Is it possible to improve the analytical approach to the evaluation of cluster-randomised trials where the complexity of the intervention demands a small number of clusters?

The case of the Triage plus Integrated TB-HIV Community Intervention Project in Lilongwe Rural, Malawi

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

George Anubi Faki Bello

June 2015

Abstract

Introduction

In this thesis, analytical approaches for the design and evaluation of cluster randomised trials are presented and reviewed. In particular, statistical power/ sample size issues relating to the design of cluster randomised trials for which only a limited number of clusters are available are assessed using a series of simulation studies. The use of computer simulation methods made it possible to investigate how well cluster randomised trials with limited numbers of clusters available can be optimised both in terms of statistical power and also the accuracy of parameter estimates. The study design conditions performing best in the simulation studies were then applied to a community intervention study involving informal healthcare providers: the 'Triage Plus integrated tuberculosis (TB) and human immuno-deficiency virus (HIV) community intervention project in Lilongwe rural, Malawi'.

Aims and objectives

The general aims of this dissertation were to:

1. investigate if it is possible to improve the analytical approach to the evaluation of cluster-randomised trials where the complexity of the intervention demands a small number of clusters and in which the primary outcome measure is a count of events occurring in a specified time interval;
2. investigate the effectiveness of engaging informal healthcare providers in integrated TB and HIV community intervention in treatment initiation rates and testing access rates, a cluster randomised trial was conducted in Malawi for which only a limited number of clusters were available to the researchers.

The specific objectives were:

1. to review cluster randomised trials and the statistical methods used in the assessment of the effectiveness of the intervention in this type of trial when the primary outcome measure is a count of events occurring in a specified time interval;
2. to assess the statistical efficiencies of different design conditions in terms of statistical power and the accuracy of parameter estimates when determining the effectiveness of complex interventions with a limited number of clusters in this situation;

3. to identify the circumstances under which each of the statistical methods would be most robust in detecting significant intervention effects or providing accurate estimates of intervention effects;
4. to apply these statistical approaches to the data collected in the cluster randomised clinical trial of community based interventions for TB and HIV (the 'Triage Plus' study);
5. to assess the effect of involving non-paid informal healthcare providers in integrated TB and HIV community interventions aimed at improving testing and treatment initiation rates for these two diseases.

Methods

Two research approaches were used in this dissertation:

1. Simulation studies were used to investigate statistical efficiencies in terms of statistical power and accuracy in parameter estimation under different study design conditions for cluster randomised trials in which the primary outcome measure is a count of the number of events occurring during a specified period of time.
2. These statistical approaches were then applied to obtain robust estimates of the effect of the test intervention using the data collected during the "Triage Plus" study. The Triage Plus intervention, implemented in rural areas of Lilongwe, involved informal healthcare providers in an integrated TB and HIV community intervention. This intervention specifically involved empowering the informal healthcare providers in disease recognition, sputum specimen collection, referral of presumptive TB cases, and conducting community TB and HIV awareness meetings.

Results

The simulation studies showed that statistical efficiency and power both varied considerably under the different design conditions investigated. Non-coverage rates within the nominal value of 5% and negligible biases in the estimated fixed effects parameters (regression coefficients) were observed for all scenarios investigated including the (minimal) 3 cluster per arm design. However, it was discovered that, in order to achieve adequate power in low incidence disease conditions such as TB treatment initiation rates, more repeated measurement times were required to achieve adequate power of 80% with a true effect size of 20% or lower (for example, 12 measurement times were needed to achieve adequate power in this situation in a 3 cluster per arm design when the ICC was 0.00154). With an ICC of 0.081

at least 9 clusters were needed to achieve adequate statistical power of ≥80% with an effect size of 20% with 6 and 12 measurement time points respectively for high and low incidence disease conditions. For an effect size of 40%, at least 3 clusters per arm were needed to achieve adequate power with 4 repeated measurement times in low incidence diseases and 3 measurement times for high incidence diseases. For ICCs of 0.321 and above, no adequate statistical power was achieved with an effect size of ≤40% in both high and low disease incidence conditions.

In the analysis of the TB services access data from the Triage plus study, the intervention significantly increased the number of presumptive TB cases accessing testing sites by 15.2% (p=0.003) in the first 12 months of the intervention; however, this was followed by a statistically non-significant reduction of 18.3% (p=0.224) when the intervention was rolled-out into the control clusters. Overall, the intervention was associated with a non-significant increase in TB treatment initiation rates of 18% (p=0.112).

In the analysis of HIV services access rates, antiretroviral therapy (ART) initiation rates increased significantly by 34.7% (p=0.048) in the intervention clusters in the first 12 months of intervention, and the ART initiation rates were similar after rolling-out the intervention to the control clusters. Overall, the intervention was associated with a 61% increase in HIV testing uptake rates (p<0.001).

Conclusion:

To achieve adequate statistical power and improved precision in parameter estimation in cluster randomised trials with a count outcome measure, with the ICC of 0.00154 the simulation results suggested that a minimum of 3 clusters per arm is required with at least 12 measurement times for the estimation of an effect size of 20% (or higher) in low incidence disease situations. However, for high incidence outcomes, a minimum of 3 clusters per arm with 3 or more measurement times may be adequate to achieve a statistical power of at least 80%. For an ICC of 0.081, at least 3 clusters per arm were needed to achieve adequate power if the effect size was 40% after 4 repeated measurement times in low incidence diseases and 3 measurement times for high incidence diseases. With ICCs of 0.321 and above, no adequate statistical power was achieved with an effect size of ≤40% in both high and low disease incidence conditions.

For the TB and HIV interventions in the "Triage Plus" study, engaging informal health care providers was clearly effective in improving TB and HIV testing uptake as well as ART

initiation. This reinforces the need for community participation in integrated TB and HIV interventions to combat the two diseases. However, for these providers to be effective in promoting TB treatment initiation, the number of sites offering TB testing and treatment initiation in rural areas should be increased to make them more accessible to the population.

ACKNOWLEDGEMENTS

A number of individuals and organisations were involved in the implementation of the Triage plus study and are too numerous to mention. In particular, I am highly indebted to theinformal health care providers for their dedication despite not receiving any incentives in the implementation of the project. To them saving lives was their priority. The project could not have progressed well without the support from community leaders (traditional authorities, group village headmenand village headmen) and the national or district level stakeholders (the National TB Programme, HIV Department, the Lilongwe District Health Office, the Lilongwe District Assembly) for their support throughout the project. Their efforts are greatly appreciated. As the project was implemented in rural Lilongwe involving members within the same communities, it is my sincere hope that the knowledge and experience gained by the various individuals will continue toflourish, thereby benefitting the whole community.

Special thanks are due to the Triage plus study team: Lifah Sanudi, Patnice Nkhonjera, Mtisunge Chobambo-Mphande, Mafayo Phiri, Chifundo Chisesa, Linda Chisi, Helecks Mtengo, and Phonex Mkuliwa for their critical roles in the implementation of the project. For the research assistants, Chifundo Chisesa, Linda Chisi, Helecks Mtengo, and Phonex John, I hope the expertise they have aquired through the project will be put into full use in future community randomised interventions. It was not easy for them to fully understand and excecute the intervention using the laid down procedures. Further thanks go to the Lilongwe District Health Office facilitators for supporting the facilitation of informal healthcare providers, health surveillance assistants and supervision of informal healthcare providers during their project implementation. Special recognition goes to Janet Chikonda for her critical role in coordinating the facilitators and ensuring the facilitators were available when needed.

In a special way, may I thank the Research for Equity and Community Health Trust management for their continued support and for nominating me to coordinate the Triage Plus study and eventually to do a PhD, the Liverpool School of Tropical Medicine for conceptualising the intervention and providing technical support throughout the implementation of the Triage Plus study, and the Norwegian Heart and Lung Patient Organisation for funding the Triage plus intervention and my PhD studentship. In conceptualising the intervention and technical and financial support that led to production of

this thesis, may I individually acknowledge the services of Professor Stephen Bertel Squire, Rachael Thompson, and Dr. Brian Faragher from the Liverpool School of Tropical Medicine, and Rasmus Malmborg of the Norwegian Heart and Lung Patient Organisation. These individuals played critical roles in ensuring proper execution of the research project.

In ensuring proper execution of the study, my key role was in designing a proper cluster randomised trial with more than one cluster per arm (i.e. changing the original study design with one cluster assigned per study arm to a design with multiple clusters (although only a limited number of clusters could be used due to a fear of contamination challenges, in fact just 3 clusters per arm). In addition, implementation of the study design, interventions and quality data management procedures was supervised in order to align to standard cluster randomised trial processes (e.g. blinding patient allocation to clusters). Simulation studies that informed the design conditions that were considered during the evaluation of the actual Triage Plus intervention were also conducted.

My big heart-felt appreciation and thanks go to my supervisors Dr Brian Faragher and Professor Bertel Squire for their words of encouragement, inspiration and technical guidance throughout the preparation of this thesis. They accorded me ample time, despite their busy schedules, in shaping the direction of the research, an experience which will be hard to forget. May I also thank Dr Bertha Simwaka, Professor Cam Bowie, Dr. Liz Corbett for their offsite supervision and encouragement. Further thanks go to Rachael Thompson and Helen Rigby for arranging all logistic support for all my annual visits and short courses attended as part of the PhD studentship.

Finally, I would like to thank my wife Ellen and my children Ishmael, Rehema, Haroon, and Munil and my parents, for giving me moral support during this hectic but adventurous time.

Table of Contents

CHAPTER 1 1
GENERAL INTRODUCTION AND CONTEXT 1
1.1 Overview of the Triage Plus study and its objectives 1
1.1.1 Introduction 1
1.1.2 Access to TB and HIV services 1
1.1.3 The relationship between poverty and TB or HIV/AIDS 3
1.1.4 The need for community participation in integrated TB and HIV interventions 4
1.1.5 The Triage Plus study and the previous studies implemented by the REACH Trust 6
1.2 Statistical power in cluster randomised studies 6
1.3 Estimation of prevalence and incidence rate ratios in cluster randomised trials 7
1.4 Aim of the study 9
1.5 Structure of the thesis 10
CHAPTER 2 11
THE TRIAGE PLUS STUDY 11
2.1 Introduction 11
2.2TB and HIV Case detection, treatment and the need for community interventions 11
2.2.1 Tuberculosis case detection and treatment 11
2.2.2 HIV and AIDS diagnosis and the need for timely ART treatment initiation 13
2.2.3 The need for an integrated approach to TB and HIV/AIDS 14
2.2.4 The need for community participation in integrated TB and HIV interventions 14
2.3 The Triage Plus Study design 15
2.3.1 The study areas 15
2.3.2 Study design 16
2.3.3 The phased intervention design 17
2.3.3.1 Number of clusters for randomisation 17
2.3.3.2 Matching of clusters 19
2.3.3.3 The randomisation process 19
2.3.3.4 Blinding of the intervention 20
2.3.3.5 The Triage Plus study implementation phases 20
2.4 Study intervention 23
2.4.1 The Intervention package 23

2.4.2 Inclusion and exclusion criteria for the informal health care providers 25
2.4.3 Conceptual basis of the intervention Framework 25
2.5 The Triage Plus study outcome measures and sample size 29
2.6 Data collection for impact evaluation 32
2.6.1 HIV and TB treatment initiation and testing data 32
2.6.2 Contextual factors (Health system and population characteristics data) 34
2.7 Study limitations 34
2.8 Conclusion 35
CHAPTER 3 36
CLUSTER RANDOMISED TRIALS 36
3.1 Introduction 36
3.2 Cluster randomisation trials and their challenges 37
3.2.1 Cluster randomisation designs 37
3.2.2 Challenges of cluster randomised trials 38
3.2.2.1 Number of clusters available for randomisation 39
3.2.2.2 Reduced statistical power due to within cluster correlation of outcomes 40
3.3 Approaches to address the limited number of clusters 44
3.4 Implications of the limited number of clusters in the Triage Plus study 45
3.5 The repeated cross-sectional design for impact evaluation 47
3.6 Measures of effect: Odds ratios, prevalence ratios and incidence rate ratios 49
3.6.1 Appropriate measures for the effectiveness of the intervention 49
3.6.2 Estimation of prevalence ratios and incidence rate ratios 51
3.6.2.1 Cox regression model 52
3.6.2.2 Poisson regression model 53
3.6.2.3 Log-binomial models 55
3.7 Analysis of clustered data using single level modelling approaches 59
3.8 Conclusion 60
CHAPTER 4 62
GENERALISED LINEAR MIXED MODELS 62
4.1 Introduction 62
4.2 Generalised linear mixed model framework 63
4.2.1 The generalised linear mixed model 63

4.2.2 Generalised linear mixed model representation in a repeated measurements design 66
4.2.3 Poisson and negative binomial models 67
4.2.3.1 The Poisson models 67
4.2.3.2 The Negative Binomial Model 71
4.2.3.3 Marginal effects Poisson model 72
4.3 Likelihood based estimation methods for incidence rate ratios 73
4.3.1 Numerical integration 73
4.3.1.1 Maximum Likelihood Estimation of GLMM using adaptive and spherical quadrature 74
4.3.1.2 Estimation of random effects and regression coefficients 76
4.3.2 Penalised quasi-likelihood 76
4.3.2.1 Estimation of generalised linear mixed models using PQL 76
4.3.2.2 Estimation of variance components: 78
4.4 Likelihood based model selection 80
4.4.1 Likelihood ratio tests 80
4.4.2 Akaike Information Criterion and Bayesian Information Criterion 81
4.5 Conclusion 82
CHAPTER 5: 83
STATISTICAL POWER ISSUES WITH CLUSTER RANDOMISED TRIALS 83
5.1 Introduction 83
5.2 Number of clusters and numbers of participants per cluster: a literature review 83
5.3 Simulation study 85
5.3.1 Rationale and objectives 86
5.3.2 Simulation procedures 86
5.3.3 The Simulation Algorithm 94
5.3.4 Assessment of statistical power estimation for different design conditions 96
5.3.5 Diagnostics and sensitivity analysis of simulation procedures 96
5.3.6 Assessing performance of the different statistical methods 97
5.4 Results 99
5.4.1 Convergence 99
5.4.2 Power estimation 104
5.4.2.1 Power estimates when the incidence of the outcome is low and ICC is 0.00154 104

5.4.2.2 Power estimates when incidence of the outcome is high and ICC is 0.00154 . 106
5.4.2.3 Effect of varying ICC on power estimation ... 108
5.4.3 Assessment of accuracy of parameter estimates ... 112
5.4.3.1 Parameter estimates and precision when the ICC was 0.00154 ... 112
5.4.3.2 Effect of varying ICC on the accuracy of parameter estimates ... 120
5.4.3.3 Coverage rates of Standard errors when ICC was 0.00154 ... 123
5.5 Conclusions ... 125
CHAPTER 6 ... 129
STATISTICAL ANALYSIS OF THE TRIAGE PLUS STUDY ... 129
6.1 Introduction ... 129
6.2 Analytical approach ... 129
6.2.1 Unit of analysis and the need for repeated measurements ... 129
6.2.2 The matched-pair design analysis ... 131
6.3 Statistical methods ... 132
6.3.1 Analysis plan ... 132
6.3.2 Adjustment for baseline outcome data and cluster-level covariates ... 134
6.3.3 Statistical modelling ... 134
6.4 Summary of characteristics of clusters ... 137
6.4.1 Baseline demographic and distribution of health facilities ... 137
6.4.2 Baseline measurement of outcome measures between intervention arms ... 143
6.4.3 Assessment of the effectiveness of the intervention in the first 12 months using unadjusted incidence rate ratios ... 144
6.5 Impact evaluation based on the generalised linear mixed modelling ... 146
6.5.1 Impact on TB and ART treatment initiation rates in the first 12 months of the intervention ... 147
6.5.1.1 Impact on TB treatment initiation rates ... 147
6.5.1.2 Impact on smear- positive TB treatment initiation rates ... 150
6.5.1.3 Impact on ART treatment initiation rates ... 152
6.5.2 Impact on secondary outcome measures in the first 12 months ... 156
6.5.2.1 Impact on TB and HIV testing access rates in the first 12 months of intervention ... 156
6.5.2.2 Impact on TB testing uptake rates in the first 12 months of the intervention . 156
6.5.2.3 Impact on HIV testing uptake rates ... 160

6.6 Impact on TB and ART treatment initiation and testing uptake rates in the second phase (11 months) of the intervention 164
6.6.1 Impact on TB and ART treatment initiation rates in the second phase (11 months) of the intervention 164
6.6.1.1 Impact on TB treatment initiation rates in the second phase (11 months) of the intervention 164
6.6.1.2 Impact on smear- positive TB treatment initiation rates in the second phase (11 months) of the intervention 167
6.6.1.3 Impact on ART treatment initiation rates in second phase (11 months) of the intervention period 171
6.6.2 Impact on TB testing uptake rates in the second phase (11 months) of the intervention 175
6.7 Summary and conclusion 178
6.8 Statistical implications of the findings of the Triage Plus study 179
CHAPTER 7 181
GENERAL DISCUSSION, POLICY IMPLICATIONS AND CONCLUSIONS 181
7.1 Introduction 181
7.2 Statistical power and parameter estimation in a cluster randomised trial with a limited number of clusters but adopting a repeat observation design 181
7.3 Statistical power of the Triage Plus study 183
7.4 How the Triage Plus study was conducted and its limitations 185
7.4.1 Research design 185
7.4.2 Study limitations 186
7.5 Effectiveness of the Triage Plus study 188
7.5.1 Intervention effect on ART treatment initiation rates and HIV testing uptake rates 188
7.5.2 Intervention effect on TB treatment initiation rates and testing uptake rates 188
7.5.3 Flexibility of the intervention 192
7.6 Policy implications 193
7.6.1 Policy implications on TB control 193
7.6.2 Policy implications on HIV transmission prevention 193
7.7 Conclusions 194
7.8 Future work 197
7.8.1 Introduction 197
7.8.2 Hierarchical model for binomial data 197

7.8.3 Bayesian estimation of incidence rate ratios 200
7.8.3.1 Introduction: 200
7.8.3.2 The Bayesian paradigm 200
7.8.3.3 Computation of posterior distributions 201
7.8.3.4 Diagnosing the convergence of posterior distributions 203
7.8.3.5 Prior distributions 205
7.8.3.6 Bayesian model selection 208
7.8.4 Future simulation work 216
8.0 Appendix 217
Appendix 8.1: A statement of declaration of the contributions made by me and other people in the whole study 217
Appendix 8.2: Data recording/collection forms 218
Appendix 8.2.1: TB treatment registration form-sample 218
Appendix 8.2.2: Anti-retroviral treatment registration form 219
Appendix 8.2.3: Presumptive TB testing registration form 220
Appendix 8.2.4: HIV testing registration form 221
Appendix 8.3 Cluster level population sizes 222
Appendix 8.4: Sample of dataset layout for TB treatment initiations (similar layout used for ART initiations, HIV testing and TB testing) 223
Appendix 8.5: Stata simulation programme 224
Appendix 8.6: Stata data processing and analysis programme 229
Appendix 8.7: Abstract presented at the 2013 Union Conference 234
9.0 References 236

LIST OF TABLES

Table 1: Poverty incidence in Lilongwe rural areas 21
Table 2: Intervention effectiveness outcome measures used in Triage Plus trial 29
Table 3: The effect of number of groups, effect size and number of time points on the convergence rate (%) based on 1000 simulated data sets 101
Table 4: The effect of the ICC, number of groups, effect size and repeated measurements on convergence rate, based on 1000 simulated data sets when the disease incidence is low: Mean outcome measure in control clusters of 15 cases per month..... 102
Table 5: The effect of the ICC, number of groups, effect size and repeated measurements on convergence rate, based on 1000 simulated data sets when the disease incidence is high: Mean outcome measure in control clusters of 70 cases per month..... 103
Table 6:The effect of the number of clusters, repeated measurement times and effect size on the statistical power of the study design (% p value<0.05) based on 1000 simulated data sets under low disease incidence – The mean outcome measure in control clusters is 15 cases per month. 105
Table 7:The effect of the number of clusters, repeated measurement times and effect size on the statistical power of the study design based on 1000 simulated data sets under high disease incidence. The mean outcome measure in control clusters is 70 cases per month..... 107
Table 8: The effect of the ICC, number of groups, effect size and repeated measurements on the statistical power, based on 1000 simulated data sets when the disease incidence is low: Mean outcome measure in control clusters of 15 cases per month 110
Table 9: The effect of the ICC, number of groups, effect size and repeated measurements on the statistical power, based on 1000 simulated data sets when the disease incidence is high: Mean outcome measure in control clusters of 70 cases per month 111
Table 10: The effect of the number of groups, time points and the effect size on how accurately the estimated effect size is close to the fixed estimate size and how the precision of the estimated effect size improves based on the 95% confidence interval after 1000 simulated data sets (using ICC=0.00154)..... 114
Table 11: The effect of the number of groups, time points and the effect size on how accurately the effect sizes were estimated based on 5% and 95% percentiles of the estimated effect size after 1000 simulated data sets..... 116
Table 12: The effect of the number of groups and repeated measurements on the accuracy of the estimated effect size based on the mean length of 95% confidence interval, based on 1000 simulated data sets 118
Table 13:The effect of the number of clusters, time points and the effect size on the percentage relative bias of how the estimated effect size differed from the fixed effect size and non-coverage of the asymptotic 95% confidence interval of the estimated effect size. 119

Table 14: The effect of the ICC, number of groups, effect size and repeated measurements on the accuracy of the estimated effect size based the mean length of 95% confidence interval, based on 1000 simulated data sets 121
Table 15: The effect of number of clusters, time points and the effect size on the non-coverage of the 95% confidence intervals of parameter estimates (regression coefficients) based on 1000 simulated data sets. 124
Table 16: Comparison of the baseline demographic characteristics between Early intervention and Delayed intervention arms 138
Table 17: Comparison of baseline crude mean counts and incidence rate ratios for TB and ART treatment initiation rates between intervention arms 143
Table 18: Comparison of baseline cluster-pair TB and ART treatment initiation rates between intervention arms 144
Table 19: Comparison of crude mean counts and incidence rate ratios in TB and ART treatment initiation rates between interventions arms in the first 12 months of intervention 145
Table 20: Comparison of cluster-pair incidence rate ratios for TB and ART treatment initiation between intervention arms in the first 12 months of the intervention 145
Table 21: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring TB treatment initiation rates in the first 12 months of the intervention 151
Table 22: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring access rates to TB treatment services among smear positive patients in the first 12 months of the intervention 153
Table 23: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring access rates to ART services in the first 12 months 155
Table 24: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring access rates to testing services for presumptive TB cases in the first 12 months 158
Table 25: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for smear -positive TB cases identified at testing sites during the first 12 months..... 159
Table 26: Maximum likelihood estimates for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring access rates to HIV testing in the first 12 months 161
Table 27: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring overall access rates to TB treatment initiation rates using the data from the second phase of the intervention (11 months) 166
Table 28: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring overall access rates in smear positive TB treatment initiation rates using the second phase (11 months) of the intervention 170

Table 29: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring overall ART initiation rates using the second phase (11 months) of the intervention..... 173
Table 30: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and coefficients models for measuring presumptive TB testing uptake rates in the second phase (11 months) of the intervention... 176

LIST OF FIGURES

Figure 1: Flow diagram for the intervention design..... 18

Figure 2: Map of Lilongwe showing the distribution of the Triage Plus study clusters between intervention arms 22

Figure 3: Triage Plus study design 23

Figure 4: The Triage Plus study implementation framework 28

Figure 5: Distribution of p-values after running 1000 simulations when the effect of the intervention was set to a Null case 98

Figure 6 Power curves when the disease incidence is low..... 109

Figure 7 Power curves when the disease incidence is high..... 109

Figure 8: Flow diagram..... 140

Figure 9: Map showing the distribution of TB and ART treatment initiation sites in the first 12 months..... 142

Figure 10: Map showing the distribution of new ART initiation sites..... 143

Figure 11: Monthly and cumulative HIV testing uptake rates per 1,000 adults..... 164

Figure 12: Monthly and cumulative TB treatment initiations rates per 10,000 adults..... 169

Figure 13: Monthly and cumulative ART treatment initiations rates per 10,000 adults... 175

Figure 14: Monthly and cumulative TB testing uptake rates per 10,000 adults..... 178

Figure 15: History plots for the converged chains..... 206

Figure 16: Density plot for the posterior distribution for the coefficient of the intervention using TB treatment initiations data from the Triage plus study..... 207

CHAPTER 1

GENERAL INTRODUCTION AND CONTEXT

1.1 Overview of the Triage Plus study and its objectives

1.1.1 Introduction

Tuberculosis (TB) and human immuno-deficiency virus (HIV) remain the major determinants of morbidity and mortality in Sub-Saharan Africa. The region includes most of the countries with the highest burden of the tuberculosis epidemic (WHO, 2009). The TB epidemic in Sub-Saharan Africa has worsened because of its interaction with HIV and AIDS (Corbett et al. 2002; Corbett et al., 2003; WHO, 2009). Of the 9.27 million TB cases in 2007 globally, 15% (1.37 million) cases were HIV-related tuberculosis, of which 79% were from the African region (WHO, 2009). Because of this interaction, these two diseases form a vicious cycle: TB is the main cause of morbidity and mortality in people with HIV; a quarter of HIV-related deaths are due to TB, and HIV enhances the progression of latent tuberculosis infection to active disease thereby driving the TB epidemic (WHO, 2009; Naidoo and Taylor, 2007;Mukadi, Maher and Harries, 2001).

Although interventions to address the two disease conditions exist, patient and health system barriers affect service utilisation. Barriers such as cost, distance, delays and repeated visits to multiple health facilities increase patient drop-out rates in the course of care seeking especially in Sub Saharan Africa (Lawn et al., 1998; Squire et al., 2005; Xu et al., 2007; Storla et al. 2008). This is particularly so among the poor who face challenges in getting a definitive diagnosis, starting on treatment and adhering to treatment regimens. To address these access challenges, the Triage plus study - a cluster randomised intervention study involving informal health care providers - was implemented in rural Malawi. Primarily, the intervention aimed at evaluating the effectiveness of engaging informal health care providers in TB and HIV testing and treatment services.

1.1.2 Access to TB and HIV services

Globally, the tuberculosis (TB) burden remains high, but its incidence rates are slowly declining (Lönnroth et al., 2010). TB treatment success rates are high internationally; at least 85% for new smear positive TB treated in National TB Programmes was reached in the 2007

cohort. TB treatment success rates surpassed the 85% target in the WHO eastern Mediterranean region (88%), the south eastern Asian region (88%), and the western Pacific region (92%) (WHO, 2009). However, treatment success rates in the African Region remain below the 85% target, and case detection rates of smear-positive TB remain below the 2005 target of 70% in many countries (WHO, 2009).

Achieving high treatment success rates and timely case detection for smear positive TB cases have long been considered key factors for controlling TB because they cut transmission rates and avert TB related deaths. However, achieving high TB case detection rates is the greatest challenge to controlling TB in many countries. In Malawi, despite the increase in TB case notifications reported at the National TB Programme from 19,496 in 1994 to 28,234 in 2003, WHO estimated that the country was only notifying 41% of the TB cases (Nyirenda, 2006; WHO, 2007a). Thus, improving TB case detection and timely initiation of TB treatment, especially for smear-positive TB cases, is the focus of most TB intervention initiatives because it is the most effective way to reduce TB transmission in the general community.

Although swift and accurate TB case detection is essential, the most commonly used approach for case detection is based on screening patients with a chronic cough at health facilities, a method which is less efficient in detecting a large proportion of smear positive TB cases (Ayles et al., 2009; Hoa et al., 2010), than the active case finding strategies that screen all at risk populations (WHO, 2011). The difficulty is to ensure sustainability of such active case finding strategies. Debate continues on the effectiveness of the active case finding strategies considering the cost and logistics involved in implementing such interventions in low-income countries (Golub et al., 2005; Uplekar and Raviglione, 2007).

In regards to access to Human Immuno-deficiency Virus (HIV) services, uptake rates for HIV testing and the number of patients put on Anti-Retroviral Therapy (ART) treatment remains low in Sub-Saharan Africa and in developing countries. Of the 7.1 million in need of ART treatment in 2006, only 28% were on ART treatment in Sub-Saharan Africa, 19% in Asia and only 14% in low and middle income countries of Eastern Europe and Central Asia (Giuliano and Vella, 2007; WHO/ UNAIDS/UNICEF, 2007).

The recently completed 'Demographic and Health Surveys in the Sub Saharan Africa' revealed that the percentages of men and women who know their sero-status ranged from 14% to 51% in men and 15% to 72% in women (see National Population Commission (Nigeria) and ICF Macro, 2009; National Statistical Office [Malawi] and ICF Macro, 2011; Central Statistical Agency [Ethiopia] and ICF International, 2012).This suggests that many of those who are infected do not know their HIV status, and, therefore, are at risk of spreading the infection or being re-infected with ART resistant strains. If such populations are not reached with preventive interventions such as HIV testing services they may end up spreading the infection or being re-infected with resistant strains. Thus, to combat the HIV pandemic, there has been an increased emphasis on prevention of new HIV infections through early HIV detection by promoting voluntary HIV testing.

Given that early ART treatment has the potential of reducing the risk of HIV transmission to an uninfected partner (Cohen and Gay, 2010; Anglemyer et al., 2011; Cohen et at, 2011), reaching all eligible patients with ART services, coupled with improved HIV testing uptake, is likely to reduce HIV transmission. With the HIV burden in Sub-Saharan Africa worsening the tuberculosis epidemic: HIV enhances the progression of latent tuberculosis infection to active disease and therefore drives the TB epidemic(Corbett et al. 2002; Corbett et al., 2003; WHO, 2009), interventions targeting both TB and HIV are essential to reduce TB and HIV transmission.

1.1.3 The relationship between poverty and TB or HIV/AIDS

Recognition of the relationship between poverty and timely access to diagnostic and treatment services from health facilities is well documented (Squire et al., 2005; Storla et al., 2008). In particular, the poor are most likely to contract TB, progress to disease, and have poor treatment outcomes but have the most difficulty accessing formal public health services. This creates the potential problem of continued TB transmission, which, in turn, maintains the risk of TB transmission among the general population (Squire et al., 2005;Lönnroth et al., 2009).

The difficulty of timely intervention amongst the poor is a result of the skewed distribution of health facilities: most are located in urban areas; so many rural poor are unable to easily gain access to health services. Additionally, the poor may not be able to afford the costs related to medical care such as transport, medication and other medical-related costs (Kemp et al., 2007; Barter et al., 2012). These costs result in those in the lowest socio-economic groups

delaying seeking health care from formal sources. These individuals, in turn, resort to informal health care sources, such as traditional healers and store-keepers. Detrimental behaviour patterns, such as delaying health care or seeking alternatives, are compounded by health system barriers, including the incorrect diagnosis of TB and HIV and, errors in referring TB and HIV symptomatic patients to the appropriate departments within the facility, both of which results in extending the time before a correct diagnosis is made (Storla et al., 2008). These factors contribute to long care-seeking pathways and high direct and indirect costs; further depleting a household's already strained financial status (Kemp et al., 2007).

TB and HIV/AIDS are chronic conditions leading to prolonged care and support and which create a vicious cycle of poverty within affected households. Poor households with TB or HIV/AIDS affected members may be caught in a damaging cycle of poverty and ill-health. By strengthening health systems (especially at the community level) and improving the awareness of TB, and general health among communities, early TB diagnosis and treatment of the most infectious forms, can occur (WHO, 2011).

1.1.4 The need for community participation in integrated TB and HIV interventions

Given the structural barriers (e.g. transportation costs) faced by the poor and the vulnerable in accessing TB and HIV/AIDS services, and the need to prevent TB and HIV in order to reduce transmission, community-based strategies involving informal healthcare providers such as store-keepers are essential for ensuring access to health services (WHO, 2011; Simwaka et al., 2012). WHO (1989, page 6) described informal healthcare providers as '*members of the communities where they work, selected by the communities, answerable to the communities for their activities, supported by the health system but not necessarily a part of its organization, and have shorter training than professional workers*'. Informal healthcare providers have been used in community interventions to improve access to services for various diseases and conditions including treatment of uncomplicated malaria in Kenya and nutrition education intervention that has resulted in improved case detection for diarrhoea, respiratory infections and malnutrition in Nepal (Curtale et al., 1995;Datiko and Lindtjørn, 2009; Marsh et al., 1999 & 2004; Simwaka et al., 2012).

In Ethiopia, extension health workers have been used in TB identification: collecting sputum and storing and transporting it to the nearest microscopic centres for TB diagnosis (Datiko

and Lindtjorn, 2009). This resulted in improved case detection: 122.2% in intervention areas compared to 69.4% in control areas, with women showing the highest rise in case detection (Datiko and Lindtjorn, 2009). In China, village doctors increased TB case detection from 36.2 to 49.9/100,000 following a similar community intervention (Xiong et al., 2007). In Malawi, storekeepers engaged in the referral of presumptive TB cases to the nearest health facility, significantly increased the smear positive TB case notification from 0.6 per 1000 people in the control areas to 1.2 per 1000 in the intervention areas (Simwaka et al., 2012).

Informal health care providers play a critical role in reaching out to individuals who might otherwise not access TB and HIV services due to factors such as inadequate knowledge of TB and HIV (Storla et al., 2008). The use of community structures, such as informal health care providers, in promoting health information at the community level does not only promote the dissemination of culturally appropriate health information, it also reduces the stigma associated with TB and HIV (Norris et al., 2006; Corkery et al., 2007; Farid et al., 2007; Mock et al., 2007; O'Donnell et al., 2012).

By equipping these informal health care providers with the necessary skills to conduct community interventions and improve access to TB and HIV services, diagnostic and treatment delays that would lead to unnecessary deaths or transmission, can be reduced (Storla et al., 2008). Working with informal healthcare providers, in improving disease recognition, health communication, treatment, and referral at the first point of access in the community, is in line with WHO recommendations (WHO, 2011).

However, engaging informal health care providers in vertical programs may lead to a duplication of efforts as the same informal healthcare providers are usually engaged to provide information on a range of diseases. In contrast, the engagement of informal health care providers to assist with multiple conditions such as TB and HIV in an integrated manner is likely to be more efficient. Studies of the clinic based integration of TB and HIV in service provision for co-infected individuals have shown promising results in improving TB and HIV health outcomes (Hermans et al., 2012).

Although evidence exists concerning the effectiveness of clinic based integration of TB and HIV services and paid community care providers (Hermans et al., 2012; Uwimana et al., 2012),

there is limited evidence that the involvement of non-paid informal health care providers in the provision of integrated TB and HIV services at the community level might lead to an improvement in access to TB and HIV services. A randomised, controlled health system intervention trial involving informal health care providers called 'Triage plus' was therefore set up to investigate their impact on access to TB and HIV services in rural Lilongwe, Malawi. Primarily, the intervention aimed to investigate whether the engagement of informal health care providers in integrated TB and HIV community interventions would be effective in improving services access rates for the two diseases, thereby guiding evidence based policy formulation on using informal providers in TB and HIV interventions at the community level. Specifically, the Triage Plus study aimed to determine the overall effect on a) TB and Anti-Retroviral Therapy (ART) treatment initiation and b) testing uptake of TB and HIV services.

1.1.5 The Triage Plus study and the previous studies implemented by the REACH Trust

This study continues the work of previous studies conducted by the Research for Equity and Community Health (REACH) Trust which focused on health issues for the poor and the vulnerable in Malawi (e.g. Simwaka et al., 2012). Previous research by the REACH Trust was conducted primarily in semi-urban areas where the effectiveness and acceptability of engaging community based structures in single disease conditions was assessed (e.g. Simwaka et al., 2012).

Due to the interaction between TB and HIV (Corbett et al. 2002; Corbett et al., 2003; WHO, 2009), this current study assessed the effectiveness of community engagement, utilising a variety of informal health care providers, to improve access to both TB and HIV services among the poor. A detailed description of the project is provided in Chapter 2.

1.2 Statistical power in cluster randomised studies

Although previous research by the REACH Trust focused on semi-urban areas, this study extends these efforts into the rural community where different challenges exist. The level of infection of *Mycobacterium tuberculosis* (MTB)is generally lower in rural than in urban settings, where new cases of TB are less common (see Ayles et al., 2009). To obtain an adequate number of TB cases, large cluster sizes are required for randomisation to different study arms, thereby creating the problem of fewer clusters available for randomisation. Because of the need to obtain adequate sample sizes for the primary outcome measures, especially TB treatment initiation rates, the Triage Plus study used 3 clusters per arm, which was achieved by subdividing the Lilongwe rural area into six large clusters.

The use of a small number of clusters for this study had its own statistical implications. First, there would be statistical challenges in achieving statistical power and obtaining robust estimates when assessing effectiveness of the intervention. These challenges stem from the need for conducting statistical analysis using clusters as units of analysis, which reduces the number of degrees of freedom for adequate statistical power (Murray, Hannan and Baker, 1996; Donner and Klar, 2000). Second, the randomisation process of clusters to different study arms may not be efficient in ensuring the even distribution of confounding factors between study arms (Murray, 1998a). To achieve adequate power and robust estimates in cluster randomised studies, previous simulation studies have recommended a minimum of 30 clusters per arm (Maas and Hox, 2005; Moineddin et al., 2007).

With only 3 clusters per arm in the Triage Plus intervention, it is likely that the statistical power to detect significant intervention effects is reduced because there is a limited number of degrees of freedom for the between-cluster residual errors (Cornifield 1978; Koespell et al., 1991; Feldman, McKinlay and Niknian, 1996; Murray, 1998a, Donner and Klar, 2000).

To improve statistical power in studies with a limited number of clusters, the use of external intracluster correlation coefficients and the subdivision of the clusters into small sub-clusters have been proposed (Feldman, McKinlay and Niknian, 1996; Murray, Hannan and Baker 1996). Hayes and Moulton (2009) caution against using external intracluster correlations of outcome measures from a different setting. Although use of sub-clusters improves power and type I error rates, Feng et al (2001) recommended the use of the original analysis clusters for valid estimates. To achieve robust estimates and adequate statistical power in the Triage plus study, alternative statistical approaches were needed.

1.3 Estimation of prevalence and incidence rate ratios in cluster randomised trials

As a measure of intervention effectiveness in cross-sectional studies, prevalence ratios or incidence rate ratios have been preferred to odds ratios. In cross-sectional studies with a common outcome, odds ratios tend to overestimate the impact of an intervention (Lee and Chia, 1993; Skov et al. 1998; McNutt et al., 2003). Estimation of prevalence ratios in cross-sectional studies or incidence rate ratios in longitudinal studies arise from different statistical models. Most notably, approaches based on Cox regression (Breslow, 1974; Lee and Chia, 1993), Poisson regression (McNutt et al., 2003; Petersen and Deddens, 2008), log-binomial regression (Zocchetti, 1995; Skov et al., 1998), and the COPY method when convergence of

log-binomial models fail (Deddens et al., 2003). These models are based on the generalised linear model (GLM) theory- a class of fixed effects models (McCullagh and Nelder, 1989). The GLM framework involves specification of a link function $g(.)$ that relates the expected response to the linear predictor $\eta_i = x_i\beta$ (where i represents the individual observations, x_i is a vector of explanatory variables with fixed effects β), and variance structure for the response variable that is assumed to follow a distribution from the exponential family (McCullagh and Nelder, 1989). A brief description of the GLM is provided in section 3.7 in Chapter 3.

The statistical approaches based on the generalised linear model theory are typically applied to data derived from clinical trials or cohort studies whose observations are independent. However, in cluster randomised trials, where clusters of individuals or communities (such as villages, schools) are assigned randomly to different intervention arms, the study observations or responses are usually correlated, so standard models can no longer be used to analyse such data (Breslow and Clayton, 1993, Murray et al., 1998a, Hayes and Moulton, 2009). Therefore, the analysis of data from cluster randomised trials needs to take into account the non-independence of the observations within the clusters, usually referred to as intracluster correlation (ICC).To account for the inherent dependency or correlation of data derived from community randomised interventions, a generalised linear mixed model (GLMM)framework, an extension of the generalised linear model theory, is usually used (Breslow and Clayton, 1993).

The GLMM framework involves incorporation of random effects in the linear predictor of the fixed effects model. This framework has several advantages, including easy incorporation of both cluster-level and individual-level covariates, possibilities of exploring variance structure, extension to various outcome types, and flexibility in addressing heterogeneity and over-dispersion (Breslow and Clayton, 1993; Omar et al., 2000). However, because of the intractability of the marginal likelihood of the response variable that is obtained by integrating out the random effects in GLMM framework, approximation methods are used (Breslow and Clayton, 1993). Approaches based on numerical integration (Rabe-Hesketh, Skrondal and Pickles, 2005) or penalised quasi-likelihood estimation (Breslow and Clayton, 1993) have been proposed to estimate the marginal likelihood. More recently, Bayesian estimation using

Markov chain Monte Carlo methods have been considered (Zhao et al., 2006). A detailed description of the GLMM framework and modelling approaches are presented in Chapter 4.

1.4 Aim of the study

There have been many studies like the Triage Plus intervention employing a limited number of clusters, a factor which is likely to have had profound consequences on both the statistical power and the statistical precision of the parameter estimates. Most empirical research has investigated power and parameter estimation issues in cluster randomised designs with at least 30 clusters per arm for valid estimates (e.g. Moineddin et al., 2007). The main aim of the present study was to investigate if it is possible to improve the analytical approach to the evaluation of cluster-randomised trials where the complexity of the intervention demands a small number of clusters and to investigate the effectiveness of engaging informal healthcare providers in integrated TB and HIV community intervention in treatment initiation rates and testing access rates.

The specific objectives of the study are:

1. To review cluster randomised trials and the statistical methods for assessing effectiveness of the intervention in cluster randomised trials.

2. To assess the statistical efficiencies of different design conditions in terms of statistical power and accuracy in parameter estimates when determining the effectiveness of complex interventions with a limited number of clusters, and to identify the circumstances under which each of the statistical methods would be more robust in detecting significant intervention effects or providing accurate estimates of intervention effects.

3. To apply these statistical approaches to the actual data from the current community based intervention on TB and HIV called 'Triage Plus'.

4. To assess the effect of involving non-paid informal healthcare providers in integrated TB and HIV community interventions in improving testing and treatment initiation rates.

1.5 Structure of the thesis

This dissertation is organised into seven chapters including this introductory Chapter. Chapter 2 describes the Triage Plus intervention study involving informal healthcare providers. Chapter 3 presents an overview of cluster randomised trials while highlighting the challenges and implications of a limited number of clusters. It also presents measures of effect, and statistical approaches applied in estimating intervention effectiveness. Chapter 4 presents statistical models in the context of a generalised linear mixed model framework in the estimation of prevalence ratios or incidence rate ratios under a correlated data structure. It also provides modelling approaches for prevalence ratios or incidence rate ratios in the context of generalised linear mixed models using the likelihood and Bayesian approaches for intervention effect estimation and model selection. Chapter 5 presents an evaluation of the statistical models using a series of simulation studies under different design conditions in order to guide the choice of statistical approaches to be pursued, while Chapter 6 presents the statistical analysis of the Triage Plus study. It also presents the results of the various statistical models, taking into account the phased design of the Triage Plus cluster randomised intervention study. Lastly, Chapter 7 offers a general discussion and draws conclusions. This final chapter ends by assessing the policy implications of this study's findings on current policy and making recommendations for further research.

CHAPTER 2

THE TRIAGE PLUS STUDY

2.1 Introduction

This Chapter provides an overview of the study design for the Triage Plus study, a TB and HIV community-based intervention implemented in rural Malawi. The intervention was participatory, involving informal healthcare providers delivering elements of the intervention in rural communities with technical support from health surveillance assistants (front line public health workers). This Chapter first extends the discussion of TB and HIV services access issues presented in Chapter 1 and the need for timely treatment initiation for the two diseases. The Chapter then explains the need for an integrated approach to TB and HIV in community interventions. A description of the study design and intervention package as well as the outcome measures and data collection methods used during the implementation of the Triage Plus study is given. Study design limitations, in terms of the number of clusters used for randomisation and the possible statistical implications, are provided. This Chapter concludes with the suggestion that there is a need for further evaluation of the statistical implications of limited cluster numbers as in the Triage Plus study design.

2.2TB and HIV Case detection, treatment and the need for community interventions

2.2.1 Tuberculosis case detection and treatment

Despite global adoption of the STOP-TB strategy to improve case detection rates and reduce TB incidence, tuberculosis case detection rates are far from reaching the global target of at least 70%. Moreover, the halving of TB prevalence and death rates by 2015 are less likely to be met in the Africa region (WHO, 2009).

Tuberculosis management and access to timely TB services is complicated by both system and patient barriers. Barriers characterised by delays and repeated visits to multiple care providers increases patient drop-out rates in the course of care seeking especially in Sub Saharan Africa (Lawn et al., 1998; Squire et al., 2005; Xu et al., 2007; Storla et al. 2008). Timely case detection and treatment initiation is further compounded by health system barriers, such as incorrect management of patients with negative smears or with no haemoptysis, and health workers failing to identify all TB patients in outpatient clinics, thereby contributing to long delays in TB case detection and treatment(Rajeswari et al., 2002, Chiang et al., 2005).

Improved TB case detection and timely diagnosis and treatment are essential for effective tuberculosis control in order to reduce the period of infectivity in the community (Golub et al., 2005). Delayed or incorrect diagnosis and treatment of tuberculosis results in severe disease and more complications that lead to increased mortality (Glynn et al., 1998). However, despite the popularity of active case finding which has remained an integral part of tuberculosis control in high risk settings, its effectiveness still remain uncertain (Golub et al., 2005). It is discouraged because of the associated high costs and weaknesses in treatment programmes (Golub et al., 2005; Uplekar and Raviglione, 2007). Given that recent TB prevalence surveys have shown that a large proportion of smear positive TB cases do not report any TB symptoms early in the disease course (Ayles et al., 2009;Hoa et al., 2010), active case finding strategies need to be reconsidered (especially in high prevalent settings for TB and HIV).

Active case finding studies recently conducted in Harare, Zimbabwe, demonstrated that the use of community based case finding approaches in high TB and HIV prevalence settings could improve TB case detection rates (Corbett et al., 2010). In Uganda, using a model based analysis of the effectiveness of active case finding strategies in tuberculosis control found an additional 1,594 TB cases over one year, and 675 deaths were averted over a 5 year period (Mupere et al., 2013). Nevertheless, the costs and logistics of implementing such strategies in resource-limited settings are unlikely to be sustainable. These approaches are more suited for select risk groups such as prisoners. New, innovative approaches for improving TB case detection in the general population are thus required, with active case finding strategies being adopted as complementary approaches.

Given that community management of undiagnosed tuberculosis remains a challenge (Wood et al., 2007;Sekandi et al., 2009), and that most new tuberculosis infections occur outside households in high incidence settings (Verver et al., 2004), adopting interventions that involve community participation is necessary for reducing tuberculosis transmission (O'Donnell et al., 2012). Such interventions may range from those that improve a community's knowledge and awareness of TB, informing them what to do when they have symptomatic TB, to interventions which ensure improved access to TB case detection and effective treatment by minimising barriers to health care access (WHO, 2011).

