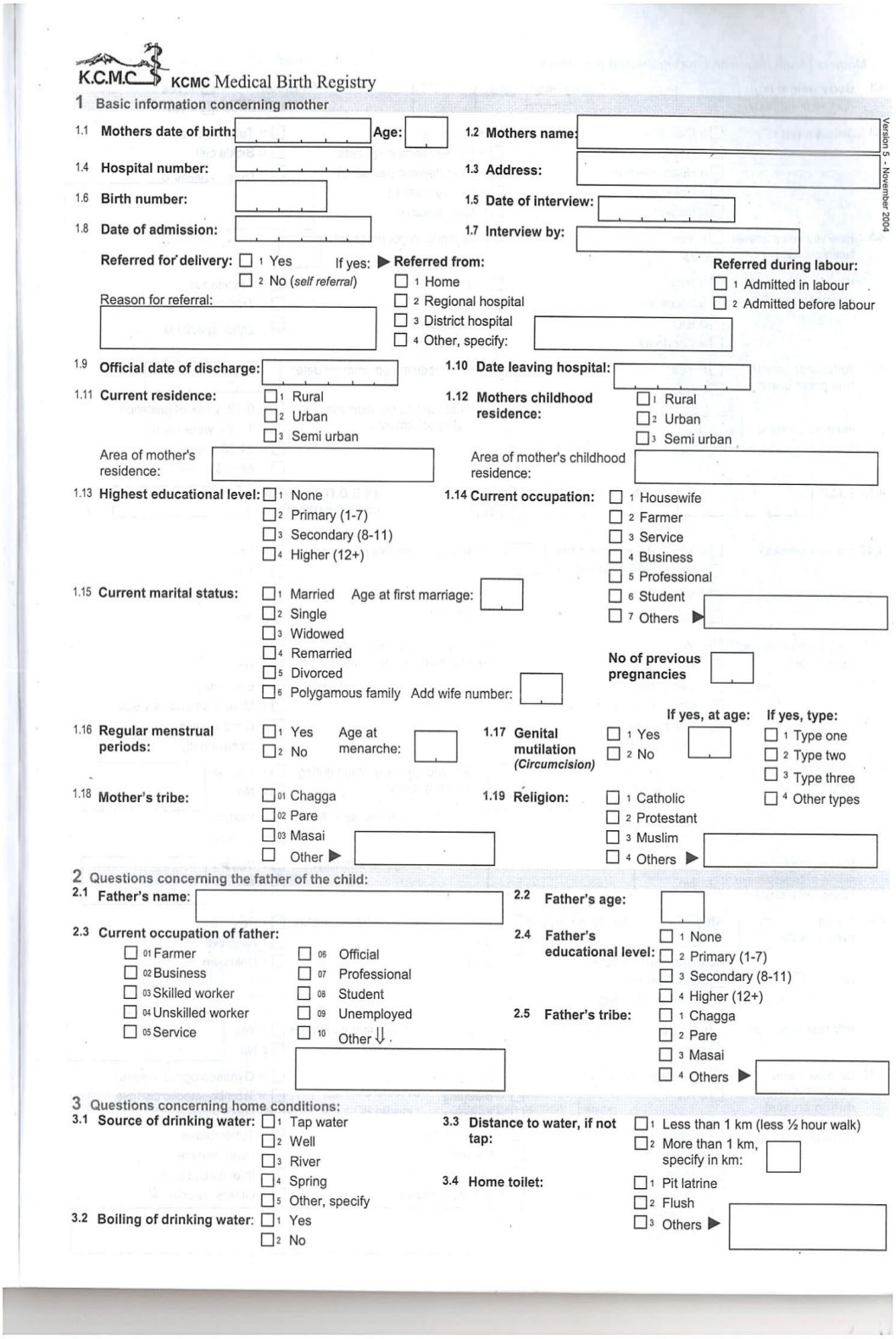
Zack, R. M., Golan, J., Aboud, S., Msamanga, G., Spiegelman, D., & Fawzi, W. (2014). Risk factors for preterm birth among hiv-infected tanzanian women: a prospective study. *Obstetrics and gynecology international*, *2014*.

Zhang, G., Srivastava, A., Bacelis, J., Juodakis, J., Jacobsson, B., & Muglia, L. J. (2018a). Genetic studies of gestational duration and preterm birth. *Best practice & research Clinical obstetrics & gynaecology*, *52*, 33–47.

