2026-06

Causal Inference for Count Treatments on Ordinal Outcomes

Ashagrie, Sharew

BAHIR DAR UNIVERSITY

COLLEGE OF SCIENCE

DEPARTMENT OF STATISTICS

CAUSAL INFERENCE FOR COUNT TREATMENTS ON ORDINAL OUTCOMES

By

Ashagrie Sharew Iyassu

30 June, 2026

Bahir Dar University, Ethiopia

BAHIR DAR UNIVERSITY

College of Science

Department of Statistics

Causal Inference for Count Treatments on Ordinal Outcomes

By

Ashagrie Sharew Iyassu

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics.

Principal supervisor: Dr. Haile Mekonen Fenta (Associate Professor)

Co- supervisors: Prof. Temesgen T. Zewotir (PhD)

Dr. Zelalem Getahun Dessie (Associate Professor)

30 June, 2026

Bahir Dar University, Ethiopia

DECLARATION

I, Ashagrie Sharew Iyassu, hereby declare that this dissertation titled **"**Causal Inference for Count Treatments on Ordinal Outcomes**"** is my own original work and has not been submitted elsewhere for any other degree or qualification.

To the best of my knowledge, all sources, data, and contributions from other works have been properly acknowledged and cited. Any collaborative work has been clearly indicated.

This dissertation was conducted under the supervision of Doctor Haile Mekonen at Bahir Dar University, Bahir Dar, Ethiopia and Oulu University, Finland, Professor Temesgen T. Zewotir at KwaZulu Natal University, Durban, South Africa, and Doctor Zelalem Getahun at Bahir Dar University, Bahir Dar and KwaZulu Natal University, Durban, South Africa

Ashagrie Sharew Iyassu		6/30/2026
Name of the candidate	Signature	Date

Bahir Dar University

College of Science

Department of Statistics

APPROVAL OF DISSERTATION FOR DEFENSE

This dissertation, titled Causal Inference for Count Treatments on Ordinal Outcomes, prepared by Ashagrie Sharew under our supervision, has been reviewed and approved by the undersigned supervisors to be submitted for oral defense in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics.

Haile Mekonnen Fenta (PhD, Associate prof.)	Haile MF	06/07/2026
Principal supervisor name	Signature	Date
Temesgen T. Zewotir (Prof)	Temesgen Zewotir	06/07/2026
Co-supervisor name	Signature	Date
Zelalem Getahun Dessie (PhD, Associate prof.)	ZG Dessie	06/07/2026
Co-supervisor name	Signature	Date

DEDICATION

This dissertation is dedicated to my wife, Nitsuh Demilie, whose unwavering support, encouragement and endless love was a source of courage for me. It is also dedicated to my lovely children, Isayas Ashagrie and Kidus Ashagrie, for their patience and love throughout my journey.

I would also like to dedicate to my father and mother for their sleepless prayer and blessing. I am here because of their love and prayer.

ACKNOWLEDGMENTS

My journey and completion of this dissertation would not have been possible without the will, mercy and blessing of almighty God and His mother St.Virgin Mary. Completing this dissertation has been a challenging yet immensely rewarding journey, and I owe my deepest gratitude to the many individuals and institutions that have supported me along the way.

First and foremost, I would like to express my deepest gratitude to my supervisors Dr. Haile Mekonnen, Professor Temesgen Zewotir, and Dr. Zelalem Getahun for their invaluable guidance, patience, and encouragement throughout my dissertation work. Their expertise, insightful feedback, and unwavering support have been instrumental in shaping this work.

I am extremely grateful to Department of Statistics at Bahir Dar University for giving the opportunity to pursue my PhD class and to Debre Markos University for sponsoring me throughout my PhD work. I would like to extend my acknowledgement to Amhara Public health Institute for supporting me.

Most importantly, I am forever indebted to my wife Nitsuh Demilie and my children Isayas Ashagrie and Kidus Ashagrie for their unconditional love, sacrifices, and belief in me. Their encouragement has been my greatest source of strength.

I must also express my deepest appreciation for my parents, brothers and sisters. Their unconditional love, prayer and encouragement have been the foundation of my resilience and success.

My heartfelt thanks also go to Dr. Dessalegn Melesse, PI for Data Analytics System and Use project, whose financial and technical support was valuable.

Finally, I thank all those who, directly or indirectly, contributed to this work. This achievement would not have been possible without their support.

SCIENTIFIC PRODUCTIONS

Papers Published

- Iyassu, Ashagrie Sharew, Haile Mekonnen Fenta, Zelalem G. Dessie, and Temesgen T. Zewotir. Identification of confounders and estimating the causal effect of place of birth on age-specific childhood vaccination. *BMC Medical Informatics and Decision Making* 24, no. 1 (2024): 406. https://doi.org/10.1186/s12911-024-02827-2.
- Iyassu, Ashagrie Sharew, Haile Mekonnen Fenta, Zelalem G. Dessie, and Temesgen T. Zewotir. Causal Effect of Count Treatment on Ordinal Outcome Using Generalized Propensity Score: Application to Number of Antenatal Care and Age Specific Childhood Vaccination. *Journal of Epidemiology and Global Health* 15, no. 1 (2025): 23. DOI: 10.1007/s44197-025-00344-7.
- Iyassu, Ashagrie Sharew, Haile Mekonnen Fenta, Zelalem G. Dessie, and Temesgen T. Zewotir. Identifying confounders and estimating the causal effect of antenatal care on age-specific childhood vaccination." *Frontiers in Public Health* 13 (2025): https://doi.org/10.3389/fpubh.2025.1420567.

Papers under review

- Ashagrie Sharew Iyassu,, Haile Mekonnen Fenta, Zelalem G. Dessie, and Temesgen T. Zewotir. Direct and Indirect Effects of Number of Antenatal Care Contacts on Age-specific Childhood Vaccination when Mediated by Institutional delivery.
- Ashagrie Sharew Iyassu,, Haile Mekonnen Fenta, Zelalem G. Dessie, and Temesgen T. Causal Mediation Analysis under Causally Ordered Mediators for the Effect of Count Exposure on Ordinal Outcome: Comparison of Weighting and G-Computation Estimation.
- Ashagrie Sharew Iyassu, Haile Mekonnen Fenta, Zelalem G. Dessie, and Temesgen T. Zewotir. Estimating the Causal Effect of Postnatal Care on Age-Specific Childhood Vaccination: A Comparison of Weighting Methods to Control Confounding Bias.
- Estimating Generalized propensity score for Count treatments: Comparison of Treatment Models.

ABSTRACT

In an observational study, the propensity score is widely used to adjust for confounding bias when the treatment is binary and, to some extent, for categorical and continuous treatments. This study introduced the application of count treatments to causal inference on ordinal outcomes. We compared different methods, including maximum likelihood for the family of count models, generalized boosted model (GBM), and covariate balancing propensity score (CBPS), to estimate generalized propensity score (GPS). In the comparison, we used effective sample size in combination with confounders balancing power and efficiency of treatment effect estimation. In the treatment effect estimation, confounders' bias was controlled using inverse probability of treatment weighting (IPTW). Two outcome models, including marginal structural modeling and considering GPS as a covariate, were compared in terms of the efficiency of treatment effect estimation. To reduce the influence of extreme IPTW, we trimmed the GPS at the 1st and 99th percentiles. We also applied path-specific effects estimation under sequential mediators while estimating the count treatments effect on the ordinal outcomes. In this case, a composite weighting method was introduced to estimate path-specific effects under sequential mediators to weight each component of the structural equation modeling of mediators and outcome. The first mediator model was weighted by IPTW generated from the treatment model; the second mediator model was weighted by the composite IPTW generated from the treatment and the first mediator model; and the outcome model was weighted by the composite IPTW generated from the treatment and mediators. The proposed weighting method was compared with simulation-based G-computation in terms of treatment effect estimation. For real-life data, we used the number of antenatal care visits (ANC) as a count treatment and age-specific childhood vaccination as an ordinal outcome. We also used a simulation study to compare the proposed models in the study. The result shows that GBM-based weighting is better than other methods to maintain a high effective sample size and efficient count treatment effect estimation while having equivalent performance with other methods in terms of covariate balancing. Marginal structural modeling produced better treatment effect estimation than using GPS as a covariate. The proposed weighting method and G-computation based on simulation have comparable performance in terms of treatment effect estimation, despite the fact that the proposed weighting method is computationally efficient. The number of ANC visits has a positive effect on the likelihood of timely childhood vaccination. In

the meantime, the effect was mediated by institutional delivery and newborn postnatal care, though the direct effect of ANC visits is higher than the indirect effects.

Keywords: Count treatment, Confounders, Ordinal outcome, Cumulative link model, Generalized propensity score, Inverse probability of treatment weighting, Marginal structural equation modeling, Generalized propensity score as covariate, Mediation analysis, Path specific effect, Sequential mediators.

TABLE OF CONTENTS

Table of Contents

DECLARATION ii

APPROVAL OF DISSERTATION FOR DEFENSE iii

DEDICATION iv

ACKNOWLEDGMENTS v

SCIENTIFIC PRODUCTIONS vi

ABSTRACT vii

TABLE OF CONTENTS ix

LIST OF APPENDICES xv

LIST OF ACRONYMS xvi

1. CHAPTER ONE: INTRODUCTION 1

1.1 Background 1

1.2 Review of Inverse Probability of Treatment Weighting (IPTW) 6

1.3 Research Motivation and Objectives 9

1.4 Contributions of the Dissertation 10

1.5 Structure of the Dissertation 10

2. CHAPTER TWO: DATA EXPLORATION 12

2.1 Sources and Descriptions of Data 12

2.2 Description of Variables 13

2.3 Exploratory Analysis 14

2.3.1 Missing Data Exploration and Management 14

2.3.2 Exploratory Analysis of Main Variables 19

CHAPTER THREE: IDENTIFYING CONFOUNDERS AND ESTIMATING EFFECTS 24

3.1. Confounder Identification for ANC and Childhood Vaccination 24

3.1.1. Background ` 24

3.1.2. Measure of Outcome Variable and Conceptual Framework 25

3.1.3. Methods of Identifying Confounders 26
3.1.4. Treatment Model 28
3.1.5. Outcome Model 30
3.1.6. Result 32
3.1.7. Discussion and Conclusion 37
3.2. Estimating the Effect of Place of Birth on Age-Specific Childhood Vaccination 39
3.2.1. Background 39
3.2.2. Methods 39
3.2.3. Result 44
3.2.4. Discussion and Conclusion 51
3.3 Causal Effect of Newborn Postnatal care on Age-specific Childhood Vaccination 53
3.3.1 Background 53
3.3.2 Description of Variables 54
3.3.3. Potential Outcome Framework and Estimating Treatment Effect 55
3.3.4. Adjusting Confounders Using Inverse Probability Treatment Weight 56
3.3.5. Covariate Balance Assessment 63
3.3.6. Statistical Model to Estimate Treatment Effect 64
3.3.7. Simulation Study 65
3.3.8. Result 66
3.3.9. Discussion and Conclusion 75
CHAPTER FOUR: ESTIMATING COUNT TREATMENTS ON ORDINAL OUTCOMES: SIMULATION AND APPLICATION 79
4.1. Estimating Generalized Propensity Scores for Count Treatments 79
4.1.1. Background 79
4.1.2. Data 79
4.1.3. Assumptions of Generalized Propensity Score Estimation 80
4.1.4. Estimation of Generalized Propensity Score 81
4.1.5. Covariate Balance Assessment 87
4.1.6. Result 88
4.1.7. Discussion and Conclusion 98
4.2. Causal Effect of Count Treatments on Ordinal Outcomes 100

4.2.1 Background 100
4.2.2 Notation and Assumption 101
4.2.3 Sensitivity Analysis 102
4.2.4 Potential Outcome Framework and Treatment Effect 103
4.2.5 Statistical Methods to Estimate Treatment Effect 103
4.2.6 Marginal Structural Model 104
4.2.7 Simulation Study 105
4.2.8 Result 107
4.2.9 Result of the Real Data 112
4.2.10 *Discussion and Conclusion* 116
4. CHAPTER FIVE: CAUSAL MEDIATION ANALYSIS 119
5.1. Direct and Indirect Effect of Count Treatments on Ordinal Outcomes 119
5.1.1 Background 119
5.1.2 Data and Methods 120
5.1.3 Mediation Analysis 121
5.1.4 Result 125
5.1.5 Discussion and Conclusion 128
5.2. Weighting vs G-Computation in Sequential Mediation 130
5.2.1. Background 130
5.2.2. Notation and Identification Assumption 132
5.2.3. Decomposition of Total Effect into Direct and Indirect Effects 133
5.2.4 Estimation 136
5.2.5 Simulation Study 139
5.2.6 Discussion and Conclusion 146
CHAPTER SIX: CONCLUSION, LIMITATION AND FUTURE WORKS 149
6.1. Conclusion 149
6.3. Future Works 151
REFERENCES 152

LIST OF TABLES

Table 2.1. Description of treatment, mediators and covariates. 13
Table 2.2. Distribution of missing values in childhood vaccination status and ANC. 16
Table 2.3. Test of missing at random (MAR) for childhood vaccination status. 16
Table 2.4. Test of MAR for number of Antenatal care visits. 18
Table 2.5. Sensitivity analysis of multiple imputations 19
Table 3.1. Model comparisons for covariate-exposure and covariates-outcome relationship. 33
Table 3.2. Result for comparisons of significant testing approaches of confounder identification. 34
Table 3.3. Results of change in estimate. 35
Table 3.4. Results for Comparison of change in estimate and significance testing methods. 36
Table 3.5. Treatment effect in proportional and partial proportional odds models. 37
Table 3.6. Estimates of performance measures for confounder identification approaches. 47
Table 3.7. Estimates of performance measures for confounder identification approaches 48
Table 3.8. Selection of link function for the outcome model. 49
Table 3.9. Effect of place of birth on age-specific childhood vaccination. 50
Table 3.10. Result of performance metrics for plasmode simulation. 67
Table 3.11. Summary results of artificial simulation for ordinal outcome. 68
Table 3.12. Results of simulation study for continuous outcome under various sample sizes. 70
Table 3.13. Distribution of stabilized weights. 73
Table 3.14. Result of proportional odds/ parallel line assumption test. 74
Table 3.15. Result for the effect of newborn postnatal care on age-specific childhood vaccination. 74
Table 3.16. Sensitivity analysis for unobserved confounders in terms of odds ratio. 75
Table 4.1. Result of covariate balance assessment in simulation study. 90
Table 4.2. Result of covariate balance assessment before and after adjusting. 95
Table 4.3. Summary of weights when ANC contacts are treatment variable. 97
Table 4.4. Distribution of GPS and GPS based weight. 108
Table 5.1. Unadjusted direct, indirect, and total effects without treatment–mediator interaction. 125
Table 5.2. Unadjusted direct, indirect, and total effects with treatment–mediator interaction. 126
Table 5.3. Adjusted direct, indirect, and total effects with treatment–mediator interaction. 127
Table 5.4. Adjusted direct, indirect, and total effects with treatment–mediator interaction. 128
Table 5.5. Result of simulation study without interaction for ordinal outcome. 141
Table 5.6. Result of simulation study with treatment-mediators interaction for ordinal outcome. 141

Table 5.7. Simulation result for binary outcome without interaction. 142
Table 5.8. Path-specific effects of the number of ANC visits on the outcome. 144
Table 5.9. Result of mediation analysis with treatment-mediators interaction. 145

LIST OF FIGURES

Figure 1.1. Flow chart to implement IPTW in causal inference. 8
Figure 2.1. Missingness conceptual framework. 15
Figure 2.2. Graphical presentation for the distribution of important variables 20
Figure 2.3. Distribution of vaccination status along with number of ANC visits. 21
Figure 2.4. Distribution of place of delivery and postnatal care by vaccination status. 22
Figure 3.1. Conceptual framework. 26
Figure 3.2. DAG for covariates, exposure and outcome. 40
Figure 3.3. Distribution of the outcome from plasmode and observed data. 44
Figure 3.4. Density plot of treatment effect estimate from plasmode simulation 45
Figure 3.5. Treatment effect for plasmode data using regression method. 46
Figure 3.6. Treatment effect from plasmode data using IPTW. 46
Figure 3.7. DAG for covariates-exposure and covariates-outcome relationship. 55
Figure 3.8. Confounders balance assessment for ordinal outcome. 69
Figure 3.9. Confounders balance assessment for continuous outcome. 71
Figure 3.10. Confounder balance assessment for actual data. 72
Figure 3.11. Density plot of propensity score as a measure of overlap assumption. 73
Figure 4.1. Sample size vs AACC of the simulation study for independent covariates. 93
Figure 4.2. Sample size vs AACC of the simulation study for correlated covariates. 93
Figure 4.3. Sample size vs effective sample size of the simulation study for independent covariates. 94
Figure 4.4. Sample size vs effective sample size of simulation study for correlated covariates. 94
Figure 4.1. Causal directed acyclic graph (DAG) for single mediator. 122
Figure 5.2. DAG for mediation analysis under ordered mediators. 133

LIST OF APPENDICES

APPENDICES 168

Appendix : Publication 168

LIST OF ACRONYMS

AACC	Average Absolute Correlation Coefficient
AF	Attributable Fraction
AIC	Akaike Information Criterion
ANC	Antenatal Care Contacts
ATE	Average Treatment Effect
BCG	Bacillus-Calmette Guerin
BIC	Bayesian Information Criterion
CBPS	Covariate Balancing Propensity Score
CDE	Controlled Direct Effect
CIE	Change In Estimate
CLM	Cumulative Link Model
DAG	Directed Acyclic Graph
DAG	Directed Acyclic Graph
EMDHS	Ethiopian Mini Demographic And Health Survey
FMoH	Federal Ministry Of Health
GBM	Generalized Boosted Model
GLM	Generalized Linear Model
GPS	Generalized Propensity Scores
IPTW	Inverse Probability Of Treatment Weighting
MAR	Missing At Random
MCAR	Missing Completely At Random
MLE	Maximum Likelihood Estimates
MNAR	Missing Not At Random
MSM	Marginal Structural Modeling
NDE	Natural Direct Effect
NIE	Natural Indirect Effect
PNC	Postnatal Care
PS	Propensity Score
PSE	Path Specific Effect
RCT	Randomized Control Trial

SMD	Standardized Mean Difference
WHO	World Health Organization
ZIP	Zero Inflation Poisson

CHAPTER ONE: INTRODUCTION

1.1 Background

Causality is connected to an action, such as a treatment or intervention applied to a unit at a particular point in time. Accordingly, the causal statement assumes that when a unit is exposed to treatment or intervention at a particular point in time, the same unit should be exposed to an alternative treatment or exposure at the same point in time [1].

Randomized control trial (RCT), which is a gold standard research design, is a random allocation of contrasted treatments or exposures to experimental units [2]. When study participants in an RCT are assigned to treatment and control groups randomly, each participant in a group would have the same probability of being assigned to the experimental or control groups. Under RCT, the baseline covariates would have a similar distribution in treated and control groups since baseline covariates did not affect treatment assignment. As a result, randomization reduces bias arising from baseline covariates, and treatment effects can be estimated by comparing the outcomes of the treatment and control groups [2, 3]. However, due to ethical and practical issues, randomized control trials are used in limited ranges[4]. An alternative non-randomized research design, called an observational study, is used to estimate the causal effect of an exposure on the outcome [5, 6]. Observational studies may lack internal validity since associations between exposure and outcome may be biased due to other factors that act as confounders or selection bias [7].

In causal analyses, the relationship between the exposure and the outcome variables under study can be altered by confounders. In this case, treatment groups and control groups that receive and do not receive the exposure are different from one another in some other essential variable that is also associated with the outcome[8]. Thus, a confounder is a variable that alters the relationship between exposure and outcome and generates a spurious treatment-outcome association.

In an observational study, confounders have to be identified and their effect controlled while estimating the association between exposure and outcome. Controlling confounders helps to get unbiased estimates of the exposure-outcome relationship[9]. Including all pretreatment covariates in any confounder-controlling methods, such as regression, introduces bias. Adding more

covariates to the model causes over-fitting and unstable coefficients due to multicollinearity[10]. A model is best when it contains the smallest number of covariates that explain the greatest amount of variance[11]. As a result, identifying confounders that potentially distort the causal effect of treatment on the outcome is imperative. However, how to identify confounders and how to deal with them is one of the challenges in observational studies. There is no common consensus criterion to identify which covariates are confounders and which are not. [11]. A common approach is to control as many pre-exposure covariates as possible[5].

Studies modified this approach by controlling all covariates significantly associated (p-value less than 0.05) with the outcome of interest [12, 13]. Others also stated that they control confounders that give a predetermined magnitude of change (10% or 15%) while estimating the relationship between exposure and outcome [12, 14].
A Directed Acyclic Graph (DAG) is also a method of identifying confounders to be controlled in the association of exposure with the outcome. It aims to identify a minimally satisfactory set of well-measured confounders that satisfy the definition of confounders to control [5].

Despite controversies about which method is better, changes in estimate and significance testing methods are widely used for confounder identification, even if significance testing is acceptable for a P-value of 0.2 or less[11]. Due to scant literature in our study setting, we identified confounders for the effect of count treatment on ordinal outcome (such as the effect of number of antenatal care contacts (ANC) on age-specific childhood vaccination) and the effect of binary treatments on ordinal outcomes (such as the effect of institutional delivery on age-specific childhood vaccination) in the subsequent sections.

To balance baseline confounders' distribution between treatment and control groups and accurately estimate treatment effect in an observational study, analysts proposed statistical methods such as matching, stratification, and regression[4]. Matching and stratification are the grouping of treatment and control groups with identical confounders' distribution. However, these methods fail for high-dimensional confounders since it is difficult to get individuals identical with all confounders. For regression adjustment, it is difficult to assess covariate balance between the treatment and control groups [15]. Although multiple regressions, which are a direct and simple way to remove bias due to confounders and easily handle high-dimensional covariates, may have

limitations and not be reliable for a large number of confounders [15, 16]. For instance, when there is a considerable difference between the treatment and control groups, regression analysis cannot make good adjustments[17]. In addition, outcome model misspecification for confounders decreases the precision of treatment effect estimation.

A new method, which has gained attention in recent decades, proposed to adjust covariate imbalance or confounders in an observational study, is the propensity score. It was first introduced by Rosenbaum and Rubin [18] ,and they defined it as the probability of treatment assignment conditional on observed baseline covariates**.**

The goal of the propensity score in an observational study is to control for measured confounders by achieving balance in characteristics between exposed and unexposed groups. By accounting for any differences in measured baseline characteristics, the propensity score mimics what would have been attained through randomization in an RCT (i.e., pseudo-randomization). In contrast to true randomization, it should be considered that the propensity score can only account for measured confounders, not for any unmeasured confounders[19].

The propensity score is intended to decrease the dimension of baseline covariates, and thus far, its estimation involves modeling the treatment on the high-dimensional baseline covariates. The estimated propensity score is correct if it balances baseline covariates [20].

Most of the studies on propensity score have focused on binary treatments proposed by Rosenbaum, P.R. and D.B. Rubin [18] where it was estimated by binary logistic regression. However, in the real world, treatments are not only binary but also categorical with more than two levels and quantitative. Imbens [21] introduced generalized propensity scores(GPS) for multi-valued treatments. Following Imbens and others [22-25] have used GPS for categorical treatments (multinomial and ordinal) with more than two levels.

Hirano and Imbens [26] have extended the GPS for continuous treatments. They used a Gaussian distribution to estimate GPS. The other scholars, Y. Zhu et al.[27] estimated GPS from continuous treatments using a boosted algorithm. Similarly, Wu et al.[28] have used matching on GPS for continuous treatments.

To the best of our knowledge, studies on causal inference using GPS focused on multi-valued treatments (categorical or ordinal) and continuous treatments where GPS is estimated by multinomial logistic regression, ordinal logistic regression, linear regression with normal or kernel density, generalized boosted algorithm, and covariate balancing propensity score. However, there is a lack of literature on estimating the causal effects of count treatment on an ordinal outcome.

In some contexts, the effect of an exposure on an outcome is accounted for by a specific set of mediators along the causal pathway connecting the exposure to the outcome [29]. In such a scenario, part or all of the exposure's effect is mediated by the mediator variable that seeks to work on mediation analysis.

In causal mediation analysis, mediators are variables that lie between treatment and an outcome, helping to elucidate the mechanism through which the exposure impacts the outcome. They act as connecting steps in the causal process: the exposure affects the mediator, and the mediator subsequently affects the outcome. This enables us to break down the total causal effect of an exposure on an outcome into the effect through mediator(s) (also called indirect effect) and the effect not through mediator(s) (also called direct effect). When used by Baron and Kenny [30], mediation analysis focused on linear models where there was no exposure-mediator interaction. Nevertheless, Robins and Greenland [31] proposed a counterfactual framework for causal inference analysis, which uses a non-parametric approach to estimate the direct and indirect effects of the exposure. The counterfactual framework allows interaction between exposure and mediator(s) as well as non-linear models.

Since the introduction of the counterfactual framework for mediation analysis, abundant methodological methods have been proposed for causal mediation analysis that allows additive and multiplicative scales, and non-linear models for survival data [32-37]. Mediation analysis can involve a single mediator or multiple mediators, according to the study setting. The multiple mediators can be independent (parallel mediators) or dependent mediators (sequential or ordered mediators), where one affects the other. In a single mediator mediation analysis, there is no other mediator except that there can exist post-exposure confounders[38]. In parallel multiple mediators, there is no mediator affected by the exposure and that goes on to affect another mediator(s). For such mediators, the analysis is carried out simultaneously [39]. In sequential mediators, which are

the focus of this study, there is a mediator(s) influenced by the exposure and in turn affects the subsequent mediator(s). For such mediators, we can apply path analysis to decompose the total exposure's effect in each of the paths[40].

Mediation analysis with causally ordered multiple mediators in previous literature has focused on either continuous or binary outcomes. We found that VanderWeele and Lim [41] and Stanghellini E, Kateri M [42] used an ordinal outcome under a single mediator. Zhou et al. [43] also used an ordinal outcome under parallel multiple mediators. This indicated that mediation analysis under causally ordered mediators in an ordinal outcome has not been done yet, which motivated this study. In addition, there has not been a study that used count exposure (for example, number of ANC contacts) in mediation analysis to our level of knowledge.

This study used the number of ANC contacts as the count treatment (treatment and exposure are alternatively used in this study), age-specific childhood vaccination/immunization (categorized as no vaccination, partial vaccination, and full vaccination) as an ordinal outcome variable, and socio-economic, household, child, mother, and community-level characteristics as covariates for the application of causal inference for count treatments on ordinal outcomes. We also used institutional delivery and newborn postnatal care as binary treatments to examine their effect on the outcome variable, and later used them as mediators in mediation analysis.

Immunization is considered one of the best cost-effective and efficient public health policies to prevent childhood morbidity and mortality, particularly in developing countries[44]. Childhood immunization played a significant role in the reduction of morbidity and mortality, saving more than three million children per year [45, 46].

The World Health Organization (WHO) urges countries to adhere to immunization schedules to achieve optimal coverage of childhood immunization and reduce the risk of vaccine-preventable illness[47]. For successful immunization coverage in preventing childhood diseases, mothers' acceptance and adherence to the vaccination schedule are important to attain high coverage and eliminate vaccine-preventable diseases [48].

In the struggle to fight diseases that can be prevented by vaccination and to reduce neonatal and under-five morbidity and mortality, the World Health Organization pledged an expanded program

on immunization in 1974, intending to immunize every child by 1990. Ethiopia launched the expanded program on immunization in 1980 with vaccines of Bacillus-Calmette-Guerin (BCG), diphtheria, pertussis, tetanus, polio, and measles. Later, in 1986, the program was revised with the target of 75% coverage and the target age group of infants less than one year old, though the progress was low in increasing immunization coverage. After the introduction of a new approach in 2003, also called reaching every district and sustainable outreach for immunization, improvement was recognized [49].

Studies conducted in Ethiopia at national and sub-national levels [50-54] reported that ANC follow-up increased the likelihood of childhood vaccination. On the contrary, Lakew, Bekele, and Biadgilign [55] reported that ANC did not affect childhood vaccination status. However, studies conducted so far have looked at the effect of ANC on childhood vaccination status regardless of when it was given, though childhood vaccination is recommended to be given right after birth within a specified schedule and time range[56]. On the other hand, studies focused on the dichotomy of childhood vaccination (vaccinated or not vaccinated). This dichotomization causes a loss of information when we consider more than one vaccine and vaccination schedule. In addition, there was a lack of literature on the causal effect of ANC on age-specific childhood vaccination alone and in the presence of a mediator(s) after controlling the effect of confounders using inverse probability of treatment weighting.

Hence, this study identified confounders for the treatment-outcome relationship. It also extended GPS estimation for count treatments and estimated their effect on ordinal outcomes using GPS-based IPTW. The count treatment effect was also estimated in the presence of single and sequential multiple binary mediators.

1.2 Review of Inverse Probability of Treatment Weighting (IPTW)

Propensity score (PS) can be used in various ways to control confounding bias, including matching with PS, stratification with PS, inverse probability weighting, and using PS as a covariate in the outcome model[18, 57, 58].

The PS matching creates matched sets of treated and untreated subjects that have similar propensity scores (PS) values. The matching ratio can vary, but one-to-one pairing is the most prevalent. The response variable is then compared between the treated and untreated subjects

within the matched set. Matching can occur with or without replacement: in matching without replacement, each treated subject is paired with only one untreated subject, while in matching with replacement, each treated subject can be paired with multiple untreated subjects[59].
However, in PS matching, unmatched cases will be discarded from the analysis, which causes a loss of information and imprecise treatment effect estimation [59, 60].

PS stratification divides a dataset into multiple strata based on PS scores. The treatment effect is analyzed within each stratum, and an overall treatment effect is calculated as a weighted average across all strata. This approach compares outcomes between treated and untreated subjects who share similar PS values and are likely to have comparable distributions of measured covariates. Stratification employs weights proportional to the number of subjects in each stratum, facilitating the estimation of the average treatment effect. The weight is calculated as $1/k$ for k equal strata. The average treatment effect (ATE) is then estimated by weighting based on the total number of treated and control subjects in each stratum[59]. The pitfall of PS stratification is that it will have biased estimation when the number of strata is small [59].

IPTW uses propensity scores to create weights. IPTW aims to form a pseudo-population where covariates and treatment assignments are independent, mimicking the conditions of randomization. The term "pseudo-population" highlights that the weighted groups differ from the actual observed population but represent what could have been sampled without confounding. This approach may seem unusual since it creates weighted groups that didn't exist in the original sample[60]. For count treatments like ANC, the IPTW is created by taking the inverse of the GPS for each treatment level to generate unstabilized weights. However, unstabilized weights can generate extreme weights due to small values of GPS when the probability of a particular treatment assignment is close to zero or due to large values of GPS when the probability of a particular treatment assignment is close to one. To counter this drawback of unstabilized weight, we can use stabilized weight [60, 61]. The stabilized weight is created by multiplying the marginal GPS (the probability of treatment assignment without confounders) by the unstabilized weights.

Sometimes the stabilized weight can be over dispersed. In such a scenario, trimming or truncating at 1st and 99th percentiles is essential for consistent treatment effect estimation [62-64]. However,

it is recommended to perform sensitivity analysis based on trimming propensity score based IPTW implementation[65].

One advantage of using IPTW instead of matching is that it requires a smaller sample size to be effective. Unlike matching methods, which often result in the loss of individuals from either the treatment or control group, weighting methods incorporate the entire sample into the final analysis to some extent [66].

The procedures we followed in this study to estimate treatment effect using IPTW are presented in Figure 1.1. In the first step, we identified the observed confounders that can modify the true effect of the treatment on the outcome using confounder identification techniques. In the second step, we estimated (G)PS using the treatment models to reduce the dimensions of the observed confounders to a single scalar. In the third step, we estimated IPTW by taking the inverse of (G)PS stabilized by the marginal propensity score (probability of the treatment without confounders). In the fourth step, we checked confounders balancing between treatment and control groups using covariate balance assessment tools. In the fifth step, the treatment effect was estimated with and without involving mediator(s).

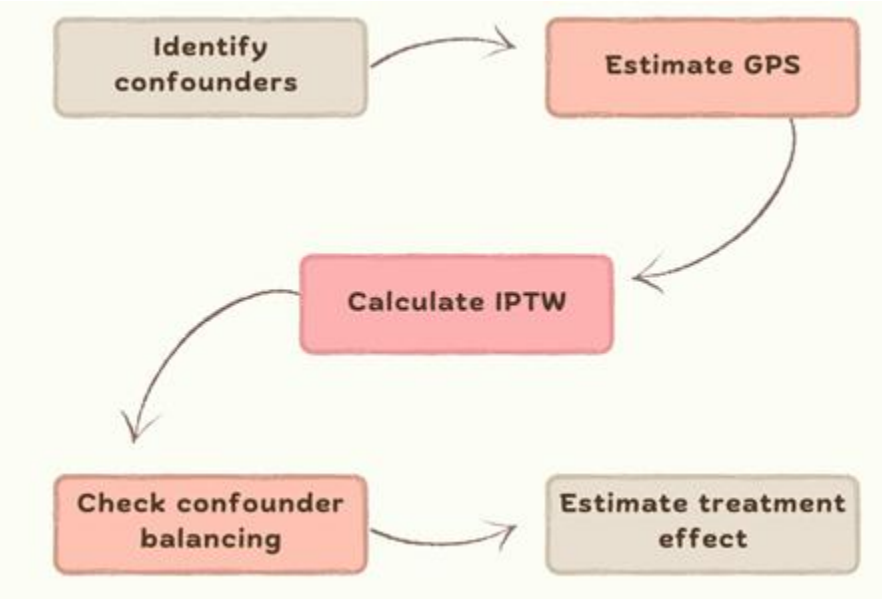


Figure 1.1. Flow chart to implement IPTW in causal inference.

1.3 Research Motivation and Objectives

In an experimental study, the true effect of the treatment can be influenced by observed and unobserved baseline confounders that are observed or unobserved before the treatment. These confounders have effects on the outcome to generate confounding bias and could be unequally distributed between treatment and control groups, which generate treatment selection bias. These confounding and selection biases can be controlled through design, like randomization, or using statistical techniques. Randomization controls both the observed and unobserved biases through random allocation of experimental units to treatment values. In this case, the researcher controls the experimental process. However, randomization cannot always be feasible for various reasons, like survey data. An observational study (quasi-experimental study) is an alternative statistical technique to deal with the treatment effect in the absence of randomization. In such a scenario, observed selection and confounding biases are controlled to resemble randomization. The presence of unobserved confounders' bias can be assessed using sensitivity analysis.

Studies so far have focused on binary treatments in an observational study followed by continuous treatments. Little has been done on multivalued treatments [22-25]. However, to our knowledge there is no literature on the use of count treatment in causal inference analysis (for example, using ANC as a treatment). In connection with this, the use of count and zero augmented models and generalized boosted models have not been practiced yet. In addition, the use of an ordinal outcome such as age-specific childhood vaccination (with ordered categories of no, partial, and complete vaccination) in treatment effect estimation is limited in the literature. Especially, estimating the causal effect of count treatment on ordinal outcome has not been done yet. Furthermore, there has been a scarcity of literature to compare composite weighting and G-computation in path-specific effect estimation under sequential mediators (for example, the effect of ANC on age-specific childhood vaccination mediated by institutional delivery and newborn postnatal care) for complete effect decomposition. Using actual household-level survey data, there is no literature on the causal effect of ANC on age-specific childhood vaccination after controlling for confounding bias with IPTW. Comparing methods for the actual data by using plasmode simulation when parametric or synthetic simulation is not practical is not abundant in the literature. Hence, the major objectives of the study were:

- To compare confounders identifying techniques and estimate treatment effect.

- To estimate the generalized propensity score for count treatments and estimate their effect on ordinal outcomes.
- To estimate the direct and indirect effect of count treatments on ordinal outcomes in the presence of single and sequential mediators.

1.4 Contributions of the Dissertation

The contributions of this dissertation work are twofold. The first contribution is, the research work will inform maternal and a child health program leaders how maternal health care seeking behavior enhances childhood vaccinations at the right vaccination schedule to prevent vaccine-preventable morbidity and mortality. The second contribution is for the scientific community who will benefit from this research work by incorporating or extending approaches and models used in this dissertation. Mainly we extended the causal inference framework to count treatments. Most research in the literature relied on binary treatments and latter extended to multivalued and continuous treatments. However, the application of count treatments is disregarded in the literature. We also used the effective sample size in combination with confounders balancing power to compare the treatment models and calculate GPS. Moreover, we estimated the path-specific effects of count treatments on ordinal outcomes under sequential mediators by applying composite weights for the mediators and outcome. Finally, we showed that childhood vaccination coverage should account for vaccination timelines instead of vaccination coverage at any age.

1.5 Structure of the Dissertation

The dissertation is structured in six chapters described as follows. Chapter one explains the introduction of the dissertation that contains background, a review of IPTW, research motivation, and objectives. Chapter two is about data and exploratory analysis, encompassing the source of data, missing data exploration and management, and visualization of key variables.

Chapter three is about techniques for identifying confounders and estimating treatment effects. The chapter contains three sub-topics where the first topic describes identifying confounders for the effect of number of ANC contacts on age-specific childhood vaccination; the second topic explores identifying confounders and estimating binary treatment effect on ordinal outcome application to the effect of institutional delivery on age-specific childhood vaccination; and the third section outlines comparison of weighting methods to control confounding bias with application to the causal effect of newborn postnatal care on age-specific childhood vaccination

and simulation data. The last two topics were included to examine the effect of institutional delivery and newborn postnatal care on age-specific childhood vaccination, so that they will act as mediators in chapter five. Chapter four deals with estimating the causal effect of count treatments on ordinal outcomes: comparison of treatment and outcome models. In this chapter, two sub-sections are included. The first sub-section explores estimating the generalized propensity score for count treatments and comparing candidate models to estimate the generalized propensity score; and the second sub-section outlines the causal effect of count treatments on ordinal outcomes and comparing candidate treatment and outcome models.

Chapter five is about causal mediation analysis, which consists of two sub-sections. The first section explores the direct and indirect effects of the number of ANC on age-specific childhood vaccination, and the second sub-section explains mediation analysis under sequential binary mediators and count treatment. Chapter six presents the conclusion, limitation of the dissertation and future works.

CHAPTER TWO: DATA EXPLORATION

2.1 Sources and Descriptions of Data

The dissertation basically used real and simulated datasets. The real dataset used in the study were obtained from the Ethiopian mini demographic and health survey (EMDHS) 2019 (https://dhsprogram.com/data/available-datasets.cfm). Specifically, data from birth records that contain all records of women aged 15-49 years with the most recent birth prior to five years of the survey were used. The data was collected from March 21, 2019, to June 28, 2019 based on a nationally representative sample to provide estimates at the national and regional levels and for urban and rural areas using two stage stratified sampling. In the survey, 5753 women with live births were interviewed to collect detail information on women's background characteristics, fertility determinants, marriage, awareness and use of family planning methods, child feeding practices, nutritional status of the children, childhood mortality, and height and weight of the children aged 0-59 months [72]. Because there was no vaccination history of dead children in the dataset, only children who were alive at the time of survey were considered for the purpose of this study.

We also used simulated datasets to compare the performance of statistical methods. The data were generated from predetermined distributions called parametric simulation. In parametric simulation, covariates, exposure and outcome data were generated from known or predefined stochastic distributions such that the generated data resembles the realistic or representative. Parameters of interest were derived from the real dataset, literature, or set by the user [67, 68]. Parametric simulation is mainly used for model development and to compare the performance of different models.

We also used Plasmode simulation to compare the performance of our methods based on data resampled from real-life datasets. Plasmode refers to a dataset created from real-life datasets where certain truths are known. Plasmode datasets are created by using an existing dataset as a template to simulate new datasets with known truths[69]. It involves resampling from a real-world dataset and incorporating a relationship between treatment and outcome defined by the investigator. By combing actual covariate data with a user-defined outcome data generation process, plasmodes preserve some of the complex relationships present in the original dataset while allowing for a specified causal effect between exposure and outcome[70]. In the outcome generation, the

parameters are determined either from the true relationship or from literature. While plasmode data contains hybrid (actual resampled covariates data and synthetic outcome data), plasmode simulation is considered as semi-parametric simulation[71].

2.2 Description of Variables

The outcome variable of the study was age-specific childhood vaccination status which had three ordered categories (No vaccination, partially and fully vaccinated). The classification was done according to the immunization schedule guideline set in Ethiopia Federal Ministry of Health (FMoH) National Expanded Program on Immunization 2016-2020 multi-year plan[49]. When a child took all vaccines at the recommended age level, it was considered as fully vaccinated; when a child missed at least one vaccine at each recommended age level, it was considered as partially vaccinated; and when a child did not take any vaccine at the recommended age level, it was considered as not vaccinated. For this study, vaccination and immunization are alternatively used. The study also involved the treatment, the mediator and socio-economic, household, mothers, and child characteristics as covariates. The nature and role of the variables are presented in Table 2.1.

Table 2.1. Description of treatment, mediators and covariates.

Level of variables	Variable	Nature of variable	Values/categories
Treatment and mediator variables	Number of ANC visits	Count	0, 1,2,.,.,,11
	Place of birth	Binary	At home/At health facility
	Newborn Postnatal care	Binary	Yes/No
Child level covariates	Birth order number	Count	1,2,3,.,.,.,11
	Child is twine	Binary	Yes/No
	Sex of child		Male/Female
Mother's level covariates	Age category of mother at birth	Categorical	15-19, 20-24,.,.,, 49-49
	Mother's education level	Categorical	No education, primary, secondary, higher
	Mother's religion	Categorical	Orthodox, protestant, Muslim, others
	Age of mother at first birth	Continuous	10,11,12,.,.,44
	Mother's marital status	Categorical	Married, widowed, divorced

Household level covariates	Ownership of radio	Binary	Yes/No
	Ownership of television	Binary	Yes/No
	Household size	Count	1,2,3,.,.,.,15
	Sex of household head	Binary	Male/Female
	Age of household head	Continuous	15,16,17,.,.,.,.,80
	Household wealth status	Categorical	Poorest, poorer, middle ,richer, richest
	Total children ever born	Count	1,2,3,.,.,.15
Community level covariates	Region	Categorical	Tigray, Afar, Amhara, Oromia, Benishangule, SNNP, Somali, Dire Dawa, Addis Ababa, Gambela
	Residence	Binary	Urban/Rural

2.3 Exploratory Analysis

2.3.1 Missing Data Exploration and Management

Missingness can be missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) [73]. MAR is a more relaxed assumption where missingness is dependent on observed covariates but independent of unobserved covariates[74]. A valid conclusion can be made when missing data are MAR by using appropriate imputation techniques such as multiple imputation[75]. The analysis is more challenging under MNAR since some relevant information are unobserved, and it needs extra untestable assumptions to proceed with the analysis. This makes the MAR assumption is a typical starting point of missing data analysis[76]. In this study, we focused on MAR.

Let Z be the treatment variable, $\boldsymbol{X}$ be P dimensional covariates, Y be the outcome variable. With the observed $\boldsymbol{X}$, let R^z and R^y are missingness indicators for Z and Y with value of 1 when Z and Y are missing and 0 when they are observed. Then missing graph also called "m-graphs"[77] for MAR assumption is presented in Figure 2.1. The hole circle indicates missing and the shaded circles indicate observed.

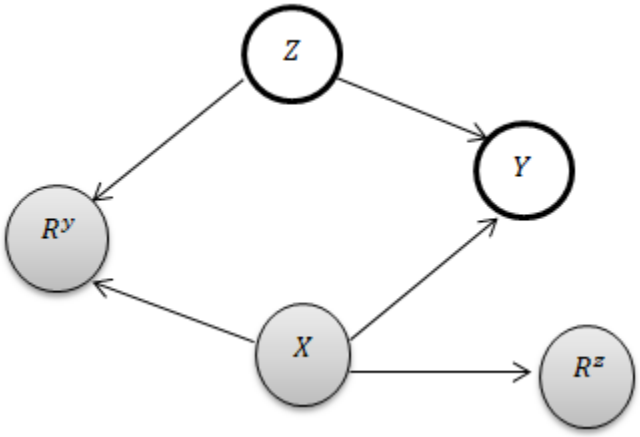


Figure 2.1. Missingness conceptual framework.

The treatment (number of ANC visits) and outcome (age-specific childhood vaccination) had missing values. For ANC, 29.87% of observations were missing and that of postnatal care, 29.9% of observations were missing. Missingness for vaccination ranges from 41.03% for BCG to 41.8% for Rotavirus 2. The overall missingness of childhood vaccination was 48.8% (Table 2.2) which indicates near to half of observations were missing. However, Graham[78] stated that multiple imputations work very well even with 50% missing of the dependent variable. The absolute bias and mean square error were smaller under all missing mechanisms for a high percentage of missingness, even up to 80% [79]. Similarly, Faria R., et al. [80] used multiple imputation for 51% of missingness and White et al. [81] used imputation for 78% of missingness.

To protect against loss of information due to complete case analysis (assuming missing completely at random), a test of missing at random (MAR) is done using regression method as follows:
First a column with missing values are dichotomized into 0 and 1 such that $R =$

$$\begin{cases} 0 \; if \; observation \; is \; observed \\ 1 \; if \; observation \; is \; missed \end{cases}$$

Second, the dichotomized variable is regressed against observed covariates using logistic regression[82]. As a result of missingness is associated with observed values (Table 2.3 for vaccination status, Table 2.4 for ANC, and similar works was done for newborn postnatal care), multiple imputation was used to obtain consistent, asymptotically efficient, and asymptotically normal estimates[83]. We used ordinal regression to impute the missing values of age-specific childhood vaccination; Poisson regression to impute the missing values of number of ANC

contacts; and logistic regression to impute the missing values of newborn postnatal care. The multiple imputation is done using *mi impute [84]* stata command. We created 50% and 30% data sets for outcome and exposure variables as per the rule of thumb of White et al[81] who suggested the number of imputed data sets should be at least as large as percentage of missing observations. The number of imputations should be at least the proportion of missing data[80]. As a result of multiple imputations, we perform sensitivity analysis on the effect of the exposure on the outcome before and after imputation such that the result is given in Table 2.5. The imputation increases data values from 3,208 to 5,150 and the coefficient of the exposure decreased from 0.576 (complete case analysis) to 0.17 (multiple imputation).

Table 2.2**.** Distribution of missing values in childhood vaccination status and ANC.

Vaccines	Observed	% missing	observed	% missing	Vaccines	Observed	% observed	missing	% missing
BCG	3,103	59.0	2,159	41.0	Pneumococcal2	3,073	58.4	2,189	41.6
DPT 1	3,077	58.5	2185	41.5	Pneumococcal3	3,071	58.4	2191	41.6
POLIO 1	3,097	58.9	2,165	41.1	Rotavirus 1	3,062	58.2	2200	41.8
DPT 2	3,076	58.5	2,186	41.5	Rotavirus 2	3,062	58.2	2200	41.8
POLIO 2	3,097	58.9	2,165	41.1	Polio inactive	3,073	58.4	2189	41.6
DPT 3	3,075	58.4	2,187	41.6	Hepatitis B 1	3,077	58.5	2185	41.5
POLIO 3	3,096	58.8	2,166	41.2	Hepatitis B 2	3,076	58.5	2186	41.5
MEASLES 1	3,082	58.6	2,180	41.4	Hepatitis B 3	3,075	58.4	2187	41.6
MEASLES 2	3,076	58.5	2,186	41.5	HiB 1	3,077	58.5	2185	41.5
POLIO 0	3,089	58.7	2,173	41.3	HiB 2	3,076	58.5	2186	41.5
Pentavalent 1	3,077	58.5	2,185	41.5	HiB 3	3,075	58.4	2187	41.6
Pentavalent 2	3,076	58.5	2,186	41.5					
Pentavalent 3	3,074	58.4	2,188	41.6					
Pneumococcal1	3,073	58.4	2,189	41.6	Over all vaccination	2,694	51.2	2,568	48.8
					ANC	3800	70	1614	30
					Postnatal car	3792	69.1	1622	29.9

Table 2.3**.** Test of missing at random (MAR) for childhood vaccination status.

Observed covariate	UOR	P value	AOR	Std. err.	P value	95% CI	
region							
Afar	1.075	0.566	0.906	0.167	0.593	0.631	1.301
Amhara	1.047	0.73	0.929	0.140	0.627	0.692	1.248
Oromia	1.029	0.819	0.939	0.153	0.697	0.682	1.291
Somali	1.332	0.025	1.118	0.211	0.554	0.772	1.618

Benishangul	0.995	0.969	0.942	0.157	0.722	0.679	1.307
SNNPR	1.269	0.058	1.145	0.191	0.418	0.826	1.587
Gambela	1.043	0.762	0.998	0.181	0.989	0.699	1.423
Harari	1.042	0.769	1.045	0.196	0.817	0.723	1.509
Addis Adaba	0.776	0.102	0.841	0.166	0.381	0.570	1.239
Dire Dawa	0.996	0.98	1.060	0.201	0.758	0.731	1.536
place_residence							
Rural	1.156	0.027	0.841	0.097	0.133	0.671	1.054
age_cat							
20-24	1.450	0.011	1.701	0.285	0.002	1.225	2.361
25-29	1.928	<0.001	2.978	0.522	<0.001	2.112	4.198
30-34	2.046	<0.001	4.478	0.903	<0.001	3.016	6.648
35-39	2.669	<0.001	9.661	2.265	<0.001	6.102	15.296
40-44	2.264	<0.001	10.62	2.997	<0.001	6.115	18.471
45-49	3.394	<0.001	29.919	11.523	<0.001	14.065	63.646
Education							
Primary	0.750	<0.001	0.901	0.074	0.203	0.767	1.058
Secondary	0.688	0.001	0.972	0.130	0.831	0.748	1.263
Higher	0.666	0.001	0.945	0.159	0.738	0.680	1.314
television							
Yes	0.778	<0.001	0.942	0.145	0.701	0.697	1.275
religion							
Catholic	0.860	0.686	0.782	0.333	0.564	0.340	1.800
Protestant	1.100	0.247	1.068	0.132	0.594	0.838	1.360
Muslim	1.099	0.146	0.924	0.102	0.475	0.745	1.147
Traditional	1.452	0.128	1.075	0.319	0.807	0.601	1.924
hsize	1.052	<0.001	0.953	0.018	0.010	0.919	0.989
age_hh	1.009	<0.001	1.008	0.004	0.019	1.001	1.015
wealth_index							
Poorer	0.899	0.194	1.047	0.109	0.658	0.854	1.284
Middle	0.891	0.188	1.085	0.120	0.461	0.873	1.349
Richer	0.955	0.606	1.175	0.138	0.167	0.934	1.479
Richest	0.721	<0.001	1.047	0.184	0.794	0.742	1.476
Number of children ever born	1.100	<0.001	7.842	0.610	<0.001	6.733	9.133
age_1st_birth	0.980	0.002	0.902	0.010	<0.001	0.882	0.922
marital_status							
Widowed	1.254	0.303	1.529	0.378	0.086	0.941	2.483
Divorced	1.242	0.108	1.714	0.261	<0.001	1.272	2.309
place_delivery							
At health facility	0.704	<0.001	0.833	0.064	0.018	0.716	0.970
BORD	1.015247	0.19	0.103	0.008	<0.001	0.089	0.121

Table 2.4. Test of MAR for number of Antenatal care visits.