2.2.2 HIV and AIDS diagnosis and the need for timely ART treatment initiation

Although HIV infection is incurable, recent advances in medical technology, such as the introduction of Highly Active Anti-Retroviral Therapy (HAART) in 1996 have generally improved the quality of life (Cohen et al., 1998; Brechtl et al., 2001) and survival rates of HIV infected individuals (Hogg et al., 1998; Palella et al. 1998). In particular, standardised clinical practice and improved availability of HIV medication have contributed to a decline in HIV related mortality rates in most affected countries (Malawi National AIDS Commission, 2007 & 2010, Ray et al., 2010).

Because of the reduced viral load in genital secretions following ART treatment (Vernazza et al., 2000; Graham et al., 2007), ART reduces HIV transmission rates (Cohen et al., 2011). HIV transmission probabilities are increased when ART initiation is delayed(Cohen et al., 2011); therefore, the early diagnosis of HIV infection followed by the timely initiation of antiretroviral treatment would most likely reduce transmission rates, the progression of HIV to AIDS and early mortality (Castilla et al., 2002, Cohen et al., 2011). Castilla and colleagues found that a late diagnosis of HIV was associated with the early onset of AIDS because those with a late diagnosis cannot benefit from antiretroviral therapy.

Effective HAART for HIV-infected persons has been inaccessible in most developing countries. Recently, free antiretroviral treatment has been scaled up following the "3 by 5" initiatives to ensure universal access to HIV services (Kim and Ammann, 2004).The increased availability of effective antiretroviral treatment has led to new insights into HIV prevention strategies, such as voluntary HIV counselling and testing. Instead of promoting only HIV infection prevention (through abstinence, being faithful among partners or the consistent use of condoms), focus is now given to the early detection of HIV by promoting HIV counselling and testing (Campbell et al., 1997; Creek et al., 2007; Preidis et al., 2013).

With the availability of effective HAART, early HIV detection is beneficial as it not only ensures the timely initiation of ART treatment but also allows HIV infected persons to access treatment, care and support. Hence, most countries affected by the epidemic have changed their national HIV policies from risk-based HIV testing to routine testing (Creek et al., 2007; Preidis et al., 2013). Thus, the comprehensive approach to HIV testing has not only been

promoted to prevent the spread of HIV but also as an essential component for one's entry into the antiretroviral treatment programme. Furthermore, as described below, the STOP-TB strategy has been unable to sufficiently contain TB incidence in the face of rising HIV prevalence. Improving early HIV detection and timely ART initiation by involving community structures such as informal healthcare providers will, in turn, help contain TB incidence (De Cock and Chaisson, 1999).

2.2.3 The need for an integrated approach to TB and HIV/AIDS

In Sub-Saharan Africa the TB epidemic has worsened because of its interaction with HIV and AIDS (Corbett et al. 2002; Corbett et al., 2003; WHO, 2009). This interaction of TB and HIV has created a need to confront the dual pandemic with an integrated approach using models that range from offering TB and HIV services at the same facility to a complete TB/HIV package that is delivered by one health care team (Harries et al., 2002, WHO, 2004; Legido-Quigley, 2013). According to the interim policy on collaborative TB and HIV activities (WHO, 2004), TB and HIV services can be integrated in two ways. First, tuberculosis services may be integrated into HIV health-care settings in order to reduce the TB burden among HIV. Second, HIV services may be integrated into tuberculosis control programmes with a view to reducing the HIV burden in patients with TB.

The integration of TB and HIV services in TB or HIV clinics is meant to improve diagnosis and treatment outcomes among co-infected patients as well as overcome challenges such as losing patients when they are referred between TB and HIV services, increased patient travel costs due to multiple visits and increased time spent in clinics (Kerschberger et al., 2012; Legido-Quigley, 2013; Hermans et al., 2012). This integration not only offers benefits for patients, health providers and the health system, but also benefits from shared resources, as the complete integration of TB and HIV services uses the same health system resources, including the drug procurement supply chain and human resources (Maher, 2010).

2.2.4 The need for community participation in integrated TB and HIV interventions

Studies conducted to assess the effectiveness of the clinic based integration of TB and HIV in service provision for co-infected individuals have shown integration to be effective in improving TB and HIV health outcomes and reducing the delay in treatment initiation (Hermans et al., 2012; Kerschberger et al., 2012). However, although the integration of TB and HIV services in resource-constrained settings is feasible, some implementation challenges still

exist such as inadequate laboratory infrastructure and skilled staff with required expertise and experience to manage TB and HIV co-infected patients(Howard and El-Sadr, 2010).

Furthermore, given the structural barriers faced by the poor and the vulnerable in accessing TB and HIV services in health facilities, the use of community based strategies involving informal healthcare providers would ensure equitable access to the health services. As described in Chapter 1 section 1.1.4, the use of informal healthcare providers in community interventions is well documented (Datiko and Lindtjørn, 2009; Marsh et al., 1999; Simwaka et al., 2012). However, there is little research involving informal healthcare providers in the provision of integrated TB and HIV services at the community level. To address this research gap, a cluster randomised intervention study involving informal health care providers called 'Triage Plus' was implemented in rural Malawi.

2.3 The Triage Plus Study design

The overall design of the study is illustrated in a flow diagram shown in Figure 1 (Campbell, Elbourne, and Altman, 2004). A detailed description of the study design follows.

2.3.1 The study areas

As indicated in Chapter 1 section 1.1.5, earlier research by the REACH Trust focused on health issues for the poor and the vulnerable in semi-urban areas (see Simwaka et al., 2012). This study continued these efforts in rural areas where different challenges exist. Thus, the Triage Plus study was implemented in rural areas of Lilongwe District.

Lilongwe is the capital city of Malawi. It is in the centre of Lilongwe District, an administrative demarcation. The city is therefore surrounded by a large rural district. Lilongwe rural is served by 18 traditional authorities with a total population of 1,230,834 (NSO, 2010).There are 4 mission hospitals, 2 community hospitals, 32 health centres, and 68 health posts in Lilongwe rural. In addition, the district is served by a referral hospital and private clinics located in the urban areas. All the mission and community hospitals located in the rural areas offer TB and HIV services (e.g. TB and ART treatment initiations). All health centres offer HIV testing while TB testing is done in 16 health facilities in the rural areas.

In general, rural areas of Lilongwe are dominated by the Chewa tribe that accounts for 99% of the total rural population (NSO, 2010) and the areas are generally poor. According to poverty mapping analysis carried out by Benson and Colleagues of the International Food

Policy Research Institute (IFPRI), Washington, DC, USA and the National Statistical Office (NSO), Malawi, respectively in 1998, 64.3% of the Malawian population live in poor households of which 76.6% are situated in rural Lilongwe (Benson et al., 2002).

2.3.2 Study design

As the intervention was meant to alter general health seeking behaviour in the intervention areas, an individually randomised clinical trial was considered inappropriate, and a cluster randomisation approach was, therefore, adopted in this study. In cluster randomised trials, distinct and identifiable groups, such as communities or schools, are randomly assigned to study conditions, and data for impact evaluation is collected from the members of those groups (Murray et al., 1998a; Hayes and Moulton, 2009). However, because the effects of the intervention were likely to persist for some time, stepped wedged (De Allegri et al., 2008) and crossover designs (Sjögren et al., 2005) were considered to be inappropriate. A parallel cluster randomised trial design was therefore used. Such parallel designs have previously been used in cluster randomised trials in the evaluation of different health intervention strategies (see Corbett et al., 2010).

Unlike parallel designs, crossover designs allow each cluster to receive both the intervention and control treatment at equal time intervals **(**Palmer et al., 1985; Chavasse et al., 1999; Sjogren et al., 2005). The within-cluster evaluation design allows a more precise intervention evaluation and requires fewer clusters than parallel designs (Hussey and Hughes, 2007).

In the stepped wedge design, a type of crossover design, different clusters switch interventions at different times, and the crossover is in one direction from control to intervention (Hussey and Hughes, 2007; Brown and Lilford, 2006, The Gambia Hepatitis Study Group, 1987; De Allegri et al., 2008). This design approach not only removes the ethical concern of withholding the intervention from some communities (Hussey and Hughes, 2007), but it is also more efficient as clusters act as their own control.

In the Triage Plus study, the areas that acted as control areas in the first 12 months of the intervention later received the intervention while the Early intervention areas continued receiving the intervention (see section 2.3.3). Although, this was similar to the stepped

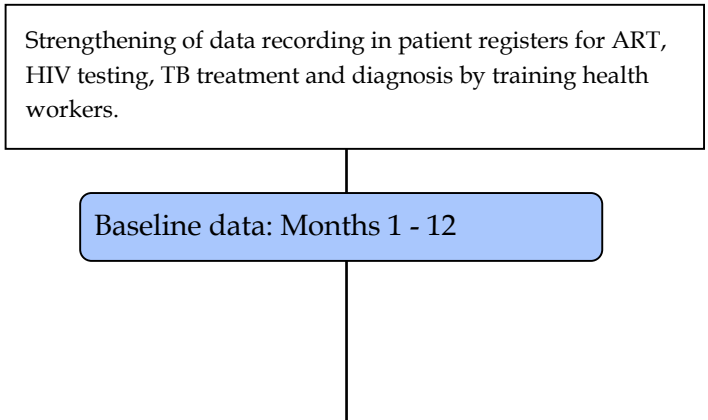
wedged design, a formal stepped wedged design randomisation was not implemented because of the limited number of clusters.

Cluster randomization has practical advantages: all subjects in the cluster are treated in the same way thereby reducing costs and ethical concerns (Donner et al., 1990). However, cluster randomisation may not permit the direct determination of the extent to which individual subjects were exposed to the intervention.

2.3.3 The phased intervention design

2.3.3.1 Number of clusters for randomisation

The Triage Plus study was implemented in rural areas of Lilongwe where the level of infection with *mycobacterium tuberculosis* is generally low compared to the urban areas. For instance, in a TB prevalence survey conducted by Ayles et al (2009), two areas (one urban and the other rural) in Zambia showed that the adjusted TB prevalence in urban areas was approximately two times higher than that of rural areas (650/100,000 (95% CI 360–940) in the rural areas vs 1200/100,000 (95% CI 750–1640 in urban areas). However, it is acknowledged that in other settings different patterns exist. A TB prevalence study conducted in India found higher TB prevalence in rural areas than in urban areas (Rao et al., 2012). To obtain an adequate number of TB cases for the primary outcome measures, such as TB treatment initiation rates, large cluster sizes were used for randomisation to different study arms. In total, 6 clusters were formed by subdividing the Lilongwe rural communities. The cluster boundaries were based on existing boundaries drawn by traditional authorities. Approximately three traditional authorities were grouped to form a cluster. This not only permitted obtaining adequate numbers of TB and HIV cases but also minimised the risk of potential contamination of the intervention between intervention arms due to reduced interactions of people between clusters due to the large size of the clusters. When small clusters with no corresponding buffer zones are used, there is a high likelihood of people between clusters interacting thereby affecting the effectiveness of the intervention (Hayes et al., 2000; Torgerson, 2001; Borm et al., 2005).Thus, the Triage Plus study had three clusters in each of the two arms.



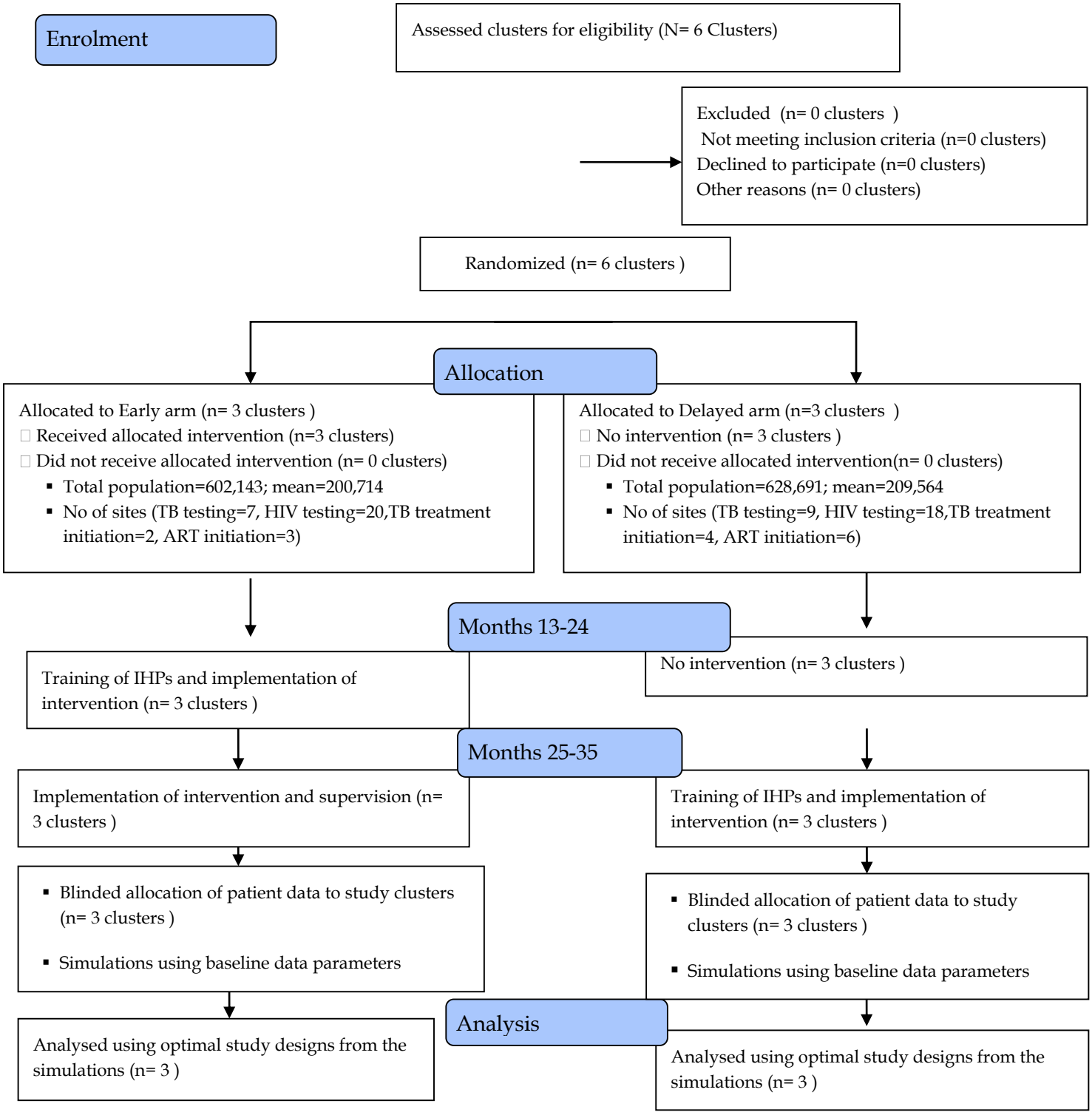


Figure 1: Flow diagram

2.3.3.2 Matching of clusters

Although randomising clusters to the interventions arms without stratification or matching improves statistical power, this approach is actually ideal when there are many clusters available for randomisation (Klar and Donner, 1997). However, with the 6 clusters available for randomisation in the Triage Plus study, obtaining balanced distribution of potential risk factors between intervention arms was difficult. Thus, before the 6 clusters were randomised to different intervention arms (an Early arm and a Delayed arm) as described below in section 2.3.3.3 and 2.3.3.5, the 6 clusters were pair-matched according to population sizes and whether or not the clusters bordered urban areas. The pair-matching was done when all the available clusters (i.e. the 6 clusters) had been enrolled in the study. Similar clusters based on population sizes and proximity to urban areas were pair-matched. One of the pair-matched pairs was then randomly assigned to the intervention arm as described in 2.3.3.3.This was done to minimise the imbalance between the two intervention arms as well as to control for the effects of urbanisation (Mc Clatchey et al., 1992; Hayes et al., 2000).

2.3.3.3 The randomisation process

A randomisation process of the pair-matched clusters was carried out in Microsoft Excel, which involved grouping the clusters into two. One cluster pair was allocated to group 1 and the other to group 2. Using a built in random number generator in each cluster pair, the pair with the lower randomised number was assigned to arm A and the other to arm B. To ensure that the final randomisation process was participatory by all the key stakeholders, then, two pieces of paper were labelled either arm A or arm B and placed in an opaque envelope of equal size and appearance and then sealed. One of the key stakeholders not involved in the randomisation process was asked to pick one envelope after the two envelopes were swapped several times. This randomisation process was discussed in detail at district and national level and the decision taken to not involve the communities in terms of where the intervention would be implemented because all communities eventually received the intervention during the full study period. Randomisation involving communities is most attractive in situations where the intervention may not be scaled out to the control communities (Ayles et al, 2013). Before selection began, it was agreed that the envelope selected would be the Early intervention arm and the other would be the Delayed intervention arm. Figure 2 shows the distribution of clusters between the intervention arms.

2.3.3.4 Blinding of the intervention

Because of the nature of the intervention implemented at community level, it was not possible to conceal either the allocation of the intervention from the communities or from the research team involved in the study. However, patient allocation to specific clusters was blindly done by an independent person who had no knowledge of the intervention.

2.3.3.5 The Triage Plus study implementation phases

In the Early intervention clusters, the schedule of the project implementation was decided randomly and each selected cluster received the full intervention before moving to the subsequent clusters. All 3 clusters in the Early intervention arm received the full intervention package within 3 months. The intervention package was implemented in the Early intervention clusters for a period of 12 months. During this time, the Delayed intervention arm continued receiving standard care. In both Early and Delayed arms, data collection continued throughout the 12 months called 'Phase I'.

Due to time constraints (inadequate time to conduct data entry, cleaning, analysis and final write up of the thesis), the Delayed intervention arm received a full intervention package for only the 11 months immediately following the initial 12 months of Phase I. The Early arm clusters continued to receive the intervention during this time. In both arms, data collection continued in the 11 month period called 'Phase II'.

Before cluster randomisation was done, baseline data for the study outcome measures and contextual factors (see section 2.6) was collected over the 12 month period before the initiation of the intervention in all clusters (called 'Baseline period'). Figure 3 summarises the phased, pair-matched, parallel cluster design used in the Triage Plus study.

In the Triage Plus study, an approximately equal target population was allocated to each of the two intervention arms. There was a total population of 602,143 (49%) in the Early intervention clusters and 628,691 (51%) in the Delayed intervention clusters based on the 2008 National population and housing census (National Statistical Office, 2010). Mean cluster population sizes were 200,714 (ranged from 166,702 to 243, 826) in the Early arm and209,564 (ranged from 167,074 to 232,433) in the Delayed arm (see Appendix 8.3 for cluster level populations). Overall, poverty levels between the two intervention arms (Early and Delayed)

were similar (see Table 1). Cluster characteristics are compared in section 6.4 of the statistical analysis (Chapter 6).

Table 1: Poverty incidence in Lilongwe rural areas

Early arm*				Delayed arm**			
Traditional authority	Total house-holds	Poverty Headcount (%)	Ultra poverty headcount (%)	Traditional authority	Total house-holds	Poverty Head-count (%)	Ultra Poverty Headcount (%)
Chadza	19,173	80.6	52.7	Kalolo	23,457	96.1	39.7
Kabudula	17,435	80.3	50.0	Chiseka	40,371	66.3	39.8
Mazengera	18,358	82.8	56.6	Chimutu	15,403	87.9	64.8
Chitekwere	6,632	75.0	42.0	Chitukula	4,824	83.9	54.7
Khongoni	15,954	75.9	45.1	Mtema	7,648	82.9	53.4
Kalumbu	11,321	89.0	66.9	Malili	14,895	74.2	44.0
Tsabango	4,885	87.1	63.3				
Kalumba	4,289	76.7	50.1				
Njewa	4,891	75.3	47.1				

*Early arm = clusters that received the intervention early;

**Delayed arm = clusters that received the intervention in the second half (next 11 months) of the study period.

Source: Malawi: An Atlas of Social Statistics, 2002

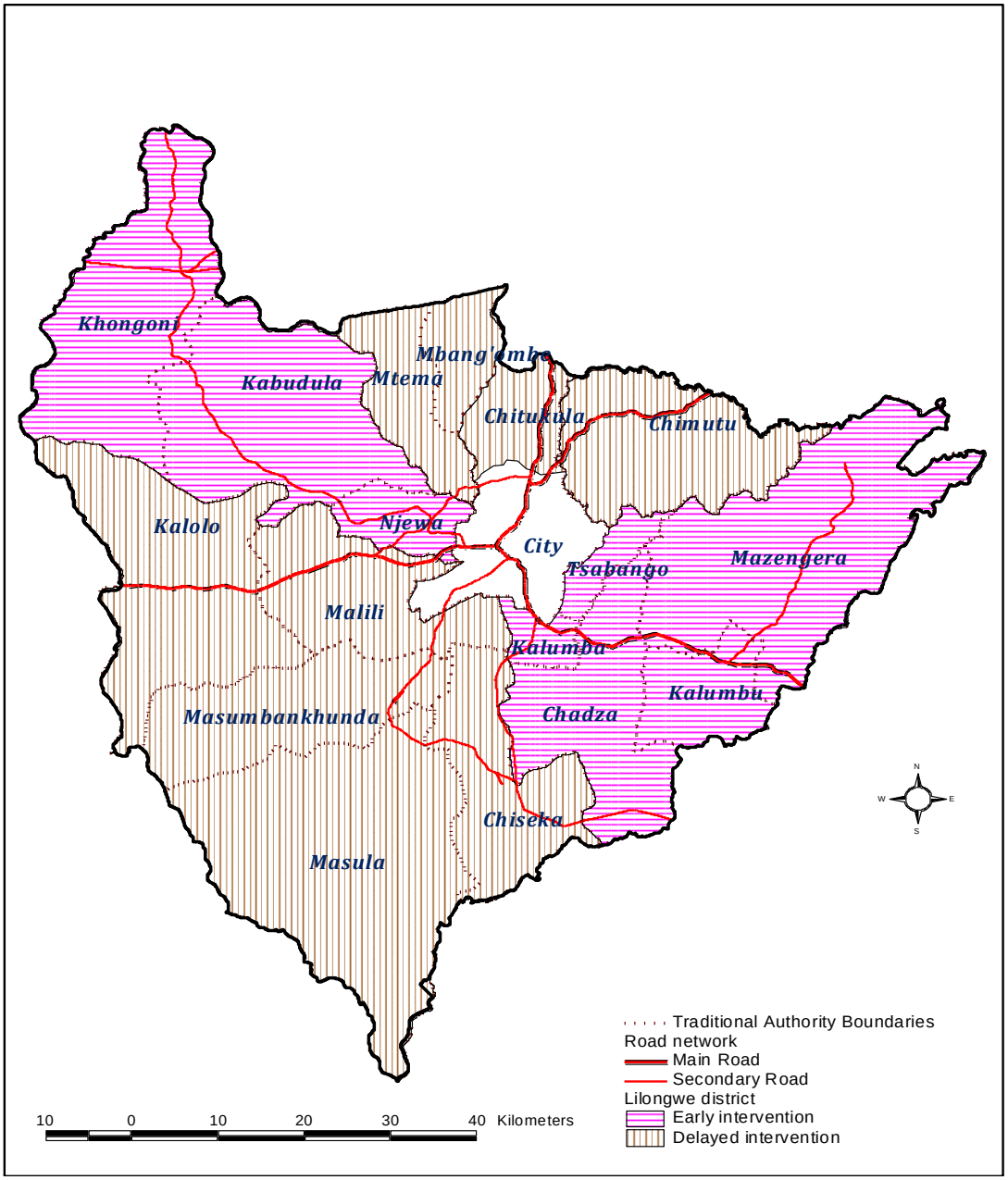


Figure 2: Map of Lilongwe showing the distribution of the Triage Plus study clusters between intervention arms.

The pink background represents clusters in the Early intervention arm and the green background represents clusters in the Delayed intervention arm. Each cluster is formed by grouping 2 to 3 traditional authorities which have purple boundary lines. The names of the traditional authorities are indicated inside the clusters. The centre of the map is the urban area which was not part of the study.

Intervention arm	Baseline	Intervention	
	Baseline period	Phase I	Phase II
	1 - 12 months	13 - 24 months	25 - 35 months
Early	Baseline data	Intervention	
Delayed	Baseline data	No intervention	Intervention

Figure 3: Triage Plus study design.

The figure shows the phased, pair-matched, parallel cluster design which was used to randomise study clusters to Early and Delayed arms. The blue colour represents the baseline period and the dark purple represents when the intervention package was implemented in the respective arms and the light purple under the Delayed intervention arm represents study period when no intervention was implemented in the Delayed arm.

2.4 Study intervention

2.4.1 The Intervention package

Informal healthcare providers are known to be the first point of call along the health-seeking pathway for the most poor and vulnerable (e.g., see Simwaka et al., 2012).The Triage Plus study, therefore, aimed to assess the potential of involving a variety of informal health care providers such as storekeepers, traditional healers and village health committees in assisting in the diagnostic and treatment process for TB and HIV at the community level.

Thus, the intervention package entailed:

a) Training informal health care providers to effectively implement the components of the intervention.

Informal health care providers were equipped with skills to implement TB and HIV interventions. They were trained to implement the following intervention components: recognising disease patterns for TB and HIV, assisting in TB sputum specimen collection, appropriately referring TB and HIV suspects to the public health system and conducting TB

and HIV community awareness meetings. Intervention materials were developed that were used in all community meetings to ensure accurate information was given, and guidelines were developed to assist informal healthcare providers with implementing specific tasks. Health surveillance assistants provided technical support to the informal healthcare providers during project implementation and linked them to the formal public health service.

At community sensitisation meetings, the health messages included information on the signs and symptoms of TB, HIV and AIDS and that treatment is provided free in public health facilities. Thus, the primary focus of the intervention was to alter the health seeking behavioural patterns and service demand thereby influencing the community's access to TB and HIV services.

Community members with TB and HIV related symptoms were advised to seek medical care in public health facilities. However for those who could not afford to do so, especially for TB, sputum smears were collected in the community by the informal healthcare providers and sent to the nearest public health facility for microscopy testing. Results were fed back to the informal health care providers by local health surveillance assistants.

b) Training front line public health personnel to effectively support informal health care providers in the project activities, and engaging local public health services and community leaders to support the intervention.

To ensure the informal health care providers received the necessary support to implement the intervention, health surveillance assistants were also trained, and they acted as a link between the informal health care providers and the public health facilities and provided supplies for sputum collection. They also provided supportive supervision to the informal healthcare providers. In addition, all local leaders in the intervention areas were sensitised to the TB and HIV issues covered during the training of informal providers. This sensitisation of local leaders enabled the informal healthcare providers to be accepted in the community. The local leaders also provided support in arranging community meetings.

2.4.2 Inclusion and exclusion criteria for the informal health care providers

The informal healthcare providers were engaged if they were a likely access point for poor and vulnerable people in the community with TB and/or HIV and if they had a physical location within the study area for at least a year. These informal providers were identified among the existing community structures already functional in the community, such as community based organisations and village health committees. The selection process was participatory. They were excluded if they were already formally trained in clinical disciplines or public health, or funded through civil services.

The informal health care providers engaged included store-keepers, patient support groups (including youth groups), home based care groups, faith-based organisations or volunteers from local non-governmental organisations (NGOs) or community based organisations (CBOs) based in the area, village health committees, or herbalists.

To ensure even distribution of informal health care providers engaged in the Triage Plus study, the number of informal providers engaged was proportionate to the cluster population size.

2.4.3 Conceptual basis of the intervention Framework

Using intervention mapping (Bartholomew et al., 2006; 2011) and the Medical Research Council (MRC) framework for the evaluation of complex interventions (Campbell et al., (2000), the TB and HIV community intervention was developed involving informal health care providers to improving access to services for the two diseases in the community. These frameworks were used in planning and developing the intervention. Using these frameworks in systematically developing the intervention improved the chances of the intervention being effective in achieving the study outcomes (Campbell et al., 2000; Bartholomew et al., 2006; 2011).

In conceptualising the Triage Plus intervention, results from previous studies on the use of informal health care providers in the delivery of health interventions (e.g. Marsh et al., 1999 & 2004; Simwaka, et al., 2012) were used. To develop the actual intervention package, results from situational analysis, qualitative baseline studies and mapping studies, and input from consultative meetings with key stakeholders were used.

This initial work provided data that was used in formulating the intervention (see Simwaka et al., 2012). This included data on community and health system barriers, TB and HIV knowledge and attitudes, existing TB and HIV burden, the optimum strategies for delivering TB and HIV health messages, the indicative capacity of different informal health care providers in the targeted areas and a matrix of specific activities that can be implemented by different types of informal health care providers.

In particular, the baseline qualitative studies helped to define potential barriers to health seeking behaviours and the review of previously used training materials informed the exact content of training materials used in the study. The situational analysis coupled with data abstraction from patient registers and existing records formed the basis for calculating sample sizes for the main study outcome measures which were strategically chosen for feasibility purposes (Mannheim, 1999). Sample sizes for the primary outcome measures of the Triage Plus study are given in section 2.5.

The conceptual framework of the Triage Plus intervention was guided by a number of theoretical frameworks.

First, the health belief model (Rosenstock, Strecher and Becker, 1988) and behavioural model of services utilisation (Phillips et al., 1998) were used for identifying and addressing barriers related to the health system and to individuals in accessing TB and HIV services. This theoretical framework primarily focused on health services access among the poor within the context of existing factors in the study communities.

Second, attribution theory (Weiner, 1986) was examined for increasing the community's understanding of the causality and control of TB and HIV/AIDS conditions. Social cognitive and learning theories (Rosenstock, Strecher and Becker, 1988;Bandura, 1971 & 1986) were examined in terms of using informal health care providers to conduct community awareness campaigns to improve a community's knowledge of TB and HIV health risks, and the resulting benefits of the adoption of preventive behaviour and seeking a diagnosis or medication for the disease, in order to build a sense of self-efficacy among the community members and increase demand for TB and HIV services.

Third, the conceptual basis of the Triage Plus intervention model, derived from the social learning theory, entailed modifying the general health seeking behaviour among community members. This was first done by engaging community leaders through sensitisation meetings at the onset of the project to ensure their better understanding of the disease conditions targeted by the project. This was followed by the engagement of informal health care providers to implement TB and HIV awareness interventions to reach the general community and promote the adoption of positive health seeking behaviours, such as seeking a diagnosis for TB symptoms or an HIV test to know their HIV status, which would, in turn, lead to accessing TB and ART treatment. As the intervention progresses, the more individuals who have adopted good health care behaviours, the more community members in turn access the health services.

The intervention was designed and implemented while recognising the sensitivities of involving informal health care providers in delivering TB and HIV activities at the community level. The initial work at different levels of project conceptualisation, design and execution helped in the implementation of the whole study. Figure 4 summarises the conceptualisation, design and implementation framework of the project.

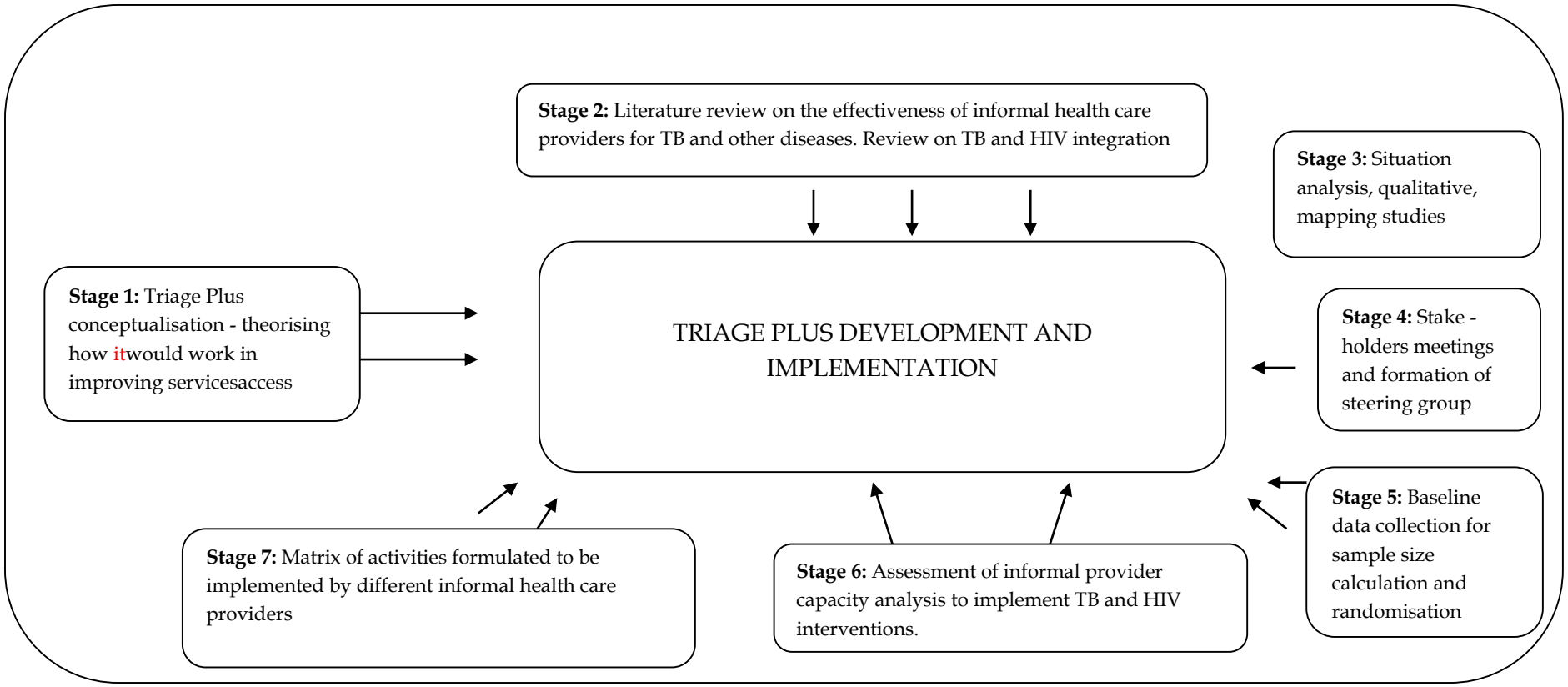


Figure 4: The Triage Plus study implementation framework.

The framework is showing the steps involved in designing and developing the intervention. Note that the order given does not necessarily reflect the order of the activities but shows the various steps involved in the development of the intervention.

2.5 The Triage Plus study outcome measures and sample size

Because of the complex nature of the Triage Plus intervention, many outcome measures were established (Table 2). The outcomes included changes in the number of people starting treatment for TB and ART, changes in TB and HIV case detection and changes in uptake rates of diagnostic tests. Because there were two diseases/conditions investigated in this study, two primary outcome measures were used, one for each of the two diseases (TB and HIV); the analyses of the two outcome measures were carried out independently. In statistical analyses involving multiple comparisons when there are two or more primary outcome measures, Bonferroni-type adjustments are usually made to ensure that the overall type I error rate is held at the conventional 5% level by setting the alpha level for statistical significance for each individual comparison at a lower level (often as low as 1%).

Table 2: Intervention effectiveness outcome measures used in Triage Plus trial
Primary outcomes*
TB treatment initiation rates
ART treatment initiation rates
Secondary outcomes*
Smear positive TB treatment initiation rates
TB testing uptake rates
HIV testing uptake rates
* Treatment initiation and testing uptake rates were based on incidence rate ratios

This analysis of multiple intervention outcomes is not only necessary to assess the effectiveness of the various intervention efforts but also to ensure that changes in the treatment initiation rates especially for TB should be preceded by changes in testing uptake rates. For instance, community awareness meetings run by informal health care providers increase appropriate knowledge on TB and HIV transmission, diagnosis and treatment, which results in changes in health seeking behaviours which increases the number of people accessing testing services and subsequently receiving treatment. Thus, treatment initiation rate changes are a consequence of changes in health seeking behaviour, testing services uptake and case detection rates. Some of these outcome measures are more likely to be sensitive in showing the effectiveness of the intervention than others, especially in TB treatment initiations, because TB treatment initiations are not as common as ART initiations. Thus, the use of multiple outcome measures in assessing the effect of the intervention allows the isolation of distinct intervention effects on the various outcomes measures.

The primary objective of the Triage Plus study was to measure the impact of engaging non-paid informal health care providers in integrated TB and HIV community interventions on TB and HIV treatment initiation rates. Although baseline summary data on TB and HIV were collected from the study areas during the study design stage to help in sample size determination, data on the clustering effects were not known. In addition, the study had a limited number of clusters available for randomisation that could potentially limit the statistical power. In this case sample size calculation to achieve adequate statistical power and the accuracy in parameter estimates was based on simulation studies that were conducted using a repeated measurement design (see Chapter 5). Because there was one primary outcome measure for each of the two diseases (TB treatment initiation and ART treatment initiations), the simulation studies investigated the adequacy of statistical power for both low incidence events (e.g. TB treatment initiations) and high incidence events (e.g. ART treatment initiations). Thus, baseline treatment initiation rates were used in the simulation process. We powered our study to be able to detect a 20% increase in the treatment initiation and testing uptake rates (see Chapter 5 for the details and the results of the simulations).

The purpose of the simulations was to identify the study design conditions that would provide adequate statistical power and accuracy of parameter estimates for both the TB and HIV outcomes that would then be used in the final analysis of the Triage Plus study. Ideally the simulation studies should have been completed before the intervention was implemented but for logistical reasons the simulation studies had to be conducted concurrently with the implementation of the intervention (the study design[1] was later refined just before the intervention started). Failure to specify the statistical analysis plan or appropriate design conditions such as the number of clusters available (Chapter 5) risks creating a situation in which only statistically significant findings can be reported. Our study was less likely to be affected by these challenges because the parameters used in the simulations were based on the baseline data and not the actual study data.

[1]The study design to include more clusters than the original two clusters was refined as part of my PhD thesis so as to be consistent with the standard practices in cluster randomised trials, though the number of clusters finally defined were still limited due to cost challenges as well as fear of contamination if very small clusters were to be defined within the study area.

The original individual level sample size determination was based on the data obtained through the situation analysis carried out at the design stage (see fourth paragraph in section 2.4.3). Thus, sample size determination was targeted at determining the minimum number of TB and HIV cases that would be needed to start treatment after the intervention in order to detect a minimum pre-determined intervention effect size.

To estimate the sample size for the TB primary outcome (i.e. TB treatment initiation rates), the number of smear positive TB cases starting treatment was used as the basis for sample size calculation. A minimum of a 20% increase in number of smear positive TB cases starting treatment from the baseline level of 50% was assumed (the WHO estimated that Malawi was detecting less than 50% of smear-positive TB cases nation-wide (WHO, 2009)). Thus, an increase in this proportion to the international case detection target of 70% in the intervention group was considered to be clinically significant.

Using the secondary data collected as part of the situation analysis, the catchment population of the Lilongwe District rural health centres was estimated to be approximately 1.1 million and the total number of TB cases registered in these rural health centres for which data was available was 936, of which 329 were smear positives. By assuming that the actual TB cases were not rising in the community year on year, the expected number of smear positive TB cases registering for treatment in each study group over a one year period was thus estimated to be 165 (329/2). This constituted 50% of all smear positive TB cases, so the actual number of smear positive TB cases occurring in each study group was projected to be 330 (165/0.50). This provided 87.5% power to detect an increase in the number of smear positive TB cases registering for treatment from 165 (50%) in the control arm to 231 (70%) in the intervention arm. Given that 35% (329/936) of the TB cases were smear positive, then a minimum of 462 TB cases (all forms of TB) were expected to start TB treatment at the end of the intervention in the intervention arm.

For HIV infected patients starting treatment, an increase from the baseline level of 15% up to 18% (i.e. a 20% proportional increase) in the intervention arm would be considered clinically significant. Malawi's 5 year ART Scale-Up plan indicated that (approx.) 85,000 new patients become eligible for ART each year across the country, in which case 8,500 new patients would

become eligible for ART in the rural areas of Lilongwe[2]. Data from ART programme in Lilongwe rural indicated that 1,300 people were already on ART treatment, a treatment initiation rate of 15% (1300/8500), and that the number of new patients eligible for ART was 8,500 (4,250 for each of the two study groups). A 20% increase in ART treatment initiations from the baseline 15% in the intervention arm was again considered clinically significant, so the available sample size provided 86.9% statistical power to such an increase in the proportion of eligible patients starting ART (in absolute terms, from 15% in the control arm to 18% in the intervention arm).

In the calculation of the sample sizes for the primary outcome measures, the design effect was not used because an estimate of the intracluster correlation was not known at the design stage. In the absence of data from other studies sufficiently similar to Triage Plus study from which to obtain estimates of the likely ICC value for inclusion in the sample size calculations, the baseline data from Triage Plus were used; these provided ICC estimates for the four different outcome measures in this study. The estimated ICC values for TB testing, ART initiation, TB treatment initiation and HIV testing rates were 0.00154, 0.0316, 0.046 and 0.081 respectively. For the simulation studies presented in Chapter 5, the smallest (0.00154) and largest (0.081) of these four ICC values were used, along with two much larger ICC values (0.321 and 0.699) taken from other similar studies, in order to test a very wide, but also realistic, range of ICC values in the simulations. The ICC of 0.321 was based on a cluster randomised trial to assess the effect of an education intervention on TB case detection and primary care of respiratory diseases in South Africa by Fairall et al, 2005 but cited and reanalysed by Clark and Bachmann, 2009. The ICC of 0.699 was based on the high ICCs implemented by Stryhn et al 2006 when they compared different approaches for determining ICCs in terms of their robustness (see also section 5.3.2 below).

2.6 Data collection for impact evaluation

2.6.1 HIV and TB treatment initiation and testing data

As part of the evaluation of the intervention, new TB and HIV cases (individuals aged 12 years and above) reporting to health facilities in the study district from January 2009 to the 31st March 2012 were included in the analysis[3]. Routine TB and ART patient treatment registers

[2]The 1.1 million people living in Lilongwe rural accounted for the 10% of the total population in Malawi.

[3]Children aged below 12 years were excluded because they rarely cough up sputum and confirmation by smear examination is impossible. For HIV, the study sought to exclude those who obtained HIV through vertical transmission.

were the primary sources of data used for measuring the primary outcome measures. Data for the TB and ART treatment initiations were collected from patient registers in all the health facilities by photographing register pages and directly entered into excel database. To ensure all cases were entered in the database, data entry verification was done by comparing number of cases in the database and those in the patient registers for each of the health facilities. The TB treatment register collected data on patient addresses, testing and treatment initiation dates, patient demographics, whether the patient had pulmonary or extra pulmonary TB, patient category (new, relapse, return after default, or others), sputum smear microscopy results and treatment outcome (cured, treatment completed, failure, died, and transferred out) (see Appendix 8.2.1). Similar critical variables were collected in ART patient treatment registers (see Appendix 8.2.2). In addition to the treatment initiation data, testing data for TB and HIV were abstracted from registers in TB and HIV testing sites using similar approaches described for TB and ART treatment initiation data collection (see Appendices 8.2.3 and 8.2.4 for data collection forms for TB and HIV testing data).

Data for TB and ART treatment initiations and TB diagnoses were extracted from patient registers from all health facilities in the district. Patients were allocated to the different clusters based on their residential address. The rationale for this was the expectation that people within clusters might be influenced by the intervention to seek care, but may choose to visit a health facility outside their cluster catchment area for a variety of reasons. However, because the HIV testing data were anonymised, data collection was confined to the rural areas and the HIV testing data were allocated to clusters based on the location of the HIV testing site, rather than by residential address of the person tested. This approach might have led to under-reporting or over-reporting of cases in a given cluster in a number of ways. First, individuals in a given cluster might have gone to a different HIV testing site outside the cluster thereby under-reporting such cases from their cluster of residence and over-reporting in the clusters where the HIV testing was sought. Second, individuals seeking HIV testing in health facilities located in the urban area might have led to underreporting of cases in the study clusters.

These data sources were chosen for several reasons: a) because the intervention was pragmatic and would be used to measure programme success rates, which is usually done using the routine data sources, b) because of the consistency of data collection methods, treatment initiation guidelines and testing methods across all health facilities whether public or not as

per policy guidelines, c) due to the extended period required for TB and HIV/AIDS treatment , these records are the only reliable source of data for tracking patients' treatment progress.

HIV and TB treatment initiation and testing data were collected at baseline (a full one year) and during the course of the intervention in both intervention arms. To strengthen correct and complete documentation of data in patient registers in health facilities, all health workers involved in documenting patient data in the patient registers (i.e. completing patient registers) were trained and further supported by regular supervisory visits.

2.6.2 Contextual factors (Health system and population characteristics data)

Community contextual factors that are likely to affect the intervention, such as health system data (distribution of TB treatment sites, microscopy sites, ART sites, and HIV testing sites by cluster and population characteristics (age and gender distribution), were documented at baseline. The number of health facilities offering TB and HIV services (i.e. treatment initiations and testing services) in each cluster was based on the number of the health facilities physically located in the respective clusters. The demographic characteristics (age and gender) of the patients for each of the conditions studied were based on the data routinely collected from patient registers. Cluster specific populations were based on the 2008 Population and Housing census conducted by the National Statistics Office (NSO, 2010). These cluster population data were used in the calculations of incidence rate ratios for the outcome measures for the Triage Plus study.

2.7 Study limitations

Because of the low level of TB infection in rural areas, the Triage Plus intervention used very large clusters. This in turn, created challenges in having adequate number of clusters for the two intervention arms (see section 2.3.3.1). The use of a limited number of clusters in the Triage Plus study posed some statistical challenges in achieving statistical power and obtaining robust estimates when assessing its effectiveness (Murray, Hannan and Baker, 1996; Donner and Klar, 2000). With this limited number of clusters, the Triage Plus study is at risk of being underpowered and, therefore, failing to detect significant effects.

To address these foreseeable challenges, alternative approaches in the analysis of the Triage Plus study have been pursued in this dissertation. In particular, the use of a repeated measurements design to increase the number of degrees of freedom for statistical power and robust parameter estimation. Although previous research has tried to assess this, most studies

used at least 5 clusters per arm and no specific guidelines on the number of repeated measurements required (e.g. Murray et al., 1998b). In this thesis we investigate the utility of taking repeated measurements in cluster randomised trials involving 3 clusters per arm in improving statistical power and the robust estimation of parameters (Chapter 5).

2.8 Conclusion

This Chapter has provided an overview of the need for engaging informal healthcare providers in community interventions to improve access to TB and HIV services. In addition, theoretic frameworks that guided the development of the Triage Plus have been briefly summarised. The frameworks helped to plan and develop the intervention in a manner that would be effective in achieving the intended objectives. The methodological implementation for the Triage Plus study in terms of study design, the development of intervention package, randomisation procedures, and data collection methods and type of data collected, was also presented. The randomisation of the pair-matched clusters was to ensure comparability between intervention arms.