Zhang, Z., Rousson, V., Lee, W.-C., Ferdynus, C., Chen, M., Qian, X., Guo, Y., et al. (2018b). Decision curve analysis: a technical note. *Annals of translational medicine*, *6*(15), 308.

Ziegler, A. (2011). *Generalized estimating equations*, vol. 204. Springer Science & Business Media.

Appendix A: KCMC medical birth registry questionnaire



K.C.M.C. KCMC Medical Birth Registry

Version 5 – November 2004

1 Basic information concerning mother

1.1 Mothers date of birth: ____ Age: ____ 1.2 Mothers name: ____

1.4 Hospital number: ____ 1.3 Address: ____

1.6 Birth number: ____ 1.5 Date of interview: ____

1.8 Date of admission: ____ 1.7 Interview by: ____

Referred for delivery: ☐ 1 Yes ☐ 2 No (*self referral*) If yes: ▶ Referred from: ☐ 1 Home ☐ 2 Regional hospital ☐ 3 District hospital ☐ 4 Other, specify: ____

Referred during labour: ☐ 1 Admitted in labour ☐ 2 Admitted before labour

Reason for referral: ____

1.9 Official date of discharge: ____ 1.10 Date leaving hospital: ____

1.11 Current residence: ☐ 1 Rural ☐ 2 Urban ☐ 3 Semi urban

1.12 Mothers childhood residence: ☐ 1 Rural ☐ 2 Urban ☐ 3 Semi urban

Area of mother's residence: ____ Area of mother's childhood residence: ____

1.13 Highest educational level: ☐ 1 None ☐ 2 Primary (1-7) ☐ 3 Secondary (8-11) ☐ 4 Higher (12+)

1.14 Current occupation: ☐ 1 Housewife ☐ 2 Farmer ☐ 3 Service ☐ 4 Business ☐ 5 Professional ☐ 6 Student ☐ 7 Others ▶ ____

1.15 Current marital status: ☐ 1 Married ☐ 2 Single ☐ 3 Widowed ☐ 4 Remarried ☐ 5 Divorced ☐ 6 Polygamous family Age at first marriage: ____ Add wife number: ____

No of previous pregnancies ____

1.16 Regular menstrual periods: ☐ 1 Yes ☐ 2 No Age at menarche: ____

1.17 Genital mutilation *(Circumcision)* ☐ 1 Yes ☐ 2 No If yes, at age: ____ If yes, type: ☐ 1 Type one ☐ 2 Type two ☐ 3 Type three ☐ 4 Other types

1.18 Mother's tribe: ☐ 01 Chagga ☐ 02 Pare ☐ 03 Masai ☐ Other ▶ ____

1.19 Religion: ☐ 1 Catholic ☐ 2 Protestant ☐ 3 Muslim ☐ 4 Others ▶ ____

2 Questions concerning the father of the child:

2.1 Father's name: ____ 2.2 Father's age: ____

2.3 Current occupation of father: ☐ 01 Farmer ☐ 02 Business ☐ 03 Skilled worker ☐ 04 Unskilled worker ☐ 05 Service ☐ 06 Official ☐ 07 Professional ☐ 08 Student ☐ 09 Unemployed ☐ 10 Other ⇓ ____