ANC_missing	UOR	z	P value	AOR	z	P value
region						
Afar	2.114	5.310	<0.001	1.237	0.540	0.589
Amhara	0.840	-1.080	0.281	0.778	-0.640	0.522
Oromia	1.475	2.750	0.006	1.038	0.100	0.919
Somali	2.830	7.370	<0.001	1.269	0.600	0.546
Benishangul	1.252	1.460	0.144	0.773	-0.660	0.510
SNNPR	1.355	2.110	0.035	1.271	0.630	0.530
Gambela	0.974	-0.160	0.870	0.548	-1.210	0.228
Harari	1.421	2.230	0.026	1.367	0.780	0.436
Addis Adaba	0.807	-1.140	0.256	0.858	-0.320	0.746
Dire Dawa	1.411	2.120	0.034	1.376	0.800	0.426
place_residence						
Rural	1.488	5.300	<0.001	1.339	1.200	0.231
age_cat						
20-24	1.897	3.560	<0.001	0.914	-0.280	0.782
25-29	2.447	5.170	<0.001	0.395	-2.590	0.010
30-34	2.445	5.050	<0.001	0.206	-3.640	<0.001
35-39	1.780	3.090	0.002	0.079	-4.680	<0.001
40-44	1.802	2.810	0.005	0.073	-4.160	<0.001
45-49	1.410	1.130	0.257	0.008	-3.920	<0.001
education						
Primary	0.753	-4.200	<0.001	1.202	1.080	0.278
Secondary	0.543	-5.030	<0.001	1.240	0.750	0.453
Higher	0.550	-3.980	<0.001	1.449	1.050	0.295
television						
Yes	0.590	-6.240	<0.001	1.328	0.860	0.390
religion						
Catholic	0.911	-0.200	0.841	0.952	-0.050	0.964
Protestant	1.237	2.200	0.028	0.705	-1.160	0.248
Muslim	2.105	9.980	<0.001	1.251	0.900	0.367
Traditional	2.374	3.450	0.001	1.557	0.740	0.459
hsize	1.142	10.560	<0.001	1.009	0.220	0.823
age_hh	0.992	-2.820	0.005	0.989	-1.420	0.155
wealth_index						
Poorer	0.687	-4.300	<0.001	0.824	-0.890	0.371
Middle	0.557	-6.050	<0.001	0.758	-1.090	0.277
Richer	0.573	-5.610	<0.001	1.031	0.130	0.899
Richest	0.433	-9.660	<0.001	0.735	-0.830	0.406
Number of Children ever born	1.167	12.570	<0.001	1.559	6.910	<0.001

age_1st_birth	0.984	-2.260	0.024	1.148	5.290	<0.001
Marital status						
Divorced	0.495	-4.060	<0.001	1.325	0.780	0.434

Table 2.5. Sensitivity analysis of multiple imputations

	Exposure	Number of obs.	Coefficient	std.err	P value	95% CI		length of CI
Complete case analysis	ANC	3028	0.576	0.042	<0.01	0.494	0.658	0.164
Multiple imputation	ANC	5150	0.17	0.008	<0.01	0.154	0.186	0.032

2.3.2 Exploratory Analysis of Main Variables

The distributions of the main study variables (treatment, mediators and outcome) are visualized in Figure 2.2. It shows that the percentage of women with no ANC visits (25.8%) is greater than each number of ANC visits, followed by the percentage of women with four ANC visits (21.6%). The percentage of women with five or more visits decreases continuously. On the other hand, the percentage of women who delivered at home (49%) is equivalent to the percentage of women who delivered at health facilities. The percentage of newborns who received a postnatal checkup (13%) is much lower than the percentage of newborns who did not receive a postnatal checkup (87%) (Figure 2.2).

The percentage distribution of childhood vaccination at the recommended age or according to the vaccination schedule shows that only 3.2% of children received the recommended doses of vaccines at the right time. This value indicates the percentage of children who received vaccines at the recommended vaccination schedules (such as at birth, at 6 weeks, at 10 weeks, at 14 weeks, and at 9 months). The largest proportions of children (81%) are partially vaccinated, which implies that they missed one or more vaccines in each vaccination schedule. A significant proportion of children (16%) do not receive any vaccines at the right time (Figure 2.2).

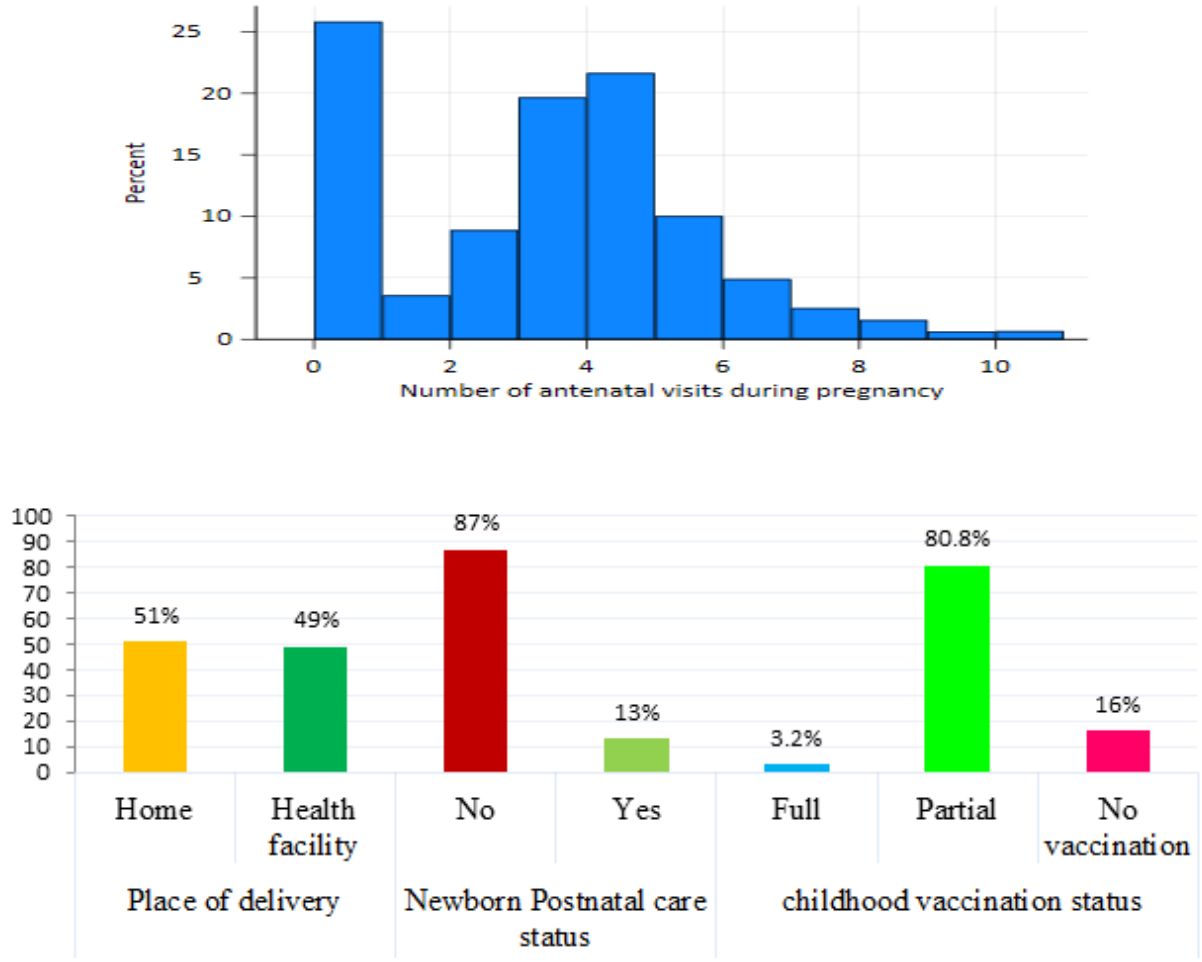


Figure 2.2. Graphical presentation for the distribution of important variables

Figure 2.3 reveals the percentage distribution of vaccination status along with the number of ANC visits. Most children whose mothers do not receive ANC visits are unvaccinated according to the recommended schedule, and the percentage of children who do not receive vaccinations decreases when mothers get ANC service. The percentages of children who are partially vaccinated are concentrated on those with 3 to 5 ANC visits. Similarly, the percentages of children with full vaccination are the highest among mothers who have four ANC visits, followed by those with three and five visits.

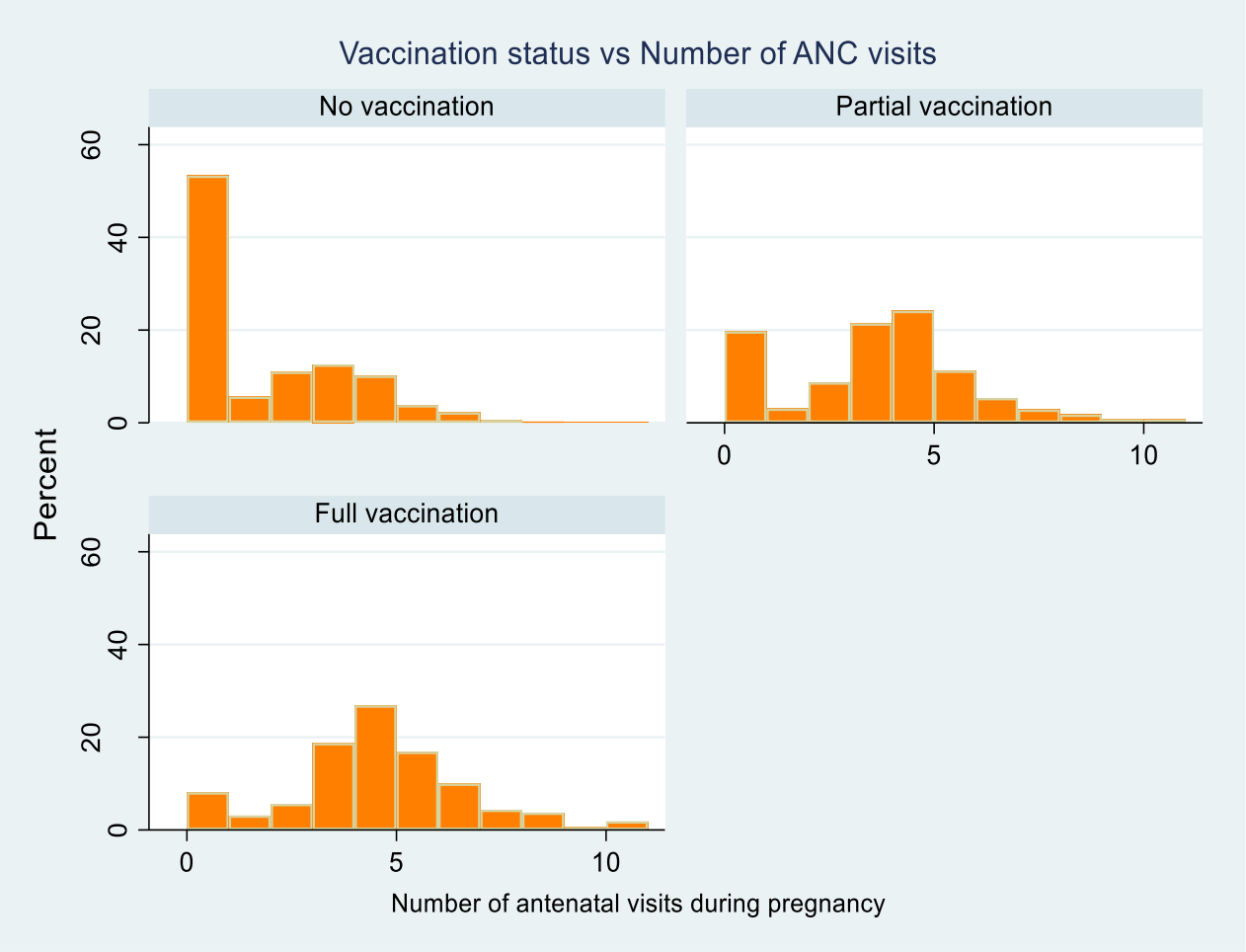


Figure 2.3. Distribution of vaccination status along with number of ANC visits.

Figure 2.4 demonstrates that the percentage of no postnatal checkup is more than the percentage of postnatal checkups among non-vaccinated children at the right schedule. On the other hand, the percentages of children with postnatal checkups are higher than the percentage of children with no postnatal checkup among fully vaccinated children on the right schedule.

In addition, the percentage of children delivered at home (23.4%) is greater than the percentage of children delivered at a health facility (7.8%) among non-vaccinated children (Figure 2.5). Similarly, the percentage of children delivered at a health facility (87%) is more than the percentage delivered at home (75.4%) among partially vaccinated children. The result further indicated that among fully vaccinated children, the percentage of children delivered at a health facility (5.2%) was more than the percentage of children born at home (1.2%).

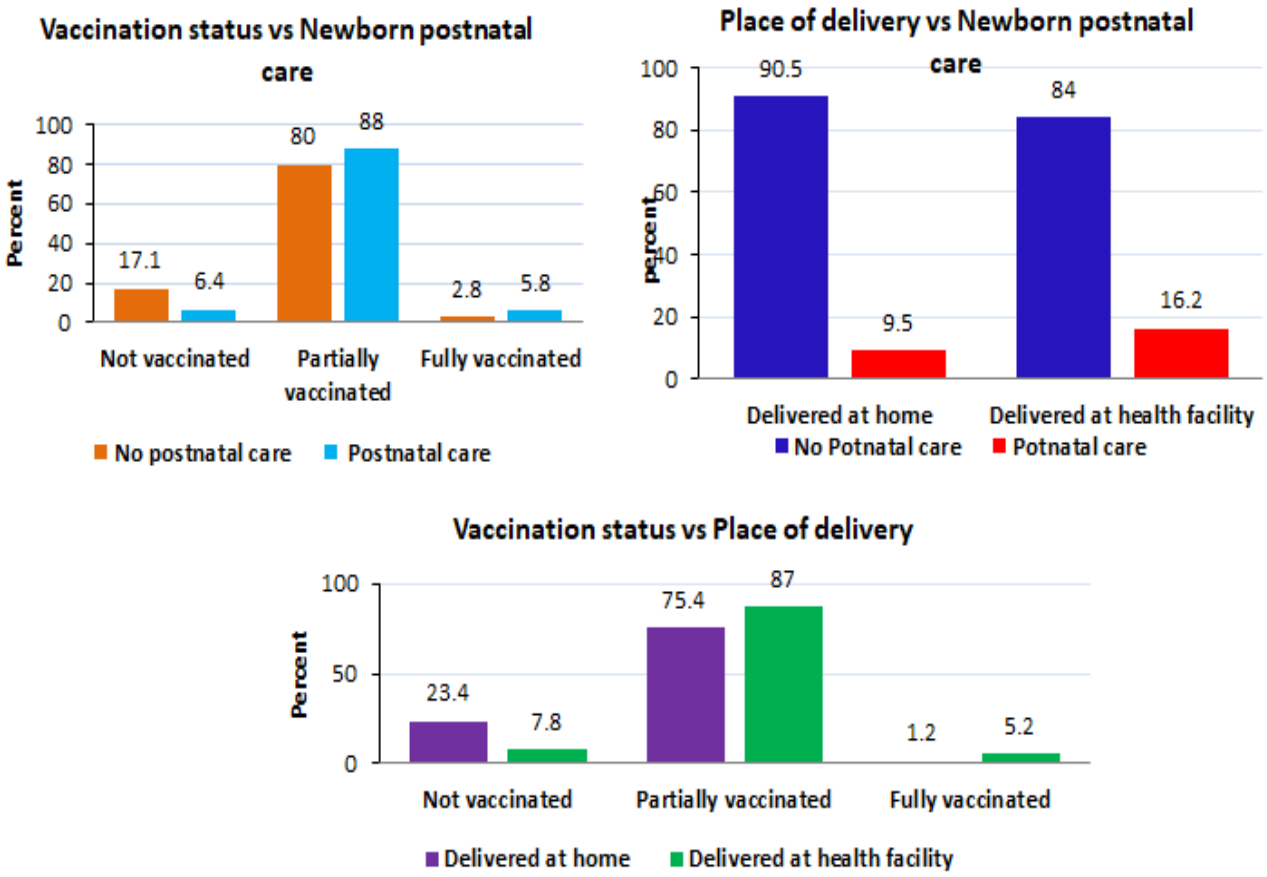


Figure 2.4. Distribution of place of delivery and postnatal care by vaccination status.

Conclusion

The study used a sample of 5,150 women and children for analysis. We employed the multiple imputation method to impute missing values for the outcome and the exposure, the mediators, and the outcome variables using Poisson, logistic, and ordinal regression. Missing at Random (MAR) was assumed and tested to impute the missing values from observed covariates. The contributing factors for outcome missingness included mothers' age category, household head age, household size, number of children ever born, age at first birth, divorce from marriage, place of delivery, and birth order number.

The number of women with one or more than six ANC visits was rare, and most women had no visits. The number of women who delivered at home and in health facilities was almost equivalent. A large number of newborns did not receive postnatal care within six weeks of birth. The findings

show that the number of children who received all recommended vaccines at each specific age was very small. On the other hand, most children missed one or more vaccines at each recommended vaccination schedule.

Among non-vaccinated children, most of their mothers had no ANC visits. Conversely, among fully vaccinated children, most of their mothers had ANC visits from three to five. The finding also implies that children who are born at health facility and receive postnatal care are more likely to get vaccines at the recommended schedule.

The preceding chapter deals with missing data management and the exploration of main variables in the study. However, the chapter does not incorporate the inferential analysis to identify confounders, methods of adjusting them, and estimate the treatment effect on the outcome. Hence, the subsequent chapter addresses what has not been done in the preceding chapter.

CHAPTER THREE: IDENTIFYING CONFOUNDERS AND ESTIMATING EFFECTS

3.1. Confounder Identification for ANC and Childhood Vaccination

3.1.1. Background `

A child who is 12-23 months old and has received all the immunizations advised by the extended immunization program is considered to have received a full vaccination [85-89]. However, for the effectiveness of vaccination, timing is very important. A timely start to vaccination is critical in the first year of life as transplacental immunity decreases fast, and timely vaccination administration has consequences for the efficacy of pediatric immunization programs[90]. Early or late administration of vaccination reduces the impact of vaccine programs on disease burden especially in high-risk groups[91].

For example, except for BCG and polio at birth, any vaccine given before six weeks has shown poor response and in some cases could be harmful to infants as they decrease the immune response of consequent doses. Hence, giving vaccines before schedule or closer to each other may lead suboptimal immune response. Conversely, the optimal level of vaccine protection may not be achieved, if a child's vaccination is delayed and the time between doses/vaccines is lengthened [92]. Individual as well as herd immunity is compromised if vaccines are given with considerable delays, which is not surprising given that outbreaks of diseases like pertussis or measles will happen[93].

In observational study, confounders have to be identified and controlled their effect while estimating the association between exposure and outcome. Controlling confounders helps to get unbiased estimates of the exposure-outcome relationship[9]. Including all pre-treatment covariates in any confounder controlling methods such as regression introduces bias. Adding more covariates to the model causes over fitting and unstable coefficients due to multicoliearity[10]. A model is best when it contains smallest number of covariates that explain the greatest amount of variance[11]. As a result identifying confounders that potentially distort the causal effect of treatment on the outcome is imperative. However, how to identify confounders and how to deal with them is one of the challenges in observational study. There is no common consensus criteria

to identify which covariates are confounders and which are not[11]. A common approach is to control as many pre-exposure covariates as possible[5]. Studies modified this approach by controlling all covariates significantly associated (p-value less than 0.05) with the outcome of interest [12, 13] mentioned in [5]. Others also stated that control confounders that give a pre-determined magnitude of change (10% or 15%) while estimating the relationship between exposure and outcome[12, 14].

Confounders that distort the causal effect of ANC contacts on age-specific or timely vaccination have not been documented yet. In addition, little has been done on comparisons of confounder identification techniques, especially with count exposure like frequency of mothers' ANC contact at health facilities. Accordingly knowing confounders and controlling them will be crucial to determining the true relationship between the frequency of ANC contacts and age-specific childhood immunization. Hence, this study was done to determine the best confounder identification technique; confounders that affect the causal effect of ANC on age-specific childhood vaccination; and estimate the effect of ANC on age-specific childhood vaccination after adjusting confounders with regression method. This section is presented according to Iyassu et al.[94] presented in the Appendix .

3.1.2. Measure of Outcome Variable and Conceptual Framework

The outcome variable is age-specific childhood vaccination status categorized into three ordered categories (no vaccination, partial vaccination and full vaccination). A child is considered fully vaccinated when receiving BCG and OPV0 at birth; DTP-HepB1-Hib1, OPV1, PCV1, and Rota1 at 6 weeks of birth; DTP-HepB2-Hib2, OPV2, PCV2, and Rota2 at 10 weeks of birth; DTP-HepB3-Hb3, OPV3, PCV3, and IPV at 14 weeks of birth, and Measles at 9 months of birth [49]. When a child received a particular vaccination, a score of 1 was given otherwise 0. With these scores, a composite index score was created with respect to expected number of vaccines at each vaccine schedule. When a child received all recommended vaccines timely in each respective vaccine schedule, it was labeled as fully vaccinated. If one or more vaccines were missed at each vaccine schedule, it was labeled as partially vaccinated and labeled as no vaccination when a child took no vaccination at each age.

We considered the ANC as the causal/treatment variable that causes the age-specific childhood vaccination status. The conceptual framework of pre-treatment covariates or possible confounders

relation with exposure and outcome is shown in Figure 3.1. These pre-treatment covariates were selected from a review of literature; and their availability in the data set was checked before adding into the framework.

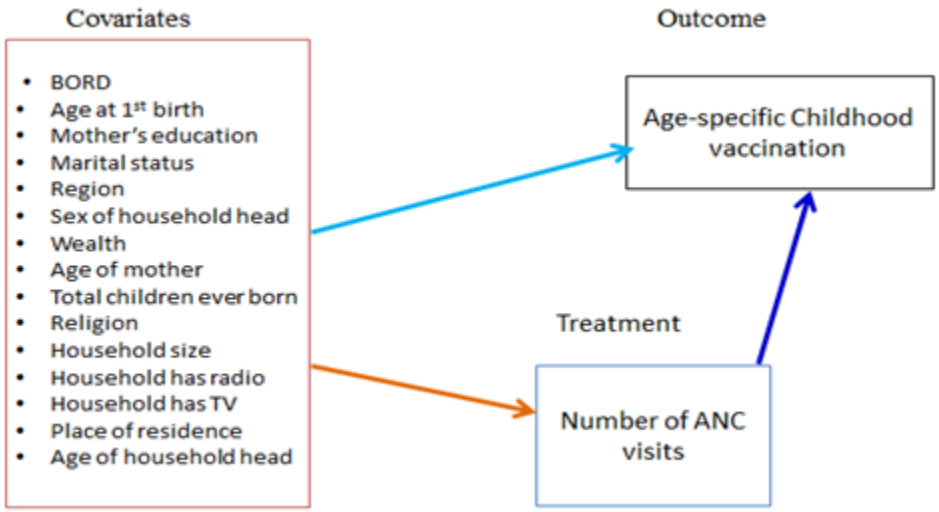


Figure 3.1. Conceptual framework.

3.1.3. Methods of Identifying Confounders

Adjusting for a set of covariates assumed to be confounders, capable of producing spurious associations between exposure and outcome, is the common method of estimating causal effects in observational studies [95]. For a variable to be confounder: it should not be in the causal pathway between exposure and outcome; and it must be unequally distributed between study subjects[96, 97]. An important step before applying statistical methods to control confounders is identifying covariates that confound the causal effect of a number of ANC on age-specific childhood vaccinations. Methods used to identify confounders are significance testing or p-value, and change in estimate.

Significance testing or the p-value method is used after identifying data-driven statistical models for cofounder-exposure and cofounder-outcome relationships. A threshold p-value =0.2 is used to identify a covariate as a cofounder[9, 98].

In this case, four approaches were compared. The first approach is selecting common significant covariates for both exposure and outcome. The second is selecting a significant covariate in exposure, outcome or both variables. The third approach is taking all pretreatment variables as

confounders. The fourth approach is control those covariates significantly associated with the outcome [99, 100].

To compare the performance of each approach, AIC, BIC, and likelihood ratio tests were used after regressing the outcome variables on selected covariates and exposure. The significance of a covariate was evaluated based on a threshold value at p=0.2 [9, 98] during bivariate analysis. Moreover, the relative change of exposures' (ANC service) effect on outcome (childhood vaccination status) was taken into account to select the approach.

Change in the estimate is a method of identifying confounders based upon the inclusion of covariate changes in the estimate of the causal effect of exposure on outcome[8] by more than the threshold value, usually 10%[101].

In divergence with significance testing, change in estimate (CIE) approaches identify covariates based on how much their control changes exposure's effect estimates regardless of significance or p-value; the observed change is supposed to measure confounding by the covariate[102]. We have used two approaches to measure the change in the estimate which are change in estimate based on the coefficient and based on the attributable fraction (AF) using odds ratio[102]. Let β_a be coefficient of exposure on outcome without a covariate**,** Z and β_z be coefficient of exposure on outcome when a covariate, Z is added on exposure-outcome relationship, then relative change in estimate due to covariate Z is estimated as:

$$\Delta = |(\beta_a - \beta_z)/\beta_a|. \qquad (3.1)$$

In applying change in estimate (CIE), covariate, Z is considered cofounder and included in the final model when $\Delta > 0.1\ or\ 10\%$ [9].

When taking into account of odds ratio, let RRa and RRu denote the estimated risk ratio with and without adjustment of covariates; then RRa/RRu is the conventional method of measuring change in estimate or change in importance. However, Greenland S, Pearce N [102] suggested AF=$(OR - 1)/OR$ is more relevant and change in estimate could be measured by |AFa-AFu|.

To apply the significance and change in estimate methods, we need to specify both the treatment and outcome models correctly. In common cause and either treatment or outcome cause

approaches require both treatment and outcome models. Whereas, the outcome cause and change in estimate methods require the outcome model specification.

3.1.4. Treatment Model

The exposure variable in this study (the number of ANC contacts of pregnant women with values ranging from 0 to 11) is count. Hence, families of count models are used to model the relationship between pre-exposure covariates and exposure. The models can be categorized into two broad families; generalized linear model (GLM) family with log link and zero-augmented family models. The GLM family includes Poisson regression and its extension called negative binomial regression. And the family of zero-augmented includes zero inflation Poisson and zero-inflated negative binomial regression[103].

Taking a scalar value y $(with\ values = 0, 1, 2, \ldots, k)$ and $\boldsymbol{x}$ is a vector of covariates, the probability density function for generalized linear model is:

$$f(y, \gamma, \emptyset) = \exp\left(\frac{y.\gamma - b(\gamma)}{\emptyset} + c(y, \emptyset)\right). \qquad (3.2)$$

Where γ is a canonical parameter and $\emptyset$ is a dispersion parameter. The functions $b(\cdot)$ and $c(\cdot)$ are known and determine which member of the family is used. Conditional mean and variance of y are given by $E(y|\boldsymbol{x}) = \mu = b'(\gamma)$ and $var(y|\boldsymbol{x}) = b''(\gamma)$ [104].The Poisson GLM is $g\big(E(y|\boldsymbol{x})\big) = \boldsymbol{x'\beta}$ with canonical link function of $g\big(E(y)\big) = \log E(y|\boldsymbol{x})$. The mean and variance are equal: $E(y) = var(y) = \mu$ and the dispersion parameter $\emptyset$, is 1. However, when there is overdispersion, $\emptyset > 1$, the Poisson GLM is negative binomial with the same canonical link function[103].

In the count outcome variable, one set of observations may be necessarily zero, and another set that may be zero according to a random event points naturally to a mixture model in which two types of zeros can occur. The relevant distribution is a mixture of an ordinary count model such as the Poisson or negative binomial with one that places all its mass at zero.

According to Lambert (1992) mentioned in [105], the zero-inflated model assumes

$$y_i \sim \begin{cases} 0 & \text{with probability } 1 - \theta_i \\ poison(\gamma) & \text{with probability } \theta_i \end{cases} \qquad (3.3)$$

The unconditional probability is given by $P(y_i = 0) = (1 - \theta_i) + \theta_i e^{\gamma_i}$ and

$$(y_i = j) = \theta_i \frac{e^{\gamma_i}}{j!} {\gamma_i}^j \ , j > 0 \ . \qquad (3.4)$$

Considering a vector of covariates, the zero-inflated model which is a mixture of two models is given by $logit(\theta) = \boldsymbol{x'_1\beta_1}$ and $\log(\gamma) = \boldsymbol{x'_2\beta_2}$, where $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ may or may not be the same. In case of over-dispersion, y_i is negative binomial with mean γ and dispersion parameter $\emptyset$ [105].

Maximum likelihood estimation

Let $y = 0,1,2,.,.,.,k$ indicates the number of ANC visits and $\boldsymbol{X}$ indicate the p-dimensional covariates, then the GLM of the Poisson distribution is outlined as:

$$P(y|\boldsymbol{X};\boldsymbol{\beta}) = \frac{[\exp(\boldsymbol{X'\beta})]^y \exp[-\exp(\boldsymbol{X'\beta})]}{y!} \ . \qquad (3.5)$$

Then the likelihood function of the conditional distribution is defined as:

$$\ell[P(y|\boldsymbol{X};\boldsymbol{\beta})] = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} P(y_i|\boldsymbol{x_i};\boldsymbol{\beta}). \qquad (3.6)$$

The log-likelihood function is then stated as:

$$\ell(y,\boldsymbol{X};\boldsymbol{\beta}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} logP((y_i|\boldsymbol{x_i};\boldsymbol{\beta})).$$

$$= \sum_{i=1}^{n}[y_i \boldsymbol{x_i}'\beta - \exp(\boldsymbol{x_i}'\beta) - \log(y_i!)].$$

This logarithmic transformation is a monotonic function which indicates the maximization of log-likelihood function $\ell(y,\boldsymbol{X};\boldsymbol{\beta})$ to estimate parameters $\widehat{\boldsymbol{\beta}}$ by taking the first derivative on $\ell(y,\boldsymbol{X};\boldsymbol{\beta})$ and equating to zero. This is given as :

$$\frac{\partial \ell(y,\boldsymbol{X};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}[y_i - \exp(\boldsymbol{x_i}'\beta)]\boldsymbol{x}_i = 0. \qquad (3.7)$$

Hence, the Hessian matrix of the model is expressed as :

$$H(y,\boldsymbol{X};\boldsymbol{\beta}) = \frac{\partial^2 \ell(y,\boldsymbol{X};\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta'}}.$$

$$= \frac{\partial[\sum_{i=1}^{n}[y_i \boldsymbol{x}_i - \exp(\boldsymbol{x_i}'\beta)\boldsymbol{x}_i]]}{\partial \boldsymbol{\beta'}}.$$

$$= -\sum_{i=1}^{n} \exp(\boldsymbol{x_i}'\beta) \, . \, \boldsymbol{x_i} \, \boldsymbol{x_i}'.$$

Since $H(.)$ is negative definite, the function is concave and Newton-Raphson iteration method is used to estimate $\widehat{\boldsymbol{\beta}}$ [106]. Similarly, the parameters of ZIP model are estimated using maximum likelihood estimation[107, 108].

3.1.5. Outcome Model

The outcome variable (age-specific childhood vaccination) is ordinal with three potential values (no vaccination coded as 0, partial vaccination coded as 1, and full vaccination coded as 2). Let $\boldsymbol{y_j}$, $j = 0, 1, 2$ be the outcome with the cumulative probability of $\gamma_j = P(Y \leq j) = \pi_1 + \pi_2 + \cdots + \pi_j$, where $\sum_{j=1}^{J} \pi_j = 1$ and π_j is the probability in the j^{th} category, then the cumulative link model with exposure, z, and a vector of covariates, $\boldsymbol{x}$ is given by [109]:

$$\gamma_j = G(u_j),\ u_j = \alpha_j - (\theta * Z + \boldsymbol{x'\beta}). \qquad (3.8)$$

Where G is the inverse link function and α_j is the thresholds or cut points which are strictly in order: $-\infty \equiv \alpha_0 \leq \alpha_1 \leq \alpha_2 \equiv \infty$.

The commonly used link function for cumulative link model is the logit which leads to

$$logit(P(y \leq j|z, \boldsymbol{x}) = log\left(\frac{P(y \leq j|z,\boldsymbol{x})}{1-P(y \leq j|z,\boldsymbol{x})}\right) = \alpha_j - (\theta * Z + \boldsymbol{x'\beta}). \qquad (3.9)$$

Then the odds ratio of the event $y \leq j$ at treatment level z_2 relative to z_1 is:

$$OR = \frac{\gamma_j(z_2)/1-\gamma_j(z_2)}{\gamma_j(z_1)/1-\gamma_j(z_1)}.$$

$$= \frac{\exp(\alpha_j - (\theta * z_2 + \boldsymbol{x'\beta}))}{\exp(\alpha_j - (\theta * z_1 + \boldsymbol{x'\beta}))}.$$

$$= \exp\left[(\alpha_j - (\theta * z_2 + \boldsymbol{x'\beta}) - (\alpha_j - (\theta * z_1 + \boldsymbol{x'\beta})\right]$$

$$= \exp(-\theta(z_2 - z_1)). \qquad (3.10)$$

Whereas for the event $y > j$, the OR= $\exp(\theta(z_2 - z_1))$.

Other link functions are [109]:

Complementary log-log link function: $\log(-\log(1 - \gamma_j) = \alpha_j - (\theta * Z + \boldsymbol{x'\beta})$.

Negative log-log link function: $-\log(-\log(\gamma_j)) = \alpha_j - (\theta * Z + \boldsymbol{x'\beta})$.

Probit link function**:** $\boldsymbol{\Phi^{-1}}(\gamma_j) = \alpha_j - (\theta * Z + \boldsymbol{x'\beta})$.

The assumption of cumulative link model is, except for the intercept, the effect of the covariate and the exposure is constant for each increase in the level of the response [110]. If this assumption fails to hold, then the general formula for cumulative link model is [109]:

$$\gamma_j = G_\lambda(u_j), u_j = \frac{g_{\boldsymbol{\alpha}}(\alpha_j) - \theta * Z - \boldsymbol{x'\beta} - \boldsymbol{w'\theta}}{\exp(z_i \boldsymbol{\varsigma})}. \quad (3.11)$$

Where $G_\lambda(.)$ is the inverse link function. It may be parameterized by the scalar parameter λ in which case we refer to ${G_\lambda}^{-1}$as a flexible link function.

$g_{\boldsymbol{\alpha}}(\alpha_j)$ parameterizes thresholds , α_j by the vector $\boldsymbol{\alpha}$ such that g restricts $\alpha_j$to be symmetric or equidistant.

$\boldsymbol{x'\beta}$ are the ordinal regression effects

$\boldsymbol{w'\theta}$ are regression effects which are allowed to depend on the response category j and they are denoted partial or non-proportional odds[111] when the logit link is applied. To include other link functions in the terminology, we denote these effects nominal effects because these effects are not integral to the ordinal nature of the data.

$\exp(z_i \boldsymbol{\varsigma})$ are scale effects since in a latent variable view, these effects model the scale of the underlying location-scale distribution.

Furthermore, different link functions of the proportional odds model were compared. The analysis was done using the R ordinal package[112].

To select best fit treatment model and best fit link function for outcome model, AIC[113], and BIC [114] were used with the formulas:

$$AIC = -2\log\left(L(\widehat{\boldsymbol{\theta}})\right) + 2k. \quad (3.12)$$

$$BIC = -2\log\left(L(\widehat{\boldsymbol{\theta}})\right) + klog(n).$$

where $L(\widehat{\boldsymbol{\theta}})$ is the likelihood value when estimating parameter $\boldsymbol{\theta}$, k is the number of parameters to be estimated and n is the sample size. The candidate model is "best" with smallest value of AIC and BIC. As shown in the formula, AIC penalizes the model for additional parameter to decrease over fitting. In BIC, the penalty increases as the sample size increases. Hence, BIC outperforms AIC for large sample size[115, 116].

Estimation of parameters

For individual i let $Y_{i1}, Y_{i2}, ., ., ., Y_{ij}$ denotes the binary indicator of the response, where Y_{ij} is 1 when the response is in j^{th} category and 0 otherwise. Then Y_{ij} is multinomial distributed as Y_{ij}~multinomial(π_{ij},1) [109, 110] with $\pi_{ij}, = P(Y_i = j|X) = P(Y_i \leq j|X) - P(Y_i \leq j-1|X)$.
The likelihood function of $\pi_{ij}\ is$:

$$P(Y_i = j|X) = \prod_{i=1}^{n} \prod_{j=1}^{J} (\pi_{ij})^{y_{ij}} \ . \tag{3.13}$$

and the log likelihood function is:

$$\begin{aligned} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}; y_{ij}) &= \sum_i \sum_j y_{ij} \log(\pi_{ij}). \\ &= \sum_i \sum_j I(y_{ij}) \log(\pi_{ij}). \\ &= \sum_i \log(\pi_{ij}). \end{aligned} \tag{3.14}$$

For logit link function, the log likelihood of $P(Y_i = j|X)\ is:$

$$\sum_i \log\left(\frac{\exp(\alpha_j - (\theta * z_2 + \boldsymbol{X_i}'\boldsymbol{\beta}))}{1+\exp(\alpha_j - (\theta * z_2 + \boldsymbol{X_i}'\boldsymbol{\beta}))}\right) - \sum_i \log\left(\frac{\exp(\alpha_{j-1} - (\theta * z_2 + \boldsymbol{X_i}'\boldsymbol{\beta}))}{1+\exp(\alpha_{j-1} - (\theta * z_2 + \boldsymbol{X_i}'\boldsymbol{\beta}))}\right). \tag{3.15}$$

And this can be extended to other link functions.
The cumulative link models (CLMs) in the ordinal package are estimated with a regularized Newton-Raphson (NR) algorithm with step-halving (line search) using analytical expressions for the Gradient and Hessian of the negative log-likelihood function. The starting values of the regression parameters are initialized at 0 [109].

3.1.6. Result

Model selection for covariate-exposure and covariate-outcome relationship

The deviance, AIC, and BIC values for candidate models are summarized in Table 3.1. The result revealed that the zero-inflated Poisson regression model has smaller deviance, AIC, and BIC values as compared to other candidate models. Hence, the zero-inflated Poisson model is the best performing to model covariate-exposure relationship (Table 3.1) for the significance testing approach.

Table 3.1 also shows the values of the model comparison methods for the link functions of the outcome model (cumulative link model). The result demonstrates that the deviance, AIC, and BIC values of the log-log link functions are smaller than those of the other link functions for the ordinal outcome model. Thus, log-log link function is used to model covariates-outcome relationship.

Table 3.1. Model comparisons for covariate-exposure and covariates-outcome relationship.

	Model	-2*Loglikilihood	AIC	BIC
Exposure	Poisson	19984.44	20068.44	20343.40
	Negative binomial	19924.87	20010.87	20292.38
	Zero inflated Poisson regression	18439.54	18611.54	19174.56
	Zero inflated negative binomial	18446.83	18620.83	19190.40
	Link functions for cumulative link model	-2*Loglikilihood	AIC	BIC
Outcome	logit	5110.16	5198.17	5486.22
	probit	5156.3	5244.3	5532.37
	Cloglog	5448.52	5536.52	5824.58
	Loglog	5017.12	5105.11	5393.17
	Cauchit	5048.44	5136.43	5424.49

Note that Aranda-Ordaz and log-gamma link functions do not converge

Deviance=-2*Loglikilihood value

Significance testing methods to identify confounders

Table 3.2 shows the result of confounder identifying approaches into two ways. The first is based on model performance metrics and the second is based on the change of treatment effect in each approach. The result shows that the AIC values of common cause, either treatment or outcome cause and outcome cause are equivalent with small variation in their BIC values. The AIC and BIC values of all pre-treatment approaches are greater than the other three approaches. The BIC value of the common cause approach is little smaller than other approaches. However, the likelihood ratio test was not significant for each pair of approaches.

On the other hand, change of estimate for the effect of ANC on age-specific childhood vaccination before and after controlling covariates for all approaches of significance testing method is also shown in Table 3.2. The unadjusted treatment effect size decreases when we adjust the covariates in each approach. The common cause, either treatment or outcome cause, and all pretreatment approaches reduced the unadjusted treatment effect equivalently. The reduction of the unadjusted treatment effect by the outcome cause is little greater than other approaches though the difference is not substantial.

The intention of identifying confounders is to control their effect and estimate the true treatment effect. The outcome cause approach generated relatively larger treatment effect size change though common cause generated relatively better model fitness due to smaller value of the BIC. This shows that there is tradeoff between treatment effect change and the BIC value. The small variation of each approach performance could be because of small variation in terms of the number of confounders included in the model for each approach.

Table 3.2**.** Result for comparisons of significant testing approaches of confounder identification.

Comparisons based on model performance metrics					
			Likelihood ratio test		
Approaches	AIC	BIC	LR.statistics	P_value	Number of covariates
Approach1: common cause	5013.65	5262.42			12
Approach2: Cause for treatment or outcome or both	5013.04	5268.36	2.6094	0.1062	13
Approach 3: Outcome cause	5013.80	5275.8	1.2349	0.2665	14
Approach4:Pretreatment	5021.55	5316.16	2.2488	0.8138	15

Comparisons based on change of treatment effect				
Approach	coefficient	risk ratio	Relative Change of coefficients	Relative Change of risk ratio(0 vs 1)
Unadjusted ANC	0.161	0.427		
Approach 1: Common cause of treatment and outcome	0.103	0.406	0.360	0.9505
Approach 2: cause of treatment or outcome or both	0.103	0.406	0.361	0.9505
Approach 3: Outcome cause	0.102	0.405	0.364	0.950
Approach 4: All pre-treatment	0.103	0.406	0.362	0.9503

Change in estimate method

Table 3.3 presents a change in the estimate of an exposure when individual covariates are included in the exposure-outcome model. The result reveals that there are six covariates to change treatment effect size by at least 7% and AF of about one. The covariates include region, place of residence, education status, existence of television at home, religion and household wealth status are identified as confounders that alter the effect of the exposure (number of Antenatal care at pregnancy) on the outcome (age specific childhood vaccination).

Table 3.3. Results of change in estimate.

Exposure	Coefficient of ANC	CIE(coeff,)	CIE(AF)
Unadjusted ANC	0.161		
Age of mothers	0.164	2%	0.281
Region	0.128	20%	2.823
place residence	0.142	12%	1.600
Education status	0.147	9%	1.173
Radio	0.157	3%	0.348
Television	0.144	10%	1.414
Religion	0.149	7%	0.981
Household size	0.161	0%	0.014
Sex of household head	0.161	0%	0.003
Age of household head	0.161	0%	0.040
Household wealth status	0.123	24%	3.282
Total children ever born	0.162	1%	0.078
Age at 1st birth	0.159	1%	0.147
Marriage status	0.161	0%	0.031
Birth order number	0.158	1%	-0.187

Comparison of significance testing and change in estimate methods

The results of comparing significance testing and change in estimate method are presented in Table 3.4. The comparison is performed using model selection techniques and change of treatment effect size. The model selection techniques favor significance testing method having smaller AIC and BIC values than change in estimate method. The difference of relative change of treatment effect on the outcome is not meaningful between significance testing method and change in estimate method. On the other way, change in estimate methods identified much smaller number of covariates as confounders compared to significance testing methods.

In addition to comparing the two methods, Table 3.4 also presented the change in estimate method at different threshold points (relative treatment effect change at 7%, 9%, and 10%). It implies that the 7% threshold point performed better than the other threshold points having the smallest values of AIC and BIC, and changes the crude treatment effect by about 37.1%. Hence, it may be important to lower the threshold point of change in estimate below 10%. In all threshold values,

the change in estimate is more conservative to keep optimal number of confounders. It keeps less than half of confounders as compared to significance testing method.

Taking significance testing as a better approach to identify confounders especially considering covariates that cause the outcome, the confounders were mothers' age category at birth, region, place of residence, mothers' education status, possession of radio and television, religion, household size, sex of household head, age of household head, household wealth status, total children ever born, age at first birth and birth order number.

Table 3.4**.** Results for Comparison of change in estimate and significance testing methods.

Comparisons based on model fitness performance metrics			
Confounder selection technique	Number of covariates	AIC value	BIC value
Change in estimate(7% threshold)	6	5277	5454
Change in estimate(9% threshold)	5	5283	5408
Change in estimate(10% threshold)	4	5283	5408
Significance testing(outcome cause)	14	5075	5285
Comparisons based on change of treatment effect size			
Confounder selection technique	Number of covariates	Coeff. of ANC	Relative change of ANC effect
Unadjusted ANC effect		0.1607	
Change in estimate(7% threshold)	6	0.1011	0.3709
Change in estimate(9% threshold)	5	0.1027	0.3609
Change in estimate(10% threshold)	4	0.1027	0.3609
Significance testing(outcome cause)	14	0.1022	0.3640

Estimating the effect of ANC on age-specific childhood vaccination

To estimate the causal effect of ANC on the cumulative probability of age-specific childhood vaccination, we first examined the multicollinearity of covariates and proportional odds assumption. Accordingly, the total number of children ever born was found collinear with other covariates and excluded from the model. In addition, age of mother at first birth and birth order number violates the proportional odds assumption, and we estimated partial proportional odds model. The model selection values and treatment effect estimate of proportional odds model and partial proportional odds model are presented in Table 3.5. The estimates are obtained after adjusting the selected confounders with regression method. The result reveals that the values of

model selection parameters and ANC effect size along with its standard error are equivalent. Thus, taking the effect size obtained from proportional odds model, the effect size of ANC on the loglog of cumulative probability of age-specific childhood vaccination was 0.114. This implies that when a pregnant woman has one more ANC visits, the loglog of cumulative probability of a child to be vaccinated at the right schedule will increase. In other words, ANC visit increases the higher order probability of age-specific childhood vaccination.

Table 3.5. Treatment effect in proportional and partial proportional odds models.

Models	AIC	BIC	ANC effect size	Std. Error	z value
Proportional odds model	5256	5445	0.114	0.0103	10.961***
Partial proportional odds model	5246	5449	0.113	0.0103	10.914***

*** Significant at 0.0001. The value of ANC effect size in table 10 is different from the effect size in Table 9 because of discarding collinear covariates.

3.1.7. Discussion and Conclusion

The purpose of identifying confounders is to get minimally sufficient covariates and control them using statistical methods such as regression, and estimate effect of the exposure on the outcome[117]. On the other hand, including all pre-treatment covariates in the regression model to adjust them causes over fitting since some covariates retained in the model may be noisy in that the model will not be reproducible for other data sets[10]. Two principal methods were explored for cofounder identification, which were significance testing and change in estimate methods.

Count models were compared concerning their performance for the relationship between pretreatment covariates and the treatment. Among others, zero-inflated Poisson regression was found the best fit. Furthermore, a cumulative link or proportional odds model with various link functions was proposed for pretreatment covariates and outcome variables relationship. Accordingly, log-log link function was found the best fit and used in significance testing and change in estimate methods.

Selecting all pretreatment covariates as confounders is one approach to confounder identification. In this approach, all covariates before the exposure should be controlled in causal inference [8]. The “common cause” approach of confounder identification is the second approach that works by

controlling all covariates that are significantly associated with the exposure and outcome[118]. The outcome cause is the third approach which states control all covariates that are significantly associated with the outcome (p-value less than 0.05) [12, 13]. The fourth approach is controlling confounders that are significant causes of exposure or outcome or both[8]. In this study, all four approaches were tested using statistical model selection techniques and the amount of treatment effect change due to each approach. The statistical model selection techniques cannot identify which method is the best except all pretreatment covariates approach poorly performed. Using treatment effect change, outcome cause approach was found the appropriate to identify confounders and control their effect. This study is consistent with other findings in the literature [99, 100, 119, 120].

Change in estimate is efficient when the cut point is set to 10% with and without adjustment of covariates [98]. However, this study shows that the threshold point can be below 10% threshold values. The number of confounders identified in change in estimate was smaller than that of the significance method. The two methods were compared for their performance using statistical model selection techniques (AIC, and BIC values) and the amount of treatment effect change. Maldonado and Greenland[98] stated that significance testing performed best when the significance value was set to 0.2. On the other hand, Talbot D et al. [121] questioned the ability of change in estimate to identify confounders due to its low ability to improve the precision of estimates. The study found that the number of ANC contacts increases the likelihood of childhood vaccination at each recommended vaccination schedule or it increases the higher level vaccination status categories.

To conclude, the outcome cause significance testing method relatively outperformed other significant testing methods in terms of amount of treatment effect size change. The change in the estimate method of confounder selection gave a smaller number of confounders and it is stricter than significance testing. It is also possible to lower the threshold point below 10% to identify confounders based on change in estimate method. Finally, using multidimensional approaches like statistical model selection and estimated effect size change are important to identify confounders in observational study.

The conclusion of this study is given solely based on real-life household survey data. However, this may not be reproducible for other similar studies since it is not verified in simulation based

study. In addition, confounders were adjusted with regression regardless of their distribution across the treatment variable. Thus, the subsequent section overcomes these limitations using binary treatment and ordinal outcome.

3.2. Estimating the Effect of Place of Birth on Age-Specific Childhood Vaccination

3.2.1. Background

According to [122, 123], three characteristics should be satisfied for a variable to be a confounder. It should be associated with exposure and outcome; it must be distributed unequally between treatment and control groups; and it should not be in the causal pathway between exposure and outcome. Temporally, confounders come prior to exposure. A covariate that comes after exposure has taken place is not a confounder since it is unable to retroactively modify the exposure[5].

In observational study, confounders have to be identified and controlled their effect while estimating the association between exposure and outcome. Controlling confounders helps to get unbiased estimates of the exposure-outcome relationship[9].

Since the use of change in estimate is questionable due to its low ability to enhance the precision of treatment effect estimate[121] ,the three significance testing approaches such as common cause, outcome cause, and all pre-treatment covariate approaches are proposed. In addition, confounders that influence the causal effect of place of birth on age-specific childhood vaccination and the level of its effect after controlling confounders has not been documented.

Hence, this study was conducted to identify confounders that result spurious associations between place of birth (exposure) and age-specific childhood vaccination status (outcome). In addition, the causal effect of an exposure on the outcome was estimated by controlling confounders using regression and inverse probability treatment weighting (IPTW). Plasmode simulation[119] was done to see the reproducibility of the identifying method and the effect of exposure on the outcome with such real data. This section is presented based on Iyassu AS et al. [120] presented in the Appendix .

3.2.2. Methods

Description of variables

In this study, the exposure variable is child's place of birth and the outcome variable is age-specific childhood vaccination status. Pre-exposure covariates such as mothers' characteristics, child characteristics, and household characteristics are considered as covariates that possibly confound the association between the exposure and the outcome. The possible relationship between covariates, exposure and outcome is presented in the DAG given in Figure 3.2. DAG is a visual representation of the assumed causal mechanisms and can help to identify covariates to adjust for confounding and control confounding bias (Figure 3.2). The arrows in the DAG show the direction of causal relationships.

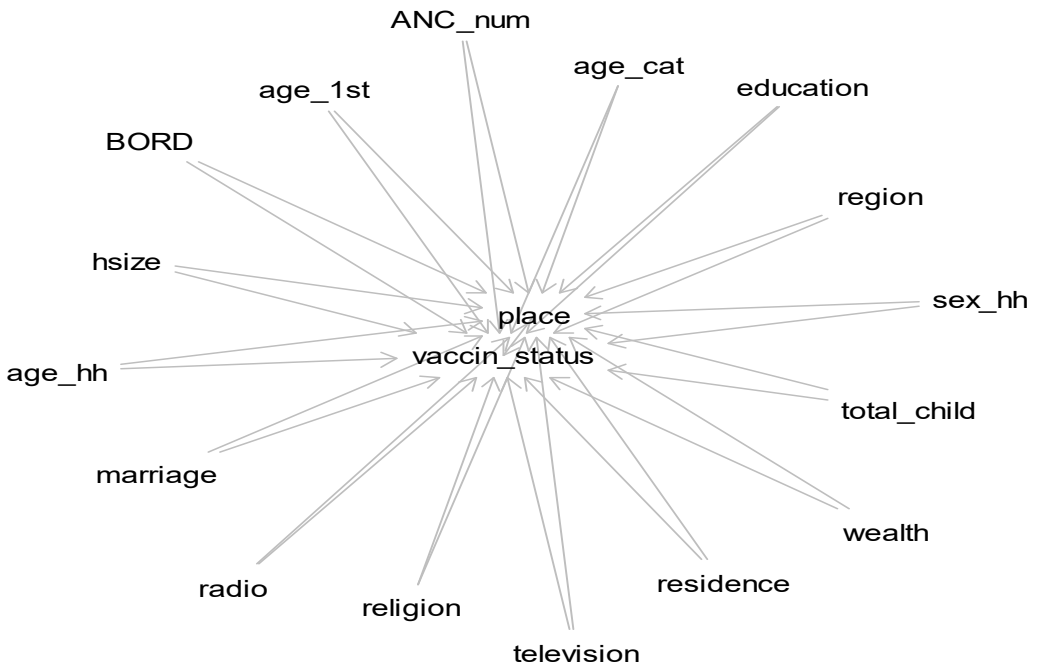


Figure 3.2**.** DAG for covariates, exposure and outcome.

Notation and assumptions

Let A be the binary exposure with values of 0 and 1, $\boldsymbol{X}$ be a P dimensional possible confounders and $Y(a)$ is the potential outcome associated with exposure, $A = a$. The following assumptions should be kept for the unbiased causal effect of the treatment on the outcome.