Furthermore, possible challenges that the Triage Plus study is likely to face in terms of achieving adequate statistical power because of the limited number of clusters used for randomisation were also presented. The chapter finally suggests the need for further research for addressing the foreseeable challenge in statistical power, which forms part of the core research area for this thesis.

CHAPTER 3

CLUSTER RANDOMISED TRIALS

3.1 Introduction

Chapter 2 presented an overview of the Triage Plus study, which involved informal health care providers implementing integrated TB and HIV interventions at the community level. The study aimed to assess whether or not engaging these informal healthcare providers would improve access to TB and HIV services in Lilongwe rural areas. To allow a better evaluation of the intervention, a cluster randomised intervention design was used to allocate three clusters to each of the 2 intervention arms (Early and Delayed arms). Although the use of cluster randomised trial designs ensure the even spread of known and unknown confounding variables between intervention arms, cluster randomised trials have their own inherent challenges, such as they are limited by the number of clusters available for randomisation and by correlations in the data derived from such cluster randomised designs (Cornfield et al., 1978; Murray, 1998a; Hayes et al., 2000).

This Chapter reviews the most commonly used cluster randomised designs for community interventions; statistical methods for assessing intervention effectiveness, as well as the challenges encountered when cluster randomised intervention designs are used (presented in section 3.2). Section 3.3 also discusses some approaches that have been used to address these challenges, and section 3.4 discusses the implications of having a limited number of clusters in the Triage Plus study in terms of intervention effect estimation.

Given the challenges in mitigating the limited number of clusters in cluster randomised designs for each of the alternative approaches, the possibility of adopting a repeated measurement design for evaluating cluster randomised designs with a limited number of clusters is discussed in section 3.5. The repeated measurement design's longitudinal nature is normally used in investigating changes occurring within the subject or cluster over time, and then inferences are made from the results about the general population (Yee and Niemeier, 1996). Repeated measurement designs may be used to avert the statistical challenges that arise due to a limited number of clusters. This Chapter further reviews the effect measures and estimation approaches used in intervention effect evaluation in section 3.6. Section 3.7 briefly presents alternative approaches for analysing hierarchically structured data and discussing the challenges inherent in them; section 3.8 concludes the Chapter.

3.2 Cluster randomisation trials and their challenges

3.2.1 Cluster randomisation designs

Cluster randomised trials are increasingly used in the evaluation of community interventions to ensure the generalisation of the intervention's effectiveness (Campbell et al., 2000; Atienza and King 2002). These cluster randomized trials, also called group randomised interventions or community randomised interventions, are comparative studies in which intact clusters of individuals are assigned to intervention or control groups. This cluster randomization has practical advantages in that all subjects in the cluster are treated in the same way, thereby reducing both costs and ethical concerns as pointed out by Donner et al. (1990). This randomisation design is usually employed largely because the nature of the intervention prevents its implementation at the individual level or for fear of treatment contamination between intervention and control groups, which results in a reduction of the effectiveness of the intervention (Hayes et al., 2000; Torgerson, 2001; Borm et al., 2005).

Cluster randomised studies have primarily employed either parallel, crossover or stepped wedged designs in allocating clusters to intervention groups. In parallel designs, clusters are randomized to either an intervention or control arm at the same time point. This approach has been used in the evaluation of interventions in cancer research (Allen et al., 2001), antismoking interventions (Cummings et al., 1998; Aveyard et al., 2001), HIV and sexually transmitted infections (Grosskurth et al., 1995; Rotheram-Borus et al., 2001), and tuberculosis interventions (Corbett et al., 2010). In the analysis of parallel cluster randomised trials, Murray (1998a), Donner and Klar (2000), Murray et al. (2004, 2008), and Varnell et al. (2004) have provided a review of analytic methods appropriate for such parallel group randomised trials.

In a crossover design, each cluster both receives the intervention and in turn a control in equal time intervals (Palmer et al., 1985; Chavasse et al., 1999; Sjogren et al., 2005). In such within-cluster evaluation designs, a more precise treatment evaluation is allowed and fewer clusters are required than in parallel designs (Hussey and Hughes, 2007). However, the time taken to complete the evaluation may be twice as long, and knowledge of how the intervention works is critical. If the time intervals between switching treatments are short then: a) the intervention may not have adequate time to take effect, b) inadequate measurements or outcomes for accurate estimates of the impact are possible, or c) there may be a carry-over effect that persists from the beginning of the first treatment (Hussey and Hughes, 2007). Thus, crossover cluster randomised trials are attractive in treatments with short term response (Pocock, 1997).

The stepped wedge design, also known as the dynamic wait-listed, delayed intervention, or phased intervention, is a type of crossover design in which different clusters switch treatments at different times and the crossover is in one direction from control to intervention (Hussey and Hughes, 2007; Brown and Lilford, 2006, The Gambia Hepatitis Study Group, 1987; De Allegri et al., 2008). In a stepped wedge design, the baseline data collection or measurement corresponds to the first time point where none of the clusters receive the intervention. Clusters are then randomised to intervention at subsequent time points and the response to the intervention measured until all sets of clusters have received the intervention (see The Gambia Hepatitis Study Group, 1987).

This design approach not only removes the ethical concern of withholding the intervention from some communities (Hussey and Hughes, 2007), but it also improves the efficiency of the intervention in that clusters act as their own control, thus fewer clusters receive the intervention. In addition, the phased assignment of clusters to the intervention leads to a high level of project implementation because it reduces the likeliness of intervention activities being thinly spread over a large geographical area due to limited human or financial resources (Brown et al., 2006). However, the approach extends the time required for the evaluation of the intervention. Furthermore, the one direction in which clusters switch treatments makes it hard to use within cluster comparisons normally employed in standard crossover designs (Hussey and Hughes, 2007). In this design, the overall effect of the intervention is made by comparing between the data points in the intervention and control sections of the wedge (Hussey and Hughes, 2007).

3.2.2 Challenges of cluster randomised trials

The use of cluster randomised trials for evaluating community interventions has received considerable attention in recent years. Just as scientific evidence from individually randomised clinical trials is considered to be the gold standard, for interventions at the group level, the community randomised trial design, where distinct groups of individuals are randomised to intervention or control arms is considered to be the gold standard. Although there is an increase in the use of community randomised trials for interventions that are better implemented at the group level, the adoption of this type of trial has in turn created methodological challenges in the design and evaluation of cluster randomised interventions (Murray, 1998; Campbell et al., 2000; Murray et al., 2004). These challenges are primarily the

limited number of clusters available for randomisation; and the reduced statistical power due to the presence of intracluster correlations (Hayes et al., 2000; Murray, 1998a).

3.2.2.1 Number of clusters available for randomisation

The first challenge in cluster randomised interventions is that the number of clusters available for randomisation is often limited as the intervention is delivered at the group level (Hayes et al., 2000). Ideally to achieve statistical power a large number of clusters are needed. However, often a small number of clusters with a large number of participants is randomised to each intervention arm (in the Triage Plus study, only 3 clusters were randomised to each intervention arm). As pointed out in Chapter 1 section 1.2 and Chapter 2 section 2.3.3.1, large numbers of clusters are commonly used for randomisation to different intervention arms to ensure adequate statistical power. However, to obtain adequate numbers of events for the primary outcome measures (in our case TB and ART treatment initiations), large number of people within the clusters are needed.

However, if there is a small number of clusters randomised to each intervention arm, it results in a number of challenges: there would be a limited number of degrees of freedom for estimating between-cluster variance for cluster-level estimates, which limits the necessary power for a valid test for the effect of the intervention (Cornfield 1978; Koepsell et al., 1991; Murray, Hannan and Baker 1996; Murray, 1998; Donner and Klar, 2000). Therefore, sample size calculation in cluster randomised trials should primarily be based on the number of clusters per condition and not only on the number of individuals (Murray, 1998a; Feldman, McKinlay and Niknian, 1996). Variations between cluster-level outcomes are more important than the within cluster variations when assessing intervention effect (Feldman, McKinlay and Niknian, 1996; Maas and Hox, 2005; Moineddin, Matheson and Glazier, 2007; Hayes and Moulton, 2009). However, the use of large cluster sizes in community interventions reduces variability in variance between clusters (Hayes and Moulton, 2009). Thus, the statistical power in a study and the precision of parameter estimates of intervention effects largely depend on the observed variability between clusters of the effect measure.

However, despite the need for more clusters for improved power in intervention effect testing, underpowered community randomised intervention studies with inadequate groups still exist (Donner et al., 1990; Simpson, Klar, and Donner, 1995; Varnell et al., 2004; Murray et al., 2008; Pals, Wiegand and Murray, 2011). A review by Pals, Wiegand and Murray (2011) of the

HIV/AIDS community intervention trials published between 2005 and 2009 showed that about half of the trials used five or fewer groups per condition. Similarly, Varnell et al (2004), found that in a review of group randomised trials published in the American Journal of Public Health and Preventive Medicine from 1998 through 2002, 15 (25%) of the trials assigned five groups or less to each treatment, and 3 (5%) had one group per treatment arm. Simpson et al (1995) in a review of the same American Journal of Public Health and Preventive Medicine about a decade before in articles published from 1990 through 1993 observed similar trends. In their review they found that eight of the trials had 8 or fewer clusters per intervention group and two had one cluster per intervention group. A review by Donner et al. (1990) for trials published from 1979 through 1989 in various journals found similar results.

With a limited number of clusters, another challenge is that the standard errors, p-values and confidence intervals of intervention effects are underestimated, though point estimation of intervention effects are unbiased (Feng et al., 2001; Maas and Hox, 2005; Moineddin, Matheson and Glazier, 2007).

Furthermore, with a small number of clusters, it is difficult to achieve the even distribution of potential confounders between intervention arms, which compromises the internal validity of the trial (Murray, 1998a; Donner and Klar, 2000; Varnell et al., 2001). Statistical adjustments of the confounding variables may reduce the imbalance of the measured variables. However, this adjustment may not remove imbalances in unmeasured confounding variables as pointed out by Kramer et al. (2009). This reduces the likelihood of making causal inferences on the effect of the intervention when comparing the study conditions (Murray, 1998a).

3.2.2.2 Reduced statistical power due to within cluster correlation of outcomes

Use of clusters or communities as units of randomisation in cluster randomised studies leads to the correlation of outcome measures within the clusters because of the similarity of values for the outcome measures taken from members of the same cluster (Hayes and Moulton, 2009).In cluster randomised interventions, the between and within group variability in the response outcomes is usually larger than that for individually randomised trials due to the correlation of outcomes within clusters (Kish, 1965; Mickey et al., 1991).

This extra variation is in part due to differences in characteristics between groups and in part by the mutual interaction of individuals within the same group as pointed out by Feng et al. (2001). Furthermore, people living within the same cluster tend to be more similar than do individuals across different clusters because they share the same environment and interact with each other (Kim et al., 2006).

For example, in a community intervention aimed at reducing waterborne diseases, individuals in the same community are more likely to use the same water source, a different source than people in other communities use. If the community's water source is contaminated with waterborne infections, then individuals using that water source would be more likely to develop the disease than the other individuals in different communities who do not use the contaminated source.

The within cluster correlation of outcomes is usually assessed by the intracluster correlation coefficient (ICC) that measures magnitude of correlation among cluster members (i.e. the proportion of the true total variation in the outcome that is attributed to differences between the clusters).Specifically, the ICC measures the correlation between outcome measures for two randomly drawn individuals from the same cluster. The ICC, denoted as ρ, can be expressed in terms of the variance components as $\rho = \sigma_b^2/(\sigma_b^2 + \sigma_w^2)$, where σ_b^2 is the between-cluster component of the variance and σ_w^2 is the within-cluster component of the variance with subscripts b and w representing between and within variation respectively (Parker, Evangelou and Eaton, 2005; Kim et al., 2006).

In cluster randomised trial designs, the within-cluster variance σ_w^2 is usually larger than the variance between clustersσ_b^2, and therefore, ICCs obtained in cluster randomised trials are usually small. For example, in the thrombosis prevention trial conducted by the Medical Research Council, an intervention targeting directly at the patient, the two components of variance were $\sigma_w^2 = 1.28$and $\sigma_b^2 = 0.0046$, and therefore, the intracluster correlation coefficient was 0.0036 (Meade et al., 1992). When the intervention is directly aimed at individual members of each cluster, then the variability of the outcome measurement is greater than the between cluster variability. However, when the interventions aim to target, say, the doctor in order to change the doctors' behaviour in the management of patients, then the ICC is likely to be higher because the estimated ICCs include the variation in the doctors' responses. For instance, the intracluster correlation was estimated to be 0.0190 in a trial of guidelines to

improve the appropriateness of general practitioners' referrals for *x* ray examinations (Oakeshott, Kerry, and Williams, 1994).

The within-cluster correlation affects the required sample size of the study, which in turn, affects the statistical power and the precision of the estimates of the effect of the intervention (Smeeth and Siu-Woon Ng, 2002). The extent to which these within-cluster correlations affect the sample size required for a study is known as the design effect (DEFF) or variance inflation factor. The design effect, defined as the ratio of the total number of individuals required using cluster randomisation to the number required using individual randomisation, is better presented in terms of the ICC and the number of individuals in each cluster (Kerry and Bland, 1998). Thus, the design effect depends on both the size of each cluster (m) and the intracluster correlation coefficient (ρ) and is given by

$$DEFF = 1 + (m - 1)\rho \qquad (3.1)$$

However, in the case where population sizes vary across clusters, cluster population m can be replaced by an 'adjusted' mean cluster size (Kim et al., 2006) defined as

$$\text{Adjusted mean cluster size, } m_{adj} = \sum_{j=1}^{N} \frac{m_j^2}{M} \qquad (3.2)$$

where m denotes the cluster sizes, M is the total population across all clusters, j = 1,…, N indexes clusters.

In this formulation of the design effect in (3.1), when m=1 the design effect would be 1 denoting an individually randomised trial. Similarly, if the intracluster correlation coefficient $\rho = 0$, indicating the absence of the within-cluster correlations, then the design effect would be 1. Using the ICC of 0.019 from the *x* ray guidelines study (Oakeshott, Kerry, and Williams, 1994) and m=60 referrals per practice, then the design effect, DEFF =1+(60−1)×0.019=2.121. Thus, to achieve the required statistical power and precision of the estimates of the effect of the intervention, the cluster randomised trial design would require just over twice as many individuals as an intervention study where individuals were randomised to the intervention arms (Kerry and Bland, 1998; Smeeth and Siu-Woon Ng, 2002).

The design effect, therefore, quantifies the extent to which the increase in variance resulting from the use of cluster sampling in a cluster randomised intervention design departs from the variance obtained under simple random sampling (Hayes and Moulton, 2009) as:

$$Design\ effect = \frac{Variance\ for\ cluster\ sampling}{Variance\ for\ simple\ random\ sampling}$$

Therefore, the larger the ICC due to increased variation between clusters, the bigger the design effect and the more individuals that need to be recruited to get the same power as that of an individually randomised study (Kerry and Bland, 1998). Thus, the statistical power in cluster randomised studies is not only affected by the number of clusters available for randomisation, but also increases with decreasing ICC (Heo and Leon, 2005). In the simulation studies performed to assess the performance of a mixed effects logistic regression model for binary outcomes under unequal cluster sizes, the statistical power decreased from 88.1% to 10.4% when the ICC were 0.0 and 0.3 respectively with the number of clusters set at 20. It was also observed that statistical power was comparable for both equal and unequal cluster sizes (Heo and Leon, 2005) though You et al (2011) observed reduced statistical power with unequal cluster sizes and that increasing the number of clusters compensated for the power loss due to variation in cluster sizes.

The presence of within-cluster variation (whatever its magnitude) makes the intracluster correlation coefficient $\rho > 0$, and in turn, the design effect > 1, which then affects the sample size needed in the cluster randomised trial design. However, the main challenge in determining sample sizes for cluster randomised trials is having an estimate of the intracluster correlation.

Given that in cluster randomised trials, a small number of clusters are commonly used with a relatively large number of individuals within each cluster, presence of even a small ICC can lead to a substantial design effect (Murray, 1998a). For example, Kim et al (2006), demonstrated that even a small ICC of 0.01 resulted in the increase of the number of individuals from 675 in an individually randomised study design to 1000 individuals in a cluster randomised design when the average cluster size approached 50.

The presence of positive correlations within or between clusters in cluster randomised studies limits statistical power to detect intervention effect (Murray, Hannan and Baker, 1996). Cluster randomised trials are therefore attractive in settings in which individual randomization is difficult or impossible (Isaakidis and Ioannidis, 2003).

Although the size of the ICC in community randomised interventions is usually very small (Gulliford, Ukoumunne, and Chinn, 1999), ignoring extra variation due to intra-cluster correlation results in under estimating the variance of the intervention effect, which, in turn, leads to an inflation of type I error rates, or the likelihood of rejecting a true null hypothesis (Feng et al., 2001, Murray et al., 2008). With the ICC in cluster randomised interventions often being small and the clusters often being fairly large, significance tests to determine whether the ICC differs from 0 usually have low power. Therefore, assuming that there is no clustering effect if the test is not significant would produce misleading results and is discouraged (Donner and Klar, 1996). Thus, standard analytical methods usually yield inflated type 1 error rates, thereby leading to false associations and misleading interpretations (Cornfield, 1978; Feng et al., 2001; Murray, 2008).

3.3 Approaches to address the limited number of clusters

To counteract the inherent challenges resulting from cluster randomised trials, in particular, the extra variation and reduced degrees of freedom that stem from the limited number of clusters available for randomisation, several approaches have been suggested (Hannan et al., 1994; Feldman, McKinlay and Niknian, 1996; Murray, Hannan and Baker 1996; Varnell et al., 2001).

First, to control for the extra variation resulting from the intracluster correlation, the use of post hoc correction with a valid external estimate of the intracluster correlation coefficient has been proposed (Varnell et al., 2001; Hannan et al., 1994). However, because the ICC measures the proportion of total variation that can be attributed to being between-cluster, Hayes and Moulton (2009) caution using the ICC when similar outcome measures are obtained from different settings. Obtaining ICC estimates in the same target population for the same measures is usually not possible because no similar studies with the same measures might have been conducted in the target areas.

A second proposal is to subdivide the original clusters into small sub-clusters and conduct an analysis at the sub-cluster level, to estimate the intervention effect (Feldman, McKinlay and

Niknian, 1996; Murray, Hannan and Baker, 1996; Varnell et al. 2001). Clusters are usually subdivided to increase the number of clusters within each intervention arm so that group level variation is captured and additional degrees of freedom for hypothesis testing are achieved (Feldman et al. 1996; Murray, Hannan and Baker, 1996; Varnell et 2001).

In a simulation study by Murray, Hannan and Baker (1996) designed to test the validity of this approach, they showed that sub-cluster level analysis, in the case of two or more clusters per intervention arm, yielded type 1 error rates with small ICCs and captured most of the group level variance. However, using this approach one risks overstating the statistical significance of the effectiveness of the intervention; therefore, to ensure nominal type I and type II error rates, they recommended using the degrees of freedom based on the unit of randomisation when assessing the effectiveness of the intervention.

Varnell et al., (2001) conducted simulation studies meant to test the validity of this approach for evaluating situations with one cluster per intervention arm. They achieved a nominal type I error rate when the ICC was close to zero or when there were fewer sub-clusters per cluster. Cluster-level variance was captured by the sub-clusters. A similar study was conducted by Feldman, McKinlay and Niknian (1996) and was aimed at increasing the residual degrees of freedom: when two experimental units were applied (using the Pawtucket Heart Health Program with large social units which used one-group-per-condition randomisation design with one community in the intervention condition and one community in the control condition) the results showed that by dividing the experimental units into sub-communities per intervention arm, the statistical power improved when the number of sub-communities as well as the number of measurement time points increased.

In studies that randomise one cluster to each study arm, one could not determine the extent to which the overall cluster variation is captured by the sub-cluster variations - the cluster variation cannot be estimated because the degrees of freedom to estimate the intervention is zero (Varnell et al., 2001; Feng et al., 2001).

3.4 Implications of the limited number of clusters in the Triage Plus study

With only 3 clusters in the Triage Plus study, the statistical power to detect significant intervention effects is likely to be reduced due to the limited number of degrees of freedom for the between cluster residual error (Cornfield 1978; Koepsell et al., 1991; Feldman,

McKinlay and Niknian, 1996; Murray1998, Donner and Klar, 2000). Furthermore, any slight random variation may limit the power of the analysis to detect an intervention effect.

However, it is worth noting that in certain situations, the availability of a sufficient number of clusters for randomization might not be practical and making major modifications to the intervention to suit its evaluation may not be recommended (Hayes et al., 2000). Having more clusters in such studies not only increases the marginal cost, which may be prohibitive, but also increases logistical challenges in its implementation.

These challenges necessitate the use of methodological approaches that will address both design and analytic issues for the valid interpretation of community intervention trials. In particular, statistical methods and design conditions that will optimize statistical power and the accuracy of the estimated intervention effect are required. A number of simulation studies have compared the various design conditions that allowed the number of clusters per arm, number of subjects per cluster, intervention effect size and the intra-cluster correlation to vary (Austin, 2007; Moineddin, Matheson and Glazier, 2007; Maas and Hox, 2005; Bennett et al., 2002). Most of these studies showed improved statistical power and parameter estimates when the number of clusters increased to at least 30 clusters in total.

Considering the cost involved in cluster randomised interventions, the benefit of adopting an effective intervention of public health importance is crucial. It is imperative to assess the relative performance of different design approaches, such as the use of repeated measurements within the same cluster over time coupled with the use of statistical modelling approaches in estimating intervention effects when the design demands a small number of clusters.

Thus, the use of correct and efficient design and evaluation procedures in interventions with a limited number of clusters is paramount (Feng et al., 2001). Identifying an appropriate statistical model that addresses the correlation inherent in cluster randomised interventions and the need for adequate degrees of freedom for adequate statistical power would provide robust estimates of intervention effectiveness.

3.5 The repeated cross-sectional design for impact evaluation

Because the commonly used likelihood based approaches in evaluating intervention effectiveness require adequate degrees of freedom (i.e. adequate number of clusters) for proper analysis, utility of repeat observations within the same cluster in cluster randomised designs with limited clusters is assessed. Thus, in evaluating the Triage Plus study, the longitudinal nature of the study is used, as it provided cross-sectional data to assess the ability of multiple data collection points to provide adequate statistical efficiency in terms of the power and precision of parameter estimates. The repeated measurements design may not only induce cluster-level autocorrelation but also increase the degrees of freedom, resulting in some gain in statistical power to detect intervention effects (Donner and Klar, 2000; Hayes and Moulton, 2009). Murray et al. (1998b) observed that if the model is correctly specified, the repeated measurements in each clusters would increase the degrees of freedom for assessing the effectiveness of the intervention, especially in a random intercept model. Although power would increase in a random coefficients model because of the repeated measurements, the number of degrees of freedom would remain the same (Murray et al., 1998b).

An investigation of the utility of repeated measurement design within the same clusters has been used in a wide range of statistical scenarios to assess the efficiency of their statistical power and their precision in parameter estimates (Feldman and McKinlay, 1994; Feldman, McKinlay, and Niknian, 1996; Murray et al., 1998b; Heo and Leon, 2009; Heo, Xue, and Kim, 2013). Feldman, McKinlay and Niknian (1996) used simulation methods to investigate if increasing the number of repeated measurements within a cluster would improve the statistical power and the precision estimates of parameters in a study design involving two experimental units consisting of sub-clusters. Statistical power was shown to have improved with an increase in the number of measurement times.

Murray et al. (1998b) used repeated measurement to assess the behaviour of type I error rates under different design conditions. Using 5 clusters per arm and 5 repeated measurement times, random intercepts and random coefficients models achieved nominal type I error rates, with the random coefficients model performing well under different data generation assumptions (random intercepts or random coefficients data structure). They further showed that at 10 groups or less the empirical sandwich method to estimate standard errors for fixed effects performed poorly, but at groups 20 or above, the sandwich estimator performed better.

In assessing the effect of study duration, frequency of observation and sample size on statistical power using published data, Raudenbush and Xiao-Feng (2001) demonstrated that statistical power increased with an increase in the frequency of observations and study duration. They also noted that having more participants included in the study was more efficient than increasing the frequency of observation for boosting statistical power. However, most of these simulation studies tended to use at least 5 clusters per arm.

Adopting a repeated measurement design such as multiple cross sectional data collection over the period of the intervention, a small number of clusters (as in the Triage Plus) would achieve a similar level of statistical power compared to a pre-post cross-sectional design with more clusters (Hedeker and Gibbons, 2006). In multiple cross sectional designs within the same clusters, the sampled individuals at each time point are usually independent (especially in large sized clusters) unlike in cohort designs where the same individuals within the clusters are measured (Feldman and McKinlay, 1994). Therefore, in repeated measurements using a cross sectional design, the outcome measures taken repeatedly in the same cluster are not perfectly correlated since ICCs in community interventions are usually small (Gulliford, Ukoumunne, and Chinn, 1999).

With these two factors (independence of samples and non-perfect correlations in outcome measures), the repeated cross sectional design in the same clusters provides more independent data than in pre-post cross- sectional designs (Hedeker and Gibbons, 2006). At each cross section data collection point, a given number of participants would be included in the study. The likelihood of the same participants being included in the subsequent cross sectional studies would be minimal especially in studies involving large clusters with large populations (Hedeker and Gibbons, 2006). Thus, studies employing cross sectional designs with more data collection time points would therefore yield more individual data for improved statistical power than when just one cross section data collection is conducted in each cluster.

Furthermore, because in repeated cross sectional designs, the sampled individuals remain independent and representative at each time of data collection, the method is less affected by both secular trends and the Hawthorne effect (i.e. where the subject's behaviour may be affected by the repeated measurement e.g. in a cohort design) as pointed out by Feldman and McKinlay (1994). However, in the health facility based data collection of the Triage Plus study, the observed number of individuals accessing health services from each cluster at each time point would be correlated. The number of events (e.g. TB cases starting treatment) would

not only change with time due to either more cases being detected earlier or the number of cases detected increasing with increased exposure to the intervention, but also the interaction of these individuals accessing health services with other community members. This, therefore, necessitates the use of modelling approaches that takes into account the correlations of the outcome measures.

The number of degrees of freedom required to detect significant effects for a given power which is achieved through a repeated measurement design, is given by (Feldman, McKinlay and Niknian, 1996)

$$\text{Degrees of freedom} = I(N-1)(T-1)$$

where
I refers to the number of treatment arms,
N is the number of clusters per arm, and
T is the number of repeated measurements (i.e. cross sectional data collection) within each cluster.

Given that the Triage Plus study has 2 intervention arms, 3 clusters per arm, and the intervention was implemented for 12 months before roll out to the Delayed arm; the number of degrees of freedom increases from 4 for a pre-post design to 44 for 12 repeated measurements. Details for model formulation for the cross sectional design or repeated measurements approach are given in Chapter 4 and Chapter 5.

3.6 Measures of effect: Odds ratios, prevalence ratios and incidence rate ratios

3.6.1 Appropriate measures for the effectiveness of the intervention

There has been intensive debate regarding the appropriateness of the prevalence odds ratio (OR) and the prevalence ratio (PR) as measures of effect in cross sectional studies, following a letter by Lee and Chia (1993) proposing the use of Cox regression models for a binary outcome in a cross-sectional study in order to estimate prevalence rate ratios. The prevalence ratio (PR) in cross-sectional studies is similar to the incidence rate ratio (IRR) in longitudinal cohort studies. They noted that prevalence odds ratios lacked intelligibility, unlike prevalence ratios, which are easy to interpret and communicate. Stromberg (1994) argued that under certain stationarity assumptions on the study population and given that the mean duration of

the disease among exposed and non-exposed subjects is known, a prevalence odds ratio can be converted into an incidence rate ratio (IRR).

However, Axelson et al. (1994) noted that such stationarity assumptions may not always work and concluded that prevalence ratios may be more interpretable and intelligible measures of effect than prevalence odds ratios. Osborn and Cattaruzza (1995), while agreeing that in some cases prevalence odds ratios may be incomprehensible compared to prevalence ratios, argued that the effect measure to be used in the study depends on the problem at hand. However, McNutt et al. (2003) and Thompson et al. (2004) as cited in Laupacis et al. (1988) noted that in public health, one is interested in estimating prevalence ratios or risk ratios and not prevalence odds ratios.

Not only do prevalence odds ratios overestimate prevalence ratios, especially when a disease outcome is common (say more than 10%), but also controlling for confounding in prevalence odds ratio results in an estimate that is even further away from the prevalence ratio than the unadjusted prevalence odds ratio (Axelson et al., 1994). However, if the disease is rare, the difference between the prevalence odds ratio and prevalence ratio is numerically negligible (Axelson et al., 1994, Skov et al. 1998). Therefore, when testing rare diseases, the prevalence ratio can be estimated by using logistic regression. However, it worth noting that in cross-sectional studies, high prevalence diseases are usually targeted, and the estimation of the prevalence ratio using logistic regression would be a poor approximation (Skov et al., 1998).

Historically, in cohort studies the Mantel-Haenszel procedure (Mantel and Haenszel, 1959) was used to estimate adjusted risk ratios if the covariates were all categorical but the procedure failed to work when some of the covariates were continuous (Deddens and Petersen 2008). Zhang and Yu (1998) proposed a formula to derive relative risks from adjusted odds ratios obtained through logistic regression as:

$$RR = \frac{OR}{(1 - P_0) + (P_0 x OR)}$$

where P_0 indicates the incidence of the outcome in the non-exposed group, and (1- P_0) in the exposed group; OR is the odds ratio, and RR is the risk ratio.

However, McNutt et al. (2003) noted that this conversion of odds ratios to relative risks would result in biased estimates if confounding is present.

These discussions led to investigators recommending alternative statistical models for directly estimating prevalence ratios and confidence intervals from cross-sectional or longitudinal studies (Wacholder 1986; Breslow 1974; Lee, 1994; Skov et al. 1998; McNutt et al., 2003; Deddens et al. 2003; Lumley et al., 2006; Savu et al., 2010). Although the use of prevalence ratios in cross-sectional studies is widely recommended, the only drawback is its lack of a practical model from which to conduct statistical analysis. Statistical methods that can directly estimate and test prevalence ratios, while adjusting for several confounders, are needed. In this section, we, therefore, review the different statistical methods for estimating intervention effectiveness using prevalence ratios as measures of intervention effect.

3.6.2 Estimation of prevalence ratios and incidence rate ratios

Estimations of prevalence ratios (PR) in cross sectional studies or incidence rate ratios (IRR) in longitudinal cohort studies arise from different statistical approaches in varying forms. Most notably, approaches based on Cox regression (Breslow, 1974; Lee, 1994), Poisson regression (McNutt et al., 2003; Petersen and Deddens, 2008), log-binomial regression (Wacholder, 1986; Zocchetti, 1995; Skov et al., 1998), and the COPY method (Deddens et al., 2003) have been used to estimate prevalence ratios. These models fall under the generalised linear models framework, a class of fixed effects models which provides a unified procedure for fitting models based on a likelihood framework (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989). These models accommodate non-normal response variables such as count or binomial data. The formulation of generalised linear models involves the specification of a linear predictor, link function and variance structure for the response outcome y_i for the individuals (McCullagh and Nelder, 1989).

The linear predictor in the GLM framework is given by:

$$\eta_i = x_i\beta \qquad (3.3)$$

where $i = 1, \ldots, n$ indexes the individual observations, x_i is a vector of explanatory variables, β is a vector of corresponding regression coefficients.

The link function $g(.)$ relates the mean value μ_i of response variable y_i [ie $\mu_i = E(y_i)$] to the linear predictor η_i

$$g(\mu_i) = \eta_i \qquad (3.4)$$

This is then concluded by specifying the variance structure of the mean μ_i, which depends on the distribution of the response variable y_i. The link function and variance of μ_i is assumed to be within the exponential family (McCullagh and Nelder, 1989).

Below are the various statistical approaches for estimating prevalence ratios (or risk ratios) assuming no clustering of individuals. These are given with the aim of identifying models to be pursued in Chapter 4 in the context of a generalised linear mixed model framework, in which clustering will be taken into account.

3.6.2.1 Cox regression model

The Cox proportional hazards regression model was designed for estimating conditional hazard ratios and survival rates in subjects with different risk periods and has been adapted to estimate prevalence ratios in cross-sectional studies (Breslow, 1974; Lee, 1994).

Given $h_0(t)$ is the baseline hazard at time t, then the individual hazard $h_i(t)$ at time t is defined as

$$h_i(t) = h_0(t)e^{x_i\beta} \qquad (3.5)$$

or, equivalently,

$$logh_i(t) = logh_0(t) + x_i\beta \qquad (3.6)$$

in a log scale, with subscript i =1,...,n indexing individual observations, x_i is a vector of explanatory variables and β is a vector of corresponding regression coefficients.

By assuming a constant risk period for all individuals, the Cox regression model estimates a cumulative incidence ratio (Breslow, 1974), which can then be adapted to estimate the

prevalence ratio in cross-sectional studies (Lee 1994). Skov et al. (1998) and Barros and Hirakata (2003) showed that point estimates derived from the Cox model were close to the true parameters with negligible bias. However, the standard errors were too large when the disease is a common one, due to the underlying distribution of the outcome following a Poisson distribution (Skov et al., 1998; Barros and Hirakata, 2003). But in cross sectional studies, prevalence data follows a binomial distribution; therefore, the variance of the coefficients derived from the Cox model are overestimated, which results in wider confidence intervals for the estimated parameters (Barros and Hirakata, 2003).

With a constant risk period and by handling ties properly, Breslow (1974) showed that the partial likelihood estimates and the estimated standard errors of the non-intercept parameters from the Cox proportional hazards regression model are similar to those from a Poisson regression model. Barros and Hirakata (2003) pointed out that the Poisson model has advantages over the Cox proportional hazards regression model in that the Cox model does not have an intercept and therefore cannot estimate probabilities. The Poisson regression model, in this instance, is preferred.

3.6.2.2 Poisson regression model

The Poisson regression model was developed for studies of rare conditions where subjects may be followed up for different time periods. However, if the exposure time for all subjects is the same, then the exponentiated coefficients in the log link function in the Poisson regression model are interpreted as incidence rate ratios (Rabe-Hesketh and Skrondal, 2012, McNutt et al., 2003).

Poisson models yield less precise estimates, especially when the outcome is common, since Poisson errors overestimate binomial errors (McNutt et al. 2003). The overestimation of binomial errors occurs because the variance for the binomial distribution reaches its maximum when the prevalence is 0.5, unlike that of the Poisson distribution which increases progressively (Coutinho et al. 2008). In correcting for the error misspecification in Poisson models, Zou (2004) proposed using a sandwich estimator to obtain the corrected variance, which improves variance estimation considerably. Barros and Hirakata (2003) also assessed Poisson regression with scale parameter adjustments with deviance and chi-square (i.e. the estimated Poisson variance is multiplied by an estimate of the overdispersion obtained by dividing the deviance or the chi-square of the estimated model by their respective residual

degrees of freedom). They concluded that the Cox and Poisson regression models with robust variance estimation performed well. However, the Cox and Poisson regression methods may give rise to predicted probabilities above unity (Deddens and Petersen, 2004; Petersen and Deddens, 2008), especially when the estimates are on the boundaries of the parameter space.

The Poisson model

Given that a Poisson distribution is assumed for the response variable y_i (i= 1,...,n), representing the number of events detected during an interval of time e.g. per month, and the average number of events are denoted by parameter μ, then the likelihood of response outcome y_i given parameter μ (Kuhn et al. 1994) is given by

$$p(y = \mu) = \frac{e^{-\mu}\mu^{y}}{y!}, y = 0, 1, \dots, ; \ \mu > 0 \qquad (3.7)$$

and

$$E(y_i) = var(y_i) = \mu.$$

Since in the Poisson model relative risk or prevalence ratio estimates are often desired instead of the mean, the expectation of the response outcome should be modelled as a function of the relative risk λ given the independent variables x_i and corresponding regression parameters β_i. Given vector Y is the number of events from a population size of$N$ thenin vector notation

$$E(Y) = N\,\lambda(X, \beta) \qquad (3.8)$$

To obtain the estimates of the β parameters, the function λ is modelled as an exponential function of independent covariates,

$$\lambda(X, \beta) = exp(X\,\beta) = \exp(x_i\beta)$$

or, equivalently,

$$\log(\mu) = x_i\beta\,, \ i = 1, 2\,..n. \qquad (3.9)$$

where μ is the mean number of events recorded in n subjects. To obtain the rates of incidence of an event, the Poisson regression model (3.9) is then given by

$$\log(\mu) = \log(N) + \beta_0 + \beta_1 x_{1i} + \ldots\ldots \qquad (3.10)$$

log (N) is used as an offset with a constant coefficient of 1 for each observation, with x_i and β as defined above. Given that the x_1 is a variable denoting the individual's exposure to the intervention, then using this approach, β_1 denotes the intervention effect on the log relative risk scale for follow up studies or logged prevalence ratios in cross sectional studies for the early compared to the delayed groups while adjusting for all other covariates in the model. The significance of the intervention effect is then assessed by testing the hypothesis $Ho: \beta_1 = 0$

3.6.2.3 Log-binomial models

Given that estimates derived from Cox/Poisson regression models are less precise, Wacholder (1986) and Zocchetti (1995) proposed to model the prevalence ratio with a logarithmic link function in the context of binomial variability, and Skov et al. (1998) called this a log-binomial model which estimates prevalence ratios directly (Wacholder, 1986; Zocchetti et al. 1995; Skov et al., 1998; Barros and Hirakata, 2003; Lumley et al., 2006; Petersen and Deddens, 2008).

If n represents independent observations assumed to come from a binomial distribution with response variable y_i taking values $1/0$ denoting the presence/ absence of the event in a subject, then the conditional probability of response variable y_i: $p(y_i = 1 \mid x_i)$ given $x_i$vector of covariates for each subject under the log-binomial regression model is given by

$$p_i = p(y_i = 1|x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \cdots)$$

and the probability of the absence of the event for each subject is:

$$p_i = p(y_i = 0|x_i) = 1 - \exp(\beta_0 + \beta_1 x_{i1} + \cdots)$$

Therefore, in the context of the independence of the observations, the likelihood function for each subject under binomial sampling distribution for data vector Y is given by (Savu, Liu and Yasui, 2010):

$$L(Y;\beta) = \prod_{i=1}^{n} (e^{\beta_0 + \beta_1 x_{i1} + \cdots})^{y_i} (1 - e^{\beta_0 + \beta_1 x_{i1} + \cdots})^{n_i - y_i} \qquad (3.11)$$

or, equivalently, the log likelihood

$$Log\ L(Y;\ \beta) = \sum_{i=1}^{n} [y_i x_i \beta + (1 - y_i)\log(1 - \exp(x_i \beta))] \qquad (3.12)$$

with n denoting the number of independent observations and $y_i$the outcome responses ($i = 1,2 \ldots, n$) indexes individual observations, the sampling distribution for data y_i given a vector of explanatory variables X, with corresponding regression coefficients β under the binomial model is given by

$$p(Y|\beta, X) = \binom{n}{y_i} (e^{\beta_0 + \beta_1 x_{i1} + \cdots})^{y_i} (1 - e^{\beta_0 + \beta_1 x_{i1} + \cdots})^{n_i - y_i} \qquad (3.13)$$

and the parameter space is defined as

$$\beta_0 + \beta_1 x_{i1} + \cdots \leq 0, i = 1,2,..,n \qquad (3.14)$$

where $\beta_0 + \beta_1 x_{i1} + \cdots$is the log-binomial model's linear predictor, and $\beta_0, \beta_1, \beta_2$ ... is a vector of model parameters to be estimated from the data (Lumley et al., 2006; Deddens and Petersen, 2008). Given x_1 is a binary exposure of interest, exp(β_1) is the adjusted prevalence ratio comparing the subjects exposed to the intervention with the non-exposed group while adjusting for the effects of other explanatory variables.

Since $p(y_i = 1|x_i)$ is a probability, the model requires constraints on vector β so that fitted probabilities are always between 0 and 1; therefore, the parameter space for a log-binomial model is a convex cone (Deddens and Petersen 2008; Savu, Liu and Yasui, 2010). Thus, in log-binomial models, the maximum likelihood estimates are derived by maximizing the parameters' space $\beta_0 + \beta_1 x_{i1} + \cdots \leq 0$ for each observation i (Wacholder, 1986; Skov et al., 1998; Deddens and Petersen, 2008), and as such, the maximum likelihood estimator will only be asymptotically Normal if the true β is within the parameter space (Lumley et al., 2006).

However, this restriction on the parameter space results in log-binomial models encountering convergence problems if the maximum likelihood estimates of the prevalence ratio (PR) parameters occur at the boundary of the restricted parameter space and the predicted probabilities are equal to 1 (Wacholder, 1986; Deddens et al., 2003). Hence, the iterative

procedures that maximize likelihood fail to converge, resulting in impermissible probabilities (Zocchetti et al., 1995; Nijem et al., 2005).

Convergence problems

Given the convergence problems of log-binomial models, Deddens et al. (2003) proposed the COPY method to approximate the maximum likelihood estimates of the prevalence ratios when log-binomial models fail to converge. The method involves expanding the original dataset to obtain copies large enough that the maximum likelihood estimates from the modified data approximate the maximum likelihood estimates in the original data. The resulting data set contains c-1 copies of the original data and 1 copy with dependent variable values interchanged. As noted by Lumley et al. (2006), this is equivalent to creating one copy of the original data set with weight w=(c-1/c)=0.999 and the other copy of the original data set with values interchanged with weight 1-w=0.001.

In this instance, a weighted log-binomial regression maximising the likelihood is performed (Savu, Liu and Yasui, 2010):

$$L_w(Y;\beta) = \prod_{i=1}^{n} (e^{\beta_0 + \beta_1 x_{i1} + \cdots})^{wy_i + (1-w)(1-y_i)} (1 - e^{\beta_0 + \beta_1 x_{i1} + \cdots})^{w(1-y_i) + (1-w)y_i}$$

or equivalently, the log likelihood

$$log\ L_w(Y,\beta) = \sum_{i=1}^{n} [wy_i + (1-w)(1-yi)]x_i\beta + [w(1-y_i) + (1-w)y_i]\log(1 - \exp(x_i\beta) \quad (3.15)$$

This approach limits the need for adjusting the standard errors and the maximum likelihood estimates (MLE) from the modified data and allows direct estimation of likelihood ratio confidence intervals (Deddens and Petersen, 2008). The approach ensures that the MLE in the modified data is always within the restricted parameter space, yields estimates that are close to the true parameters and is superior to the Poisson and Cox proportional hazards models (Deddens et al. 2003 & 2008; Petersen and Deddens, 2006). A simulation study conducted by Savu, Liu and Yasui, (2010) confirmed the robustness of the COPY method against link function misspecification.

Yu and Wang (2008) proposed using the SAS Nonlinear Programming (NLP) procedure for estimating prevalence ratios. This procedure results in estimates being within an acceptable range as it explicitly imposes the constraints and, therefore, can be used as an alternative to the COPY method (Yu and Wang, 2008). When they compared the NLP procedure with PROC GENMOD using the log-binomial and COPY methods, they showed that the convergence rates, bias and mean square error estimates were comparable between the NLP procedure and the COPY method. However, the two methods share the same problem of data modification since the NLP procedure first uses the COPY method to generate initial values for the PROC NLP procedure and then uses PROC NLP to obtain the maximum likelihood estimates (Yu and Wang, 2008). Hence, the COPY method is preferred when the log-binomial model fails to converge.

Skov et al. (1998) proposed a method based on logistic regression and robust estimation of standard errors, called 'GEE-logistic'. According to Skov and colleagues, by duplicating every case in the data set to a non-case observation so that the new data set has three groups of cases as follows: cases, original non-cases and the new non-cases that have just been duplicated (e.g. if there were 50 cases and 100 non cases then the 50 cases would be duplicated and the final data set will be 50 cases, 100 original non-cases and 50 non-cases that have been duplicated making 150 non-cases), Schouten et al. (1993) proposed fitting the log-binomial model parameters using logistic regression such that

$$p(y_i = 1|x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \cdots)}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \cdots)} \quad (3.16)$$

where $i = 1, \ldots, n$ indexes the individual observations, $x_i$is a vector of covariates, β are the corresponding regression coefficients to be estimated.

However, Skov et al. (1998) argued that due to correlations within the new data set, as a result of the case duplications, and the fact that the standard logistic model cannot maximize over the same parameter space that a log-binomial model uses to obtain maximum likelihood estimates, the generalized estimation equation approach must be used (Liang and Zeger, 1986). Although the GEE-logistic models produced point estimates that were close to the true parameter (with negligible bias) and had correct type I error rates, the prevalence rates

produced were greater than one; therefore, Skov et al. (1998) preferred the log-binomial model for estimation of prevalence ratios.

3.7 Analysis of clustered data using single level modelling approaches

The generalised linear models framework presented in section 3.6.2 assumes that the observations of the response variable y_i are independent of each other, which is not the case in clustered data because the within cluster correlation (intracluster correlation) is not zero. Ignoring the clustering of the data leads to underestimating the variance and thereby overstating the statistical significance (Omar and Thompson, 2000).

Different techniques have been used in these single level models to address the underestimation of standard errors.

Fixed effects approach

In the fixed effect approach, dummy variables are included in the individual level model to account for the between cluster differences and allow intercepts and slopes to vary across clusters or covariates that are believed to influence individual responses. However, the regression approach used, in this case the Analysis of Co-variance (ANCOVA), is not feasible with many clusters, because we are adjusting for baseline measures which may result in too many parameters to be estimated. Furthermore, it does not permit an assessment of the association between cluster-level variables between clusters and how they influence the relationship between lower units and the response variable (Kreft and Leeuw, 1998). In addition, this approach may not fully account for cluster effects, as covariates at the cluster level may not be fully measured.

Marginal effects approach

In the marginal effects approach, design effects are used to adjust for standard errors and the construction of cluster level marginal models are used to address the dependencies between clusters (Liang and Zeger, 1986). In this approach, the GLM framework presented in section 3.6.2 is extended to model correlated data by introducing second-order variance components into the generalised linear model's estimating equation – a modelling approach also referred to as generalised estimation equations (GEE) which models the marginal or population averaged estimates (Liang and Zeger, 1986). Although the estimates of the coefficients and their standard errors in marginal models are fairly robust to misspecification of the correlation

structure (Liang and Zeger, 1986), the approach fails to investigate the nature of the between-group variability (Breslow and Clayton, 1993). If the overall effect is of interest, then it is more appropriate to model the averaged marginal probabilities of success rates over all clusters being studied (Hu et al., 1998). However, due to the limited number of clusters usually available in cluster randomised trials (Hayes et al., 2000), reliable asymptotic properties for robust standard errors may not be achieved leading to unreliable type I errors (Murray et al., 1998a; Omar and Thompson, 2000).

Thus, single level models are not suitable for cluster randomised trials because they do not model the variation in the outcome variable between clusters. The inherent correlations of observations resulting from cluster randomised trials or repeated measurements are better accounted for by using generalised linear mixed models (GLMM) or multilevel models (Breslow and Clayton, 1993) presented in Chapter 4.