2.4 Father's educational level: ☐ 1 None ☐ 2 Primary (1-7) ☐ 3 Secondary (8-11) ☐ 4 Higher (12+)

2.5 Father's tribe: ☐ 1 Chagga ☐ 2 Pare ☐ 3 Masai ☐ 4 Others ▶ ____

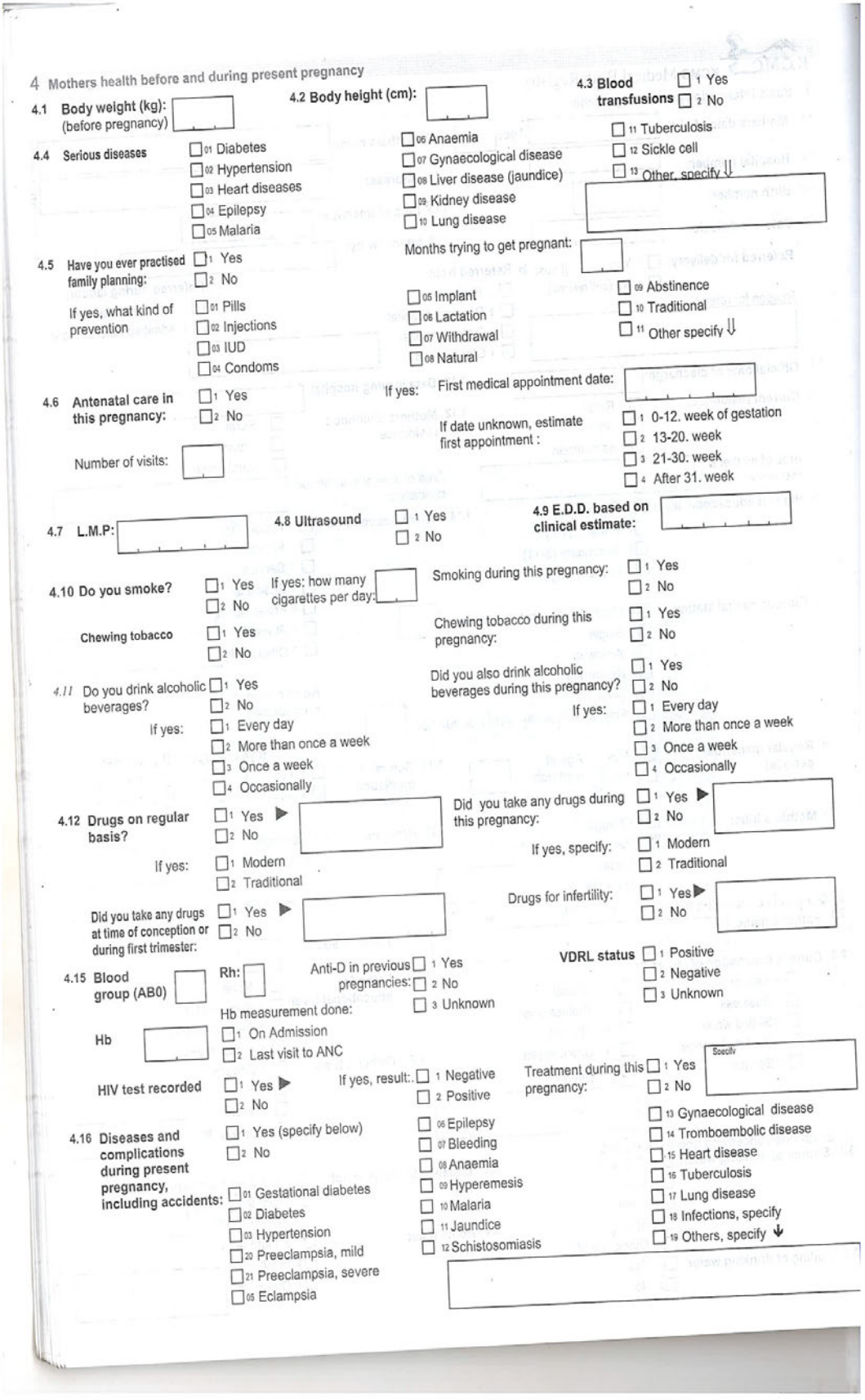
3 Questions concerning home conditions:

3.1 Source of drinking water: ☐ 1 Tap water ☐ 2 Well ☐ 3 River ☐ 4 Spring ☐ 5 Other, specify

3.2 Boiling of drinking water: ☐ 1 Yes ☐ 2 No

3.3 Distance to water, if not tap: ☐ 1 Less than 1 km (less ½ hour walk) ☐ 2 More than 1 km, specify in km: ____

3.4 Home toilet: ☐ 1 Pit latrine ☐ 2 Flush ☐ 3 Others ▶ ____



4 Mothers health before and during present pregnancy

4.1 Body weight (kg): (before pregnancy)

4.2 Body height (cm):

4.3 Blood transfusions ☐ 1 Yes ☐ 2 No

4.4 Serious diseases
☐ 01 Diabetes
☐ 02 Hypertension
☐ 03 Heart diseases
☐ 04 Epilepsy
☐ 05 Malaria
☐ 06 Anaemia
☐ 07 Gynaecological disease
☐ 08 Liver disease (jaundice)
☐ 09 Kidney disease
☐ 10 Lung disease
☐ 11 Tuberculosis
☐ 12 Sickle cell
☐ 13 Other, specify ⇓