Unconfoundness assumption: This is also called ignorability assumption where the treatment assignment is independent of the potential outcome given the set of measured covariates: $Y(a) \perp A|\boldsymbol{X}$ [99]. When propensity score is used as covariate adjustment rather than conditioning on them, $\boldsymbol{X}$ can be replaced by propensity score denoted by $e(x)$.

Positivity assumption: $0 < P(A = a|\boldsymbol{X}) < 1, \forall a \in 0,1$, this is an overlap or common support assumption that requires every study participant has a chance to be in any of the treatment conditions. The probability of treatment assignment for any participant is neither zero nor one under treatment conditions [18].

Stable unit treatment value assumption (SUTVA): the potential outcomes of i^{th} subject are not influenced by the potential outcome of j^{th} subject for $i \neq j$, and each unit receives the same version of the treatment[1]

Potential outcome framework and treatment effect

For causal inference under the potential outcome framework [124], the potential outcome is the possible outcome under different treatment conditions. For binary treatment, such as place of delivery,$A = \{0,1\}$, a subject has two potential outcome, $Y(1)$ for $A = 1$ (when a subject received treatment) and $Y(0)$ when $A = 0$ (when a subject does not receive treatment). The outcome is observed only under one, and not under both treatment conditions. If a subject receives treatment level $A = 1$, then $Y\ (1)$ is observed but $Y\ (0)$ is unobserved; if a subject receives treatment level $A = 0$, then $Y(0)$ is observed but $Y(1)$ is unobserved. Under the potential outcome framework, the observed outcome is written as:

$$Y = \ (1 - A) * Y\ (0) + A * Y\ (1). \qquad (3.15)$$

The treatment effect at individual level is $Y_i(1) - Y_i(0)$.

Because it is impossible to calculate change of individual-level treatment effect, we need to estimate average or population-level treatment effects change under the assumption of unconfoundness or exchangeability. Thus, the average treatment effect between treated and untreated subjects is denoted and given by: $\Delta ATE = E\big(Y(1)\big) - E\big(Y(0)\big)$.

In the presence of measured baseline covariates that could result spurious relationship between treatment and outcome, two approaches are used to control confounding effects: the first approach uses ordinal regression model conditional on treatment and confounders, and the second approach uses inverse probability treatment weighting which is called marginal structural modeling.

Let $Y_j, j = 0,1,2$ be the status of age-specific childhood vaccination, A be place of delivery (0 for home delivery and 1 for institutional delivery), then the proportional odds model is given by [112, 125]:

For regression approach: $\mathrm{g}(\mathrm{E}(\mathrm{p}(\mathrm{Y} \leq \mathrm{j/a}, \mathbf{X}))) = \alpha_j - (\beta_1 a + \boldsymbol{\beta}_2{}^T \boldsymbol{X})$. (3.16)

where $\mathrm{g}(.)$ is the link function of the proportional odds model, $\boldsymbol{\beta}_2$ is a p-dimensional vector of parameters and $\boldsymbol{X}$ is a p-dimensional matrix of confounders.

$ATE = \mathrm{g}(\mathrm{E}(\mathrm{p}(\mathrm{Y} \leq \mathrm{j}/1, \mathbf{X}))) - \mathrm{g}(\mathrm{E}(\mathrm{p}(\mathrm{Y} \leq \mathrm{j}/0, \mathbf{X})))$ in terms of risk difference.

$ATE = \mathrm{g}(\mathrm{E}(\mathrm{p}(\mathrm{Y} \leq \mathrm{j}/1, \mathbf{X})))/\mathrm{g}(\mathrm{E}(\mathrm{p}(\mathrm{Y} \leq \mathrm{j}/0, \mathbf{X})))$ in terms of risk ratio and can be extended to odds ratio.

For marginal structural modeling, the weighted proportional odds model is:

$$\mathrm{g}(\mathrm{E}(\mathrm{p}(\mathrm{Y} \leq \mathrm{j/a}))) = \alpha_j - \beta_1 a. \quad (3.17)$$

In this case, each value of the subject is weighted by IPTW estimated by using WeightIt R function [126] and covariate balancing was checked using cobalt R package [127].

Identification of confounders

Different criteria are used to select cofounders. The first criterion is pre-exposure criteria [128]. In this criterion, one can consider and control all covariates that come prior to exposure under study. The second method is significance testing criteria (ST) with cutoff P-values fixed at less than or equal to 0.2. In this method, two approaches are used: common cause[118] that states a covariate is a confounder when it causes the exposure and outcome; outcome cause that states one can consider a covariate to be a confounder when it is a cause of the outcome. The outcome cause criteria may contain covariates associated with both exposure and outcome and covariates associated the outcome but unrelated to the exposure [99]. Using the notation used by Shortreed & Ertefaie [99]), let $\boldsymbol{X}_\ell$ be a set of all pre-treatment covariates, $\boldsymbol{X}_p$ be a set of common cause covariates, and $\boldsymbol{X}_q$ be a set of outcome cause covariates. These three different covariates were included in equation (3.10) while using the regression method to adjust confounders and estimate treatment effect. While using equation (3.11) to estimate the treatment effect, we used propensity score to control confounders estimated after fitting the generalized linear model given as follows:

$$g(E(P(A = 1/\boldsymbol{X}; \hat{\theta}))) = \hat{\theta}_0 + \widehat{\boldsymbol{\theta}}^T \boldsymbol{X}. \quad (3.18)$$

$g(.)$ is the link function for binary response such as logit link functions, and $\mathbf{X}$ is one of $\boldsymbol{X}_\ell$, $\boldsymbol{X}_p$ and $\boldsymbol{X}_q$, and $\widehat{\boldsymbol{\theta}}$ is a vector of estimated coefficients**.**

Plasmode simulation

Plasmode simulation starts resampling of exposure and covariates from the original data. Then the outcome data is generated from resampled exposure and covariates. The parameters or effect sizes in the simulation process are estimated from the original data by modeling the relationship of outcome with exposure and covariates. Sometimes the parameters can be defined by the user. Plasmode simulation is viewed as semi-parametric simulation because exposure and covariates are generated naturally from unknown parametric distribution whereas; the outcome is simulated from known distributions and resampled exposure and covariate[71, 119]. The approach depends on resampling from the practical exposure and covariates data without modification in all simulated datasets to preserve the relations among these variables and complex data structure. It has also the advantage of a user-specified exposure effect. Moreover, the simulated datasets are used to compare variable selection strategies for confounder adjustment via the propensity score[119].
The plasmode data are simulated [71], with the following procedures:

1. Exposure and covariates structures are generated by resampling from an original data
2. Generating the outcome that incorporates
   2.1 Determining outcome generating model: linear regression with normal distribution is used and categorized into ordered categories
   2.2 Determining exposure and covariate effect by estimation from original data and individual specification
   2.3 Generating outcome data using chosen outcome model, effects, and resampled covariates and exposure.

The challenges of plasmode simulation are specifying the number of plasmode data sets (N), resampling techniques (Resampling without replacement, m out of n, and resampling with replacement, n out of n), and resampling size (m). So far there is no defined rule to choose the sampling techniques, to determine simulation and resample sizes[71]. In this study, the size of the simulated data set or the number of repetitions (N) is specified to be 500 as used by [71, 119, 129, 130]. Sampling without replacement technique is implemented. The m out of n resampling without replacement also called subsampling requires fewer assumptions and prevents bootstrap failure[71, 131]. In addition, the size of the resample (m) is 1000.

Comparison of methods

To compare the performance of methods from simulated data, we used different multi-dimensional performance measurement metrics illustrated by [68]. These include bias, mean square error, empirical standard error, average model standard error, and bias eliminated coverage. For each performance measurement metrics estimate, Monte Carlo standard error is also estimated. The estimates of each performance measures from simulation study is generated using simsum R function from rsimsum package[132, 133] The estimand of the study is treatment effect on the outcome.

3.2.3. Result

Result of plasmode simulation

To make sure plasmode simulation generates realistic data that resemble the observed data, we compared the distribution of the simulated and observed data. We simulated 500 artificial datasets for the outcome each containing 1000 observation with the proportion of childhood vaccination status matched with the observed data. In Figure 3.3, we presented the bar chart for the distribution of each category of the outcome. It shows the proportions of each category are similar between simulated and observed data sets.

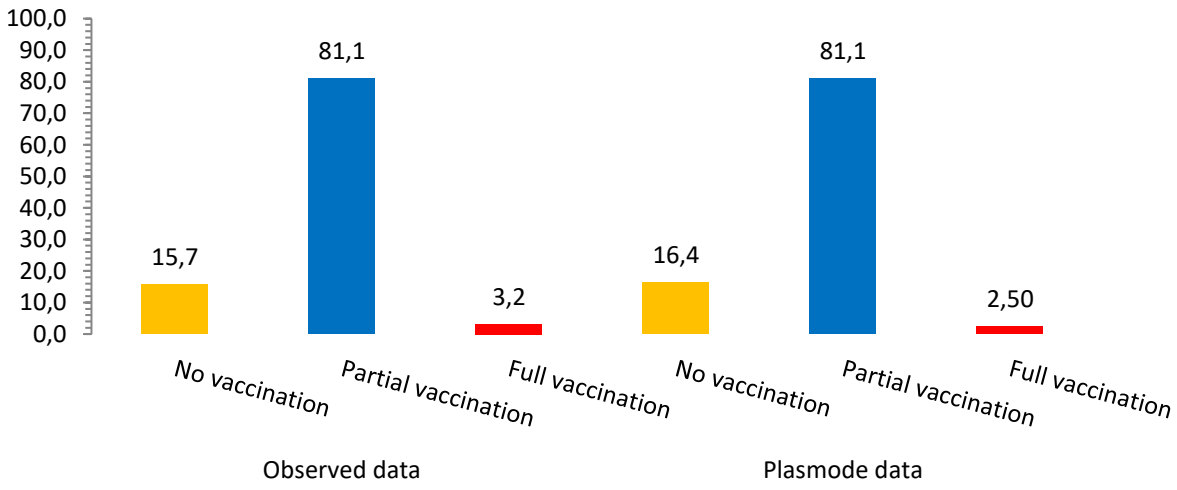


Figure 3.3. Distribution of the outcome from plasmode and observed data.

Evaluation of confounder selection strategy

We used simulated data for the covariate selection strategy to be included in propensity score function or to control with the regression method. Figure 3.4 is the histogram of treatment effect

estimator distribution from simulation study. The distribution of the estimator for each confounder selection techniques indicates, the estimator is nearly symmetrical. Thus, we are able to estimate other performance methods assuming the estimator normally distributed.

Figures 3.5 and 3.6 also show the result of treatment effect for the three proposed covariate selection approaches using simulated data. Figure 3.5 presents the treatment effect before adjusting confounders, outcome cause (out.cause), common cause (com.cause), and all pre-treatment covariates (All.covariates) adjusted by regression. The distribution of crude effects is far from the three approaches. However, the distribution of treatment effect is similar for the three approaches. Figure 3.6 also shows similar result after adjusting with propensity score-based IPTW.

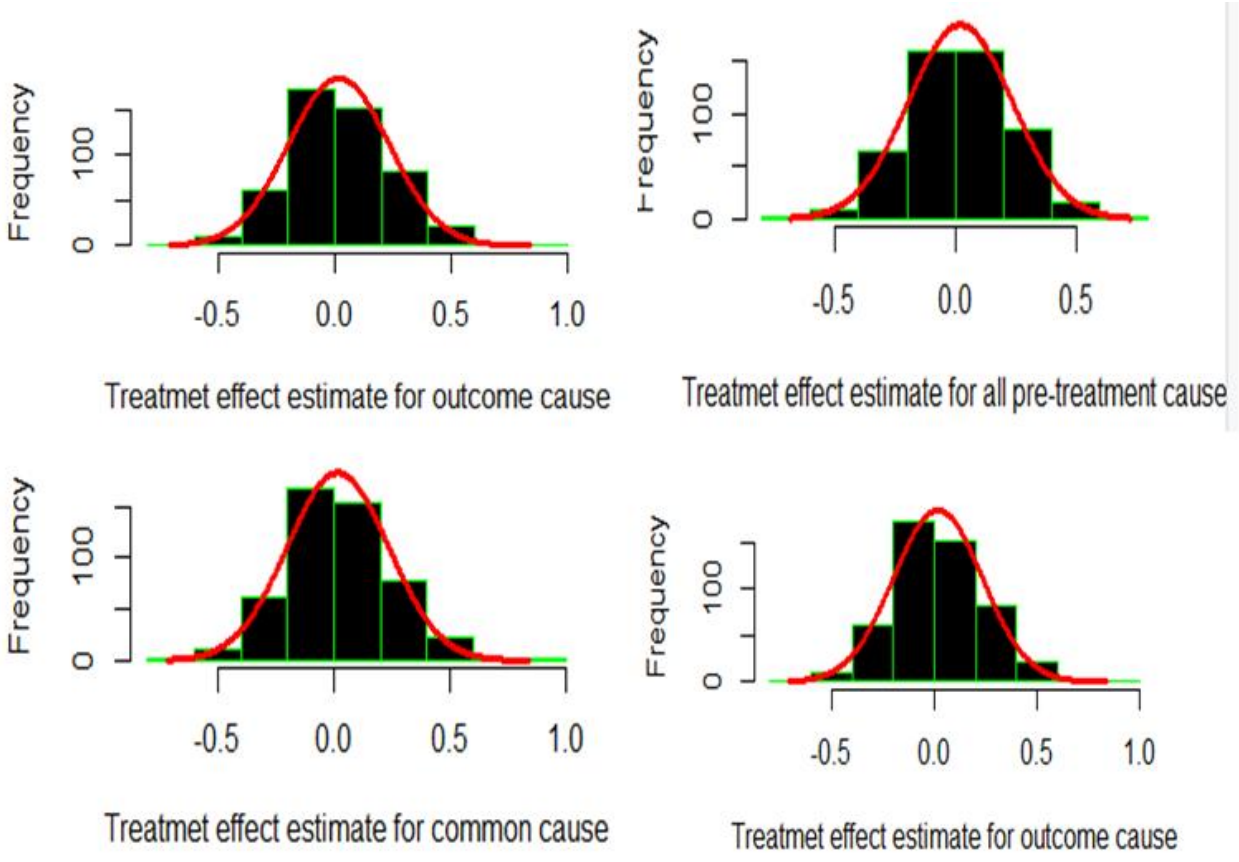


Figure 3.4**.** Density plot of treatment effect estimate from plasmode simulation

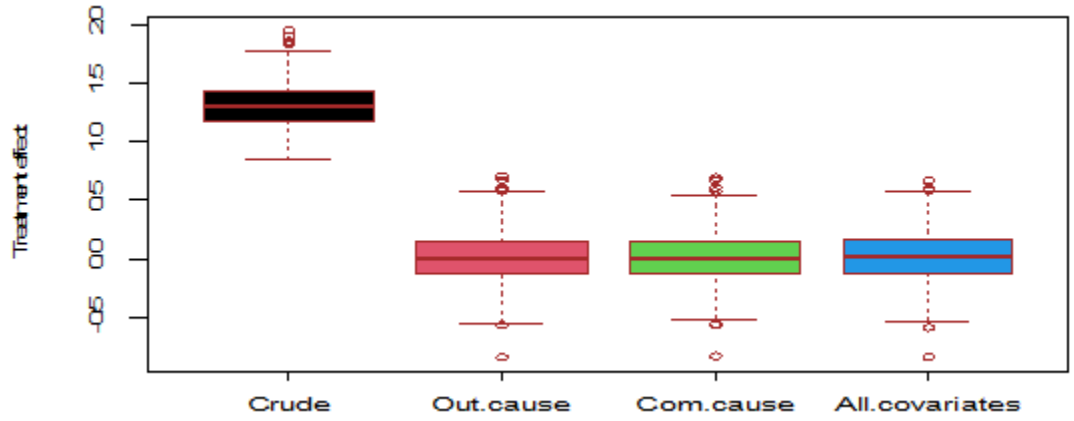


Figure 3.5. Treatment effect for plasmode data using regression method.

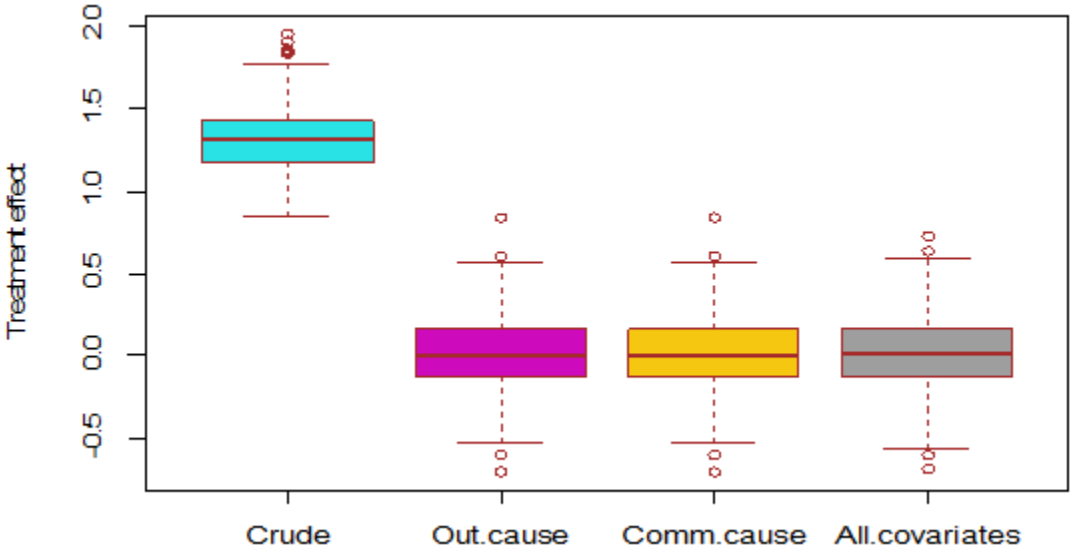


Figure 3.6. Treatment effect from plasmode data using IPTW.

Table 3.6 demonstrates the estimate of method performance metrics for simulation study using IPTW from 500 simulated data sets for each confounder selection approach. In the generation of artificial outcomes from real covariates, we fixed the true treatment effect at 0.5 for all approaches and confounder adjustment methods. The average treatment effect after adjusting confounders was almost similar for all approaches with negligible difference between outcome and common cause. The true treatment effect was highly reduced when taking confounders into account and adjusting with IPTW.

The absolute value of bias for all-pretreatment covariates was higher than outcome and common cause covariates. Its uncertainty is also higher than the other two approaches. The absolute difference of bias between outcome and common cause covariates is 0.001 with approximately equal uncertainty values. The bias eliminated coverage of outcome and common cause covariates was 88.8% and that of all pre-treatment covariates is 89.2% with equal Monte Carlo standard error for all approaches. The empirical standard error which is the square root of the variance of an estimator and measure of the efficiency of an estimator are equal in outcome cause and all-pre-treatment covariates. The empirical standard error of common cause is little higher than other approaches. On the other hand, the model based standard error which is the average of the standard error of estimators from 500 simulation data is equal in outcome and common cause confounder selection approaches. However, its value is little higher in all pre-treatment confounder selection approach than outcome and common cause confounder selection approaches. The magnitude of relative percent error in standard error in common cause is higher than the other two approaches. The difference of power of test at 5% between outcome and common cause confounder selection approaches is very small (difference=0.002). Considering all method performance measures and taking the aggregates, outcome cause confounder identification approach outperformed common cause and all pre-treatment covariates approach.

Table 3.6. Estimates of performance measures for confounder identification approaches.

Performance measure	Outcome cause	Common cause	All pre-treatment
Average treatment effect estimate	0.019	0.018	0.022
Bias	-0.481 (0.01)	-0.482 (0.01)	-0.978 (0.02)
Coverage	0.236(0.019)	0.250 (0.019)	0.22 (0.002)
Bias eliminated coverage	0.888 (0.014)	0.888 (0.014)	0.892 (0.014)
Empirical standard error	0.216(0.007)	0.218(0.007)	0.216 (0.007)
MSE	0.2784 (0.010)	0.2794 (0.010)	1.003 (0.019)
Model based standard error	0.167 (0.003)	0.167 (0.003)	0.168 (0.003)
Relative % error in standard error	-22.83 (2.45)	-23.48(2.43)	-22.20 (2.47)
Power of 5% test	0.114 (0.014)	0.116 (0.014)	0.108 (0.014)

Table 3.7 also shows the estimate of average treatment effect and performance measurement metrics for confounder identification approaches when confounders are adjusted using regression. The bias of treatment effect for outcome cause covariates is little smaller than the other two approaches (Table 3.7). The bias eliminated coverage of outcome cause covariates somehow more than the other two approaches. Similarly, the estimator in outcome cause covariates was more efficient than all pre-treatment covariates with smaller values of empirical standard error and model based standard error.

Table 3.7. Estimates of performance measures for confounder identification approaches

Performance measure	Outcome cause	Common cause	All pre-treatment
Average treatment effect estimate	0.020	0.021	0.021
Bias	-0.780 (0.009)	-0.791 (0.009)	-0.790 (0.010)
Coverage	0.040 (0.009)	0.040 (0.009)	0.050 (0.010)
Bias eliminated coverage	0.960 (0.009)	0.958 (0.009)	0.954 (0.009)
Empirical standard error	0.211 (0.007)	0.211 (0.007)	0.217 (0.007)
MSE	0.652(0.015)	0.653 (0.015)	0.654 (0.015)
Model based standard error	0.209(0.0003)	0.208 (0.003)	0.213 (0.003)
Relative % error in standard error	-0.847 (3.14)	-1.399 (3.12)	-1.72 (3.11)
Power of 5% test	0.042 (0.009)	0.046 (0.009)	0.052 (0.010)

The result when confounders are adjusted using regression method implies that outcome cause covariates performed better than the common cause and all pre-treatment covariates. On the other hand, the bias and mean square error of adjusting confounders with IPTW are smaller than adjusting confounders with regression method.

Result from the actual data

Evaluation of confounder selection strategies

In order to estimate the causal effect of place of birth on age-specific childhood vaccination, we first compared the link functions of cumulative link model for marginal structural modeling

(MSM), and multiple regression of the outcome on the treatment and covariates. We compared the performance of link functions using AIC and BIC model selection criteria. The values of AIC and BIC are presented in Table 3.8. The logit and probit link functions have equivalent AIC and BIC values in MSM. The negative log-log link function outperformed in the multiple regressions followed by logit link for all confounder identification techniques. For the purpose of consistency in treatment effect estimation between MSM and multiple regression, we choose logit link function for both effect estimating statistical methods.

Table 3.8. Selection of link function for the outcome model.

Methods	Link function	IPTW AIC	BIC	Regression AIC	BIC
All pretreatment	Logit	5205	5225	5078	5353
	Loglog	5212	5231	5004	5279
	Cloglog	5223	5242	5433	5708
	Probit	5205	5224	5125	5400
	Cauchit	5217	5237	5037	5312
Outcome cause	Logit	5278	5298	5116	5306
	Loglog	5287	5307	5072	5262
	Cloglog	5289	5309	5439	5628
	Probit	5276	5296	5158	5348
	cauchit	5294	5314	5097	5286
Common cause	Logit	5271	5291	5116	5292
	Loglog	5281	5301	5072	5248
	Cloglog	5282	5302	5433	5708
	Probit	5270	5289	5158	5335
	cauchit	5288	5308	5096	5272

After fixing the outcome model link function, we estimated the effect size of the treatment on the outcome for all confounder identification techniques using MSM and regression. Table 3.9 contains the effect of treatment (place of birth) on outcome (age-specific childhood vaccination) along with their standard errors for unadjusted, outcome cause, common cause, and all pre-treatment covariates adjusted by regression and IPTW. Before using IPTW for MSM, we checked covariate balance (using absolute standardized mean difference) and positivity of propensity score for the three confounder selection approaches. The covariates are balanced and propensity score values range from 0.021 to 0.999.

Unadjusted treatment effect on the log odds of the cumulative probability of the outcome is 1.3. When adjusted for outcome cause, common cause, and all pre-treatment covariates using the

cumulative link model, the effect dropped down to 0.53, 0.54, and 0.51 respectively. When using IPTW, the effect dropped down to 0.36, 0.37 and 0.46 respectively. To compare confounder identifying approaches, almost all brought the same reduction of crude effect in the regression technique although it looks like all pre-treatment gave the largest reduction with higher standard error. When using IPTW, the outcome cause approach brought the largest reduction with relatively the smallest standard error. On the other hand, IPTW reduced the unadjusted treatment effect with a smaller standard error and narrower confidence interval as compared to the regression method. This shows that the IPTW confounder adjustment method is preferable to the regression method.

Based on outcome cause confounder selection approach, number of antenatal care services, age of household head, age at first birth, household size, total number of children ever born, birth order number, region, place of residence, religion, mother's education status, ownership of television and radio, and household wealth status were the confounders for the causal effect of place of delivery on age-specific childhood vaccination.

Table 3.9. Effect of place of birth on age-specific childhood vaccination.

Confounder adjustment method	Confounder selection approach	Average treatment effect	Standard error of treatment effect	95% CI	
Regression	Unadjusted treatment effect	1.3	0.081	1.16	1.5
	Outcome cause	0.53	0.099	0.34	0.73
	Common cause	0.54	0.099	0.34	0.73
	All pre-treatment covariates	0.51	0.101	0.32	0.71
IPTW	Outcome cause	0.36	0.072	0.23	0.50
	Common cause	0.37	0.073	0.23	0.52
	All pre-treatment covariates	0.46	0.075	0.31	0.60

Effect of place of birth on age-specific childhood vaccination

After identifying and adjusting confounders, the next step was estimating the treatment effect on the outcome. Choosing outcome cause approach for confounder selection and using the IPTW confounder adjustment method, the log odds of institutional delivery versus home delivery is given as follows:

$$logit(P(y \leq j|A = a) = \alpha_j - 0.36a,\ a = 0,1. \quad (3.19)$$

Then the odds ratio of the event, $y \leq j\ is, OR = exp^{-0.36} = 0.70$ and that of $Y > j$ is , $OR = exp^{0.36} = 1.43$. This implies that institutional delivery decreases the lower level of vaccination status by 30%. On the other hand, institutional delivery increases the likelihood of higher level of vaccination status (partial from no vaccination and full vaccination from partial vaccination) by 43%.

3.2.4. Discussion and Conclusion

This study was done to identify confounders and estimate the causal effect of place of birth on age-specific childhood vaccination. In the process of estimating exposure's causal effect on the outcome, we have to be curious about extraneous/cofounding covariates that could alter the effect of the exposure on the outcome. Most studies so far focused on confounder identification and treatment effect estimation when the outcome is continuous or binary. However, to the best of our knowledge, no literature that dealt with when the outcome is ordinal, especially the causal effect of place of birth on age-specific childhood vaccination.

The study used regression and propensity score-based inverse probability treatment weighting to control confounding bias. A common approach is to control as many covariates as possible that are observed before the treatment [5]. However, including all pre-treatment covariates in any confounder controlling methods such as regression introduces bias. Adding more covariates to the model causes over fitting and unstable coefficients due to multicollinearity [10]. A model is best when it contains the smallest number of covariates that explain the greatest amount of variance[11]. To prove this assertion, we proposed three approaches of confounder selection to control their effect. These are all pre-treatment covariates, the common cause of treatment and outcome covariates (real confounders) and outcome cause covariates. As described in [99, 100], the outcome cause covariates included real cofounders and predictors of outcome variable.

To show the reproducibility of our confounder selection approach, we used plasmode simulation as used in [71, 119]. A key advantage of statistical plasmode simulations is their ability to maintain the intricate structure of real-world data by resampling covariate information from an actual dataset. For these simulations to be effective, it is essential to have a suitable representative dataset, which forms the foundation for the entire plasmode simulation study[71]. We resampled 1000 data sets without replacement 500 times out of 5150 total data sets. A latent continuous outcome was

simulated from a known covariate structure and manually added true treatment effect (to be constant across confounder selection approaches). The effect of each covariate is determined from the actual relationship of observed data. The performance of each confounder selection approaches is compared using multi-dimensional performance metrics as described by Morris et.al. [68] using IPTW and regression based confounder adjusting.

The result from plasmode data set showed that there were no notable differences across confounder selection approaches but it looks outcome cause approach gives better results. On the other hand, common cause covariates are the subset of outcome cause covariates. Hence, we used outcome cause covariates to estimate the causal effect of place of delivery on timely childhood vaccination. Even though the regression and IPTW gave comparable results, it seems IPTW can give better results for confounder adjustment. However, in the plasmode data, there was a huge difference between the true treatment effect and the treatment effect after covariate adjustment which is too far from the treatment effect from observed data. One of the challenges to generating artificial ordinal outcomes is that there is no direct method to simulate it. We rather simulated continuous latent outcome and then cut in to ordered categories based on the quintiles of observed data. In this process, there might be loss of information, and the treatment effect in observed and simulated data could deviate.

The result from the observed data set, outcome cause gave relatively better result as compared to other approaches. This is consistent with the findings in the literature [99, 100, 119]. Finally, after identifying confounders using outcome cause approach and adjusting them using IPTW, the causal effect of place of birth on age-specific childhood vaccination is estimated. The result shows institutional delivery enhances the timely vaccination of children.

To conclude, in the causal inference, the identification of confounders is an important step before attempting to estimate exposure's effect on outcome. It is important to identify confounders and adjust them with regression, weighting, or other matching methods. In the study, outcome cause covariates (common causes and predictors of the outcome) results relatively better performance than other proposed methods in terms of treatment effect. Propensity score based IPTW helps to see the balanced confounders between treatment and control groups. It also gave better treatment effect as compared to regression adjustment of covariates as demonstrated in this study.

Institutional delivery enhances the likelihood of a newborn baby to get vaccine as per the standard schedule.

The limitation of this study is that it only uses the IPTW estimated from logistics regression as the treatment variable is binary. It does not account various weighting methods including machine learning (GBM and super learner), CBPS and entropy balancing. The next section deals with comparison of various weighting methods to control confounding bias and estimate treatment effect on the outcome.

3.3 Causal Effect of Newborn Postnatal care on Age-specific Childhood Vaccination

3.3.1 Background

Enhancing child health has been at the forefront of global public health initiatives for the past four decades. Reducing child mortality has been a fundamental focus of the global health agenda, as outlined in the Millennium Development Goals (MDGs) and the Sustainable Development Goals (SDGs)[134]. Vaccinations have been demonstrated to significantly enhance child growth by decreasing instances of stunting and underweight conditions[135]. Immunization not only saves the lives of infants but also enhances productivity, educational outcomes, and economic growth[136, 137].

Postnatal care refers to the care provided to a mother and her newborn baby following childbirth, extending up to 42 days after delivery. The postnatal period encompasses the days and weeks after childbirth and is a crucial time for both mothers and newborns. During the first six weeks, every mother and baby should have at least four postnatal checkups. These visits occur on the first day (within 24 hours), on day three (48–72 hours), between days seven and fourteen, and at six weeks (42 days)[138]. Most maternal and neonatal deaths occur during childbirth and the postnatal period. It is estimated that 75% of all neonatal deaths happen within the first week of life, and one million newborns dying within the first 24 hours after childbirth[139].

Previous studies showed that postnatal care (PNC) was a predictor of complete childhood vaccination [140-144]. These studies identified PNC as a predictor of full vaccination: 1) by using binary logistic regression where vaccination was classified as full and incomplete after merging no vaccination and partial vaccination as one category. This can result biased estimates such that

missing one vaccine is equivalent to no vaccination where the risk of morbidity is different in both cases. In addition, the classification was done regardless of the schedule of vaccination where the vaccination is given at any age until 23 months. 2) The childhood vaccination status was regressed on individual, household and community level predictors. In this scenario, the effect of PNC could be observed by adjusting other factors using regression which does not tell that how much of the other predictors are matched in full and incomplete vaccinated children.

The parametric and machine learning weighting methods were compared in terms of generating optimal causal treatment effect. The comparison is done using plasmode (hybrid) simulation and parametric (artificial) simulation data.

3.3.2 Description of Variables

Outcome variable

Like the previous sections, the outcome variable is age-specific childhood vaccination status which had three ordered categories (No vaccination, partially vaccinated and fully vaccinated).

Treatment variable

The treatment variable is newborn PNC. It is binary with value of 1 when a newborn has PNC within 42 days of birth, and 0 otherwise.

Covariates

Individual, household and community level covariates which came before the treatment and have possible mix up effect on the outcome variable were identified from literature and based on their availability are presented in the DAG (Figure 3.7). DAG is built to portray prior knowledge related to specific causal relationships. DAGs can pinpoint variables that, when controlled for during the design or analysis phases, can effectively eliminate confounding and certain types of selection bias. DAGs consist of variables (or nodes), such as treatments, health outcomes, or covariates, and arrows (or edges) that illustrate known or suspected causal relationships between these variables [145].

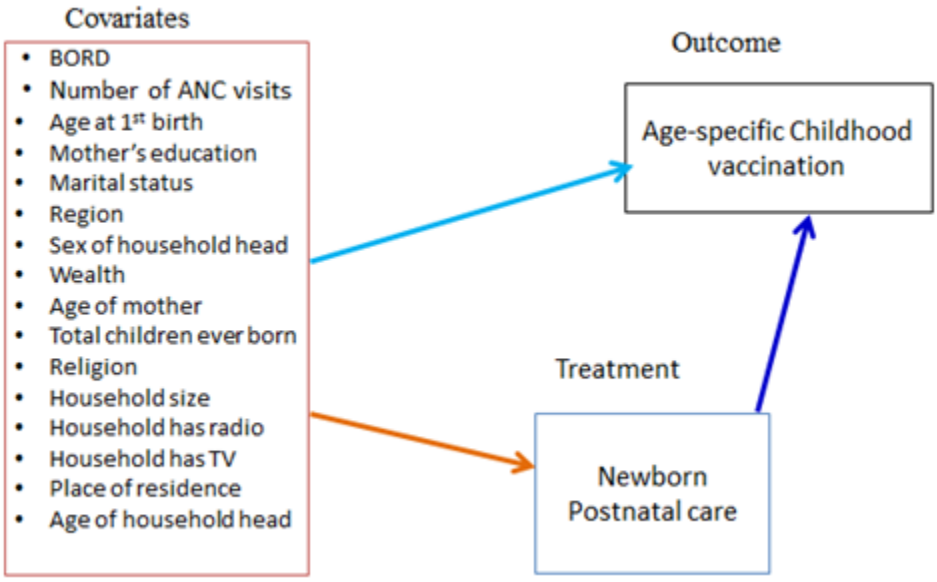


Figure 3.7. DAG for covariates-exposure and covariates-outcome relationship.

3.3.3. Potential Outcome Framework and Estimating Treatment Effect

Let $n = n_1 + n_0$ samples are drawn from population of size N= $N_1 + N_0$, where N_1 is population size of treatment group, and N_0 population size of control group for binary treatment T_i taking values of 1 when treatment is assigned to i^{th} individual (treatment group) and values of 0 when treatment is not assigned to i^{th} individual (control group). In the sample, we have n_1 treated individuals and n_0 control individuals. Let $\boldsymbol{X}$ be a matrix of P covariates; $\boldsymbol{X_{ip}}$ is the value of p^{th} covariates for i^{th} individual so that $\boldsymbol{X_i} = [X_{i1}, X_{i2}, X_{i3}, \ldots, X_{ip}]$ is the row vector of covariates for i^{th} individual. Finally, we have a pair of potential outcomes for binary treatment T_i such that $Y_i(1)$ is the potential outcome of i^{th} individual when it is exposed to the treatment and $Y_i(0)$ is the potential outcome of i^{th} individual when it is not exposed to the treatment. For each individual the treatment effect is $Y_i(1) - Y_i(0)$. However, we cannot observe both $Y_i(1)$ and $Y_i(0)$ simultaneously for a single individual. Rather we observe $T_i, Y_i, and\ \boldsymbol{X_i}$ simultaneously. Thus, the observed outcome for each individual i is:

$$Y_i = Y_i(1)T_i + (1 - T_i)Y_i(0). \qquad (3.20)$$

Because the individual level treatment effect is not identifiable, we use the population level Average Treatment Effect (ATE) which is defined as [59, 146, 147] :

$$ATE = E[(Y(1) - Y(0)]. \quad (3.21)$$

Similarly, the Average Treatment on the Treated (ATT) is defined as:

$$ATT = E[Y(1) - Y(0)|T = 1]. \quad (3.22)$$

The treatment effect may be heterogeneous if it affects the outcome for individuals differently due to different distribution of covariate between treatment and control groups such as difference in sex, age, economic status etc. In such conditions, we need to find the conditional average treatment effect defined as[59]:

$$CATE = E[Y(1) - Y(0)|\boldsymbol{X} = \boldsymbol{x}]. \quad (3.23)$$

and conditional average treatment effect on the treated defined as:

$$CATT = E[Y(1) - Y(0)|T = 1, \boldsymbol{X} = \boldsymbol{x}]. \quad (3.24)$$

3.3.4. Adjusting Confounders Using Inverse Probability Treatment Weight

Confounders are the subset of covariates that have mix up effect with the treatment variable. They will result spurious relationship between treatment and outcome unless they are controlled by design such as randomization or statistical methods. Adjusting for confounders is critical to accurately estimate the treatment effect on the outcome. In observational study, observed confounders can be adjusted by regression and propensity score methods [59].

Outcome regression method is used to estimate conditional effect of the treatment accounting for covariate imbalance between treatment arms [148, 149]. However, standard adjustment by parametric regression models is sensitive to model misspecification[150]. Moreover, in regression adjustment, researchers often struggle to determine if the adjusted effect relies on extrapolation. If the distributions of multivariate covariates among participants in different treatment conditions are significantly different, the overlap in multidimensional space may be limited. This means there may be no treated participants who are comparable to untreated participants across all observed covariates, making any comparisons between the two groups potentially reliant on extrapolation[60].

An alternative to regression adjustment is to use inverse probability of treatment weights (IPTWs) to address biases caused by observed confounders. IPTWs are used to construct a pseudo-population where confounders and treatment assignment are independent, a characteristic we

would anticipate under randomization. In this study, we formed IPTWs in different ways including propensity score, generalized boosted modeling, covariate balancing propensity score, entropy balancing, and neural network to compare their performance.

Propensity score based IPTW

Propensity score is widely used method to adjust confounding in treatment effect estimation. It leverages the balancing property of the propensity score, which ensures that the distribution of confounders is identical between treatment and control groups when conditioned on the propensity score[18, 147]. IPTW is constructed by first estimating each individual's probability, the so called propensity score, of receiving each respective treatment value conditional on observed confounders, and taking the reciprocal of the estimated probability [60]. That is, for average treatment effect, each individual receives weights of $\frac{T}{P(T=1|\boldsymbol{X})}+\frac{1-T}{1-P(T=1|\boldsymbol{X})}$ for $T=0,1$. These weights are not stabilized. However, when the probability of receiving the treatment level is unlikely or gets to zero, the weights will be large which often increases the variance of the treatment effect estimate. To solve the large weight problem, we can use stabilized weights that can be constructed by the base line probability of selecting treatment to the probability of selecting the treatment given observed confounders[60].

That is the stabilized IPTW $=\frac{T*P(T=1)}{P(T=1|\boldsymbol{X})}+\frac{(1-T)(1-P(T=1)}{1-P(T=1|\boldsymbol{X})}$ for $T=0,1.$ (3.25)

Generalized Boosted Model based IPTW

GBM fits a piecewise constant model to predict treatment indicator, using multiple regression trees combined iteratively. Starting with one tree, each new tree is added to best fit the residuals from the previous iteration and maximize log likelihood. Predictions from each tree are scaled down to enhance model smoothness. While more iterations increase flexibility, they can lead to over fitting. To prevent this, GBM selects an optimal number of trees to minimize external criteria, such as out-of-sample prediction error or balance in covariates between treatment and control groups[25, 151].

GBM can be used to estimate weights of binary treatments with the "*optimal iteration (number of trees) for estimating the propensity scores set to be the one that minimizes a stopping rule criterion*

based on the difference between the weighted distributions of the pretreatment covariates in the two treatment conditions"[25].

Let $\boldsymbol{X}$ is a p-dimensional pretreatment confounders, T is a treatment indicator, and $e(\boldsymbol{x})$ is propensity score. Then the GBM algorithm is illustrated as follows according to [151]. The expected log-likelihood function of the Bernoulli distribution of binary treatment is:

$$E\big(\ell(p)\big) = E(t * log\big(e(\boldsymbol{x})\big) + (1-t)\log([1-e(\boldsymbol{x})]|\boldsymbol{X}). \quad (3.26)$$

The expectation is with respect to all future participants that might experience the same treatment assignment procedure. The conventional method uses the logistic transform of $e(\boldsymbol{x})$ to simplify the analysis.

$$e(\boldsymbol{x}) = \frac{1}{1+\exp[-g(\boldsymbol{x})]}. \quad (3.27)$$

If we substitute equation (3.27) to equation (3.26), the log-likelihood function in terms of $g(\boldsymbol{x})$ and $\ell(g)$ is :

$$E\big(\ell(g(x)\big) = E(t * g(\boldsymbol{x}) - \log\{[1+\exp[g(\boldsymbol{x})]\}|\boldsymbol{X}). \quad (3.28)$$

Let's assume that we have an initial estimate that maximizes equation (3.18) that we call $\hat{g}(\boldsymbol{x})$. The commonly sample log-odds $\hat{g}(\boldsymbol{x}) = \log(\frac{\bar{y}}{1-\bar{y}})$ offers the initial estimate. We want to advance this initial estimate by including a small adjustment to it. That is, we want to find an $h(x)$ such that:

$$E\big(\ell(\hat{g}+\lambda h)\big) > E\big(\ell(\hat{g})\big). \quad (3.29)$$

This new upgrading provides an increase in the expected log-likelihood and, therefore, we can update our current guess as:

$$\hat{g}(\boldsymbol{x}) \leftarrow \hat{g}(\boldsymbol{x}) + \lambda h(\boldsymbol{x}). \quad (3.30)$$

for some step size $\boldsymbol{\lambda}$**.** The next step is how to find $h(\boldsymbol{x})$ that satisfies equation (3.19).

$$h(\boldsymbol{x}) = \frac{\partial}{\partial g(x)} E(L(g(x))) = E\left(t - \frac{1}{1+exp[-g(x)]}|\boldsymbol{X}\right).$$
$$= E(t - e(\boldsymbol{x})|\boldsymbol{X}). \quad (3.31)$$

Given a regression tree estimate for h, the update expression in Equation (3.30) shows that we simply need to do a line search for the λ that offers the greatest increase in the log-likelihood. The tree is a piecewise constant model. It partitions the participants according to their features into regions, $T_1, T_2, ., ., ., T_K$. Then the optimal adjustment to $\hat{g}(\boldsymbol{x})$ can be obtained separately for each region. That means for all observations that fall into the k^{th} partition, $x \in T_K$.

$$h(\boldsymbol{x}) = \arg max_{\theta} \sum_{x_i \in T_K} t_i \left[\hat{g}(\boldsymbol{x_i}) + \theta\right] - \log\{1 + \exp[\hat{g}(\boldsymbol{x_i}) + \theta]\}.$$

$$\frac{\sum_{x_i \in T_k} t_i - e(\boldsymbol{x_i})}{\sum_{x_i \in T_k} e(\boldsymbol{x_i})[1 - e(\boldsymbol{x_i})]}. \quad (3.32)$$

Where $\arg max_{\theta}$ indicates the value of θ that minimizes the sum. To avoid intensive computation, the estimate in the last line of equation (3.32) is based on maximizing a second-order Taylor approximation of the first line. It is very stable as long as not all estimated probabilities are 0 or 1 in any terminal node. In those cases, we set $h(\boldsymbol{x}) = 0$ for that region.The details of boosting algorithm is presented in [151].

Covariate balancing propensity scores based IPTW

One of the limitations of parametric logistic model to estimate propensity score is that the model may be misspecified. The covariate balancing method introduced by [20]models the treatment while optimizing the covariate balance, offers robustness to minor model misspecification with respect to balancing confounders compared to direct maximum likelihood estimation. Moreover, although the model is correctly specified, the covariate balancing method can improve the covariate balance in observed data sets and potentially improve the accuracy of estimated treatment effects over conventional logistic regression. Alongside maximizing model likelihood, the covariate balancing method includes a balance condition for the weighted means of covariates during parameter estimation. This approach employs a generalized method of moments or an empirical likelihood estimation framework [152] to find estimates that best optimize the likelihood function while simultaneously satisfying the balance condition[153].

Consider a random sample of n observations, a binary treatment variable T_i for each unit i, and a k-dimensional column vectors $\boldsymbol{X}_i$ of observed confounders, then a parametric propensity score model is $\pi_{\beta}(\boldsymbol{X}_i) = P(T_i = 1|\boldsymbol{X}_i)$, with $\boldsymbol{\beta} \epsilon \Theta$ is an L-dimensional column unknown parameters. The most commonly choice model with $L = K$ is the logistic model given by:

$$\pi_{\beta}(\boldsymbol{X}_i) = \frac{\exp(\boldsymbol{X_i}^T \boldsymbol{\beta})}{1 + \exp(\boldsymbol{X_i}^T \boldsymbol{\beta})}. \quad (3.33)$$

The maximum log-likelihood estimate of $\boldsymbol{\beta}$ to predict the treatment assignment using the estimated propensity score is:

$$\widehat{\boldsymbol{\beta}}_{MLE} = \arg\max_{\boldsymbol{\beta}\epsilon\Theta} \sum_{i=1}^{n} T_i \, log\{\pi_\beta(\boldsymbol{X}_i)\} + (1 - T_i)\log\{1 - \pi_\beta(\boldsymbol{X}_i)\}. \quad (3.34)$$

Considering $\pi_\beta(.)$ is continuously differentiable with respect to β such that:

$\frac{\partial}{\partial \boldsymbol{\beta}}\left[T_i log\{\pi_\beta(\boldsymbol{X}_i)\} + (1 - T_i)\log\{1 - \pi_\beta(\boldsymbol{X}_i)\right]$ exists, then the first order score condition is:

$$\frac{1}{n}\sum_{i=1}^{n} S_\beta\,(T_i, \boldsymbol{X}_i) = 0 \;, where \;\; S_\beta(T_i, \boldsymbol{X}_i) = \frac{T_i \pi_\beta{}'(\boldsymbol{X}_i)}{\pi_\beta(\boldsymbol{X}_i)} - \frac{(1-T_i)\pi_\beta{}'(\boldsymbol{X}_i)}{1-\pi_\beta(\boldsymbol{X}_i)} \quad (3.35)$$

$$\text{and } \pi_\beta{}'(\boldsymbol{X}_i) = \frac{\partial \pi_\beta(\boldsymbol{X}_i)}{\partial \boldsymbol{\beta}}.$$

Equation (3.35) can be taken as the condition that balances the particular function of covariates. However, this standard approach may face a difficulty when a propensity score model is misspecified. Then the CBPS overcomes the missspecification of the parametric propensity score model by manipulating the twofold characteristics of the propensity score as a covariate balancing score and the conditional probability of treatment assignment[20].

Covariate balancing property is operationalized using inverse propensity score weighting:

$$E\left[\frac{T_i \widetilde{\boldsymbol{X}_\iota}}{\pi_\beta(\boldsymbol{X}_i)} - \frac{(1-T_i)\widetilde{\boldsymbol{X}_\iota}}{1-\pi_\beta(\boldsymbol{X}_i)}\right] = 0. \quad (3.36)$$

Where $\widetilde{\boldsymbol{X}_\iota} = f(\boldsymbol{X}_i)$ is an M-dimensional vector-valued measurable function of $\boldsymbol{X}_i$ [20].

For the score condition in equation (3.35), $\widetilde{\boldsymbol{X}_\iota} = \pi_\beta{}'(\boldsymbol{X}_i)$, as a result more weights are given to covariates that predict the treatment assignment in propensity score model. Nevertheless, equation (3.36) holds for any functions of covariates so that expectation exists. Hence, taking $\widetilde{\boldsymbol{X}_\iota} = \boldsymbol{X}_i$ balances the first moment of the covariates even the propensity score model is misspecified.

CBPS is estimated using the moment condition based on covariate balancing property under generalized method of moments (GMM)[152]. When a number of parameters are equal to the number of moment conditions, CBPS are considered to be *just identified.* For example, CBPS is just identified if we consider covariate balancing condition given in equation (3.36) by setting $f(\boldsymbol{X}_i) = \boldsymbol{X}_i$ [20]. However, the CBPS is over identified when we combine score condition in equation (3.25) and covariate balancing condition in equation (3.36) because the number of moment conditions $L + M$ exceeds the model parameters, L [20]

Formally, the sample covariate balancing moment condition in equation (3.35) is defined as[20]:

$$\frac{1}{n}\sum_{i=1}^{n} \hbar_{\beta}\,(T_i, \boldsymbol{X}_i)\widetilde{\boldsymbol{X}_i} = 0 \;\; , where \;\; \hbar_{\beta}(T_i, \boldsymbol{X}_i) = \frac{T_i - \pi_{\beta}(\boldsymbol{X}_i)}{\pi_{\beta}(\boldsymbol{X}_i)\left(1-\pi_{\beta}(\boldsymbol{X}_i)\right)}. \tag{3.37}$$

For over identified CBPS, and following [152], the efficient GMM estimator of the parameters is given by [20]:

$$\hat{\beta}_{GMM} = \arg\min_{\boldsymbol{\beta}\epsilon\Theta}\left\{\overline{g_{\beta}}(T, \boldsymbol{X})^{T}\Sigma_{\beta}(T, \boldsymbol{X})^{-1}\overline{g_{\beta}}(T, \boldsymbol{X})\right\}. \tag{3.38}$$

Where $\overline{g_{\beta}}(T, \boldsymbol{X})$ is the moment condition sample mean:

$$\overline{g_{\beta}}(T, \boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} g_{\beta}\,(T_i, \boldsymbol{X}_i). \tag{3.39}$$

And $g_{\beta}(T_i, \boldsymbol{X}_i)$ is the combination of all moment conditions:

$$g_{\beta}(T_i, \boldsymbol{X}_i) = \begin{pmatrix} S_{\beta}(T_i, \boldsymbol{X}_i) \\ \hbar_{\beta}(T_i, \boldsymbol{X}_i)\widetilde{\boldsymbol{X}_i} \end{pmatrix}. \tag{3.40}$$

The chosen consistent covariance estimator for $g_{\beta}(T_i, \boldsymbol{X}_i)$ is defined as:

$$\Sigma_{\beta}(T, \boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} E\left\{g_{\beta}(T_i, \boldsymbol{X}_i)g_{\beta}(T_i, \boldsymbol{X}_i)^{T}|\boldsymbol{X}_i\right\}. \tag{3.41}$$

For logistic regression propensity score model ($\pi_{\beta}(\boldsymbol{X}_i) = logit^{-1}(\boldsymbol{X}_i{}^{T}\beta)$, we have[20]:

$$\Sigma_{\beta}(T, \boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \pi_{\beta}(\boldsymbol{X}_i)\left\{1-\pi_{\beta}(\boldsymbol{X}_i)\right\}\boldsymbol{X}_i\boldsymbol{X}_i{}^{T} & \boldsymbol{X}_i\widetilde{\boldsymbol{X}_i}^{T} \\ \boldsymbol{X}_i\widetilde{\boldsymbol{X}_i}^{T} & \left[\pi_{\beta}(\boldsymbol{X}_i)\left\{1-\pi_{\beta}(\boldsymbol{X}_i)\right\}\right]^{-1}\boldsymbol{X}_i\widetilde{\boldsymbol{X}_i}^{T} \end{pmatrix}. \tag{3.42}$$

The optimal value of the parameter is estimated for the just identified CBPS using equation (3.38) excluding the score condition such that this objective function equals 0.