3.8 Conclusion

The Chapter has provided an overview of cluster randomised designs and the challenges inherent in such designs. Furthermore, the Chapter has also given an overview of the approaches that have been used to mitigate the challenges, especially when only a limited number of clusters are available for randomisation. The statistical efficiency implications of the Triage Plus study have been highlighted and ways to avert the statistical challenges have been proposed. In particular, the use of multiple measurements within the clusters over time during the intervention period was suggested. The utility of this approach for analysing studies with cluster numbers as low as 3 is assessed in a series of simulations presented in Chapter 5.

Furthermore, the Chapter has reviewed estimation approaches for prevalence (or risk ratios) in the evaluation of interventions. Although prevalence ratios are estimated directly when using log-binomial regression, it encounters convergence problems when the parameter estimates are at the boundaries of the parameter space. Instead, Poisson regressions models can be adapted to model prevalence ratios for cross sectional studies by ensuring a constant risk for all individuals or clusters (i.e. measurements in each cluster need to be taken at equal time intervals) and incidence rate ratios in longitudinal cohort designs. Because the Triage Plus study primarily collected monthly count data during the study period, Poisson regression models with incidence rate ratios as measures for assessing the effectiveness of the

intervention (with some adjustments for over-dispersion by using robust standard errors) are the ideal choice for analysing the data from the Triage Plus study described in Chapter 6.

CHAPTER 4

GENERALISED LINEAR MIXED MODELS

4.1 Introduction

This thesis focuses on the assessment of the effectiveness of an integrated TB and HIV community intervention (Triage Plus) that used a cluster randomised design. The clusters were randomised into Early and Delayed intervention arms. Data on TB and HIV treatment initiations and testing uptake were repeatedly collected in each of the clusters at baseline and during the intervention period. Given that repeated observations from each cluster over time are more correlated than across clusters, this type of data is referred to as hierarchical or multilevel data. In this multilevel form, the repeated measurements (level 1) are nested within the clusters (level 2).

As the data derived from this structure are correlated, the standard models, such as generalised linear models - a class of fixed effects models (McCullagh and Nelder, 1989) presented in section 3.6 in Chapter 3, cannot be used to analyse the data. To account for the correlations of these observations, generalised linear mixed model (GLMM) approaches have been proposed which have random effects incorporated in the linear predictor (Breslow and Clayton, 1993). The GLMM is an extension of the generalised linear model theory (McCullagh and Nelder, 1989). Generalised linear mixed models are also known as multilevel models (Goldstein and Rasbash, 1996) and generalised linear random effects models (Stiratelli, Laird and Ware, 1984).

These multilevel models accommodate the extra-variability inherent in longitudinally repeated measures or cluster designs (Breslow and Clayton, 1993). The influence of cluster level explanatory variables is easily assessed in multilevel modelling while controlling for differences in explanatory variables in lower levels. Thus, the use of generalised linear mixed modelling allows for the consideration of both epidemiological reasons, enabling the quantification of the need for contextual factors in assessing intervention effectiveness, and statistical reasons, improving estimation. Since the effects of the intervention may vary between clusters, the use of multilevel modelling resolves this problem by allowing an analysis of the effects that vary by cluster.

In this Chapter, the statistical models for analysing hierarchically structured data, specifically the repeated measurement data from a cluster randomised trial, are discussed. Furthermore, estimation approaches for these statistical models are also discussed. The aim of this Chapter is therefore to continue the review of statistical methods briefly presented in Chapter 3 with the aim of identifying the statistical models that can then be pursued in the simulations in Chapter 5 and the actual analysis of the Triage Plus study data in Chapter 6.

This Chapter is organised as follows: section 4.2 presents a generalised linear mixed model framework in the context of a repeated measurement design within the clusters. Section 4.2 also presents the GLMM representation of Poisson and Negative binomial models for the analysis of count data and log-binomial models for hierarchically structured data that are binomially distributed. Section 4.3 discusses estimation approaches for incidence rate ratios in multilevel models using likelihood approaches. Section 4.4 concludes the chapter.

4.2 Generalised linear mixed model framework

4.2.1 The generalised linear mixed model

The generalised linear mixed model is an extension of the GLM theory where random effects are incorporated into the linear predictor (3.3) in Chapter 3. The resulting generalised linear mixed model includes the usual fixed effects for the regression coefficients and the random effects that model the correlations in the data.

In data derived from a cluster randomised design denoted by (y_{ij}, x_{ij}) , y_{ij} represents observations of the response measured in an individual $i = 1, \dots, n$ in cluster $j = 1, \dots, N$, and x_{ij} is a vector of explanatory variables. For the repeated measurement data design within each cluster according to notation by Rabe-Hesketh et al., (2005), $i = 1, \dots, T$ indexes the individual measurement time-points and $j = 1, \dots, N$ indexes clusters in which repeated measurements are taken. Because the Triage Plus study adopted a repeated measurement design, the repeated measurement notation given by Rabe-Hesketh et al., (2005) has been used in this thesis. However, for the simulation studies carried in Chapter 5, the notation i $(i = 1, \dots, n)$ indexes the number of simulations in a given design condition investigated. Similar notation has also been used in Chapter 7 under the Future Work section (section 7.8) to denote the iterations $(i = 1, \dots, n)$.

The generalised linear mixed model is composed of three parts:

(i) A distribution component for the response variable.

In the distribution component of the generalised linear mixed model, the conditional distribution of response y_{ij}, given the covariates x_{ij} and the random intercepts u_j for cluster j, is assumed to follow a distribution from the exponential family of distributions (McCullagh and Nelder, 1989) with probability density function given as

$$f(y_{ij}|x_{ij}, \theta_{ij}, u_{j,}\phi) = \exp\left\{\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right\} \qquad (4.1)$$

where:
functions b(.) and c(.)are known functions and are specific to the distribution of the exponential family (i.e. the same form for all y_{ij}),
θ_{ij} is the natural parameter and is the natural or canonical parameter for y_{ij},
ϕ is scale parameter,and
$a(\phi)$ is a fixed dispersion parameter, which is 1 for Poisson and binomially distributed data (Breslow, 1996).

(ii) The structural assumptions specifying the conditional expectation of y_{ij}, given the covariates and random effects, is given by

$$\eta_{ij} = E(y_{ij}|x_{ij}, u_{j,}) = x_{ij}\beta + u_j \qquad (4.2)$$

where $i = 1, \dots, T$ indexes the repeated measurement time points, $j = 1, \dots, N$ indexes clusters, x_{ij} is a vector of explanatory variables, β is a vector of corresponding regression coefficients, and u_j is a random intercept for cluster j.

(iii) A monotonic differentiable link function $g(.)$ converts the expected value mean $\mu_{ij}$of the response variable y_{ij}(ie $E(y_{ij}|x_{ij}, u_j) = \mu_{ij}$) to the linear predictor η_{ij}

$$g(\mu_{ij}) = \eta_{ij} \qquad (4.3)$$

The inverse of the link function $h = g^{-1}$ is also called the 'response function', so that $\mu_{ij} = h(\eta_{ij})$.

By transformation, the unit cumulant function $b(.)$ relates the natural parameter θ_{ij} to the mean μ_{ij} as well as to the variance of y_{ij} as

$$\mu_{ij} = E(y_{ij}|x_{ij}, u_j) = \exp(\theta_{ij})$$

so that

$$\theta_{ij} = \theta(\mu_{ij}) \text{and } Var(y_{ij}|x_{ij}, u_j) = \sigma^2(\mu_{ij}) = \frac{\phi}{a_{ij}} v(\mu_{ij}).$$

where σ^2 is the variance of the responses y_{ij} given the covariates and random effects.

Given $g(\mu_{ij}) = \theta(\mu_{ij})$ then $g(.)$ is the natural link function that yields $\theta_{ij} = \eta_{ij}$.
The parameters are determined as:

$$\mu_{ij} = b'(\theta_{ij}) = \frac{\partial b(\theta_{ij})}{\partial \theta} = \exp(\theta_{ij}),$$

$$\sigma^2(\mu_{ij}) = \phi v(\mu_{ij})$$

$$v(\mu_{ij}) = b''(\theta_{ij}) = \frac{\partial^2 b(\theta_{ij})}{\partial \theta^2}$$

Thus, $var(\mu_{ij})$ has two components: one that depends on a scale parameterϕ and external factors, and the second that relates the variance to the mean $v(\mu_{ij})$, which determines the form of the distribution and is called the 'unit variance function'.

This implies that for a continuous response,$v(\mu_{ij}) = 1$ indicates a Normal distribution with the mean μ_{ij} and the scale parameter $\phi = \sigma^2$. In a discrete response, $v(\mu_{ij}) = \mu_{ij}$ implies a Poisson distribution, and $v(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$ implies a binomial distribution, with $a(\phi) = \phi = 1$ in both Poisson and binomial distributions (McCullagh and Nelder, 1989). Specifying a

variance structure in GLM or GLMM indicates that a distributional form is specified (Clark and Thayer, 2004).

In addition to the distribution assumption (4.1), the structural assumption (4.2) and link function (4.3) of the cluster level random effects in a generalised linear mixed model are assumed to be independent and Normally distributed with mean $E(u_j) = 0$ and with an unknown covariance matrix $cov(u_j) = D(\theta)$.

4.2.2 Generalised linear mixed model representation in a repeated measurements design

This section describes the generalised linear mixed model under the repeated measurements design. This is an extension of the previous equations under the generalised linear modelling framework that assumed the responses were independent (see section 3.6)

Given y_{ij} is a response measured at time point $i$in cluster j, x_{ij} is a covariate vector with fixed effects (such as intervention status, measurement occasion variable, interaction terms between intervention and time, and cluster or individual level covariates) and $W_{ij}$is a covariate vector associated with random effects u_j (eg. random intercepts and random coefficients) the responses y_{ij} are conditionally independent with mean $E(y_{ij}|\, u_j) = \mu_{ij}$ and variance $var(y_{ij}|u_j) = \phi a_{ij}^{-1} v(\mu_{ij})$, where a_{ij} is a dispersion parameter,$v(.)$ and ϕ are the unit variance functionand scale parameter respectively. According to Breslow and Clayton (1993), the general form of a generalised linear mixed model with repeated measurements is given by

$$g(\mu_{ij}) = x_{ij}\beta + W_{ij}u_j \qquad (4.4)$$

The expectation of the conditional distribution of the response variable, given the random effects, is given by

$$E(y_{ij}|u_j, x_{ij}, \beta) = \mu_{ij} = g^{-1}(x_{ij}\beta + W_{ij}u_j) \qquad (4.5)$$

The random effects u_j are assumed to be independent and Normally distributed as $N(0, D(\theta))$ i.e. mean $E(u_j) = 0$, and unknown covariance matrix given as $cov(u_j) = D(\theta)$ and depends on an unknown vector θ of variance components. The differentiable link function $g(.)$ is a

known monotonic function that relates the mean μ_{ij} to the linear predictor $\eta_{ij} = x_{ij}\beta + W_{ij}u_j$, with coefficient β remaining constant over measurement time-points i. To account for the correlation of the repeated measurements within clusters as well as the overdispersion in the Poisson distributed data, the random effects u_{ij} for level one (i.e. repeated measurements) can be included in the model (see section 4.2.3.1 model (4.9)).

The GLMM in (4.4) can be extended in two ways. First, if $W_{ij} = 1$, then the model is a random intercepts model that allows each cluster or group to have its own random intercept but same slope over time and is given as

$$\eta_{ij} = x_{ij}\beta + u_j$$
$$u_j \sim N(0, \sigma^2) \quad (4.6)$$

If $W_{ij} = (1, x_{ij})$, then the GLMM has both random intercept and random slope parameters and is given as

$$\eta_{ij} = x_{ij}\beta + u_j + x_{ij}u_j$$
$$u_j \sim N(0, \sigma^2) \quad (4.7)$$

If the coefficients of fixed effects β are allowed to vary over the measurement time, then the GLMM becomes a generalised varying-coefficient mixed model (Hastie and Tibshirani, 1993; Lu and Zhang, 2009). If the effect modifier is time, then the model in (4.4) is a time-varying coefficient model (Lee and Shaddick, (2007), indicating a special case of the varying-coefficient models of Hastie and Tibshirani (1993).

4.2.3 Poisson and negative binomial models

For the Triage Plus study, Poisson and or negative binomial regression models with mixed effects were the natural choice for analysis because the number of new TB and HIV cases accessing treatment and testing services over the intervention period were recorded each month.

4.2.3.1 The Poisson models

In modelling the Triage Plus study data, there are two possible model structures (random intercepts (see model (4.6) or random coefficients (see model (4.7)) that needed to be fitted to

the data to assess which of these models better fit the data. By using model selection criteria presented in section 4.4 for likelihood based model selection a correct model structure would be identified. Below is the description of the two model structures for the mixed effects Poisson models. In addition, to the two model structures, marginal effects models would also be used to obtain population averaged estimates of intervention effectiveness.

The random intercepts Poisson model

Random intercepts Poisson models (see model (4.6) have been used to model correlations resulting from clustering effects and repeated observations in the same clusters, with the assumption that the effect of the intervention is the same across all clusters (Breslow and Clayton, 1993; Rabe-Hesketh and Skrondal, 2008). Therefore, in modelling longitudinal data, the levels are defined by the repeated measurement occasions nested within clusters (Snijders, 1996).

Given y_{ij} are the observed events at time point $i$in cluster j, which are assumed to be Poisson distributed with mean μ_{ij}, and $n_{ij}$is the number of people at risk in cluster j at time i, then generalised linear random intercepts model for data that follow a Poisson distribution is given as:

$$y_{ij}|x_{ij}, u_j \sim Poisson\left(\mu_{ij}\right)$$

$$log\ \mu_{ij} = \log\left(n_{ij}\right) + \beta_0 + \beta_1 x_{ij} + u_j \quad (4.8)$$

$$u_j \sim N(0, \sigma^2)$$

where $x_{ij}$is a vector of covariates including interaction terms, u_j is arandom interceptsfor cluster j, β_0 is the mean number of events in the control arm, β_1 is the regression coefficient for an intervention variable contrasting intervention clusters (treat=1) and control clusters (treat=0),and $\log(n_{ij})$ is an offset. The random effects (i.e. the random intercepts)capture the effects of unknown or unmeasured cluster level variables and are assumed to be independent across clusters. Thus, $(\beta_0 + u_j)$ in the model is the random intercept for cluster j.

To account for overdispersion and within cluster variations, a random intercept at level one (Breslow and Clayton, 1993) is included and the model in (4.8) is extended to

$$y_{ij}|x_{ij}, u_j, u_{ij} \sim Poisson\left(\mu_{ij}\right)$$

$$log\ \mu_{ij} = \log(n_{ij}) + \beta_0 + \beta_1 x_{ij} + u_{ij} + u_j \qquad (4.9)$$

$$u_j \sim N(0, \sigma_j^2)$$

$$u_{ij} \sim N(0, \sigma_{ij}^2)$$

where β_0 and β_1 are the intercept and intervention regression coefficients respectively, $u_j$arethe between cluster random effects with variance σ_j^2, u_{ij} are level one random effects with variance σ_{ij}^2. The subscripts j and ij are to distinguish cluster level and level one variance parameters respectively. The level one random effects u_{ij} models both the extra-Poisson variability within clusters and level one variability due to repeated measurements, which are assumed to be independently Normally distributed (Gibbons et al., 2008). In random intercepts models, the cluster level error component remains constant across repeated measurements. However, the error resulting from repeated measurements varies between clusters and time points. The resulting intra-cluster correlation (ICC) between two repeated measurements within a cluster is given in section 3.2.2.2 as

$$ICC = \rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \qquad (4.10)$$

where σ_b^2 and $\sigma_w^2$are the between and within cluster variance componentsrespectively. In a multilevel model for Poisson distributed data, σ_b^2 is the variance of the random intercept denoted as σ_j^2 in model (4.9) (Moineddin et al, 2007) and the within cluster variance σ_w^2 is given as (Clark and Bachmann, 2009; Nakagawa and Schielzeth, 2010):

$$\sigma_w^2 = \ln\left(\frac{1}{exp(\beta_0)} + 1\right)$$

Using formula (4.10), the ICC in Poisson distributed data with overdispersion without covariates is therefore estimated as:

$$\frac{\sigma_b^2}{\sigma_b^2 + \phi.\ln\left(\frac{1}{exp(\beta_0)} + 1\right)} \qquad (4.11)$$

where σ_b^2 is the random intercepts variance, ϕ is the overdispersion scale parameter derived by dividing the Pearson chi-square by their degrees of freedom, and $exp(\beta_0)$ is the mean count of the outcome measure when all covariates are zero. Because of the inclusion of the intercept β_0 in the calculation of ICC for Poisson distributed data, the estimated ICC from Poisson data depends on the intercept β_0 since a log transformation used in count data is not variance stabilising (Nakagawa and Schielzeth, 2010).

If the ICC is close to 1, then this reflects increased within-cluster variation, necessitating the need for using random effects models. This allows us to successfully quantify and remove the corresponding variability. Very low values of ICC close to zero, indicate low within-cluster variation and imply that the use of random effects models will not help improve the model.

The random coefficients Poisson model

In random intercepts models, the dependence of repeated observations within a cluster is well accounted for but it assumes that the effect of the actual intervention is constant in all clusters. However, to allow the effect of the intervention to vary randomly between clusters (see model in (4.7)), the random intercepts model (4.9) is extended to

$$y_{ij}|x_{ij}, u_j \sim Poisson\left(\mu_{ij}\right)$$

$$log\ \mu_{ij} = \log(n_{ij}) + \beta_0 + \beta_1 x_{1j} + u_{ij} + u_j x_{1j} + \beta_2 x_{2ij} \quad (4.12)$$

where:

i indexes the repeated measurements within cluster j, with random effects u_{ij} and u_j as defined in model (4.9)(Guo and Zhao, 2000 page 445; Thompson, Warn and Turner, 2004 page 395), and the effect $\beta_1 + u_j$ of the intervention (x_{1j}) is allowed to vary over the clusters (Rabe-Hesketh and Skrondal, 2008). β_0 and β_1 are regression coefficients as defined in model (4.8). The explanatory variables x_{1j} and x_{2ij} represent the intervention and other explanatory variables measured at cluster level and at each time point within clusters respectively.The two random effects have bivariate Normal distribution with mean zero and unknown covariance matrix D given by

$$D = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}, \sigma_{21}^2 = \sigma_{12}^2$$

where σ_{11}^2 is the variance of the random intercept, σ_{22}^2 is the variance of the random coefficients, σ_{21}^2 and σ_{12}^2 are the covariances between the random intercepts and random coefficients. In this context, σ_{11}^2 corresponds to the variance σ_b^2 in model (4.10) and σ_j^2 in model (4.9). Using the notation given in the covariance matrix D above, then the ICC term in (4.11) controlling for all covariates can be formulated as

$$\frac{\sigma_{11}^2}{\sigma_{11}^2 + \phi . \ln\left(\frac{1}{exp(X\beta)} + 1\right)} \qquad (4.13)$$

where X and β are the vectors of covariates and coefficients respectively.

According to Rabe-Hesketh and Skrondal (2008), the correlation between the random intercept and slope is given by

$$\rho_{21} = \frac{\sigma_{21}^2}{\sqrt{\sigma_{11}^2 * \sigma_{22}^2}} \qquad (4.14)$$

where ρ_{21} is the correlation of the random intercept and slope.

If there is no correlation between the two random effects, using the random coefficient model may not improve the model fit; when they are correlated, however, it may be necessary to fit the random coefficient model to obtain improved estimates since clusters with larger mean event counts per unit time tend to have larger slopes when the correlation is positive (Rabe-Hesketh and Skrondal, 2008 page 160).

4.2.3.2 The Negative Binomial Model

Although the Poisson distribution is commonly used in modelling discrete events that are Poisson distributed, its restrictive assumption of equality of the mean and variance fails in certain situations, thereby leading to biased standard errors. This results in over-dispersion: the conditional variance exceeds the conditional mean. To capture the extra variability associated with discrete count data, Poisson-gamma and negative binomial models are used to account for the greater than Poisson variability in estimating incidence rate ratios (Hausman, Hall and Griliches, 1984; Chin and Quddus, 2003; Ntzoufras, 2008 page 315). The two models are related in that the marginal likelihood of the negative binomial model (i.e. the marginal negative binomial model) is derived by integrating out the random effects of the Poisson-gamma model and the two models give similar results when fitted to the count data

with over-dispersion (Ntzoufras, 2008). Since the negative binomial models have a mixture of both Poisson and gamma distributions, the between-cluster variations are accounted for when a negative binomial model is fitted to the count data though variations resulting from time are not considered (Chin and Quddus, 2003). To adequately allow for cluster-specific variations and variations resulting from the repeated measures of count data, the random effect negative binomial model can be used (Chin and Quddus, 2003). This is achieved by introducing in the model a random effects term into the relationship between the expected numbers of events at a given time and the covariates (Chin and Quddus, 2003).

Because the negative binomial and Poisson regression models have the same mean structure, the negative binomial is considered to be a generalisation of the Poisson model, and it has an extra dispersion parameter which is greater than zero for modelling the over-dispersion in count data. Usually the estimates of the coefficients under the Poisson and negative binomial models are similar but have different estimates of variance, with the negative binomial model having a wider sampling distribution leading to wider confidence intervals for the parameter estimates (Rabe-Hesketh & Skrondal, 2008). Since analysing mixed effects Poisson regression models with robust standard errors provides similar results to using negative binomial models (Rabe-Hesketh & Skrondal, 2012, page 712), the mixed effects Poisson regression model is preferred because it is relatively simpler to implement in Stata software.

4.2.3.3 Marginal effects Poisson model

Unlike in multilevel or hierarchical modelling where the dependence in the repeated measurements or clustered data is explicitly modelled by using random effects, marginal or population-averaged modelling usually estimates the model parameters as if the data was not clustered, and the clustering is only taken into account when estimating standard errors.

The GLM framework used in non-correlated data presented in section 3.6.2 and the marginal modelling presented in section 3.7 were then used to model correlated data by introducing second-order variance components into the generalised linear model estimating equation – a modelling approach also referred to as generalised estimation equations (GEE) (Liang and Zeger, 1986). The GEE estimates marginal effects with corresponding robust sandwich-based standard errors while taking into account the dependence of the repeated measurements within clusters (Rabe-Hesketh and Skrondal, 2012). Fitting the generalised linear Poisson model (3.10) with robust standard errors when estimating marginal effects results in

comparable results (Rabe-Hesketh and Skrondal, 2012). In this thesis, therefore, the GLM with robust standard errors was used in marginal effects modelling. The regression coefficients and their standard errors in marginal models are fairly robust to misspecification of the correlation structure (Liang and Zeger, 1986). However, marginal modelling fails to investigate the nature of the between-group variability (Breslow and Clayton, 1993) which is correctly achieved when the random effects modelling is used.

Because of this, model formulation in marginal modelling is slightly different from multilevel modelling. However, because of the similarity in expectations between marginal and conditional random effects models in Poisson models, the exponentiated regression coefficient $\exp(\beta)$ has both marginal and cluster-specific interpretations as the incidence rate ratio (Rabe-Hesketh and Skrondal, 2008 page 381). To improve the estimates in marginal effects models, robust standard errors are used to estimate all regression coefficients (Zou, 2004).

4.3 Likelihood based estimation methods for incidence rate ratios

To make statistical inferences using the GLMM (4.4) involves the estimation of fixed effects and random effects variance components. However, the marginal likelihood of the response obtained by integrating out the random effects generally does not have a closed-form (Breslow and Clayton, 1993; Turtz et al., 2003). Several approximation methods for estimating the marginal likelihood have been used in literature. The most commonly used methods based on the likelihood estimation approach are numerical integration (Rabe-Hesketh, Skrondal and Pickles, 2005) or penalised quasi-likelihood (PQL) (Breslow and Clayton, 1993). More recently, Bayesian estimation using Markov chain Monte Carlo methods have been used (Zhao et al., 2006). This section we present these various approximation methods in the estimation of incidence rate ratios for longitudinal data or prevalence ratios for cross-sectional data.

4.3.1 Numerical integration

Numerical integration methods based on Gauss- Hermite quadrature have been used to obtain an approximation for the marginal likelihood (Hartzel et al., 2001). This type of numerical integration requires more quadrature points to approximate the normal distribution (Rabe-Hesketh, Skrondal and Pickles, 2002). Numerical integration of the generalised linear mixed models using adaptive quadrature is computationally efficient, requiring fewer quadrature points to achieve accurate parameter estimation (Rabe-Hesketh,

Skrondal and Pickles, 2002). It performs better than the Gauss- Hermite quadrature in a variety of situations, including large cluster sizes and high intracluster correlations (Rabe-Hesketh, Skrondal and Pickles, 2005).

4.3.1.1 Maximum Likelihood Estimation of GLMM using adaptive and spherical quadrature

Given a responsey_{ij}measured at time point i in cluster $j$is conditionally independent given the random intercepts u_j and u_{ij} for clusters and repeated measurement time points respectively, x_{ij} is a vector of explanatory variables, the joint probability of the responses in a cluster is given by

$$\Pr(y_{ij}|x_{ij}, u_j) = \prod_{i=1}^{T} pr(\, y_{ij}|x_{ij}, u_j) \qquad (4.15)$$

with y_{ij}, x_{ij} and u_j as defined above. However, for simplification purposes, only the cluster level random effects are considered in the above equation as well as the following equations for estimating the maximum likelihood of GLMM. The marginal joint probability of the responses y_{ij} in a given cluster and integrating out the random intercept is given by

$$\Pr(y_{ij}\,|x_{ij}) = \int \phi(u_j; 0, \sigma^2) \prod_{i=1}^{T} pr(\, y_{ij}|x_{ij}, u_j)\, \mathrm{d}u_j \qquad (4.16)$$

where

$\phi(u_j; 0, \sigma^2)$ is the normal density of random intercept u_j with mean 0 and varianceσ^2 (Rabe-Hesketh, Skrondal and Pickles, 2002 & 2012).

$\prod_{i=1}^{T} pr(\, y_{ij}|x_{ij}, u_j)$is the conditional likelihood contribution (i.e. joint probability) of y_{ij} for cluster j, given the random effects and covariates.

The marginal likelihood for all clusters is

$$L(\beta, \sigma^2) = \prod_{j=1}^{N} \mathrm{pr}(y_{ij}\,|x_{ij}) \qquad (4.17)$$

The log-likelihood of β is given by

$$logL(\beta) = \prod_{j=1}^{N} log\ pr(\ y_{ij}|x_{ij}) \qquad (4.18)$$

Using a Newton-Raphson algorithm, the vector of model coefficients β and random effects variance parameterσ^2 are estimated by finding the values of the parameters that maximise the likelihood (Rabe-Hesketh, Skrondal and Pickles, 2012).

When using the Gauss-Hermite quadrature, the random effects $u_j$in (4.16) are substituted with Gauss-Hermite quadrature locations e_r and the standardised normal density $\phi(v_j = u_j/\sigma)$with weights w_r for the rth location($r = 1, \dots, R$). That is, the integration is done over v_j instead of u_j. The approximation of the marginal probability (4.15) at level 2 becomes

$$\Pr(y_{ij}\,|x_{ij}) = \int \phi(v_j) \prod_{i=1}^{T} pr(\ y_{ij}|x_{ij}, v_j)\,\mathrm{d}v_j \approx \sum_{r=1}^{R} \Pr \prod_{\mathrm{i}=1}^{\mathrm{T}} (y_{ij}|x_{ij}, e_r) w_r \quad (4.19)$$

The method works well if $\Pr(y_{ij}\,|x_{ij})$ is a polynomial of degree up to $2R - 1$.

The procedure involves approximating the Normal distribution of the random effects with a discrete distribution with a given number of integration points. By taking into account the properties of the integrand $\phi(v_j) \prod_{i=1}^{T} pr(\ y_{ij}|x_{ij}, v_j)\,\mathrm{d}v_j$ coupled with large sample sizes, the ratio of the integrand to the posterior density is well approximated by a low-degree polynomial (Rabe-Hesketh, Skrondal and Pickles, 2005).

$$\Pr(y_{ij}\,|x_{ij}) = \int \phi(u_j; \mu_j, \sigma^2) \left(\frac{\phi(v_j) \prod_{i=1}^{T} pr(\ y_{ij}|x_{ij}, v_j)}{\phi(u_j; \mu_j, \sigma^2)} \right) \mathrm{d}v_j \qquad (4.20)$$

where $\mu_j$is the mean of the posterior density. By changing v_j to $u_j = (v_j - \mu_j)/\sigma$ as variable for integration and applying standard quadrature rules (Rabe-Hesketh, Skrondal and Pickles, 2002, 2005), the model (4.19) gives

$$\Pr(y_{ij}|x_{ij}) = \sum_{r=1}^{R} \pi_{\,\mathrm{jr}} \prod_{\mathrm{i}=1}^{\mathrm{T}} (y_{ij}|x_{ij}, e_r) w_r \quad (4.21)$$

where $\pi_{\,\mathrm{jr}} = \sqrt{2\,\pi}\ \sigma \exp\left(\frac{{e_r}^2}{2}\right) \phi(\mu_j + u_j e_r) w_r$.

Rabe-Hesketh, Skrondal and Pickles (2002) generalised the adaptive quadrature to multilevel modelling including multilevel random coefficients models.

4.3.1.2 Estimation of random effects and regression coefficients

Turtz and Kauermann (2003) estimated random effects u_j using the posterior probability given data $y_{ij}$and vector of regression coefficientsβ and vector of natural parameters θ as defined above in model (4.1) as

$$p(u_j|y_{ij},\beta,\theta) = \frac{p(y_{ij}|u_j,\beta)p(u_j,\theta)}{\int p(y_{ij}|u_j,\beta)p(u_j,\theta)du_j} \qquad (4.22)$$

with $p(u_j,\theta)$ being a density of the mixing normal distribution $N(0,\sigma^2)$ and the posterior mean of u_j (Tutz and Kauermann, 2003; Rabe-Hesketh and Skrondal, 2008) defined as

$$u_j = \int u_j p(u_j|y_{ij}\,;x_{ij})du_j \qquad (4.23)$$

which are then estimated using adaptive spherical quadrature. Once the random effects $u_j$are known, the regression coefficients β are then obtained by maximising the log-likelihood (4.18).

4.3.2 Penalised quasi-likelihood

Marginal quasi-likelihood (MQL) has been used as an approximation procedure for likelihood (Goldstein & Rasbash, 1996). Breslow and Clayton (1993) proposed to approximate the likelihood using Laplace's method for integral approximation commonly referred to as penalised quasi-likelihood (PQL). PQL is a modification of MQL that involves updating the random effects u_j with their current estimates ${u_j}^{(m)}$ at iteration m. The PQL is an approximate inference technique for estimating generalised linear mixed models. It uses weighted least-squares algorithms in the estimation of parameters in the mean together with likelihood equations from an approximating normal model in estimating variance components (Dean, Ugarte and Militino, 2004). Both the MQL and PQL rely on the Taylor expansion to achieve the approximation.

4.3.2.1 Estimation of generalised linear mixed models using PQL

Given a response y_{ij} measured at time point i in cluster j that is conditionally independent given a vector of random effects u that are assumed to be independent and Normally distributed as $N\big(0, D(\theta)\big)$ with unknown covariance matrix $D = D(\theta)$ which depends on an unknown vector θ of variance components and ϕ is the dispersion parameter, the integrated quasi-likelihood function (Breslow and Clayton, 1993; Dean, Ugarte and Militino, 2004) used to estimate (β, θ) is

$$|D|^{-1/2} \int \exp\left\{-\frac{1}{2\phi}\sum_{ij} d_{ij}\big(y_{ij}, \mu_{ij}\big) - \frac{1}{2}\mu_i D^{-1} u\right\} \mathrm{du} \qquad (4.24)$$

where

$$d_{ij}\big(y_{ij}, \mu_{ij}\big) = -2\int_{y_{ij}}^{\mu_{ij}} \frac{y_{ij} - \mu_{ij}}{v(\mu_{ij})} \mathrm{d}\mu_{ij}$$

is a conditional deviance measure of fit for fixed effects β given random effects u_j and covariance matrix $cov(u) = D$ with $v(.)$ and μ_{ij} being the variance function of the mean given the random effects and the mean at time i and cluster j. The integrated quasi-likelihood in (4.24) is not tractable, and, therefore, an alternative is to use Laplace approximation to give a penalised quasi-likelihood as

$$PQL(\beta, u) = -\frac{1}{2\phi}\sum_{ij} d_{ij}\big(y_{ij}, \mu_{ij}\big) - \frac{1}{2}\mu_i D^{-1} u \qquad (4.25)$$

Breslow and Clayton (1993) then replaced the PQL with its quadratic expansion of the exponent term of the integrand in the quasi-likelihood function (4.24) to find its maximum point before integration. This is to ensure that the term in the exponent of (4.24) has continuous second order partial derivatives with respect to random effects u_j.

To ensure that an iterative weighted least-squares algorithm is employed to estimate fixed effects in the normal mixed effects model, Breslow and Clayton defined a vector Y with $y_{ij} = \eta_{ij} + \big(y_{ij} - \mu_{ij}\big)g'(\mu_{ij})$ where $g'\big(\mu_{ij}\big) = \frac{1}{\mu_{ij}}$.

The associated normal theory model is given as

$$Y = x_{ij}\beta + W_{ij}u_j + \varepsilon \qquad (4.26)$$

where $\varepsilon \sim N(0, Z^{-1})$, $Z = diag\{var(y_{ij}|u)[g'(\mu_{ij})]^2\}^{-1}$, $u \sim N(0, D)$, with ε and u_j areindependent to each other.

According to Breslow and Clayton (1993) using matrix terminology, the estimate of fixed effects β is

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \qquad (4.27)$$

and its estimated asymptotic variance is given by

$$var(\hat{\beta}) = (X^T V^{-1} X)^{-1} \qquad (4.28)$$

Given the data, the vector of random effects u are estimated as empirical Bayes estimates of their posterior mean given the dataas

$$\hat{u} = DW_{ij}^T V^{-1}(Y - X\hat{\beta}) \qquad (4.29)$$

where $V = Z^{-1} + W_{ij} D W_{ij}^T$, X is the vector of covariates, the superscript T denotes the time at which measurement of the response variable is taken.

Because the standard errors of the estimated $\hat{u}$ do not take into account the additional variability that arises when estimating the unknown vector θ of variance components, the standard errors, therefore, may be underestimated (Breslow and Clayton, 1993; Dean, Ugarte and Militino, 2004).

4.3.2.2 Estimation of variance components:

In estimating variance components, Breslow and Clayton (1993) used restricted maximum likelihood (REML) estimation equations given as

$$\frac{1}{2}\left[(Y - X\hat{\beta})^T V^{-1} \frac{\partial V}{\partial \theta_k} V^{-1}(Y - X\hat{\beta}) - tr\left(P \frac{\partial V}{\partial \theta_k}\right)\right] = 0, k = 1,2,3, \ldots (4.30)$$

where

$P = V^{-\frac{1}{2}}(I - H)V^{-\frac{1}{2}}$,

$H = V^{-\frac{1}{2}}X(X^TV^{-1}X)^{-1}X^TV^{-1/2}$ is a projection *hat* matrix.

The corresponding asymptotic variance of $\hat{\theta}$ is given by

$$\mathcal{I}_{ks}^{-1} = \frac{1}{2}tr\left(P\frac{\partial V}{\partial \theta_k}P\frac{\partial V}{\partial \theta_s}\right), k, s = 1,2,3, \ldots (4.31).$$

where $\mathcal{I}$ is the Fisher information matrix, the subscripts k and s are the k, s-th element of the matrix taking values 1,2,3, ...

In deriving the penalized quasi-likelihood and the corresponding equations shown above, Breslow and Clayton (1993) involved several adhoc adjustments and approximations which performed well because the equations are the REML equations under the normal theory linear model. These approximations are likely to improve as the outcome responses become normally distributed.

Estimation algorithm

The PQL was estimated using the iterative generalised least squares algorithm (Goldstein, 1991); therefore, it consists of the following steps (Dean, Ugarte and Militino, 2004):

1. Initial estimates of the parameters are given ($\beta_0^{(0)}$, $\beta_1^{(0)}$, $\sigma^{2(0)}$.
2. The PQL then proceeds by first solving for $(\hat{\beta}, \hat{u})$ iteratively using (4.28) and (4.30) for β and u respectively while the variance components are fixed.
3. Then, the variance components are updated.
4. Steps 2 and 3 are repeated until the convergence of both estimates for fixed effects and variance components.

Inferences for parameters are then made by comparing the standardised parameter estimates to the asymptotic standard normal distribution.

4.4 Likelihood based model selection

In real life, there are many candidate models available for analysis of the given data. One is, therefore, faced with challenges in choosing the most suitable model among the candidate models. A common approach used in model selection is a stepwise strategy that involves comparing different model candidates on how well they predict the data when guided by tests based on approximate asymptotic P values or information criteria. The following model selection criteria were used in this dissertation during the analysis of the Triage Plus study presented in Chapter 6.

4.4.1 Likelihood ratio tests

In the likelihood framework, likelihood ratio tests have been used extensively for model selection because of their ability to test for multiple parameters and are often applied with large data sets (Berkhof and Snijders, 2001). The likelihood ratio test assesses the goodness-of-fit between two models by comparing a relatively more complex model to a simpler model when fitting a given dataset.

For instance, it can be used to assess whether the random coefficients model fits the data better than the random intercepts model. In terms of the variable selection process, then the more complex model is the one with more variables and or interaction terms are included in the model and the simpler model is the one with fewer variables or without interaction terms.

Determination of the likelihood ratio test is based on the difference in the log-likelihoods between the complex and simpler model. Thus, the likelihood ratio test statistic is defined as:

$$LR = 2(L_1 - L_0) \qquad (4.32)$$

with L_0 and L_1 being the maximised log-likelihood value estimates for the null (simpler model) and the alternative (complex or saturated) model respectively. The likelihood ratio test is asymptotically chi-square distributed with degrees of freedom equal to the number of additional parameters in the complex model (Berkhof and Snijders, 2001). However, since the test statistic does not have a chi-square distribution when testing variance parameters, and since variance parameters are nonnegative thereby making the p values conservative , the asymptotically correct p value of the likelihood ratio chi-square test is divided by 2 (Snijders and Bosker, 1999; Berkhof and Snijders, 2001).

Correct parameter estimation in random effects modelling depends on whether the data are better fitted using the random intercepts or random coefficients models. To select an appropriate model for the random effects modelling, the following sequential procedure is used.

(i) Select an appropriate model structure (random intercepts or random coefficients) using likelihood ratio tests of the null hypothesis, that we have variance parameters of zero, at the 5% level of significance, and use model fit criteria after including all possible covariates of interest into the model

(ii) Trim the model covariates through model variable selection and only include variables remaining statistically significant in subsequent model fittings.

4.4.2 Akaike Information Criterion and Bayesian Information Criterion

In likelihood based model selection, the Akaike Information Criterion (AIC) (Akaike, 1981) and Bayesian Information Criterion (BIC) (Schwarz, 1978) have also been popular in comparing maximum likelihood models; therefore, they have been used in model selection. Developed by Hirotsugu Akaike in 1971, the AIC was developed as an estimator of the Kullback-Leibler information function and is a measure of the loss of information when an incorrect model is fitted to the data. The AIC is, therefore, a measure of the goodness of fit of an estimated statistical model (Akaike, 1981; Hoeting et al., 2006). It is not necessarily a test of the model, as in hypothesis testing, but a test between models. The AIC is given by

$$AIC = -2 * ln(likelihood) + 2 * p \qquad (4.33)$$

Unlike the AIC, which uses a constant 2 to weight the number of parameters p in the model, the Bayesian Information Criterion uses the log of the number of observations to weight the parameters.

The BIC is given by

$$BIC = -2 * ln(likelihood) + ln(N) * p \qquad (4.34)$$

where p is the number of parameters fitted in the model, thereby providing a trade-off between model fit and complexity (number of parameters). The best model for a given data

set is determined by ranking all the potential models according to their AIC or BIC, and the one with the lowest AIC or BIC is taken to be the best model.

4.5 Conclusion

In cluster randomised trials with a repeated measurement design within the same clusters, the use of multilevel modelling averts the challenges inherent in such designs and ably accommodates the correlations resulting from the cluster design as well as the repeated measurements within the same clusters. This is achieved by including random effects in the linear predictor.

However, the use of multilevel modelling has brought its own challenges, such as the need for sophisticated statistical methods such as the numerical methods for nonlinear mixed models because of the intractability of the marginal likelihood (Breslow and Clayton, 1993, Hedeker and Gibbons, 2006). However, the development of statistical programmes has eased implementation of such methods.

In the analysis of discrete data, such as counts or binomial data, the common likelihood based estimation methods presented in this chapter include numerical integration by adaptive quadrature and penalised quasi-likelihood. Although the PQL estimation is faster than adaptive quadrature, as it is implemented in the iterative generalised least squares algorithm, it has two shortcomings. First, in certain situations the approximations derived from PQL methods may underestimate both the random and fixed effects (Breslow and Lin, 1995), especially in binary response variables with small cluster sizes and high intracluster correlation (Rodriguez and Goldman, 2001). Second, as pointed out by Rabe-Hesketh, Skrondal and Pickles (2005), PQL estimation does not involve likelihood and, therefore, likelihood based inference, such as likelihood ratio tests and likelihood based confidence intervals, cannot be determined. Thus, in this dissertation, numerical integration based on adaptive quadrature is adopted and is implemented using a Stata programme.

CHAPTER 5:

STATISTICAL POWER ISSUES WITH CLUSTER RANDOMISED TRIALS

5.1 Introduction

Chapter 3 presented a review of cluster randomised designs and the challenges encountered when cluster randomised designs are used. In particular, challenges in achieving adequate statistical power due to a limited number of clusters available for randomisation, were highlighted. Different approaches for mitigating these challenges were briefly reviewed including use of a repeated measurement design. The chapter further reviewed statistical methods and appropriate measures for assessing the effectiveness of the intervention. These statistical methods were further reviewed and presented in Chapter 4.

This chapter focuses on simulation methods for assessing the utility of the repeated measurement design in cluster randomised trials with limited number of clusters. In particular, the simulation studies aim to assess statistical efficiencies for different design conditions in terms of statistical power and accuracy in parameter estimates when determining the effectiveness of complex interventions with a limited number of clusters. In this way, the circumstances under which each of the statistical methods would be more robust in detecting significant intervention effects or providing accurate estimates of intervention effects, can be identified.

The chapter is organised as follows: section 5.2 briefly reviews the issues concerning the number of clusters and participants per cluster; section 5.3 presents the details of the simulation study as well as the specific objectives of the simulations; section 5.4 presents the results of the simulations and the conclusions made.

5.2 Number of clusters and numbers of participants per cluster: a literature review

Data obtained from cluster randomised interventions are usually correlated; therefore, multilevel modelling approaches must be used for analysis (Breslow and Clayton, 1993; Murray, 1998a; Gelman and Hill, 2007). These multilevel models, also known as hierarchical models, are ideal for assessing how contextual factors affect the outcome of interest (O'Campo, 2003). The maximum likelihood estimation methods commonly used in multilevel modelling are asymptotic, and these methods assume that the sample sizes will be large enough for

accurate estimates. However, in randomised studies that use cluster as the unit for treatment allocation, a common drawback is that the number of clusters available is usually limited, as evidenced by several studies (Murray, Varnell and Blitstein, 2004; Pals, Wiegand and Murray, 2011; Simwaka et al., 2012). Adding more clusters may in itself raise distinct challenges, ranging from the increased costs incurred by the addition of clusters to problems of contamination (Torgerson, 2001; Borm et al., 2005) if the clusters are too small. Furthermore, it may not be possible to increase the number of clusters due to there being only a limited number of clusters in the study area, or the outcome of interest may have low prevalence or incidence necessitating larger clusters to achieve an adequate sample size in each cluster. However, it is an inescapable fact that having only a limited number of clusters raises questions about the statistical power as well as the accuracy of the effect size estimates and their standard errors obtained in such studies (Moineddin, Matheson and Glazier, 2007).

To counteract the problem of a limited number of clusters, researchers have tended to opt for using a small number of clusters with a sufficiently large number of individuals sampled per cluster: the aim is to compensate for the limited number of available clusters as pointed out by Feldman, McKinlay, and Niknian (1996). However, simulation studies using multilevel modelling have suggested that the number of clusters per intervention arm is more important than the sample sizes within the clusters for achieving adequate statistical power and the correct estimation of parameters and their variance components (Mass and Hox, 2004; Moineddin, Matheson and Glazier, 2007). Although fixed effects regression coefficients are not biased, the estimated standard errors and variance components are biased, with non-coverage rates above the nominal value in the presence of only a small number of clusters (Bennett, Parpia and Cousens, 2002; Moineddin, Matheson and Glazier, 2007), thereby affecting type I error rates. Other simulation studies have indicated that a minimum of 30 clusters may be needed for adequate statistical power and valid parameter estimates (Moineddin, Matheson and Glazier, 2007; Maas and Hox, 2005).

Despite the clear relationship between number of clusters used and statistical efficiency, cluster randomised interventions with a small number of clusters continue to be reported in the literature (Donner et al., 1990; Simpson, Klar, and Donner, 1995; Varnell et al., 2004; Murray et al., 2008; Pals, Wiegand and Murray, 2011) because of the challenges involved in having more clusters (as was the case for the Triage Plus study with 3 clusters per arm described in Chapter 2). Statistical power and robustness problems have been reported in intervention studies with a small number of clusters (Bennett, Parpia and Cousens, 2002;

Murray et al., 1998b). Most of the studies that have assessed statistical power have involved one or two time intervals for the outcome measurement and at least 5 clusters per intervention arm. However, little is known about cluster randomised studies with a panel data structure with only 3 clusters per arm and with low incidence disease conditions (e.g. TB). In such longitudinal studies, the analytical levels are defined by the measurement occasions at a lower level nested within clusters at a higher level (Murray et al., 1998b; Heo and Leon, 2009; Heo et al., 2013). There is limited research into the minimum number of clusters, measurement occasions, detectable effect sizes and disease prevalence / incidence necessary to achieve optimal statistical efficiency in terms of power and accuracy of parameter estimates in the context of multilevel modelling. There is a clear need, therefore, for further investigations into the required design conditions necessary for achieving optimal statistical efficiency to guide the statistical analysis of such studies when the number of clusters is small.

In this dissertation, simulation studies have been used to assess statistical efficiencies under different design conditions such as a varied number of clusters, measurement occasions, effect sizes and prevalence/ incidence of the outcomes of interest. Although simulation studies have inherent limitations, such as being sensitive to assumptions about the variability of the outcome and the statistical Normality of random effects (Arnold et al., 2011), the results found in these simulations may inform the analytical approach for actual data derived from studies with limited number of clusters (and in particular the analysis of the Triage Plus study data).