4.5 Have you ever practised family planning: ☐ 1 Yes ☐ 2 No

If yes, what kind of prevention
☐ 01 Pills
☐ 02 Injections
☐ 03 IUD
☐ 04 Condoms
☐ 05 Implant
☐ 06 Lactation
☐ 07 Withdrawal
☐ 08 Natural
☐ 09 Abstinence
☐ 10 Traditional
☐ 11 Other specify ⇓

Months trying to get pregnant:

4.6 Antenatal care in this pregnancy: ☐ 1 Yes ☐ 2 No

If yes: First medical appointment date:

If date unknown, estimate first appointment :
☐ 1 0-12. week of gestation
☐ 2 13-20. week
☐ 3 21-30. week
☐ 4 After 31. week

Number of visits:

4.7 L.M.P:

4.8 Ultrasound ☐ 1 Yes ☐ 2 No

4.9 E.D.D. based on clinical estimate:

4.10 Do you smoke? ☐ 1 Yes ☐ 2 No
If yes: how many cigarettes per day:

Smoking during this pregnancy: ☐ 1 Yes ☐ 2 No

Chewing tobacco ☐ 1 Yes ☐ 2 No

Chewing tobacco during this pregnancy: ☐ 1 Yes ☐ 2 No

4.11 Do you drink alcoholic beverages? ☐ 1 Yes ☐ 2 No
If yes:
☐ 1 Every day
☐ 2 More than once a week
☐ 3 Once a week
☐ 4 Occasionally

Did you also drink alcoholic beverages during this pregnancy? ☐ 1 Yes ☐ 2 No
If yes:
☐ 1 Every day
☐ 2 More than once a week
☐ 3 Once a week
☐ 4 Occasionally

4.12 Drugs on regular basis? ☐ 1 Yes ▶ ☐ 2 No
If yes:
☐ 1 Modern
☐ 2 Traditional

Did you take any drugs during this pregnancy: ☐ 1 Yes ▶ ☐ 2 No
If yes, specify:
☐ 1 Modern
☐ 2 Traditional

Did you take any drugs at time of conception or during first trimester: ☐ 1 Yes ▶ ☐ 2 No

Drugs for infertility: ☐ 1 Yes ▶ ☐ 2 No

4.15 Blood group (AB0)

Rh:

Anti-D in previous pregnancies: ☐ 1 Yes ☐ 2 No ☐ 3 Unknown

VDRL status ☐ 1 Positive ☐ 2 Negative ☐ 3 Unknown

Hb measurement done:
☐ 1 On Admission
☐ 2 Last visit to ANC

Hb

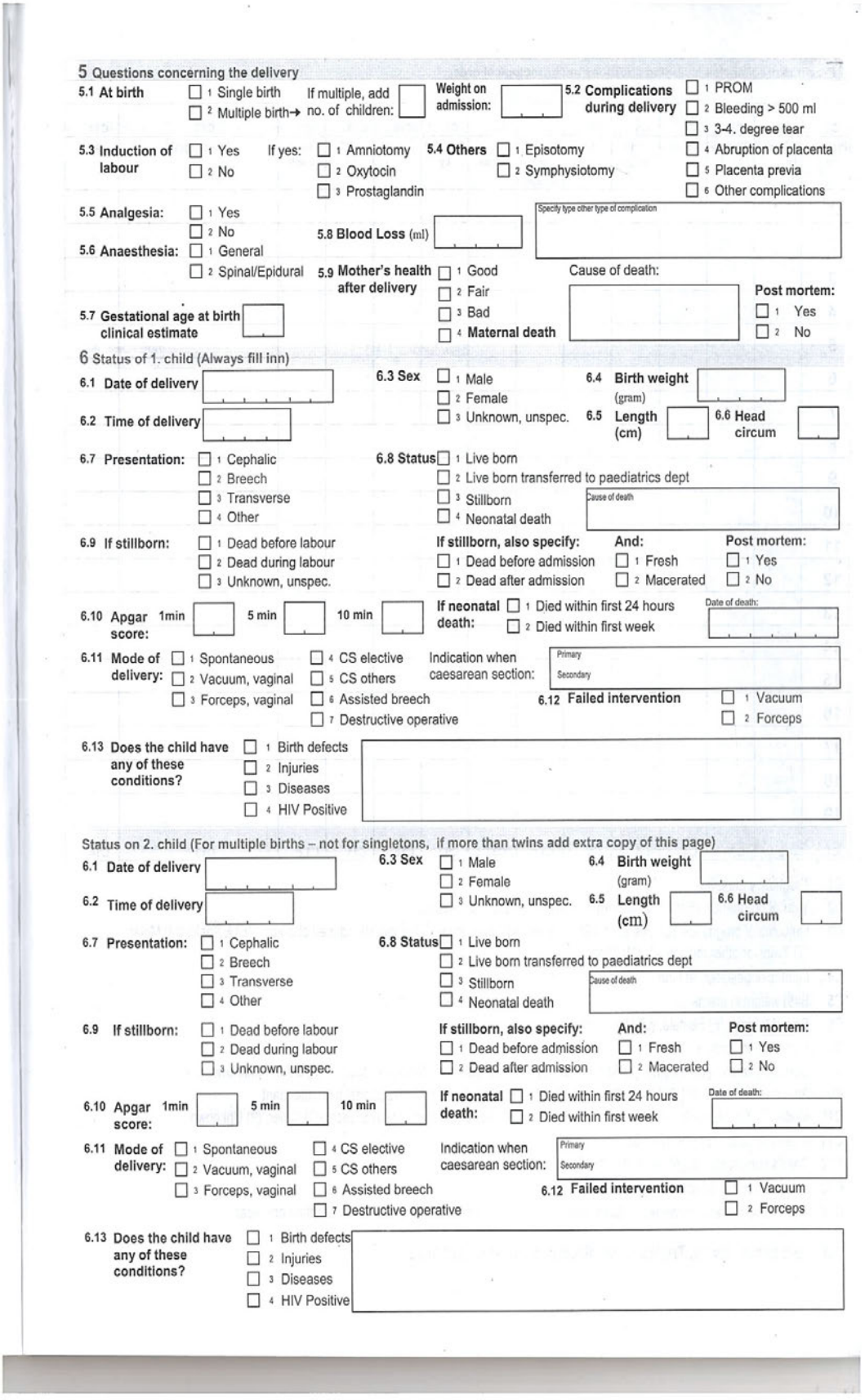
HIV test recorded ☐ 1 Yes ▶ ☐ 2 No

If yes, result: ☐ 1 Negative ☐ 2 Positive

Treatment during this pregnancy: ☐ 1 Yes ☐ 2 No
Specify

4.16 Diseases and complications during present pregnancy, including accidents: ☐ 1 Yes (specify below) ☐ 2 No

☐ 01 Gestational diabetes
☐ 02 Diabetes
☐ 03 Hypertension
☐ 20 Preeclampsia, mild
☐ 21 Preeclampsia, severe
☐ 05 Eclampsia
☐ 06 Epilepsy
☐ 07 Bleeding
☐ 08 Anaemia
☐ 09 Hyperemesis
☐ 10 Malaria
☐ 11 Jaundice
☐ 12 Schistosomiasis
☐ 13 Gynaecological disease
☐ 14 Tromboembolic disease
☐ 15 Heart disease
☐ 16 Tuberculosis
☐ 17 Lung disease
☐ 18 Infections, specify
☐ 19 Others, specify ⇓



5 Questions concerning the delivery

5.1 At birth ☐ 1 Single birth ☐ 2 Multiple birth→ If multiple, add no. of children: ☐ Weight on admission: ☐

5.2 Complications during delivery ☐ 1 PROM ☐ 2 Bleeding > 500 ml ☐ 3 3-4. degree tear ☐ 4 Abruption of placenta ☐ 5 Placenta previa ☐ 6 Other complications

5.3 Induction of labour ☐ 1 Yes ☐ 2 No If yes: ☐ 1 Amniotomy ☐ 2 Oxytocin ☐ 3 Prostaglandin

5.4 Others ☐ 1 Episotomy ☐ 2 Symphysiotomy

5.5 Analgesia: ☐ 1 Yes ☐ 2 No

5.6 Anaesthesia: ☐ 1 General ☐ 2 Spinal/Epidural

5.8 Blood Loss (ml) ☐

Specify type other type of complication

5.9 Mother's health after delivery ☐ 1 Good ☐ 2 Fair ☐ 3 Bad ☐ 4 Maternal death

Cause of death:

Post mortem: ☐ 1 Yes ☐ 2 No

5.7 Gestational age at birth clinical estimate ☐

6 Status of 1. child (Always fill inn)

6.1 Date of delivery ☐

6.2 Time of delivery ☐

6.3 Sex ☐ 1 Male ☐ 2 Female ☐ 3 Unknown, unspec.

6.4 Birth weight (gram) ☐

6.5 Length (cm) ☐ 6.6 Head circum ☐

6.7 Presentation: ☐ 1 Cephalic ☐ 2 Breech ☐ 3 Transverse ☐ 4 Other

6.8 Status ☐ 1 Live born ☐ 2 Live born transferred to paediatrics dept ☐ 3 Stillborn ☐ 4 Neonatal death

Cause of death

6.9 If stillborn: ☐ 1 Dead before labour ☐ 2 Dead during labour ☐ 3 Unknown, unspec.

If stillborn, also specify: ☐ 1 Dead before admission ☐ 2 Dead after admission

And: ☐ 1 Fresh ☐ 2 Macerated

Post mortem: ☐ 1 Yes ☐ 2 No

6.10 Apgar score: 1min ☐ 5 min ☐ 10 min ☐

If neonatal death: ☐ 1 Died within first 24 hours ☐ 2 Died within first week

Date of death: ☐

6.11 Mode of delivery: ☐ 1 Spontaneous ☐ 2 Vacuum, vaginal ☐ 3 Forceps, vaginal ☐ 4 CS elective ☐ 5 CS others ☐ 6 Assisted breech ☐ 7 Destructive operative

Indication when caesarean section: Primary Secondary

6.12 Failed intervention ☐ 1 Vacuum ☐ 2 Forceps

6.13 Does the child have any of these conditions? ☐ 1 Birth defects ☐ 2 Injuries ☐ 3 Diseases ☐ 4 HIV Positive

Status on 2. child (For multiple births – not for singletons, if more than twins add extra copy of this page)

6.1 Date of delivery ☐

6.2 Time of delivery ☐

6.3 Sex ☐ 1 Male ☐ 2 Female ☐ 3 Unknown, unspec.

6.4 Birth weight (gram) ☐

6.5 Length (cm) ☐ 6.6 Head circum ☐

6.7 Presentation: ☐ 1 Cephalic ☐ 2 Breech ☐ 3 Transverse ☐ 4 Other

6.8 Status ☐ 1 Live born ☐ 2 Live born transferred to paediatrics dept ☐ 3 Stillborn ☐ 4 Neonatal death

Cause of death

6.9 If stillborn: ☐ 1 Dead before labour ☐ 2 Dead during labour ☐ 3 Unknown, unspec.

If stillborn, also specify: ☐ 1 Dead before admission ☐ 2 Dead after admission

And: ☐ 1 Fresh ☐ 2 Macerated

Post mortem: ☐ 1 Yes ☐ 2 No

6.10 Apgar score: 1min ☐ 5 min ☐ 10 min ☐

If neonatal death: ☐ 1 Died within first 24 hours ☐ 2 Died within first week

Date of death: ☐

6.11 Mode of delivery: ☐ 1 Spontaneous ☐ 2 Vacuum, vaginal ☐ 3 Forceps, vaginal ☐ 4 CS elective ☐ 5 CS others ☐ 6 Assisted breech ☐ 7 Destructive operative

Indication when caesarean section: Primary Secondary

6.12 Failed intervention ☐ 1 Vacuum ☐ 2 Forceps

6.13 Does the child have any of these conditions? ☐ 1 Birth defects ☐ 2 Injuries ☐ 3 Diseases ☐ 4 HIV Positive

Appendix B: Ethical clearance certificate

CRERC FORM 07

TUMAINI UNIVERSITY

KILIMANJARO CHRISTIAN MEDICAL COLLEGE
P. O. Box 2240, MOSHI, Tanzania

RESEARCH ETHICAL CLEARANCE CERTIFICATE

No.2424

Research Proposal No. 1189

Study Title: Predictors of preterm birth and perinatal death in northern Tanzania based on a zonal hospital birth registry data between 2000 and 2017

Study Area: KCMC birth registry

P. I Name: Innocent B. Mboya

Coinvestigators: Michael Mahande,Henry G. Mwambi, Joseph Obure

Institution (s): University of KwaZulu Natal

The Proposal was approved by CRERC on: 26th June,2019

Duration of Study: One year

From: 26th June,2019 to 26th June,2020

Name: BEATRICE Z. TEMBA [redacted]
Secretary – CRERC

Name : PROF.MRAMBA NYINDO [redacted]
Chairman – CRERC