For over identified CBPS under GMM, the null hypothesis for the propensity score is correctly specified is tested using chi-square distribution with Hansen's J-statistic [20]:

$$J = n\left\{\overline{g}_{\hat{\beta}_{GMM}}(T, \boldsymbol{X})^{T}\Sigma_{\hat{\beta}_{GMM}}(T, \boldsymbol{X})^{-1}\overline{g}_{\hat{\beta}_{GMM}}(T, \boldsymbol{X})\right\} \rightarrow \mathcal{X}^{2}{}_{(L+M).} \tag{3.43}$$

Entropy balancing

Entropy balancing is a preprocessing method that enables researchers to generate balanced samples for estimating treatment effects. This process involves a reweighting scheme that assigns a scalar weight to each sample unit, ensuring that the reweighted groups meet specific balance constraints related to the sample moments of the covariate distributions. These constraints guarantee that the reweighted groups align precisely on the defined moments. The weights obtained from entropy balancing can be applied to any standard model the researcher chooses for analyzing outcomes in the reweighted data. The subsequent effect analysis functions similarly to that using survey

sampling weights or weights derived from a logistic propensity score model. This preprocessing step helps reduce model dependence in later analyses, as entropy balancing orthogonalizes the treatment indicator relative to the covariate moments included in the reweighting.[146].

Let the weighted counterfactual of the control group at treatment level $t = 1$ is [146, 154]:

$$E(Y(0)|T = 1) = \frac{\sum_{\{i|T=0\}} Y_i w_i}{\sum_{\{i|T=0\}} w_i}. \quad (3.44)$$

where w_i is the entropy balancing weight for each control unit, obtained by reweighting system that minimizes the entropy distance measures.

$$\min_{w_i} H(w_i) = \sum_{\{i|T=0\}} w_i \log(w_i/q_i). \quad (3.45)$$

subject to balance constraints

$$\sum_{\{i|T=0\}} w_i c_{ir}(\boldsymbol{X_i}) = m_r, with\ r \in 1,2,.,.,.,R \quad \text{and}$$

$$\sum_{\{i|T=0\}} w_i = 1 \text{ and}$$

$$w_i \geq 0 \ \ \forall i \text{ such that } T = 0.$$

Where $q_i = \frac{1}{n_0}$ is a base weight $c_{ir}(\boldsymbol{X_i}) = m_r$ is the set of R balance constraints executed on the covariate moments the reweighted control group. See the details elsewhere in [146, 154].

Super learner based IPTW

While parametric models are valuable in various contexts, they rely on strong assumptions that may not hold true in real-world scenarios. To address the limitations of these models, modern statistical and machine learning techniques including supper learner(SL) have been introduced, which impose less restrictive assumptions [155]. Supper learner is an ensemble learning algorithm that combines multiple candidate learners with weighted contributions to create the optimal estimator, aiming to minimize a specified loss function[155].

It is a method for selecting the optimal prediction algorithm from a set of user-defined models (In our case: stepwise model selection based on AIC (sl.stepAIC), generalized linear model (sl.glm), generalized additive model (sl.gam), and k nearest neighborhood (sl.knn)). It depends on the choice of a loss function and a library of candidate algorithms. SL evaluates the performance of these candidates using V-fold cross-validation. For each algorithm, it averages the estimated risks across the validation sets to obtain the cross-validated risk. These estimates are then used to calculate the best weighted linear convex combination of the candidates, which has the smallest

estimated risk. This weighted combination is applied to the full dataset to generate a new set of predicted values, known as the SL estimator[156].

3.3.5. Covariate Balance Assessment

The goal of IPTW is to balance the distribution of confounders between treatment and control groups after weighting. Standardized mean difference (SMD) is the widely used statistic to evaluate the balance of confounders. The unweighted SMD and weighted SMD can be obtained as [59, 157]:

$$\text{Unweight SMD} = \left| \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{\frac{S^2{}_1 + S^2{}_0}{2}}} \right|. \quad (3.46)$$

$$\text{Weighted SMD} = \left| \frac{\bar{X}_{1w} - \bar{X}_{0w}}{\sqrt{\frac{S^2{}_{1w} + S^2{}_{0w}}{2}}} \right|.$$

Where $\bar{X}_1 \; and \; \bar{X}_0$ are the samples means of treatment and control arms with the corresponding variances of $S^2{}_1 \; and \; S^2{}_0$ respectively. For dichotomous variable $\bar{X}_1 \; and \; \bar{X}_0$ are replaced by sample proportions of $P_1 \; and \; P_0$ respectively. $S^2{}_1 \; and \; S^2{}_0$ are also replaced by $P_1(1 - P_1)$ and $P_0(1 - P_0)$ respectively. The SMD greater than 0.1 represents the unbalance between treatment and control groups[158]. There are also other confounder balance assessment measures including Kolmogorov–Smirnov (KS) [159] test statistics, variance ratio with threshold value of 1 [160], and correlation coefficient [27, 161]. Balance assessment can be presented using bar chart[162] , boxplots [159], and entire distribution [3]. In this study, we used "love plot" to show balance graphically [59] using love.plot R functions from cobalt R package [127]

Assumptions of causal inference

When using weighting to control for bias due to confounding, several key assumptions must be met [59, 60, 64]:

Exchangeability (or ignorability): this is no unmeasured confounding assumption. Exchangeability assumption requires that the potential outcomes and treatment are independent, either marginally or conditionally. Mathematically, conditional exchangeability can be defined as $E[Y(a) \mid X] = E[Y(a) \mid A = a, X]$ for all values of treatment a. This assumption is unfortunately not testable. However, one can use sensitivity analysis to test if there is unmeasured

confounder[163]. In a sensitivity analysis, the researcher assumes specific values for the magnitude of the unobserved confounder and calculates how much the effect would change if that confounder were present.

Positivity: the positivity assumption states that every participant in the study must have none zero probability to receive either of the treatment level. That is none of the predicted probabilities that measure propensity score and then IPTW can be 0 and 1 conditional on observed confounders. Mathematically, positivity assumption is $0 < P(T = 1|\boldsymbol{X}) < 1$.

Correct Model Specification: the model used to estimate the propensity scores to estimate IPTW must be correctly specified. This includes having the correct functional form and including all relevant variables.

Sensitivity analysis

Sensitivity analysis is mostly used in causal inference to quantify the robustness of observed exposure-outcome association for unmeasured or uncontrolled confounders[164]. The sensitivity analysis can be quantified with the method called E-value. "*The E-value is the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to fully explain away a specific treatment–outcome association"[164].* Mathematically, the E-value is defined as[164, 165]:

E-value=$RR_{obs} + \sqrt{RR_{obs}(RR_{obs} - 1)}$, RR_{obs} is observed risk ratio.
E-value=$OR_{obs} + \sqrt{OR_{obs}(OR_{obs} - 1)}$, OR_{obs} is observed odds ratio.

3.3.6. Statistical Model to Estimate Treatment Effect

The outcome variable in this study is age-specific childhood vaccination which is ordinal with three categories (no vaccination, partial vaccination and full vaccination). We used weighted cumulative link model [109, 166] also called marginal structural model(MSM) [167]. The marginal structural model where the causal effects of a treatment can be consistently estimated using IPTW in ordinal outcome is specified as: $g(E(P(Y \leq j|T = t))) = \beta_j - \beta_1 t$ for $t = 0,1$.

where $g(.)$ is one of the link function in cumulative link model [109, 166] , β_j is the threshold or intercept, β_1 is the linear effect of the treatment (newborn postnatal care in our case) on the

outcome (age-specific childhood vaccination in our case), and $P(Y \leq j|T = t)$ is the cumulative probability of the outcome that falls in jth category. The Newton-Raphson method is used for finding the maximum likelihood estimates (MLE) of parameters in a CLM. The method iteratively refines parameter estimates by using the gradient (first derivative) and Hessian (second derivative) of the log-likelihood function. Essentially, it finds the "best fit" parameters for the model by maximizing the likelihood of observing the given data.

Using logit link function and odds ratio scale, the ATE is estimated as $\exp(-\beta_1)$. The ATE can also be estimated using risk ratio given by $\frac{P(Y \leq j|T=1)}{P(Y \leq j|T=0)}$.

3.3.7.Simulation Study

In order to validate the performance of weighting methods, we used two types of simulation study. The first one is plasmode simulation[71, 119] and the second one is parametric simulation [71]. Plasmode refers to a dataset created from real data where certain truths are known. A plasmode dataset are created by using an existing dataset as a template to simulate new datasets with known truths[69]. It involves resampling from a real-world dataset and incorporating a relationship between treatment and outcome defined by the investigator. By combing actual covariate data with a user-defined outcome data generation process, plasmodes preserve some of the complex relationships present in the original dataset while allowing for a specified causal effect between exposure and outcome[70]. In the outcome generation, the parameters are determined either from the truth relationship or from literature. While plasmode data contains hybrid (actual resampled data and synthetic outcome data), plasmode simulation is considered as semi-parametric simulation[71]. In this study, we resampled 1000 cases with replacement from original covariates and treatment, and we generated the outcome 500 times to estimate treatment effect and associated performance metrics using each of the proposed weighting methods. The parameters to generate the outcome are determined from the actual relationships.

In parametric simulation, covariates, exposure and outcome data are generated from known or predefined stochastic distributions such that the generated data resembles the realistic or representative. Parameters of interest are derived from the real data, literature, or set by the user[67, 68]. Parametric simulation is mainly used for model development and to compare the performance of different models. In this study, data were generated at sample size of 1000 repeated 500 times.

In parametric simulation, the binary treatment (treat), latent Gaussian outcome, and five ($X_1 - X_5$) independent standard normal random covariates, and five binomial random variables ($X_6 - X_{10}$) are generated. Let P be the probability of receiving the treatment, it is generated as:

$$log\left(\frac{P}{1-P}\right) = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_{10} X_{10} \text{ , treat=rbinom(n, size = 1, P).} \quad (3.47)$$

$$Y = \beta_{treat} * treat + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{10} X_{10} + \epsilon, \epsilon \sim N(0,1).$$

The coefficients $\alpha_1 - \alpha_{10}$ and $\beta_1 - \beta_{10}$ were determined from our real data exposure-covariate and covariates-outcome relationship. The true treatment effect was set to be 0.8. Continuous covariates were generated by accounting small to moderate correlation within them to consider the real world scenario. To generate the ordinal outcome, the latent Gaussian random variable was categorized at 16th and 96th quintiles to have equivalent percentage with the actual data.

To validate our finding in ordinal outcome, we tested on the continuous outcome and binary treatment at various sample sizes (n=500, 1000, 2000) each repeated 1000 times. Independently and identically distributed confounders were generated; the treatment and the outcome variables were also generated as follows:

$$log\left(\frac{P}{1-P}\right) = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_7 X_7. \quad (3.48)$$

$$Y = \beta_{treat} * treat + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_7 X_7 + \epsilon, \epsilon \sim N(0,1).$$

$$\alpha_1 - \alpha_7 = 0.2, -0.25, 0.6, -0.4, -0.35, -0.5, 0.7.$$

$\beta_1 - \beta_3 = log(2),\ \beta_4 - \beta_6 = log(1.75), \beta_7 = \log(1.5)$ which were adapted from Austin[168] and $\beta_{treat} = 0.8$.

To compare the performance of weighting methods, we used performance measurement metrics including bias, empirical standard error, mean square error, model standard error [68].

3.3.8. Result

Results of plasmode simulation

Table 3.10 summarizes the results of plasmode simulation for all weighting methods. The results revealed that the treatment effect size on the log odds of cumulative probability is small across all weighting methods with relatively higher in CBPS than other weighting methods. When we compare the weighting methods using performance measuring metrics, GBM weighting generated relatively the smallest empirical standard error, mean square error, and second smallest bias in absolute value. CBPS weighting generated relatively smaller bias in absolute value and the highest empirical standard error and mean square. However, the difference of performance metrics across all weighting methods is not substantial.

Table 3.10. Result of performance metrics for plasmode simulation**.**

Metrics	Weighting methods				
	PS weight	GBM weight	CBPS weight	Ebal weight	Super weight
point estimate	0.007	0.008	0.012	0.007	0.008
SE	0.162 (0.001)	0.164 (0.001)	0.217 (0.002)	0.162(0.0002)	0.164 (0.001)
Bias	-0.593 (0.008)	-0.592 (0.008)	-0.588 (0.009)	-0.593 (0.008)	-0.592 (0.008)
Empirical SE	0.175 (0.006)	0.174 (0.005)	0.271 (0.006)	0.175 (0.005)	0.175 (0.005)
MSE	0.382 (0.009)	0.380 (0.009)	0.420 (0.011)	0.382 (0.009)	0.381 (0.009)

PS weight= propensity score based weight, GBM weight= Generalized boosted model based weight, CBPS weight= Covariate balancing propensity score based weight, ebal weight= entropy balancing based weight, super weight= super learner based weight, Values in bracket are standard errors.

Results of parametric simulation

To further compare the performance of the weighting methods, the results of the artificial data simulation for ordinal outcome are presented in Table 3.11. The result contains the point estimates of the treatment effect on the log odds of cumulative probability along with performance metrics. The result indicates that the magnitudes of point estimate bias, empirical standard error, and mean square error are relatively the smallest in the GBM weighting across all sample sizes. It follows that super learner relatively performed the best in most of performance metrics next to GBM. However, the GBM method failed to balance confounders between the treatment and control groups which show performance metrics-confounders balance trade off in GBM weighting method. The entropy balancing based weight perfectly balanced all confounders between the treatment and control groups (see Figure 3.8). Propensity score based weighting balanced all

confounders across all sample sizes. CBPS balanced all confounders at the largest sample size (n=2000). Super learner weighting methods failed to balance confounders at sample size of 1000 and some confounders are at the threshold point for sample sizes of 500 and 2000.

Table 3.11. Summary results of artificial simulation for ordinal outcome.

Sample size	Metrics	Weighting methods				
		PS	GBM	CBPS	Entropy	Super learner
500	point estimate	-0.020	-0.010	-0.012	-0.011	-0.011
	Standard error	0.261 (0.001)	0.300 (0.001)	0.261 (0.001)	0.257 (0.001)	0.262 (0.001)
	Bias	-0.820 (0.012)	-0.808 (0.009)	-0.812 (0.011)	-0.811 (0.011)	-0.811 (0.011)
	Empirical standard error	0.382 (0.008)	0.291 (0.006)	0.349 (0.008)	0.346 (0.008)	0.347 (0.008)
	Mean squared error	0.817 (0.022)	0.737 (0.016)	0.780 (0.019)	0.778 (0.019)	0.778 (0.019)
1000	point estimate	-0.017	-0.005	-0.015	-0.012	-0.006
	Standard error	0.181 (0.0003)	0.218 (0.0003)	0.185 (0.0003)	0.181 (0.0002)	0.200 (0.0002)
	Bias	-0.817 (0.010)	-0.805 (0.007)	-0.815 (0.009)	-0.812 (0.009)	-0.806 (0.007)
	Empirical standard error	0.309 (0.007)	0.225 (0.005)	0.292 (0.006)	0.284 (0.006)	0.232 (0.005)
	Mean squared error	0.763 (0.017)	0.702 (0.012)	0.749 (0.016)	0.741 (0.015)	0.704 (0.012)
2000	point estimate	-0.015	-0.0101	-0.014	-0.014	-0.013
	Standard error	0.125 (0.0001)	0.140 (0.0001)	0.127 (0.0001)	0.125 (0.0001)	0.127 (0.0001)
	Bias	-0.815 (0.007)	-0.810 (0.005)	-0.814 (0.006)	-0.814 (0.006)	-0.813 (0.0063)
	Mean squared error	0.709 (0.011)	0.683 (0.008)	0.703 (0.011)	0.698 (0.010)	0.700 (0.010)
	Empirical standard error	0.213 (0.005)	0.164 (0.004)	0.201 (0.004)	0.188 (0.004)	0.198 (0.004)

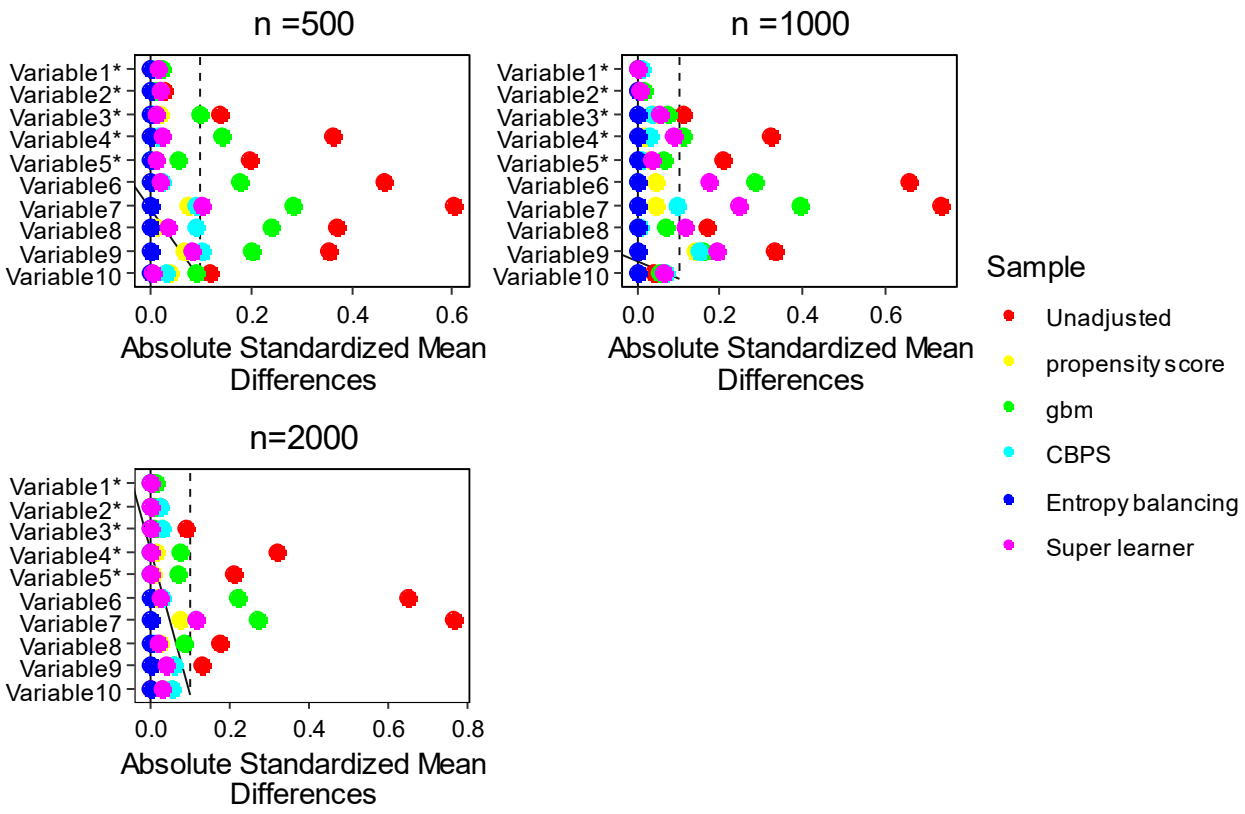


Figure 3.8. Confounders balance assessment for ordinal outcome.

Because the treatment effect is negative in the simulation study with an ordinal outcome, possibly due to transforming the latent continuous outcome into ordinal categories, we conducted an additional simulation study using a continuous outcome under different sample sizes. The results are presented in Table 3.12. We also extended the sample size to 5000. The entropy balancing outperformed in terms of bias and mean square error for sample sizes from 500-2000. It also outperformed in terms of mean square error while super learner outperformed in terms of bias. The result shows that GBM outperformed in terms empirical standard errors.

The performances of weighting methods are also compared in terms of confounders balance assessment graphically (Figure 3.9). It shows that entropy balancing weighting matched the treatment and control groups perfectly for all sample sizes. Propensity score also balances confounders for all sample sizes. CBPS and supper learner weighting do not balance confounders between treatment and control groups at sample sizes of 500 and 1000 but they balanced the confounders at the largest sample sizes of 2000 and 5000. GBM cannot balance most of

confounders at lower sample size and improves to large sample size. The finding shows that GBM cannot balance confounders for lower sample size and when the treatment and confounders relationship is high even for high sample sizes.

Table 3.12. Results of simulation study for continuous outcome under various sample sizes.

Sample size	Metrics	Weighting methods				
		PS	GBM	CBPS	Entropy balancing	Super learner
500	point estimate	0.914	1.07	0.901	0.811	0.942
	Standard error	0.166 (0.0002)	0.184 (0.0001)	0.169 (0.0001)	0.170 (0.0001)	0.170 (0.0001)
	Bias	0.114 (0.005)	0.266 (0.004)	0.101 (0.004)	0.011 (0.004)	0.142 (0.004)
	Empirical standard error	0.146 (0.003)	0.114 (0.003)	0.138 (0.003)	0.137 (0.003)	0.138 (0.003)
	Mean squared error	0.034 (0.001)	0.084 (0.002)	0.029 (0.001)	0.019 (0.001)	0.039 (0.002)
1000	point estimate	0.891	1.20	0.940	0.803	1.05
	Standard error	0.129 (0.0001)	0.141 (0.001)	0.130 (0.001)	0.128 (0.0001)	0.136 (0.0001)
	Bias	0.091 (0.004)	0.403 (0.003)	0.140 (0.004)	0.003 (0.004)	0.25 (0.003)
	Empirical standard error	0.121 (0.003)	0.091 (0.002)	0.115 (0.003)	0.114 (0.002)	0.093 (0.002)
	Mean squared error	0.023 (0.001)	0.171 (0.002)	0.033 (0.001)	0.013 (0.001)	0.071 (0.002)
2000	point estimate	0.962	1.03	0.987	0.798	0.967
	Standard error	0.087 (0.000)	0.092 (0.000)	0.088 (0.000)	0.087 (0.000)	0.088 (0.000)
	Bias	0.162 (0.003)	0.233 (0.002)	0.187 (0.003)	-0.002 (0.002)	0.167 (0.003)
	Empirical standard error	0.088 (0.002)	0.065 (0.001)	0.083 (0.002)	0.076 (0.002)	0.081 (0.002)
	Mean squared error	0.034 (0.001)	0.058 (0.001)	0.042 (0.001)	0.006 (0.0003)	0.034 (0.001)
5000	point estimate	0.779	0.947	0.808	0.805	0.798
	Standard error	0.057 (0.000)	0.059 (0.000)	0.058 (0.000)	0.057 (0.000)	0.057 (0.000)
	Bias	-0.021 (0.002)	0.147 (0.001)	0.008 (0.002)	0.005 (0.002)	-0.002 (0.002)
	Empirical standard error	0.052 (0.001)	0.042 (0.001)	0.050 (0.001)	0.047 (0.001)	0.051 (0.001)
	Mean squared error	0.003 (0.0001)	0.023 (0.000)	0.003 (0.0001)	0.002 (0.0001)	0.003 (0.0001)

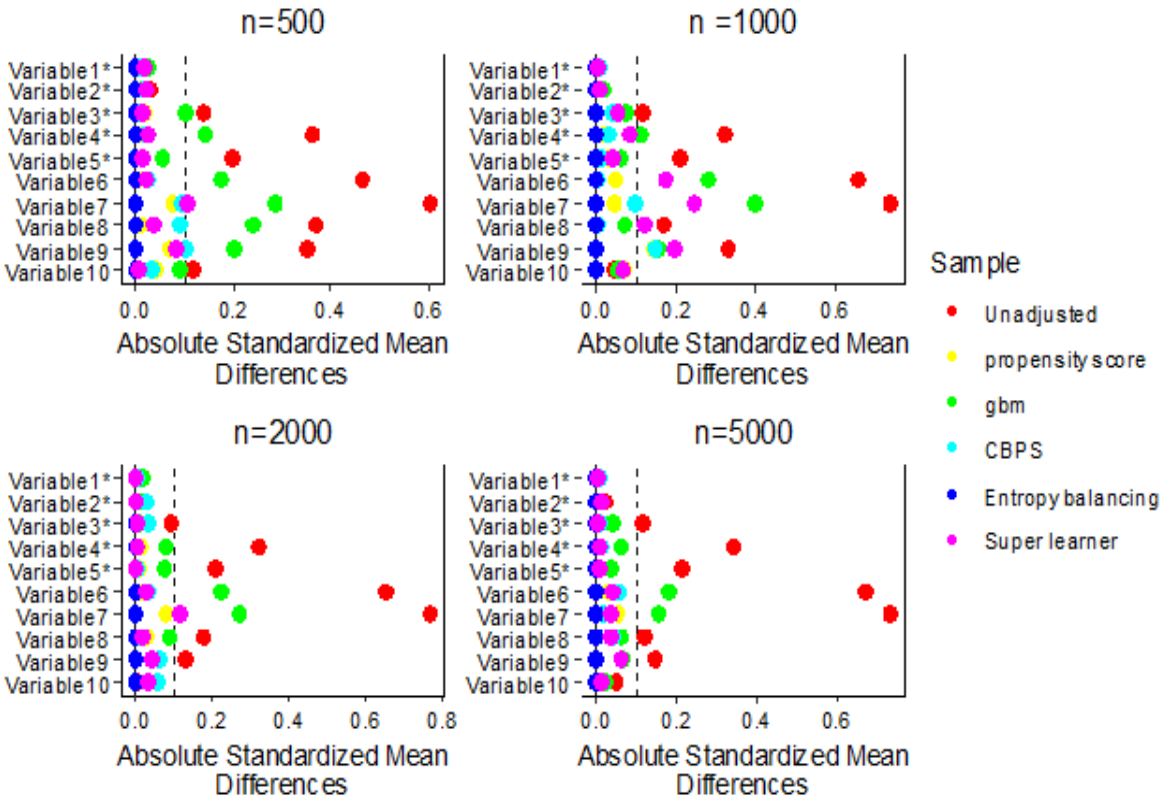


Figure 3.9**.** Confounders balance assessment for continuous outcome.

The result of performance measures in both plasmode and parametric simulation for ordinal outcome indicates GBM based weighting performed well in most of performance measuring metrics while it did not balance confounders. For continuous outcome, entropy balancing outperformed in terms of treatment effect estimation and confounders balance. However, the difference of each metrics among all weighting methods and all simulations is not notable, and one can say that all weighting methods demonstrated equivalent performance measures except GBM fails to balance confounders when the relationship between confounders and treatment is high. Nevertheless, the confounders balance was done from single estimate which was not from repetitions of all simulations.

Result of the actual data

Estimating confounding effect and balance assessment

Before estimating the average treatment effect, we need to balance the distribution of confounders between the treatment and control groups. To identify confounders, we first performed bivariate analysis and selected important variables at a 0.2 significance level. We then conducted

multivariable analysis. Adding confounders on the bivariate relationship of treatment and outcome, the result shows that the confounders reduced the treatment effect on log odds of cumulative probability by 42% (from 0.954 to 0.553). Meanwhile, institutional delivery, frequency of antenatal care visits, age of the household head, birth order number, household size, age at first birth, age category at the most recent birth, mother's educational status, access to media, household wealth status, region, and place of residence were identified as confounders that have a significant relationship with the outcome.

Figure 3.10 demonstrates the balance status of confounders between the treatment and control groups after weighting. As shown in the figure, confounders were perfectly balanced after weighting with the entropy balancing method, followed by the GBM-based weighting method. The frequency of ANC visits (ANC_num) is not balanced in the CBPS weighting methods. In the propensity score-based weighting method, ANC_num is at the boundary of the threshold point. The graph further indicates that before adjusting for confounders with the weighting methods, four confounders were unbalanced between the treatment and control groups. Notably, ANC_num is the primary selection variable included in the treatment group.

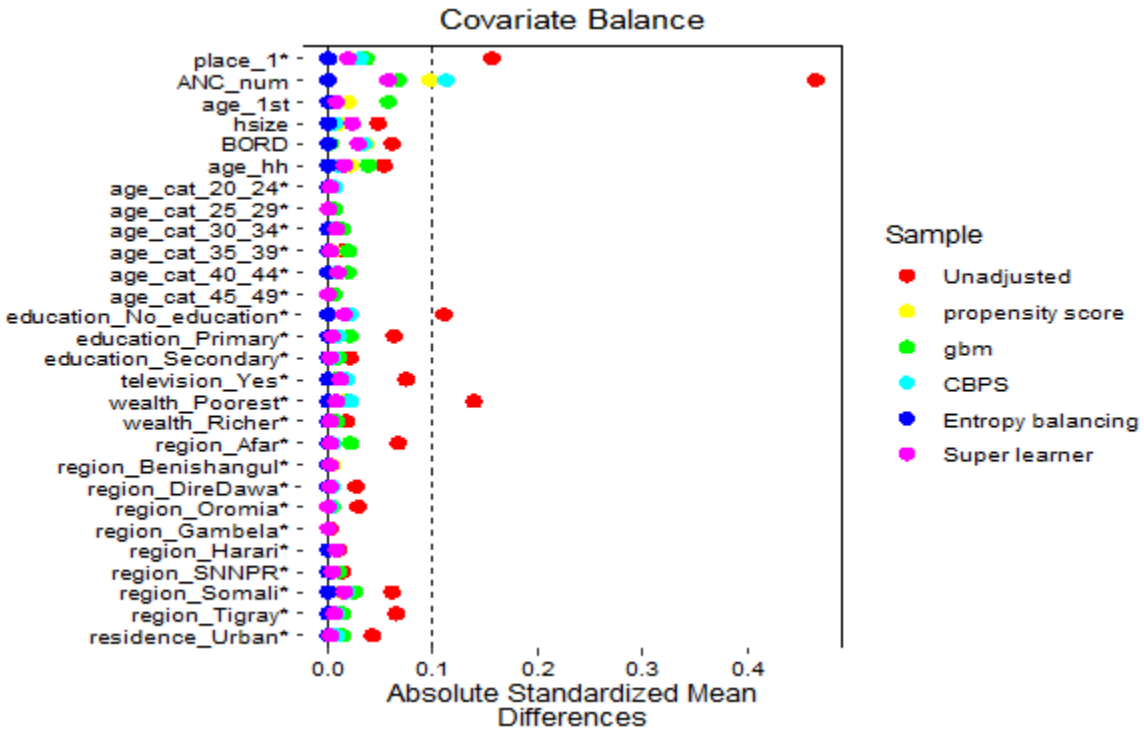


Figure 3.10**.** Confounder balance assessment for actual data.

Figure 3.11 also presents the visualization of the overlap or common support assumption of the propensity scores estimated from four methods. It shows that there is sufficient overlap of propensity scores between the treatment and control groups, implying that the treatment has matches in the control group.

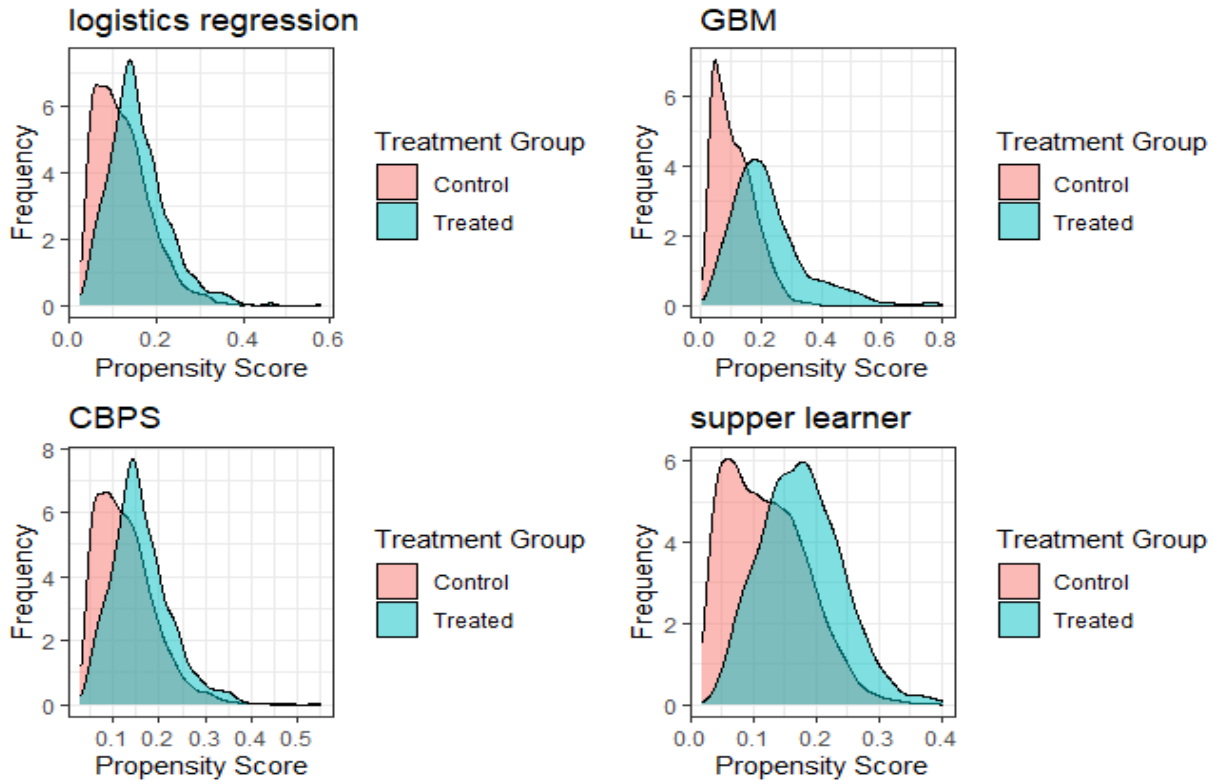


Figure 3.11. Density plot of propensity score as a measure of overlap assumption.

In addition, the distribution of stabilized weights generated from all methods is examined, as presented in Table 3.13. The results indicate that there are no extreme weights across all proposed methods.

Table 3.13. Distribution of stabilized weights.

Weighting methods	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
PS weight	0.219	0.934	0.979	0.997	1.039	3.083
GBM weight	0.159	0.915	0.961	0.962	1.031	3.554
CBPS weight	0.230	0.938	0.983	0.996	1.031	3.554
ebal weight	0.065	0.933	0.992	1.00	1.05	3.742
Super learner	0.317	0.925	0.976	0.978	1.042	2.845

Average treatment effect

Before estimating the average treatment effect of newborn postnatal care on age-specific childhood vaccination, we tested the proportional odds or parallel line assumption of the cumulative link model. The findings are presented in Table 3.14. The evidence demonstrated that the assumption is satisfied.

Table 3.14**.** Result of proportional odds/ parallel line assumption test.

Weighting methods	logLik	AIC	LRT	Pr(>Chi)
PS weight	-2836.3	5680.7	0.58756	0.4434
GBM weight	-2784.7	5577.3	1.5643	0.211
CBPS weight	2890	5788	0.61464	0.433
Entropy weight	-2896.5	5801	0.5774	0.4473
Super Learner weight	-2836.3	5680.7	0.58756	0.4434

As a result, the findings of the treatment effect on the log-odds of cumulative probability are revealed in Table 3.15 for all methods. The results include the average treatment effect (ATE), along with its standard error, 95% confidence interval, and significance level. The ATE effect size ranges from 0.522 in entropy balancing to 0.634 in the super learner method. The standard error of the estimate is the smallest in the entropy-based weight (Std. Error = 0.116) and the highest in the GBM-based weight (Std. Error = 0.131). Following the highest standard error, the confidence interval of the ATE size in the GBM-based weight is the widest among all methods. However, the ATE sizes are not substantial among the PS-based weight, GBM-based weight, and entropy balancing. Considering the odds ratio for entropy balancing, newborn postnatal care increased the likelihood of achieving a higher-level category of age-specific childhood vaccination by 68%, followed by 75% in GBM-based weighting.

Table 3.15**.** Result for the effect of newborn postnatal care on age-specific childhood vaccination.

Weighting methods	ATE	Odds ratio	Std. Error	z value	95% CI of ATE	
PS weight	0.591	1.81	0.118	5.01***	0.389	0.884
GBM weight	0.560	1.75	0.131	4.28***	0.307	0.820
CBPS weight	0.606	1.83	0.120	5.04***	0.373	0.844
Entropy weight	0.522	1.68	0.116	4.51***	0.297	0.751
Super Learner weight	0.634	1.88	0.126	5.02***	0.389	0.884

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sensitivity analysis

The results of the sensitivity analysis measured with the e-value are demonstrated in Table 3.16. The findings indicate that the minimum strength of association between an unobserved confounder and the outcome, in terms of odds ratio is 3.01 (95% CI: 2.23, 3.99) to fully explain away the observed odds ratio of 1.81 in the PS-based weight. In other weighting methods, the minimum odds ratio of the association between unobserved confounders and the outcome that would fully explain away the observed odds ratio ranges from 2.76 (95% CI: 2.03, 3.66) in entropy balancing to 3.07 (95% CI: 2.6, 4.08) in CBPS. The results in Table 3.16 imply that if an unmeasured confounder had an odds ratio ranging from 2.76 to 3.07 with the outcome, it could potentially account for the observed association, suggesting that the association may not be causal. Conversely, if no such confounder exists, or if unmeasured confounders are weaker than this e-value, the observed association (1.68–1.83) can be considered more robust.

Table 3.16**.** Sensitivity analysis for unobserved confounders in terms of odds ratio.

Weighting methods	Observed OR	e-value	95% CI of e-value	
PS	1.81	3.01	2.23	3.99
GBM	1.75	2.90	2.06	3.97
CBPS	1.83	3.07	2.26	4.08
Entropy	1.68	2.76	2.03	3.66
Super Learner	1.88	3.04	2.22	4.07

3.3.9. Discussion and Conclusion

For binary treatment, inverse probability treatment weighting can be estimated using different methods, including conventional logistic regression, machine learning algorithms, and entropy balancing. In this study, we addressed interconnected problems: comparing weighting methods and estimating the average treatment effect on the outcome. Both parametric and plasmode simulations were used to generate synthetic and hybrid data from predefined distributions for all covariates, treatment, and outcome, as well as artificial outcome data from actual covariates and treatment, respectively.

The findings of the study imply that, in both simulated datasets, the difference between performance measurement metrics is not meaningful. However, the performance of GBM is better than that of the other methods across most performance in the plasmode simulation and parametric

simulation for ordinal outcome. This finding is consistent with other research indicating that GBM is less biased than logistic regression and CBPS under complex relationships[169].

However, GBM fails to balance all covariates between the treatment and control groups after weighting in the parametric simulation for both ordinal and continuous outcome. When we compared the weighting methods using covariate balancing status with simulated data, entropy balancing perfectly balanced all covariates after weighting and outperformed in treatment effect estimation for continuous outcome. A study comparing CBPS and GBM showed that CBPS performed better than the GBM method when there is no complex association between the outcome or exposure and covariates. However, when the relationship is complex, the reverse is true[20, 153]. In line with our study, entropy balancing weight was found to perform best among machine learning methods in terms of balancing baseline characteristics, achieving perfect covariate balance. However, it produced results similar to those of conventional logistic regression[170]. In entropy balancing, unit weights vary smoothly to retain important information and avoid the need for continuous balance checking over propensity score models, which may stochastically balance the pre-treatment covariates [146]. The results of GBM-based weighting produced contradictory outcomes between treatment effect estimation and covariate balancing in the simulation study. Findings in the literature support the scenario that better covariate balance does not necessarily lead to lower bias. For example, logistic regression often yielded the lowest absolute mean difference but resulted in larger variance. On the other hand, boosted CART often produced a larger absolute mean difference but achieved better bias reduction [171].

For the real-life dataset, estimating the average treatment effect of newborn postnatal care on age-specific childhood vaccination, the confounders were balanced in all weighting methods except for CBPS, which did not balance the frequency of ANC visits, placing it at the boundary of the threshold point for propensity score-based weighting. Entropy balancing perfectly balanced the confounders, followed by GBM. Supporting our findings, a study in the literature showed that GBM outperformed logistic regression in confounder balancing[172].

The study found that newborn postnatal care significantly influenced childhood vaccination status across all weighting methods. The standard error of the average treatment effect in entropy balancing was less than the other weighting methods, resulting in a narrower confidence interval and smaller effect size. Newborn postnatal care increased the likelihood of timely childhood

vaccination by 68% in entropy balancing and by 88% in the super learner weighting method. In the GBM weighting method, newborn postnatal care increased the chance of timely vaccination by 75%. Some studies in the literature support our findings, although they do not adopt a marginal structural model approach. For example, a study in India found that postnatal care was significantly associated with timely vaccination, with an adjusted odds ratio of 1.48[142].
Maternal non-postnatal care lowered odds of childhood complete vaccination with an adjusted odds ratio of 0.26 [86]. Another study in Myanmar also showed that mothers who had postnatal care within 48 hours had completely vaccinated children [173].
The weighting methods were also compared in terms of sensitivity for unobserved or uncontrolled confounders. The sensitivity analysis was examined using the e-value [164]. In entropy balancing, the e-value was the smallest, implying that the treatment-outcome association could be more easily explained away by unobserved confounders compared to other weighting methods. Other weighting methods yielded equivalent high e-values, indicating that the odds ratio of unobserved-outcome association should be at least 3 to explain away treatment-outcome relationship.

To sum up, among the five proposed weighting methods, entropy balancing achieved perfect confounder balancing, the smallest effect size, and the lowest sensitivity value except in the simulation study with continuous outcome. GBM failed to balance covariates in the simulation study. CBPS did not balance all confounders in the actual data. However, except covariate balancing power, all weighting methods have equivalent performance in terms of treatment effect estimation. It is essential to implement multiple criteria to compare the performance of weighting methods before estimating treatment effect in causal inference using weighting method.
Newborn postnatal care significantly influenced timely childhood vaccination while ensuring that children are comparable in terms of their maternal and household characteristics.

Chapter three focuses on identifying appropriate confounders selection technique for both binary and count treatments, and estimating binary treatment effect on ordinal outcome after controlling confounding bias using IPTW. Chapter four deals with extending the binary treatment effect on ordinal outcome using IPTW and the causal effect of count treatment on ordinal outcome using conditional regression approach to the causal effect of count treatment on ordinal outcome after controlling confounding bias with IPTW generated by GPS. Maximum likelihood and machine

learning algorithms are compared for treatment model to estimate GPS and two outcome (MSM and conditional) models are compared to estimate count treatment effect on ordinal outcome.

CHAPTER FOUR: ESTIMATING COUNT TREATMENTS ON ORDINAL OUTCOMES: SIMULATION AND APPLICATION

4.1. Estimating Generalized Propensity Scores for Count Treatments

4.1.1. Background

Much of the studies on propensity scores in causal inference have focused on binary treatments proposed by Rosenbaum, P.R. and D.B. Rubin[18] where it was estimated by binary logistic regression. However, in real world, treatments are not only binary but also categorical with more than two levels and quantitative. Imbens[21] introduced generalized propensity score for multi-valued treatments. Following Imbens, others [174] also have used generalized propensity score for categorical treatments (multinomial and ordinal) with more than two levels.

Hirano and Imbens[26] have extended the GPS for continuous treatments. They used Gaussian distribution to estimate GPS. Y. Zhu et al.[27] estimated GPS from continuous treatments using boosted algorithm. Similarly, Wu et al.[28] have used matching on GPS for continuous treatments. Most of the studies on GPS for quantitative treatments focused on continuous treatments estimated by multiple linear regression with normal or kernel density and machine learning algorithms [16-18]. However, there is scarcity of literature on estimating GPS for count treatments.

Hence, this study focused on extending GPS for count treatments. Candidate treatment models were compared in terms of covariate balancing using GPS based weight, effective sample size after weighting and variability of generated GPS based weights.

4.1.2. Data

Simulation study

In this study, we used simulation data to compare different methods of estimating GPS for count treatments. In our simulations, we used 10 pre-treatment covariates under different sample size (500, 1000, 2000, 3000, 4000, and 5000) scenarios. The sample sizes were chosen in such a way that the inference is valid for different sample sizes. The samples were randomly selected from population size of 1000000. To have diversity of covariates: five of them were continuous, simulated from normal distributions with mean 0 and variance of 1($x_1, x_2, x_6, x_9, x_{10} \sim N(0,1)$), three covariates were binary (with common to rare outcomes), simulated from binomial distributions ($x_3 \sim bin(n, 0.3)$, $x_4 \sim bin(n, 0.09)$, $x_5 \sim bin(n, 0.5)$), one is generated from uniform

distribution, ($x_7 \sim U(0,3)$) , and the other is count generated from Poisson distribution ($x_8 \sim poisson(1.5)$).

The treatment variable denoted by T, is count generated from Poisson distribution $T \sim poisson(\lambda = E(T|\boldsymbol{X}))$. The coefficients are chosen based on [34] such that covariates influence treatment selection and the mean is not large. The treatment variable is generated from Poisson distribution with mean function given by:

$$E(T/\boldsymbol{X}) = exp(0.25 * x1 + 0.15 * x2 + 0.02 * x3 + x4 + 0.12 * x5 + 0.15 * x6 + 0.11 * x7 + 0.2 * x8 + 0.25 * x9 + 0.1 * x10)). \quad (4.1)$$

In addition to the independent covariates stated above, correlated covariates were also generated to consider the existence of real world situations. The covariates are generated as:

$x_1 \sim N(0,1))$, $x_2 \sim N(0,1)) + 0.5\, x_1$, $x_3 \sim N(0,1) + x_2$, $x_4 \sim bin(n, 0.3)$, $x_5 \sim bin(n, \exp(x_4)/(1 + \exp(x_4)))$, $x_6 \sim N(0,1)$, $x_7 \sim U(0,3) + x_6$, $x_8 \sim poisson(1.5)$, $x_9 \sim N(0,1)$, $x_{10} \sim N(0,1) + x_9$.

From household survey data, we used number of ANC contacts as count treatment. Fifteen baseline covariates such as age category of mothers, region, place of residence, mothers' education level, ownership of radio and television, religion, household size, sex of household head, age of household head, wealth status, total children ever born, age at first birth, marital status and birth order number that could affect the probability of number of ANC visits were considered**.**

4.1.3. Assumptions of Generalized Propensity Score Estimation

Let T be the count treatment with values $t = 0, 1, 2, \ldots, k$; $\boldsymbol{X}$ is a matrix of pretreatment covariates; and $r(t,x)$ is GPS score estimated by $f_{T|\boldsymbol{X}}(t|x)$. Then the following assumptions should be satisfied while estimating GPS [18, 26].

Unconfoundedness: this states that with the same values of $r(t,x)$**,** the probability of treatment assignment does not depend on pretreatment covariates: $\boldsymbol{X} \perp \boldsymbol{I}(T = t)|r(t,x)$. This means that given estimated GPS, treatment assignment to an individual need not be biased by unmeasured pretreatment covariates.

Positivity: this is also called overlap assumption. The estimated propensity score must be between 0 and 1: $0 < r(t,x) < 1$. That means, each subject in the study should have none zero probability to be assigned to the treatment or control group.

Correct specification of propensity score model: though formally untestable, the statistical model, $f_{T|X}(t|x)$, to estimate propensity score to balance covariates needs to be correctly specified. In reality, there are many propensity scores that can get rid of covariate imbalance between groups [18], so this assumption is well thought-out met if covariate balance is attained between treatment and control groups [159].

4.1.4. Estimation of Generalized Propensity Score

The unconfoundedness assumption requires adjusting the differences in the distribution of a set of pre-treatment covariates and removes all biases in comparison of treatment effect on the outcome. GPS is a scalar function which has a balancing property similar to binary treatment propensity score. Estimating this scalar function removes all biases caused by differences in covariates[26]. We used maximum likelihood, parametric and nonparametric methods of CBPS [20, 175], and generalized boosted model(GBM) [27, 151] to estimate the GPS for count treatment which is an extension of multivalued treatments [21, 176] and continuous treatments [26]. Maximum likelihood methods include generalized linear models with log link functions (Poisson and negative binomial) and zero truncated models such as zero inflated Poisson.
Allowing the treatment of interest to take on integer values 0 to k, the treatment space is given by $T = \{0,1,2, \dots, k\}$.

Following the definition of Imbens [21], the GPS is the conditional probability of a particular treatment assignment given baseline covariates:

$$r(t, \boldsymbol{X}) = P(T = t|\boldsymbol{X} = x). \qquad (4.2)$$

Methods used to estimate GPS

In this study, maximum likelihood method (family of count models) and machine learning algorithms were used to estimate GPS and compared their performance in terms of covariate balancing power and effective sample size.

Maximum likelihood methods

Given the treatment variable is count, Poisson regression and negative binomial regression are used from generalized linear model families and zero inflated Poisson regression is used from

zero-augmented models. For the treatment variable T, the generalized linear model is given by [103]:

$$f(t;\lambda,\varphi) = exp\left(\frac{(t.\lambda - b(\lambda)}{\varphi} + c(t,\varphi)\right). \quad (4.3)$$

With λ is a canonical parameter that depends on covariates through linear prediction; and φ is a dispersion parameter. The functions $b(.)$ and $c(.)$ are known and determine members of GLM families.

The conditional mean the treatment, $E(T|\boldsymbol{X}) = \mu$, on covariates is modeled as:

$$g(\mu) = \boldsymbol{X}^t\boldsymbol{\beta}\ . \quad (4.4)$$

$g(.)$ is log link function ($g(\mu) = log(\mu)$) and βs are parameters estimated by maximum likelihood methods as illustrated in section 3.1. Poisson distribution is used when mean and variances are equal and negative binomial distribution is used when there is excess zero or over dispersion. Both Poisson and negative binomial regression uses the same log-linear mean function, $log(\mu) = \boldsymbol{X}^t\boldsymbol{\beta}$, with different assumptions.

Zero inflated Poisson (ZIP) regression deals with excess zeros in count response where the excess zeros are modeled independently. In ZIP distribution, the observation zero is observed with probability of π and a Poisson ($\boldsymbol{\mu}$) random variable is observed in T with probability $1-\pi$.

Thus, the occurrence of treatment, T, distributed as follows[177]:

$$T = \begin{cases} 0\ with\ probabilty\ \pi + (1-\pi)\exp(-\mu) \\ k\ with\ probability\ of\ (1-\pi)\frac{\exp(-\mu)\mu^t}{t!} \end{cases},k > 0. \quad (4.5)$$

Considering vector of covariates, the zero inflated model which is a mixture of two models is given by $logit(\pi) = \boldsymbol{x'_1\beta_1}$ and $\log(\mu) = \boldsymbol{x'_2\beta_2}$. $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ may or may not be the same[105].

Procedures to estimate GPS:

1. We estimated parameters based on the assumption of GLM and zero-augmented models using glm function from stats R package [178] for Poisson, glm.gb function from MASS R package for negative binomial, and zeroinfl function from pscl R package [103]. Then get estimate of $E(T|\boldsymbol{X})$ which is $\hat{\mu}$
2. Estimate scalar GPS, $\hat{r}(T,\boldsymbol{X})$ using dpois() and dnbinom() R functions for Poisson and negative binomial and vector of GPS for each treatment, $i = 0,1,2,.,.,.,k$ using zero inflated Poisson regression using equation (4.4).
3. Estimate the marginal probability, $\hat{r}(T)$ also called stabilizing factor using steps 1 and 2.

4. Estimate the inverse probability weight such that $w = \frac{1}{\hat{r}(T,\boldsymbol{X})}$. However to avoid extreme weights due to small values of probability or propensity score, the weight is stabilized by marginal propensity score given by $w_s = \frac{\hat{r}(T)}{\hat{r}(T,\boldsymbol{X})}$.
5. Check covariate balance; examine effective sample size after balancing covariates and estimate treatment effect on outcome using weighted regression, using weight in step 4.

Covariate balancing generalized propensity score

While estimating propensity score, the propensity score models should be correctly specified, otherwise there is substantial bias of treatment effect [179]. This contradicts the purpose of propensity score to reduce the dimension of baseline covariates so that treatment and control groups are balanced with respect to baseline covariates. To overcome the problem of models misspecification, K. Imai and M. Ratkovic [20] introduced covariate balancing propensity core (CBPS), which is robust method of propensity score estimation for binary treatments such that covariate balancing is optimized. CBPS works under the frame work of generalized methods of moments or empirical likelihood.

The CBPS achieves the double characteristics of the propensity score as a covariate balancing score and the conditional probability of treatment assignment. "*CBPS dramatically improves the poor empirical performance of propensity score matching and weighting methods reported in the literature*" [20].

Fong, et al. [175] extended CBPS to continuous treatments and introduced covariate balancing generalized propensity score (CBGPS) that minimizes the correlation between treatment and covariates. They developed parametric CBGPS and non-parametric (npCBGPS) approaches which showed, with simulation study, the superiority over maximum likelihood estimation under misspecification. Parametric CBGPS approach follows closely to the MLE method, and a nonparametric extension, npCBGPS, places no parametric restrictions on the GPS and on the marginal distribution of the treatment [175]. This npCBGPS gives researchers a method to directly derive weights without giving a functional form to the propensity scores [175]. Hence, we applied CBGPS and npCBGPS for count treatments and compared its performance with standard maximum likelihood methods or count models. The mathematical formulation of the parametric

CBGPS for continuous treatment is presented in [175] and we defined the mathematical aspect of npCBGPS as described in [175].