5.3 Simulation study

Because the findings of the simulation studies were intended primarily to inform which analytical approach should be used for the actual data derived from the Triage Plus study, and as the primary outcome in this study was a count variable, the Poisson generalised linear mixed models presented in Chapter 4 were used to assess statistical efficiencies under different design conditions. In particular, the relative efficiencies of Poisson GLMM regression methods were evaluated using likelihood estimation approaches in the estimation of intervention effect size. To run the simulations, the baseline dataset was used as a 'training dataset' to generate the random sample datasets to obtain the desired parameter specifications such as the ICC and mean response per group (see Appendix 8.5 for the simulation code).

The Triage Plus study described in Chapter 2 used only 3 clusters per arm, and the design involved the collection of repeated observations in each cluster over time. The simulation studies reported below generated time-series cross-sectional panel data over a 12 month

intervention period in line with the Triage Plus study. Thus, the maximum units of analysis used were 3 x 12 = 36 cluster-months per intervention arm (varying from 3x2 to 3x12 cluster-months for 2 to 12 repeated measurement occasions respectively). These simulations were used to investigate the likely statistical power and efficiencies of the Triage Plus study design and then used in the actual analysis of the Triage Plus study to assess the effectiveness of community engagement in improving access to TB and HIV services.

The simulations were then extended to involve other likely scenarios for comparison. Since simulation studies sometimes use design conditions that may not be applicable in field study situations, they do not only allow for estimation of accuracy measures of the parameter estimates of interest compared to the known true estimates; but they may also provide a better understanding of the intended statistical approach to be used (Hodgson and Burke, 2000). The simulations implemented in this Chapter provided a broader understanding of the analytical approaches needed for the analysis of the data derived from the Triage Plus study.

5.3.1 Rationale and objectives

Given the limited number of clusters used in the Triage Plus study, the objective of the current study was to use simulation methods to investigate the relative statistical efficiencies of different cluster configurations, in terms of statistical power, and the accuracy of parameter estimates and variance components. Additionally, the statistical efficiencies of different statistical methods for the estimation of relative risks or prevalence ratios were investigated under different design conditions for cluster randomised trials.

Specifically, the objectives of the simulation studies were to:

i. determine the optimal design conditions of cluster randomised trials with only a limited number of clusters available, such as Triage Plus, as well as the statistical power and accuracy of parameter estimates for fixed effects parameters and variance components (with their corresponding standard errors).

ii. determine the effect of disease prevalence, under different statistical configurations, on the statistical power and accuracy of parameter estimates in cluster randomised trials with a limited number of clusters.

5.3.2 Simulation procedures

Generation of data

To correspond closely with the Triage Plus study, simulations were constructed with repeated measurements in each cluster using desired parameter specifications. The distributions and parameters specified included the mean response per group and the desired effect size. Since the simulations represented a cluster randomised trial design with repeated measurements, covariate correlation structures, which summarised a correlation between clusters, and repeated measurements within clusters were also specified (Arnold et al., 2011). More specifically, the parameter specifications in the statistical models were estimated from the real data obtained from the Triage Plus study and used to generate the random sample data sets by fitting the model (5.3) without covariates to obtain the model intercept and the standard deviation of the random intercept parameter $\sqrt{\sigma_{11}^2}$. To induce overdispersed data within the clusters, an overdispersion scale parameter ϕ obtained from the baseline data was specified during the generation of the data (see section on generation of Poisson outcome data below).

The random outcome y_{ij} representing the outcome of interest for the i^{th} measurement occasionin j^{th}cluster was generated in the context of multilevel modelling. The covariates considered in the simulation were the intervention and the secular trends. A separate indicator variable, *treat*, was used to indicate whether the cluster received the intervention (treat=1) or control (treat=0) and for secular trend, denoted by variable *time,* for the repeated measurements (time=2, 3, 4, 5, 6, 9, 12). The number of clusters was equally distributed between intervention and control arms during the simulation. For the reproducibility of the simulated data, the response variables were generated using the same *seed* number in Stata by including a Stata command '*seed(3321)*' as option in the simulation syntax.

To generate the random outcome, a generalised linear mixed model (GLMM) framework was used to account for the correlations resulting from the randomised cluster design as well as the repeated measurements from each cluster.

The general form of the GLMM framework used is given in (4.4) in section 4.2.2 and repeated here as

$$g(\mu_{ij}) = x_{ij}\beta + W_{ij}u_j \qquad (5.1)$$

where μ_{ij} is the expectation of response outcome y_{ij} given a covariate vector with fixed effects x_{ij} and a covariate vector W_{ij} with random effects u_j (in our case we used a random intercept where $W_{ij} = 1$). Use of the random intercepts model reduced the simulation time

needed; more importantly, this approach was adequate to answer the simulation questions investigated, making random coefficients modelling less necessary in our case. The function $g(.)$ is a differentiable link function that relates the expectation μ_{ij} with the linear predictor of the model. The random effects u_j are Normally distributed with mean zero and unknown variance-covariance structure.

Generation of Poisson outcome data

To be in line with the Triage Plus cross-sectional panel structure, conditional Poisson distributed outcome data were generated in Stata version 11.2. Given y_{ij} is an outcome response at repeated measurement occasion i $(i = 2, \dots, T)$ indexing time points within clusters, $j(j = 1, \dots, N)$ indexing clusters, which is conditionally Poisson and distributed with mean μ_{ij}, the statistical model used in the simulations is given in model (4.9) in Chapter 4 section 4.2.3.1 repeated here as

$$y_{ij}|x_{ij}, u_j, u_{ij} \sim Poisson\left(\mu_{ij}\right)$$

$$log\ \mu_{ij} = \log\left(n_{ij}\right) + \beta_0 + \beta_1 x_{ij} + u_{ij} + u_j \qquad (5.2)$$

$$u_j \sim N(0, \sigma_j^2)$$

$$u_{ij} \sim N(0, \sigma_{ij}^2)$$

where:

y_{ij} denotes the number of events measured at time point$i$ in cluster$j$ and is conditionally Poisson distributed. The time points were varied from 2 to 12.

$\mu_{ij}$is the mean count for cluster j at time point i,

$\log\left(n_{ij}\right)$ is an offset with a constant coefficient of 1,

β_0 is the mean number of events in the control arm at baseline time point,

β_1 is the estimated intervention effect comparing intervention clusters and control clusters,

x_{ij} is a covariate for intervention status (i.e. treated or not),

u_j is a random effect for individual cluster j with a variance σ_j^2 to model between cluster variation

u_{ij} is a random effect with variance σ_{ij}^2 at time point i and cluster j to model both extra-Poisson and variability within clusters due to repeated measurements. The subscripts j and ij are used to distinguish cluster level and level one variance parameters respectively.

The random effects are assumed to be Normally distributed with a mean of zero and a known standard deviation.

The intracluster correlations (ICC)resulting from these random effects affect statistical power as well as the accuracy of the parameter estimates (Moineddin, Matheson and Glazier, 2007; Parker, Evangelou and Eaton, 2005). The ICC was then estimated using the formula (4.13) repeated here as:

$$\frac{\sigma_{11}^2}{\sigma_{11}^2 + \phi.\ln\left(\frac{1}{exp(\beta_0)} + 1\right)}$$

where σ_{11}^2 is the random intercepts variance, ϕ is the overdispersion scale parameter derived by dividing the Pearson chi-square by their degrees of freedom, and $exp(\beta_0)$ is the mean count of the outcome measure.

For the simulation studies, four different ICC values were considered in total. The first two values, 0.00154 and 0.081, were the smallest and largest values found in the baseline data from the Triage Plus study; in order to test a very wide, but also realistic, range of ICC values in the simulations, two additional, much larger ICC values (0.321 and 0.699) were taken from other similar studies (see below). The estimated standard deviation of the random intercepts parameter $\sqrt{\sigma_{11}^2}$, overdispersion scale parameter ϕ and the model intercept β_0 were 0.005, 0.5 and 2.71 respectively for the ICC of 0.00154. For the ICC of 0.081, the estimated standard deviation of the random intercepts parameter $\sqrt{\sigma_{11}^2}$, overdispersion scale parameter ϕ and the model intercept were 0.1405, 3.493 and 2.71 respectively. The Stata analysis output below shows the parameters that were used in the calculation of baseline ICCs using the formula in (4.13) (in this case an ICC of 0.081).

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
clustid: Identity				
sd(_cons)	**.1405399**	**.0458503**	**.0741482**	**.2663783**

LR test vs. Poisson regression: chibar2(01) = **34.31** Prob>=chibar2 = **0.0000**

```
. scalar b0=log(15)

. di b0
2.7080502
```

To investigate the effect of varying intracluster correlation coefficients on statistical power, further simulations were run using the baseline ICC of 0.081 compared to other ICC scenarios of 0.321 and 0.699. The standard deviations of the random intercept parameters $\sqrt{\sigma_{11}^2}$ for the ICCs (0.081, 0.321 and 0.699) were 0.1405, 0.140, and 1.459 respectively. The overdispersion scale parameter ϕ and the within cluster variances were determined as described above and the corresponding ICCs were determined using formula (4.13). As pointed out in section 2.5 in Chapter 2, the ICC of 0.081 was based on the baseline data from the Triage Plus study. The ICC of 0.321 was based on a cluster randomised trial to assess the effect of an education intervention on TB case detection and primary care of respiratory diseases in South Africa which was reanalysed by Clark and Bachmann, 2009. The ICC of 0.699 was based on the high ICCs implemented by Stryhn et al (2006). This high ICC was applied here to assess power issues when there are very high ICCs.

Thus, to generate the outcome variable y_{ij}, the fixed effects parameters and the standard deviations for the random intercepts parameter $\sqrt{\sigma_{11}^2}$, and an estimate of the overdispersion scale parameter ϕ were specified. The overdispersion parameter was specified in the model in order to generate an overdispersed count data. The intervention effect size β_1 was specified based on pre-determined intervention effects of 20%.

To estimate the remaining 2 parameters (β_0, and σ_{11}), baseline datasets for the Triage Plus study was used to estimate the parameters by fitting a model without covariates (Arnold et al., 2011) as

$$y_{ij}|x_{ij}, u_j, u_{ij} \sim Poisson\left(\mu_{ij}\right)$$

$$log\ \mu_{ij} = \log(n_{ij}) + \beta_0 + u_{ij} + u_j \qquad (5.3)$$

In the model (5.2),parameter β_1 is the parameter of interest, while the constant β_0 represents the average mean count of events in the control clusters at baseline time point.

Once the desired parameters were estimated, 12 repeated measurements, datasets and random effects were then generated. Cluster level data was created by expanding the data sets to the required number of clusters, and corresponding random effects were generated for

each cluster. The outcome response $y_{ij}$was then simulated using models (5.2), and Poisson mixed effects regression models (5.2) were fitted to the data set with indicators contrasting the intervention or control specified in the model. Regressions coefficients were saved for further analysis.

Design conditions investigated

In the simulation of data sets, the following factors were allowed to vary: number of clusters per arm, effect size in the intervention arm, number of repeated measurement points and incidence rate ratios as follows:

i. The number of clusters per arm was set at 3, 4, 5, 6, 9, and 12, with the 3 clusters per arm case corresponding to the Triage Plus study. The remaining cases represent the common number of clusters used in intervention studies and investigated in similar simulation studies involving hierarchical modelling.

ii. The intervention effect sizes were based on incidence rate ratio (IRR) and specified as 1.1, 1.2, 1.3, 1.4, 1.5, 1.6 and 1.8 by varying the regression coefficient of the variable contrasting intervention and control (the effect size of 1.2 or 20%increase in uptake represented the desired effect of the intervention in the Triage Plus study). The maximum of an IRR of 1.8 was considered to be the highest effect size that can be obtained in such interventions (though in certain instances, higher effect sizes than the 1.8 may arise). The other intervention effect sizes were included for comparison. These intervention effects were fixed during each of the simulations which were then compared to the estimated intervention effects after 1000 simulations. In investigating the effect of varying ICCs on statistical power estimation and the accuracy of parameter estimates, effect sizes of 1.2, 1.4, 1.6 and 1.8 are reported for better clarity in comparing the different ICCs in the same table.

iii. As the risk of encountering estimation problems increases as the number of data collection (time)points increases (due to the reduced number of potential observations at each fixed time) (Turtz and Kauermann, 2003), the intervention's effect was evaluated at time points for repeated outcome measurement, set at 2, 3, 4, 6 and 12. Basic baseline and post-intervention study design were represented by 2

measurement points: 12 measurement points represent the desired 12 month period for the project implementation used in Triage Plus.

iv. There were four different outcome measures investigated in the Triage Plus study (TB and ART treatment initiation rates as primary outcomes, HIV and TB testing uptake rates as secondary outcomes) with varied incidence rates. Incidence (or prevalence) of the investigated outcomes affects the statistical power required to detect a significant effect of the intervention and the accuracy of the parameter estimates (Raudenbush and Liu, 2000; Moineddin, Matheson and Glazier, 2007). Statistical power and the accuracy of the parameter estimates and their standard errors were also assessed by varying disease incidence based on the actual data derived from Triage Plus study (Dean et al. 2004; Burton et al., 2006, Arnold et al., 2011). This was done by varying the mean number of events (new or incident cases of TB and HIV treatment initiation rates) in the control clusters (Delayed arm), which corresponded to the intercept of the regression model. Because the incidence of TB treatment initiations is very low in Malawi and that TB testing uptake is higher[4], for the simulations TB treatment initiations and TB testing uptake were considered to be examples of low incidence and high incidence disease events. Therefore, the baseline number of new cases of TB treatment initiations and testing uptake of 15 and 70 respectively were used in the simulations.

Number of simulations

Although Burton et al (2006) provided formulae for determining the required number of simulations, the required number of replications in this study was determined by gradually increasing this number until no further changes in the parameter estimates were observed. In general, about 500 simulations were required to achieve adequately converged parameter estimates, but 1000 simulations were finally adopted and used for each of the conditions investigated in order to align with most reported simulations (Maas and Hox, 2005; Moineddin, Matheson and Glazier, 2007; Heo et al., 2013). All simulated data sets were then saved for further analysis.

[4] Only 20% of cases who take up TB diagnosis (which is TB smear microscopy) are eventually diagnosed TB cases in Malawi, and only a (variable and unknown) proportion of these proceed to start treatment.

Quantities of interest in each simulation

The parameters of interest stored after fitting Poisson mixed effects regression models included fixed effects regression coefficients (β_i) assumed to be constant across time points and their associated simulation standard errors (SE_i). The mean estimate $\overline{\beta}$ over the n simulations conducted were obtained under each of the conditions investigated as a measure of the true estimate of interest given as

$$\overline{\beta} = \sum_{i=1}^{n} \beta_i / n$$

where i ($i = 1, \ldots, n$) in this caserepresents number of simulations conducted to estimate the regression coefficients of the intervention variable (in our case, n=1000) in a given design condition investigated.

The mean estimates derived from the regression coefficients were then exponentiated to obtain incidence rate ratios. To show the levels of uncertainty around these estimates, empirical standard errors (calculated as the standard deviation of the parameter estimates over all the simulations) and the mean of the within simulation standard errors of the parameter estimates were calculated according to Burton et al. (2006).

The empirical standard error was calculated as

$$SE(\hat{\beta}) = \sqrt{[\frac{1}{n-1}] \sum_{i=1}^{n} (\beta_i - \overline{\beta})}$$

and the within simulation standard errors as

$$SE(\hat{\beta}) = \sum_{i=1}^{n} SE(\beta_i) / n$$

In addition, non-parametric summary measures of uncertainty using the 5% and 95% percentiles of distribution were obtained.

5.3.3 The Simulation Algorithm

In implementing the simulation study, the procedures described by Arnold et al (2011) and Heo and Leon (2009) were adopted. The covariates for each of the simulations were fixed by simulating them once.

The steps used can be summarised as follows:

1. Generate data by specifying the underlying hierarchical model with sample sizes, covariates, expressed parameters and random effects that are Normally distributed with mean zero and known variance-covariance structure.

    a. Generate a set of clusters with half of them assigned to the intervention arm and half to the control arm.

```
set  obs  `nclust'

gen n=_n

generate trt=0

recode   trt (0=1) if n>((`nclust')/2)          /*generate intervention variable with equal
                                                distribution, 0=control, 1=intervention cluster*/
```

    b. Generate cluster-level random effects using mean zero and a known variance-covariance structure (i.e. using ICCs of 0.00154 and 0.081) based on the baseline data sets (i.e. training datasets) for the Triage Plus study (see section 5.3.2).

```
drawnorm u2, mean(0) sd(0.005)                  /*generate cluster level random effects
                                                 using ICC from Triage Plus*/
```

    c. Generate a repeated measurement data structure with corresponding random effects for each measurement occasion using the baseline sample as in (b).

```
expandcl `time', generate(newcl) cluster(temp1)

drop newcl

replace n= _n

gen time=n-((temp1-1)*`time')

gen cluster = temp1

drop temp1

drawnorm u2, mean(0) sd(0.005)                /*generate cluster level random effects
                                               using ICC from Triage Plus*/
```

d. Estimate the mean of the outcome (β_0) in the control areas using the baseline data sets.

e. Fit the generalised linear mixed effects Poisson model (*xtmepoisson*) to the simulated data while fixing the intervention effect estimate.

```
xtmepoisson count trt if time<=`time'|| cluster:, cov(ex) iterate(100)

        scalar b_trt   = _b[trt]

        scalar se_trt  = _se[trt]
```

2. Simulate the response outcome (y_{ij}) for each of the cluster and measurement occasions using the appropriate model as described in section 5.3.2 after specifying the effect size and number of repeated measurement time points. After each simulation, store the appropriate parameters and standard errors (see in (e) above) for future analysis.

3. Repeat steps 1 and 2 a total of 1000 times.

   a. Estimate the statistical power (section 5.3.4)by calculating the test statistic (z test) and p-value. The test statistic is derived by dividing the parameter estimate by its standard error and using *normprob* function in Stata to obtain p-values. Thus, the p-value for a two-sided significant test is calculated using:

$$p\ value = 2[1 - \Phi(Z)],$$

where $Z = \frac{b}{sb}$ and Φ is the standard normal cumulative distribution function.

b. Estimate power by calculating the proportion of the total observations for which the p-value is ≤ 0.05.
c. Check if the simulations were calculated correctly by simulating the Null case to ensure that the statistical power is shown to be a specified significance level and that the distribution of the p- values is uniform.
d. Determine the accuracy of the parameter estimates by calculating all indicators for statistical efficiency using the formulae described in section 5.3.6.

5.3.4 Assessment of statistical power estimation for different design conditions

Lack of statistical significance in any particular evaluation could occur either because the intervention was not effective in producing the desired impact (i.e. the true effect size was smaller than anticipated) or because the study design made it unlikely that a biologically real effect would be detected. Hence, estimation of the actual statistical power of a study is critical in distinguishing between these two alternatives in the presence of statistical non-significance. Statistical power, defined as the probability of obtaining a statistically significant effect size estimate given that there is a biologically real effect in the population being studied, is dimensionless and as such is a better design parameter for comparing different designs using different outcome measures. Due to the design challenges of the Triage Plus study, where there were only 3 clusters in the intervention arm and 3 clusters in the control arm, the statistical power of the Triage Plus intervention was investigated through a simulation study.

The parameter estimates of the 1000 saved simulations completed under each of the studied design conditions were used to calculate test statistics (z test) and p-values as described by Arnold et al (2011). Statistical power for each design condition was then estimated as a proportion of the total observations for which the p-value is less than the conventional significance level of 5% ($\alpha = 0.05$).

5.3.5 Diagnostics and sensitivity analysis of simulation procedures

The formal development and evaluation of new diagnostic procedures for simulation methods is not within the remit of this dissertation. Nevertheless, the statistical methods for the simulations conducted may be sensitive to the different parameter values and covariates used in their evaluation. Hence, it is necessary to test or monitor the behaviour patterns of the parameter estimates to ensure that the models used are reasonably acceptable. Sensitivity analysis involves carrying out a series of diagnostic tests in which different parameter values

are adjusted to see how the changes affect the posterior results. By showing how the model responds to changes in the parameters used, sensitivity analysis is a useful tool in model evaluation. Thus, sensitivity analysis refers to the estimation and adjustment of model parameters, and it allows for interactions to improve agreement between model output and the data set.

To confirm that the simulations were correctly done, a simulation of a Null case (H0: B1=0) was carried out to demonstrate that the distribution of the *p*-values was uniform. In general, the simulations with a Null case were uniformly distributed despite having some minor variations at around a p value of 0.8 and 1 (see Figure 5).

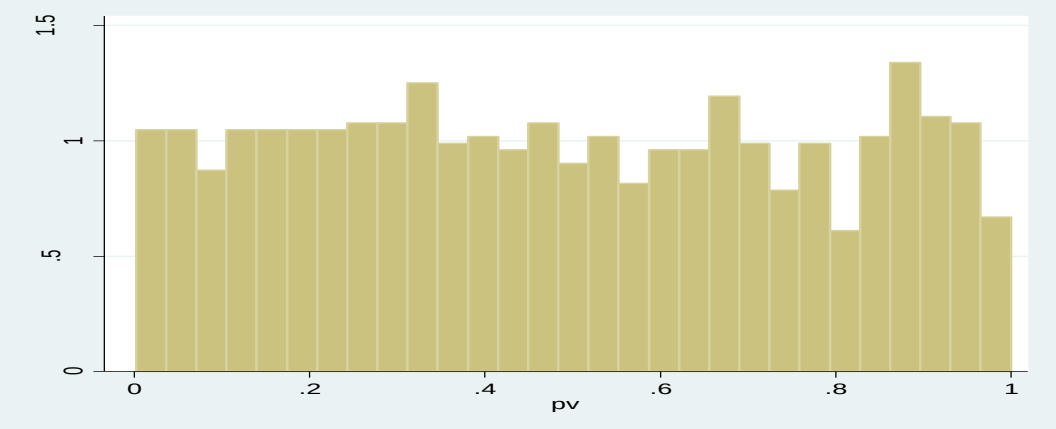


Figure 5: Distribution of p-values after running 1000 simulations when the effect of the intervention was set to a Null case

5.3.6 Assessing performance of the different statistical methods

As statistical efficiency may vary with design conditions, statistical efficiency under different design conditions using various performance measures was assessed (Maas & Hox, 2005; Moineddin, Matheson & Glazier, 2007; Burton et al., 2006; Arnold et al., 2011; Collins, Schafer and Kam, 2001). Performance measures used to evaluate efficiencies in parameter estimation included bias, mean square error, the average length of the 95% confidence intervals of parameters estimates and the coverage of the 95% confidence intervals as applied in similar studies. The simulated results were compared against the true values used to generate the data. Measures used to evaluate the performance of the statistical methods under varying design conditions are presented below.

The bias of parameter estimates

In assessing the accuracy of the statistical methods in parameter estimation, percentage relative bias as well as the standardised bias as described by Burton et al (2006) were calculated. If $\hat{\beta}$ is the estimate of the true population parameter β, the percentage relative bias for the true parameters is given as:

$$\frac{\hat{\beta} - \beta}{\beta} \text{x } 100$$

and the standardised bias is given as:

$$\frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \text{x } 100$$

where $\hat{\beta}$ is the mean estimate over n simulations. According to Collins et al (2001), any standardized bias greater than 40% is considered significant. In this thesis, only the percentage relative bias is used in assessing the accuracy in parameter estimation as it is commonly reported in simulation studies (see Moineddin, Matheson and Glazier, 2007).

The 95% coverage of standard errors

The accuracy of the standard errors of the parameter estimates was assessed by establishing the coverage of the 95% confidence interval of the asymptotic standard normal distribution in each simulated data set. The 95% coverage is a proportion of the times the 95% confidence interval of an estimate of interest includes the true population estimate. The coverage indicator was calculated and set to 1 if the confidence interval contained the true value and 0 if otherwise (Maas and Hox, 2005; Moineddin, Matheson and Glazier, 2007; Burton et al., 2006). Correct coverage should be within two standard errors of the nominal coverage probability (Gelman and Hill, 2007) given as:

$$\hat{\beta}_i \pm Z_{1-\alpha/2} SE(\hat{\beta}_i)$$

The mean square error

Because mean square error determination incorporates both bias and uncertainty around the parameter estimates, it provides an overall measure of the accuracy of parameter estimates (Burton et al., 2006; Collins, Schafer and Kam, 2001). The smaller the MSE estimate (ideally less than 5% of the estimate) the better the fit. Because the mean square error is not commonly reported in simulation studies, it was not used in this thesis. The MSE is given as:

$$(\hat{\beta} - \beta)^2 + (SE(\hat{\beta}))^2$$

Average length of confidence interval

The average length of the 95% confidence interval for the parameter estimate was used to assess efficiency in the estimation of parameter estimates. The shorter the average length of the interval, the more precise the estimates, hence the greater the efficiency and statistical power (Collins, Schafer and Kam, 2001; Burton et al., 2006). The average length of the 95% confidence interval is given by:

$$\text{Average length of the 95\% CI of}\hat{\beta} = \left(\sum_{i=1}^{n} 2Z_{1-\alpha/2}SE(\beta_i)\right)/n$$

5.4 Results

5.4.1 Convergence

Convergence rate was defined as the proportion of simulations that converged, and its properties were investigated under different scenarios. Model convergence rates were high overall in all scenarios investigated and varied from 99.2% to 100%. Unsurprisingly, there was no clear pattern in convergence rates after varying the number of clusters per arm, number of repeated measurements, effect size and/or incidence of the disease conditions and ICC confirming the findings of previous studies (Moineddin, Matheson and Glazier, 2007). Moineddin and colleagues found that convergence rates increased to 100%, regardless of the correlations structure, when group size was at least 30 individuals. In general, convergence rates increased with increasing ICC, number of groups and group sizes (Moineddin, Matheson and Glazier, 2007). However, subtle observations were noted in this study, in that increasing the number of clusters per arm still produced slight convergence failure rates (<1%) especially when the incidence of the disease conditions and repeated measurement times increased in a situation when the ICC was 0.00154 (Table 3). For ICC values of 0.081 and above, similar high convergence rates were observed when the number of clusters per arm and effect sizes were varied (Tables 4 & 5). It is interesting to note that overall convergence rates were high even when effect sizes were as low as 10% and with high ICC levels due to increased background incidence rate as the ICC increases leading to improved convergence rates (Amatya, Bhaumik and Gibbons, 2013). However, there was a convergence problem noted at the effect size of 10% when 12 clusters were used in low incidence disease

conditions such as TB treatment uptake rates when the ICC was 0.00154 (Table 3). Further simulation work to investigate this finding will be needed in future. However, this observation was not observed when the ICCs were 0.081 and above potentially due to the increased background incidence rate noted by Amatya, Bhaumik and Gibbons (2013).

Table 3: The effect of number of groups, effect size and number of time points on the convergence rate (%) based on 1000 simulated data sets

Number of time points	Mean monthly outcome measure in control clusters of 15 cases							Mean monthly outcome measure in control clusters of 70 cases						
	Effect size (incidence rate ratio)							Effect size (incidence rate ratio)						
	1.1	1.2	1.3	1.4	1.5	1.6	1.8	1.1	1.2	1.3	1.4	1.5	1.6	1.8
3 clusters per arm														
2	100	100	100	99.8	100	99.8	99.9	100.0	100.0	100	100.0	99.9	99.9	100.0
3	100	100	100	100	100	99.8	99.9	100.0	99.9	99.8	100.0	99.8	100.0	100.0
4	100	100	99.8	99.8	100	100	100	100.0	100.0	99.9	99.9	100	100.0	100.0
6	100	99.8	100	100	100	100	100	100.0	99.9	99.6	99.9	99.9	99.8	99.8
12	100	100	99.8	100	100	100	100	99.7	99.8	99.8	99.9	99.9	99.8	99.8
4 clusters per arm														
2	100	100	100	100	100	100	99.9	99.9	99.9	100	99.9	100	100	99.9
3	100	99.9	100	99.9	100	100	100	99.8	99.9	100	100	99.9	99.9	99.8
4	100	100	99.8	99.9	99.9	100	100	100	100	100	99.9	100	99.9	100
6	100	100	100	100	100	100	100	99.8	99.8	99.5	99.9	99.6	100	99.9
12	99.9	99.8	99.9	100	99.9	99.8	99.9	99.8	99.8	99.9	99.7	99.9	99.8	99.8
5 clusters per arm														
2	100	100	100	99.9	100	100	100	99.7	100	100	99.9	99.7	100	99.7
3	99.8	100	100	100	99.9	99.9	99.9	99.9	100	99.8	99.7	99.9	100	100
4	99.8	99.9	99.9	99.8	99.9	100	100	99.9	99.9	99.8	99.9	99.9	99.9	100
6	100	99.8	99.6	99.9	100	100	99.7	99.9	99.8	99.6	99.9	99.7	99.9	99.8
12	100	100	99.9	100	99.9	100	99.9	99.9	99.9	99.7	99.9	100	99.9	99.4
6 clusters per arm														
2	100	100.0	100	100.0	100	99.9	99.9	100.0	100.0	100	100.0	99.9	99.9	99.9
3	100	99.9	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100	100.0	100	99.8	99.6
4	99.8	99.9	100	100.0	99.9	99.9	99.9	100.0	99.9	99.8	99.8	99.8	99.5	99.9
6	99.9	99.9	99.9	99.7	100	100.0	100.0	99.5	99.6	99.8	99.4	99.9	99.8	99.9
12	99.9	99.9	100	99.9	99.6	99.8	99.9	99.7	99.7	99.7	99.6	100	99.8	99.7
9 clusters per arm														
2	99.9	99.9	100	99.8	99.9	100.0	99.8	99.9	100.0	99.9	99.9	100	100.0	99.9
3	99.9	99.8	100	100.0	100	100.0	100.0	99.9	99.8	100	99.9	99.8	100.0	99.8
4	99.8	100.0	100	100.0	100	99.9	100.0	99.8	100.0	99.6	99.9	99.8	99.9	99.9
6	99.9	100.0	99.7	99.9	100	99.7	100.0	99.6	99.7	99.4	99.2	99.9	99.8	99.4
12	100.0	99.5	99.8	99.7	99.8	100.0	99.8	99.7	99.9	99.4	99.3	99.5	99.1	99.8
*12 clusters per arm**														
2	-	100.0	100	100.0	100	99.8	99.8	99.9	100.0	99,9	100.0	99.5	99.7	99.5
3	-	100.0	100	100.0	100	99.9	99.9	100.0	100.0	99.8	99.8	99.7	99.9	99.7
4	-	99.9	99.9	99.7	99.9	99.9	99.9	99.8	99.8	99.8	99.8	99.8	99.9	99.7
6	-	99.9	99.9	99.9	99.7	99.9	99.9	99.9	99.6	99.8	99.4	99.5	99.6	99.8
12	-	99.8	99.8	100.0	99.7	99.7	100.0	99.4	99.5	99.7	99.0	99.6	99.2	99.7

ICC=0.00154

* No convergence was achieved at the effect size of 10%, hence no estimates are given

Table 4: The effect of the ICC, number of groups, effect size and repeated measurements on convergence rate, based on 1000 simulated data sets when the disease incidence is low: Mean outcome measure in control clusters of 15 cases per month

		ICC=0.081				ICC = 0.321				ICC = 0.699			
		Effect size (incidence rate ratio)				Effect size (incidence rate ratio)				Effect size (incidence rate ratio)			
	Effect	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8
3 clusters	Time												
	2	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100	100.0
	3	100.0	99.9	99.8	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100	100.0
	4	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100	100.0	100	100.0
	6	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100	99.9
	12	99.9	99.9	99.8	99.8	100.0	100.0	100.0	99.9	99.9	100.0	99.9	100.0
4 clusters													
	2	100.0	99.9	100.0	99.8	100.0	100.0	100.0	100.0	99.9	100.0	100.0	99.9
	3	99.7	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0
	4	100.0	99.9	100.0	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	6	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	99.9
	12	99.9	100.0	99.8	99.8	99.9	100.0	100.0	100.0	99.9	100.0	100.0	100.0
5 Clusters													
	2	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
	3	99.9	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	99.8	99.9	100.0
	4	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
	6	99.7	100.0	99.9	100.0	100.0	100.0	100.0	100.0	99.9	99.8	100.0	99.9
	12	100.0	99.7	99.9	99.8	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0
6 Clusters													
	2	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0
	3	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	100.0	99.8	99.9
	4	99.9	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0
	6	99.9	100.0	99.9	99.9	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0
	12	99.9	99.9	99.8	99.5	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
9 Clusters													
	2	100.0	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9
	3	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0	99.9	99.9	100.0	100.0
	4	100.0	100.0	99.6	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	6	99.6	100.0	99.9	99.9	100.0	100.0	100.0	100.0	99.9	100.0	99.7	99.9
	12	99.5	100.0	99.6	99.7	99.9	99.6	100.0	100.0	99.9	99.9	99.9	99.9
12 Clusters													
	2	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	100.0
	3	100.0	99.9	99.9	100.0	100.0	99.9	100.0	100.0	100.0	99.9	100.0	100.0
	4	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0
	6	99.9	99.8	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.7	100.0
	12	100.0	99.9	99.6	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	99.9

Table 5: The effect of the ICC, number of groups, effect size and repeated measurements on convergence rate, based on 1000 simulated data sets when the disease incidence is high: Mean outcome measure in control clusters of 70 cases per month

		ICC=0.081				ICC = 0.321				ICC = 0.699			
		Effect size (incidence rate ratio)				Effect size (incidence rate ratio)				Effect size (incidence rate ratio)			
	Effect	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8
3 clusters	Time												
	2	100.0	99.8	99.9	99.7	100.0	100.0	99.9	99.9	99.9	99.9	99.9	99.9
	3	99.8	99.8	99.4	100.0	99.9	99.9	99.8	100.0	100.0	100.0	100.0	99.9
	4	99.9	99.8	99.6	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9
	6	99.8	99.8	99.5	99.6	99.8	100.0	99.9	100.0	100.0	99.7	99.9	100.0
	12	99.8	98.3	97.9	97.0	99.8	100.0	100.0	100.0	100.0	99.9	99.8	99.6
4 clusters													
	2	99.9	99.8	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0
	3	99.8	99.7	99.5	99.8	99.9	99.8	100.0	100.0	100.0	100.0	99.9	99.7
	4	99.9	99.6	99.8	99.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
	6	99.3	99.6	99.5	99.1	99.7	99.8	100.0	99.9	99.8	99.8	99.9	99.9
	12	98.7	98.5	97.6	98.0	100.0	100.0	100.0	100.0	99.7	99.6	99.6	99.9
5 Clusters													
	2	99.8	100.0	100.0	99.9	100.0	99.8	99.8	100.0	99.9	100.0	100.0	100.0
	3	99.9	99.9	99.6	99.9	100.0	99.9	100.0	99.9	100.0	100.0	99.9	100.0
	4	99.7	99.7	99.4	99.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	6	99.4	98.9	99.7	99.4	99.9	100.0	100.0	100.0	100.0	99.9	99.8	99.9
	12	98.1	98.3	97.9	98.3	100.0	100.0	100.0	100.0	99.8	99.9	99.7	99.8
6 Clusters													
	2	99.9	99.8	100.0	99.8	99.9	99.9	99.9	100.0	99.9	100.0	100.0	100.0
	3	100.0	99.9	99.9	99.9	99.9	100.0	100.0	99.7	100.0	99.8	99.9	100.0
	4	99.4	99.9	99.9	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	6	99.3	99.5	99.6	98.2	100.0	100.0	99.8	99.9	99.8	99.4	99.9	99.7
	12	98.2	98.2	98.0	96.9	100.0	100.0	100.0	100.0	99.9	99.8	100.0	100.0
9 Clusters													
	2	99.9	99.8	100.0	100.0	99.8	99.9	100.0	100.0	99.9	100.0	99.8	100.0
	3	99.6	99.9	99.5	99.6	99.8	99.9	99.9	99.8	99.8	100.0	99.9	100.0
	4	99.6	99.5	99.3	99.3	100.0	100.0	100.0	100.0	100.0	100.0	99.7	99.9
	6	99.5	99.5	99.2	99.1	100.0	99.9	99.8	99.7	99.6	99.8	99.7	99.9
	12	98.1	95.8	97.3	96.4	100.0	100.0	100.0	99.9	99.8	100.0	99.8	99.9
12 Clusters													
	2	100.0	100.0	100.0	100.0	100.0	99.9	99.8	99.9	99.7	99.8	100.0	100.0
	3	99.9	99.6	99.8	99.7	100.0	100.0	99.9	100.0	99.8	100.0	100.0	99.8
	4	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0
	6	99.7	99.8	99.5	99.1	99.9	99.9	99.9	99.9	99.8	99.6	99.7	99.7
	12	99.5	98.7	97.7	98.4	99.9	100.0	100.0	99.9	99.7	99.8	99.6	99.8

5.4.2 Power estimation

Statistical power determination was calculated as the proportion of the total observations for which the observed p value was ≤ 0.05. Conventionally, when detecting a significant effect a power of at least 80% is considered adequate (Cohen, 1992).

The results for estimating statistical power for different scenarios are presented in Tables 6 - 9. As statistical properties vary according to the prevalence or incidence of the conditions being studied (Moineddin, Matheson and Glazier, 2007), the results are presented according to the incidence levels of the studied disease conditions, which is in line with the different study outcomes of the Triage Plus study. The disease conditions are categorised as low incidence (e.g. TB treatment initiation rates) and high incidence (access rates to HIV testing, TB testing or ART treatment initiation rates). Statistical power estimates for ICCs of 0.081 and higher are presented in Tables 8 and 9 for low and high incidence diseases respectively.

5.4.2.1 Power estimates when the incidence of the outcome is low and ICC is 0.00154

The Triage Plus study had 3 clusters per arm and a desired effect size of 20%. Under these circumstances, the simulation study indicates that the Triage Plus study would have had only 24% statistical power to detect a 20% improvement in the detection and initiation of TB treatment if only 2 repeated measurement times had been used in each cluster: a pre- and post- evaluation design. However, statistical power would have increased with increasing effect sizes and the number of repeated measurement times. For instance, at least 80% statistical power would have been attained with an effect size of at least 50% and only 2 time points. The same percentage of power (>80%) would also be achieved with the desired effect size of 20% if there are 12 repeated measurement times. However, with an effect size as low as 10%, inadequate power is achieved even with 12 repeated measurement times (37% power at 12 measurements).

With an effect size of 20% and two assessment time points, varying the number of clusters per intervention arm produced an improvement in estimated power from 24% with 3 clusters to 34%, 36%, 42%, 61% and 70% for 4, 5, 6, 9 and 12 clusters respectively. Estimated power increased to at least 80% when the number of repeated measurement times was set at 6 or 12 with 5 or 6 clusters per arm. Thus, using 9 or12 clusters per arm required less repeated measurement times (i.e. 4 and 3 measurement times for 9 and 12 clusters per arm respectively) to achieve adequate power to detect an effect size of 20%.

By increasing the effect size to >20%, adequate power was achieved with less repeated measurement times: the minimum number of assessment times required decreased as the number of clusters per arm increased. With an effect size of 10%, none of the design conditions gave adequate power even with an increased number of clusters per arm or repeated measurement times (Table 6).

Table 6:The effect of the number of clusters, repeated measurement times and effect size on the statistical power of the study design (% p value<0.05) based on 1000 simulated data sets under low disease incidence – The mean outcome measure in control clusters is 15 cases per month.

Clusters and time points	Effect size (incidence rate ratio)						
3 clusters per arm	1.1	1.2	1.3	1.4	1.5	1.6	1.8
2	9.8	24.0	46.2	68.3	83.7	92.2	98.7
3	12.3	35.5	64.7	85.1	94.4	97.6	99.9
4	15.8	41.9	76.5	93.1	98.5	99.6	100
6	20.2	58.7	88.9	98.4	99.9	100.0	100.0
12	37.0	88.9	99.8	100.0	100.0	100.0	100.0
4 clusters per arm							
2	11.8	33.9	58.1	79.4	90.8	96.1	99.8
3	15.3	45.0	77.4	90.7	98.0	99.5	100.0
4	18.5	54.4	86.5	97.3	99.7	99.9	100.0
6	26.2	74.1	95.4	99.6	100.0	100.0	100.0
12	45.2	94.5	99.8	100.0	100.0	100.0	100.0
5 clusters per arm							
2	12.3	36.2	67.2	84.7	95.1	98.2	99.7
3	17.2	51.6	84.8	96.9	99.8	100.0	100.0
4	22.4	64.9	93.0	99.3	100.0	100.0	100.0
6	30.8	82.7	98.2	100.0	99.9	100.0	100.0
12	54.0	98.7	100.0	100.0	100.0	100.0	100.0
6 clusters per arm							
2	15.6	41.9	75.7	88.9	97.2	99.3	100.0
3	23.5	61.3	88.1	97.9	99.5	100.0	100.0
4	25.6	72.2	95.7	99.7	100.0	100.0	100.0
6	38.2	89.2	99.1	100.0	100.0	100.0	100.0
12	65.2	99.3	100.0	100.0	100.0	100.0	100.0
9 clusters per arm							
2	20.7	60.5	88.1	97.6	99.3	99.9	100.0
3	29.4	77.4	96.7	99.7	100.0	100.0	100.0
4	40.5	87.1	99.1	100.0	100.0	100.0	100.0
6	50.9	97.0	100.0	100.0	100.0	100.0	100.0
12	79.1	99.5	100.0	100.0	100.0	100.0	100.0
*12 clusters per arm**							
2	-	69.6	94.5	99.1	100.0	100.0	100.0
3	-	86.1	99.2	100.0	100.0	100.0	100.0
4	-	94.9	99.7	100.0	100.0	100.0	100.0
6	-	99.2	100.0	100.0	100.0	100.0	100.0
12	-	100.0	100.0	100.0	100.0	100.0	100.0

ICC=0.00154

* No convergence was achieved at the effect size of 10%, hence no estimates are given

5.4.2.2 Power estimates when incidence of the outcome is high and ICC is 0.00154

The Triage Plus study also included outcome measures for disease conditions with different incidence rates over the intervention period (e.g. TB testing uptake rates, ART initiation rates among eligible HIV positives individuals, HIV testing uptake rates, etc.), power was therefore also assessed for high incidence conditions. Without loss of generality, the effect of the intervention for improving uptake rates of TB testing, in terms of presumptive TB cases newly accessing TB microscopy testing, was evaluated as a proxy for high incidence outcomes.

Using TB microscopy testing uptake rates as the outcome measure of interest, with a mean number of presumptive TB cases accessing TB testing = 70 in the control clusters (based on the baseline sample), the simulation results showed that 3 clusters per arm provided over 77% power to detect a 20% improvement in uptake when two repeated measurement times were used. Statistical power increased to >95% when the effect size was increased to >20% even with only two repeated measurement times and 3 clusters per arm. Estimated power was at least 90% with an effect size of 20% when 3 or more repeated measurements times were used. However, with an effect size of 10% and just 3 clusters per arm, adequate power was only achieved when 12 repeated measurement times were used.

When the number of clusters per arm was varied, an estimated power level of 80% or greater was achieved with 2 repeated measurements and an effect size >20%. With an effect size of only 10%, estimated power of ≥80% was achieved with 4 - 12 repeated measurements for 4 or more clusters per arm.

Thus, in high incidence outcomes, adequate power to detect an intervention effect of 20% or greater can be achieved even when the number of clusters per study arm is as low as 3; more clusters per arm or repeated measurement times are required when the effect size is below 20% (Table 7).

The relationship between estimated power and the number of repeated measurement points in the 3 clusters per arm study design for different effect sizes is presented in Figure 6 for low incidence diseases and Figure 7 for high incidence diseases. Figure 6 clearly shows that with a true effect size of up to 20% (e.g. improving service access to TB treatment), statistical power of at least 80% can only be achieved with at least 12 repeated measurements times; however, with effect sizes of ≥40%, power levels of 80% or greater can be achieved with only 3 repeated

measurement times. With high incidence conditions (e.g. access to TB microscopy testing) shown in Figure 7, adequate statistical power of 80% or greater is achieved with 12 repeated measurements for effect sizes of 10%, but fewer repeated measurements times (in some situations as few as 3 times) may be required when the effect size is 20% or greater .

Table 7:The effect of the number of clusters, repeated measurement times and effect size on the statistical power of the study design based on 1000 simulated data sets under high disease incidence. The mean outcome measure in control clusters is 70 cases per month.

Clusters and measurement times	Effect size (incidence rate ratio)						
3 clusters per arm	1.1	1.2	1.3	1.4	1.5	1.6	1.8
2	28.7	77.4	96.0	99.3	99.9	100.0	100.0
3	43.6	90.1	99.3	100.0	100.0	100.0	100.0
4	52.3	95.7	100.0	99.9	100.0	100.0	100.0
6	69.1	99.2	100.0	100.0	100.0	100.0	100.0
12	95.1	100.0	100.0	100.0	100.0	100.0	100.0
4 clusters per arm							
2	36.0	87.0	98.5	99.8	100.0	100.0	100.0
3	51.9	95.2	99.9	100.0	100.0	100.0	100.0
4	63.6	99.3	100.0	100.0	100.0	100.0	100.0
6	80.7	99.9	100.0	100.0	100.0	100.0	100.0
12	98.3	100.0	100.0	100.0	100.0	100.0	100.0
5 clusters per arm							
2	44.6	91.0	99.7	99.8	99.9	100.0	100.0
3	63.5	98.1	100.0	100.0	100.0	100.0	100.0
4	73.1	99.8	100.0	100.0	100.0	100.0	100.0
6	87.1	100.0	100.0	100.0	100.0	100.0	100.0
12	99.0	100.0	100.0	100.0	100.0	100.0	100.0
6 clusters per arm							
2	52.9	95.2	99.7	99.9	100.0	100.0	100.0
3	68.0	98.7	100.0	100.0	100.0	100.0	100.0
4	82.7	99.8	100.0	100.0	100.0	100.0	100.0
6	93.4	100.0	100.0	100.0	100.0	100.0	100.0
12	99.9	100.0	100.0	100.0	100.0	100.0	100.0
9 clusters per arm							
2	67.4	98.3	100.0	100.0	100.0	100.0	100.0
3	85.6	99.7	100.0	100.0	100.0	100.0	100.0
4	92.4	100.0	100.0	100.0	100.0	100.0	100.0
6	90.0	100.0	100.0	100.0	100.0	100.0	100.0
12	99.9	100.0	100.0	100.0	100.0	100.0	100.0
12 clusters							
2	76.9	99.8	100.0	100.0	100.0	100.0	100.0
3	91.3	100.0	100.0	100.0	100.0	100.0	100.0
4	97.7	100.0	100.0	100.0	100.0	100.0	100.0
6	99.7	100.0	100.0	100.0	100.0	100.0	100.0
12	100	100.0	100.0	100.0	100.0	100.0	100.0

* ICC=0.00154

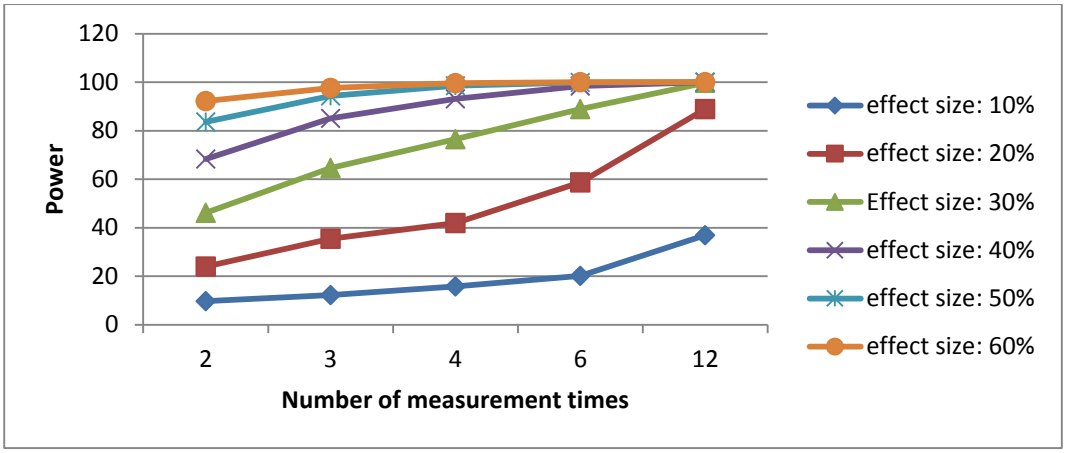


Figure 6 **Power curves when the disease incidence is low:**

The power curves show the relationship between power and the number of repeated measurement time points at different effect sizes (10%, 20%, 30%, 40%, 50% and 60%) when 3 clusters per arm are used using 1000 simulated datasets, when incidence is low. The power curves indicate that statistical power increases as the number of repeated measurement times increases and the increase is higher when the effect size is at least 20%.