Suppose T_i is the treatment indicator for the i^{th} individual whose support is $\boldsymbol{\mathcal{T}} \sqsubseteq \boldsymbol{\mathcal{R}}$, $\boldsymbol{X}_i$ is a K-dimensional observed covariates such that $\boldsymbol{X}_i \in \boldsymbol{\mathcal{R}}^K$, a sample of observation $\{\boldsymbol{X}_i, T_i\}$ randomly selected from joint distribution $f(\boldsymbol{X}_i, T_i)$. Assuming the ignorability and common support assumption and to estimate the npCBGPS, let's begin from centering and orthogonilizing the covariates and the treatment [175].

$$\boldsymbol{X}_i{}^* = \frac{\boldsymbol{X}_i - \bar{\boldsymbol{X}}}{\sqrt{\boldsymbol{S}_{\boldsymbol{X}}}}. \qquad (4.6)$$

$$T_i{}^* = \frac{T_i - \bar{T}}{\sqrt{S_T}}.$$

Where $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}_{\boldsymbol{X}}$ are sample mean vector and sample covariance of $\boldsymbol{X}_i$ respectively. $\bar{T}$ and S_T are the sample mean and variance of T respectively.

The npCBGPS does not require the treatment model is correctly specified. Instead, it uses an empirical likelihood method to choose inverse stabilized weight based on generalized propensity score and ensures covariate balances [175].

The stabilized weight is defined as:

$$w_i = \frac{f(T_i{}^*)}{f(T_i{}^* | \boldsymbol{X}_i{}^*)}. \qquad (4.7)$$

Where there are no parametric restrictions on GPS, $f(T_i{}^* | \boldsymbol{X}_i{}^*)$ and stabilizing marginal distribution, $f(T_i{}^*)$. The expectation of the stabilized weight over joint distribution is unity such that $E(w_i) = 1$.

After weighting with w_i, $T_i{}^*$ and $\boldsymbol{X}_i{}^*$ are uncorrelated (similarly, T_i and $\boldsymbol{X}_i$ are also uncorrelated).Mathematically:

$$E(\, w_i T_i{}^* \boldsymbol{X}_i{}^* \,) = E(T_i{}^*) E(\boldsymbol{X}_i{}^*) = 0. \qquad (4.8)$$

Similarly, weighting with w_i reserves the marginal means of $\boldsymbol{X}_i{}^*$ and $T_i{}^*$ which gives us:

$E(\, w_i \boldsymbol{X}_i{}^*) = E(\boldsymbol{X}_i{}^*) = 0$ and $E(\, w_i T_i{}^*) = E(T_i{}^*) = 0$.

The sample constraints on the mean of w_i,the marginal means of $T_i{}^*$, $\boldsymbol{X}_i{}^*$ and the cross products of $T_i{}^* \boldsymbol{X}_i{}^*$ gives rise to:

$$\sum_{i=}^{n} w_i \, g(\, T_i{}^*, \boldsymbol{X}_i{}^*) = 0 \text{ and } \sum_{i=1}^{n} w_i - n = 0. \qquad (4.9)$$

Where $g(T_i{}^*, \boldsymbol{X}_i{}^*) = (T_i{}^*, \boldsymbol{X}_i{}^*, T_i{}^* \boldsymbol{X}_i{}^*)^T$ with dimension of $2K + 1$.

While maximizing empirical likelihood of observing the data, w_i is chosen that satisfies the moment conditions in equation (4.7).

From equation (4.7), $f(T_i^*, \boldsymbol{X}_i^*) = \frac{1}{w_i} f(T_i^*)\, f(\boldsymbol{X}_i^*)$.

Hence, the likelihood function is defined as balances [175]:

$$\prod_{i=1}^{n} f(T_i^*, \boldsymbol{X}_i^*) = \prod_{i=1}^{n} \frac{1}{w_i} f(T_i^*)\, f(\boldsymbol{X}_i^*)\ . \qquad (4.10)$$

This empirical likelihood function is maximized by taking w_i subject to constraints[180]:

$\sum_{i=}^{n} w_i\, g(T_i^*, \boldsymbol{X}_i^*) = 0\ ,\ \sum_{i=1}^{n} w_i = n\ , \sum_{i=}^{n} w_i\, T_i^*\, \boldsymbol{X}_i^* = 0$ and $w_i \geq 0$.

This is same as maximizing:

$$\sum_{i=1}^{n} \big(\log f(\boldsymbol{X}_i^*) + \log f(T_i^*) - \log w_i\big). \qquad (4.11)$$

Where only the last term includes w_i. So that the estimation of npCBGPS is reduced to obtain:

$$\arg\min_{w \in \mathcal{R}^n} \sum_{i=1}^{n} \log w_i.$$

subject to the above constraints. w_i is chosen in this method estimates the stabilizing inverse GPS weight with the wanted covariate balancing properties built in the constraints [175].

The w_i is estimated using *empirical algorithm* and a *penalized imbalance* approaches [175].

The numerical algorithm uses Lagrange multiplier method to solve numerically the optimization problem[180]. Constructing the Lagragian [175]:

$$\mathcal{L}(w_i, \lambda, \gamma) = \sum_{i=1}^{n} log\ w_i + \lambda(n - \sum_{i=1}^{n} w_i) + \gamma^T \sum_{i=1}^{n} w_i g(T_i^*, \boldsymbol{X}_i^*). \qquad (4.12)$$

where λ and γ are Lagrange multipliers. Then the first order conditions are [175]:

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{1}{w_i} - \lambda + \gamma^T g(T_i^*, \boldsymbol{X}_i^*) = 0. \qquad (4.13)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = n - \sum_{i=1}^{n} w_i = 0.$$

$$\frac{\partial \mathcal{L}}{\partial \gamma} = \sum_{i=1}^{n} w_i g(T_i^*, \boldsymbol{X}_i^*) = 0.$$

Summing $w_i \frac{\partial \mathcal{L}}{\partial w_i} = 0$ over i to get $\lambda = 1$ then substitute into $\frac{\partial \mathcal{L}}{\partial w_i}$ and solve for w_i gives:

$$w_i = \frac{1}{1 - \gamma^T g(T_i^*, \boldsymbol{X}_i^*)}. \qquad (4.14)$$

Therefore, the constrained optimization problem is solved by unconstrained maximization:

$$\arg\max_{\gamma \in \mathcal{R}^k} \sum_{i=1}^{n} log\left(1 - \gamma^T g(T_i{}^*, \boldsymbol{X}_i{}^*)\right). \quad (4.15)$$

This is relatively straightforward. However, through optimization, the value inside the logarithm $(1 - \gamma^T g(T_i{}^*, \boldsymbol{X}_i{}^*))$ can be negative. In this case, the second order Taylor series approximation is used to the log around $1/n$. After we get the solution, w_i is obtained by $\frac{1}{1-\gamma^T g(T_i{}^*, \boldsymbol{X}_i{}^*)}$. Similarly, the penalized imbalance method is described in [175].

Generalized boosted model

One of the challenges in estimating GPS using regression is specifying the unknown functional relationship between covariates and the treatment. Generalized boosted modeling (GBM), also known as generalized boosted regression, solves the variable specification problem. McCaffrey et al. [151]first used the generalized boosting model (GBM) for the estimation of the propensity score, which they stated it works well in practice. GBM is machine learning; data-adaptive algorithm which uses regression trees as weak predictors and captures nonlinear and interactive effects of the covariates [181].

Given $r(t, \boldsymbol{X})$ is estimated GPS with covariate matrix $\boldsymbol{X}$, the GBM adds a regression tree (Say $\hat{h}(x)$) to the fitted $g(E(t|\boldsymbol{x}))$ (say $\hat{g}((t|\boldsymbol{x}))$ to obtain a better fit. The added regression tree is found by fitting the residuals of the estimated $\hat{g}(t|\boldsymbol{x})$ versus $\boldsymbol{X}$.

The $g(E(t|\boldsymbol{X}))$ is restructured by $\hat{g}(t|\boldsymbol{x}))$ + $\boldsymbol{\lambda}\ \hat{h}\ (x)$, where $\boldsymbol{\lambda}$ is known as the shrinkage factor or the learning rate. Normally the smaller the λ $(usually\ less\ than\ 1)$, the smoother the estimated propensity score. McCaffery et al. [151] suggested using 0.001 or 0.005 for the shrinkage parameter.

For count treatment T and matrix of covariates $\boldsymbol{X}$, GBM fits the log linear model for treatment on covariates in the general model form [17]:

$$g(E(T|\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x}). \quad (4.16)$$

Where $g(.)$ is log-link function and $m(\boldsymbol{x})$ is the mean function of log of T given $\boldsymbol{X}$. This mean function is estimated using a more flexible machine learning, GBM, algorithm which fits regression trees until the model is sufficiently fit the data [25, 27]. GBM automatically chooses important covariates, nonlinear, and interaction terms to correctly estimate the mean function hence providing better estimates of the GPS [25, 27]. With the estimated mean function, stabilized IPTW can be obtained and implemented just like in the maximum likelihood weighting procedure

given above. We fitted GBM using gbm R function from GBM package [182] with an interaction of three and shrinkage factor 0.005.

4.1.5. Covariate Balance Assessment

An important component of propensity score analysis is covariate balance assessment. In causal inference matching and weighting using propensity score are widely used to balance covariates. In this study we focused on IPTW using GPS to balance covariates. Imbens [21] and Zhang et al.[183] proposed weights can be obtained from GPS and defined as $\frac{r(T)}{r(T,X)}$ where the numerator is used to stabilize weight which is obtained from marginal density function, and estimated by fitting intercept only model.

Robins et al. [167] reminded that for quantitative exposures unstabilized weights (i.e., those with $r(T) = 1$) produces extreme weight and will have infinite variance and would not be used.

Based on GPS weights, there are two types of covariate balance diagnostics such as blocking based diagnostics [26, 161, 184]and correlation based diagnostics [27, 161]. The blocking diagnostic which is same as binary treatment uses absolute standardized mean difference between strata before and after balancing with GPS based weight. However, blocking focused stratifying treatments into strata where there may be loss of information while stratifying. In this study we used correlation based diagnostics using GPS-based weight for covariate balance assessment.

As introduced by Zhu et al[27], GPS-based weighted correlation between quantitative treatment and baseline covariates can show that the baseline covariates and quantitative exposure are independent. The authors recommended that confounding effect of covariates is small when the Average Absolute Correlation Coefficient (AACC) is less than 0.1. Covariate balance assessment was done using bal.tab and love.plot R functions from cobalt R package [127].

In addition to assessing covariate balance using AACC, we used effective simple size as a measure of performance for proposed treatment models. Effective sample size gives an important measure of inequality between treatment groups and a potential loss in accuracy from weighting[25]. A very small value of effective sample size relative to actual sample size implied a large disparity in the weights of the treatment group and that a small number of study subjects received very high weights as compared to the majority of study subjects in the sample. Among treatment models that yield equal covariate balance, the one with maximum effective sample size is preferred [25]. The effective sample size (ESS) can be estimated using the following formula [25, 185]:

$$Ess = \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} {w_i}^2}. \quad (4.17)$$

where w is the weight.

4.1.6. Result

Result for simulation study

Covariate balance assessment

The results of covariate balance assessments measured by ACC for individual covariates are presented in Table 4.1 and covariate balance assessments in terms of AACC is depicted in Figures 4.1 and 4.2.

For the sample size of 500, eight covariates were not balanced before weighting which absolute value of correlation coefficient are greater than the threshold value of 0.1. However, after adjusting with GPS weight obtained from GLM, only one covariate is unbalanced. Similarly, with GPS weights generated by GBM, two covariates are unbalanced (Table 4.1). The CBGPS and npCBGPS based weights balanced all covariates.

For the sample size of 1000, seven baseline covariates are unbalanced. After weighting with GLM based weight, all covariates are balanced. In each of other weighting methods, one covariate is unbalanced. Similarly, for sample size of 3000, six baseline covariates are unbalanced. Weighting by CBGPS based weights, two covariates are unbalanced. However, all other weighting methods balanced all covariates.

At the sample size of 4000 and 5000, GLM and GBM based weights balanced all covariates. But CBGPS and npCBGPS fail to balance some of the covariates.

For correlated covariates, at lower sample size (n=500), all weighting methods failed to balance at least one covariate. However, as the sample size increases, the GLM and GBM based weights consistently balanced all covariates whereas CBGPS and npCBGPS failed to do so (Table 4.1).

Considering AACC for the overall covariate balance assessment, the results are presented in Figure 4.1 for independent covariates and Figure 4.2 for correlated covariates. The AACC from GLM and GBM based weighting is below the threshold point (AACC=0.1) for both independent and correlated covariates. For CBPS, the AACC is above the threshold point for sample size of 4000 under independent covariates and well below the threshold point for correlated covariates across

all sample sizes. The AACC value is above the threshold point for sample size of 4000 and 5000 under independent covariates and sample size of 2000 under correlated covariates.

In addition to ACC and AACC, we used effective sample size to compare the treatment models and the result is presented in Figure 4.3 and 4.4. For GBM based weighting, the effective sample size is consistently increasing as the actual sample size is increasing. In addition, the size of the effective sample size in GBM weighting is more than other treatment models under independent and correlated covariates except at sample size of 5000 under correlated covariates. The effective sample size from other treatment models weight is inconsistent in both covariate conditions.

Table 4.1. Result of covariate balance assessment in simulation study for independent and correlated covariates.

Independent covariates															
	n=500					n=1000					n=2000				
covariate	Unadjusted correlation	GPS based Weight adjusted correlation				Unadjusted correlation	GPS based Weight adjusted correlation				Unadjusted correlation	GPS based Weight adjusted correlation			
		GLM	GBM	CBGPS	npCBGPS		GLM	GBM	CBGPS	npCBGPS		GLM	GBM	CBGPS	npCBGPS
X1	0.305	0.0529	0.083	0.081	0.048	0.233	0.009	0.089	0.038	0.083	0.065	0.134	0.010	0.0278	0.007
X2	0.133	0.0985	0.043	0.002	0.021	0.196	0.004	0.082	0.044	0.069	0.141	0.061	0.030	0.0377	0.001
X3	0.038	0.0985	0.039	0.030	0.005	0.011	0.059	0.035	0.029	0.005	0.039	0.031	0.011	0.0823	0.006
X4	0.529	0.0387	0.018	0.069	0.085	0.539	0.075	0.033	0.109	0.188	0.527	0.081	0.004	0.2598	0.001
X5	0.027	0.0296	0.027	0.001	0.003	0.066	0.022	0.006	0.028	0.024	0.025	0.158	0.009	0.0288	0.003
X6	0.158	0.0339	0.070	0.028	0.025	0.193	0.003	0.102	0.023	0.066	0.153	0.031	0.019	0.0503	0.004
X7	0.131	0.1417	0.110	0.068	0.021	0.082	0.028	0.063	0.036	0.032	0.080	0.086	0.024	0.0098	0.004
X8	0.242	0.0685	0.099	0.075	0.039	0.265	0.050	0.095	0.072	0.092	0.282	0.006	0.01	0.2003	0.002
X9	0.272	0.0611	0.114	0.078	0.044	0.181	0.050	0.071	0.032	0.060	0.267	0.021	0.002	0.0899	0.002
X10	0.122	0.0451	0.061	0.033	0.020	0.123	0.041	0.077	0.042	0.043	0.109	0.055	0.001	0.073	0.002
	n=3000					n=4000					n=5000				
covariate	Unadjusted correlation	GPS based Weight adjusted correlation				Unadjusted correlation	GPS based Weight adjusted correlation				Unadjusted correlation	GPS based Weight adjusted correlation			
		GLM	GBM	CBGPS	npCBGPS		GLM	GBM	CBGPS	npCBGPS		GLM	GBM	CBGPS	npCBGPS
X1	0.051	0.022	0.013	0.054	0.001	0.014	0.002	0.012	0.085	0.066	0.025	0.012	0.009	0.018	0.373

X2	0.132	0.008	0.018	0.027	0.002	0.149	0.040	0.003	0.176	0.025	0.157	0.039	0.028	0.029	0.121
X3	0.031	0.006	0.021	0.051	0.001	0.011	0.018	0.005	0.216	0.095	0.017	0.002	0.001	0.0001	0.309
X4	0.537	0.013	0.022	0.160	0.001	0.546	0.040	0.003	0.209	0.190	0.548	0.034	0.025	0.151	0.216
X5	0.023	0.005	0.014	0.018	0.004	0.038	0.034	0.024	0.085	0.082	0.032	0.001	0.009	0.011	0.093
X6	0.176	0.004	0.006	0.102	0.001	0.169	0.019	0.012	0.121	0.008	0.150	0.011	0.005	0.006	0.039
X7	0.096	0.011	0.002	0.022	0.004	0.102	0.007	0.027	0.172	0.052	0.101	0.015	0.012	0.030	0.132
X8	0.282	0.015	0.003	0.051	0.003	0.249	0.010	0.023	0.069	0.318	0.276	0.043	0.006	0.074	0.144
X9	0.279	0.034	0.015	0.117	0.001	0.260	0.039	0.015	0.172	0.314	0.250	0.034	0.009	0.082	0.258
X10	0.105	0.025	0.002	0.018	0.002	0.118	0.002	0.006	0.145	0.113	0.127	0.005	0.010	0.037	0.323

Correlated covariates

	n=500					n=1000					n=2000				
covariate	Unadjusted correlation	GPS based Weight adjusted correlation				Unadjusted correlation	GPS based Weight adjusted correlation				Unadjusted correlation	GPS based Weight adjusted correlation			
		GLM	GBM	CBGPS	npCBGPS		GLM	GBM	CBGPS	npCBGPS		GLM	GBM	CBGPS	npCBGPS
X1	0.209	0.137	0.019	0.004	0.104	0.110	0.017	0.073	0.066	0.038	0.089	0.032	0.049	0.000	0.079
X2	0.201	0.038	0.073	0.017	0.075	0.241	0.061	0.134	0.119	0.084	0.151	0.037	0.014	0.039	0.059
X3	0.122	0.135	0.037	0.062	0.091	0.153	0.024	0.082	0.052	0.052	0.126	0.011	0.008	0.003	0.112
X4	0.562	0.085	0.002	0.146	0.166	0.529	0.017	0.032	0.104	0.183	0.522	0.022	0.022	0.153	0.099
X5	0.199	0.150	0.066	0.072	0.019	0.167	0.048	0.070	0.061	0.061	0.148	0.028	0.023	0.035	0.236
X6	0.250	0.107	0.072	0.117	0.013	0.241	0.039	0.094	0.086	0.083	0.258	0.023	0.018	0.049	0.089
X7	0.249	0.091	0.071	0.085	0.101	0.217	0.044	0.101	0.095	0.075	0.230	0.021	0.004	0.019	0.104
X8	0.161	0.071	0.144	0.051	0.049	0.301	0.003	0.044	0.097	0.104	0.286	0.020	0.008	0.007	0.089
X9	0.294	0.005	0.090	0.072	0.052	0.225	0.039	0.078	0.111	0.079	0.258	0.042	0.002	0.006	0.050

X10	0.186	0.174	0.014	0.019	0.114	0.109	0.028	0.016	0.002	0.039	0.157	0.040	0.034	0.028	0.266
	n=3000					n=4000					n=5000				
covariate	Unadjusted correlation	GPS based Weight adjusted correlation				Unadjusted correlation	GPS based Weight adjusted correlation				Unadjusted correlation	GPS based Weight adjusted correlation			
		GLM	GBM	CBGPS	npCBG PS		GLM	GBM	CBGPS	npCBG PS		GLM	GBM	CBGPS	npCBG PS
X1	0.095	0.023	0.005	0.042	0.057	0.073	0.053	0.017	0.009	0.052	0.112	0.007	0.002	0.032	0.001
X2	0.158	0.035	0.006	0.017	0.018	0.166	0.068	0.011	0.003	0.082	0.197	0.022	0.015	0.005	0.001
X3	0.102	0.018	0.023	0.038	0.015	0.084	0.031	0.011	0.017	0.022	0.119	0.019	0.008	0.028	0.002
X4	0.532	0.017	0.020	0.143	0.016	0.549	0.029	0.014	0.091	0.044	0.523	0.040	0.041	0.145	0.016
X5	0.191	0.031	0.003	0.01	0.009	0.187	0.014	0.024	0.026	0.133	0.152	0.017	0.033	0.071	0.004
X6	0.273	0.019	0.014	0.086	0.003	0.253	0.045	0.004	0.006	0.061	0.280	0.034	0.005	0.047	0.003
X7	0.269	0.016	0.017	0.021	0.004	0.258	0.060	0.007	0.006	0.113	0.267	0.032	0.003	0.027	0.004
X8	0.265	0.009	0.028	0.094	0.024	0.237	0.010	0.002	0.031	0.021	0.262	0.019	0.009	0.003	0.002
X9	0.278	0.080	0.036	0.011	0.016	0.275	0.021	0.025	0.078	0.056	0.279	0.032	0.011	0.020	0.002
X10	0.177	0.015	0.007	0.089	0.001	0.144	0.017	0.001	0.051	0.012	0.176	0.025	0.007	0.016	0.005

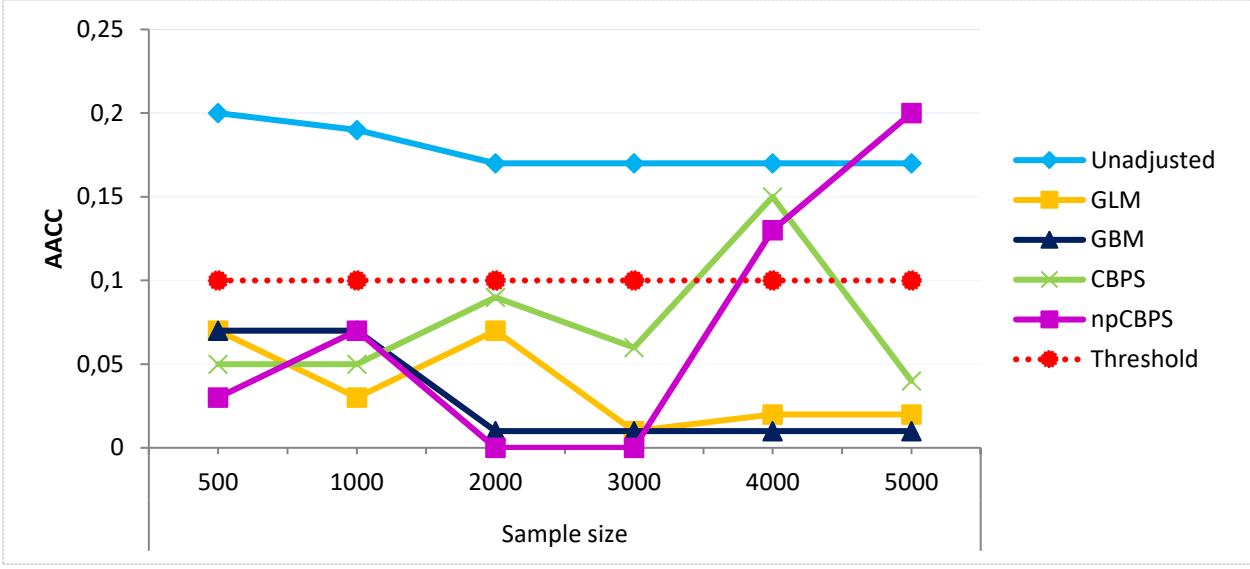


Figure 4.1. Sample size vs AACC of the simulation study for independent covariates.

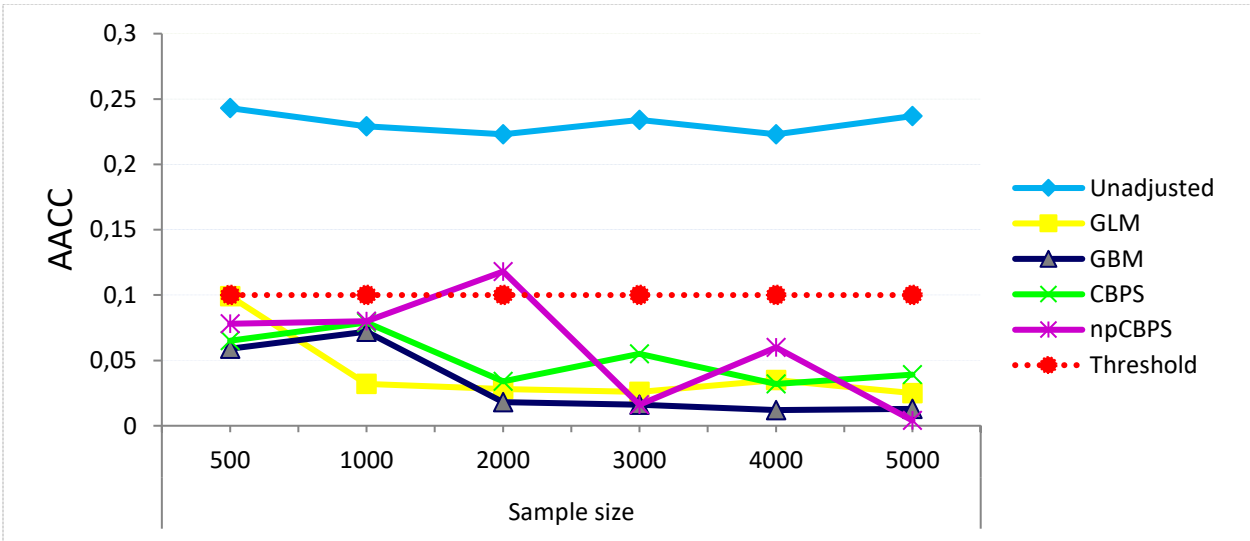


Figure 4.2. Sample size vs AACC of the simulation study for correlated covariates.

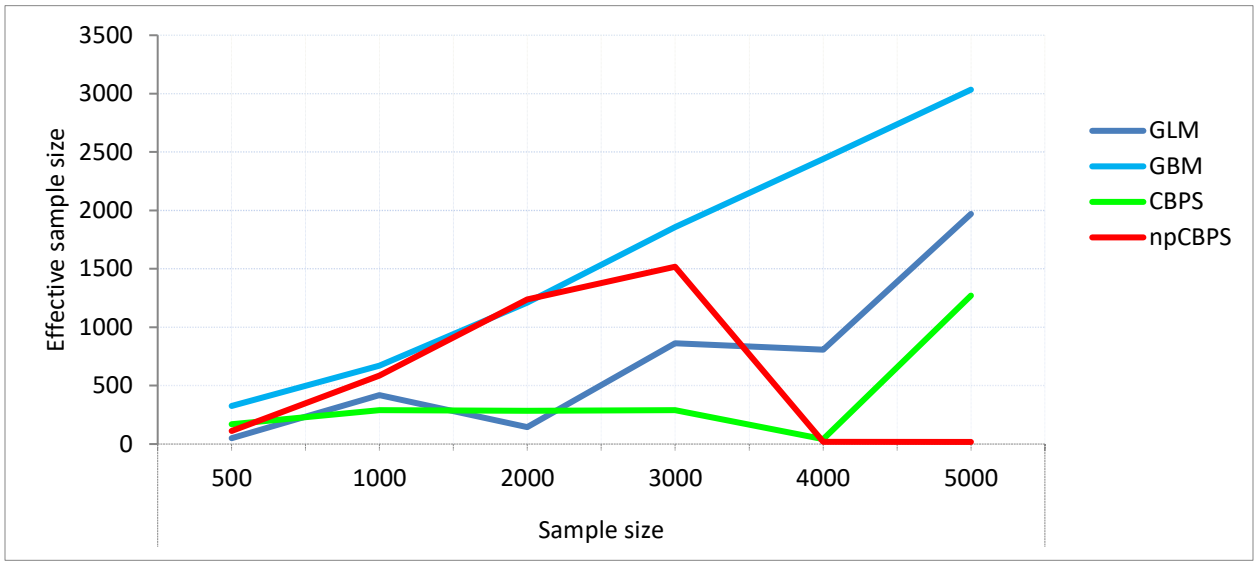


Figure 4.3. Sample size vs effective sample size of the simulation study for independent covariates.

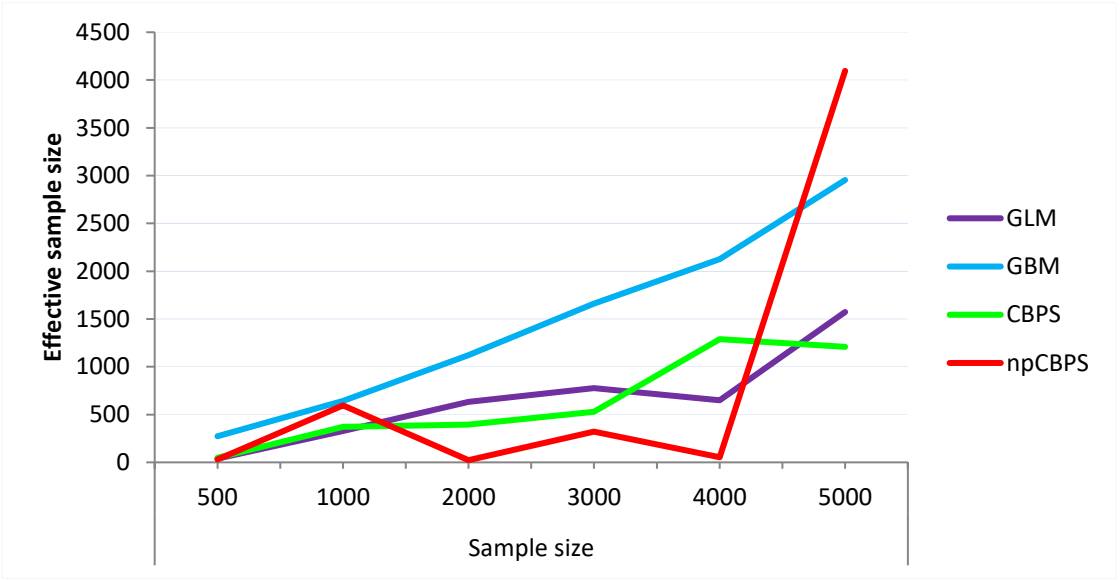


Figure 4.4. Sample size vs effective sample size of simulation study for correlated covariates.

Result of the actual data

This section presents the result of covariate balance assessment and summary of GPS and GPS based weights generated by GLM, zero-inflated model, GBM, CBGPS and npCBGPS.

Covariate balance assessment

Considering number of ANC contacts as count treatment, we did covariate balance assessment using correlation before and after adjusting with GPS based weight. The result is presented in

Table 4.2. Before adjusting with GPS weight, there is imbalance with respect to some covariates (ACC is greater than 0.1; and AACC value is 0.121). This implied that there are covariate imbalances among treatment values before weighting.

When we weight with stabilized IPTW generated by GLM with Poisson distribution, poorest wealth category was unbalanced (ACC=0.113) with AACC of 0.027. This indicates on average baseline covariates are balanced. GLM with Poisson distribution generates effective sample size of 1919. However, when we use stabilized IPTW generated by GLM using negative binomial distribution, all covariates are balanced and the effective sample size is 2142. This entails that negative binomial distribution give better covariate balance and maximum effective sample size as compared to Poisson distribution.

The result in Table 4.2 also showed that when we estimate stabilized GPS based weight with zero-inflated Poisson (ZIP) regression, it looks ownership of television (ACC=0.106) and richest wealth category are unbalanced even if the AACC is 0.032. This denoted that estimating GPS with ZIP gives balanced covariates on average with effective sample size of 3012. GBM balanced all covariates with effective sample size of 3148. Likewise, CBGPS balanced all covariates with effective sample size of 1851 which is less than effective sample size of all other methods. Relatively, npCBGPS gave more unbalanced covariates as compared to other methods though it balances on average (AACC=0.037) with effective sample size of 2536.

Table 4.2**.** Result of covariate balance assessment before and after adjusting with GPS weight.

covariate	Unadjusted correlation	Weight adjusted absolute correlation					
		Poisson	NB	Zip	GBM	CBGPS	npCBGPS
age_cat_15-19	0.036	0.001	0.024	0.010	0.029	0.001	0.010
age_cat_20-24	0.028	0.023	0.036	0.007	0.012	0.002	0.011
age_cat_25-29	0.037	0.007	0.017	0.026	0.016	0.003	0.011
age_cat_30-34	0.003	0.014	0.019	0.025	0.012	0.001	0.001
age_cat_35-39	0.016	0.006	0.001	0.004	0.025	0.003	0.007
age_cat_40-44	0.046	0.017	0.038	0.005	0.036	0.014	0.014
age_cat_45-49	0.043	0.015	0.005	0.018	0.014	0.002	0.013
Addis Adaba	0.208	0.025	0.057	0.067	0.034	0.008	0.060
Afar	0.117	0.049	0.039	0.010	0.065	0.002	0.037

Amhara	0.053	0.028	0.040	0.008	0.010	0.007	0.018
Benishangul	0.032	0.011	0.027	0.004	0.003	0.006	0.011
Dire Dawa	0.126	0.003	0.061	0.015	0.008	0.003	0.031
Gambela	0.042	0.032	0.042	0.032	0.027	0.002	0.014
Harari	0.064	0.067	0.055	0.024	0.066	0.004	0.022
Oromia	0.026	0.006	0.023	0.010	0.005	0.003	0.004
SNNPR	0.058	0.023	0.008	0.007	0.010	0.006	0.018
Somali	0.255	0.070	0.013	0.046	0.027	0.035	0.078
Tigray	0.129	0.000	0.038	0.033	0.043	0.006	0.042
Residence_Urban	0.332	0.001	0.063	0.081	0.032	0.009	0.099
education_Higher	0.231	0.005	0.045	0.045	0.029	0.005	0.068
No education	0.327	0.034	0.009	0.073	0.004	0.016	0.101
Primary	0.119	0.050	0.043	0.006	0.032	0.007	0.036
Secondary	0.201	0.018	0.052	0.085	0.035	0.013	0.067
radio_Yes	0.173	0.002	0.028	0.028	0.012	0.005	0.051
television_Yes	0.372	0.012	0.080	0.106	0.021	0.015	0.110
religion_Catholic	0.017	0.015	0.019	0.012	0.016	0.007	0.006
Muslim	0.160	0.045	0.029	0.039	0.015	0.017	0.050
Orthodox	0.223	0.024	0.024	0.062	0.028	0.016	0.070
Protestant	0.028	0.040	0.069	0.016	0.035	0.005	0.009
hsize	0.187	0.011	0.001	0.053	0.057	0.027	0.058
sex_hh_Male	0.015	0.052	0.034	0.002	0.035	0.008	0.003
age_hh	0.017	0.034	0.017	0.022	0.022	0.001	0.006
wealth_Middle	0.005	0.053	0.077	0.020	0.035	0.004	0.002
wealth_Poorer	0.047	0.052	0.044	0.034	0.020	0.004	0.014
wealth_Poorest	0.367	0.113	0.055	0.076	0.047	0.024	0.114
wealth_Richer	0.086	0.048	0.053	0.025	0.021	0.008	0.031
wealth_Richest	0.389	0.002	0.086	0.115	0.011	0.015	0.116
total_child	0.237	0.032	0.017	0.059	0.024	0.025	0.075
age_1st	0.145	0.016	0.015	0.041	0.018	0.001	0.044
marriage_Divorced	0.020	0.009	0.009	0.004	0.006	0.000	0.008

marriage_Married	0.014	0.009	0.020	0.001	0.017	0.001	0.005
marriage_Widowed	0.000	0.003	0.019	0.006	0.028	0.002	0.001
BORD	0.212	0.049	0.031	0.050	0.004	0.014	0.067
Average correlation	0.121	0.027	0.034	0.032	0.025	0.008	0.037
Adjusted effective sample size		1919	2142	3012	3148	1851	2536

Summary of weights

Table 4.3 presents distributions of GPS based weights or stabilized IPTW estimated by various methods. The largest weight is found in Poisson regression (max. weight=35.991), second and third largest weights are observed in CBGPS (max.weight=29.57) and negative binomial distribution (max. weight=28.276). The least maximum weight is observed in npCBGPS (max. weight=0.0083). The minimum weight is approximately non-zero except in npCBGPS. There is no more variation in the median weight except npCBGPS, and the same is true for the mean weight. The standard deviation of weight is relatively the largest in CBGPS followed by Poisson distribution and negative binomial distribution. An important statistical method to compare variability is coefficient of variation (CV). Its result showed that weights generated by CBGPS (CV=1.34) and GLM with Poisson distribution (CV=1.297) as well as negative binomial distribution (CV=1.185) generated more variability consecutively. The least weight variability is observed in GBM (CV=0.796) and ZIP (CV=0.843).

Table 4.3. Summary of weights when ANC contacts are treatment variable.

Summary measures	Poisson	Negative binomial	Zero-inflated Poisson	GBM	CBGPS	npCBGPS
Min	0.004	0.090	0.025	0.011	0.008	0.0000
1st Qu.	0.530	0.662	0.595	0.566	0.644	0.0001
Median	0.905	0.787	0.859	0.909	0.867	0.0002
Mean	1.059	1.029	0.968	0.974	1.147	0.0002
3rd Qu	1.111	1.009	1.093	1.086	1.119	0.0002
Max.	35.991	28.276	18.781	14.67	29.57	0.0083
Sd	1.374	1.220	0.815	0.776	1.53	0.0002
cv	1.297	1.185	0.843	0.796	1.34	1.0153

4.1.7. Discussion and Conclusion

Though propensity score analysis for binary treatments [18, 186], for ordinal treatments [187], for multiple treatments [20, 21, 23, 188] and for continuous treatments [26-28, 161, 175, 183] have been done so far, there remains a lack of literature on how to estimate propensity score specific to count treatments, and no well-studied methods to conduct propensity score weighting as covariate balance. Hence, this study extended estimating a scalar balancing score called GPS for count treatments and we compared various methods to estimate GPS for such treatment.

Both the simulation study and actual data from household survey have been used for demonstration. The simulation was done to represent all data types (continuous, binary, uniform and discrete) for various sample sizes. Treatment variable was simulated using Poisson process in all samples size scenarios.

Various methods have been tested their performance in terms of covariate balancing power, effective sample size and distribution of stabilized IPTW or GPS based weights. Weighted correlation between treatment and covariates was used as a measure of covariate balance. The absolute correlation between the treatment and each covariate is a global measure and can tell whether the whole matched set is balanced. Based on the balancing condition, the correlations between the exposure and pre-exposure covariates should, on average, be equal to zero if covariate balance is achieved[28]. Peter C Austin [161] suggested that correlation based covariate balancing diagnostics have better performance and are simpler to summarize and report.

When the AACC and Maximum ACC (MACC) are less than 0.1, the confounding effect of covariates is small; when the AACC and MACC are between 0.1 and 0.3, the confounding effect is medium; and when AACC and MACC are greater than 0.55, the confounding effect is large[27].

On the other hand, McCaffrey et al. [151] recommended that the GBM's iterative process for binary treatments creates highly variable weights and small effective sample sizes. McCaffrey et al. [25] provided tutorials to estimate propensity scores and propensity score weights with GBM and procedures to evaluate balance and overlap of pretreatment covariates for multiple treatments. Y. Zhu et al [27] extended GBM to continuous treatments and applied IPTW to estimate treatment effect.

Imai and Ratkovic [20] used CBPS for multiple treatments to optimize covariate balance. They reported that CBPS radically improves the poor empirical performance of propensity score matching and weighting methods reported in the literature.

Fong et al.[175] extended CBPS to continuous treatments and named it covariate balancing generalized propensity score (CBGPS).They developed parametric or non-parametric approach of which minimizes correlation between treatment and covariates. They also suggested that, with simulation study, the approaches out performed standard maximum likelihood methods.

In our simulation study, GLM and GBM based weighting consistently balanced covariates at least on average under both simulation conditions but CBGPS failed to balance covariates on average at sample size of 4000 under independent covariates and the AACC was higher than GLM and GBM. Similarly, npCBGPS failed to balance covariates on average at sample size of 4000 and 5000 under independent covariates and at sample size of 2000 under correlated covariates.

GBM consistently retained maximum effective sample size and consistently increased along with actual sample sizes under both simulation conditions, whereas the effective sample size in other methods was not consistently increasing along actual sample sizes. When effective sample size is high, it implies the weights are less variable, and the weighted estimate has high precision, alike an unweighted sample.

For the actual data, the average absolute correlation coefficient (AACC) decreases after weighting with GPS weight for all methods. Using weighted AACC threshold value of 0.1; all covariates were balanced for all proposed methods. However, GBM method retained maximum effective sample size followed by zero-inflated Poisson method. The GPS based weights generated by GBM and Zip are more consistent than other methods.

To conclude, this study proposed extending propensity score analysis for count treatments and examine covariate balance using GPS based weighted correlation and effective sample size after controlling covariates difference across treatment values. We also compared various treatment models for GPS estimation for count treatments given pre-treatment covariates. GBM outperformed in terms of giving smallest weighted AACC, especially for high sample sizes, more consistent GPS based weight and maximum effective sample size both in simulation studies and actual data analysis followed by zero-inflated Poisson regression for actual data analysis.

One can estimate the causal effect of count treatment on outcome by controlling pre-treatment covariates and/or confounders with GPS weight. However, it is necessary to compare various count models and boosted algorithms such as GBM to estimate GPS for count treatment although we found GBM works well in our study. In addition to covariate balance assessment, effective sample

size after balancing covariates across treatment values should be used as GPS model performance metrics.

This study used to compare count treatment models for GPS estimation in terms of covariate balance assessment and effective samples size. However, it did not consider treatment effect estimation on outcome. Hence, the subsequent section addressed the issue of estimating count treatment effect on ordinal outcome and two outcome models were compared.

4.2. Causal Effect of Count Treatments on Ordinal Outcomes

4.2.1 Background

RCT, which is a gold standard research design, is a random allocation of contrasted treatments or exposures to experimental units [2]. When study participants in RCT are assigned to treatment and control groups randomly, each participant in a group would have the same probability. Under RCT, the baseline covariates would have similar distribution in treated and control groups since baseline covariates did not affect treatment assignment. As a result, randomization reduces bias arising from baseline covariates and treatment effects can be estimated by comparing the outcome of treatment and control groups [2, 3].

In order to balance baseline covariates distribution between treatment and control groups and accurately estimate treatment effect in observational study, analysts proposed statistical methods such as matching, stratification, and regression [4]. Matching and stratification is grouping of treatment and control groups with identical covariate distribution. However, these methods fail for high dimensional covariates since it is difficult to get individuals identical with all covariates. For regression adjustment, it is difficult to assess covariate balance between treatment and control groups [15]. Although multiple regressions, which is a direct and simple way to remove bias due to confounders and easily handle high dimensional covariates, it may have limitations and not be reliable [15, 16]. In addition, when there is a considerable difference between treatment and control groups, regression analysis cannot make good adjustments [17] .

A new method, which has gained attention in recent decades, proposed to adjust covariate imbalance or confounders in observational study is propensity score. It was first introduced by Rosenbaum and Rubin [18] and they define propensity score as the probability of treatment

assignment conditional on observed baseline covariates**.** Propensity score allows observational study to mimic the characteristics of an RCT.

Many propensity score methods have been proposed to make causal inferences from observational data. These are matching with propensity score, stratification with propensity score, inverse probability weighting, and covariate adjustment using propensity score [18, 57, 58].

To the best of our knowledge, most of the studies on causal inference using GPS focused on multi-categorical treatments (categorical or ordinal) and continuous treatments where GPS is estimated by multinomial logistic regression, ordinal logistic regression, linear regression with normal or kernel density, and other machine learning algorithms such as generalized boosted algorithm and covariate balancing propensity score. However, there is a scarcity of literature on estimating the causal effects of count treatments on ordinal outcomes.

Hence, this study focused on choosing outcome models while estimating the causal effect of count treatments on ordinal outcome variables using generalized propensity scores. The performance of outcome models was compared using multi-dimensional performance metrics[68] from simulated and actual data.

4.2.2 Notation and Assumption

We have a random sample of size n, indexed by $i = 1, 2, 3, \dots, n$. Let T_i denotes the count treatment (in our case number of ANC contacts) applied on individual i with domain $\mathcal{T} = \{0,1,2, \dots, k\}$, $Y_i(t)$ is the potential outcome for individual i as a result of treatment assignment at $T_i = t_i$, and $\boldsymbol{x}_i$ is a P-dimensional vector of observed baseline covariates. For the simplicity of notation, we drop the subscript i. Following Imbens [21] and Hirano and Imbens [26], the probability mass function of treatment assignment given the covariates is:

$r(t, \boldsymbol{x}) = P(T = t|\boldsymbol{x} = x), t \in \mathcal{T}$. Then the generalized propensity score is $R = r(t, \boldsymbol{x})$. To estimate the causal effect of a treatment from observational data, the following assumption are assumed satisfied.

No unmeasured confounding assumption: the treatment assignment is independent of potential outcome given generalized propensity score [26]: $Y(t) \perp T|r(t, \boldsymbol{x})$. That is for every t, $P(t|r(t, x), Y(t)) = P(t|r(t, \boldsymbol{x})), for\ t \in \mathcal{T}$ [27].

Positivity assumption: $0 < r(t, \boldsymbol{x}) < 1, \forall t \in \mathcal{T}$, this is an overlap or common support assumption that requires every study participant to have a chance to be in any of the treatment conditions. The probability of treatment assignment for any participant is neither zero nor one under treatment conditions [18]. A generalized propensity score outside of [0.01,0.99] was trimmed [62, 63] to prevent over-dispersion of weights and for consistent treatment effect estimate.

Stable unit treatment value assumption (SUTVA): the potential outcomes of i^{th} subject are not influenced by the potential outcome of j^{th} subject, where $i \neq j$, and each unit receives the same version of the treatment. This is an assumption of no interference and consistency [1].

4.2.3 Sensitivity Analysis

The no unmeasured confounding assumption is alternatively called "exchangeability" assumption which states treatment and control groups have the same potential outcome [189] given confounders are controlled. This means that treatment and control groups are exchangeable having the same propensity score and the probability of the outcome would be the same had either group been exposed. Notably, exchangeability also entails that there are no unmeasured confounders that imbalance the treatment and control groups [64, 190, 191]. Sensitivity analysis is used to measure the existence of unmeasured confounders or hidden bias [192]. It examines the level of strength that unmeasured confounding factors would need to reach in order to "explain away" the observed association. In other words, it assesses how strongly the unmeasured confounders must be related to both the treatment and the outcome for the treatment-outcome association to be considered non-causal[164].

To deal with unmeasured confounding in causal inference research, VanderWeele and Ding[164] developed an easily calculated statistical tool called the E-Value. The E-value is the minimum strength of association that an unmeasured confounder would need to have with both the exposure and the outcome to fully explain away observed treatment-outcome association, conditional on the measured covariates[164, 193]. For observed risk ratio (RR) of observed exposure-outcome association, the E value is calculated as follows [164, 193-195]:

$$E\ value = RR_{obs} + \sqrt{RR_{obs}(RR_{obs} - 1)}. \quad (4.18)$$

For $RR_{obs} < 1$, we use $1/RR_{obs}$ in the formula.

The larger E value implies strong association between exposure and outcome, which suggests that the unmeasured cofounder-exposure and unmeasured confounder-outcome association should be

large to explain away exposure-outcome association. On the other hand, the smaller the E value indicates that weakly associated unmeasured E value is required to explain away the exposure-outcome association[193].

In this study, the adjusted risk ratio for two treatment levels, x_i and x_j is calculated using *adjrr* stata command [196], the point and confidence interval estimate of the E-value are estimated using e_value R function from tipr package[197, 198].

4.2.4 Potential Outcome Framework and Treatment Effect

For causal inference under the potential outcome framework [124], the potential outcome is the possible outcome under different treatment conditions.

If the treatment is continuous or count such as the number of women's ANC visits, the potential outcomes,$Y(t),\ t \in \mathcal{T},$ represent the outcome that would result if an individual is exposed to treatment level t that varies in the set $\mathcal{T}$. If a subject received treatment $T = t_i$, the observed potential outcome would be $Y(t_i)$ but the potential outcome, $Y(t_j),$ with alternative treatments, t_j for $i \neq j$ (all $k-1$ treatments) would be unobserved which would be a counterfactual outcome.

The change of the outcome for two different treatment levels for quantitative or multivalued treatments is $Y_i(t_j) - Y_i(t_i)$. However, we cannot observe the potential outcome under the treatment state for those observed in the alternative treatment state, and we are unable to observe the potential outcome under the alternative treatment state for those observed in the treatment state. This unobservable potential outcome makes us unable to estimate the change of treatment causal effect[199]. This is called "the fundamental problem of causal inference" as described by Holland [200].

Because it is impossible to calculate individual-level treatment effect change, we need to estimate average or population-level treatment effects change under the assumption of unconfoundness or exchangeability. For a particular treatment level $T = t$, average treatment effect (ATE) is: $ATE = E(Y(t))$, and for two different treatment levels, the change of average treatment effect is: $\Delta ATE = E\left(Y(t_j)\right) - E(Y(t_i)) = \mu(t_j) - \mu(t_i)$. This is the mean change of the outcome if all individuals are exposed to t_j instead of t_i[35].

4.2.5 Statistical Methods to Estimate Treatment Effect

Covariate adjustment using GPS

It is also called propensity score regression and comprises of regressing the outcome on the treatment and estimated GPS. Given Y is the outcome, T is the treatment and $R = r(t, \boldsymbol{X})$ is the estimated GPS, then a dose-response function which is the conditional expectation of Y given T and R is given by polynomial equation with degree two or more [26]:

$$E(Y/T, R) = \beta_0 + \beta_1 T + \beta_2 T^2 + \beta_3 R + \beta_4 R^2 + \beta_5 T. \quad (4.19)$$

This dose–response function indicates the average outcome if all subjects received treatment at $T = t$. The average treatment effect (ATE) at $T = t$ denoted by μ(t) is estimated as:

$$\hat{\mu}(t) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 r(t, \boldsymbol{X}_i) + \hat{\beta}_4 r^2(t, \boldsymbol{X}_i) + \hat{\beta}_5 t r(t, \boldsymbol{X}_i)\right). \quad (4.20)$$

$$\Rightarrow \hat{\mu}(t) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 \overline{r(t, \boldsymbol{X}_i)} + \hat{\beta}_4 \overline{r^2(t, \boldsymbol{X}_i)} + \hat{\beta}_5 t \overline{r(t, \boldsymbol{X}_i)}.$$

For two treatment levels t_i and t_j, the average change of treatment effect can be obtained by:

$$\Delta ATE = \mu(t_j) - \mu(t_i). \quad (4.21)$$

The average treatment effect on the treated (ATT) uses the same formula (4.19 and 4.20) except the average on the propensity score is obtained for each treatment level, $T = t$, for n_t study units instead of averaging for all study participants, n [201].

However, in our study, the outcome variable is ordinal. Thus, we used a cumulative link model for the dose-response function given in equation (4.12). Let $u_j = P(Y \leq j) = \pi_1 + \pi_2 + \cdots + \pi_j$, where $\sum_{j=1}^{J} \pi_j = 1$ are cumulative probabilities, π_j is probability in the j^{th} category, then the cumulative link dose-response model is given by:

$$g(u_j) = \beta_j - (\beta_1 T + \beta_2 T^2 + \beta_3 R + \beta_4 R^2 + \beta_5 TR). \quad (4.22)$$

Where $u_j = E(P(Y \leq j|T, R))$ and $g(.)$ is the link function.

4.2.6 Marginal Structural Model

Robins et al. [167] introduced marginal structural modeling where the causal effects of a treatment can be consistently estimated using IPTW which is obtained by the inverse of GPS. However, a direct inverse of GPS ($IPTW = 1/r(T, \boldsymbol{X})$) causes infinite variances and should not be used. Instead, stabilized weights $\left(wt = \frac{r(T)}{r(T,X)}\right)$ is used to estimate dose-response function. Once the GPS based stabilized weights is estimated, the weighted regression of the outcome on the exposure is modeled, and from the fitted outcome model, the average treatment effect can be estimated by calculating the expected outcome for each level of quantitative treatment[161]. When we weight

by the generalized propensity score, a sample in which covariates are balanced across all treatment groups is created, and then we estimate the average outcome at $T = t$ in the sample such that $E\{Y(t)\}$ [188].

The marginal structural model (MSM) for potential outcome, $Y(t)$ or weighted regression model for continuous outcome model is[27]:

$$E\big(Y(t)\big) = \beta_0 + \beta_1 t. \quad (4.23)$$

For ordinal outcome, the MSM is given by:

$$g(E(P(Y \leq j|t))) = \hat{\beta}_j - \hat{\beta}_1 t. \quad (4.24)$$

Where $g(.)$ is the link function, β_j is the weighted intercept of dose-response function for j^{th} group, and $\hat{\beta}_1$ is a weighted slop of dose-response function. The multiplicative change of two treatment effects at t_i and t_j, for $\gamma_j(t) = P(Y \leq j|t)$, say for probit link function is risk ratio (RR) of the outcome $Y \leq j$ at t_i and t_j.

$$\text{Risk ratio (RR)} = \gamma_j\big(t_j\big)/\gamma_j(t_i). \quad (4.25)$$

The risk ratio in equation (4.16) is equivalent to change of average treatment effect (at population level assuming all individuals received $t_i \; and \; t_j$ for weighted linear models. However, the treatment effect and risk ratio of two treatments on the treated is estimated by using weights at each treatment level, $T = t$.

In addition, the risk difference of the outcome at two treatment levels t_i and t_j is given by:

$$\text{Risk difference (RD)} = \gamma_j\big(t_j\big) - \gamma_j(t_i). \quad (4.26)$$

Comparison of methods

In order to compare the performance of treatment effect estimator methods, we used multiple performance measurement metrics including (root) mean square error, empirical standard error, model standard error, and bias. The formulas and description of these performance measure metrics are detailed in Morris et al. [68]. Using the performance measuring metrics; we selected robust method for treatment effect estimation method from the simulation study and actual data.

4.2.7 Simulation Study

Data generation

In this study, we used simulation study to compare the performance of methods to estimate the causal effect of count treatments on ordinal outcomes. The baseline covariates ($\boldsymbol{X}$), the treatment (T) and the outcome (Y) are generated in two scenarios.

In the first scenario, 10 baseline independent covariates, treatment and outcome are generated; in the second scenario, a mixture of independent and correlated covariates, treatment and outcome are generated as follows:

Scenario I:

- $X_1, X_2, X_6, X_9, X_{10} \sim N(0,1), X_3 \sim Bernoulli(0.5), X_4 \sim Bernoulli(0.3), X_5 \sim Bernoulli(0.09)$, $X_7 \sim uniform(0,3), X_8 \sim Poisson(1.5)$.
- $g\big(E(T|\boldsymbol{X})\big) = 0.02x_1 + 0.1x_2 + 0.05x_3 + x_4 + 0.12 * x_5 + 0.15 * x_6 + 0.11 * x_7 + 0.12 * x_8 + 0.22 * x_9 + 0.01 * x_{10}$.
- $E(Y|T, \boldsymbol{X}) = 0.5 * T + b1 * x_1 + b2 * x_2 + b3 * x_3 + b4 * x_4 + b5 * x_5 + b6 * x_6 + b7 * x_7 + b8 * x_8 + b9 * x_9 + b10 * x_{10}$.

Scenario II:

- $X_1 \sim N(0,1), X_2 \sim N(0,1) + 0.2X_1, X_3 \sim N(0,1), X_4 \sim N(0,1) + 0.5X_2, X_5 \sim N(0,1), X_6 \sim N(0,1) + X_5, X_7 \sim N(0,1) + 0.25X_5$, $X_8 \sim NN(0,1), X_9 \sim N(0,1)$, $X_{10} \sim N(0,1) + X_8$.
  - $g\big(E(T|\boldsymbol{X})\big) = 0.02x_1 + 0.1x_2 + 0.05x_3 + x_4 + 0.12 * x_5 + 0.15 * x_6 + 0.11 * x_7 + 0.12 * x_8 + 0.22 * x_9 + 0.01 * x_{10}$.
  - $E(Y|T, \boldsymbol{X}) = 0.5 * T + b1 * x_1 + b2 * x_2 + b3 * x_3 + b4 * x_4 + b5 * x_5 + b6 * x_6 + b7 * x_7 + b8 * x_8 + b9 * x_9 + b10 * x_{10}$.

Where $1 = log(1.25), b2 = log(1.5), b3 = log(1.75) b4 = log(2), b5 = log(1.25),\ b6 = log(1.5), b7 = \log(1.75)$, $b8 = \log(2)$, $b9 = \log(1.75)$, $b10 = \log(2)$.

In both scenarios $g(.)$ is the Poisson link function, $E(Y|T, \boldsymbol{X})$ is latent/unobserved continuous variable for ordinal category. The coefficient of outcome generation was adapted from Austin and Stuart [202] labeled as low, medium, high and very high. To generate ordinal outcome variable, first normal continuous latent variable was generated from normal distribution ($Y^* \sim N(E(Y\ |T, \boldsymbol{X}), 1)$); and then categorized into three ordered categories. All process of generating ordinal response variable was done according to Carsey and Harden [203].

Monte Carlo simulation

From large population of size 1000000 [161], we draw a random sample size of 500,1000, 3000 and 5000 which were adapted from Imai and Ratkovic [204]. Each sample size was repeated 1000 times and the average treatment effect on the outcome was estimated across 1000 iterations of the simulation for each sample size. The average treatment effect for 1000 iterations was obtained. In addition, model standard error, empirical standard error, mean square error (MSE) and biases were obtained to compare different methods of treatment effect estimate[68]. The true treatment effect was 0.5.

Description of the real data

To check the validity of methods from the simulation study, we estimated the causal effect of number of ANC contacts (treatment) on age-specific childhood vaccinations (outcome) for children less than five years.

Fifteen baseline confounders such as age category of mothers, region, place of residence, mothers' education level, ownership of radio and television, religion, household size, sex of household head, age of household head, wealth status, total child ever born, age at first birth, marital status and birth order number were controlled. A total of 5,150 sample women in the survey are considered in the study.

Due to the ordered nature of the outcome, we used cumulative link model as described above. On the other hand, the treatment variable is number of pregnant women's ANC contacts which is used as response variable for GPS estimation using a family of count models (Poisson model, Negative binomial model, zero inflated models) and GBM.

4.2.8 Result

Result of simulation study

Test of positivity assumption

The result in Table 4.4 presents the summary of GPS and GPS-based IPTW before and after trimming GPS for the independent covariates and a mix of the independent and correlated covariates. Trimming is done to have non-zero and non-one probability without losing more

information. The assumption is almost violated in all sample sizes since the minimum value of GPS is almost zero in all cases. Besides, some GPS-based weights are inflated in some cases such as at 4000 and 5000 sample sizes for both covariate cases. The IPTW at the independent covariate case is larger, especially at higher sample sizes, than the IPTW in the other covariates case. As a result, GPS is trimmed which the summary is also shown in Table 4.4. After trimming, the positivity assumption is satisfied. In addition, the GPS-based IPTW is gets smaller than before trimming. This makes the IPTW to be stable.

Table 4.4. Distribution of GPS and GPS based weight.

Sample size	Scenario I							
	Before trimming				After trimming			
	GPS		IPTW		GPS		IPTW	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
500	0.0004	0.689	0.0003	44.52	0.011	0.689	0.003	24.02
1000	0.003	0.535	0.00001	13.16	0.010	0.535	0.0001	11.54
2000	0.0012	0.629	0.00001	13.32	0.0103	0.629	0.00001	11.33
3000	0.0004	0.559	0.0001	41.62	0.010	0.559	0.00001	11.32
4000	0.0002	0.590	0.0000	90.30	0.010	0.590	0.0000	20.23
5000	0.0004	0.606	0.0000	61.78	0.0102	0.606	0.0000	10.11
	Scenario II							
500	0.003	0.962	0.0000	23.35	0.011	0.962	0.000	23.35
1000	0.001	0.984	0.0000	14.74	0.0103	0.984	0.0000	10.42
2000	0.0011	0.984	0.0000	44.67	0.011	0.984	0.0000	12.49
3000	0.001	0.993	0.0000	75.68	0.0105	0.982	0.0000	22.70
4000	0.0001	0.981	0.0000	165.89	0.0109	0.981	0.0000	18.88
5000	0.0002	0.986	0.0000	242.96	0.0101	0.986	0.0000	25.83

Comparison of outcome models (MSM vs GPS as covariate)

The results of the simulation study for the ordinal outcome are presented in Table 4.5. The results correspond to independent covariates (Scenario I) and a mixture of independent and correlated covariates (Scenario II). The findings indicated that the model standard errors in MSM for Scenario I are smaller than those in GPS across all sample sizes. Similarly, except at $n = 3000$, the standard errors in MSM for Scenario II are smaller than those in GPS, although the differences are not substantial for either scenario. The MSE of MSM for Scenario I is smaller than that of GPS across all sample sizes, except at $n = 2000$. Likewise, the MSE of MSM for Scenario II is smaller than that of GPS across all sample sizes except $n = 500$. Furthermore, the absolute value of the bias of the treatment effect under MSM is smaller than that under GPS for Scenario I across all sample

sizes, except at $n = 2000$. For Scenario II, the absolute bias of MSM is smaller than that of GPS except at $n = 500$.

We also reran the simulation study for the continuous outcome from which the ordinal outcome was generated, and the results are revealed in Table 4.6. Contrary to the ordinal outcome, the model standard error of MSM is higher than GPS as covariates in all sample sizes except n=2000 and both scenarios. The MSE of MSM under scenario I is smaller than that of GPS as a covariate for all sample sizes. However, the MSE of MSM under scenario II is higher than that of GPS as a covariate except when n=2000 and n=3000. The absolute value of the bias of treatment effect estimation using MSM is smaller than that of GPS as a covariate for all sample sizes under scenario I and at the sample sizes of n=2000 and n=3000 (Table 4.6).

The results of the simulation study presented in Tables 4.5 and 4.6 indicate that the MSE and the magnitude of biases for treatment effect estimation are consistently smaller in the MSM than in the generalized propensity score (GPS) model when covariates are independent, and this holds for both ordinal and continuous outcomes. Under a mixture of independent and correlated covariates, the performance of the MSM in terms of model standard error, MSE, and bias is not consistent across ordinal and continuous outcomes. This finding implies that the MSM performs well in terms of model performance metrics for ordinal and continuous outcomes when the covariates are generated independently. Similarly, the MSM performs well for an ordinal outcome when the covariates are a mixture of independent and correlated variables; however, its performance requires further verification for continuous outcomes.

Table 4.5. Results of the simulation study for ordinal outcome under scenario I and II.

		Scenario I					Scenario II				
n	Outcome models	Treatment effect	model std.err	EMPSE	MSE	Bias	Treatment effect	model std.err	EMPSE	MSE	Bias
500	MSM	-0.452	0.054 (0.0001)	0.045 (0.001)	0.908 (0.003)	-0.952 (0.001)	-0.424	0.061 (0.000)	0.032 (0.001)	0.854 (0.002)	-0.923 (0.001)
	GPS as covariate	-0.459	0.057 (0.001)	0.040 (0.001)	0.922 (0.002)	-0.959 (0.001)	-0.410	0.057 (0.001)	0.030 (0.001)	0.829 (0.00)	-0.910 (0.001)
1000	MSM	-0.385	0.035 (0.000)	0.032 (0.001)	0.784 (0.002)	-0.885 (0.001)	-0.477	0.042 (0.000)	0.040 (0.001)	0.957 (0.002)	-0.977 (0.001)
	GPS as covariate	-0.415	0.039 (0.000)	0.025 (0.001)	0.838 (0.001)	-0.915 (0.001)	-0.503	0.044 (0.001)	0.031 (0.00)	1.01 (0.002)	-1.00 (0.001)
2000	MSM	-0.518	0.024 (0.000)	0.025 (0.001)	1.037 (0.002)	-1.018 (0.001)	-0.426	0.028 (0.000)	0.028 (0.001)	0.858 (0.002)	-0.926 (0.001)
	GPS as covariate	-0.425	0.029 (0.000)	0.022 (0.001)	0.857 (0.001)	-0.925 (0.001)	-0.452	0.029 (0.000)	0.022 (0.005)	0.907 (0.001)	-0.952 (0.001)
3000	MSM	-0.354	0.019 (0.000)	0.020 (0.001)	0.729 (0.001)	-0.854 (0.001)	-0.449	0.024 (0.000)	0.028 (0.001)	0.902 (0.002)	-0.949 (0.009)
	GPS as covariate	-0.464	0.024 (0.000)	0.021 (0.001)	0.929 (0.001)	-0.964 (0.001)	-0.468	0.024 (0.000)	0.023 (0.001)	0.938 (0.001)	-0.968 (0.001)
4000	MSM	-0.353	0.017 (0.000)	0.020 (0.004)	0.728 (0.001)	-0.853 (0.001)	-0.418	0.021 (0.000)	0.020 (0.001)	0.844 (0.001)	-0.918 (0.001)
	GPS as covariate	-0.422	0.020 (0.000)	0.019 (0.004)	0.851 (0.001)	-0.922 (0.001)	-0.495	0.022 (0.000)	0.021 (0.001)	0.991 (0.001)	-0.995 (0.001)
5000	MSM	-0.340	0.015 (0.000)	0.017 (0.004)	0.705 (0.001)	-0.840 (0.001)	-0.458	0.018 (0.000)	0.021 (0.001)	0.918 (0.001)	-0.958 (0.001)
	GPS as covariate	-0.444	0.018 (0.000)	0.018 (0.004)	0.892 (0.001)	-0.944 (0.001)	-0.492	0.019 (0.000)	0.019 (0.004)	0.984 (0.001)	-0.991 (0.001)

Table 4.6. Results of the simulation study for continuous outcome under scenario I and II.

n	Estimates	Scenario I					Scenario II				
		Treatment effect	model std.err	EMPSE	MSE	Bias	Treatment effect	model std.err	EMPSE	MSE	Bias
500	MSM	0.639	0.056 (0.001)	0.041 (0.001)	0.021 (0.004)	0.139 (0.001)	0.826	0.086 (0.001)	0.044 (0.001)	0.109 (0.001)	0.326 (0.001)
	GPS as covariate	0.666	0.046 (0.000)	0.027 (0.001)	0.028 (0.003)	0.166 (0.001)	0.612	0.038 (0.000)	0.020 (0.004)	0.013 (0.001)	0.112 (0.001)
1000	MSM	0.542	0.041 (0.000)	0.034 (0.001)	0.003 (0.001)	0.042 (0.011)	0.925	0.048 (0.000)	0.039 (0.001)	0.182 (0.001)	0.425 (0.001)
	GPS as covariate	0.683	0.033 (0.000)	0.019 (0.004)	0.034 (0.002)	0.183 (0.001)	0.600	0.014 (0.000)	0.008 (0.002)	0.010 (0.001)	0.100 (0.002)
2000	MSM	0.601	0.025 (0.000)	0.000 (0.000)	0.010 (0.000)	0.101 (0.000)	0.818	0.030 (0.000)	0.000 (0.000)	0.102 (0.000)	0.319 (0.000)
	GPS as covariate	-3.13	0.480 (0.002)	0.279 (0.006)	13.27 (0.064)	-3.63 (0.009)	-3.70	0.176 (0.001)	0.095 (0.002)	17.62 (0.025)	-4.20 (0.003)
3000	MSM	0.585	0.024 (0.0000)	0.018 (0.004)	0.008 (0.001)	0.085 (0.001)	0.814	0.028 (0.000)	0.022 (0.005)	0.099 (0.004)	0.314 (0.001)
	GPS as covariate	0.045	0.006 (0.000)	0.002 (0.001)	0.570 (0.001)	-0.755 (0.001)	0.033	0.003 (0.000)	0.001 (0.000)	0.588 (0.000)	-0.767 (0.000)
4000	MSM	0.579	0.021 (0.000)	0.016 (0.003)	0.006 (0.001)	0.079 (0.005)	0.819	0.026 (0.000)	0.019 (0.004)	0.102 (0.004)	0.319 (0.001)
	GPS as covariate	0.658	0.017 (0.000)	0.010 (0.002)	0.025 (0.001)	0.158 (0.001)	0.598	0.008 (0.000)	0.004 (0.001)	0.010 (0.000)	0.098 (0.001)
5000	MSM	0.591	0.019 (0.000)	0.014 (0.003)	0.008 (0.001)	0.091 (0.004)	0.849	0.022 (0.000)	0.017 (0.004)	0.122 (0.004)	0.349 (0.005)
	GPS as covariate	0.705	0.015 (0.000)	0.009 (0.002)	0.042 (0.001)	0.205 (0.003)	0.606	0.007 (0.000)	0.004 (0.001)	0.011 (0.000)	0.106 (0.001)

4.2.9 Result of the Real Data

Comparison of GPS models based on log-likelihood ratio test

The comparison of candidate treatment models is done using the log-likelihood ratio test for non-nested models. The result of the likelihood ratio test in Table 4.7 showed that the zero-inflated Poisson model is better than GLM with Poisson and negative binomial link functions, and the negative binomial link function is better than the Poisson link function.

Table 4.7**.** Comparisons of count models for the actual data.

	Vuong z-statistic	H_A	p-value
Raw	-3.131	negative binomial > Poisson	0.00087
AIC-corrected	-3.131	negative binomial > Poisson	0.00090
BIC-corrected	-3.131	negative binomial > Poisson	0.00087
Raw	-19.905	Zero inflated Poisson > Poisson	< 2.2e-16
AIC-corrected	-18.886	Zero inflated Poisson > Poisson	< 2.2e-16
BIC-corrected	-15.551	Zero inflated Poisson > Poisson	< 2.2e-16
Raw	-21.832	Zero inflated Poisson >negative binomial	<2.2e-16
AIC-corrected	-20.670	Zero inflated Poisson > negative binomial	< 2.2e-16
BIC-corrected	-16.866	Zero inflated Poisson > negative binomial	< 2.2e-16

Vuong Non-Nested Hypothesis Test-Statistic:
(test-statistic is asymptotically distributed N(0,1) under the
null that the models are indistinguishible)

The distribution of GPS and GPS-based IPTW is presented in Table 4.8. The result shows that positivity assumption is hardly met before trimming GPS. As a result we trimmed the GPS at 0.01 from bottom and 0.99 from top.

Table 4.8. Distribution of GPS and stabilized weights for the actual data.

model	Before trimming				After trimming			
	GPS		IPTW		GPS		IPTW	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
Poisson	0.00004	0.495	0.003	24.001	0.0101	0.470	0.004	14.09
NB	0.0002	0.473	0.09	28.275	0.0101	0.473	0.09	9.63
ZIP	0.0006	0.941	0.941	0.0253	0.010	0.941	0.025	18.78
GBM	0.00004	0.3589	0.0103	14.736	0.0102	0.359	0.0103	7.98

GPS=Generalized Propensity score, IPTW= Inverse probability of treatment weighting,
NB=Negative Binomial, ZIP= Zero inflated Poisson Model

Comparison of GPS and outcome models in terms of treatment effect estimation

The results to compare GPS and outcome models using performance metrics are revealed in Table 4.9. In MSM, all GPS models result in almost similar performance metrics with little variation. However, the treatment effect in GBM-based GPS-IPTW is higher than the other GPS models. On the other hand, when we use GPS as a covariate in the outcome model, the GBM-based GPS results better performance as compared to other count models. Comparing outcome models, GPS as a covariate performs better than MSM when GPS is estimated by Poisson regression. However, for other GPS models, MSM performs better than GPS as a covariate with smaller model standard error, narrower confidence interval, smaller RMSE, and smaller absolute bias. While ZIP and GBM give comparable performance metrics in MSM, the effect of the number of ANC on age-specific childhood vaccination is higher in GBM than in ZIP. As a result, it is good to use GBM as a treatment model and MSM as an outcome model.

Table 4.9**.** Result of number of ANC on age specific childhood vaccination

GPS models	Outcome models	ANC effect	model std.err	p value	CI.ll	CI.ul	length of CI	MSE	RMSE	Bias
Poisson	MSM	0.118	0.012	< 2e-16	0.095	0.141	0.046	0.002	0.040	-0.016
	GPS as covariate	0.187	0.010	< 2e-16	0.167	0.208	0.041	0.004	0.067	-0.013
negative binomial	MSM	0.133	0.010	< 2e-16	0.113	0.153	0.040	0.005	0.069	-0.009
	GPS as covariate	0.193	0.011	< 2e-16	0.171	0.215	0.044	0.004	0.066	-0.014
ZIP	MSM	0.111	0.010	< 2e-16	0.091	0.132	0.041	0.002	0.040	-0.011
	GPS as covariate	0.138	0.012	<2e-16	0.115	0.161	0.045	0.006	0.079	-0.017
GBM	MSM	0.138	0.012	<2e-16	0.116	0.161	0.046	0.002	0.040	-0.011
	GPS as covariate	0.206	0.011	< 2e-16	0.185	0.226	0.041	0.006	0.078	-0.016

In addition to multiple model performance comparison metrics, the two outcome models were also compared using AIC, BIC and deviance values. The result in Table 4.10 conveys the AIC and BIC values of MSM are smaller than GPS as covariate. In addition, the deviance of MSM is smaller

than GPS as covariate. These statistical model comparison techniques implied MSM is better than GPS as covariate cumulative link outcome models.

Table 4.10**.** Comparison of outcome models for GBM based GPS models.

Outcome models	no of parameters	AIC	BIC	Deviance
MSM	3	5115.1	5134.5	5109
GPS as covariate	4	5389.5	5415.6	5381.4

Sensitivity analysis

The result of the sensitivity analysis for the chosen number of ANC contacts and all outcome levels is presented in Table 4.11. The sensitivity analysis is done using the point and 95% confidence interval of the E-values calculated from the risk ratio. No ANC contact is considered as a control group, and one or more ANC contacts are considered as exposure groups. For the one ANC contact relative to no ANC contact, the E-values are 1.70 (95% CI: 1.62, 1.77), 1.28 (95% CI: 1.23, 1.32), and 2.22 (95% CI: 2.01, 2.44) for no vaccination, incomplete vaccination, and fully/complete vaccination, respectively. For four ANC contacts relative to no ANC contacts, the E-values are 4.02 (95% CI: 3.43, 4.68), 1.56 (95% CI: 1.47, 1.64), and 7.06 (95% CI: 5.44, 9.11) for no vaccination, incomplete vaccination, and complete vaccination, respectively. A similar result is also presented for eight ANC contacts relative to no ANC contacts. The result shows that unmeasured covariates should be strongly correlated with the outcome and the exposure to nullify the exposure-outcome association, especially for higher-order outcome levels.

Table 4.11. Result of sensitivity analysis using E value.

	Treatment levels (number of ANC contacts)			
Outcome levels	0 (control)	1	4	8
0(No vaccination)		1.70 (1.62,1.77)	4.02 (3.43, 4.68)	13.03 (9.02, 18.73)
1(Incomplete vaccination)		1.28 (1.23, 1.32)	1.56 (1.47,1.64)	1.55 (1.48, 1.61)
2(Fully/completvaccination)		2.22 (2.01, 2.44)	7.06 (5.44, 9.11	21.61 (14.48, 32.14)

Estimating the effect of number of ANC contacts on age-specific childhood vaccination

The proportional odds assumption is done using the nominal_test R function, and we got the result of likelihood ratio test statistics 2.33 and p-value 0.127. This shows that the proportional odds assumption is not violated. Thus, the treatment effect is constant across all levels of the outcome. After fitting MSM using the probit link function, the estimated dose-response function is given as follows:

MSM: $E(g(u(t)) = \mathbf{\Phi}^{-1}(u(t)) = \hat{\beta}_j - 0.138t$, where $u(t) = P(Y \leq j|t), j = 0, 1, 2$. Then the estimated probability to be in jth category is $\hat{u}(t) = P(\hat{y} \leq j|t) = \mathbf{\Phi}(\hat{\beta}_j - 0.138t)$ where $\mathbf{\Phi}$ is the standard normal cumulative distribution functions. The positive coefficient indicates the treatment (number of ANC contacts) increases the high ordered category of the response (age- specific childhood vaccination), i.e., the likelihood of being in higher category increases when number of ANC increases. For two treatment (number of ANC contacts) values t_i and t_j, the estimated risk change of cumulative probability is $P(\hat{y} \leq j|t_j - t_i) = \mathbf{\Phi}(-0.138(t_j - t_i))$.

Similarly, the estimated risk ratio of two treatments values, t_i and t_j, is:

$P(\hat{y} \leq j|t_j)/P(\hat{y} \leq j|t_i) = \mathbf{\Phi}(\hat{\beta}_j - 0.138t_j)/\mathbf{\Phi}(\hat{\beta}_j - 0.138t_i)$, where $\hat{\beta}_j = \{-0.691, 2.32\}$

The results of the adjusted relative effect of one or more ANC contacts with respect to no ANC contacts are presented in Table 4.12. The result shows that when the number of ANC contacts increased, the probability of getting timely vaccination increased. For example, one ANC contact decreased the chance of no vaccination by 16.9% and increased the chance of partial and full vaccination by 5% and 43.2%, respectively. Four ANC contacts decreased the chance of no vaccination by 55.4% and increased the chance of partial and full vaccination by 15% and 3.8 times of no vaccination, respectively (Table 4.12). The same is true when the number of ANC contacts increased; the chance of getting no vaccination at the scheduled time decreased, and the chance of all vaccines being administered at the right schedule increased.

Table 4.12. Result of risk ratio (RR) of one or more ANC contacts relative to no ANC contacts.

	Outcome levels

Treatment	0		1		2	
combination	RR	95% CI	RR	95% CI	RR	95% CI
0 vs1	0.831 (0.011)	0.809, 0.853	1.05(0.006)	1.0373, .0620	1.432(0.05)	1.34, 1.53
0 vs 4	0.436 (0.030)	(0.381,0.498)	1.15(0.017)	(1.114, 1.181)	3.80(0.461)	(2.99, 4.82)
0 vs 6	0.262(0.0309)	(0.208, .330)	1.17(0.017)	1.13, 1.198	6.70(1.106)	(4.85, 9.26)
0 vs 8	0.148 (0.026)	0.104, 0.209	1.143(0.013)	1.117, 1.169	11.06(2.195)	(7.5, 16.32)
0 vs 11	0.055 (0.016)	(0.031, .098)	1.039(0.025)	(0.992, 1.089)	20.89 (4.80)	(13.31, 32.77)

4.2.10 *Discussion and Conclusion*

In the study, we focused on selecting the appropriate outcome model for an ordinal outcome and estimating the causal effect of a count treatment. A Monte Carlo simulation study with sample sizes of 500, 1000, 3000, and 5000, randomly selected from the population of size 1000000[161] and household survey data from the Ethiopian Demographic and Health Survey (EDHS 2019) were used. The treatment variable is generated from the Poisson distribution, and the outcome variable is generated from a normally distributed latent variable with a mean of the linear combination of treatment and covariates and a standard deviation of one. The simulation was done for independent covariates as well as a mixture of independent and correlated covariates. We also repeated the simulation for a continuous outcome to determine whether or not our finding is consistent for a continuous outcome. GPS as a covariate[26] and Marginal structural modeling [167] were used to fit the dose-response function.

For both simulated and real data, the assumption of positivity is hardly met (minimum GPS∼0 and maximum GPS∼1). To meet the assumption and avoid extreme weights that result biased treatment effects, the GPS is trimmed at 1st percentile from the bottom and 99th percentile from the top[62, 63]. This resulted stable weight, better covariate balance, and effective sample size. Trimming increases the proportion of overlap propensity score distribution and reduces unmeasured confounders[205]. Trimming may be important, especially for IPTW but needs to routinely perform sensitivity analysis based on trimming for any propensity score based analysis[65].

IPTW is sensitive to misspecification of propensity score model that will result bias treatment effect [65, 191, 206]. On the other hand, the better predictive model may not give an actual treatment effect estimate[207, 208]. As a result, we compared GPS candidate models in terms of

treatment effect estimation for the actual data. In addition to parametric count models, we used GBM that can automatically consider non-linear and interaction terms, and choose important covariates[27, 151].

In the simulation study, it was found that the model performance metrics in MSM were less than GPS as covariates for the ordinal outcome under the two covariate scenarios. Besides, the MSE and magnitude of bias in MSM were less than GPS as covariate in the continuous outcome and independent covariates. However, the result was inconclusive in continuous outcome under the second scenario of the covariates.

For actual data, the zero inflated and the GBM GPS models performed comparably in terms of model performance metrics. However, the treatment effect size is different in both GPS models where zero inflated model over estimates treatment effect for MSM outcome model and under estimates in GPS as covariate outcome model.

For the actual data, we chose the probit link function similar to simulation study for cumulative link model of age-specific childhood vaccination [112]. It was found that model performance metrics in MSM were less than covariate adjustment using GPS. Stürmer T. et al. [65] recommended not to use propensity score as continuous covariates since its validity depends on correctly specifying two models (Propensity score and outcome models). They concluded that it does not give the specific advantage as compared to other propensity score implementations (weighting, matching, and stratification). Another study showed that the bias produced by propensity score regression adjustment was smaller than propensity score weighting. However, the two propensity score methods perform better when the sample size is moderate and large [188]. Matching, stratification and IPTW based on propensity score differ from covariate adjustment using propensity score in that they separate the design of the study from the analysis of the study; this separation does not occur when we use covariate adjustment using propensity score [3].

The study also showed that the number of ANC services enhances the probability of timely childhood vaccination.

To sum up, while estimating GPS for count treatments, it is important to use GBM, whether the GPS model is correctly specified or misspecified using parametric models. It is also important to use MSM to accurately estimate the count treatment effect on an ordinal outcome.

Finally, it was found that the number of ANC contacts increased the probability of age-specific complete childhood vaccination. This means when a woman gets more ANC services, it is likely she will vaccinate her child on the vaccination schedule. This could be because a woman gets more ANC services; it is more likely that she will deliver at a health facility, and the newborn baby gets a vaccine at birth, as well as a woman getting education on when and where she will vaccinate her baby.

In some scenarios, there can be post-treatment confounders, also called mediators that are affected by the treatment. These post-treatment confounders are in the causal pathway between the treatment and the outcome, where some or all of the treatment effect goes through the mediator(s). Hence, chapter four does not consider the role of mediator(s) that divide the treatment effect into direct and indirect effects. Chapter five addresses the effect of single and causally ordered mediators for the effect of count treatment on ordinal outcome.

CHAPTER FIVE: CAUSAL MEDIATION ANALYSIS

5.1. Direct and Indirect Effect of Count Treatments on Ordinal Outcomes

5.1.1 Background

Mediation analysis is commonly used in many fields of study such as epidemiology, medicine, social science, economics, agriculture and health sciences[209]. It aims to determine the extent to which the effect of an exposure on an outcome is accounted for by a specific set of mediators along the causal pathway connecting the exposure to the outcome[29].

Mediation analysis disentangles the total causal effect of an exposure on outcome into the effect through mediator(s) also called indirect effect) and the effect not through mediator(s) also called direct effects. When used by Baron and Kenny [30], mediation analysis focused on linear models where there is no exposure-mediator interaction. Nevertheless, Robins and Greenland [31] proposed counterfactual framework for causal inference analysis, which uses non-parametric approach to estimate direct and indirect effect of the exposure. The counterfactual framework allows interaction between exposure and mediator(s) as well as non-linear models.

Since the introduction of counterfactual framework for mediation analysis, abundant methodological methods have been proposed for causal mediation analysis that allows additive and multiplicative scale, and non-linear models for survival data [32-37].

This study investigated the direct and indirect effect of number of ANC contacts on age-specific childhood vaccination when mediated by institutional delivery. ANC is a routine health care service offered by skilled professionals to pregnant women assuming mothers' and newborn baby's health is in optimum condition[48]. At the time of ANC visit, other than examinations and medication, health education and counseling about postnatal care and child immunization are given to pregnant women. A study showed that the decision on child vaccination begins at the time of pregnancy, and childhood complete vaccination is associated with maternal health care utilization such as ANC [210]. Noh Won et al[211] reported that mothers who attended a minimum of four ANC visits are more likely for their children to complete basic vaccination. Adedire et al[212] also

suggested that when pregnant women who attended ANC services would improve childhood vaccination and consequently childhood immunization coverage.

Studies conducted in Ethiopia at national and sub-national levels [50-54] reported that ANC follow-up increased the likelihood of childhood vaccination. On the contrary Lakew, Bekele, and Biadgilign [55] reported that ANC did not affect childhood vaccination status. However, studies conducted so far have identified the effect of ANC on childhood vaccination status regardless of when it was given though childhood vaccination is recommended to be given right after birth within a specified schedule and time range[56]. In addition, there was lack of literature on the direct and indirect effect of ANC on age-specific childhood vaccination when mediated by institutional delivery.

Institutional delivery is considered the mediator or can be called post-treatment confounder. It is shown that skilled birth assistance during delivery[85, 213], and delivery services[214] are associated with complete childhood vaccination status. Hence, this study is done to estimate the direct, indirect, and total effect of number of ANC contacts on age-specific childhood immunization status when mediated by institutional delivery.

5.1.2 Data and Methods

Description of variables

Four types of variables are considered in the study. These are outcome, exposure, mediator and confounding variables and each of them are described as follows.

Outcome variable

The outcome variable was age-specific childhood vaccination status administered according to the immunization schedule specified in the Ethiopian immunization schedule[49].

Treatment variable

The treatment variable is the number of pregnant women's ANC contacts. The exact number of ANC contacts is considered and its effect on age-specific childhood vaccination status under mediator is examined.

Mediator variable

Institutional delivery is considered as a mediator variable. The assumption put in the study is ANC associated with institutional delivery and outcome, and institutional delivery associated with age-specific childhood vaccination status.

Confounding variables

The net effect of ANC and subsequent mediator on childhood immunization is difficult to evaluate because women who visited health center for ANC service are different in terms of observed and/or unobserved maternal and household characteristics[215]. These characteristics not only limited or encouraged women not to visit or to visit health facilities for ANC service but also affects the childhood vaccination status. As a result, these variables can be a source of confounding bias that need to be controlled.

In this study covariates associated with the outcome [216], may or may not be associated with exposure and mediator, which are real confounders are controlled using cumulative link models for exposure-outcome and mediator-outcome relationship, as well as logistic regression for the exposure-mediator relationship. A threshold value of p=0.2 was used for a bivariate relationship and then included in a multivariable relationship [9, 98]. Based on the threshold value age of household head, age of mother at first birth, household size, total child ever born, birth order number, region, place of residence, mother's religion, age of mother at recent birth, mother's education status, ownership of radio and television, and household wealth status.

5.1.3 Mediation Analysis

Mediation analysis aims to decompose the causal effect of an exposure on the outcome into a direct effect that does not go through a mediator and an indirect effect that goes through a mediator (intermediate outcome). Baron and Kenny[30] explained statistical mediation analysis and their work has been influential for several years. They showed how to identify the direct and indirect effects using sequential regression analysis. However, the original Baron and Kenny (1986) method did not consider covariates and the interaction of exposure and mediator as well as non-linear relationship between exposure and outcome[217].

Techniques of mediation analysis increased quickly over the past decades. Literature in causal inference has elucidated the assumptions associated with the estimation of direct and indirect

effects and prolonged the traditional methods to settings of potential outcomes that include the interaction of exposure and mediator as well as non-linear relationships.

For this study Y is age-specific childhood vaccination (ordinal), M is place of delivery, T is number of ANC contacts, $\mathbf{C}$ is a matrix of base line confounders, then the causal direct acyclic graph is presented in Figure 5.1.

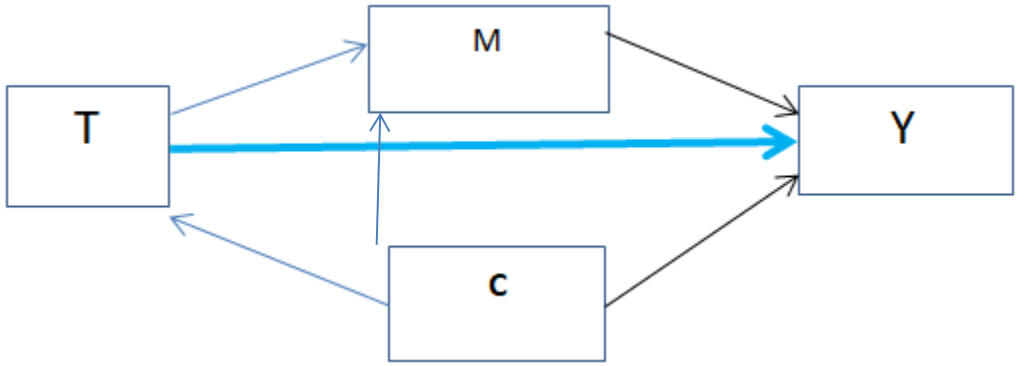


Figure 4.1. Causal DAG for single mediator.

Given the specification of confounders, exposure, mediator, and outcome, the mediator and outcome models are specified as follows with treatment and mediator interaction effect:

$$g(p(M = 1|T, \boldsymbol{C})) = \alpha_0 + \alpha_1 t + \boldsymbol{\alpha_2}^t \boldsymbol{C}\,. \quad (5.1)$$

$$g\big(P(Y \leq j|t, m, \boldsymbol{C})\big) = \beta_j - \big(\beta_1 t + \beta_2 m + \beta_3 tm + \boldsymbol{\beta_2}^t \boldsymbol{C}\big). \quad (5.2)$$

$g(.)$ is the link function for binary and ordinal outcome, and t is the transpose. The parameters in the models are estimated with maximum likelihood estimation.

For the above mediation and outcome model specification, and to find the controlled direct effect (CDE), natural direct effect (NDE), and natural indirect effect (NIE), the following identification assumptions are assumed true. There is no unmeasured confounding for exposure-outcome, mediator-outcome, and exposure-mediator relationships. In addition, no confounder influences the mediator that is affected by the exposure.

Estimating natural direct and indirect effects

For binary mediator and ordinal outcome, the natural direct effect (NDE), the natural indirect effect (NIE), controlled direct effect (CDE) and total effect (TE) are estimated with odds ratio scale. Under identification assumption and rare outcome assumption holds true, the NIE, NDE and CDE can be obtained as [41, 43]:

$$NIE^{OR} = \frac{1+\exp(\alpha_0+\alpha_1 t^*+\boldsymbol{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t+\boldsymbol{\alpha_2}^T\boldsymbol{C})} \times \frac{1+\exp(\beta_2+\beta_3\boldsymbol{t}+\alpha_0+\alpha_1 t+\boldsymbol{\alpha_2}^T\boldsymbol{C})}{1+\exp(\beta_2+\beta_3\boldsymbol{t}+\alpha_0+\alpha_1 t^*+\boldsymbol{\alpha_2}^T\boldsymbol{C})}. \quad (5.3)$$

$$NDE^{OR} \approx \frac{1+\exp(\alpha_0+\alpha_1 t^*+\boldsymbol{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t+\boldsymbol{\alpha_2}^T\boldsymbol{C})} \times \frac{1+\exp(\beta_2+\beta_3\boldsymbol{t}+\alpha_0+\alpha_1 t+\boldsymbol{\alpha_2}^T\boldsymbol{C})}{1+\exp(\beta_2+\beta_3\boldsymbol{t}+\alpha_0+\alpha_1 t^*+\boldsymbol{\alpha_2}^T\boldsymbol{C})}. \quad (5.4)$$

$$CDE^{OR} = \exp[(\beta_1 + \beta_3 m)(t - t^*)] . \quad (5.5)$$

$$TE^{OR} = NDE \times NIE. \quad (5.6)$$

Proof

Using pearl [218] mediation formula:

$$E\left(P\left(Y_{tM_{(t)}} \leq j|\ \boldsymbol{C}=\boldsymbol{c}\right)\right) = \sum_{\boldsymbol{m}} E[Y \leq j|T=t, M=m, \boldsymbol{C}=\boldsymbol{c}]\ P(M=m|T=t, \boldsymbol{C}=\boldsymbol{c}). \quad (5.7)$$

And

$$E\left(P\left(Y_{tM_{t^*}} \leq j|\ \boldsymbol{C}=\boldsymbol{c}\right)\right) = \sum_{\boldsymbol{m}} E[Y \leq j|T=t, M=m, \boldsymbol{C}=\boldsymbol{c}]\ P(M=m|T=t^*, \boldsymbol{C}=\boldsymbol{c}).$$

In addition, under a rare outcome assumption [41, 43], we have:

$$\log\left[\frac{P(Y \leq j|t,m,\boldsymbol{C})}{1-P(Y \leq j|t,m,\boldsymbol{C})}\right] \approx \log\left(P(Y \leq j|x,m,\boldsymbol{C})\right) = \beta_j - \left[\beta_1 t + \beta_2 m + \beta_3 tm + \beta_4{}^T\boldsymbol{C}\right]. \quad (5.8)$$

Taking the reverse order of the inequality, we have $P(Y > j|t,m,\boldsymbol{C}) \approx \exp(-\beta_j + \beta_1 t + \beta_2 m + \beta_3 tm + \beta_4{}^T\boldsymbol{C})$.

using equations (5.7) and (5.8):

$$\log\{\frac{P\left(Y_{t,M_t} > j\middle|\boldsymbol{C}\right)}{P\left(Y_{t,M_t} \leq j\middle|\boldsymbol{C}\right)} \approx \log(P(Y_{t,M_t} > j|\boldsymbol{C}).$$

$$= \log\{\sum_m P(\ Y_{t,m} > j|M_t = m, \boldsymbol{C})\boldsymbol{P}(M_t = m|\boldsymbol{C})\}.$$

$$= \log\{\sum_m P(\ Y > j|t, m, \boldsymbol{C})\boldsymbol{P}(M = m|t, \boldsymbol{C})\}.$$

$$\approx \log\{\sum_m \exp(-\beta_j + \beta_1 t + \beta_2 m + \beta_3 tm + \beta_4{}^T\boldsymbol{C})\boldsymbol{P}(M = m|t, \boldsymbol{C})\}.$$

$$= \log\{\exp\left(-\beta_j + \beta_1 t+\beta_4{}^T\boldsymbol{C}\right)\sum_{\boldsymbol{m}} \exp\left(\beta_2 m + \beta_3 tm\right) P(M = m|t, \boldsymbol{C})\}.$$

$$= \log\{\exp\left(-\beta_j + \beta_1 t+\beta_4{}^T\boldsymbol{C}\right) E\left\{\exp\left(\left(\beta_2 + \beta_3 t\right)M\right)\middle|t, \boldsymbol{C}\right\}.$$

For a binary mediator,M,

$$\sum_m \exp(\beta_2 m + \beta_3 tm)\ P(M = m|t, C).$$

$$= \exp\left(\beta_2 * 0 + \beta_3 t * 0\right)P(M = 0|t, C)+\exp\left(\beta_2 * 1 + \beta_3 t * 1\right)P(M = 1|t, C).$$

Thus, $E\left\{\exp\left(\left(\beta_2+\beta_3 t\right)M\right)\middle|t,\boldsymbol{C}\right\}=\boldsymbol{exp}\left(\beta_2+\beta_3 t\right)P(M=1|t,C)+P(M=0|t,C)$.

We know that, $P(M=1|t,C)=\exp(\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})\Big/1+\exp(\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})$ and

$$P(M=0|t,C)=1-P(M=1|t,C)=\frac{1}{1+\exp(\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})}.$$

Then $E\left\{\exp\left(\left(\beta_2+\beta_3 t\right)M\right)\middle|t,\boldsymbol{C}\right\}$.

$$=exp\left(\beta_2+\beta_3 t\right)\frac{\exp(\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})}+\frac{1}{1+\exp(\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})}.$$

$$=\frac{1+\exp(\beta_2+\beta_3 t+\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})}.$$

Thus, we get

$$\log\{\frac{P\left(Y_{t,M_t}>j\middle|\boldsymbol{C}\right)}{P\left(Y_{t,M_t}\leq j\middle|\boldsymbol{C}\right)}\approx\exp\left(-\beta_j+\beta_1 t+{\beta_4}^T\boldsymbol{C}\right).\frac{1+\exp(\beta_2+\beta_3 t+\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t+{\alpha_2}^T\boldsymbol{C})}.$$

In the same way,

$$\log\{\frac{P\left(Y_{t,M_{t^*}}>j\middle|\boldsymbol{C}\right)}{P\left(Y_{t,M_t}\leq j\middle|\boldsymbol{C}\right)}\approx\exp\left(-\beta_j+\beta_1 t+{\beta_4}^T\boldsymbol{C}\right).\frac{1+\exp(\beta_2+\beta_3 t+\alpha_0+\alpha_1 t^*+{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t^*+{\alpha_2}^T\boldsymbol{C})}.$$

Similarly,

$$\log\{\frac{P\left(Y_{t^*,M_{t^*}}>j\middle|\boldsymbol{C}\right)}{P\left(Y_{t^*,M_{t^*}}\leq j\middle|\boldsymbol{C}\right)}\approx\exp\left(-\beta_j+\beta_1 t^*+{\beta_4}^T\boldsymbol{C}\right).\frac{1+\exp(\beta_2+\beta_3 t^*+\alpha_0+\alpha_1 t^*+{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t^*+{\alpha_2}^T\boldsymbol{C})}.$$

This leads to

$$NIE^{OR}=\frac{1+\exp(\alpha_0+\alpha_1 t^*+\boldsymbol{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t+\boldsymbol{\alpha_2}^T\boldsymbol{C})}\times\frac{1+\exp(\beta_2+\beta_3 t+\alpha_0+\alpha_1 t+\boldsymbol{\alpha_2}^T\boldsymbol{C})}{1+\exp(\beta_2+\beta_3 t+\alpha_0+\alpha_1 t^*+\boldsymbol{\alpha_2}^T\boldsymbol{C})}.$$

$$NDE^{OR}\approx\frac{1+\exp(\alpha_0+\alpha_1 t^*+\boldsymbol{\alpha_2}^T\boldsymbol{C})}{1+\exp(\alpha_0+\alpha_1 t+\boldsymbol{\alpha_2}^T\boldsymbol{C})}\times\frac{1+\exp(\beta_2+\beta_3 t+\alpha_0+\alpha_1 t+\boldsymbol{\alpha_2}^T\boldsymbol{C})}{1+\exp(\beta_2+\beta_3 t+\alpha_0+\alpha_1 t^*+\boldsymbol{\alpha_2}^T\boldsymbol{C})}.$$

$$CDE^{OR}=\exp[(\beta_1+\beta_3 m)(t-t^*)].$$

$$TE^{OR}=NDE\times NIE.$$

t^* is the treatment level contrary to t. When there is no interaction between treatment and mediator, β_3 will be zero.

Given the odds ratio of natural indirect effect and the odds ratio of natural direct effect and with rare outcome assumption, the proportion of exposure's effect mediated is given by [36]:

$$PM = \frac{OR^{NDE}(OR^{NIE}-1)}{OR^{NDE}*OR^{NIE}-1}. \quad (5.9)$$

5.1.4 Result

Mediation effect of the mediator before adjusting confounders

The effect of the mediator on each level of the treatment is shown in Table 5.1. The mediation effect is calculated using the odds ratio in the absence of treatment-mediator interaction. The control treatment level is no ANC contact, and at least one ANC contact is considered as treatment value. The effect of ANC first contact through institutional delivery (indirect effect) increased by 10%, and its natural direct effect increased by 31% as compared to no ANC contact. Similarly, the natural indirect effect of four ANC contacts on the outcome is 51%, and its natural direct effect increased by 2.94 as compared to no ANC contact. The natural indirect effect of eight ANC contacts is 80%, and the natural indirect effect of eight ANC contacts on the outcome increased by 8.67 as compared to no ANC contact.

The odds ratio of natural direct effect and total effects slowly increased from one to four ANC contacts and quickly increased afterward. The proportion of the odds ratio of the treatment effect mediated by mediators increased from 0.3 for first ANC contact to 0.44 for four ANC contacts, and 0.47 for eight ANC contacts (Table 5.1).

Table 5.1. Unadjusted direct, indirect, and total effects without treatment–mediator interaction.

Treatment (ANC) level 0 (control)	OR of NIE	OR of NDE	OR of TE	Proportion mediated
1	1.1	1.31	1.44	0.30
2	1.23	1.72	2.11	0.35
3	1.37	2.25	3.08	0.40
4	1.51	2.94	4.45	0.44
5	1.63	3.86	6.27	0.46
6	1.71	5.05	8.64	0.47
7	1.77	6.62	11.71	0.48
8	1.80	8.67	15.64	0.47
9	1.83	11.35	20.75	0.48
10	1.84	14.88	27.37	0.47
11	1.85	19.49	36.00	0.47

The result in Table 5.2 reveals exposure's natural indirect effect, natural direct effect, total effect, and controlled direct effect in the presence of treatment-mediator interaction before adjusting the confounders' effect. The interaction of treatment and mediator had a negative effect on the outcome and slowly decreased the likelihood of the indirect effect of the treatment through the mediator. However, the natural direct effect of the treatment on the outcome increased across treatment levels as compared to the control level i.e. zero. The odds ratio of the natural direct effect quickly increased after four ANC contacts, and after eight contacts, the odds ratio increased further (from 9.99 at eight contacts to 24.55 at eleven ANC contacts). The odds ratio of controlled direct effect at home delivery (mediator=0) increased for increasing exposure's level. Furthermore, given institutional delivery (mediator=1), the odds ratio of controlled direct effect increased across increasing exposure levels (OR=1.23, 2.32, 5.36 at first, four, and eight ANC contacts respectively). It looks like the odds ratio of controlled direct effect for health facility delivery is less than home delivery due to the negative effect of interaction effects.

Table 5.2. Unadjusted direct, indirect, and total effects with treatment–mediator interaction.

Treatment (ANC) level 0 (control)	OR of NIE	OR of NDE	OR of TE	CDE(m=0)	CDE(m=1)
1	1.12	1.32	1.49	1.38	1.23
2	1.25	1.75	2.18	1.9	1.52
3	1.34	2.33	3.13	2.61	1.88
4	1.39	3.1	4.32	3.6	2.32
5	1.39	4.14	5.76	4.95	2.86
6	1.34	5.54	7.44	6.82	3.53
7	1.27	7.43	9.44	9.39	4.35
8	1.18	9.99	11.81	12.93	5.36
9	1.09	13.46	14.67	17.81	6.62
10	1	18.16	18.16	24.53	8.17
11	0.91	24.55	22.42	33.78	10.07

Direct and indirect effect of the exposure when confounders are adjusted

The natural indirect effect (NIE), natural direct effect (NDE), and total effect (TE) of the exposure on the outcome after adjusting for the confounders' effect are measured with odds ratio. The results are presented in Table 5.3, which shows that the likelihood of natural indirect effect of ANC contact increased from 3% for first ANC contact to 14% for four ANC contacts and 31% for eight ANC contacts when compared to no ANC contact. Similarly, the odds ratio

of the natural direct effect of ANC contact on the higher-order vaccination status increased by 21% for the first ANC contact, by 2.14 for four ANC contacts, and by 4.57 for eight ANC contacts when compared to no ANC contact. On the other hand, the odds ratio of total exposure's effect on the outcome increased by 25% for the first ANC contact, by 2.62 for four ANC contacts, and by 5.99 for eight ANC contacts as compared to no ANC contact. Similarly, the proportion of exposure's effect mediated by institutional delivery increased from 0.15 for the first ANC contact to 0.21 for four ANC contacts and to 0.28 for eight ANC contacts (Table 5.3).

Table 5.3. Adjusted direct, indirect, and total effects with treatment–mediator interaction.

Treatment (ANC) level 0 (control)	OR of NIE	OR of NDE	OR of TE	Proportion mediated
1	1.03	1.21	1.25	0.15
2	1.06	1.46	1.55	0.16
3	1.1	1.77	2.17	0.19
4	1.14	2.14	2.62	0.21
5	1.18	2.58	3.17	0.23
6	1.23	3.13	3.84	0.25
7	1.27	3.78	4.80	0.27
8	1.31	4.57	5.99	0.28
9	1.35	5.53	7.46	0.30
10	1.38	6.68	9.24	0.31
11	1.41	8.08	11.39	0.32

We also added exposure-mediator interaction in the outcome model and estimated odds ratio (OR) of natural direct and indirect effects as well as total effect. We also estimated controlled direct effects for the mediator levels. The results are presented in Table 5.4. It shows that the odds ratio of the natural indirect effect did not vary much from the first to four contacts. Later to four ANC contacts, the odds ratio of the natural indirect effect decreased slowly from 1.1 to 0.86. The odds ratio of natural direct effect increased from 1.22 for the first ANC contact to 2.27 for four ANC contacts and to 5.26 for eight ANC contacts compared to no ANC contacts. Similarly, the odds ratio of total effect increased from 1.15 for the first contact to 2.49 for four contacts and to 5.34 for eight ANC contacts compared to no ANC contacts.