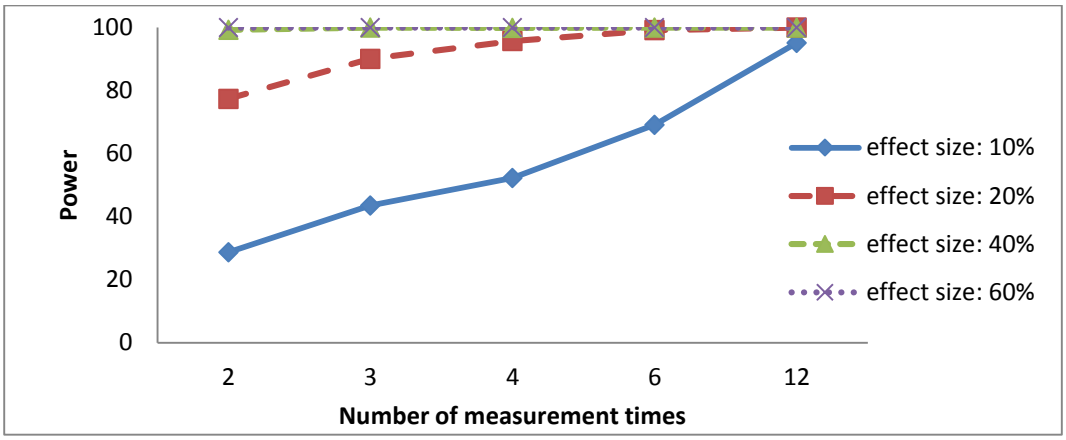


Figure 7 Power curves when the disease incidence is high.

The power curves show the relationship between power and the number of repeated measurement time points at different effect sizes (10%, 20%, 40% and 60%) when 3 clusters per arm are used using 1000 simulated datasets, when incidence is high. The curves indicate that statistical power increases as the number of repeated measurement times increases. With high incidence diseases, adequate statistical power is achieved even at low effect sizes of 10% after 12 measurement time points.

5.4.2.3 Effect of varying ICC on power estimation

To investigate the effect on statistical power estimation due to increasing the intracluster correlations when disease incidence is low (15 events per month), the ICCs were varied from

0.081 to 0.699. In general, statistical power decreased with increasing ICCs and decreasing number of clusters. With a very high ICC (0.699), no adequate statistical power was achieved in all scenarios investigated. For an ICC of 0.321, adequate statistical power was achieved when the number of clusters was 12 and the effect size was 60%. The statistical power increased with increasing number of repeated measurement times (Table 8). With an ICC of 0.081, adequate statistical power was achieved with 4 repeated measurement time points and an effect size of 40% in a 3 cluster per arm design. For lower effect sizes (20% and below), adequate statistical power was achieved when the number of clusters was 9 with 12 repeated measurement times (Table 8).

In high incidence disease situations (70 events per month), no adequate statistical power was achieved in all scenarios investigated when the ICC was 0.699. However, when the ICC was varied up to just 0.321, higher statistical power was achieved with high effect sizes and adequate power was attained when the number of clusters was 6 with 3 repeated measurement time points at an effect size of 80%. For an effect size of 60%, 12 clusters per arm were needed to achieve adequate statistical power. No adequate statistical power was achieved when the effect size was 40% or less even after increasing the number of clusters to 12 per arm (Table 9). With an ICC of 0.081, adequate statistical power was achieved with only 3 repeated measurement time points in a 3 cluster per arm design when the effect size was 40%. For an effect size of 20%, a minimum of 9 clusters per arm and 6 repeated measurement times were needed to achieve statistical power of at least 80% (Table 9).

Table 8: The effect of the ICC, number of groups, effect size and repeated measurements on the statistical power, based on 1000 simulated data sets when the disease incidence is low: Mean outcome measure in control clusters of 15 cases per month

		ICC=0.081				ICC = 0.321				ICC = 0.699			
		Effect size (incidence rate ratio)				Effect size (incidence rate ratio)				Effect size (incidence rate ratio)			
	Effect	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8
3 clusters	Time												
	2	30.7	63.3	81.0	89.3	20.6	30.8	44.6	60.3	16.7	20.1	22.8	23.4
	3	36.1	71.2	89.2	95.4	21.6	33.0	47.2	57.9	16.8	19.7	20.9	24.0
	4	44.6	79.7	92.8	97.7	24.8	36.6	51.6	61.4	18.7	20.4	23.5	25.9
	6	47.6	83.6	96.3	99.6	25.4	37.2	48.2	62.8	17.4	18.9	22.6	24.4
	12	50.9	87.8	98.0	99.6	22.6	39.6	50.8	63.8	18.7	21.5	24.7	25.5
4 clusters													
	2	36.1	68.3	84.9	93.2	20.0	34.0	50.4	66.1	12.1	15.9	19.4	22.6
	3	41.7	78.4	92.9	97.2	19.4	32.1	51.4	67.9	12.0	14.2	19.4	22.7
	4	45.8	84.6	96.2	99.3	21.4	35.0	50.3	63.7	14.9	17.6	18.3	18.6
	6	51.1	90.4	98.7	99.8	20.0	39.6	56.0	70.9	15.7	16.5	18.0	23.0
	12	55.6	93.9	99.5	100.0	21.1	35.9	52.5	69.2	15.0	15.6	16.5	23.3
5 Clusters													
	2	39.0	76.1	89.1	94.2	20.8	38.7	52.1	68.3	10.5	12.4	17.3	18.0
	3	43.6	86.1	95.6	98.7	16.4	35.3	55.4	73.6	13.1	15.8	15.7	21.3
	4	51.5	90.1	97.6	99.6	19.4	38.5	55.9	72.7	12.0	14.0	16.8	21.1
	6	54.1	93.7	99.5	100.0	19.1	39.2	57.0	71.9	12.8	13.3	14.4	19.6
	12	57.0	96.3	100.0	100.0	18.6	39.1	58.1	75.0	12.6	13.7	15.9	19.2
6 Clusters													
	2	40.1	79.1	91.0	94.9	18.3	36.3	59.6	74.9	11.0	13.9	17.8	18.4
	3	52.5	87.7	96.9	99.0	19.5	37.8	60.9	81.2	9.1	14.1	15.1	21.4
	4	52.5	91.2	98.3	99.7	20.2	41.4	62.5	76.8	12.1	12.4	15.2	19.7
	6	62.4	97.6	99.8	100.0	18.3	41.8	61.9	78.4	11.7	15.5	16.2	19.9
	12	65.7	97.9	100.0	100.0	20.6	42.8	63.0	79.5	10.7	14.9	16.7	20.4
9 Clusters													
	2	52.1	86.7	94.5	98.6	20.2	48.2	69.9	88.9	8.9	12.8	16.9	23.6
	3	64.7	93.5	98.8	99.8	22.5	45.3	75.8	91.6	9.2	13.1	15.6	22.0
	4	70.6	98.0	99.8	99.9	22.0	48.2	73.6	90.5	9.0	13.4	17.7	22.4
	6	75.3	99.3	100.0	100.0	21.8	49.7	76.0	91.8	9.8	14.2	17.4	24.1
	12	80.0	99.9	100.0	100.0	22.2	56.1	79.3	92.0	9.8	13.0	17.2	23.6
12 Clusters													
	2	59.4	93.7	96.9	98.8	18.7	55.1	81.7	95.7	9.6	13.6	20.0	25.6
	3	70.8	96.9	99.7	99.6	23.5	58.9	84.7	95.1	8.8	15.2	16.0	26.8
	4	73.1	99.0	99.8	100.0	25.0	59.7	86.2	96.4	9.8	12.1	18.3	25.4
	6	81.8	100.0	100.0	100.0	23.3	61.8	87.6	97.3	9.9	13.1	18.1	24.9
	12	86.9	100.0	100.0	99.7	27.0	63.1	86.1	96.2	8.8	12.9	18.1	28.3

Table 9: The effect of the ICC, number of groups, effect size and repeated measurements on the statistical power, based on 1000 simulated data sets when the disease incidence is high: Mean outcome measure in control clusters of 70 cases per month

		ICC=0.081				ICC = 0.321				ICC = 0.699			
		Effect size (incidence rate ratio)				Effect size (incidence rate ratio)				Effect size (incidence rate ratio)			
	Effect	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8
3 clusters	Time												
	2	42.8	79.1	95.3	98.3	22.8	39.2	48.3	62.3	15.8	20.4	21.3	23.7
	3	46.5	84.3	97.0	99.4	24.4	37.9	49.6	66.5	21.1	19.7	24.3	23.2
	4	51.9	87.9	98.1	99.8	23.7	35.7	53.0	65.4	20.5	20.5	21.8	21.7
	6	51.3	89.8	98.4	99.9	23.3	35.2	54.3	63.8	18.6	23.9	21.4	23.8
	12	51.0	87.7	98.4	99.8	23.1	38.4	51.5	63.6	18.8	18.4	21.6	27.1
4 clusters													
	2	49.2	85.5	96.8	99.2	21.5	34.3	56.6	69.6	14.9	17.0	17.7	22.1
	3	49.6	91.0	99.7	99.7	21.0	35.0	56.2	70.6	16.4	17.2	19.6	21.5
	4	55.5	93.1	99.3	100.0	19.9	37.0	53.0	72.3	15.3	16.1	17.3	19.4
	6	56.9	91.8	99.8	100.0	23.8	37.9	55.2	70.8	14.5	17.5	19.7	22.4
	12	56.9	94.5	99.7	100.0	22.9	38.4	54.7	68.6	14.2	18.0	17.6	23.1
5 Clusters													
	2	52.4	91.8	98.1	99.6	19.6	40.7	56.6	75.3	13.5	13.4	16.4	19.4
	3	60.8	94.3	99.7	100.0	19.7	43.2	59.4	75.7	11.7	14.8	15.9	19.3
	4	58.8	96.4	99.9	100.0	22.0	35.1	58.0	75.2	14.4	14.4	18.1	19.9
	6	59.2	96.7	99.9	100.0	18.4	40.4	61.2	76.3	14.6	14.1	16.2	20.1
	12	63.3	96.5	99.9	100.0	20.9	40.6	59.2	75.6	13.7	15.3	17.4	22.0
6 Clusters													
	2	57.5	93.4	98.3	99.6	17.7	40.4	62.5	78.0	13.9	12.6	16.3	19.7
	3	62.5	97.2	99.8	100.0	23.4	41.9	66.9	81.8	10.4	13.4	16.5	20.0
	4	64.0	97.1	100.0	100.0	20.5	43.0	67.4	80.6	11.7	14.9	14.9	20.8
	6	67.3	98.2	99.8	100.0	19.5	42.7	64.3	82.7	11.7	12.1	17.0	21.2
	12	66.6	98.6	100.0	100.0	22.1	40.7	63.6	82.4	12.0	12.4	14.3	21.2
9 Clusters													
	2	69.9	96.9	99.6	99.8	21.4	50.7	78.2	90.9	10.6	13.4	18.4	25.1
	3	76.4	99.2	100.0	100.0	22.4	53.3	78.8	92.0	8.9	12.7	16.9	22.1
	4	77.1	99.7	100.0	100.0	21.4	54.9	78.2	91.3	10.4	13.6	17.8	22.5
	6	80.7	99.8	100.0	100.0	23.3	53.4	79.4	92.8	11.4	13.6	17.0	20.6
	12	82.2	100.0	100.0	100.0	22.5	54.0	77.5	92.0	9.7	13.7	21.0	22.2
12 Clusters													
	2	77.8	99.3	100.0	100.0	23.8	59.5	86.2	96.1	9.5	12.8	18.2	26.0
	3	84.0	99.7	100.0	100.0	24.9	60.7	87.8	97.2	9.9	15.0	18.0	28.1
	4	85.7	100.0	100.0	100.0	23.3	62.9	87.0	95.5	9.1	13.2	18.3	25.6
	6	90.1	100.0	100.0	100.0	27.9	62.9	87.0	97.0	9.5	13.8	19.2	24.5
	12	88.6	100.0	100.0	100.0	25.3	63.4	87.0	97.6	8.4	13.8	18.1	29.0

5.4.3 Assessment of accuracy of parameter estimates

5.4.3.1 Parameter estimates and precision when the ICC was 0.00154

To better understand the statistical issues relating to different design scenarios, probabilistic simulations were used to assess the accuracy of parameter estimates (Gelman & Hill, 2007). These simulations showed negligible bias (deviation from a true estimate) in the estimation of fixed effects parameters of <1% in all scenarios evaluated for high incidence disease conditions of 70 events per month (Table 10). In low incidence diseases of 15 events per month, rates of bias of up to 2% were observed especially in a 3 cluster per arm design with 2 measurements times (Table 10).The results are similar to simulation studies done by Mass & Hox (2005) and Bennett et al (2002) who also found negligible bias in the estimation of fixed effects parameters. However, Moineddin et al (2007) found biases for fixed effect parameters as high as 3.7% using multilevel logistic regression models.

However, in an assessment of 95% confidence intervals and 5% and 95% percentiles, the results from this study indicated a clear pattern, with the precision of parameter estimates (based on the widths of 95% confidence intervals and 5% and 95% percentiles) improving as the number of clusters per arm was increased and as the number of measurement times in each cluster was increased (Tables 11 - 13). To clearly show how precision was related to different design conditions, Table 13 presents the average length of the 95% confidence intervals of the parameter estimates calculated as an absolute difference between the upper and lower limits of the confidence intervals around the estimate. The overall precision of parameter estimates improves with an increased number of clusters and repeated measurements times and a higher incidence of the studied disease condition.

Average 95% confidence interval length improved from 0.729 to 0.275 in low disease incidence using 3 clusters per arm, an effect size of 20% and 2 to 12 repeated measurement times. For high disease incidence, precision improved from 0.332 to 0.128. The greatest reductions in average 95% confidence interval length occurred using 12 repeated measurements and 12 clusters (0.137 for low incidence outcomes, 0.064 for high incidence outcomes).

It is important to note that the average length of the 95% confidence intervals increased with increasing effect size and was more evident in low incidence outcomes. With a low incidence outcome, 3 clusters per arm and 6 measurement occasions, the average length of the 95% confidence intervals increased from 0.368, when the effect size was 10%, to 0.541, when the

effect size was 80%. Similarly, when there was a high incidence outcome and 3 clusters per study arm, the average length of the 95% confidence intervals increased from 0.171, when the effect size was 10%, to 0.252, when the effect size was 80%.

The findings from these simulations show that to achieve precise estimates, which will in turn provide gains in statistical efficiency and hence statistical power, at least 6 repeated measurement times are required to improve precision by 50% and that the gain in precision increases as the incidence of the outcome increases (Table 13).

Table 10: The effect of the number of groups, time points and the effect size on how accurately the estimated effect size is close to the fixed estimate size and how the precision of the estimated effect size improves based on the 95% confidence interval after 1000 simulated data sets (using ICC=0.00154)

	Low incidence outcomes- Mean outcome measure in control clusters of 15 cases per month							High incidence outcomes- Mean outcome measure in control clusters of 70 cases per month						
	Effect size (incidence rate ratio)*							Effect size (incidence rate ratio)*						
	1.1	1.2	1.3	1.4	1.5	1.6	1.8	1.1	1.2	1.3	1.4	1.5	1.6	1.8
3 clusters per arm														
2	1.114 (0.826-1.508)	1.214 (0.9-1.634)	1.319 (0.988-1.767)	1.428 (1.074-1.905)	1.519 (1.148-2.018)	1.626 (1.233-2.151)	1.823 (1.392-2.393)	1.103 (0.961-1.267)	1.203 (1.048-1.381)	1.301 (1.138-1.488)	1.403 (1.231-1.601)	1.505 (1.325-1.711)	1.599 (1.407-1.819)	1.802 (1.591-2.042)
3	1.109 (0.873-1.412)	1.206 (0.953-1.528)	1.312 (1.042-1.654)	1.413 (1.126-1.776)	1.512 (1.209-1.895)	1.615 (1.294-2.019)	1.811 (1.456-2.256)	1.102 (0.987-1.231)	1.200 (1.075-1.339)	1.303 (1.171-1.450)	1.401 (1.263-1.556)	1.501 (1.352-1.668)	1.603 (1.448-1.775)	1.797 (1.627-1.987)
4	1.101 (0.896-1.355)	1.201 (0.981-1.472)	1.304 (1.069-1.593)	1.401 (1.152-1.704)	1.508 (1.244-1.830)	1.610 (1.331-1.949)	1.807 (1.501-2.178)	1.102 (1.002-1.213)	1.200 (1.094-1.317)	1.303 (1.188-1.430)	1.402 (1.281-1.534)	1.503 (1.374-1.644)	1.602 (1.467-1.750)	1.801 (1.652-1.963)
6	1.101 (0.933-1.301)	1.199 (1.019-1.411)	1.298 (1.106-1.524)	1.399 (1.196-1.638)	1.499 (1.283-1.753)	1.594 (1.367-1.859)	1.800 (1.550-2.091)	1.101 (1.019-1.190)	1.202 (1.114-1.298)	1.302 (1.209-1.403)	1.398 (1.301-1.503)	1.503 (1.399-1.616)	1.599 (1.490-1.717)	1.802 (1.680-1.932)
12	1.102 (0.981-1.487)	1.200 (1.071-1.346)	1.299 (1.161-1.454)	1.398 (1.252-1.562)	1.500 (1.345-1.673)	1.598 (1.436-1.779)	1.801 (1.621-2.000)	1.101 (1.044-1.163)	1.201 (1.139-1.266)	1.301 (1.235-1.371)	1.402 (1.333-1.476)	1.500 (1.426-1.578)	1.601 (1.523-1.682)	1.801 (1.715-1.891)
4 clusters per arm														
2	1.111 (0.860-1.440)	1.211 (0.941-1.563)	1.312 (1.024-1.685)	1.416 (1.111-1.811)	1.510 (1.187-1.925)	1.614 (1.273-2.054)	1.827 (1.451-2.307)	1.102 (0.977-1.245)	1.206 (1.070-1.359)	1.299 (1.158-1.459)	1.401 (1.251-1.569)	1.504 (1.345-1.683)	1.604 (1.436-1.793)	1.801 (1.616-2.007)
3	1.107 (0.900-1.364)	1.208 (0.986-1.482)	1.316 (1.079-1.608)	1.407 (1.156-1.716)	1.509 (1.243-1.835)	1.616 (1.334-1.959)	1.816 (1.507-2.192)	1.101 (0.999-1.214)	1.201 (1.093-1.320)	1.302 (1.188-1.427)	1.403 (1.281-1.538)	1.501 (1.372-1.643)	1.603 (1.467-1.752)	1.802 (1.653-1.966)
4	1.101 (0.922-1.316)	1.198 (1.007-1.426)	1.301 (1.097-1.545)	1.405 (1.187-1.665)	1.502 (1.272-1.775)	1.602 (1.361-1.887)	1.806 (1.538-2.122)	1.100 (1.012-1.196)	1.202 (1.109-1.304)	1.302 (1.203-1.410)	1.402 (1.296-1.518)	1.503 (1.392-1.624)	1.601 (1.483-1.728)	1.804 (1.675-1.943)
6	1.101-(0.953-1.273)	1.201 (1.043-1.383)	1.300 (1.132-1.495)	1.406 (1.226-1.612)	1.506 (1.315)	1.606 (1.405-1.836)	1.807 (1.587-2.059)	1.099 (1.028-1.175)	1.202 (1.125-1.284)	1.299 (1.218-1.384)	1.400 (1.314-1.492)	1.500 (1.409-1.597)	1.600 (1.504-1.702)	1.803 (1.697-1.915)
12	1.099 (0.992-1.217)	1.199 (1.086-1.325)	1.300 (1.179-1.433)	1.399 (1.271-1.540)	1.499 (1.363-1.648)	1.600 (1.457-1.758)	1.800 (1.643-1.972)	1.101 (1.050-1.154)	1.201 (1.147-1.258)	1.299 (1.242-1.360)	1.402 (1.341-1.466)	1.499 (1.435-1.566)	1.600 (1.531-1.671)	1.801 (1.726-1.879)
5clusters per arm														
2	1.101 (0.875-1.387)	1.201 (0.959-1.508)	1.305 (1.046-1.631)	1.400 (1.127-1.743)	1.505 (1.215-1.869)	1.601 (1.294-1.982)	1.809 (1.471-2.229)	1.101 (0.989-1.228)	1.204 (1.083-1.339)	1.301 (1.173-1.445)	1.403 (1.268-1.554)	1.502 (1.357-1.663)	1.604 (1.452-1.772)	1.800 (1.634-1.983)
3	1.103 (0.916-1.331)	1.206 (1.005-1.449)	1.306 (1.092-1.562)	1.412 (1.184-1.684)	1.513 (1.272-1.801)	1.610 (1.356-1.914)	1.815 (1.534-2.149)	1.103 (1.012-1.203)	1.200 (1.103-1.306)	1.299 (1.196-1.412)	1.401 (1.291-1.520)	1.503 (1.386-1.631)	1.600 (1.478-1.732)	1.800 (1.667-1.945)
4	1.102 (0.940-1.293)	1.203 (1.029-1.407)	1.303 (1.118-1.519)	1.404 (1.207-1.635)	1.505 (1.297-1.747)	1.608 (1.389-1.863)	1.808 (1.566-2.089)	1.101 (1.022-1.186)	1.200 (1.116-1.291)	1.299 (1.209-1.396)	1.401 (1.306-1.504)	1.500 (1.400-1.607)	1.599 (1.494-1.712)	1.799 (1.683-1.923)
6	1.101 (0.968-1.254)	1.205 (1.061-1.369)	1.301 (1.148-1.475)	1.400 (1.238-1.584)	1.504 (1.334-1.696)	1.605 (1.426-1.808)	1.809 (1.611-2.033)	1.098 (1.034-1.167)	1.204 (1.136-1.277)	1.300 (1.227-1.378)	1.400 (1.323-1.482)	1.503 (1.421-1.589)	1.603 (1.516-1.695)	1.799 (1.704-1.899)
12	1.100 (1.004-1.205)	1.200 (1.097-1.312)	1.303 (1.193-1.424)	1.403 (1.287-1.529)	1.501 (1.380-1.634)	1.603 (1.474-1.743)	1.803 (1.662-1.957)	1.098 (1.053-1.145)	1.201 (1.152-1.251)	1.301 (1.250-1.355)	1.400 (1.345-1.456)	1.500 (1.443-1.561)	1.599 (1.538-1.663)	1.801 (1.735-1.871)

	Low incidence outcomes- Mean outcome measure in control clusters of 15 cases per month							High incidence outcomes- Mean outcome measure in control clusters of 70 cases per month						
	Effect size (incidence rate ratio)							*Effect size (incidence rate ratio)*						
	1.1	*1.2*	*1.3*	*1.4*	*1.5*	*1.6*	*1.8*	*1.1*	*1.2*	*1.3*	*1.4*	*1.5*	*1.6*	*1.8*
6clusters per arm														
2	1.106 (0.894-1.372)	1.203 (0.976-1.486)	1.307 (1.067-1.604)	1.403 (1.147-1.719)	1.509 (1.240-1.839)	1.609 (1.325-1.956)	1.807 (1.495-2.187)	1.103 (0.999-1.218)	1.203 (1.091-1.326)	1.300 (1.184-1.429)	1.399 (1.275-1.536)	1.501 (1.369-1.647)	1,605 (1.467-1.760)	1.804 (1.650-1.972)
3	1.105 (0.933-1.311)	1.206 (1.022-1.426)	1.308 (1.112-1.541)	1.412 (1.203-1.659)	1.508 (1.288-1.768)	1.611 (1.379-1.884)	1.808 (1.553-2.107)	1.100 (1.017-1.190)	1.200 (1.110-1.298)	1.300 (1.206-1.403)	1.397 (1.297-1.505)	1.503 (1.397-1.617)	1.602 (1.490-1.722)	1.802 (1.680-1.935)
4	1.097 (0.949-1.270)	1.199 (1.040-1.384)	1.302 (1.132-1.499)	1.402 (1.222-1.611)	1.500 (1.309-1.720)	1.600 (1.399-1.831)	1.802 (1.579-2.058)	1.102 (1.030-1.179)	1.201 (1.124-1.282)	1.299 (1.217-1.387)	1.402 (1.315-1.495)	1.502 (1.411-1.599)	1.599 (1.502-1.702)	1.801 (1.696-1.914)
6	1.100 (0.978-1.239)	1.204 (1.073-1.351)	1.301 (1.161-1.458)	1.401 (1.253-1.567)	1.500 (1.344-1.675)	1.604 (1.439-1.789)	1.803 (1.623-2.004)	1.102 (1.043-1.164)	1.199 (1.137-1.266)	1.301 (1.233-1.372)	1.401 (1.331-1.476)	1.500 (1.425-1.578)	1.602 (1.523-1.684)	1.799 (1.712-1.889)
12	1.101 (1.014-1.196)	1.201 (1.107-1.302)	1.300 (1.200-1.408)	1.399 (1.293-1.514)	1.501 (1.390-1.622)	1.601 (1.483-1.728)	1.801 (1.672-1.940)	1.100 (1.058-1.143)	1.201 (1.156-1.247)	1.300 (1.253-1.349)	1.400 (1.351-1.452)	1.500 (1.447-1.554)	1.600 (1.544-1.657)	1.801 (1.740-1.864)
9clusters per arm														
2	1.110 (0.931-1.324)	1.209 (1.018-1.438)	1.309 (1.106-1.550)	1.408 (1.193-1.664)	1.508 (1.280-1.779)	1.613 (1.373-1.898)	1.811 (1.546-2.123)	1.101 (1.017-1.193)	1.199 (1.109-1.297)	1.301 (1.206-1.403)	1.401 (1.299-1.511)	1.503 (1.395-1.620)	1.600 (1.487-1.722)	1.799 (1.674-1.934)
3	1.104 (0.961-1.269)	1.205 (1.052-1.381)	1.308 (1.144-1.497)	1.408 (1.234-1.607)	1.508 (1.325-1.718)	1.608 (1.416-1.826)	1.807 (1.596-2.046)	1.102 (1.034-1.175)	1.201 (1.128-1.279)	1.302 (1.224-1.385)	1.403 (1.321-1.491)	1.501 (1.413-1.594)	1.602 (1.510-1.700)	1.803 (1.702-1.911)
4	1.102 (0.979-1.240)	1.201 (1.070-1.349)	1.304 (1.164-1.461)	1.399 (1.251-1.565)	1.501 (1.345-1.676)	1.605 (1.439-1.790)	1.800 (1.618-2.004)	1.100 (1.041-1.162)	1.201 (1.138-1.268)	1.301 (1.234-1.372)	1.401 (1.330-1.476)	1.501 (1.426-1.580)	1.601 (1.522-1.684)	1.799 (1.713-1.890)
6	1.101 (1.00-1.213)	1.200 (1.092-1.320)	1.302 (1.186-1.429)	1.404 (1.281-1.538)	1.503 (1.373-1.645)	1.603 (1.466-1.752)	1.806 (1.656-1.969)	1.102 (1.054-1.152)	1.202 (1.151-1.255)	1.298 (1.243-1.355)	1.401 (1.343-1.461)	1.500 (1.438-1.564)	1,600 (1.535-1.668)	1.800 (1.729-1.873)
12	1.099 (1.027-1.176)	1.200 (1.123-1.282)	1.300 (1.218-1.388)	1.400 (1.313-1.493)	1.502 (1.410-1.600)	1.603 (1.506-1.705)	1.800 (1.695-1.913)	1.100 (1.066-1.135)	1.200 (1.164-1.238)	1.300 (1.261-1.340)	1.399 (1.359-1.441)	1.501 (1.457-1.545)	1.601 (1.556-1.648)	1.800 (1.750-1.851)
12 clusters per arm**														
2	-	1.202 (1.038-1.392)	1.303 (1.129-1.506)	1.406 (1.220-1.621)	1.510 (1.313-1.737)	1.609 (1.402-1.847)	1.809 (1.582-2.069)	1.100 (1.027-1.180)	1.200 (1.121-1.284)	1.300 (1.217-1.389)	1.401 (1.313-1.496)	1.503 (1.409-1.604)	1.602 (1.503-1.708)	1.802 (1.694-1.916)
3	-	1.200 (1.068-1.350)	1.301 (1.160-1.460)	1.402 (1.252-1.571)	1.501 (1.342-1.679)	1.603 (1.435-1.790)	1.805 (1.620-2.011)	1.100 (1.040-1.162)	1.199 (1.136-1.266)	1.300 (1.232-1.371)	1.400 (1.327-1.476)	1.500 (1.424-1.579)	1.601 (1.521-1.685)	1.803 (1.715-1.896)
4	-	1.201 (1.087-1.328)	1.302 (1.180-1.436)	1.401 (1.271-1.545)	1.501 (1.364-1.652)	1.601 (1.456-1.760)	1.803 (1.643-1.980)	1.099 (1.048-1.152)	1.201 (1.146-1.259)	1.300 (1.242-1.361)	1.399 (1.337-1.464)	1.500 (1.435-1.568)	1.600 (1.531-1.672)	1.800 (1.724-1.879)
6	-	1.202 (1.108-1.305)	1.304 (1.203-1.413)	1.403 (1.297-1.519)	1.504 (1.392-1.626)	1.603 (1.485-1.732)	1.798 (1.668-1.937)	1.100 (1.059-1.144)	1.200 (1.155-1.246)	1.300 (1.253-1.349)	1.400 (1.350-1.452)	1.501 (1.448-1.555)	1.603 (1.547-1.661)	1.802 (1.741-1.866)
12	-	1.201 (1.134-1.271)	1.300 (1.229)	1.400 (1.325-1.480)	1.500 (1.421-1.584)	1.601 (1.517-1.690)	1.799 (1.707-1.896)	1.100 (1.071-1.130)	1.201 (1.170-1.234)	1.301 (1.267-1.335)	1.401 (1.366-1.438)	1.500 (1.463-1.539)	1.601 (1.561-1.641)	1.801 (1.757-1.845)

* Incidence rate ratios (1.1 to 1.8) on top row were specified and therefore were fixed, the incidence rate ratios with 95% confidence intervals in brackets for the measurements times of 2 – 12 were estimated using the 1000 simulations for accessing accuracy of the estimation of the specified intervention effect sizes based on the 95% confidence intervals.

* *No convergence was achieved at the effect size of 10%, hence no estimates are given

Table 11: The effect of the number of groups, time points and the effect size on how accurately the effect sizes were estimated based on 5% and 95% percentiles of the estimated effect size after 1000 simulated data sets.

	Low incidence outcomes- Mean outcome measure in control clusters of 15 cases per month							High incidence outcomes- Mean outcome measure in control clusters of 70 cases per month						
	*Effect size (incidence rate ratio)**							*Effect size (incidence rate ratio)**						
	1.1	*1.2*	*1.3*	*1.4*	*1.5*	*1.6*	*1.8*	*1.1*	*1.2*	*1.3*	*1.4*	*1.5*	*1.6*	*1.8*
3 clusters per arm														
2	1.114 (0.861-1.423)	1.214 (0.967-1.512)	1.319 (1.048-1.655)	1.428 (1.151-1.809)	1.519 (1.213-1.895)	1.626 (1.290-2.020)	1.823 (1.462-2.267)	1.103 (0.987-1.235)	1.203 (1.076-1.335)	1.301 (1.168-1.440)	1.403 (1.260-1.559)	1.505 (1.355-1.665)	1.599 (1.441-1.765)	1.802 (1.625-1.986)
3	1.109 (0.914-1.346)	1.206 (0.993-1.441)	1.312 (1.099-1.562)	1.413 (1.179-1.678)	1.512 (1.269-1.799)	1.615 (1.362-2.920)	1.811 (1.523-2.150)	1.102 (1.005-1.205)	1.200 (1.100-1.310)	1.303 (1.195-1.416)	1.401 (1.278-1.522)	1.501 (1.374-1.631)	1.603 (1.480-1.742)	1.797 (1.667-1.945)
4	1.101 (0.914-1.299)	1.201 (1.020-1.407)	1.115 (1.115-1.515)	1.401 (1.195-1.633)	1.508 (1.297-1.770)	1.610 (1.388-1.870)	1.807 (1.557-2.108)	1.102 (1.021-1.186)	1.200 (1.112-1.289)	1.303 (1.208-1.405)	1.402 (1.304-1.506)	1.503 (1.404-1.609)	1.602 (1.494-1.717)	1.801 (1.681-1.927)
6	1.101 (0.966-1.252)	1.199 (1.054-1.371)	1.298 (1.137-1.472)	1.399 (1.225-1.593)	1.499 (1.318-1.700)	1.594 (1.405-1.801)	1.800 (1.593-2.012)	1.101 (1.032-1.169)	1.202 (1.133-1.277)	1.302 (1.223-1.383)	1.398 (1.318-1.485)	1.503 (1.422-1.596)	1.599 (1.151-1.670)	1.802 (1.704-1.905)
12	1.102 (0.999-1.203)	1.200 (1.093-1.304)	1.299 (1.186-1.417)	1.398 (1.285-1.522)	1.500 (1.377-1.632)	1.598 (1.477-1.737)	1.801 (1.663-21.95)	1.101 (1.056-1.149)	1.201 (1.154-1.254)	1.301 (1.249-1.359)	1.402 (1.346-1.462)	1.500 (1.436-1.560)	1.601 (1.537-1.672)	1.801 (1.733-1.873)
4 clusters per arm														
2	1.111 (0.895-1.365)	1.211 (0.972-1.475)	1.312 (1.060-1.600)	1.416 (1.151-1.728)	1.510 (1.234-1.833)	1.614 (1.328-1.941)	1.827 (1.509-2.192)	1.102 (1.000-1.212)	1.206 (1.096-1.322)	1.299 (1.183-1.421)	1.401 (1.285-1.533)	1.504 (1.374-1.644)	1.604 (1.464-1.747)	1.801 (1.654-1.961)
3	1.107 (0.935-1.305)	1.208 (1.025-1.420)	1.316 (1.108-1.551)	1.407 (1.182-1.652)	1.509 (1.289-1.762)	1.616 (1.388-1.892)	1.816 (1.559-2.111)	1.101 (1.017-1.190)	1.201 (1.111-1.297)	1.302 (1.205-1.401)	1.403 (1.302-1.516)	1.501 (1.394-1.613)	1.603 (1.493-1.714)	1.802 (1.680-1.928)
4	1.101 (0.949-1.257)	1.198 (1.047-1.369)	1.301 (1.131-1.484)	1.405 (1.228-1.611)	1.502 (1.317-1.700)	1.602 (1.407-1.826)	1.806 (1.595-2.059)	1.100 (1.029-1.177)	1.202 (1.126-1.282)	1.302 (1.220-1.383)	1.402 (1.317-1.493)	1.503 (1.415-1.594)	1.601 (1.495-1.708)	1.804 (1.698-1.910)
6	1.101 (0.972-1.229)	1.201 (1.065-1.344)	1.300 (1.152-1.452)	1.406 (1.262-1.566)	1.506 (1.355-1.682)	1.606 (1.451-1.796)	1.807 (1.632-2.003)	1.099 (1.041-1.158)	1.202 (1.142-1.266)	1.299 (1.234-1.369)	1.400 (1.332-1.472)	1.500 (1.427-1.576)	1.600 (1.520-1.676)	1.803 (1.720-1.891)
12	1.099 (1.013-1.193)	1.199 (1.106-1.298)	1.300 (1.197-1.404)	1.399 (1.299-1.513)	1.499 (1.389-1.612)	1.600 (1.481-1.729)	1.800 (1.677-1.943)	1.101 (1.060-1.142)	1.201 (1.156-1.245)	1.299 (1.253-1.347)	1.402 (1.351-1.454)	1.499 (1.446-1.551)	1.600 (1.545-1.657)	1.801 (1.737-1.866)
5 clusters per arm														
2	1.101 (0.913-1.309)	1.201 (0.904-1.434)	1.305 (1.091-1.557)	1.400 (1.171-1.675)	1.505 (1.265-1.781)	1.601 (1.352-1.894)	1.809 (1.530-2.133)	1.101 (1.009-1.201)	1.204 (1.102-1.300)	1.301 (1.194-1.415)	1.403 (1.288-1.525)	1.502 (1.385-1.632)	1.604 (1.482-1.730)	1.800 (1.666-1.937)
3	1.103 (0.952-1.280)	1.206 (1.050-1.389)	1.306 (1.135-1.498)	1.412 (1.226-1.613)	1.513 (1.322-1.719)	1.610 (1.391-1.834)	1.815 (1.579-2.061)	1.103 (1.027-1.177)	1.200 (1.124-1.284)	1.299 (1.210-1.388)	1.401 (1.310-1.494)	1.503 (1.408-1.606)	1.600 (1.500-1.700)	1.800 (1.688-1.908)
4	1.102 (0.971-1.253)	1.203 (1.061-1.360)	1.303 (1.152-1.471)	1.404 (1.246-1.580)	1.505 (1.331-1.690)	1.608 (1.429-1.800)	1.808 (1.607-2.019)	1.101 (1.036-1.174)	1.200 (1.129-1.271)	1.299 (1.224-1.370)	1.401 (1.327-1.482)	1.500 (1.419-1.581)	1.599 (1.517-1.686)	1.799 (1.706-1.902)
6	1.101 (0.995-1.223)	1.205 (1.083-1.331)	1.301 (1.173-1.439)	1.400 (1.261-1.556)	1.504 (1.371-1.658)	1.605 (1.463-1.757)	1.809 (1.657-1.983)	1.098 (1.047-1.153)	1.204 (1.148-1.259)	1.300 (1.240-1.358)	1.400 (1.332-1.466)	1.503 (1.435-1.572)	1.603 (1.534-1.673)	1.799 (1.726-1.884)
12	1.100 (1.026-1.187)	1.200 (1.119-1.284)	1.303 (1.218-1.399)	1.403 (1.311-1.502)	1.501 (1.403-1.604)	1.603 (1.501-1.711)	1.803 (1.692-1.924)	1.098 (1.061-1.136)	1.201 (1.160-1.238)	1.301 (1.263-1.342)	1.400 (1.356-1.441)	1.500 (1.454-1.551)	1.599 (1.551-1.650)	1.801 (1.750-1.855)
6 clusters per arm														
2	1.106 (0.930-1.313)	1.203 (1.007-1.400)	1.307 (1.102-1.538)	1.403 (1.189-1.649)	1.509 (1.189-1.649)	1.609 (1.371-1.866)	1.807 (1.545-2.108)	1.103 (1.021-1.185)	1.203 (1.112-1.296)	1.300 (1.205-1.402)	1.399 (1.297-1.502)	1.501 (1.390-1.624)	1,605 (1.489-1.728)	1.804 (1.678-1.939)
3	1.105 (0.956-1.266)	1.206 (1.046-1.368)	1.308 (1.138-1.488)	1.412 (1.237-1.606)	1.508 (1.237-1.606)	1.611 (1.420-1.822)	1.808 (1.587-2.044)	1.100 (1.031-1.170)	1.200 (1.124-1.279)	1.300 (1.221-1.388)	1.397 (1.319-1.481)	1.503 (1.420-1.592)	1.602 (1.508-1.699)	1.802 (1.703-1.902)
4	1.097 (0.973-1.234)	1.199 (1.062-1.343)	1.302 (1.156-1.465)	1.402 (1.255-1.564)	1.500 (1.255-1.564)	1.600 (1.433-1.778)	1.802 (1.614-1.995)	1.102 (1.046-1.160)	1.201 (1.140-1.264)	1.299 (1.235-1.365)	1.402 (1.329-1.472)	1.502 (1.429-1.577)	1.599 (1.520-1.681)	1.801 (1.717-1.887)
6	1.100 (0.996-1.204)	1.204 (1.099-1.319)	1.301 (1.190-1.424)	1.401 (1.282-1.533)	1.500 (1.282-1.533)	1.604 (1.470-1.754)	1.803 (1.660-1.962)	1.102 (1.051-1.150)	1.199 (1.149-1.252)	1.301 (1.242-1.359)	1.401 (1.345-1.459)	1.500 (1.438-1.561)	1.602 (1.537-1.667)	1.799 (1.729-1.867)
12	1.101 (1.028-1.172)	1.201 (1.128-1.277)	1.300 (1.214-1.385)	1.399 (1.316-1.490)	1.501 (1.316-1.490)	1.601 (1.50-1.706)	1.801 (1697-1.913)	1.100 (1.069-1.134)	1.201 (1.164-1.236)	1.300 (1.263-1.340)	1.400 (1.360-1.440)	1.500 (1.455-1.543)	1.600 (1.554-1.646)	1.801 (1.752-1.848)

	Low incidence outcomes- Mean outcome measure in control clusters of 15 cases per month							High incidence outcomes- Mean outcome measure in control clusters of 70 cases per month						
	Effect size (incidence rate ratio)							*Effect size (incidence rate ratio)*						
	1.1	*1.2*	*1.3*	*1.4*	*1.5*	*1.6*	*1.8*	*1.1*	*1.2*	*1.3*	*1.4*	*1.5*	*1.6*	*1.8*
9 clusters per arm														
2	1.110 (0.964-1.282)	1.209 (1.055-1.389)	1.309 (1.140-1.493)	1.408 (1.239-1.603)	1.508 (1.320-1.722)	1.613 (1.418-1.831)	1.811 (1.594-2.048)	1.101 (1.034-1.178)	1.199 (1.128-1.276)	1.301 (1.221-1.386)	1.401 (1.318-1.483)	1.503 (1.411-1.602)	1.600 (1.506-1.693)	1.799 (1.170-1.901)
3	1.104 (0.984-1.233)	1.205 (1.078-1.346)	1.308 (1.179-1.447)	1.408 (1.261-1.558)	1.508 (1.354-1.675)	1.608 (1.448-1.781)	1.807 (1.631-1.992)	1.102 (1.047-1.163)	1.201 (1.145-1.263)	1.302 (1.234-1.369)	1.403 (1.337-1.475)	1.501 (1.425-1.577)	1.602 (1.527-1.682)	1.803 (1.719-1.886)
4	1.102 (0.997-1.208)	1.201 (1.091-1.316)	1.304 (1.185-1.422)	1.399 (1.276-1.537)	1.501 (1.371-1.637)	1.605 (1.474-1.750)	1.800 (1.642-1.960)	1.100 (1.051-1.150)	1.201 (1.148-1.255)	1.301 (1.246-1.361)	1.401 (1.343-1.463)	1.501 (1.440-1.566)	1.601 (1.536-1.670)	1.799 (1.731-1.878)
6	1.101 (1.019-1.186)	1.200 (1116-1.292)	1.302 (1.207-1.400)	1.404 (1.305-1.506)	1.503 (1.402-1.615)	1.603 (1.492-1.720)	1.806 (1.691-1.927)	1.102 (1.064-1.142)	1.202 (1.159-1.245)	1.298 (1.254-1.342)	1.401 (1.353-1.449)	1.500 (1.451-1.551)	1,600 (1.548-1.650)	1.800 (1.742-1.859)
12	1.099 (1.042-1.161)	1.200 (1.137-1.265)	1.300 (1.232-1.371)	1.400 (1.325-1.471)	1.502 (1.427-1.579)	1.603 (1.527-1.681)	1.800 (1.719-1.891)	1.100 (1.073-1.127)	1.200 (1.172-1.230)	1.300 (1.268-1.331)	1.399 (1.364-1.433)	1.501 (1.466-1.536)	1.601 (1.565-1.639)	1.800 (1.761-1.839)
*12 clusters per arm***														
2	-	1.202 (1.069-1.348)	1.303 (1.173-1.453)	1.406 (1.257-1.570)	1.510 (1.345-1.687)	1.609 (1.444-1.787)	1.809 (1.628-2.012)	1.100 (1.038-1.167)	1.200 (1.132-1.267)	1.300 (1.232-1.366)	1.401 (1.328-1.474)	1.503 (1.430-1.586)	1.602 (1.522-1.685)	1.802 (1.720-1.890)
3	-	1.200 (1.096-1.320)	1.301 (1.186-1.424)	1.402 (1.279-1.533)	1.501 (1.375-1.636)	1.603 (1.465-1.744)	1.805 (1.659-1.959)	1.100 (1.051-1.148)	1.199 (1.149-1.250)	1.300 (1.246-1.356)	1.400 (1.341-1.464)	1.500 (1.438-1.565)	1.601 (1.535-1.666)	1.803 (1.732-1.875)
4	-	1.201 (1.110-1.298)	1.302 (1.203-1.395)	1.401 (1.299-1.508)	1.501 (1.395-1.611)	1.601 (1.483-1.715)	1.803 (1.665-1.933)	1.099 (1.058-1.142)	1.201 (1.157-1.248)	1.300 (1.252-1.350)	1.399 (1.346-1.439)	1.500 (1.447-1.556)	1.600 (1.544-1.658)	1.800 (1.737-1.860)
6	-	1.202 (1.127-1.288)	1.304 (1.221-1.388)	1.403 (1.316-1.490)	1.504 (1.412-1.604)	1.603 (1.509-1.710)	1.798 (1.691-1.904)	1.100 (1.068-1.135)	1.200 (1.163-1.237)	1.300 (1.262-1.339)	1.400 (1.361-1.442)	1.501 (1.457-1.544)	1.603 (1.557-1.648)	1.802 (1.754-1.854)
12	-	1.201 (1.144-1.257)	1.300 (1.243-1.357)	1.400 (1.340-1.460)	1.500 (1.435-1.564)	1.601 (1.528-1.672)	1.799 (1.723-1.877)	1.100 (1.077-1.126)	1.201 (1.176-1.227)	1.301 (1.274-1.327)	1.401 (1.370-1.430)	1.500 (1.471-1.532)	1.601 (1.569-1.634)	1.801 (1.766-1.836)
ICC=0.00154														

* Incidence rate ratios (1.1 to 1.8) on top row were specified and therefore were fixed, the incidence rate ratios with 95% confidence intervals in brackets for the measurements times of 2 – 12 were estimated using the 1000 simulations for accessing accuracy of the estimation of the specified intervention effect sizes based on the 5% and 95% percentiles.