The odds ratio of the controlled direct effect of health facilities delivery increased from 1.15 for the first ANC contact to 1.75 for four ANC contacts, and to 3.06 for eight ANC contacts. Similarly,

the odds ratio of controlled direct effect of home delivery increased from 1.26 for first contact to 2.51 for four contacts, and to 6.30 for eight ANC contacts. The odds ratio of the controlled direct effect of institutional delivery is less than that of home delivery because of the negative effect of the exposure-mediator interaction on the outcome.

Table 5.4**.** Adjusted direct, indirect, and total effects with treatment–mediator interaction.

Treatment (ANC) level	OR of NIE	OR of NDE	OR of TE	OR of CDE for m=1	OR of CDE for m=0
0 (control)					
1	1.04	1.22	1.27	1.15	1.26
2	1.07	1.5	1.6	1.32	1.58
3	1.09	1.84	2.01	1.52	1.99
4	1.1	2.27	2.49	1.75	2.51
5	1.09	2.79	3.07	2.01	3.16
6	1.08	3.44	3.73	2.32	3.97
7	1.05	4.26	4.49	2.66	5.00
8	1.01	5.26	5.34	3.06	6.30
9	0.97	6.52	6.3	3.53	7.92
10	0.7	8.08	5.69	4.06	9.97
11	0.86	10.03	8.58	4.66	12.55

5.1.5 Discussion and Conclusion

The finding revealed that the four conditions of Baron and Kenny's [30] approach are satisfied. The exposure (number of ANC contacts) is significantly associated with the outcome even if it is not a necessary condition[219]; the exposure is significantly associated with the mediator (place of delivery); the exposure is significantly associated with the outcome controlling mediator effect; and the mediator is significantly associated with the outcome after controlling the exposure. Although the associations are all positive, the association of exposure-mediator interaction is negative and significant.

In the study, no ANC is used as the control exposure value, and one or more contacts are considered as treatment values. The comparisons of ANC contacts effect are compared against no ANC contacts. After estimating the odds ratio of natural direct, indirect, and total effects as well as controlled indirect effects [41, 43] for each individual, we use the average to know the overall effect.

For the first ANC contact, the adjusted odds ratio is 1.03 through the mediator and 1.21 directly with the total odds ratio of 1.25. For four ANC contacts, the adjusted odds ratio of the indirect exposure's effect is 1.14 and that of direct exposure's effect is 2.14 with the total odds ratio of exposure's effect is 2.62. Similarly, the odds ratio of direct and indirect exposure's effect increased across increasing treatment values. Institutional delivery mediated 15% of the first ANC contact effect, 21% of four ANC contacts effect, and 28% of eight ANC contacts effect as compared to no ANC contact. This implies that the odds ratio of natural direct effect is higher than the odds ratio of natural indirect effects such that place of delivery partially mediated the effect of the number of ANC on timely childhood vaccination.

Due to the negative effect of exposure-mediator interaction, the odds ratio of natural indirect and total effects decreased for an increasing number of ANC contacts. However, the natural direct effect increased more than the model without the interaction effect. The odds ratio of controlled direct effect for both levels of the mediator increased for an increasing number of ANC contacts.

To conclude, the exposure and the mediator are found significantly associated with the outcome, and the exposure is also significantly associated with the mediator before and after adjusting the confounders' effect. The unadjusted effect of the exposure and mediator on the outcome is reduced by thirty-seven and forty-four percent after adding confounders in the regression which indicates the essence of adjusting confounders' effect before estimating treatment effects on the outcome.

Institutional delivery had a mediating role on the effect of ANC on age-specific childhood vaccination status. Because some portion of the ANC contact's effect went through the institutional delivery though the larger portion of its effect went directly to the outcome.

The exposure-mediator interaction negatively influences the outcome and reduced the indirect or mediation effect in turn increasing exposure direct effect of increasing exposure level (number of ANC contacts). The interaction effect also increased the control direct effect at both levels of mediator values. However, the control direct effect at control mediator value (deliver at home) is greater than the control direct effect for mediator event value (deliver at health facility). This indicated that if a pregnant woman had more ANC contacts but delivered at home, still it is possible she could vaccinate her newborn child on the schedule.

The study focused on decomposing the total effect of count treatment on ordinal outcome into direct and indirect effects under a single mediator. In addition, confounders' bias is controlled using regression method. However, there can be another post-treatment confounders affected by the treatment and may or may not be affected by the mediator. It is also ideal to control the confounding effects using weighting method and decomposing the total effect into all available path specific effects. Hence, the next section addresses all the limitations of the previous section.

5.2. Weighting vs G-Computation in Sequential Mediation

5.2.1. Background

Mediation analysis can involve single mediator or multiple mediators according to the study setting. The multiple mediators can be independent (parallel mediators) or dependent mediators (sequential or ordered mediators) where one affects the other. In a single mediator mediation analysis, there is no any other mediator except there can exist post-exposure confounders[38]. In parallel multiple mediators, there is no mediator affected by the exposure, and that goes on to affect another mediator(s). For such mediators, the analysis is carried out simultaneously [39]. In sequential mediators, which are the focus of this study, there is mediator(s) influenced by the exposure and in turn affects the subsequent mediator. For such mediators, we can apply path analysis to decompose total exposure's effect in each of the paths[40].

For causally ordered mediators, different approaches of estimating mediated effects through specific paths have been discussed in the literature. Depending on decomposing total effects into number of paths, the existing approaches are categorized as "two-way decomposition", "partial decomposition", and "complete decomposition"[220]. In "two-way decomposition", the total exposure's effect is separated into direct and indirect effects by accounting all mediators as one unit[39, 221]. With k ordered mediators, the "partial decomposition" approach involves decomposing the total exposure's total effect into $k+1$ distinct paths [221-223]. In "complete decomposition" also called "fine-decomposition" with k ordered mediators, the total effect is decomposed into 2^k particular paths[220, 224]. In complete decomposition, the total effect is decomposed into all available paths provided that identifiably assumption which is discussed in the subsequent section is not violated.

Investigators in the literature have used various methods to estimate direct and indirect effects, as well as path specific effects under the setting of multiple mediators. VanderWeele and Vansteelandt[39] used regression and weighting methods to estimate direct and indirect effects considering all mediators at a time. Daniel et al [224] used sensitivity analysis methods to estimate the bounds of path specific effects under normality assumption using G-computation. Lin and VanderWeele [225] used interventional approach using regression to estimate all path specific effect under two causally ordered mediators. Tai and Lin [220] have also extended interventional approach to generalized interventional approach using regression and G-computation to estimate path specific effects under multiple ordered mediators. On the other hand, Tai et al. [226] proposed reclassify fully mediated interaction technique for sequential effect decomposition under causally ordered mediator. Zugna et al. [227] compared inverse odds ratio, inverse probability weighting and imputation techniques to estimate marginal and conditional direct and indirect effects under two-way decomposition. They also used extended imputation for three-way decomposition under two sequential mediators. Similarly, Steen et al. [223] used flexible approach that combines imputation and weighting method to estimate natural direct and indirect effects. Moreover, Tai et al. [228] used G-computation with Monte Carlo simulation to estimate path specific effects under causally ordered mediators.

All the studies stated and the references therein focused on regression, weighting and G-computation and a combination of either regression with weighting or regression with G-computation methods to estimate direct and indirect effects as well as path specific effects. However, to the level of our knowledge, there has not been any study that used different weighting methods to control pre-treatment and pre-mediator confounders, and G- computation simultaneously to compare their performance.

Mediation analysis with causally ordered multiple mediators in previous literatures have focused on either continuous or binary outcome. We found that VanderWeele and Lim [41] used ordinal outcome under single mediator; and Zhou et al. [43] used ordinal outcomes under parallel multiple mediators. This indicated that mediation analysis under causally ordered mediators in ordinal outcome is under developed. In addition, the investigators used a count exposure which was overlooked in the literature according to our level of knowledge.

Thus, this study applies complete decomposition under sequential mediators; and proposes a new weighting strategy and compares it to G-computation via simulation and actual data.

5.2.2. Notation and Identification Assumption

For a causal mediation model under ordered mediators presented in Figure 5.2, T is count treatment, M_1 is the first mediator, M_2 is the second mediator, Y is an ordinal outcome, and $\boldsymbol{C}$ is a P dimensional confounders. An exposure causes all mediators and the outcome, the first mediator causes the second mediator and the outcome and the second mediator also causes the outcome.
In the counter factual framework, let t and t^* are two treatment levels, then:

i. $Y(t, M_1(t), M_2(t.M_1(t)))$, is the potential outcome of Y when $T = t$, $M_1 = m_1$ at $T = t$, $M_2 = m_2$ at $T = t$ and $M_1 = m_1$at $T = t$
ii. $Y(t^*, M_1(t^*), M_2(t^*, M_1(t^*)))$ is the potential outcome of Y when $T = t^*$, $M_1 = m_1$ at $T = t^*$, $M_2 = m_2$ at $T = t^*$ and $M_1 = m_1$at $T = t^*$
iii. $Y(t, M_1(t^*), M_2(t^*, M_1(t^*)))$ is the potential outcome of Y when $T = t$, $M_1 = m_1$ at $T = t^*$, $M_2 = m_2$ at $T = t^*$ and $M_1 = m_1$ at $T = t^*$
iv. $Y(t, M_1(t), M_2(t^*, M_1(t^*)))$ is the potential outcome of Y when $T = t$, $M_1 = m_1$ at $T = t$, $M_2 = m_2$ at $T = t^*$ and $M_1 = m_1$at $T = t^*$
v. $Y(t, M_1(t), M_2(t, M_1(t^*)))$ is the potential outcome of Y when $T = t$, $M_1 = m_1$ at $T = t$, $M_2 = m_2$ at $T = t$ and $M_1 = m_1$at $T = t^*$

To identify the potential outcomes and estimate direct and indirect effects of an exposure on the outcome, the identification assumptions are required. These assumptions are necessary to identify the causal effects from observational data and estimate path-specific effects without bias.

1. **Consistency assumption**: the potential outcome of the outcome is equal to the observed value of the outcome for a given exposure and mediator values. That is $(t, m_1, m_2) = Y$, when $T = t$, $M_1 = m_1$ and $M_2 = m_2$. The potential outcome of mediators' value is equal to the observed mediators' value for the given exposure and mediator values. That is $M_2(t, m_1)$ =M_2 when $T = t$, and $M_1 = m_1$,and $M_1(t)$ =M_1 when $T = t$ [41].
2. **Positivity assumption**: the conditional probability of M_2 given $M_1, T, \boldsymbol{C}$, and the conditional probability of M_1 given T and **C** is none zero. Mathematically, it can be

expressed as $P(M_2|M_1 = m_1, T = t, \boldsymbol{C}) > 0$, and $(M_1|T = t, \boldsymbol{C}) > 0$, for $T = t, t^*$ [29, 40]. In addition, the conditional probability of T given $\boldsymbol{C}$ is none zero.

3. **No unmeasured confounders assumption**: this is alternatively called "conditional exchangeability" assumption [38, 226]. Under ordered mediators, there are four assumptions described as follows [39, 225]:

   3.1.1. No unmeasured/uncontrolled confounders to explain away exposure-outcome association (mathematically, it is expressed as $Y(t, m_1, m_2) \perp T|\boldsymbol{C}$)

   3.1.2. No unmeasured/uncontrolled confounders to explain away mediator-outcome association (Mathematically, it is expressed as $Y(t, m_1, m_2) \perp (M_1, M_2)|T, \boldsymbol{C}$)

   3.1.3. No unmeasured/uncontrolled confounders to explain away exposure-mediator association (Mathematically, it is expressed as $(M_1(t), M_2(t)) \perp T|\boldsymbol{C}$)

   3.1.4. Cross-world independence assumption: It assumes no unmeasured/uncontrolled confounders affected by exposure to explain away mediator-outcome association (Mathematically, it is expressed as $Y(t, m_1, m_2) \perp (M_1(t^*), M_2(t^*))|\boldsymbol{C}$) [221].

The cross-world independence assumption is a strong assumption. When this assumption is violated, the option is to consider the mediator-outcome confounder as an additional mediator [29, 221]. Figure 5.2 shows M_1 is affected by T and is an intermediate confounder of $M_2 - Y$ association, and shows the violation of cross-world independence assumption when M_2 is considered the only mediator. However, the first mediator is considered as a mediator jointly with the second mediator.

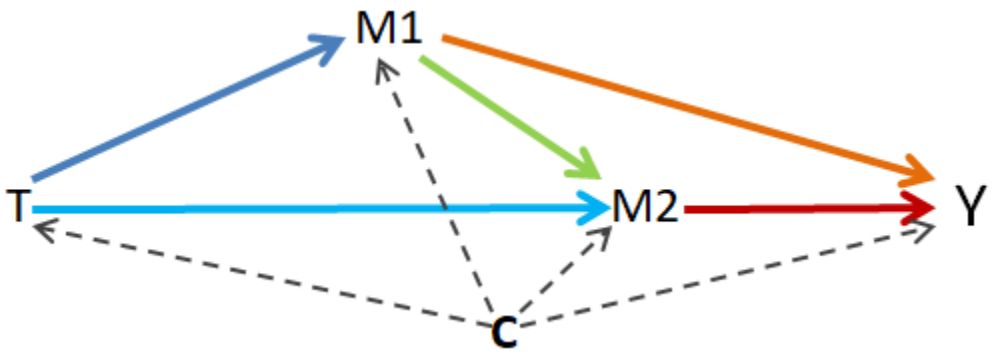


Figure 5.2. DAG for mediation analysis under ordered mediators.

5.2.3. Decomposition of Total Effect into Direct and Indirect Effects

Figure 5.2 shows causal relationship between exposure and ordinal outcome, exposure and binary mediators and binary mediators and ordinal outcome. We described three approaches for decomposing the total effect into natural direct and indirect effects: (1) two-way decomposition, (2) partial decomposition, and (3) complete decomposition. Each approach offers a progressively more detailed understanding of mediation pathways, especially under sequential mediators.

Approach 1: Two-way decomposition

In two-way decomposition, in the presence of ordered mediators, the total effect is decomposed into natural direct and indirect effects by considering the mediators as one single block of mediators. The natural indirect effect is the portion of the total effect mediated through M_1 and M_2 jointly including the interaction effect between exposure and mediators; and the natural direct effect is the portion of the effect not mediated by the M_1 and M_2 [29, 39, 221]. Using the potential outcome values, the natural direct effect (NDE) and natural indirect effects (NIE) conditional on confounders, $\boldsymbol{C}$, can be defined in terms of risk ratio(RR) and risk difference(RD) scale as follows. It can also be extended to odds(OR) ratio scale [41, 43]. The marginal NDE and NIE can also be obtained when covariates are controlled using propensity score weighting.

$$NDE_{RR} = P(Y(t, M_1(t^*), M_2(t^*, M_1(t^*) \leq j|\boldsymbol{C})/P(Y(t^*, M_1(t^*), M_2(t^*, M_1(t^*) \leq j|\boldsymbol{C}). \quad (5.10)$$

$$NDE_{RD} = P(Y(t, M_1(t^*), M_2(t^*, M_1(t^*) \leq j|\boldsymbol{C}) - P(Y(t^*, M_1(t^*), M_2(t^*, M_1(t^*) \leq j|\boldsymbol{C}). \quad (5.11)$$

$$NIE_{RR} = P(Y(t, M_1(t), M_2(t, M_1(t) \leq j|\boldsymbol{C})/P(Y(t, M_1(t^*), M_2(t^*, M_1(t^*) \leq j|\boldsymbol{C}). \quad (5.12)$$

$$NIE_{RD} = P(Y(t, M_1(t), M_2(t, M_1(t) \leq j|\boldsymbol{C}) - P(Y(t, M_1(t^*), M_2(t^*, M_1(t^*) \leq j|\boldsymbol{C}). \quad (5.13)$$

The total effect is the product of the natural direct and indirect effects in terms of RR and OR scale; and it is the sum of natural direct and indirect effects in terms of additive (RD) scale.

The two-way decomposition of the total effect into a joint natural direct and indirect effect is found if the assumptions stated above hold with respect to the joint mediator, M_1 and M_2

Approach 2: Partial decomposition

In the two-way decomposition, there are only two path-specific effects (direct and indirect effects) that do not account for path-specific effects across each and combined sequential mediators. In the partial decomposition (partial sequential decomposition for causally ordered mediators) with k

mediators, the total effect is decomposed into $k+1$ paths where k paths represent indirect effects and one path represents the direct effect [220, 221, 223].

For two sequential mediators M_1 and M_2, we have a three-way decompositions (fine grained decomposition) [221, 229] such as the effect: (i) through paths involving neither M_1 nor M_2 ($T \rightarrow Y$) (ii) through pathways not involving M_1 ($T \rightarrow M_2 \rightarrow Y$), and (iii) through pathways involving M_1 (the combination of $T \rightarrow M_1 \rightarrow M_2 \rightarrow Y$ and $T \rightarrow M_1 \rightarrow Y$, shortly denoted as $T \rightarrow M_1 Y$).

In the three-way decomposition involving two sequential mediators, the natural indirect effects in equations (5.10) and (5.11) can be decomposed into the conditional or marginal partial indirect effect through M_1 and conditional or marginal partial indirect effect through M_2 alone as follows

$$NIE_{RR} = NIE_{RR}(T \rightarrow M_1 Y) * NIE_{RR}(T \rightarrow M_2 \rightarrow Y). \quad (5.14)$$

$$NIE_{RD} = NIE_{RR}(T \rightarrow M_1 Y) + NIE_{RR}(T \rightarrow M_2 \rightarrow Y). \quad (5.15)$$

$$NIE_{RR} = P(Y(t, M_1(t), M_2(t, M_1(t) \leq j|\boldsymbol{C})/P(Y(t, M_1(t^*), M_2(t, M_1(t^*) \leq j|\boldsymbol{C}) * P(Y(t, M_1(t^*), M_2(t, M_1(t^*) \leq j|\boldsymbol{C})/P(Y(t, M_1(t^*), M_2(t^*, M_1(t^*) \leq j|\boldsymbol{C}). \quad (5.16)$$

$$NIE_{RD} = P(Y(t, M_1(t), M_2(t, M_1(t) \leq j|\boldsymbol{C}) - P(Y(t, M_1(t^*), M_2(t, M_1(t^*) \leq j|\boldsymbol{C}) + P(Y(t, M_1(t^*), M_2(t, M_1(t^*) \leq j|\boldsymbol{C}) - P(Y(t, M_1(t^*), M_2(t^*, M_1(t^*) \leq j|\boldsymbol{C}). \quad (5.17)$$

In order to estimate these effects, two assumptions should be satisfied, that is no unmeasured confounding of M_1-M_2 associations and the lack of confounders of this association in turn affected by the exposure [29].

Approach 3: Complete effect decomposition

In the settings with two sequentially ordered mediators, M_1 and M_2, as shown in Figure 5.2, there are four possible pathways from exposure (T) to outcome (Y): through M_1 alone, through M_2 alone, through both M_1 and M_2, and through neither M_1 nor M_2. From these pathways, we can define four path-specific effects: through each $M_1 (PSE_{T \rightarrow M_1 \rightarrow Y})$ denoted by PSE_1, through $M_2 (PSE_{T \rightarrow M_2 \rightarrow Y})$ denoted by PSE_2, through M_1 and then $M_2 (PSE_{T \rightarrow M_1 \rightarrow M_2 \rightarrow Y})$ denoted by PSE_{12} and neither M_1nor M_2 $(PSE_{T \rightarrow Y})$ denoted by PSE_0.

Generally, with K ordered mediators, the number of path specific effects (PSEs) is 2^K [224, 225]. Due to its complexity with large K, we focused on two ordered mediators. Using risk difference (RD), risk ratio (RR) and odds ratio (OR) scale, the four path specific effects are defined as follows [225, 228].

$$RD_{PSE_0} = \Phi(t_1, 0,0,0) - \Phi(0,0,0,0).$$
$$OR(RR)_{PSE_0} = \Phi(t_1, 0,0,0)/\Phi(0,0,0,0). \quad (5.18)$$
$$RD_{PSE_1} = \Phi(t_1, t_2, 0,0) - \Phi(t_1, 0,0,0).$$
$$OR(RR)_{PSE_1} = \Phi(t_1, t_2, 0,0)/\Phi(t_1, 0,0,0). \quad (5.19)$$
$$RD_{PSE_2} = \Phi(t_1, t_2, t_3, 0) - \Phi(t_1, t_2, 0,0).$$
$$OR(RR)_{PSE_2} = \Phi(t_1, t_2, t_3, 0)/\Phi(t_1, t_2, 0,0). \quad (5.20)$$
$$RD_{PSE_{12}} = \Phi(t_1, t_2, t_3, t_4) - \Phi(t_1, t_2, t_3, 0). \quad (5.21)$$
$$OR(RR)_{PSE_{12}} = \Phi(t_1, t_2, t_3, t_4)/\Phi(t_1, t_2, t_3, 0).$$

Where $\Phi(t_1, t_2, t_3, t_4)$ is defined as risk or odds of $(Y(t_1, M_1(t_2), M_2(t_3, M_1(t_4))$, PSE_0 is the same as NDE whereas PSE_1 , PSE_2, PSE_{12} are components of NIE.

In terms of risk difference, the total risk difference (RD_{TE}) is the sum of each path specific effects, and in risk ratio or odds ratio scale, the total effect is the product of each path specific effect.

$$RD_{TE} = RD_{PSE_0} + RD_{PSE_1} + RD_{PSE_2} + RD_{PSE_{12}}.$$
$$OR(RR)_{TE} = RD_{PSE_0} * OR(RR)_{PSE_1} * OR(RR)_{PSE_2} * OR(RR)_{PSE_{12}}. \quad (5.22)$$

5.2.4 Estimation

Given count exposure, binary sequential mediators, and ordinal outcomes under the aforementioned assumptions, $Y(t_1, M_1(t_2), M_2(t_3, M_1(t_4)))$ can be non-parametrically defined using mediation formula as [224, 225, 228] :

$$Y(t_1, M_1(t_2), M_2(t_3, M_1(t_4))) = \sum_{m2=0}^{1} \sum_{m1=0}^{1} p(Y \leq j | \boldsymbol{C} = c,\ T = t_1, M_1 = m_1, M_2 = m_2) \times P(M_1 = m_1 |\ \boldsymbol{C} = c, T = t_2) \times \sum_{m'_1=0}^{1} P(M_2 = m_2 |\boldsymbol{C} = c,\ T = t_3, M_1 = m'_1) \times P(M_1 = m'_1 | \boldsymbol{C} = c,\ T = t_4). \quad (5.23)$$

If M_1 and M_2 are continuous, the summation in equation (5.23) is replaced by integrals.

To estimate each path specific effects regression, weighting and G-computation methods can be used.

Regression method

Consider settings of two treatment levels ($T = t_1$ and $T = t_2$), mediator values ($M_1 = m_1$ and $M_2 = m_2$) and j^{th} category of the outcome, the structural equation modeling(SEM) with two way interactions can be written as follows:

$$g(P(Y \leq j|T = t, M_1 = m_1, M_2 = m_2, \boldsymbol{C})) = \theta_j - [\theta_1 t + \theta_2 m_1 + \theta_3 m_2 + \theta_4 t m_1 + \theta_5 t m_2 + \boldsymbol{\theta_c}^T \boldsymbol{C}]. \quad (5.24)$$

$$g(P(M_2 = m_2|T = t, M_1 = m_1, \boldsymbol{C})) = \beta_0 + \beta_1 t + \beta_2 m_1 + \beta_3 t m_1 + \boldsymbol{\beta_c}^T \boldsymbol{C}. \quad (5.25)$$

$$g(P(M_1 = m_1|T = t, \boldsymbol{C})) = \gamma_0 + \gamma_1 t + \boldsymbol{\gamma_c}^T \boldsymbol{C}. \quad (5.26)$$

Where $g(.)$ is the link function, $\boldsymbol{\theta_c}$, $\boldsymbol{\beta_c}$ and $\boldsymbol{\gamma_c}$ are vectors of coefficients.

Combining equation (5.23) and equation (5.24-5.26), the four path specific effects in terms of RD under rare outcome assumption are given as follows [43, 225]:

$$RD_{PSE_0} = \exp\{[\theta_1 + \theta_5(\beta_0 + \beta_2\gamma_0 + \beta_2\boldsymbol{C}^T\gamma_c + \boldsymbol{\beta_c}^T\boldsymbol{C}) + \theta_4(\gamma_0 + \boldsymbol{\gamma_c}^T\boldsymbol{C}] + [\theta_4\gamma_1 + \theta_5(\beta_1 + \beta_3\gamma_0 + \beta_3\boldsymbol{C}^T\boldsymbol{\gamma}_c + \beta_2\gamma_1)]t_1 + \theta_5\beta_3\gamma_1{t_1}^2\}(t_2 - t_1). \quad (5.27)$$

$$RD_{PSE_1} = exp\{\theta_2\gamma_1 + \theta_4\gamma_1 t_2\}(t_2 - t_1). \quad (5.28)$$

$$RD_{PSE_2} = \exp\{\theta_3(\beta_1 + \beta_3\gamma_0 + \beta_3\boldsymbol{C}^T\gamma_c) + \theta_5(\beta_1 + \beta_3\gamma_0 + \beta_3\boldsymbol{C}^T\gamma_c)t_2 + \theta_3\beta_3\gamma_1 t_2 + \theta_5\beta_3\gamma_1 t_2 t_0\}(t_2 - t_1). \quad (5.29)$$

$$RD_{PSE_{12}} = \exp\{\theta_3\beta_2\gamma_1 + [(\theta_5\beta_2\gamma_1 + \theta_3\beta_3\gamma_1)]t_2 + \theta_5\beta_3\gamma_1{t_1}^2\}(t_2 - t_1). \quad (5.30)$$

If there is no exposure mediators interactions, θ_4, θ_5, and β_3 will be zero so that the equations will be simplified further. On the other hand, it is also possible to add three way interactions such as exposure-mediator one-mediator two interactions as well as mediator one-mediator two interactions.

Weighting method

In weighted based estimation of path specific effects, first requires estimating inverse-probability treatment weights which can be obtained from regression models for exposure and mediators

[230]. In this case, three regressions have been used: the first one is regression of exposure conditional on confounders, the second is regression of first mediator conditional on exposure and confounders, and the third is regression of second mediator conditional on exposure, first mediator and confounders[221]. In this study, the proposed stabilized inverse-probability weights were estimated as follows:

$$w^T = \frac{P(T=t)}{P(T=t|\boldsymbol{C})}\ , w^{M_1} = \frac{P(M_1=m_1|T=t)}{P(M_1=m_1|,T=t,\boldsymbol{C})}\ ,\ w^{M_2} = \frac{P(M_2=m_2|T=t,M_1=m_1)}{P(M_2=m_2|,T=t,M_1=m_1,\boldsymbol{C})}.$$

Having estimated stabilized weights, marginal structural model (MSM) was used to estimate the marginal path specific effects. The MSM is used to estimate the counterfactuals after controlling the effect of confounding through weighting rather than regression.

Hence, the weighted SEM modeling described in equations (5.1) and (5.2) can be re-written as:

$$g(P(Y \leq j|T=t, M_1=m_1, M_2=m_2)) = \theta_j - [\theta_1 t + \theta_2 m_1 + \theta_3 m_2 + \theta_4 t m_1 + \theta_5 t m_2]$$
weighted by $w^T * w^{M_1} * w^{M_2}$. (5.31)

$g(P(M_2=m_2|T=t, M_1=m_1, \boldsymbol{C})) = \beta_0 + \beta_1 t + \beta_2 m_1 + \beta_3 t m_1$ weighted by $w^T * w^{M_1}$.(5.32)

$g(P(M_1=m_1|T=t, \boldsymbol{C})) = \gamma_0 + \gamma_1 t$ weighted by w^T. After estimating parameters, we applied mediation formula stated in equation (14) and estimated path specific effects using equations (5.27)-(5.30) except there is no coefficient for the confounders, **C,** and parameters are estimated using weighted regression. It is also possible to use double robust method by incorporating unbalanced confounders in the weighted outcome and mediators' models.

G-computation

G-computation was first introduced by Robin in 1986 to estimate causal effect of an exposure on time varying outcome in the presence of time varying confounders [231]. However, at this time G-computation is widely used in non-mediation causal analysis[232-234] and mediated causal analysis[228, 235] in the absence of time varying exposure and covariates. G-computation is used to estimate counterfactual outcomes under different exposure levels after fitting the joint distribution of data[234, 236].

In the literature, G-computation has been used in three scenarios. In the first scenario, it is used to find the average treatment effect (ATE) with the average of the difference between two estimated

counterfactual outcomes [233, 234]. This means that when $\widehat{y}_i(t_1) = E(Y|t = t_1, \boldsymbol{C})$ and $\widehat{y}_i(t_2) = E(Y|t = t_2, \boldsymbol{C})$, then the ATE is $E(\widehat{y}_i(t_2) - \widehat{y}_i(t_1)) = \frac{1}{n}(\sum_i(\widehat{y}_i(t_2) - \widehat{y}_i(t_1))$. In our case, the counterfactuals of the outcome, $\Phi(t_1, t_2, t_3, t_4)$, for each individual are estimated using equations (5.22-5.24). Then the path specific effects are estimated using equations (5.16-5.19) using the average of estimated counterfactuals.

The second scenario is marginal structural equation modeling (MSM). The MSM of G-computation states that after estimating the counterfactuals from the fitted models then regress the entire set of counterfactuals on the treatment. Then the coefficient of the treatment represents the causal treatment effect [230].

The third scenario is simulation approach of G-computation which was used to estimate interventional path specific effects in mediation analysis under causally ordered mediators [220, 228]. In this scenario, first fit the models in (5.24-5.26). Using the estimated parameters in the first step, random samples of counterfactuals of the mediators ($G_1(t_2)$ and $G_2(t_3, G_1(t_4))$ are estimated using random samples of covariates for each fixed exposure level. Then the random samples of counterfactuals of the outcome $Y(t_1, G_1(t_2), G_2\big(t_3, G_1(t_4)\big))$ are estimated using the random samples of covariates and estimated counterfactuals of the mediators for fixed exposure levels. Then compute the average of $Y(t_1, G_1(t_2), G_2\big(t_3, G_1(t_4)\big))$ which is $\Phi(t_1, G_1(t_2), G_2(t_3, G_1(t_4)))$ in our case and estimate each interventional path specific effects based on equations (9-12). The detail procedures are illustrated elsewhere in [228]. Alternatively, the counterfactuals of mediators and outcome can be estimated from randomly simulated estimated parameters from their sampling distributions (assuming multivariate normal distribution) where the detail procedure is found in [220]. To compare the performance of candidate methods, we used biases, standard errors and confidence intervals of each PSE estimated by parametric bootstrapping [43, 228].

5.2.5 Simulation Study

Data generation

To compare the performance of proposed methods for count treatment, binary mediators, and ordinal outcome, we conducted a simulation study. In the simulation study, data were generated in two ways. The first is without treatment-mediator interaction, and the second is with treatment-

mediator interaction. While estimating PSEs, we investigated the bias, standard errors, and confidence intervals, where the standard errors and confidence intervals were estimated using bootstrap methods. Each data points were generated with the sample sizes of 500, 1000, and 2000 [237]. To estimate the standard errors and 95% confidence intervals, we run the bootstrap sample 1000 times.

Four covariates, one from normal distribution ($X_1 \sim N(0.1)$), one from binomial distribution ($X_2 \sim Bernoulli(0.5)$) , one from uniform distribution ($X_3 \sim U(0,1)$) and one from Poisson distribution ($X_4 \sim Poisson(1.5)$) were generated.

The treatment was generated as: $g(E(T|\boldsymbol{X})) = 0.5x_1 + 0.2x_2 + 0.75x_3 + x_4$ where $g(.)$ is Poisson link function.

The mediators were generated as follows:

$g\big(p(M_1 = 1|T = t, \boldsymbol{X})\big) = 0.5t + 0.33x_1 + 0.25x_2 + 0.15x_3 + 0.13x_4.$

$g\big(p(M_2 = 1|T = t, M_1 = m_1, \boldsymbol{X})\big) = 0.4t + 0.25m_1 + tm_1 + 0.3x_1 + 0.5x_2 + 0.12x_3 + 0.11x_4$, where $g(.)$ is the logit link.

The latent continuous variable for the ordinal outcome was generated as follows:

$E(Y|T = t, M_1 = m_1, M_2 = m_2, \boldsymbol{X}) = 0.6t + b_1m_1 + tm_1 + b_2m_2 + tm_2 + b_3x_1 + b_4x_2 + b_5x_3 + b_6 * x_4$,where $b_1 = log(1.25), b_2 = log(1.5), b_3 = log(1.75), b_4 = log(2), b_5 = log(1.25),\ b_6 = log(1.2)$.

After generating the latent continuous outcome, it was cut into three ordered categories at 60th and 90th percentiles such that the first category represents 60%, the second category represents 30%, and the third category represents 10% of observations.

Result of simulation study

The simulation results for the ordinal outcome are presented in Tables 5.5 and 5.6 for different sample sizes. The results consist of path-specific effects, biases, and standard errors for effect estimation, as well as 95% confidence intervals of the effect estimates. For cases with no interaction, the biases and standard errors of the total effect of an exposure on the outcome in G-computation were less than those of the weighting method across all sample sizes. However, the difference is not apparent for each path-specific effect (Table 5.5).

The results of the simulation study with exposure-mediators interaction are presented in Table 5.6. The results show that the bias and standard errors of the weighting method were less than those of G-computation for all sample sizes and treatment-mediators interaction. The 95% confidence interval of the weighting method was also narrower than the G-computation. Similarly, the standard error of the direct effect of the weighting method was less than the G-computation (Table 5.6).

Table 5.5**.** Result of simulation study without interaction for ordinal outcome.

N	Method	PSE	Effects	bias	std. error	95% CI	
500	G-computation	PSE0	1.46	0.0310	0.172	1.24	1.90
		PSE1	1.01	-0.0007	0.016	0.979	1.05
		PSE2	1.01	-0.0001	0.010	0.994	1.03
		PSE12	1.01	-0.0002	0.002	0.998	1.006
		TE	1.49	0.029	0.169	1.264	1.925
	Weighting	PSE0	1.69	0.0300	0.232	1.381	2.269
		PSE1	1.03	-0.0001	0.014	1.003	1.056
		PSE2	1.04	0.002	0.022	1.009	1.095
		PSE12	1.01	-0.0001	0.003	0.999	1.009
		TE	1.82	0.035	0.243	1.480	2.441
1000	G-computation	PSE0	1.57	0.018	0.116	1.398	1.860
		PSE1	1.01	-0.0002	0.007	0.994	1.021
		PSE2	1.02	-0.0006	0.010	1.006	1.048
		PSE12	1.01	-000002	0.001	0.999	1.003
		TE	1.62	0.017	0.115	1.444	1.906
	Weighting	PSE0	1.53	0.081	0.240	1.255	2.144
		PSE1	1.02	-0.001	0.011	1.006	1.049
		PSE2	1.01	0.001	0.031	0.939	1.059
		PSE12	1.01	-0.0001	0.003	1.001	1.011
		TE	1.59	0.089	0.280	1.239	2.275
2000	G-computation	PSE0	1.50	0.003	0.052	1.414	1.627
		PSE1	1.02	-0.00001	0.005	1.007	1.025
		PSE2	1.02	0.0002	0.005	1.015	1.035
		PSE12	1.01	-0.0001	0.0007	0.999	1.002
		TE	1.57	0.003	0.0518	1.474	1.682
	Weighting	PSE0	1.26	0.0764	0.297	0.810	1.827
		PSE1	1.06	-0.003	0.024	1.022	1.112
		PSE2	1.09	-0.003	0.036	1.028	1.170
		PSE12	1.01	-0.0003	0.003	1.001	1.012
		TE	1.46	0.066	0.288	1.035	2.051

Table 5.6. Result of simulation study with treatment-mediators interaction for ordinal outcome.

N	Method	PSE	Effects	bias	std. error	95% CI	
500	G-computation	PSE0	2.79	1.277	2.796	1.927	8.268
		PSE1	1.01	-0.0003	0.009	0.986	1.023
		PSE2	1.01	-0.002	0.009	0.997	1.033

		PSE12	1.01	-0.0005	0.003	1.00	1.011
		TE	2.84	1.275	2.794	1.979	8.316
	Weighting	PSE0	2.92	0.226	0.924	2.225	4.914
		PSE1	1.04	0.001	0.021	1.010	1.091
		PSE2	1.10	0.003	0.039	1.038	1.195
		PSE12	1.01	-0.0004	0.005	1.002	1.020
		TE	3.37	0.273	1.071	2.554	5.678
1000	G-computation	PSE0	3.39	0.031	1.283	2.596	7.66
		PSE1	1.01	-0.0001	0.004	0.998	1.012
		PSE2	1.02	-0.001	0.011	1.006	1.047
		PSE12	1.01	-0.0001	0.001	1.00	1.004
		TE	3.49	0.03	1.282	2.701	7.717
	Weighting	PSE0	3.64	0.017	0.760	2.884	5.354
		PSE1	1.03	0.0001	0.015	1.009	1.066
		PSE2	1.14	0.002	0.029	1.087	1.204
		PSE12	1.01	-0.0003	0.004	1.003	1.017
		TE	4.33	0.022	0.938	3.387	6.534
2000	G-computation	PSE0	3.94	0.017	0.770	3.122	5.604
		PSE1	1.01	-0.0001	0.003	1.00	1.012
		PSE2	1.02	-0.0003	0.005	1.005	1.026
		PSE12	1.01	-0.0005	0.001	1.00	1.002
		TE	4.02	0.017	0.768	3.203	5.68
	Weighting	PSE0	3.40	0.044	0.325	2.877	4.140
		PSE1	1.10	-0.005	0.025	1.045	1.141
		PSE2	1.09	0.0006	0.014	1.062	1.12
		PSE12	1.02	-0.0009	0.004	1.007	1.023
		TE	4.133	0.016	0.428	3.456	5.128

The simulation study was also repeated for a binary outcome without interaction, and the result was presented in Table 5.7. It showed that the bias and standard error of total effect estimates were equivalent between G-computation and weighting methods for all sample sizes. However, the 95% confidence interval of the weighting method was narrower than the G-computation. Similarly, the bias and standard error of the direct effect in the weighting method were smaller than the G-computation, while they were equivalent for the indirect effects (Table 5.7).

Table 5.7. Simulation result for binary outcome without interaction.

N	Method	PSE	Effects	bias	std. error	95% CI	
500	G-computation	PSE0	1.14	-0.013	0.094	1.03	1.39
		PSE1	1.01	-0.001	0.008	0.981	1.02
		PSE2	1.01	0.001	0.006	0.997	1.02
		PSE12	1.01	0.004	0.001	0.999	1.01
		TE	1.15	0.012	0.090	1.04	1.39
	Weighting	PSE0	1.19	0.008	0.069	1.08	1.35
		PSE1	0.99	-0.001	0.006	0.981	1.01
		PSE2	1.02	-0.004	0.009	1.01	1.04
		PSE12	1.001	-0.001	0.001	1.00	1.01
		TE	1.23	0.007	0.065	1.11	1.36

1000	G-computation	PSE0	1.13	0.004	0.045	1.05	1.23
		PSE1	0.999	-0.001	0.002	0.995	1.01
		PSE2	1.01	-0.006	0.005	0.993	1.01
		PSE12	1.00	0.001	0.001	0.998	1.02
		TE	1.13	0.003	0.044	1.06	1.23
	Weighting	PSE0	1.193	0.002	0.043	1.12	1.29
		PSE1	1.004	-0.004	0.004	0.997	1.01
		PSE2	1.002	0.002	0.007	0.991	1.02
		PSE12	1.000	-0.001	0.001	1.00	1.01
		TE	1.20	0.004	0.039	1.13	1.29
2000	G-computation	PSE0	1.13	-0.001	0.041	1.06	1.22
		PSE1	1.00	-0.001	0.004	0.989	1.01
		PSE2	1.01	-0.001	0.004	0.999	1.01
		PSE12	0.99	-0.001	0.001	0.999	1.01
		TE	1.13	-0.001	0.039	1.07	1.22
	Weighting	PSE0	1.16	-0.004	0.035	1.09	1.23
		PSE1	1.02	-0.001	0.005	1.01	1.02
		PSE2	1.02	-0.001	0.004	1.01	1.02
		PSE12	1.01	-0.001	0.004	1.00	1.02
		TE	1.20	-0.007	0.038	1.11	1.26

Result of the actual data

In the real-world data for this study, the number of ANC visits was used as count exposure; institutional delivery and newborn postnatal care were used as first and second binary sequential mediators. Age-specific childhood vaccination was used as ordinal outcome with level of no vaccination when a child did not take any vaccine at the given child age and as per vaccine schedule, partial vaccination when a child did not take all vaccines at the given age and vaccine schedule, and full vaccination when a child received all vaccines at the given age and vaccine schedule.

Table 5.8 shows the result of mediation analysis for the actual data. The result contains the path-specific effect of the exposure on the outcome by using no ANC as a control value, twice, three times, and four times of ANC follow-up as exposure values. With zero and two treatment combinations, the weighting method produced smaller bias and standard error, while G-computation produced smaller bias and standard error when number of ANC contacts increased. However, the difference between G-computation and weighting is not meaningful in terms of performance metrics.

The result further conveyed that the total effect of the number of ANC visits increased as the number of contacts increased, as compared to no contacts. The direct effect also increased as the number of contacts increased. In both estimation methods, the indirect effect of an exposure on

the outcome through the first mediator was larger than the indirect effects through the second mediator and the first mediator, and then the second mediator. The indirect effect of an exposure through the second mediator was higher than the indirect effect through mediator one and mediator two, even though the two indirect effects are small (Table 5.8).

In mediation analysis with treatment-mediators interaction, the result is presented in Table 5.9. The comparison of G-computation and weighting method did not change from the no-interaction effect. In the meantime, the total and direct exposure effect increased, as the exposure values increased as compared to the control value. The indirect effect through the second mediator and the first mediator through the second mediator is still low (Table 5.9).

Furthermore, the result reveals that, the indirect effect through institutional delivery is consistently larger than the other indirect effects. This suggests that institutional delivery has a critical mediation role. The indirect effect through institutional delivery and then through newborn postnatal care is consistently lower than other indirect effects. The direct effect of ANC is larger than any of other indirect effects

Table 5.8. Path-specific effects of the number of ANC visits on the outcome.

Exposure level combination	methods	PSE	Effects	bias	std. error	95% CI	
0 vs 2	G-computation	PSE0	1.43	0.005	0.071	1.30	1.57
		PSE1	1.06	-0.001	0.012	1.03	1.08
		PSE2	1.02	0.001	0.005	1.01	1.03
		PSE12	1.43	0.005	0.071	1.30	1.57
		TE	2.20	0.021	0.217	1.82	2.66
	Weighting	PSE0	1.52	0.006	0.099	1.35	1.73
		PSE1	1.22	0.001	0.033	1.15	1.28
		PSE2	1.03	-0.001	0.008	1.01	1.04
		PSE12	1.01	-0.001	0.002	1.00	1.01
		TE	1.91	0.006	0.116	1.70	2.18
0 vs 3	G-computation	PSE0	1.71	0.011	0.127		
		PSE1	1.08	-0.001	0.019	1.05	1.12
		PSE2	1.03	0.001	0.008	1.02	1.05
		PSE12	1.01	-0.001	0.001	1.00	1.01
		TE	1.91	0.013	0.142	1.66	2.20
	weighting	PSE0	1.88	0.014	0.185	1.56	2.27
		PSE1	1.38	0.004	0.065	1.26	1.51
		PSE2	1.05	-0.001	0.015	1.02	1.08
		PSE12	1.01	-0.001	0.004	1.00	1.02
		TE	2.76	0.020	0.257	2.34	3.39

0 vs 4	G-computation	PSE0	2.04	0.020	0.203	1.67	2.48
		PSE1	1.11	-0.001	0.025	1.06	1.16
		PSE2	1.04	0.001	0.011	1.02	1.06
		PSE12	1.01	-0.001	0.002	1.00	1.01
		TE	2.36	0.025	0.235	1.96	2.86
	Weighting	PSE0	2.32	0.027	0.306	1.82	2.98
		PSE1	1.56	0.008	0.101	1.38	1.76
		PSE2	1.08	-0.001	0.025	1.03	1.13
		PSE12	1.01	-0.001	0.007	0.999	1.03
		TE	3.96	0.051	0.502	3.17	4.94

Table 5.9**.** Result of mediation analysis with treatment-mediators interaction.

Exposure level combination	methods	PSE	Effects	bias	std. error	95% CI	
0 vs 2	G-computation	PSE0	1.38	0.005	0.081	1.16	1.49
		PSE1	1.10	-0.001	0.022	1.06	1.14
		PSE2	1.02	0.001	0.008	1.01	1.04
		PSE12	1.01	0.001	0.002	1.01	1.01
		TE	1.48	0.004	0.086	1.32	1.65
	Weighting	PSE0	1.21	0.004	0.059	1.09	1.34
		PSE1	1.13	-0.001	0.031	1.06	1.19
		PSE2	1.47	0.020	0.021	1.05	1.96
		PSE12	1.00	-0.000	0.000		
		TE	1.99	0.022	0.230	1.61	2.56
0 vs 3	G-computation	PSE0	1.57	0.014	0.133	1.33	1.86
		PSE1	1.15	-0.002	0.032	1.088	1.22
		PSE2	1.03	0.001	0.014	1.01	1.07
		PSE12	1.01	-0.001	0.002	1.00	1.01
		TE	1.86	0.015	0.156	1.57	2.18
	Weighting	PSE0	1.47	0.023	0.263	1.03	2.09
		PSE1	1.74	0.008	0.199	1.40	2.17
		PSE2	1.06	-0.002	0.030	1.01	1.13
		PSE12	1.01	-0.001	0.006	1.00	1.02
		TE	2.73	0.027	0.440	1.97	3.74
0 vs 3	G-computation	PSE0	1.89	0.021	0.207	1.53	2.34
		PSE1	1.19	-0.005	0.043	1.12	1.28
		PSE2	1.05	0.002	0.020	1.02	1.09
		PSE12	1.01	-0.001	0.002	0.996	1.01
		TE	2.36	0.028	0.260	1.89	2.89
	Weighting	PSE0	2.22	0.071	0.567	1.39	3.59
		PSE1	1.59	0.020	0.208	1.24	20.52
		PSE2	1.09	0.001	0.051	1.01	1.22

PSE12	1.01	0.001	0.012	0.987	1.04
TE	3.89	0.094	0.875	2.50	5.99

5.2.6 Discussion and Conclusion

This study focused on using two sequential binary mediators with four path specific effects. Studies on mediation analysis relayed on binary exposure. On the other hand, studies on mediation analysis with ordinal outcome are limited except VanderWeele and Lim [41] and [42]used ordinal outcome under single mediator; and Zhou et al. [43] used ordinal outcome under parallel mediators. Thus, this study used count exposure for mediation analysis on ordinal outcome under sequential mediators which has not been done to the level of our knowledge. This is the first study applying complete path-specific decomposition under causally ordered mediators for count exposures and ordinal outcomes.

G-computation via simulation was used to estimate path specific effects under sequential mediators and complete effect decomposition [220, 224, 228]. On the other hand, inverse probability weighting was used to estimate marginal direct and indirect effects under two-way decomposition[29, 39]. In this study, we proposed weighting of each marginal structural equation modeling where the first mediator equation was weighted by inverse probability weight of exposure's regression on covariates, the second mediator equation was weighted by the product of exposure's weight and first mediators weight, and the outcome equation was weighted by the product of invers probability weights generated by the exposure, first mediator and second mediator. All weights were stabilized. To test the performance of the proposed weight, we compared its performance with G-computation via simulation using a simulation study and actual household survey data for count exposure, two binary sequential mediators, and an ordinal outcome. We also repeated the simulation study for the binary outcome for triangulation purposes.

In the simulation study for an ordinal outcome without exposure-mediators interaction, the bias and standard errors of the direct and total effects estimates were smaller in G-computation as compared to the weighting method. However, the differences between these performance metrics were not meaningful. With exposure-mediators interaction, the bias and standard errors of the weighting method were smaller than the G-computation method with a narrower 95% confidence interval. For a binary outcome without interaction, the bias and standard errors of the weighting method were smaller than the G-computation.

For the real data, we first examine the causal effects of the number of ANC visits (exposure), institutional delivery (first mediator), and newborn postnatal care (second mediator) on age-specific childhood vaccination (outcome). We found that the adjusted effect of the exposure on the outcome was 0.185, the effect of the first mediator on the outcome was 0.491, and the effect of the second mediator on the outcome was 0.566. The effects were significant at 95% confidence intervals. In the mediation analysis, we used no ANC visit as the control value, and twice, three times, and four times ANC visits as exposure values. In both interaction and without interaction of exposure-mediators, the biases and standard errors of effect estimates were equivalent in the G-computation and weighting method. In both methods, the path-specific effects increased as the exposure value increased. However, the exposure's indirect effect on the outcome through institutional delivery was higher than the other two indirect effects (effect through newborn postnatal care and then the effect through institutional delivery and then newborn postnatal care).

In conclusion, simulation-based G-computation performed better than proposed weighting method in the absence of treatment-mediator interaction though the difference was not substantial. The proposed weighting method outperformed the simulation–based G-computation in two-way treatment-mediators interaction. Both methods performed equivalently in actual data analysis. However, the G-computation via simulation is computationally intensive to obtain the conditional effect estimates after adjusting for confounders via regression, while the weighting approach offers a computationally efficient strategy for marginal path-specific effect estimation under realistic causal scenarios. In addition, G-computation via simulation uses simulation of either confounders or their parametric estimates, which means the actual values of the confounders or their parametric estimates are not used in path specific effect estimates. Thus, it is important to use the proposed weighting methods to estimate path-specific effects under multiple sequential mediators.

The direct causal effect of number of ANC contacts on age-specific childhood vaccination is higher than each of the indirect effects. However, some of its effects are shared through institutional delivery, newborn postnatal care, and combination of them. The institutional delivery was more important mediator as compared to newborn postnatal care and a combination of them.

The limitations of this section were the simulation was not done for model misspecification of the outcome, exposure and mediators or a combination of them. The simulation was not also tested

with high dimensional confounders. Three way interactions were not also tested. Controlled direct effect will also be a topic in other study setting. Finally, the sensitivity analysis for unmeasured confounders was not done. It could also be important to replicate the simulation for continuous outcome.

CHAPTER SIX: CONCLUSION, LIMITATION AND FUTURE WORKS

6.1. Conclusion

The primary objective of this dissertation work is extending the causal inference to count treatment where it was disregarded in the literature. Different approaches of causal inference for count treatment have been examined under ordinal outcome. Parametric and plasmode simulations have been employed to compare proposed statistical methods. The study also has used the mini-Ethiopian Demographic and Health Survey (MEDHS 2019) data to triangulate and apply the proposed methods for the real world datasets.

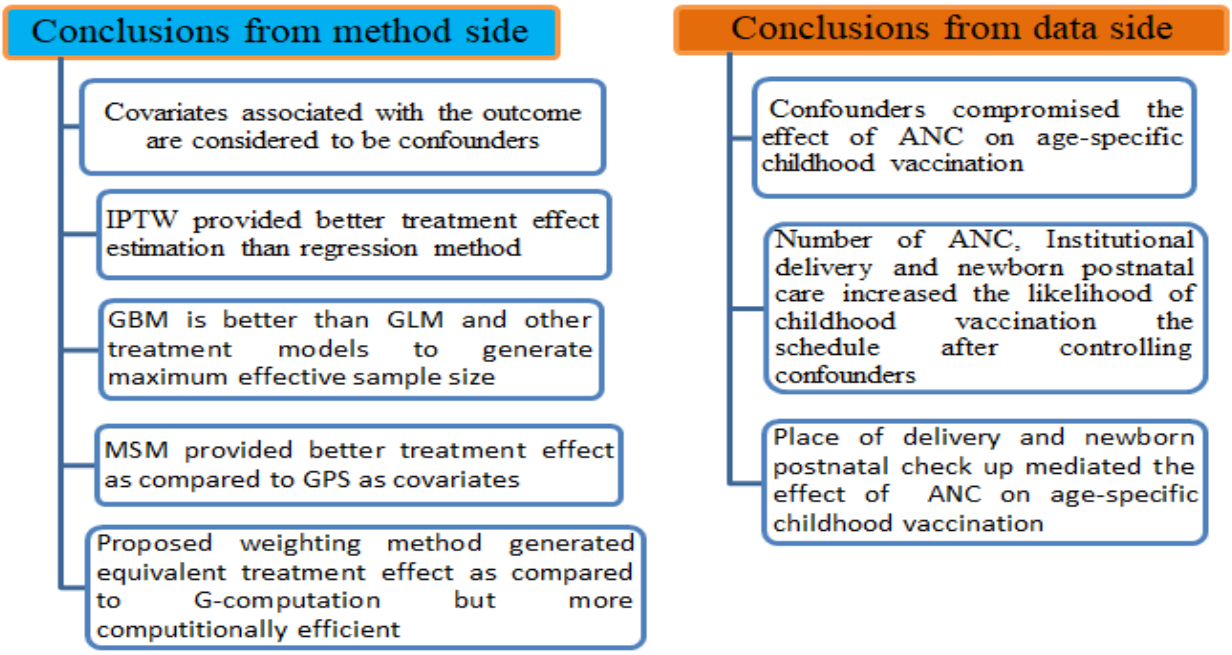
Exploring and identifying confounders in causal inference are the important steps before adjusting their confounding bias and estimate the true treatment causal effect on the outcome. The finding in this dissertation shows that under binary and count treatments, covariates that are associated with the outcome are the real confounders and generate accurate treatment effect estimation. In the meantime change in treatment effect estimate methods of confounder identification selects smaller number of confounders than the significance testing method which implied that change in estimate method is more conservative approach of confounder identification.