** No convergence was achieved at the effect size of 10%, hence no estimates are given

Table 12: The effect of the number of groups and repeated measurements on the accuracy of the estimated effect size based on the mean length of 95% confidence interval, based on 1000 simulated data sets

	Low incidence outcomes- Mean outcome measure in control clusters of 15 cases per month							High incidence outcomes- Mean outcome measure in control clusters of 70 cases per month						
				Effect size (incidence rate ratio)							Effect size (incidence rate ratio)			
	1.1	1.2	1.3	1.4	1.5	1.6	1.8	1.1	1.2	1.3	1.4	1.5	1.6	1.8
3 clusters per arm														
2	0.682	0.729	0.778	0.830	0.870	0.919	1.00	0.305	0.332	0.350	0.370	0.387	0.411	0.451
3	0.539	0.574	0.612	0.650	0.685	0.725	0.800	0.245	0.263	0.278	0.293	0.315	0.327	0.360
4	0.460	0.492	0.524	0.553	0.587	0.618	0.677	0.212	0.223	0.241	0.253	0.270	0.283	0.311
6	0.368	0.391	0.418	0.443	0.493	0.492	0.541	0.171	0.184	0.194	0.202	0.218	0.227	0.252
12	0.257	0.275	0.293	0.310	0.328	0.343	0.379	0.119	0.128	0.136	0.143	0.152	0.159	0.176
4 clusters per arm														
2	0.580	0.622	0.661	0.700	0.738	0.781	0.856	0.268	0.288	0.301	0.318	0.338	0.357	0.391
3	0.464	0.496	0.529	0.561	0.592	0.625	0.685	0.215	0.227	0.239	0.257	0.271	0.285	0.313
4	0.395	0.419	0.447	0.478	0.503	0.526	0.583	0.183	0.195	0.208	0.221	0.231	0.245	0.268
6	0.319	0.341	0.363	0.387	0.409	0.431	0.472	0.147	0.159	0.166	0.178	0.188	0.198	0.217
12	0.224	0.240	0.254	0.269	0.285	0.300	0.329	0.103	0.111	0.118	0.125	0.132	0.140	0.153
5 clusters per arm														
2	0.512	0.549	0.585	0.616	0.654	0.687	0.758	0.239	0.257	0.272	0.285	0.306	0.320	0.349
3	0.415	0.444	0.471	0.500	0.529	0.558	0.615	0.191	0.203	0.216	0.229	0.245	0.255	0.278
4	0.354	0.378	0.401	0.427	0.449	0.476	0.523	0.164	0.175	0.186	0.199	0.207	0.218	0.341
6	0.286	0.308	0.327	0.346	0.362	0.382	0.422	0.133	0.141	0.151	0.159	0.168	0.179	0.195
12	0.200	0.214	0.232	0.243	0.255	0.269	0.295	0.093	0.99	0.105	0.111	0.118	0.124	0.136
6 clusters per arm														
2	0.478	0.510	0.537	0.572	0.600	0.631	0.692	0.219	0.235	0245	0.261	0.278	0.289	0.321
3	0.378	0.404	0.429	0.455	0.481	0.505	0.554	0.173	0.187	0.197	0.208	0.220	0.231	0.255
4	0.322	0.345	0.367	0.389	0.411	0.431	0.478	0.149	0.158	0.170	0.180	0.188	0.199	0.218
6	0.261	0.278	0.297	0.313	0.331	0.350	0.382	0.120	0.129	0.138	0.145	0.153	0.161	0.177
12	0.183	0.195	0.208	0.220	0.232	0.245	0.268	0.085	0.091	0.096	0.101	0.107	0.113	0.124
9 clusters per arm														
2	0.393	0.419	0.444	0.471	0.499	0.525	0.576	0.176	0.188	0.198	0.212	0.225	0.235	0.260
3	0.308	0.329	0.353	0.373	0.393	0.409	0.450	0.141	0.150	0.161	0.170	0.181	0.190	0.209
4	0.261	0.279	0.297	0.314	0.332	0.351	0.386	0.121	0.130	0.138	0.147	0.154	0.162	0.177
6	0.213	0.228	0.243	0.258	0.272	0.287	0.314	0.098	0.104	0.112	0.118	0.126	0.133	0.144
12	0.149	0.159	0.170	0.180	0.190	0.199	0.218	0.069	0.074	0.079	0.082	0.088	0.092	0.102
*12 clusters arm**														
2	-	0.354	0.377	0.401	0.423	0.445	0.487	0.153	0.162	0.172	0.183	0.194	0.205	0.222
3	-	0.282	0.300	0.319	0.337	0.355	0.391	0.122	0.130	0.139	0.149	0.155	0.164	0.180
4	-	0.241	0.256	0.274	0.289	0.304	0.337	0.104	0.112	0.119	0.126	0.133	0.141	0.154
6	-	0.197	0.209	0.222	0.234	0.247	0.270	0.085	0.091	0.096	0.102	0.108	0.114	0.125
12	-	0.137	0.160	0.155	0.163	0.172	0.190	0.059	0.064	0.068	0.072	0.076	0.080	0.088

ICC=0.00154

* No convergence was achieved at the effect size of 10%, hence no estimates are given

Table 13:The effect of the number of clusters, time points and the effect size on the percentage relative bias of how the estimated effect size differed from the fixed effect size and non-coverage of the asymptotic 95% confidence interval of the estimated effect size.

	Low incidence- Mean count in control clusters of 15 cases per month							High incidence - Mean count in control clusters of 70 cases per month						
	Bias (% Relative bias ($\frac{\hat{\theta}-\theta}{\theta}$**x 100**))							Bias (% Relative bias ($\frac{\hat{\theta}-\theta}{\theta}$**x 100**))						
	Effect size (incidence rate ratio)							Effect size (incidence rate ratio)						
	1.1	1.2	1.3	1.4	1.5	1.6	1.8	1.1	1.2	1.3	1.4	1.5	1.6	1.8
3 clusters per arm														
2	0.014 (1.3)	0.014 (1.1)	0.019 (1.43)	0.028 (2.0)	0.019 (1.26)	0.026 (1.6)	0.023 (1.25)	0.003 (0.27)	0.003 (0.21)	0.001 (0.08)	0.003 (0.21)	0.005 (0.33)	-0.001 (-0.05)	0.002 (0.09)
3	0.009 (0.82)	0.006 (0.49)	0.012 (0.91)	0.013 (0.93)	0.012 (0.82)	0.015 (0.91)	0.011 (0.6)	0.002 (0.17)	0.0 (-0.03)	0.003 (0.21)	0.001 (0.10)	0.001 (0.09)	0.003 (0.17)	-0.003 (-0.14)
4	0.001 (0.11)	0.001 (0.08)	0.004 (0.33)	0.001 (0.04)	0.008 (0.54)	0.01 (0.61)	0.007 (0.39)	0.002 (0.20)	0.0 (0.04)	0.003 (0.26)	0.002 (0.11)	0.003 (0.19)	0.002 (0.12)	0.001 (0.05)
6	0.001 (0.12)	-0.001 (-0.09)	-0.002 (-0.15)	-0.001 (-0.05)	-0.001 (0.0)	-0.006 (-0.38)	0.0 (-0.02)	0.001 (0.09)	0.002 (0.18)	0.002 (0.17)	0.0 (-0.13)	0.003 (0.23)	-0.001 (-0.04)	0.002 (0.09)
12	0.002 (0.17)	0.0 (0.01)	-0.001 (-0.07)	-0.002 (-0.13)	0.0 (0.0)	-0.002 (-0.011)	0.001 (0.03)	0.001 (0.13)	0.001 (0.07)	0.001 (0.09)	0.002 (0.17)	0.0 (0.01)	0.001 (0.05)	0.001 (0.05)
4 clusters per arm														
2	0.011 (1.0)	0.011 (0.95)	0.12 (0.90)	0.016 (1.16)	0.010 (0.67)	0.014 (0.90)	0.027 (1.53)	0.002 (0.19)	0.006 (0.47)	0.0 (-0.05)	0.001 (0.05)	0.004 (0.29)	0.004 (0.27)	0.001 (0.04)
3	0.007 (0.63)	0.008 (0.66)	0.016 (1.24)	0.007 (0.53)	0.009 (0.62)	0.016 (0.98)	0.016 (0.90)	0.001 (0.10)	0.001 (0.10)	0.002 (0.13)	0.003 (0.24)	0.001 (0.08)	0.003 (0.21)	0.002 (0.14)
4	0.001 (0.07)	-0.002 (-0.18)	0.001 (0.11)	0.005 (0.36)	0.002 (0.12)	0.002 (0.12)	0.006 (0.34)	0.0 (0.01)	0.002 (0.19)	0.002 (0.19)	0.002 (0.18)	0.003 (0.22)	0.001 (0.06)	0.004 (0.23)
6	0.001 (0.13)	0.001 (0.06)	0.0 (0.03)	0.006 (0.40)	0.006 (0.38)	0.006 (0.35)	0.007 (0.38)	-0.001 (-0.07)	0.002 (0.13)	-0.001 (-0.1)	0.0 (0.0)	0.0 (-0.01)	0.0 (0.0)	0.003 (0.16)
12	-0.001 (-0.12)	-0.001 (0.0)	0.0 (-0.02)	-0.001 (-0.07)	-0.001 (-0.07)	0.0 (0.02)	0.0 (0.01)	0.08 (0.07)	0.001 (0.09)	-0.001 (-0.06)	0.002 (0.14)	-0.001 (-0.06)	0.0 (-0.02)	0.001 (0.05)
5clusters per arm														
2	0.001 (0.06)	0.001 (0.11)	0.005 (0.35)	0.0 (0.0)	0.005 (0.33)	0.001 (0.04)	0.009 (0.50)	0.001 (0.12)	0.004 (0.31)	0.001 (0.11)	0.003 (0.24)	0.002 (0.13)	0.004 (0.23)	0.0 (-0.02)
3	0.003 (0.31)	0.006 (0.53)	0.006 (0.43)	0.012 (0.83)	0.013 (0.88)	0.010 (0.63)	0.015 (0.83)	0.003 (0.27)	0.0 (0.03)	-0.001 (-0.07)	0.001 (0.05)	0.003 (0.002)	0.0 (-0.01)	0.0 (0.0)
4	0.002 (0.20)	0.003 (0.25)	0.003 (0.20)	0.004 (0.32)	0.005 (0.33)	0.008 (0.52)	0.008 (0.45)	0.001 (0.11)	0.0 (0.03)	-0.001 (-0.09)	0.001 (0.09)	0.0 (-0.01)	-0.001 (-0.07)	-0.001 (-0.07)
6	0.001 (0.13)	0.005 (0.40)	0.001 (0.10)	0.0 (0.03)	0.004 (0.25)	0.005 (0.34)	0.009 (0.52)	-0.002 (-0.14)	0.004 (0.33)	0.0 (0.03)	0.0 (0.03)	0.003 (0.17)	0.003 (0.19)	-0.001 (-0.07)
12	0.0 (-0.02)	0.0 (-0.03)	0.003 (0.25)	0.003 (0.21)	0.001 (0.09)	0.003 (0.16)	0.003 (0.18)	-0.002 (-0.16)	0.001 (0.05)	0.001 (0.11)	0.0 (-0.03)	0.0 (0.0)	-0.001 (-0.04)	0.001 (0.08)
6 clusters per arm														
2	0.006 (0.56)	0.003 (0.27)	0.007 (0.54)	0.003 (0.21)	0.009 (0.58)	0.009 (0.54)	0.007 (0.37)	0.003 (0.25)	0.003 (0.21)	0.0 (0.02)	-0.001 (-0.06)	0.001 (0.06)	0.005 (0.28)	0.004 (0.21)
3	0.005 (0.49)	0.006 (0.52)	0.008 (0.63)	0.012 (0.87)	0.008 (0.56)	0.011 (0.68)	0.008 (0.44)	0.0 (-0.01)	0.0 (0.01)	0.0 (0.04)	-0.003 (-0.2)	0.003 (0.22)	0.002 (0.11)	0.002 (0.13)
4	-0.003 (-0.23)	-0.001 (-0.05)	0.002 (0.18)	0.002 (0.17)	0.0 (0.0)	0.0 (0.02)	0.002 (0.12)	0.002 (0.15)	0.001 (0.04)	-0.001 (-0.05)	0.002 (0.15)	0.002 (0.15)	-0.001 (-0.06)	0.001 (0.08)
6	0.0 (0.04)	0.004 (0.33)	0.001 (0.09)	0.001 (0.09)	0.0 (0.03)	0.004 (0.25)	0.003 (0.18)	0.002 (0.16)	-0.001 (-0.04)	0.001 (0.05)	0.002 (0.12)	0.0 (-0.02)	0.002 (0.10)	-0.001 (-0.08)
12	0.001 (0.11)	0.001 (0.06)	0.0 (0.0)	-0.001 (-0.07)	0.001 (0.07)	0.001 (0.0)	0.001 (0.05)	0.0 (-0.02)	0.001 (0.06)	0.0 (0.03)	0.0 (0.02)	0.0 (-0.01)	0.0 (-0.01)	0.001 (0.05)
9 clusters per arm														
2	0.010 (0.9)	0.009 (0.75)	0.007 (0.54)	0.008 (0.57)	0.009 (0.58)	0.013 (0.82)	0.011 (0.60)	0.002 (0.14)	-0.001 (-0.05)	0.001 (0.07)	0.001 (0.06)	0.003 (0.20)	0.0 (0.02)	-0.001 (-0.06)
3	0.004 (0.35)	0.005 (0.43)	0.008 (0.63)	0.008 (0.56)	0.008 (0.56)	0.008 (0.48)	0.007 (0.39)	0.002 (0.17)	0.001 (0.1)	0.002 (0.17)	0.003 (0.23)	0.001 (0.03)	0.002 (0.14)	0.003 (0.19)
4	0.002 (0.16)	0.001 (0.1)	0.002 (0.18)	-0.001 (-0.09)	0.0 (0.0)	0.005 (0.30)	0.0 (0.03)	0.0 (-0.03)	0.001 (0.11)	0.001 (0.07)	0.001 (0.08)	0.001 (0.05)	0.001 (0.06)	-0.001 (-0.04)
6	0.001 (0.12)	0.0 (0.04)	0.001 (0.09)	0.004 (0.25)	0.0 (0.03)	0.003 (0.16)	0.006 (0.31)	0.002 (0.14)	0.002 (0.17)	-0.002 (-0.17)	0.001 (0.07)	0.0 (0.01)	0.0 (0.01)	0.0 (0.0)
12	-0.001 (-0.11)	0.0 (-0.04)	0.0 (0.0)	0.0 (0.0)	0.001 (0.07)	0.003 (0.16)	0.0 (0.02)	0.0 (0.0)	0.0 (0.03)	0.0 (0.0)	-0.001 (-0.06)	0.001 (0.04)	0.001 (0.08)	0.0 (-0.01)
12 clusters per arm*														
2	-	0.002 (0.16)	0.003 (0.26)	0.006 (0.44)	0.010 (0.65)	0.009 (0.54)	0.009 (0.48)	0.0 (0.04)	0.0 (-0.03)	0.0 (0.0)	0.001 (0.08)	0.003 (0.22)	0.002 (0.13)	0.002 (0.09)
3	-	0.0 (0.03)	0.001 (0.09)	0.002 (0.14)	0.001 (0.05)	0.003 (0.16)	0.005 (0.26)	0.0 (-0.03)	-0.001 (-0.05)	0.0 (-0.03)	0.0 (-0.03)	0.0 (-0.01)	0.001 (0.05)	0.003 (0.18)
4	-	0.001 (0.11)	0.002 (0.14)	0.001 (0.10)	0.001 (0.07)	0.001 (0.04)	0.003 (0.19)	-0.001 (-0.11)	0.001 (0.09)	0.0 (0.0)	-0.001 (-0.08)	0.0 (0.0)	0.0 (0.01)	0.0 (0.0)
6	-	0.002 (0.19)	0.004 (0.28)	0.003 (0.22)	0.004 (0.27)	0.003 (0.21)	-0.002 (-0.14)	0.0 (0.04)	0.0 (0.0)	0.0 (0.02)	0.0 (0.02)	0.001 (0.04)	0.003 (0.16)	0.002 (0.11)
12	-	0.001 (0.08)	0.0 (0.01)	0.0 (0.03)	0.0 (0.01)	0.001 (0.06)	-0.001 (-0.07)	0.0 (0.02)	0.001 (0.11)	0.001 (0.05)	0.001 (0.09)	0.0 (0.02)	0.001 (0.04)	0.001 (0.03)

ICC=0.00154

* No convergence was achieved at the effect size of 10%, hence no estimates are given

5.4.3.2 Effect of varying ICC on the accuracy of parameter estimates

To assess the effect of varying ICC values on the accuracy of parameter estimates, the average length of the 95% confidence intervals of the parameter estimates were used; these showed clear patterns when the design conditions were varied (see section 5.4.3.1). In general, the accuracy of parameter estimates decreased with increasing ICCs (Table 14). Within a given ICC, the accuracy of parameter estimates improved with increasing number of clusters, time points and effect size. This was most evident when the ICC was 0.081. The average length of the 95% confidence intervals reduced on average by 50% when the measurement time points were changed from 2 to 12 for any given number of clusters and when the ICC was 0.081. However, with ICCs of 0.321 and higher, there were minimal changes in the average length of the 95% confidence intervals even after varying the number of clusters, measurement time points and effect sizes (Table 14). These findings suggest that precise estimates of the estimated parameters can only be achieved if the ICCs are not too high and that at least 12 repeated measurement times are required to improve precision by 50% (Table 14).

Table 14: The effect of the ICC, number of groups, effect size and repeated measurements on the accuracy of the estimated effect size based the mean length of 95% confidence interval, based on 1000 simulated data sets																	
		Low incidence outcomes- Mean outcome measure in control clusters of 15 cases per month								High incidence outcomes- Mean outcome measure in control clusters of 70 cases per month							
		ICC=0.081				ICC = 0.321				ICC = 0.081				ICC = 0.321			
		Effect size (incidence rate ratio)				Effect size (incidence rate ratio)				Effect size (incidence rate ratio)				Effect size (incidence rate ratio)			
	Effect	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8
3 clusters	Time																
	2	0.985	1.033	1.142	1.268	1.375	1.706	1.896	2.109	0.544	0.641	0.722	0.821	1.244	1.477	1.728	1.977
	3	0.679	0.751	0.881	0.923	1.362	1.547	1.829	2.035	0.489	0.559	0.625	0.714	1.256	1.469	1.663	1.894
	4	0.566	0.642	0.727	0.788	1.328	1.560	1.739	1.987	0.451	0.536	0.616	0.676	1.294	1.478	1.713	1.869
	6	0.501	0.572	0.645	0.724	1.315	1.532	1.720	1.868	0.439	0.516	0.591	0.666	1.258	1.487	1.726	1.850
	12	0.453	0.515	0.590	0.661	1.303	1.476	1.693	1.858	0.434	0.501	0.583	0.657	1.256	1.433	1.633	1.869
4 clusters																	
	2	0.816	0.869	1.010	1.132	1.317	1.518	1.742	1.939	0.495	0.590	0.647	0.726	1.176	1.359	1.573	1.739
	3	0.596	0.671	0.755	0.377	1.282	1.414	1.639	1.845	0.442	0.514	0.580	0.645	1.165	1.344	1.564	1.733
	4	0.517	0.581	0.671	0.274	1.213	1.431	1.611	1.796	0.425	0.490	0.569	0.631	1.163	1.348	1.529	1.782
	6	0.461	0.517	0.603	0.220	1.192	1.392	1.534	1.800	0.415	0.479	0.554	0.621	1.150	1.339	1.559	1.757
	12	0.414	0.498	0.556	0.203	1.153	1.347	1.520	1.727	0.405	0.473	0.536	0.607	1.155	1.344	1.507	1.722
5 Clusters																	
	2	0.762	0.881	0.905	1.071	1.187	1.427	1.554	1.760	0.475	0.546	0.613	0.672	1.088	1.274	1.439	1.608
	3	0.538	0.609	0.706	0.753	1.123	1.293	1.515	1.687	0.407	0.474	0.552	0.609	1.097	1.252	1.401	1.595
	4	0.462	0.545	0.602	0.671	1.113	1.299	1.502	1.656	0.394	0.453	0.525	0.591	1.061	1.223	1.406	1.594
	6	0.431	0.495	0.558	0.623	1.095	1.281	1.412	1.626	0.384	0.448	0.509	0.572	1.058	1.238	1.400	1.588
	12	0.391	0.452	0.520	0.584	1.071	1.265	1.414	1.572	0.381	0.438	0.502	0.571	1.059	1.240	1.420	1.609
6 Clusters																	
	2	0.645	0.706	0.828	0.938	1.100	1.279	1.447	1.619	0.430	0.494	0.553	0.625	1.022	1.165	1.325	1.500
	3	0.485	0.562	0.637	0.718	1.087	1.227	1.370	1.566	0.386	0.447	0.506	0.568	1.005	1.179	1.367	1.480
	4	0.438	0.498	0.560	0.633	1.020	1.191	13.84	1.527	0.375	0.429	0.493	0.553	0.990	1.159	1.336	1.460
	6	0.393	0.452	0.518	0.579	1.004	1.183	1.358	1.529	0.361	0.421	0.475	0.533	0.965	1.486	1.316	1.469
	12	0.363	0.422	0.485	0.542	1.005	1.176	1.315	1.508	0.356	0.415	0.468	0.524	0.974	1.151	1.295	1.451

		Low incidence outcomes- Mean outcome measure in control clusters of 15 cases per month								High incidence outcomes- Mean outcome measure in control clusters of 70 cases per month							
		ICC=0.081				ICC = 0.321				ICC = 0.081				ICC = 0.321			
		Effect size (incidence rate ratio)				Effect size (incidence rate ratio)				Effect size (incidence rate ratio)				Effect size (incidence rate ratio)			
	Effect	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8	1.2	1.4	1.6	1.8
9 Clusters																	
	2	0.524	0.603	0.792	0.770	0.915	1.077	1.198	1.344	0.366	0.414	0.476	0.532	0.849	0.973	1.127	1.277
	3	0.412	0.465	0.526	0.587	0.886	1.020	1.165	1.315	0.319	0.373	0.429	0.48	0.830	0.963	1.102	1.231
	4	0.358	0.418	0.475	0.542	0.866	1.004	1.151	1.288	0.314	0.365	0.412	0.461	0.820	0.977	1.101	1.232
	6	0.336	0.388	0.443	0.492	0.844	0.982	1.116	1.262	0.308	0.354	0.407	0.456	0.829	0.959	1.099	1.225
	12	0.310	0.362	0.414	0.460	0.819	0.988	1.114	1.236	0.299	0.350	0.399	0.449	0.810	0.967	1.073	1.213
12 Clusters																	
	2	0.439	0.479	0.572	0.636	0.796	0.927	1.060	1.173	0.316	0.357	0.405	0.461	0.736	0.858	0.974	1.090
	3	0.354	0.416	0.467	0.518	0.771	0.892	1.014	1.153	0.286	0.330	0.375	0.427	0.729	0.835	0.961	1.093
	4	0.325	0.371	0.429	0.475	0.740	0.877	0.995	1.125	0.277	0.322	0.369	0.415	0.723	0.849	0.964	1.069
	6	0.293	0.342	0.391	0.435	0.739	0.869	0.978	1.111	0.269	0.313	0.359	0.404	0.725	0.842	0.953	1.091
	12	0.276	0.321	0.367	0.411	0.733	0.841	0.963	1.081	0.266	0.307	0.353	0.397	0.715	0.828	0.957	1.077

5.4.3.3 Coverage rates of Standard errors when ICC was 0.00154

The accuracy of the simulated data parameter estimates were evaluated using coverage rates; 95% confidence intervals were computed for the parameter estimates from each data set obtained from the different scenarios studied. The coverage rate was defined as the proportion of times the true parameter was within the specified (95%) confidence intervals (Burton et el, 2006). 'Adequate coverage' was thus defined as a coverage rate equal to or greater than the specified 95% (see Collins, Schafer and Kam, 2001). Using this coverage rate, the failure rate of the true specified parameter (non-coverage) was defined as 5%.

Table 15 presents the non-coverage rates of the asymptotic 95% confidence intervals of the parameter estimates of interest based on 1000 simulated data sets for each of the studied scenarios. Overall, the effects of varying the number of clusters, measurement times, effect size and incidence of the outcomes were small, with non-coverage rates ranging from 2.8% to 6%. With 3 clusters per arm, non-coverage rates were approximately close to the 5% nominal value across most of the different effect sizes when the number of measurement times was 4 and the outcome measure incidence was low (range 4.2% to 5.0%) and also when the number of measurement times was 3 and the outcome measure incidence was high (range 3.5% to 5.1%).

In general, the close clustering of the non-coverage rates around the 5% nominal value clearly demonstrates that the standard errors were correctly estimated for each of the different scenarios investigated.

Table 15: The effect of number of clusters, time points and the effect size on the non-coverage of the 95% confidence intervals of parameter estimates (regression coefficients) based on 1000 simulated data sets.

	% non-coverage of standard errors based on low incidence conditions of 15 cases per month							% non-coverage of standard errors based on high incidence conditions of 70 cases per month						
	Effect size (incidence rate ratio)							Effect size (incidence rate ratio)						
	1.1	1.2	1.3	1.4	1.5	1.6	1.8	1.1	1.2	1.3	1.4	1.5	1.6	1.8
3 clusters per arm														
2	4.8	4.0	4.4	4.1	3.9	4.3	4.2	4.4	4.6	4.7	4.1	4.1	4.4	4.1
3	4.1	4.0	4.1	4.2	4.3	3.3	3.8	4.6	3.9	3.7	4.9	5.1	4.8	3.5
4	4.5	4.2	4.8	4.8	4.2	4.3	5.0	3.4	4.2	4.2	3.8	2.8	4.4	4.0
6	3.8	4.0	5.1	4.5	4.1	4.4	4.4	4.1	4.7	5.2	4.6	3.9	4.4	4.1
12	4.7	3.7	4.1	3.4	3.0	4.1	3.6	4.0	3.5	3.6	4.4	4.6	5.8	3.9
4 clusters per arm														
2	3.8	4.8	4.6	4.9	5.0	4.3	4.6	3.6	4.7	4.7	4.6	4.3	4.0	4.0
3	3.9	4.8	5.9	5.7	5.0	5.0	5.0	4.4	4.1	5.2	3.9	3.9	3.9	4.1
4	4.2	3.5	4.3	4.6	2.8	4.1	4.2	5.0	4.7	3.3	3.9	3.1	5.9	4.1
6	4.0	4.1	5.0	4.2	4.3	4.2	3.5	4.1	3.4	5.1	3.4	3.6	3.2	3.8
12	4.3	3.9	4.6	3.8	3.9	3.3	4.1	3.6	4.1	4.2	5.1	4.1	3.7	5.4
5 clusters per arm														
2	4.2	4.1	3.8	4.6	4.5	4.3	4.0	3.8	4.8	3.7	5.4	5.8	4.5	3.9
3	3.9	4.1	3.7	4.2	4.1	4.2	3.6	3.2	3.8	4.8	5.0	5.1	4.1	4.6
4	4.4	5.2	4.6	4.6	4.4	4.4	4.4	4.8	4.7	4.1	3.8	3.9	3.0	4.5
6	4.3	4.1	4.3	5.7	4.4	4.2	3.7	4.4	4.0	3.4	5.2	4.6	3.4	4.8
12	3.3	3.5	4.3	3.0	3.1	3.3	3.9	4.4	4.0	3.7	3.1	4.0	4.3	3.2
6clusters per arm														
2	4.4	4.3	4.7	5.7	5.3	4.5	4.8	3.6	4.3	4.3	3.7	5.4	4.8	5.7
3	5.0	4.3	4.4	4.2	4.9	4.2	5.6	4.5	5.4	5.5	3.5	3.9	4.0	4.5
4	4.8	4.9	6.0	4.7	4.0	4.4	3.9	4.2	3.3	3.7	3.5	4.3	4.1	3.5
6	5.1	4.1	4.6	4.5	5.1	4.0	4.6	4.5	3.9	5.4	4.8	3.9	4.6	3.9
12	3.8	3.7	4.4	3.5	4.9	4.6	4.0	3.4	3.5	3.8	3.7	4.1	4.1	4.1
9 clusters per arm														
2	4.7	4.1	4.1	4.0	4.2	3.8	4.5	4.2	4.4	5.3	3.3	5.6	5.2	4.3
3	4.7	4.9	4.0	4.0	4.3	4.8	4.7	4.4	4.8	5.2	4.3	5.0	4.4	3.1
4	4.6	3.8	4.5	4.7	3.8	4.0	4.9	5.1	5.3	4.5	3.9	4.4	5.3	5.4
6	3.2	3.2	3.0	3.8	3.1	4.4	3.1	4.0	4.1	4.5	4.1	4.3	3.7	3.7
12	4.5	4.0	3.6	4.8	4.6	4.2	3.4	4.6	4.5	4.3	5.5	4.9	4.4	4.4
*12 clusters per arm**														
2	-	3.8	4.5	4.1	4.1	3.7	4.5	6.0	5.8	4.6	4.8	4.2	4.9	4.1
3	-	3.9	2.9	5.0	4.5	3.8	3.7	4.3	4.2	3.8	5.6	5.2	4.7	5.3
4	-	3.9	3.5	4.1	3.3	3.2	4.1	3.8	4.6	5.0	4.0	4.5	4.7	4.9
6	-	4.1	3.6	3.8	4.1	4.1	3.9	3.9	4.7	3.8	4.3	3.9	3.5	4.2
12	-	5.0	4.4	3.7	4.2	4.0	4.5	4.3	3.0	3.3	5.0	3.8	4.2	4.4

ICC=0.00154

* No convergence was achieved at the effect size of 10%, hence no estimates are given

5.5 Conclusions

This Chapter has investigated the conceptual and practical issues relating to the attainment of adequate statistical power and efficiency for a range of cluster randomised study designs. The simulation studies presented provide a wealth of findings that will help to inform the choice of design conditions for a cluster randomised study in circumstances where the number of clusters is both limited and small. By using simulation approaches, different design conditions could be investigated in a way that would not normally be possible using real field trials.

It was apparent that there was no gain in evaluating the different design conditions of interest within the context of the Triage Plus study using different statistical approaches (e.g. using negative binomial models for count data), so this option was not pursued. In particular, random effects models analysed with Poisson regression models and robust standard error would provide similar results to using negative binomial models. Use of robust standard errors reduces the over-dispersion usually encountered in Poisson regression models (Rabe-Hesketh & Skrondal, 2012, page 712). As highlighted in section 4.2.3.2, the estimates of the regression coefficients in Poisson and negative binomial models are similar but differ in variance estimates, and therefore use of the robust standard errors in marginal effects models corrects the difference (Rabe-Hesketh & Skrondal, 2008). The only drawback of using the Poisson models is that they may give rise to predicted probabilities that are greater than 1 when the estimates are on or are very close to the boundaries of the parameter space (Deddens and Petersen, 2004); this problem was resolved in this study by simply increasing the number of iterations used to reduce high rates of non-convergence rates (Table 15).

Furthermore, because the primary outcome measure collected in the Triage Plus study was a count, the use of log-binomial regression models, presented under the Future Work section 7.8.2 in Chapter 7,would not have been ideal in the current study despite producing efficient estimates and accurate measures of statistical uncertainty in estimating prevalence or risk ratios (Wacholder, 1986; Zocchetti, 1995; Skov et al., 1998; Savu, Liu & Yasu, 2010).

The simulation study findings reported here show that the gain in statistical efficiency and power varied considerably under the different design conditions investigated. As expected, nominal non-coverage rates as well as negligible biases in the estimation of regression coefficients of fixed effect parameters were observed in all scenarios investigated, including the 3 cluster per arm design. However, in order to achieve precise estimates that would in turn provide some gains in statistical efficiency and hence power, if at least 6 repeated measurement times are used with an ICC of 0.00154, precision improves by as much as 50%

compared to using just two time points, and this gain in precision is greatest for high incidence outcomes. In high incidence disease conditions, with effect sizes of as low as 10%, adequate power is still achievable, particularly if at least 6 repeated measurements are taken over time.

By contrast, the simulations also clearly show that more measurement times are required to achieve adequate power (≥80%) with effect sizes of 20% or lower in low incidence situations. With effect sizes of 10% or lower in low incidence disease conditions, no adequate power levels or efficient parameter estimates were achieved in any of the design conditions investigated, implying a substantial (and probably totally unacceptable) risk of a statistical type II error (i.e. of drawing a wrong negative conclusion) when analysing the effectiveness of an intervention using a small number of clusters when both the population effect size and disease incidence are low. Possibly the most surprising finding was that, in low incidence disease conditions and with an effect size of 10%, the simulations encountered convergence problems even when the number of measurement times was increased to 12; these challenges are likely to be even more severe with very low effect sizes (<10%). Moineddin, Matheson and Glazier (2007) also encountered lowest convergence rates at disease prevalence rates of 10%. Further research will be needed in future to investigate this finding. However, for ICC values ≥0.081, such convergence challenges were not observed potentially due to the fact that larger values of the between cluster variance increase the background incidence rate leading to improved convergence rates (Amatya, Bhaumik and Gibbons, 2013).

Simulations carried out to investigate the effect of varying the intracluster correlations on statistical power estimation indicated that statistical power decreased with increasing ICC. With an ICC of 0.699, the simulations indicated that no adequate statistical power could be achieved when the number of events per month was ≤70. However, with ICCs of ≤0.081, adequate statistical power could be achieved with 4 repeated measurement time points and an effect size of 40% in a 3-cluster per arm design when the disease incidence rate was low, and with only 3 repeated time points when the mean number of events per month was 70. At effect sizes of ≤20%, adequate statistical power was achieved when the number of clusters was 9 with 12 measurement time points for low incidence disease rates and 6 time points for high incidence disease rates.

Although the simulations reported in this dissertation did not assess the effect of varying group level explanatory variables on power and parameter estimation accuracy, the results

obtained are comparable to previous studies (Bennett et al., 2002; Moineddin, Matheson and Glazier, 2007). However, unlike previous studies, our study has further demonstrated that by employing a repeated measurement approach in the design of the cluster randomised trials, adequate statistical properties (e.g. statistical power) can be achieved in situations where only a limited number of clusters is available for randomisation provided the incidence of the outcome of interest is not too small and the ICC is not too high. In the simulations, estimates of variance components and model intercepts from baseline samples from the actual Triage Plus study were used. Assessment of various explanatory variables would have been essential in assessing statistical efficiency, but these analyses were not included. Including them in the simulations would have meant adding more computer processing time considering the varied conditions considered in the simulations.

The findings presented in this Chapter have a number of implications worth noting:

1. In situations where the number of available clusters is both small and limited, study designs that do not involve a longitudinal data structure (i.e. do not involve several post-intervention assessment times) risk being considerably underpowered and are likely to provide poor estimates of the effectiveness of the intervention (effect size). Some researchers avert the challenge of a limited number of clusters by analysing the data as if it were not derived from a cluster randomised design (i.e. derived from a conventional individual-participant randomised design), a practice that risks overstating the significance of the interventions effects (Murray, 1998a), resulting in wrong decisions and poorly informed policy directions. It is worth noting that it is only necessary to increase the number of clusters to about 6 in each arm of the study and to have more than one post-intervention assessment time to achieve optimal statistical properties. However, although the use of random effects modelling with variable adjustments accounted for the heterogeneity and unknown confounding effects, the use of more clusters is recommended to ensure an even spread of covariates between the treatment arms (and hence to ensure internal validity).

2. Power estimation in conjunction with the use of the average length of the 95% confidence intervals of parameter estimates (calculated as an absolute difference between the limits around the estimate) was found to be critical in assessing the varied design conditions. Using power estimation alone did not provide clear patterns in determining optimal design conditions especially when the ICCs were moderate. The simulations showed that precise estimates were achieved with increasing the number of clusters and measurement times for

each of the effect sizes and disease incidence and decreased with increasing intracluster correlations.

3. Given that statistical efficiency is most affected by the incidence of the disease condition and the intervention effect size, especially in rare disease conditions such as TB treatment initiations, when the number of clusters is limited the use of more measurement times can achieve optimal statistical efficiency and power in the context of large cluster sizes as pursued here (the Triage Plus study, from which the baseline data sets were derived, used large cluster sizes).

4. However, because there is often (financial and other) constraints that will limit the actual number of post-intervention assessment times available for random effects modelling (Best, Mason and Li, 2011), in order to ensure internal validity of the trial results, the findings of this simulation study suggest that a minimum of 9 clusters per arm and a minimum of 4 measurement times in low incidence diseases and low ICC (in our case an ICC of 0.00154) are needed with an effect size of ≤20% to obtain valid estimates using mixed effects Poisson models. However, for high incidence outcomes, as low as 3 clusters per arm and at least 3 measurement times may be adequate to achieve statistical power of at least 80%. With an ICC of 0.081, more clusters and measurement times were needed to achieve adequate statistical power in both high and low incidence disease conditions.

5. In complex study designs requiring a number of factors to be considered for sample size determination, sample size calculations cannot be conducted using the existing formulae and extended formulae may prove to be too complex to be of practical value. Simulation methods, such as those employed in this study, may need to be used to overcome these issues where relevant data are available to do so.

CHAPTER 6

STATISTICAL ANALYSIS OF THE TRIAGE PLUS STUDY

6.1 Introduction

Different statistical methods were presented in Chapters 3 and 4. Some of these methods were used in the simulation studies presented in Chapter 5. In this Chapter, the analytical framework is applied, which the simulation methods showed to be optimal, on the Triage Plus integrated TB-HIV community cluster-randomised intervention study. In particular, the robustness of the statistical methods in assessing the effectiveness of the Triage Plus intervention were assessed by using marginal and random effects models. The circumstances under which the statistical methods would be more robust in detecting significant intervention effects when applied to real situations are assessed.

Given that the second part of the thesis focuses on assessing the effectiveness of engaging informal healthcare providers in integrated TB and HIV interventions at community level, the chapter further presents results of the Triage Plus study as well as its implications for TB and HIV control. As the phased intervention approach was used in implementing the Triage Plus intervention (see section 2.3.3.5 in Chapter 2), the results presented in this Chapter are presented according to the phases. This Chapter further presents a summary of the findings concerning the involvement of informal health care providers in TB and HIV community interventions. This Chapter concludes with the statistical implications of the Triage Plus study findings and gives cautious interpretations of the non-significant findings.

6.2 Analytical approach

6.2.1 Unit of analysis and the need for repeated measurements

In cluster randomised interventions, analysis is usually done at the cluster level. Since the number of events (e.g. TB treatment initiations) at cluster level on each of the different measurement occasions were recorded, our unit of analysis is the measurement occasion within each cluster (i.e. cluster x measurement occasions units). Cluster level summaries are usually robust and provide valid results, unlike individual level analysis (Hayes and Moulton, 2009). Using cluster- measurement occasion summaries improves statistical power, especially

in studies with a limited number of clusters as observed in the simulation studies presented in Chapter 5.

When there is only a moderate number of clusters, Thompson, Warner and Turner (2004) noted that inferences made from asymptotic approximations are less robust. Considering that multilevel modelling is most attractive with more clusters (Feldman, McKinlay, and Niknian, 1996; Maas and Hox, 2005; Moineddin, Matheson and Glazier, 2007; Snijders and Bosker, 1999) and that the Triage Plus study has only limited number of clusters, the repeated-measurement nature of our data improves the number of degrees of freedom and, therefore, the statistical power, especially when marginal or random intercepts models are used (Murray et al., 1998b). As pointed out by Murray et al (1998b), power still improves in random coefficients models, even though the degrees of freedom remain constant with the increasing number of repeated measurements from the same clusters. By using the longitudinal structure of data collection as described by Feldman, McKinlay, and Niknian (1996), as shown in the simulation studies presented in Chapter 5 that statistical power and efficiency in parameter estimations can be increased without necessarily increasing the number of clusters, especially in high incidence outcomes such as ART initiation rates.

The use of repeated measurements, sometimes called time-series cross-sectional data, in the same clusters in the context of a mixed effects framework has the advantage of allowing the assessment of whether or not the effect of the intervention increases over time as the intervention becomes established in the community (Gelman and Hill, 2007).

In such modelling approaches, the incorporation of separate, random intercept and slope terms for each time point may be necessary to distinguish intervention effects from secular trends (Hayes and Moulton, 2009). The approach allows for greater flexibility in modelling (studying) cluster-level and time effects when accounting for correlations resulting from the clustering and repeated measurements (Gelman and Hill, 2007). As pointed out by Gueorguieva and Krystal (2004), the flexibility in modelling correlation and variance patterns is necessary since variations in response outcomes between clusters may occur over time and during the implementation of the intervention: variations may be greater immediately after the initiation of the intervention or later in the study when the intervention has had an opportunity to 'bed-in' (i.e. if there is a lag between intervention implementation and first signs of effect).

6.2.2 The matched-pair design analysis

The matched-pair design implemented in the Triage Plus study was meant to minimise differences in baseline characteristics between intervention arms which improves precision and statistical power as well as trial results credibility. However, the matching on surrogate variables that were expected to be correlated with the outcome measures was not effective in minimising between-group variation in baseline outcome measures such as TB and ART treatment initiation rates (see Table 18). The pair-matching was less robust due to there being only a limited number of clusters (Diehr et al., 1995; Hayes and Moulton, 2009).

The matched-pair design analysis usually decreases statistical power, especially when the number of pairs is less than 10 (Diehr et al., 1995, Klar and Donner, 1997; Feng et al., 2001). Klar and Donner (1997) noted that with a limited number of matched-pairs (as in the Triage Plus study) obtaining balanced matches on all potential factors may be difficult. Therefore, small differences, though important in the context of TB and HIV community interventions, may be missed if a matched (conditional) analysis is conducted. In simulation studies, which sought to compare the performance of unmatched (unconditional) and matched (conditional) analysis methods on data from a matched study design, Diehr et al (1995) concluded that the use of unmatched methods was more efficient in the presence of fewer than 10 cluster-pairs.

In contrast, Donner, Taljaard and Klar (2007), in their investigation of the effect of individual-level risk factors on outcomes of interest in data arising from a matched-pair design, showed that unmatched analyses may result in biased estimates of the regression coefficients. They argued, however, that in community intervention trials with a small number of large clusters, using unmatched analysis would generally be valid and efficient. Therefore, since Triage Plus had only 3 matched-pairs (giving it only 2 degrees of freedom for testing the effect of the intervention), the loss of degrees of freedom associated with matched analysis outweighs the reduction in variance thereby resulting in reduced power compared with an unmatched design. The matched design would only provide increased power if the matching was highly effective due to the high correlation of outcome measures within matched pairs (Hayes and Moulton, 2009). Assessing the variation of the effect of the intervention in matched pairs is not possible as it depends on effect size and the intrinsic between-cluster variability which cannot be separated without additional assumptions (Hayes and Moulton, 2009). Given that the matching done in the Triage Plus study did not match on the baseline outcome measures (i.e. TB and HIV treatment initiations and testing uptake rates) and that the matching carried out on the surrogate variables was less effective, the use of unmatched analysis methods was

clearly optimal for the Triage Plus study data analysis; the matching done only increased the face validity of the study (Klar and Donner, 1997; Feng et al., 2001). Bennett et al (2002) also noted that point estimates of the intervention effect size derived from paired and unpaired analyses were similar. Thus, unmatched analyses were used to improve the statistical power of the Triage Plus study data analysis, as this study was effectively a large scale community intervention trial with a small number of clusters (Donner, Taljaard and Klar, 2007; Feng et al., 2001; Diehr et al., 1995). The matching, conducted at the randomisation stage of the study, did help to balance the two arms of the study with respect to population size but not necessarily across the baseline outcome measures, as evidenced by the lack of homogeneity between the matched pairs on these measures (Table 18).

6.3 Statistical methods

6.3.1 Analysis plan

As the intervention clusters only received the intervention package after baseline data was collected from each cluster-pair for at least 12 months prior to the intervention, the beginning of the intervention period was defined as the month when one of the clusters in a cluster-pair initiated intervention activities. The intervention was implemented in the cluster in each matched pair assigned to the Early intervention arm during the first 12 months of the study period, while the other cluster in each matched pair was by definition assigned to the Delayed intervention arm; clusters in the Delayed intervention arm received the intervention in the final 11 months of the study period. Due to time constraints, the Delayed clusters could not receive the intervention for the 12 months period. Monthly data on number of TB and ART patients starting treatment (primary outcomes)was repeatedly measured in all 6 clusters: overall, each cluster had 23 data measurements which corresponded to the monthly data collection (i.e. 12 months in Phase I and 11 months in Phase II). The repeated measurements in each cluster allowed for the implementation of multilevel modelling with repeated measurements nested within clusters.

Because the Early intervention arm received the intervention package in the first 12 months while the Delayed intervention arm acted as a control arm, our analysis is conducted in two phases. First, the first 12 months' data are used to estimate the effect of the intervention by comparing treatment initiation rates and testing uptake rates between the two arms adjusted for the baseline data and cluster-level characteristics. Second, the next 11 months' data are

used to assess if the estimated intervention effect changed with the introduction of the intervention to the Delayed intervention arm. Thus, two scenarios are possible:

1. If the effect of the intervention peaks early within the first few months, then there would be a reduced effect size between the first 12 months and the next 11 months because the introduction of the intervention to the Delayed intervention arm narrows the gap in the effect size or remains constant.
2. If the effectiveness of the intervention is established with time, then the narrowing effect in effect size reduction would be less pronounced between the two study arms.

Guided by the primary and secondary outcomes of the Triage Plus study, as defined at the study design stage (section 2.5), and by the need for clarity, the results for the primary outcome measures are presented first followed by those for the secondary outcome measures. Our intervention was expected to promote uptake rates for TB and HIV testing through TB and HIV community sensitisations and referrals made at the community level by the informal health care providers. These increased testing uptake rates were then expected to eventually lead to improved TB and HIV treatment initiation rates. The primary outcome measures considered in the first phase are therefore TB and ART treatment initiation rates. Uptake rates for TB and HIV testing are the secondary outcome measures in the analysis.