Under binary treatment, there is no best weighting method to adjust confounders except entropy balancing consistently balances confounders in simulation and actual data. In the simulation study, it is found that there is tradeoff between confounder balancing between treatment and control groups and accuracy of treatment effect estimation when using GBM based weighting.

To estimate GPS for count treatment, different statistical techniques have been tested and compared their performance in terms of balancing confounders, effective sample size after weighting and treatment effect estimation. The methods include maximum likelihood (including GLM for count models and zero augmented models), gradient boosted model (GBM) and generalized methods of moments or empirical likelihood (CBPS and npCBPS). Except npCBPS, the proposed methods balanced confounders on average in the simulation study and all methods in the actual data. However, GBM retained the largest effective sample size after weighting with IPTW generated from GPS. In addition, the GBM outperformed other methods in count treatment effect estimation on the outcome.

The dissertation also explored mediation analysis using single and sequential binary mediators. The actual data analysis shows that the causal effect of number of ANC on age-specific childhood vaccination was mediated by institutional delivery such that the causal effect of the treatment was decomposed into direct and indirect effects. In sequential mediators, a weighting method was proposed and compared with simulation based G-computation to estimate path specific effects. The weighting method involves sequentially weighting the outcome and mediators model with weights generated by treatment and each mediator and then by multiplying each weight to create composite weight for the outcome and second mediator. The finding shows that the proposed weighting method generated equivalent or better treatment effect estimation as compared to G-computation. In addition, the proposed weighting method is computationally efficient as compared to G-computation which is computationally intensive in the case of sequential mediators. Finally, the causal effect of ANC is disentangled into the indirect effect through institutional delivery, newborn postnatal care, a combination of institutional delivery and newborn postnatal care as well as direct effect.

Generally, the overall conclusion is summarized in the following figure. The conclusions are presented in the data side and method side.



6.2. Limitations

Certain limitations of the dissertation require noting. The first limitation is the dissertation focused on average treatment effect that does not consider other variants of treatment effects such as average treatment effect on the treated and control. The second limitation is the proposed methods are applied and tested when controlling confounders using IPTW generated from (G)PS. However, there are other variants of confounder adjustment including matching, stratifying, and others based on (G)PS. Hence, all conclusions in the study may not be reproducible for other variants of controlling confounders' bias using (G)PS. The third limitation is there is no parametric distribution to generate/simulate ordinal outcome. We rather generate latent continuous outcome from Gaussian distribution and then reclassify the continuous outcome to generate the ordered outcome. In this case, the conclusions may not be generalizable to other outcome types. In addition, there is no unique set up to generate/simulate artificial data including sample size, type and number of covariates, and the effect size of each covariate and treatment. Hence, the conclusions of this study may not be generalizable to other simulation set ups. Misspecification of treatment and outcome models and complex relationships are not tested to compare the proposed model performance. Plasmode simulation is not also replicated for different sample sizes.

Besides, the study does not consider other variants of ordinal models including adjacent-Categories and Continuation-Ratio ordinal models.

6.3. Future Works

The current form of the proposed methods in this dissertation, like any other methods, presents several opportunities for further improvements to enhance its general usage. However, whatever modifications or extensions intended at this stage shall be addressed in future research works. In generalized propensity score estimation for count treatment and confounders, it will be important to use support vector machines, k-nearest neighbors, random forests, artificial neural networks [238] and other machine learning methods suitable for count treatment variables. Estimating count treatment effect using variant of matching algorism based on generalized propensity score will be the future work. In addition, considering count treatment for various outcome type and data nature will be the topic. The application of machine learning algorithms for ordinal outcome is unnoticed in the literature that will be the potential topic in future research.

REFERENCES

1. Imbens, G.W. and D.B. Rubin, *Causal inference in statistics, social, and biomedical sciences*. 2015: Cambridge University Press.
2. Cook, T.D., D.T. Campbell, and W. Shadish, *Experimental and quasi-experimental designs for generalized causal inference*. Vol. 1195. 2002: Houghton Mifflin Boston, MA.
3. Austin, P.C., *An introduction to propensity score methods for reducing the effects of confounding in observational studies.* Multivariate behavioral research, 2011. 46(3): p. 399-424.
4. Deaton, A. and N. Cartwright, *Understanding and misunderstanding randomized controlled trials.* Social science & medicine, 2018. 210: p. 2-21.
5. Ranapurwala, S.I., *Identifying and addressing confounding bias in violence prevention research.* Current epidemiology reports, 2019. 6: p. 200-207.
6. Tennant, P.W., et al., *Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations.* International journal of epidemiology, 2021. 50(2): p. 620-632.
7. Rothman, K.J., *Validity in epidemiologic studies.* Modern epidemiology, 2008: p. 128-147.
8. VanderWeele, T.J., *Principles of confounder selection.* European journal of epidemiology, 2019. 34: p. 211-219.
9. Lee, P.H. and I. Burstyn, *Identification of confounder in epidemiologic data contaminated by measurement error in covariates.* BMC medical research methodology, 2016. 16: p. 1-18.
10. Zhang, Z., *Too much covariates in a multivariable model may cause the problem of overfitting.* Journal of thoracic disease, 2014. 6(9).
11. Tong, S. and Y. Lu, *Identification of confounders in the assessment of the relationship between lead exposure and child development.* Annals of Epidemiology, 2001. 11(1): p. 38-45.
12. Wiebe, D.J., *Homicide and suicide risks associated with firearms in the home: a national case-control study.* Annals of emergency medicine, 2003. 41(6): p. 771-782.
13. Culyba, A.J., et al., *Association of future orientation with violence perpetration among male youths in low-resource neighborhoods.* JAMA pediatrics, 2018. 172(9): p. 877-879.
14. Branas, C.C., et al., *Investigating the link between gun possession and gun assault.* American journal of public health, 2009. 99(11): p. 2034-2040.
15. Zhao, Q.-Y., et al., *Propensity score matching with R: conventional methods and new features.* Annals of translational medicine, 2021. 9(9).
16. Posner, M.A., et al., *Comparing standard regression, propensity score matching, and instrumental variables methods for determining the influence of mammography on stage of diagnosis.* Health Services and Outcomes Research Methodology, 2001. 2: p. 279-290.

17. Rubin, D.B., *Using propensity scores to help design observational studies: application to the tobacco litigation.* Health Services and Outcomes Research Methodology, 2001. 2: p. 169-188.
18. Rosenbaum, P.R. and D.B. Rubin, *The central role of the propensity score in observational studies for causal effects.* Biometrika, 1983. 70(1): p. 41-55.
19. Fu, E.L., et al., *Merits and caveats of propensity scores to adjust for confounding.* Nephrology Dialysis Transplantation, 2019. 34(10): p. 1629-1635.
20. Imai, K. and M. Ratkovic, *Covariate balancing propensity score.* Journal of the Royal Statistical Society Series B: Statistical Methodology, 2014. 76(1): p. 243-263.
21. Imbens, G.W., *The role of the propensity score in estimating dose-response functions.* Biometrika, 2000. 87(3): p. 706-710.
22. Lechner, M., *Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies.* Review of Economics and Statistics, 2002. 84(2): p. 205-220.
23. Imai, K. and D.A. Van Dyk, *Causal inference with general treatment regimes: Generalizing the propensity score.* Journal of the American Statistical Association, 2004. 99(467): p. 854-866.
24. Tchernis, R., M. Horvitz-Lennon, and S.L.T. Normand, *On the use of discrete choice models for causal inference.* Statistics in medicine, 2005. 24(14): p. 2197-2212.
25. McCaffrey, D.F., et al., *A tutorial on propensity score estimation for multiple treatments using generalized boosted models.* Statistics in medicine, 2013. 32(19): p. 3388-3414.
26. Hirano, K. and G.W. Imbens, *The propensity score with continuous treatments.* Applied Bayesian modeling and causal inference from incomplete-data perspectives, 2004. 226164: p. 73-84.
27. Zhu, Y., D.L. Coffman, and D. Ghosh, *A boosting algorithm for estimating generalized propensity scores with continuous treatments.* Journal of causal inference, 2015. 3(1): p. 25-40.
28. Wu, X., et al., *Matching on generalized propensity scores with continuous exposures.* Journal of the American Statistical Association, 2022: p. 1-29.
29. Popovic, M., et al., *Applied causal inference methods for sequential mediators.* 2021.
30. Baron, R.M. and D.A. Kenny, *The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.* Journal of personality and social psychology, 1986. 51(6): p. 1173.
31. Robins, J.M. and S. Greenland, *Identifiability and exchangeability for direct and indirect effects.* Epidemiology, 1992. 3(2): p. 143-155.
32. VanderWeele, T.J. and S. Vansteelandt, *Conceptual issues concerning mediation, interventions and composition.* Statistics and its Interface, 2009. 2(4): p. 457-468.

33. VanderWeele, T.J., *Marginal structural models for the estimation of direct and indirect effects.* Epidemiology, 2009. 20(1): p. 18-26.
34. Imai, K., L. Keele, and D. Tingley, *A general approach to causal mediation analysis.* Psychological methods, 2010. 15(4): p. 309.
35. VanderWeele, T.J., *Bias formulas for sensitivity analysis for direct and indirect effects.* Epidemiology, 2010. 21(4): p. 540-551.
36. VanderWeele, T.J. and S. Vansteelandt, *Odds ratios for mediation analysis for a dichotomous outcome.* American journal of epidemiology, 2010. 172(12): p. 1339-1348.
37. Valeri, L. and T.J. VanderWeele, *Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros.* Psychological methods, 2013. 18(2): p. 137.
38. De Stavola, B.L., et al., *Mediation analysis with intermediate confounding: structural equation modeling viewed through the causal inference lens.* American journal of epidemiology, 2015. 181(1): p. 64-80.
39. VanderWeele, T. and S. Vansteelandt, *Mediation analysis with multiple mediators.* Epidemiologic methods, 2014. 2(1): p. 95-115.
40. Albert, J.M., et al., *Generalized causal mediation and path analysis: Extensions and practical considerations.* Statistical Methods in Medical Research, 2019. 28(6): p. 1793-1807.
41. VanderWeele, T.J., Y. Zhang, and P. Lim, *Brief report: mediation analysis with an ordinal outcome.* Epidemiology, 2016. 27(5): p. 651-655.
42. Stanghellini, E. and M. Kateri, *Exact mediation analysis for ordinal outcome and binary mediator.* Epidemiology, 2022. 33(6): p. 840-842.
43. Zhou, Y., et al., *Causal Mediation Analysis for an Ordinal Outcome with Multiple Mediators.* Structural Equation Modeling: A Multidisciplinary Journal, 2024. 31(2): p. 205-216.
44. Abir, T., et al., *The impact of antenatal care, iron–folic acid supplementation and tetanus toxoid vaccination during pregnancy on child mortality in Bangladesh.* PloS one, 2017. 12(11): p. e0187090.
45. Wang, H., et al., *Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015.* The Lancet, 2016. 388(10053): p. 1725-1774.
46. Mbengue, M.A.S., et al., *Determinants of complete immunization among senegalese children aged 12–23 months: evidence from the demographic and health survey.* BMC public health, 2017. 17: p. 1-9.
47. Who, *Children: reducing mortality*. 2018.

48. Okafor, A.T., *Antenatal Care and Maternal Sociocultural Determinants of Childhood Immunization in Northern Nigeria*. 2019, Walden University.
49. Health, F.M.o., *Ethiopia national expanded programme on immunization*. 2015, BMJ Publishing Group FMOE Addis Ababa.
50. Sullivan, M.-C., et al., *Minding the immunization gap: family characteristics associated with completion rates in rural Ethiopia.* Journal of Community Health, 2010. 35: p. 53-59.
51. Lake, M.W., et al., *Factors for low routine immunization performance: a community-based cross-sectional study in Dessie town, south Wollo zone, Ethiopia, 2014.* Adv Appl Sci, 2016. 1(1): p. 7-17.
52. Tesfaye, T.D., W.A. Temesgen, and A.S. Kasa, *Vaccination coverage and associated factors among children aged 12–23 months in Northwest Ethiopia.* Human vaccines & immunotherapeutics, 2018. 14(10): p. 2348-2354.
53. Gualu, T. and A. Dilie, *Vaccination coverage and associated factors among children aged 12–23 months in debre markos town, Amhara regional state, Ethiopia.* Advances in Public Health, 2017. 2017.
54. Wado, Y.D., M.F. Afework, and M.J. Hindin, *Childhood vaccination in rural southwestern Ethiopia: the nexus with demographic factors and women's autonomy.* The Pan African Medical Journal, 2014. 17(Suppl 1).
55. Lakew, Y., A. Bekele, and S. Biadgilign, *Factors influencing full immunization coverage among 12–23 months of age children in Ethiopia: evidence from the national demographic and health survey in 2011.* BMC public health, 2015. 15: p. 1-8.
56. Mbengue, M.A.S., et al., *Vaccination coverage and immunization timeliness among children aged 12-23 months in Senegal: a Kaplan-Meier and Cox regression analysis approach.* The Pan African Medical Journal, 2017. 27(Suppl 3).
57. Austin, P.C. and M.M. Mamdani, *A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use.* Statistics in medicine, 2006. 25(12): p. 2084-2106.
58. Rosenbaum, P.R., *Model-based direct adjustment.* Journal of the American statistical Association, 1987. 82(398): p. 387-394.
59. Shi, A.X., P.N. Zivich, and H. Chu, *A comprehensive review and tutorial on confounding adjustment methods for estimating treatment effects using observational data.* Applied Sciences, 2024. 14(9): p. 3662.
60. Thoemmes, F. and A.D. Ong, *A primer on inverse probability of treatment weighting and marginal structural models.* Emerging Adulthood, 2016. 4(1): p. 40-59.

61. Hernán, M.Á., B. Brumback, and J.M. Robins, *Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men*. 2000, LWW. p. 561-570.
62. Crump, R.K., et al., *Dealing with limited overlap in estimation of average treatment effects.* Biometrika, 2009. 96(1): p. 187-199.
63. Lechner, M. and A. Strittmatter, *Practical procedures to deal with common support problems in matching estimation.* Econometric Reviews, 2019. 38(2): p. 193-207.
64. Cole, S.R. and M.A. Hernán, *Constructing inverse probability weights for marginal structural models.* American journal of epidemiology, 2008. 168(6): p. 656-664.
65. Stürmer, T., et al., *Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs.* Journal of internal medicine, 2014. 275(6): p. 570-580.
66. Olmos, A. and P. Govindasamy, *A practical guide for using propensity score weighting in R.* Practical assessment, research, and evaluation, 2015. 20(1).
67. Burton, A., et al., *The design of simulation studies in medical statistics.* Statistics in medicine, 2006. 25(24): p. 4279-4292.
68. Morris, T.P., I.R. White, and M.J. Crowther, *Using simulation studies to evaluate statistical methods.* Statistics in medicine, 2019. 38(11): p. 2074-2102.
69. Gadbury, G.L., et al., *Evaluating statistical methods using plasmode data sets in the age of massive public databases: an illustration using false discovery rates.* PLoS genetics, 2008. 4(6): p. e1000098.
70. Vaughan, L.K., et al., *The use of plasmodes as a supplement to simulations: a simple example evaluating individual admixture estimation methodologies.* Computational statistics & data analysis, 2009. 53(5): p. 1755-1766.
71. Schreck, N., et al., *Statistical plasmode simulations–Potentials, challenges and recommendations.* Statistics in Medicine, 2024. 43(9): p. 1804-1825.
72. EPHI, I., *Ethiopian Public Health Institute (EPHI)[Ethiopia] and ICF.* Ethiopia Mini Demographic and Health Survey 2019: Key Indicators, 2019.
73. Rubin, D.B., *Inference and missing data.* Biometrika, 1976. 63(3): p. 581-592.
74. Fielding, S., et al., *Simple imputation methods were inadequate for missing not at random (MNAR) quality of life data.* Health and Quality of Life Outcomes, 2008. 6: p. 1-9.
75. Rubin, D.B. *Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse*. in *Proceedings of the survey research methods section of the American Statistical Association*. 1978. American Statistical Association Alexandria, VA, USA.

76. Leurent, B., et al., *Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial.* Pharmacoeconomics, 2018. 36(8): p. 889-901.
77. Ji, F., S. Rabe-Hesketh, and A. Skrondal, *Diagnosing and handling common violations of missing at random.* psychometrika, 2023. 88(4): p. 1123-1143.
78. Graham, J.W., *Missing data analysis: Making it work in the real world.* Annual review of psychology, 2009. 60: p. 549-576.
79. Lee, J.H. and J.C. Huber Jr, *Evaluation of multiple imputation with large proportions of missing data: how much is too much?* Iranian journal of public health, 2021. 50(7): p. 1372.
80. Faria, R., et al., *A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials.* Pharmacoeconomics, 2014. 32(12): p. 1157-1170.
81. White, I.R., P. Royston, and A.M. Wood, *Multiple imputation using chained equations: issues and guidance for practice.* Statistics in medicine, 2011. 30(4): p. 377-399.
82. Fairclough, D.L., *Design and analysis of quality of life studies in clinical trials*. 2010: Chapman and Hall/CRC.
83. Allison, P.D. *Handling missing data by maximum likelihood*. in *SAS global forum*. 2012. San Diego, CA, USA:.
84. Corpopration, S., *Stata Multiple-imputation Reference Manual: Release 11*. 2009: Stata Press.
85. Anichukwu, O.I. and B.O. Asamoah, *The impact of maternal health care utilisation on routine immunisation coverage of children in Nigeria: a cross-sectional study.* BMJ open, 2019. 9(6): p. e026324.
86. Budu, E., et al., *Maternal healthcare utilsation and complete childhood vaccination in sub-Saharan Africa: a cross-sectional study of 29 nationally representative surveys.* BMJ open, 2021. 11(5): p. e045992.
87. Dheresa, M., et al., *Child vaccination coverage, trends and predictors in Eastern Ethiopia: Implication for sustainable development goals.* Journal of Multidisciplinary Healthcare, 2021: p. 2657-2667.
88. Farida, F., V. Widyaningsih, and B. Murti, *The Effect of Maternal Education and Antenatal Care on Basic Immunization Completeness in Children aged 12-23 Months in Asian and African: Meta-Analysis.* Journal of Maternal and Child Health, 2020. 5(6): p. 614-628.
89. Jimma, M.S., et al., *Full vaccination coverage and associated factors among 12-to-23-Month Children at Assosa Town, Western Ethiopia, 2020.* Pediatric Health, Medicine and Therapeutics, 2021: p. 279-288.
90. Heininger, U., K. Stehr, and J. Cherry, *Serious pertussis overlooked in infants.* European journal of pediatrics, 1992. 151: p. 342-343.

91. Heininger, U. and M. Zuberbühler, *Immunization rates and timely administration in pre-school and school-aged children.* European journal of pediatrics, 2006. 165: p. 124-129.
92. Guidance, E., *Scientific panel on childhood immunisation schedule: Diphtheria-tetanus-pertussis (DTP) vaccination.* Stockholm, November, 2009.
93. Paget, W., et al., *A national measles epidemic in Switzerland in 1997: consequences for the elimination of measles by the year 2007.* Eurosurveillance, 2000. 5(2): p. 17-20.
94. Iyassu, A.S., et al., *Identifying confounders and estimating the causal effect of antenatal care on age-specific childhood vaccination.* Frontiers in Public Health, 2025. 13: p. 1420567.
95. Pearl, J. and A. Paz, *Confounding equivalence in causal inference.* Journal of Causal Inference, 2014. 2(1): p. 75-93.
96. Kamangar, F., *Confounding variables in epidemiologic studies: basics and beyond.* Archives of Iranian medicine, 2012. 15(8): p. 508.
97. Meuli, L. and F. Dick, *Understanding confounding in observational studies.* European Journal of Vascular and Endovascular Surgery, 2018. 55(5): p. 737.
98. Maldonado, G. and S. Greenland, *Simulation study of confounder-selection strategies.* American journal of epidemiology, 1993. 138(11): p. 923-936.
99. Shortreed, S.M. and A. Ertefaie, *Outcome-adaptive lasso: variable selection for causal inference.* Biometrics, 2017. 73(4): p. 1111-1122.
100. Jacob, D., *Variable Selection for Causal Inference via Outcome-Adaptive Random Forest.* arXiv preprint arXiv:2109.04154, 2021.
101. Greenland, S., *Invited commentary: variable selection versus shrinkage in the control of multiple confounders.* American journal of epidemiology, 2008. 167(5): p. 523-529.
102. Greenland, S. and N. Pearce, *Statistical foundations for model-based adjustments.* Annual review of public health, 2015. 36: p. 89-108.
103. Zeileis, A., C. Kleiber, and S. Jackman, *Regression models for count data in R.* Journal of statistical software, 2008. 27(8): p. 1-25.
104. Zeileis, A., C. Kleiber, and S. Jackman, *Regression models for count data in R.* Journal of statistical software, 2008. 27: p. 1-25.
105. Agresti, A., *Foundations of linear and generalized linear models*. 2015: John Wiley & Sons.
106. Sumi, S.N., N.C. Sinha, and M.A. Islam, *Generalized linear models for analyzing count data of rainfall occurrences.* SN Applied Sciences, 2021. 3(4): p. 481.
107. Long, D.L., et al., *A marginalized zero-inflated Poisson regression model with overall exposure effects.* Statistics in medicine, 2014. 33(29): p. 5151-5165.
108. Eggers, J., *On statistical methods for zero-inflated models*. 2015.

109. Christensen, R.H.B., *Cumulative link models for ordinal regression with the R package ordinal.* Submitted in J. Stat. Software, 2018. 35: p. 1-46.
110. Kim, D., *Categorical data analysis, by AGRESTI, ALAN*. 2013, Oxford University Press.
111. Peterson, B. and F.E. Harrell Jr, *Partial proportional odds models for ordinal response variables.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 1990. 39(2): p. 205-217.
112. Christensen, R.H.B., *Cumulative link models for ordinal regression with the R package ordinal.* Submitted in J. Stat. Software, 2018. 35.
113. Akaike, H., *A new look at the statistical model identification.* IEEE transactions on automatic control, 2003. 19(6): p. 716-723.
114. Schwarz, G., *Estimating the dimension of a model.* The annals of statistics, 1978: p. 461-464.
115. Acquah, H.d.-G., *Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship.* 2010.
116. Acquah, H.d.-G., *A bootstrap approach to evaluating the performance of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in selection of an asymmetric price relationship.* 2012.
117. Greenland, S., J. Pearl, and J.M. Robins, *Causal diagrams for epidemiologic research.* Epidemiology, 1999. 10(1): p. 37-48.
118. Glymour, M.M., J. Weuve, and J.T. Chen, *Methodological challenges in causal research on racial and ethnic patterns of cognitive trajectories: measurement, selection, and bias.* Neuropsychology review, 2008. 18: p. 194-213.
119. Franklin, J.M., et al., *Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases.* Computational statistics & data analysis, 2014. 72: p. 219-226.
120. Iyassu, A.S., et al., *Identification of confounders and estimating the causal effect of place of birth on age-specific childhood vaccination.* BMC Medical Informatics and Decision Making, 2024. 24(1): p. 406.
121. Talbot, D., et al., *The change in estimate method for selecting confounders: A simulation study.* Statistical methods in medical research, 2021. 30(9): p. 2032-2044.
122. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern epidemiology*. Vol. 3. 2008: Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia.
123. Porta, M., *A dictionary of epidemiology*. 2014: Oxford university press.
124. Rubin, D.B., *Causal inference using potential outcomes: Design, modeling, decisions.* Journal of the American Statistical Association, 2005. 100(469): p. 322-331.
125. Smith, T.J., D.A. Walker, and C.M. McKenna, *An exploration of link functions used in ordinal regression.* Journal of Modern Applied Statistical Methods, 2020. 18(1): p. 20.

126. Greifer, N., *Using weightit to estimate balancing weights*.
127. Greifer, N., *Covariate balance tables and plots: a guide to the cobalt package.* Accessed March, 2020. 10: p. 2020.
128. Rubin, D.B., *For objective causal inference, design trumps analysis.* 2008.
129. Atiquzzaman, M., et al., *Using external data to incorporate unmeasured confounders: a plasmode simulation study comparing alternative approaches to impute body max index in a study of the relationship between osteoarthritis and cardiovascular disease.* J Stat Res, 2020. 54(2): p. 131-145.
130. Desai, R.J., et al., *Evaluating the use of bootstrapping in cohort studies conducted with 1: 1 propensity score matching—A plasmode simulation study.* Pharmacoepidemiology and Drug Safety, 2019. 28(6): p. 879-886.
131. Bickel, P.J. and A. Sakov, *On the choice of m in the m out of n bootstrap and confidence bounds for extrema.* Statistica Sinica, 2008: p. 967-985.
132. Gasparini, A., *rsimsum: Summarise results from Monte Carlo simulation studies.* Journal of Open Source Software, 2018. 3(26): p. 739.
133. Gasparini, A., I.R. White, and M.A. Gasparini, *Package 'rsimsum'.* 2024.
134. Fukuda-Parr, S., *From the Millennium Development Goals to the Sustainable Development Goals: shifts in purpose, concept, and politics of global goal setting for development.* Gender & Development, 2016. 24(1): p. 43-52.
135. Bogler, L., et al., *Estimating the effect of measles vaccination on child growth using 191 DHS from 65 low-and middle-income countries.* Vaccine, 2019. 37(35): p. 5073-5088.
136. Shen, A.K., R. Fields, and M. McQuestion, *The future of routine immunization in the developing world: challenges and opportunities.* Global Health: Science and Practice, 2014. 2(4): p. 381-394.
137. Utazi, C.E., et al., *High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries.* Vaccine, 2018. 36(12): p. 1583-1591.
138. Organization, W.H., *Postnatal care for mothers and newborns: Highlights from the World Health Organization 2013 Guidelines.* Postnatal Care Guidelines, 2015. 4.
139. WHO, *Newborns: Improving survival and well-being*. 2020.
140. Yenit, M.K., Y.A. Gelaw, and A.M. Shiferaw, *Mothers' health service utilization and attitude were the main predictors of incomplete childhood vaccination in east-central Ethiopia: a case-control study.* Archives of Public Health, 2018. 76: p. 1-9.
141. Asresie, M.B., G.W. Dagnew, and Y.A. Bekele, *Changes in immunization coverage and contributing factors among children aged 12–23 months from 2000 to 2019, Ethiopia: Multivariate decomposition analysis.* PLoS One, 2023. 18(9): p. e0291499.

142. Schön, M., et al., *How to ensure full vaccination? The association of institutional delivery and timely postnatal care with childhood vaccination in a cross-sectional study in rural Bihar, India.* PLOS Global Public Health, 2022. 2(5): p. e0000411.
143. Bantie, B., et al., *Mapping geographical inequalities of incomplete immunization in Ethiopia: a spatial with multilevel analysis.* Frontiers in Public Health, 2024. 12: p. 1339539.
144. Desalew, A., et al., *Incomplete vaccination and its predictors among children in Ethiopia: a systematic review and meta-analysis.* Global Pediatric Health, 2020. 7: p. 2333794X20968681.
145. Digitale, J.C., J.N. Martin, and M.M. Glymour, *Tutorial on directed acyclic graphs.* Journal of Clinical Epidemiology, 2022. 142: p. 264-267.
146. Hainmueller, J., *Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies.* Political analysis, 2012. 20(1): p. 25-46.
147. Lee, J., et al., *Inverse Probability of Treatment Weighting with Deep Sequence Models Enables Accurate treatment effect Estimation from Electronic Health Records.* arXiv preprint arXiv:2406.08851, 2024.
148. Hahn, J., *On the role of the propensity score in efficient semiparametric estimation of average treatment effects.* Econometrica, 1998: p. 315-331.
149. Heckman, J.J., H. Ichimura, and P. Todd, *Matching as an econometric evaluation estimator.* The review of economic studies, 1998. 65(2): p. 261-294.
150. Rubin, D.B., *Using multivariate matched sampling and regression adjustment to control bias in observational studies.* Journal of the American Statistical Association, 1979. 74(366a): p. 318-328.
151. McCaffrey, D.F., G. Ridgeway, and A.R. Morral, *Propensity score estimation with boosted regression for evaluating causal effects in observational studies.* Psychological methods, 2004. 9(4): p. 403.
152. Hansen, L.P., *Large sample properties of generalized method of moments estimators.* Econometrica: Journal of the econometric society, 1982: p. 1029-1054.
153. Setodji, C.M., et al., *The right tool for the job: choosing between covariate-balancing and generalized boosted model propensity scores.* Epidemiology, 2017. 28(6): p. 802-811.
154. Hainmueller, J. and Y. Xu, *Ebalance: A Stata package for entropy balancing.* Journal of Statistical Software, 2013. 54: p. 1-18.
155. Ju, C., et al., *Propensity score prediction for electronic healthcare databases using super learner and high-dimensional propensity score methods.* Journal of Applied Statistics, 2019. 46(12): p. 2216-2236.
156. Polley, E.C. and M.J. Van der Laan, *Super learner in prediction.* 2010.

157. Yucel Karakaya, S.P. and I. Unal, *Balance diagnostics in propensity score analysis following multiple imputation: A new method.* Pharmaceutical Statistics, 2024. 23(5): p. 763-777.
158. Austin, P.C., *Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples.* Statistics in medicine, 2009. 28(25): p. 3083-3107.
159. Austin, P.C. and E.A. Stuart, *Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies.* Statistics in medicine, 2015. 34(28): p. 3661-3679.
160. Zhang, Z., et al., *Balance diagnostics after propensity score matching.* Annals of translational medicine, 2019. 7(1): p. 16.
161. Austin, P.C., *Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures.* Statistical methods in medical research, 2019. 28(5): p. 1365-1377.
162. Rosenbaum, P.R. and D.B. Rubin, *Reducing bias in observational studies using subclassification on the propensity score.* Journal of the American statistical Association, 1984. 79(387): p. 516-524.
163. Robins, J.M., *Association, causation, and marginal structural models.* Synthese, 1999. 121(1/2): p. 151-179.
164. VanderWeele, T.J. and P. Ding, *Sensitivity analysis in observational research: introducing the E-value.* Annals of internal medicine, 2017. 167(4): p. 268-274.
165. Vale, C.C.R., N.K.d.O. Almeida, and R.M.V.R.d. Almeida, *On the use of the E-value for sensitivity analysis in epidemiologic studies.* Cadernos de Saúde Pública, 2021. 37: p. e00294720.
166. Agresti, A. and C. Tarantola, *Simple ways to interpret effects in modeling ordinal categorical data.* Statistica Neerlandica, 2018. 72(3): p. 210-223.
167. Robins, J.M., M.A. Hernan, and B. Brumback, *Marginal structural models and causal inference in epidemiology.* Epidemiology, 2000: p. 550-560.
168. Austin, P.C., *A comparison of 12 algorithms for matching on the propensity score.* Statistics in medicine, 2014. 33(6): p. 1057-1069.
169. Choi, B.Y., et al., *Power comparison for propensity score methods.* Computational Statistics, 2019. 34: p. 743-761.
170. Kwak, D., et al., *Comparing Machine Learning and Advanced Methods with Traditional Methods to Generate Weights in Inverse Probability of Treatment Weighting: The INFORM Study.* Pragmatic and Observational Research, 2024: p. 173-183.

171. Cannas, M. and B. Arpino, *A comparison of machine learning algorithms and covariate balance measures for propensity score matching and weighting.* Biometrical Journal, 2019. 61(4): p. 1049-1072.

172. Stone, C.A. and Y. Tang, *Comparing propensity score methods in balancing covariates and recovering impact in small sample educational program evaluations.* Practical Assessment, Research, and Evaluation, 2013. 18(1).

173. Phway, P., et al., *Continuum of care of mothers and immunization status of their children: A secondary analysis of 2015–2016 Myanmar Demographic and Health Survey.* Public Health in Practice, 2022. 4: p. 100335.

175. Fong, C., C. Hazlett, and K. Imai, *Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements.* The Annals of Applied Statistics, 2018. 12(1): p. 156-177.

176. Lechner, M., *Identification and estimation of causal effects of multiple treatments under the conditional independence assumption*. 2001: Springer.

177. Yusuf, O., T. Bello, and O. Gureje, *Zero inflated poisson and zero inflated negative binomial models with application to number of falls in the elderly.* Biostatistics and Biometrics Open Access Journal, 2017. 1(4): p. 69-75.

178. Chambers, J., T. Hastie, and D. Pregibon. *Statistical models in S*. in *Compstat: proceedings in computational statistics, 9th symposium held at Dubrovnik, Yugoslavia, 1990*. 1992. Springer.

179. Smith, J.A. and P.E. Todd, *Does matching overcome LaLonde's critique of nonexperimental estimators?* Journal of econometrics, 2005. 125(1-2): p. 305-353.

180. Owen, A.B., *Empirical likelihood*. 2001: Chapman and Hall/CRC.

181. Hastie, T., et al., *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. 2009: Springer.

182. Ridgeway, G., *Generalized Boosted Models: A guide to the gbm package.* Update, 2007. 1(1): p. 2007.

183. Zhang, Z., et al., *Causal inference with a quantitative exposure.* Statistical methods in medical research, 2016. 25(1): p. 315-335.

184. Bia, M. and A. Mattei, *A Stata package for the estimation of the dose-response function through adjustment for the generalized propensity score.* The Stata Journal, 2008. 8(3): p. 354-373.

185. Ridgeway, G., et al., *Toolkit for Weighting and Analysis of Nonequivalent Groups: A guide to the twang package.* Vignette, 2021. 2021: p. 26.

186. Joffe, M.M. and P.R. Rosenbaum, *Invited commentary: propensity scores.* American journal of epidemiology, 1999. 150(4): p. 327-333.

187. Chen, J., et al., *Estimating individualized treatment rules for ordinal treatments.* Biometrics, 2018. 74(3): p. 924-933.
188. Feng, P., et al., *Generalized propensity score for estimating the average treatment effect of multiple treatments.* Statistics in medicine, 2012. 31(7): p. 681-697.
189. Hernán, M.A. and J.M. Robins, *Estimating causal effects from epidemiological data.* Journal of Epidemiology & Community Health, 2006. 60(7): p. 578-586.
190. Igelström, E., et al., *Causal inference and effect estimation using observational data.* J Epidemiol Community Health, 2022. 76(11): p. 960-966.
191. Chesnaye, N.C., et al., *An introduction to inverse probability of treatment weighting in observational research.* Clinical Kidney Journal, 2022. 15(1): p. 14-20.
192. Liu, W., S.J. Kuramoto, and E.A. Stuart, *An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research.* Prevention science, 2013. 14: p. 570-580.
193. Gaster, T., et al., *Quantifying the impact of unmeasured confounding in observational studies with the E value.* BMJ medicine, 2023. 2(1).
194. Chung, W.T. and K.C. Chung, *Key concepts in clinical epidemiology: the use of the E-value for sensitivity analysis.* Journal of Clinical Epidemiology, 2023. 163: p. 92-94.
195. Ding, P. and T.J. VanderWeele, *Sensitivity analysis without assumptions.* Epidemiology, 2016. 27(3): p. 368-377.
196. Norton, E.C., M.M. Miller, and L.C. Kleinman, *Computing adjusted risk ratios and risk differences in Stata.* The Stata Journal, 2013. 13(3): p. 492-509.
197. D'Agostino McGowan, L., *Sensitivity analyses for unmeasured confounders.* Current Epidemiology Reports, 2022. 9(4): p. 361-375.
198. Lucy, D. and A. McGowan, *tipr: An R package for sensitivity analyses for unmeasured confounders.* Journal of Open Source Software, 2022. 7(77): p. 4495.
199. Winship, C., *Counterfactuals and causal inference: Methods and principles for social research*. 2007: Cambridge University Press.
200. Holland, P.W., *Statistics and causal inference.* Journal of the American statistical Association, 1986. 81(396): p. 945-960.
201. Williamson, E., et al., *Propensity scores: from naive enthusiasm to intuitive understanding.* Statistical methods in medical research, 2012. 21(3): p. 273-293.
202. Austin, P.C. and E.A. Stuart, *Estimating the effect of treatment on binary outcomes using full matching on the propensity score.* Statistical methods in medical research, 2017. 26(6): p. 2505-2525.

203. Carsey, T.M. and J.J. Harden, *Monte Carlo simulation and resampling methods for social scientists*. 2011, Inter-University Consortium for Political and Social Research. Chapel Hill, NC.
204. Imai, K. and M. Ratkovic, *Robust estimation of inverse probability weights for marginal structural models.* Journal of the American Statistical Association, 2015. 110(511): p. 1013-1023.
205. Stürmer, T., et al., *Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study.* American journal of epidemiology, 2010. 172(7): p. 843-854.
206. Kang, J.D. and J.L. Schafer, *Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data.* 2007.
207. Lunceford, J.K. and M. Davidian, *Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study.* Statistics in medicine, 2004. 23(19): p. 2937-2960.
208. Schafer, J.L. and J. Kang, *Average causal effects from nonrandomized studies: a practical guide and simulated example.* Psychological methods, 2008. 13(4): p. 279.
209. Gabriel, E.E., et al., *Cross-direct effects in settings with two mediators.* Biostatistics, 2023. 24(4): p. 1017-1030.
210. Nozaki, I., M. Hachiya, and T. Kitamura, *Factors influencing basic vaccination coverage in Myanmar: secondary analysis of 2015 Myanmar demographic and health survey data.* BMC public health, 2019. 19: p. 1-8.
211. Noh, J.-W., et al., *Factors affecting complete and timely childhood immunization coverage in Sindh, Pakistan; A secondary analysis of cross-sectional survey data.* PloS one, 2018. 13(10): p. e0206766.
212. Adedire, E.B., et al., *Immunisation coverage and its determinants among children aged 12-23 months in Atakumosa-west district, Osun State Nigeria: a cross-sectional study.* BMC public health, 2016. 16: p. 1-8.
213. Budu, E., et al., *Maternal healthcare utilization and full immunization coverage among 12–23 months children in Benin: a cross sectional study using population-based data.* Archives of Public Health, 2021. 79: p. 1-12.
214. Regassa, N., Y. Bird, and J. Moraros, *Preference in the use of full childhood immunizations in Ethiopia: the role of maternal health services.* Patient preference and adherence, 2019: p. 91-99.
215. Dixit, P., L.K. Dwivedi, and F. Ram, *Strategies to improve child immunization via antenatal care visits in India: a propensity score matching analysis.* PloS one, 2013. 8(6): p. e66175.
216. Brookhart, M.A., et al., *Variable selection for propensity score models.* American journal of epidemiology, 2006. 163(12): p. 1149-1156.

217. Böhnke, J.R., *Explanation in causal inference: methods for mediation and interaction*. 2016, Taylor & Francis.
218. Pearl, J., *The mediation formula: A guide to the assessment of causal pathways in nonlinear models.* Causality: Statistical perspectives and applications, 2012: p. 151-179.
219. Pardo, A. and M. Román, *Reflections on the Baron and Kenny model of statistical mediation.* Anales de psicologia, 2013. 29(2): p. 614-623.
220. Tai, A.-S. and S.-H. Lin, *Complete effect decomposition for an arbitrary number of multiple ordered mediators with time-varying confounders: A method for generalized causal multi-mediation analysis.* Statistical Methods in Medical Research, 2023. 32(1): p. 100-117.
221. VanderWeele, T.J., S. Vansteelandt, and J.M. Robins, *Effect decomposition in the presence of an exposure-induced mediator-outcome confounder.* Epidemiology, 2014. 25(2): p. 300-306.
222. Huang, Y.-T. and H.-I. Yang, *Causal mediation analysis of survival outcome with multiple mediators.* Epidemiology, 2017. 28(3): p. 370-378.
223. Steen, J., et al., *Flexible mediation analysis with multiple mediators.* American journal of epidemiology, 2017. 186(2): p. 184-193.
224. Daniel, R.M., et al., *Causal mediation analysis with multiple mediators.* Biometrics, 2015. 71(1): p. 1-14.
225. Lin, S.-H. and T. VanderWeele, *Interventional approach for path-specific effects.* Journal of Causal Inference, 2017. 5(1): p. 20150027.
226. Tai, A.-S., L.-H. Liao, and S.-H. Lin, *On the Conventional Definition of Path-specific Effects: Fully Mediated Interaction With Multiple Ordered Mediators.* Epidemiology, 2022. 33(6): p. 817-827.
227. Zugna, D., et al., *Applied causal inference methods for sequential mediators.* BMC Medical Research Methodology, 2022. 22(1): p. 301.
228. Tai, A.-S., et al., *G-Computation to Causal Mediation Analysis With Sequential Multiple Mediators—Investigating the Vulnerable Time Window of HBV Activity for the Mechanism of HCV Induced Hepatocellular Carcinoma.* Frontiers in Public Health, 2022. 9: p. 757942.
229. Avin, C., I. Shpitser, and J. Pearl, *Identifiability of path-specific effects.* 2005.
230. Snowden, J.M., S. Rose, and K.M. Mortimer, *Implementation of G-computation on a simulated data set: demonstration of a causal inference technique.* American journal of epidemiology, 2011. 173(7): p. 731-738.
231. Robins, J., *A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect.* Mathematical modelling, 1986. 7(9-12): p. 1393-1512.

232. Chatton, A., et al., *G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study.* Scientific reports, 2020. 10(1): p. 9219.

233. Le Borgne, F., et al., *G-computation and machine learning for estimating the causal effects of binary exposure statuses on binary outcomes.* Scientific reports, 2021. 11(1): p. 1435.

234. Ren, J., et al., *Comparing G-Computation, Propensity Score-Based Weighting, and Targeted Maximum Likelihood Estimation for Analyzing Externally Controlled Trials with Both Measured and Unmeasured Confounders: A Simulation Study.* 2022.

235. Wang, A. and O.A. Arah, *G-computation demonstration in causal mediation analysis.* European journal of epidemiology, 2015. 30: p. 1119-1127.

236. Naimi, A.I., S.R. Cole, and E.H. Kennedy, *An introduction to g methods.* International journal of epidemiology, 2017. 46(2): p. 756-762.

237. Iyassu, A.S., et al., *Causal Effect of Count Treatment on Ordinal Outcome Using Generalized Propensity Score: Application to Number of Antenatal Care and Age Specific Childhood Vaccination.* Journal of Epidemiology and Global Health, 2025. 15(1): p. 23.

238. Sidumo, B., E. Sonono, and I. Takaidza, *Count Regression and Machine Learning Techniques for Zero-Inflated Overdispersed Count Data: Application to Ecological Data.* Annals of Data Science, 2024. 11(3): p. 803-817.

APPENDICES

Appendix : Publication

Iyassu *et al. BMC Medical Informatics and Decision Making* (2024) 24:406
https://doi.org/10.1186/s12911-024-02827-2

BMC Medical Informatics and Decision Making

RESEARCH Open Access

Identification of confounders and estimating the causal effect of place of birth on age-specific childhood vaccination

Ashagrie Sharew Iyassu[1,5*], Haile Mekonnen Fenta[1,2,3], Zelalem G. Dessie[1,4] and Temesgen T. Zewotir[4]

Abstract

Background In causal analyses, some third factor may distort the relationship between the exposure and the outcome variables under study, which gives spurious results. In this case, treatment groups and control groups that receive and do not receive the exposure are different from one another in some other essential variables, called confounders.

Method Place of birth was used as exposure variable and age-specific childhood vaccination status was used as outcome variables. Three approaches of confounder selection techniques such as all pre-treatment covariates, outcome cause covariates, and common cause covariates were proposed. Multiple logistic regression was used to estimate the propensity score for inverse probability treatment weighting (IPTW) confounder adjustment techniques. The proportional odds model was used to estimate the causal effect of place of birth on age-specific childhood vaccination. To validate the result obtained from observed data, we used a plasmode simulation of resampling 1000 samples from actual data 500 times.

Result Outcome cause and common cause confounder identification techniques gave comparable results in terms of treatment effect in the plasmode data. However, outcome causes that contain common causes and predictors of the outcome confounder identification gave relatively better treatment effect results. The treatment effect result in the IPTW confounder adjustment method was better than that of the regression adjustment method. The effect of place of birth on log odds of cumulative probability of age-specific childhood vaccination was 0.36 with odds ratio of 1.43 for higher level vaccination status.

Conclusion It is essential to use plasmode simulation data to validate the reproducibility of the proposed methods on the observed data. It is important to use outcome-cause covariates to adjust their confounding effect on the outcome. Using inverse probability treatment weighting gives unbiased treatment effect results as compared to the regression method of confounder adjustment. Institutional delivery increases the likelihood of childhood vaccination at the recommended schedule.

Keywords Confounder, Causal inference, Plasmode simulation, Place of birth, Vaccination

*Correspondence:
Ashagrie Sharew Iyassu
statashe@gmail.com
[1]College of Science, Bahir Dar University, Bahir Dar, Ethiopia
[2]Center for Environmental and Respiratory Health Research, Population Health, University of Oulu, Oulu, Finland
[3]Biocenter, University of Oulu, Oulu, Finland
[4]School of Mathematics, Statistics & Computer Science, University of KwaZulu Natal, Durban, South Africa
[5]Debremarkos University, Debre Markos, Ethiopia

Journal of Epidemiology and Global Health (2025) 15:23
https://doi.org/10.1007/s44197-025-00344-7

RESEARCH ARTICLE

Causal Effect of Count Treatment on Ordinal Outcome Using Generalized Propensity Score: Application to Number of Antenatal Care and Age Specific Childhood Vaccination

Ashagrie Sharew Iyassu[1,5] · **Haile Mekonnen Fenta**[1,2,3] · **Zelalem G. Dessie**[1,4] · **Temesgen T. Zewotir**[4]

Received: 23 November 2024 / Accepted: 26 December 2024

Abstract

Background Many of the studies in causal inference using propensity scores relied on binary treatments where it is estimated by logistic regression or machine learning algorithms. Since 2000s, attention has been given for multiple values (categorical) and continuous treatments and the propensity score associated with such treatments is called generalized propensity score (GPS). However, there is scant literature on the use of count treatments in causal inference. Besides, effective sample size, after weighting, along with other methods has not been practiced for GPS model performance measure. The study was done with the aim of using count treatments in causal inference; select appropriate GPS and outcome models for such treatment and ordinal outcome.

Method A family of count models and a generalized boosted model (GBM) were used for GPS estimation. Their performance was measured in terms of covariate balancing power, effective sample size and the average treatment effect after GPS-based weighting. Marginal structural modeling (MSM) and covariate adjustment using GPS were used to estimate treatment effect on ordinal outcome. Stabilized inverse probability treatment weighting was used for covariate balancing assessment. Monte Carlo simulation study at various sample sizes with 1000 replication and household survey data were used in the study.

Result GPS was trimmed at 1% and 99% which gave better results as compared to untrimmed results. The generalized boosted model performed well both in simulation and actual data producing a larger effective sample size and smaller metrics when estimating average treatment effect on the outcome. The MSM was found better than GPS as a covariate in the outcome model.

Conclusion It is important to trim GPS when it approaches zero or one without loss of more information due to trimming. Effective sample size after weighting should be used along with other methods such as correlation and absolute standardized mean differences for GPS model selection. GBM should be used for GPS estimation for count treatments. MSM is important for the outcome model when weighting GPS method is used. Finally, the number of antenatal care services had an increasing effect on the probability of age-specific childhood vaccination.

Keywords Generalized propensity score · Ordinal regression · Monte Carlos simulation · Effective sample size · Antenatal Care

1 Introduction

Imbens and Rubin [1] stated causality as it is connected to an action such as treatment or intervention applied to a unit at a particular point in time. A unit can be a person or any physical entity and the same person or physical entity at different time points. Accordingly, the causal statement assumes that when a unit is exposed to treatment or intervention at a particular point in time, the same unit should be exposed to an alternative treatment or exposure at the same point in time.

Randomized control trial (RCT), which is a gold standard research design, is a random allocation of contrasted

✉ Ashagrie Sharew Iyassu
statashe@gmail.com

[1] College of Science, Bahir Dar University, Bahir Dar, Ethiopia

[2] Population Health, Center for Environmental and Respiratory Health Research, University of Oulu, Oulu, Finland

[3] University of Oulu, Oulu, Finland

[4] School of Mathematics, Statistics & Computer Science, University of KwaZulu Natal, Durban, South Africa

[5] Debremarkos University, Debre Markos, Ethiopia

Published online: 12 February 2025

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 30 May 2025
DOI 10.3389/fpubh.2025.1420567

OPEN ACCESS

EDITED BY
Roberto Dias de Oliveira,
State University of Mato Grosso do Sul, Brazil

REVIEWED BY
Shahzad Ali Khan,
Health Services Academy, Pakistan
Andrea da Silva Santos,
Federal University of Grande Dourados, Brazil

*CORRESPONDENCE
Ashagrie Sharew Iyassu
satashe@gmail.com

RECEIVED 20 April 2024
ACCEPTED 07 May 2025
PUBLISHED 30 May 2025

CITATION
Iyassu AS, Mekonnen Fenta H, Dessie ZG and Zewotir TT (2025) Identifying confounders and estimating the causal effect of antenatal care on age-specific childhood vaccination. *Front. Public Health* 13:1420567. doi: 10.3389/fpubh.2025.1420567

Identifying confounders and estimating the causal effect of antenatal care on age-specific childhood vaccination

Ashagrie Sharew Iyassu[1,2]*, Haile Mekonnen Fenta[1,3], Zelalem G. Dessie[1,4] and Temesgen T. Zewotir[4]

[1]College of Science, Bahir Dar University, Bahir Dar, Ethiopia, [2]Department of Statistics, Debre Markos University, Debre Markos, Ethiopia, [3]Population Health, Center for Environmental and Respiratory Health Research, University of Oulu, Oulu, Finland, [4]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Background: Immunization is an efficient and cost-effective public health program. It averts millions of child deaths per year. It is taken as one of the main interventions that can be used to achieve the third Sustainable Development Goal, which is to end preventable deaths of newborns and under-five children by 2030. The study was done with the aim of identifying appropriate confounder identification methods and examining confounders for the causal effect of a number of antenatal care visits on age-specific childhood vaccination.

Methods: A family of generalized linear models with log link functions was used to model the covariate and the number of antenatal care association. A cumulative link model was used to model the number of antenatal care and covariate-age-specific childhood vaccination associations. AIC and BIC values were used to compare models. Significance testing methods and change in estimate methods were used to identify covariates that confound the effect of a number of antenatal care on age-specific childhood vaccinations.

Result: A zero-inflated Poisson model was selected to model covariate–exposure association, and a proportional odds model with a log link was selected to model the outcome variable. Among significance testing methods, the common cause approach yielded smaller values of BIC and a smaller number of covariates. However, the likelihood ratio test showed no difference between the common cause and other approaches. A change in the estimate method is more conservative at a 10% cut point, which selects a smaller number of confounders. However, the significance testing method was better performed than the change in estimate method.

Conclusion: The significance testing method with a *p*-value of less than or equal to 0.2 performed better than a change in estimate method at a 10% cut point of effect change for confounder identification. Mothers' age at first birth, region, place of residence, education status of mothers, presence of radio and television in the household, religion, household size, wealth status, total children ever born, and birth order number are identified as confounders.

KEYWORDS

antenatal care, childhood immunization, confounders, significance testing, change in estimate