Marginal effects Poisson regression models as well as random intercepts and random coefficients generalised linear mixed effects Poisson models were used in order to identify the most appropriate model for assessing the impact of the intervention in regards to improving TB and ART treatment initiation rates. The relative effect of each covariate on the outcome of interest was assessed while adjusting for all the covariates in the model. Random effects (generalised linear mixed modelling) methods are adopted because of their suitability in the presence of repeated measurements and clustering effects in cluster randomised designs (Murray et al., 1998b; Best, Mason and Li, 2011). Such models are usually used for large numbers of clusters with a small number of repeated measurement times, but are used here to increase the degrees of freedom and to improve statistical power as demonstrated in Chapter 5.

6.3.2 Adjustment for baseline outcome data and cluster-level covariates

Because only a limited number of clusters were used in the Triage Plus study, the cluster randomisation process did not adequately achieve an even distribution of either the baseline outcome measures or cluster level covariates between the two study arms (Table 17).Adjustment for the estimated intervention effects for baseline data collected over 12 months prior to the implementation of the intervention was therefore made to control for any baseline imbalances in outcome measures between the intervention arms. This adjustment reduces between cluster variation in the study outcome measure thereby increasing study power and precision in the estimated effect sizes (Hayes and Moulton, 2009). The most robust method for adjusting for baseline differences in outcome measures is to include the baseline variable in the regression model as an independent covariate (Hayes and Moulton, 2009). However, for the cluster level covariates, baseline adjustment only included covariates in the regression model that improved log likelihood estimation (see section 4.4.1 in Chapter 4).

6.3.3 Statistical modelling

To estimate the effectiveness of community engagement in improving TB and HIV services access, concurrent comparisons of the outcome measures between the intervention arms using likelihood based approaches were performed. Estimation of the effectiveness of the intervention and investigation of heterogeneity in cluster level TB and ART treatment initiations rates as well as TB and HIV testing uptake rates were investigated using generalised linear mixed effects Poisson models with some adjustment for cluster level covariates. Because observations from the same cluster are usually correlated, cluster level random effects to account for the autocorrelation between observations were included. This approach allows for both within and between cluster variability (Breslow and Clayton, 1993).

The statistical model

If y_{ij} is an outcome response for cluster $j(j = 1, \dots, 6)$ at measurement time point i $(i = 1, \dots, 12)$, which is conditionally Poisson and distributed with mean μ_{ij}, the general statistical model used to assess effectiveness of the intervention is given by

$$y_{ij}|x_{ij}, u_j, u_{ij} \sim Poisson\left(\mu_{ij}\right)$$

$$\log\left(\mu_{ij}\right) = log(n_{ij}) + \beta_0 + \beta_1 treat_{ij} + \beta_2 Base_{ij} + \beta_3 Time_{ij} + \beta_4 treat_{ij} * Time_{ij} + \beta_5 C_{ij} + u_j + u_j treat_{ij} + u_{ij} \quad (6.1)$$

where:

μ_{ij} is the mean count at time *i* for cluster *j;*

the constant β_0 is the grand mean count when all covariates and random effects terms are zero (Omar and Thompson, 2000);

$treat_{ij}$ is a variable indicating the intervention status for the cluster and takes values 1 for an Early intervention cluster and 0 for a Delayed intervention cluster;

the fixed effects $Base_{ij}$ are the baseline ART and TB treatment initiation rates or testing uptake rates for HIV and TB at time 0 (i.e. at baseline) in cluster j are the systematic differences between intervention groups existing at baseline and persisting throughout the intervention;

$Time_{ij}$ is the repeated measurement time variable at a given time point within a cluster and therefore measures both time trends for repeated measurements and secular trends occurring in both intervention groups over the intervention periods (Feldman, McKinlay and Niknian, 1996);

the $treat_{ij} * Time_{ij}$ isan interaction term representing intervention effects that develop over time during intervention.

In this specification, if there are only two time intervals, the design reverts to the standard pre-post intervention design with pre-post measures nested within clusters.

To compare the longitudinal trends between two intervention arms in either TB and ART treatment initiation rates or TB and HIV testing uptake rates, an interaction term between intervention status and time is included in the model. Thus, β_1 and β_4 in the model are the coefficients of interest as they represent the intervention effect after adjusting for the other covariates and random effects.

In our analyses, just like other cluster level covariates, intervention status, time and interactions between time and intervention status were treated as fixed effects.

Because there were different cluster-level population sizes, $\log(n_{ij})$ is an offset (covariate with regression coefficient set to 1) used to obtain model based cluster-specific estimates taking into account the cluster-level population (Rabe-Hesketh and Skrondal, 2012, page 724). The random cluster-level intercept u_j accounts for between cluster variations resulting from the repeated measurements within each cluster and the cluster randomised design respectively (Hayes and Moulton, 2009; Hu et al., 1998). The random intercept u_{ij} models the overdispersion in the count data. All the random effects were assumed to be Normally distributed with mean zero and an unknown variance-covariance matrix. The C_{ij} represents any cluster level covariates considered in the model that are likely to affect the outcome of interest (e.g. distribution of health facilities, gender distribution). Therefore, β_5 in model (6.1) is a vector of regression coefficients for the covariates included in the C_{ij}, and fixed effects $\beta_0$to $\beta_4$are single estimates. Appendix 8.6 details the Stata analysis programme used.

Instead of using generalised estimating equations in determining marginal models, Poisson generalised linear models with robust variance estimators were used to adjust for clustering because of the robustness issues of using generalised estimating equations with a small number of clusters (Donner, Eliasziw and Klar, 1994; Moore and Tsiatis, 1991; Zou, 2004). With 12 measurement times, there are adequate degrees of freedom (about 44) to ensure asymptotic properties of the sandwich estimators to avert the misspecification of the covariance matrix (Liang and Zeger, 1986).

Although economic status, the distance a person has to travel to health facilities and gender all affect access rates to services (in particular TB initiation rates) (Storla et al., 2008, Kemp et al., 2007; Needham et al., 2001), the C_{ij} only included gender distribution and the distribution of health facilities (i.e. number of health facilities offering TB and HIV services in each of the 6 clusters were documented and this actedas a proxy for service proximity) as covariates in the model. In addition, factors such as poor living and working conditions (associated with a high risk of TB transmission), prevalence of HIV (HIV impairs the host's defences against TB infection and disease), prevalence of malnutrition, smoking, diabetes, alcohol abuse and indoor air pollution (Lönnroth et al., 2009) were also not considered. This is because the primary outcome data was collected from health registers that do not capture such risk factors as economic status and actual distance travelled by all patients.

To improve convergence and accuracy in the estimation of parameters, covariates are standardised to mean zero and standard deviation 1 by centering each covariate about its mean (Gelman and Hill, 2007). This standardisation was unfeasible for the likelihood based approaches in the Stata software, as the centering of covariates around mean zero was found to be impossible because the covariates were averaged and fixed across time points, and therefore centering of such covariates within clusters would result in having zeros in each time point. Likewise, centering the embedded time variable that indexes repeated measurement occasions in Stata did not offer any improvement in any of the parameter estimates (see Appendix 8.4 for the data layout used for final analysis).

Monthly and cumulative treatment initiations and testing **uptake rates**

In addition to the statistical modelling used in assessing the effectiveness of community engagement in increasing access to TB and HIV treatment initiations and testing uptake, monthly and cumulative access rates for TB and HIV services between the two arms were also calculated and plotted using cluster level populations as denominators. The graphical presentation of the monthly and cumulative treatment and testing uptake rates over time helped to show the differences in TB and HIV treatment initiations and testing uptake rates between the study arms. These were then confirmed by fitting the marginal and random effects statistical models with covariate adjustments.

6.4 Summary of characteristics of clusters

The Consort diagram shown in Figure 8 summarises the overall cluster characteristics and the outcome measures (i.e. number of individuals starting TB and ART as well as those accessing testing services for TB and HIV) between the Early and Delayed arms. A detailed description of the results of the Triage Plus study follows.

6.4.1 Baseline demographic and distribution of health facilities

Table 16 summarises the baseline demographic characteristics by cluster (including the gender distribution usually associated with service access) in the two study intervention arms (clusters for the Early arm were 1, 3 and 4; clusters for Delayed arm were 2, 5 and 6).

Overall, the gender distribution at baseline for each of the outcome measures was similar across clusters. However, the proportion of female TB patients starting treatment at baseline

was high in cluster 4 compared to the other clusters. The proportion of females accessing HIV testing was higher in cluster 2 than the other clusters.

Mean cluster population was greater in the Delayed intervention areas (209,564) than the Early intervention areas (200,714). Cluster level population ranged from 166,702 to 243,826 in the Early intervention areas and from 167,074 to 232,433 in the Delayed intervention areas.

Table 16: Comparison of the baseline demographic characteristics between Early intervention and Delayed intervention arms

	Early Intervention			**Delayed Intervention**		
	Defined Clusters with the traditional authorities in each cluster					
Demographic factors	1	3	4	2	5	6
Gender distribution for total TB cases starting treatment at baseline (% females)	46.5	34.7	53.1	48.3	46.9	43.6
Gender distribution for TB testing uptake at baseline (% females)	52.0	47.0	52.0	51.0	53.0	53.0
Gender distribution for ART patients starting treatment at baseline (% females)	60.4	59.8	65.3	60.6	60.3	63.0
Gender distribution for HIV testing uptake at baseline (% females)	54.3	53.5	49.4	64.8	45.1	51.6
Mean cluster catchment population at baseline	243,826	191,615	166,702	167,074	229,184	232,433

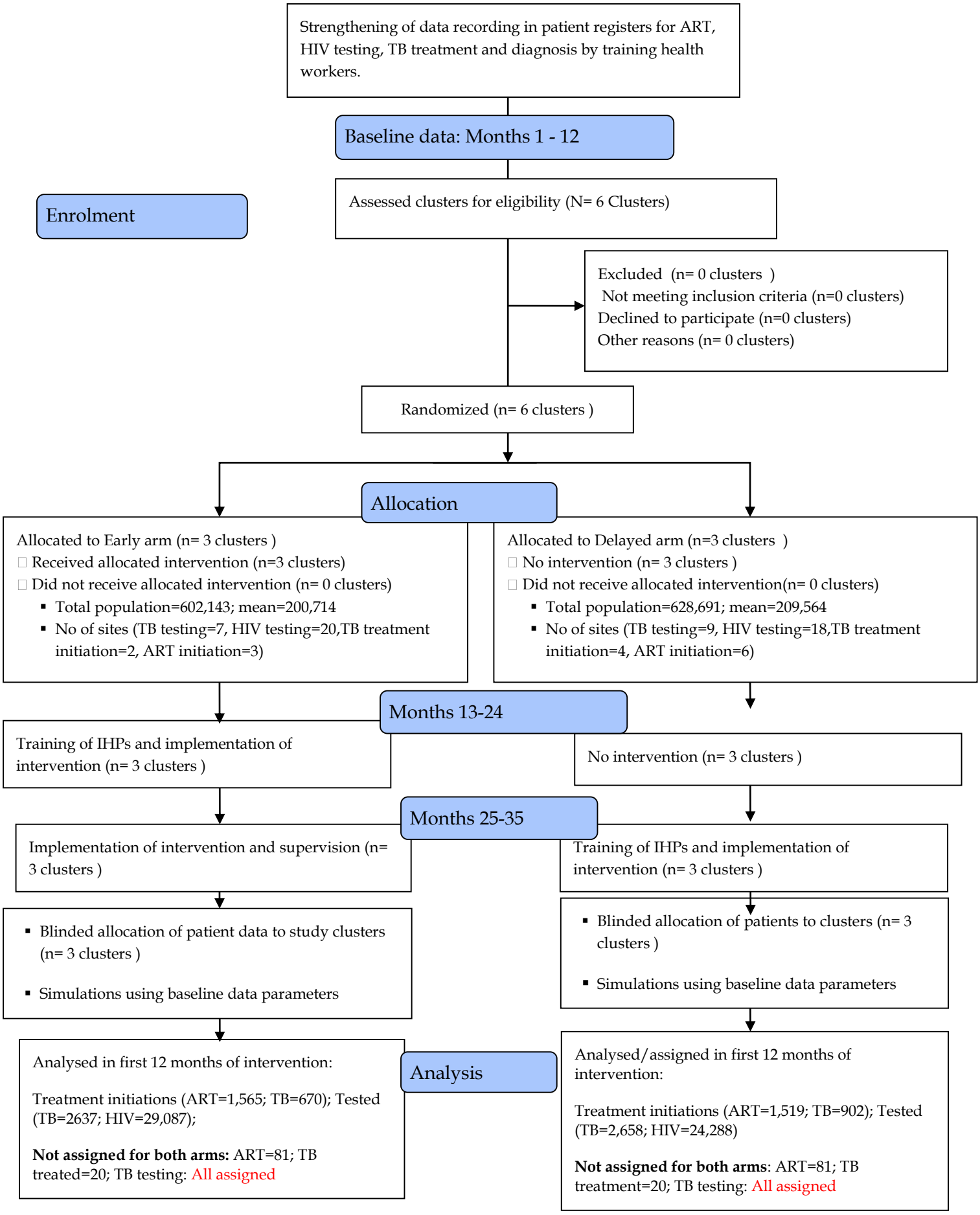


Figure 8: Flow diagram

In terms of the distribution of health facilities offering TB and HIV services, Figures 8, 9 and 10 summarise the distribution of health facilities offering TB and HIV services in the study areas. At baseline, the number of health facilities initiating treatment for TB and ART in the Early arm were two and three sites respectively (Figure 8); in the Delayed arm, four health facilities initiated TB treatment and ART was initiated in six health facilities. The number of health facilities offering HIV testing in the rural areas was twice that of TB testing (Figure 8).After the first 12 months of the study period, new health facilities, which were also evenly spread between intervention arms, were opened to offer ART treatment initiation services.

Both of the intervention arms were served by urban health facilities such as Bwaila Hospital, Light House and Martin Preuss, as well as by other private and public health facilities.

Figure 9 shows the distribution of health facilities offering TB and ART treatment initiation services during the first 12 months of the Triage Plus study, and Figure 10 shows the distribution of health facilities when new ART initiation sites were opened between the two intervention arms.

Because the disparity in the distribution of health facilities offering TB and ART treatment initiation and testing services was likely to affect service access rates, this imbalance was adjusted for by including number of sites as a covariate in all models used in subsequent analyses when assessing the effectiveness of the intervention.

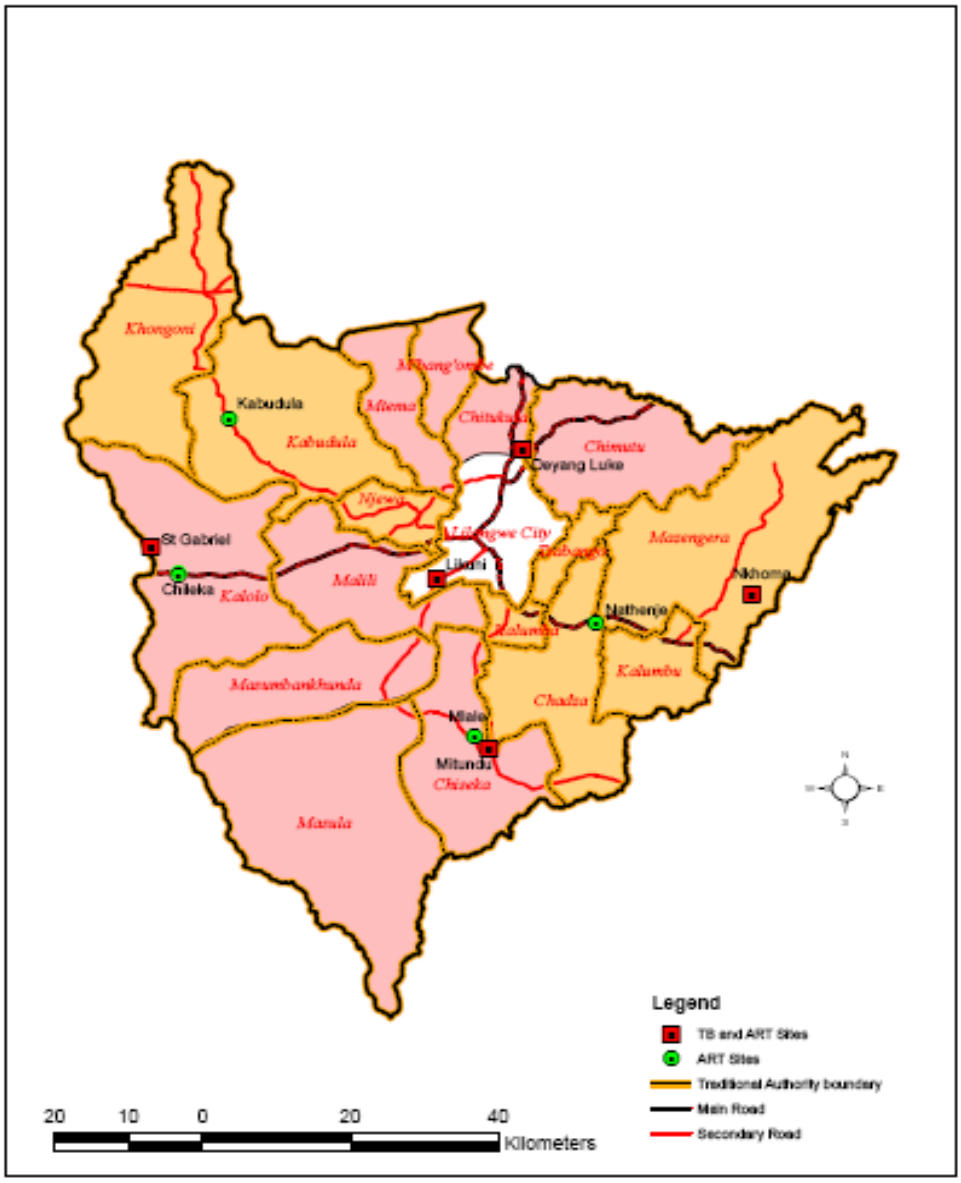


Figure 9: Map showing the distribution of TB and ART treatment initiation sites in the first 12 months.

The pink background represents clusters in the Early intervention arm and the orange background represents clusters in the Delayed intervention arm. The circles in green background represent health facilities initiating ART treatment and the squares with red background represent health facilities initiating both TB and ART treatment.

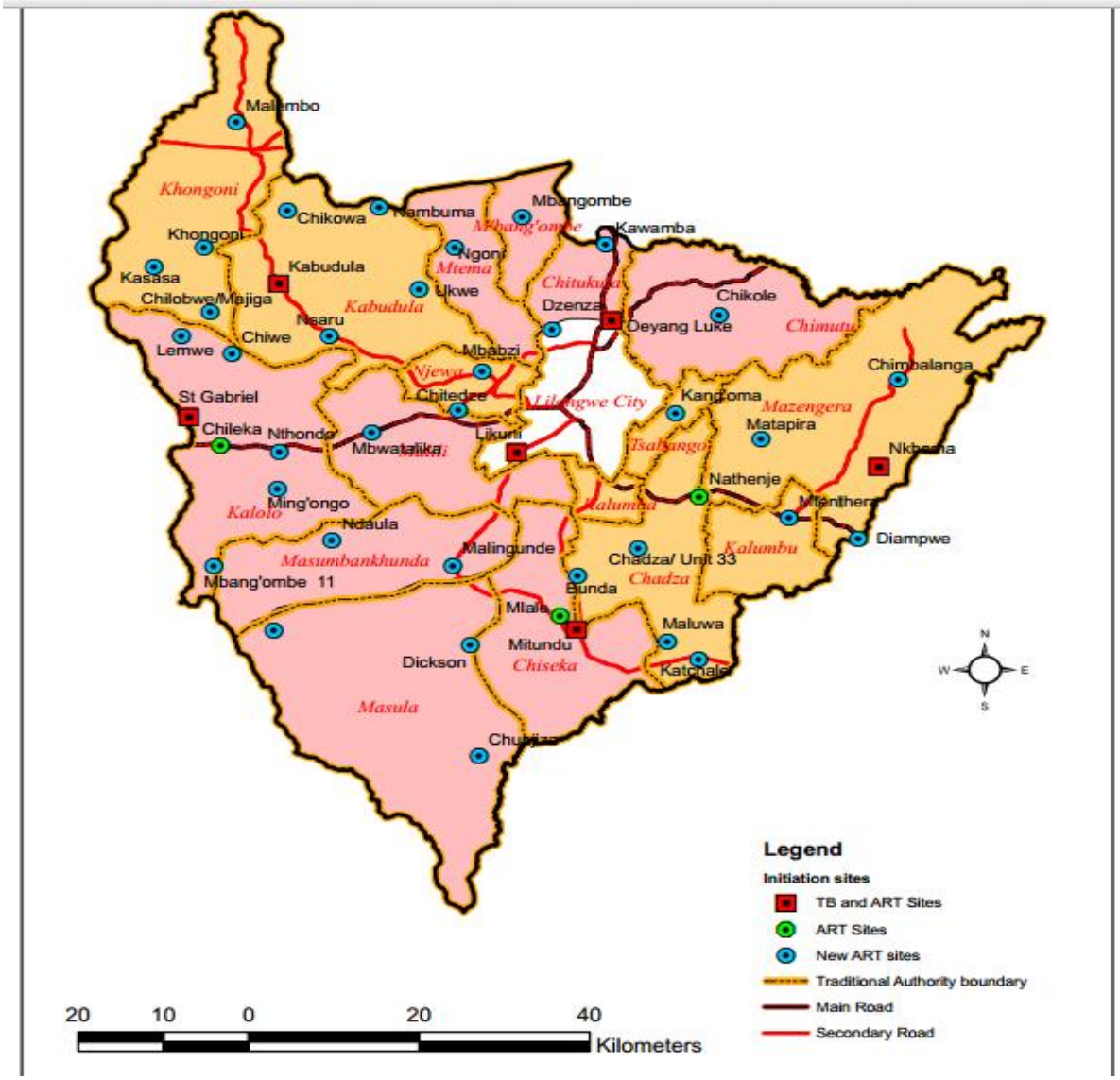


Figure 10: Map showing the distribution of new ART initiation sites.

The pink background represents clusters in the Early intervention arm and the orange background represents clusters in the Delayed intervention arm. The circles with blue background are the new health facilities initiating ART and the circles in green background are the old ART initiating sites. The sites shown in red colour initiate both TB and ART treatment.

6.4.2 Baseline measurement of outcome measures between intervention arms

Tables 17 and 18 summarise the unadjusted estimates of incidence rate ratios (IRR) and the mean counts for baseline TB and ART treatment initiation rates by intervention arm and by cluster-pair. The incidence rate ratio for smear-positive TB cases starting treatment was 0.847, suggesting some imbalance in this rate at the baseline between the two intervention arms with the Early arm having lower smear-positive TB cases starting treatment than the Delayed arm, but baseline total TB and ART treatment initiation rates were similar between the two intervention arms (IRR = 0.904 and 1.039 respectively).

Because of the observed baseline differences in treatment initiation rates between the intervention arms (especially for smear- positive TB cases), TB and ART treatment initiation rates between matched cluster-pairs were further explored. Clear variability was detected between cluster-pairs, with cluster-pair 2 having lower incidence rate ratios than both of the other cluster-pairs. TB treatment initiation rates varied from an IRR of 0.663 in pair 2 to 1.318 in pair 3. This again necessitated baseline adjustments in subsequent analyses when estimating the effect of the intervention.

Table 17: Comparison of baseline crude mean counts and incidence rate ratios for TB and ART treatment initiation rates between intervention arms

	Mean number of new cases/month (SD)		Incidence rate per 100,000 population		
	Early intervention*	Delayed intervention*	Early intervention	Delayed intervention	IRR**
Total TB cases starting treatment at baseline	15.0 (4.97)	17.3 (5.02)	111.93	123.75	0.904
Smear-positive TB cases starting treatment at baseline	4.1 (0.85)	5.2 (1.20)	31.39	37.06	0.847
ART cases starting treatment at baseline	32.97 (6.43)	33.2 (8.29)	245.95	236.68	1.039

* Early intervention arm cluster IDs: 1,3, & 4; Delayed arm cluster IDs: 2,5,& 6

** Incidence Rate Ratio (IRR) = incidence rate in Early intervention/incidence rate in Delayed intervention

Table 18: Comparison of baseline cluster-pair TB and ART treatment initiation rates between intervention arms									
	Incidence rateper 100,000 population								
	Pair 1 (Cluster IDs: 3 & 5)*			Pair 2 (Cluster IDs: 1 & 6)*			Pair 3 (Cluster IDs: 4 & 2)*		
	Early intervention	Delayed intervention	IRR**	Early intervention	Delayed intervention	IRR**	Early intervention	Delayed intervention	IRR**
TB cases starting treatment at baseline	68.89	84.65	0.814	104.17	157.03	0.663	172.76	131.08	1.318
Smear-positive TB cases starting treatment at baseline	23.48	31.85	0.737	24.20	43.02	0.562	50.99	35.91	1.420
ART cases starting treatment at baseline	213.45	195.48	1.092	257.56	282.66	0.911	266.34	229.24	1.162

* Cluster IDs 1,3 and 4 are Early arm clusters and clusters IDs 2.5 and 6 are delayed arm clusters.

** Incidence Rate Ratio (IRR) = incidence rate in Early intervention/incidence rate in Delayed intervention.

6.4.3 Assessment of the effectiveness of the intervention in the first 12 months using unadjusted incidence rate ratios

Tables 19 and 20 summarise the TB and ART treatment initiation rates over the first 12 months of the intervention. In the first 12 months, a key feature of these results is the high variability between clusters-pairs and between intervention arms in the unadjusted TB and ART treatment initiation rates. Overall, TB treatment initiation incidence rate ratios were <1 indicating less TB cases starting treatment in the Early intervention arm than the Delayed intervention arm. There was an increase in number of patients starting ART (IRR = 1.221).

Unadjusted analyses of the TB and ART treatment initiation rates in the first 12 months of the intervention between the cluster-pairs (Table 20) show similar patterns of variability. Cluster-pair 2 reported low incidence rate ratios, especially for TB treatment initiation rates, compared to the other pairs. This suggests that TB and ART treatment initiation rates were low in the Early intervention cluster compared to the Delayed intervention cluster in this cluster-pair.

Table 19: Comparison of crude mean counts and incidence rate ratios in TB and ART treatment initiation rates between interventions arms in the first 12 months of intervention					
	Mean number of new cases/month (SD)		Incidence rate per 100,000 population		
	Early intervention*	Delayed intervention*	Early intervention	Delayed intervention	IRR**
TB cases starting treatment in the first 12 months	15.0 (5.0)	19.0 (10.2)	106.79	138.38	0.772
Smear-positive TB cases starting treatment in the first 12 months	3.6 (2.2)	6.0 (3.2)	25.91	41.51	0.624
ART patients starting treatment s	36.2 (13.7)	33.5 (11.3)	4711.34	3859.13	1.221

* Early intervention arm cluster IDs: 1,3, & 4; Delayed arm cluster IDs: 2,5,& 6

** Incidence Rate Ratio (IRR) = incidence rate in Early intervention/incidence rate in Delayed intervention

Table 20: Comparison of cluster-pair incidence rate ratios for TB and ART treatment initiation between intervention arms in the first 12 months of the intervention									
	Incidence rate per 100,000 population								
	Pair 1 (Cluster IDs: 3 & 5)*			Pair 2 (Cluster IDs: 1 & 6)*			Pair 3 (Cluster IDs: 4 & 2)*		
	Early inter-vention	Delayed inter-vention	IRR**	Early inter-vention	Delayed inter-vention	IRR**	Early inter-vention	Delayed interventio n	IRR**
TB cases starting treatment in first 12 months	101.77	87.27	1.166	84.90	189.73	0.447	144.57	127.49	1.134
Smear-positive TB cases starting treatment in first 12 months	24.53	40.58	0.604	20.51	47.33	0.433	35.39	32.92	1.075
ART patients starting treatment in first 12 months	278.68	183.70	1.517	262.48	282.23	0.930	223.75	266.35	0.840

* Cluster IDs 1,3 and 4 are Early arm clusters and clusters IDs 2.5 and 6 are delayed arm clusters.

** Incidence Rate Ratio (IRR) = incidence rate in Early intervention/incidence rate in Delayed intervention

6.5 Impact evaluation based on the generalised linear mixed modelling

Examination of the intervention effects between cluster-pairs (Table 20) revealed considerable variation in TB and ART treatment initiation rates, suggesting that the effect of the intervention varied between cluster-pairs. It has already been widely recognised that statistical approaches that take into account the observed variability between clusters are necessary in the assessment of the effectiveness of the intervention (Thompson, Warn and Turner, 2004; Donner and Klar, 2000; Murray, 1998a; Omar and Thompson, 2000). Generalised linear mixed effects models presented in Chapter 4 were therefore used to account for the observed variability in intervention effect between cluster-pairs while adjusting for cluster-level baseline TB and ART treatment initiation rates and characteristics. Final estimates for the effect of the intervention are based on the appropriate model that better fitted the data based on the model selection criteria given below (see also section 4.4 in Chapter 4) and the pairing of clusters is not taken into account in the analysis.

Model selection criteria used to identify models that best fit the data

To identify the appropriate model, model selection procedures presented in section 4.4 of this dissertation were used, including likelihood ratio tests, BIC and AIC. Different models were fitted for assessing the effectiveness of the intervention, in regards to TB and ART treatment initiation rates, using generalised linear mixed Poisson models described in section 4.2.1. Appropriate model structures were selected by fitting models based on marginal effects or population averages, and random effects models. Random intercepts and random coefficients models were used in the random effects modelling. The appropriate model was selected based on the model selection criteria presented in section 4.4.

The covariates considered in the model selection process included baseline treatment initiation rates or testing uptake rates, gender distribution and the distribution of health facilities physically located in the study areas. The covariates were averaged and fixed across time points (see Appendix 8.4 for sample data used for the analysis).In addition, interaction terms between the intervention indicator and time were also included to assess if the effect of the intervention changed with time. All covariates that were significant or affected the estimates of the intervention were included in the final model.

In selecting the appropriate model, the results for TB treatment initiation rates showed no differences in model fitting between the marginal models and random effects models. All

model selection criteria using either likelihood ratio tests, AIC or BIC were similar for the different models with similar log likelihood estimates. Fitting the random coefficients model did not make any (statistically significant) difference compared to the random intercepts model (likelihood ratio test χ^2 = 2.44, p=0.118). Similarly, when fitting the random effects models to assess ART treatment initiation rates, no improvement was gained over the marginal model (likelihood ratio test χ^2 = 1.68, p=0.642). However, because of the clustering effects in studies involving repeated measurements and with cluster designs and the improvements noted in the estimated standard errors when the random intercepts models were fitted (compared to the marginal models), the random effects models were the most preferred. In all marginal models, standard errors and confidence intervals of the fixed effects were estimated based on sandwich estimators to account for repeated measurement occasions in the data as well as over-dispersion (Liang and Zeger, 1986).

6.5.1 Impact on TB and ART treatment initiation rates in the first 12 months of the intervention

In this section, the results from the first 12 months of the study period are presented, during which only the Early intervention arm received the intervention and the Delayed intervention arm continued to have standard care. The incidence rate ratios determined here assesses the actual effect of engaging informal healthcare providers in integrated TB and HIV community interventions.

6.5.1.1 Impact on TB treatment initiation rates

For TB treatment initiation rates (new TB cases registered for treatment in the first 12 months), the marginal effects model with a single covariate for intervention status estimated the incidence rate ratio (IRR) for the effect of the intervention as 0.824 (95% CI: 0.513 to 1.34). This is slightly higher than the unadjusted incidence rate ratios of 0.772 shown in Table 19 potentially due to differences in use of the cluster level populations: Table 19 used cluster level populations directly to estimate the incidence rate ratios while the current estimate (IRR=0.824) was derived using monthly counts of new TB cases with differences in cluster level populations fixed when the offset was used. Using a random intercepts and random coefficients models, the IRR were 0.876 (95% CI:0.583 to 1.315) and 0.876 (95% CI: 0.586 to 1.316) respectively.

Thus, using the marginal model without adjusting for covariates revealed that persons in Early intervention clusters were on average 17.6% less likely (per month) to initiate TB treatment than their counterparts in the Delayed intervention clusters, but this difference was not significant at the 5% level (p=0.424 for marginal model). Using the random intercepts model, TB patients starting treatment (per month and per cluster) were 12.4% less likely to initiate treatment (again, not statistically significant). The standard deviation of the random intercepts parameter ($\sqrt{\sigma_{11}^2}$) was estimated to be 0.243 (95% CI: 0.132 to 0.488).

When baseline TB treatment initiation rates were considered, the incidence rate ratios based on the marginal effects and random intercepts were respectively 0.938 (95% CI: 0.650 to 1.354) and 0.963 (95% CI: 0.695 to 1.333), neither of which were statistically significant. A reduction in the variability of the impact estimate was observed when the baseline TB treatment initiation rates were adjusted for the random intercepts standard deviation of 0.180 (95% CI: 0.093 to 0.350). The log likelihood, AIC and BIC values were respectively -236.788, 481.575, and 490.682.

When the baseline gender distribution and the distribution of health facilities initiating TB treatment in rural areas were included in the model, the effect size estimates based on the marginal and random intercepts models were respectively 1.07 (95% CI: 0.803 to 1.413) and 1.08 (95% CI: 0.787 to 1.476), suggesting about 7 - 8% more cases per month were detected in the Early intervention areas. Inclusion of baseline gender distribution and the distribution of health facilities offering TB treatment initiations in the model resulted in a further reduction in the variability of the impact estimate in the random intercepts model with standard deviation of 0.112 (95% CI: 0.050 to 0.253), suggesting an improvement in the model fit. The corresponding log likelihood, AIC and BIC values were respectively -234.554, 481.108, and 494.768.

To assess whether the effect of the intervention varied over the duration of intervention, interaction terms between measurement occasion and intervention status were included in the model; the results obtained are presented in Table 21. This inclusion did not produce a marked improvement in the model fit (values for log likelihood, AIC and BIC were respectively -234.554, 481.108, and 494.768). The corresponding fixed effects interaction term was not statistically significant. Using the marginal model, the IRR for the intervention increased from 1.07 to 1.18 (i.e. 1.204 * 0.981) after adjusting for the covariates (including the interaction terms). This suggests that the intervention increased the likelihood of initiating

treatment for TB cases in the Early intervention group by 18% compared to the Delayed intervention group, although this estimated effect size remained statistically non-significant. Nevertheless, this does give a clear indication that a model that allows for the effect of the intervention to vary with measurement occasions is superior to one that assumes that the effect of the intervention is constant over duration of the intervention.

Only baseline TB treatment initiation rates and the distribution of the health facilities remained statistically significant, indicating that increase in number of health facilities offering TB treatment initiations was also associated with an increase in TB treatment initiations, which in turn, depended on the baseline TB treatment initiations in each cluster. The intervention effect size estimate slightly improved when a random intercepts model was fitted: IRR increased from 1.18 to IRR 1.20 (1.219*0.981), suggesting a 20% increase in the number of TB cases starting treatment in the Early intervention areas compared to the Delayed intervention areas (Table 21). No further improvement over the random intercepts model was detected when a random coefficients model was fitted (Table 21). In both the random intercepts and random coefficients models, the distribution of health facilities and baseline TB treatment initiation rates were significantly associated with TB treatment uptake rates.

To determine the level of clustering of the outcome measures (i.e. TB treatment initiation rates), the intracluster correlation coefficients were estimated using the formula given in (4.13) repeated here as:

$$ICC = \frac{\sigma_{11}^2}{\sigma_{11}^2 + \phi . \ln\left(\frac{1}{exp(\beta_0 + \beta_1 x_{ij} + ..)} + 1\right)}$$

where σ_{11}^2 is the random intercepts variance, and ϕ is the overdispersion scale parameter derived by dividing the Pearson chi-square by its degrees of freedom. Using the estimated standard deviation of the random intercepts parameter $\sqrt{\sigma_{11}^2}$ of 0.112 and the overdispersion scale parameter ϕ of 3.478, the intracluster correlation coefficient was 0.067 (Table 21) suggesting moderate clustering of the outcome measures. The estimated ICC in TB treatment initiations is comparable to the based ICCs (i.e. 0.00154 and 0.081) used in the simulation studies presented in Chapter 5.

Using the random intercepts model, a unit increase in the number of health facilities was associated with an estimated 24% increase in the TB treatment initiations per cluster while

controlling for the other covariates. An extra TB case initiated on TB treatment at baseline was associated with an estimated 3% increase in TB treatment initiations per cluster. The results also suggest that female gender distribution was associated with an 18% increase in TB treatment initiations though not significant and that the TB treatment initiations did not change much with increased time of the intervention (0.4%), controlling for the other covariates and was again not significant at the 5% level.

6.5.1.2 Impact on smear- positive TB treatment initiation rates

To assess whether the engagement of informal healthcare providers increased the number of new smear-positive TB patients starting treatment during the first 12 months of the intervention, marginal, random intercepts and coefficients models were fitted to the data. Covariate adjustment was made for baseline smear-positive TB treatment initiation rates, gender and health facility distribution. In addition, an interaction term between intervention status and time was included in the model as well as a time to see if there was any change in the smear-positive TB treatment initiation rates over time (Table 22). The IRRs for the marginal, random intercepts and coefficients models were all identical; variation between clusters was zero while the correlation between the two random effects was 0.422, suggesting the need for fitting the random coefficient model despite model fit indices being similar across the three models fitted (Table 22).

When the effect of the intervention was assessed by using the marginal, random intercepts and random coefficients models with a single covariate for intervention status estimated the incidence rate ratio for the effect of the intervention as 0.622 (95% CI: 0.427 to 0.905) for the marginal effects model, 0.647 (95% CI: 0.450 to 0.930) for the random intercepts and 0.646 (95% CI: 0.448 to 0.931) for the random coefficients models.

Table 21: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring TB treatment initiation rates in the first 12 months of the intervention

	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	1.204	0.957, 1.515	0.112	1.219	0.835, 1.779	0.305	1.239	0.881,1.744	0.218
Base[5]	1.025	1.003, 1.048	0.025	1.027	1.001, 1.053	0.040	1.025	1.000,1.050	0.046
Occasion[6]	1.004	0.978, 1.031	0.748	1.004	0.983, 1.026	0.694	1.004	0.983,1.026	0.694
Gender	1.255	0.306, 5.140	0.752	1.181	0.084, 16.50	0.902	0.912	0.107,7.743	0.932
Treat*Month[7]	0.981	0.944, 1.012	0.335	0.981	0.950, 1.014	0.253	0.981	0.950,1.014	0.253
H. facility[8]	1.247	1.096, 1.419	0.001	1.244	1.053, 1.470	0.010	1.248	1.067,1.458	0.005
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.112			0.000	
$\sqrt{\sigma_{22}^2}$								0.179	
ICC[10]								0.0674	
Model fit indices									
Log likelihood		-236.975			-233.781			-232.561	
AIC		483.952			481.108			483.123	
BIC		495.335			494.768			503.613	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline TB treatment initiation rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of health facilities offering TB treatment initiations; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

When all the covariates were included in the model (baseline smear-positive TB treatment initiations, time, gender, distribution of health facilities initiating TB treatment and interaction terms),the results indicate that the intervention significantly reduced TB smear-positive cases starting treatment on average by 28% (IRR=0.722*1.002 = 0.723, $p<0.05$) when the marginal effects model was fitted, suggesting that fewer smear-positive cases were detected in the Early intervention arm than in the Delayed arm (potentially due to more health facilities initiating

TB treatment (see Figure 8 in section 6.4.1)) . However, using the random intercepts and random coefficients models, the reductions in smear-positive cases starting treatment were not significant at the 5% level (Table 22). In the marginal model, all covariates included in the model were significantly associated with the number of smear-positive TB cases starting TB treatment except for the interaction terms between intervention status and time (Table 22).

When the random intercepts model with all covariates were included, the standard deviation of the random intercepts parameter $\sqrt{\sigma_{11}^2}$ was 0.0 and the estimated intracluster correlation coefficient was 0.0, suggesting no between cluster variation in smear-positive TB treatment initiations rates (Table 22). The estimated overdispersion scale parameter ϕ for the smear-positive TB treatment initiations was 3.085.

Using the random effects models, only baseline TB treatment initiation rates and time period of the intervention were significantly associated with smear-positive TB patients starting treatment with an estimated 20% increase and 5% decrease respectively. Although a unit increase in female gender distribution was associated with approximately a four times increase in number of smear-positive TB cases starting treatment per cluster, the increase had a borderline statistical significance (p=0.07) for the two random effects models after controlling for the other covariates. Increasing the number of health facilities initiating TB treatment did not result in increasing the number of smear-positive TB cases starting treatment (Table 22).

6.5.1.3 Impact on ART treatment initiation rates

To assess the effect of the intervention on ART treatment initiation rates, the first 12 months data following the intervention were used. Employing the marginal effects, random intercepts and random coefficients models with only intervention status fitted showed that the intervention increased ART treatment initiation rates by 12.8% (IRR=1.128; 95% CI: 0.895 to 1.422; p=0.308) in the marginal model, and 11.2% (IRR=1.112; 95% CI: 0.891 to 1.389; p=0.34) for the random intercepts and random coefficients models. These estimated effect sizes slightly improved to 13.1% (IRR=1.131; 95% CI: 0.919 to 1.391; p=0.244) in the marginal model and 11.5% (IRR=1.115; 95% CI: 0.907 to 1.370; p=303) for both the random intercepts and random coefficients models when baseline ART initiation rates were added to the model. Finally, when an improved model was fitted with intervention status, baseline ART initiation rates, gender distribution and distribution of health facilities, the incidence rate ratios were 1.301 (95% CI: 1.101 to 1.536, p=0.002), 1.286 (95% CI: 1.060 to 1.559, p=0.011) and 1.316 (95%

CI: 1.086 to 1.1595, p=0.005) for the marginal model, random intercepts and random coefficients models respectively. Thus, the intervention was associated with an estimated 30.1%, 28.6% and 31.6% increase in ART treatment initiations when the marginal, random intercepts and random coefficients models were fitted after controlling for the other covariates.

Table 22: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring access rates to TB treatment services among smear positive patients in the first 12 months of the intervention

	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	0.722	0.540, 0.967	0.029	0.722	0.452, 1.154	0.174	0.722	0.452, 1.154	0.174
Base[5]	1.198	1.113, 1.289	<0.001	1.198	1.067, 1.344	0.002	1.198	1.067, 1.344	0.002
Occasion[6]	0.952	0.921, 0.985	0.005	0.952	0.916, 0.990	0.014	0.952	0.916, 0.990	0.014
Gender	3.912	1.640, 9.335	0.002	3.913	0.883, 17.327	0.072	3.913	0.883, 17.327	0.072
Treat*Month[7]	1.002	0.926, 1.085	0.959	1.002	0.940, 1.068	0.949	1.002	0.940, 1.068	0.949
H. facility[8]	0.932	0.879, 0.987	0.017	0.932	0.803, 1.080	0.348	0.932	0.803, 1.080	0.348
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.000			0.000	
$\sqrt{\sigma_{22}^2}$								0.000	
ICC[10]								0.000	
Model fit indices									
Log likelihood		-157.203			-157.203			-157.203	
AIC		324.406			330.406			334.406	
BIC		335.790			348.620			357.173	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline smear positive TB treatment initiation rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of health facilities offering TB treatment initiations; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation coefficient.

Using these marginal, random intercepts and random coefficients models, the distribution of health facilities and intervention status were the only covariates that were significantly associated with ART initiation rates. A unit increase in number of health facilities was associated with a 19.4% (IRR=1.194, 95% CI: 1.069 to 1.333, p=0.002), 20% (IRR=1.200, 95% CI: 1.030 to 1.399, p=0.020) and 23.3% (IRR=1.233, 95% CI: 1.104 to 1.378, p<0.001)increase in the number of ART patients starting treatment (p=0.002) for the marginal, random intercepts and random coefficients models.

To assess whether the effect of the intervention varied with time, an interaction term was included in the model between time and intervention status. The overall incidence rate ratio increased from 1.301 to 1.347 (IRR=1.356*0.993=1.347) using the marginal model, from 1.286 to 1.332 (IRR=1.341*0.993=1.332) in the random intercepts model and from 1.316 to 1.362 (IRR=1.372*0.993) in the random coefficients model and were significant at the 5% level (Table 23). Only the distribution of health facilities initiating ART was significantly associated in increasing the number of patients starting ART for all the three models. However, fitting the random coefficients model and the inclusion of an interaction term did not improve the model fit (see model fit indices in Table 23).

The intracluster correlation coefficient estimated after fitting the random intercepts model adjusting for all covariates was 0.0402, suggesting moderate correlation in ART treatment initiation rates which is comparable to the baseline ICCs used in the simulation studies (Table 23). The corresponding standard deviation for the random intercept parameter $\sqrt{\sigma_{11}^2}$ and the overdispersion scale parameter ϕ used to estimate the ICC were 0.056 and 3.486 respectively.

In all the three models fitted, baseline ART initiation rates and time period of the intervention were less likely to be associated with number of patients starting ART and were only associated with an estimated 0.3% increase and 0.8% decrease in the number of patients starting ART respectively. The results further suggest that female gender distribution was associated with reduced access to ART initiations and ranged from an estimated 67% when the marginal model was fitted to 91% reduction for a random coefficients model but was not significant at 5% level.

Table 23: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring access rates to ART services in the first 12 months									
	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	1.356	1.003, 1.833	0.048	1.341	1.053, 1.706	0.017	1.372	1.078, 1.746	0.010
Base[5]	1.003	0.993, 1.013	0.567	1.003	0.994, 1.011	0.545	0.998	0.989, 1.008	0.741
Occasion[6]	0.992	0.962, 1.024	0.619	0.992	0.976, 1.008	0.344	0.992	0.976, 1.008	0.344
Gender	0.319	0.002, 44.68	0.650	0.328	0.002, 59.145	0.674	0.086	0.00, 37.145	0.429
Treat*Month[7]	0.993	0.949, 1.040	0.778	0.993	0.971, 1.016	0.572	0.993	0.971, 1.016	0.572
H. facility[8]	1.194	1.069, 1.333	0.002	1.200	1.030, 1.399	0.019	1.233	1.104, 1.378	<0.001
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.056			0.000	
$\sqrt{\sigma_{22}^2}$								0.108	
ICC [10]								0.0402	
Model fit indices									
Log likelihood		-285.874			-285.034			-283.855	
AIC		581.748			586.068			587.710	
BIC		593.131			604.281			610.477	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline ART initiation rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of health facilities offering ART treatment initiations; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

6.5.2 Impact on secondary outcome measures in the first 12 months

In this section, the secondary outcome measures of the intervention are analysed. The secondary outcome measures recorded included the number of new clients/patients who accessed TB and HIV testing services. All health facilities in the rural areas offered HIV testing and counselling services, but TB microscopy testing is done only in selected health facilities in rural and urban areas.

Numbers of presumptive TB cases were obtained from all TB testing sites. Individuals accessing TB testing were allocated to the study clusters based on their residential address. However, because HIV testing registers do not record an individual's personal residential address, data collection for this measure was mainly confined to sites located in rural areas. Thus, HIV testing registers from health facilities in the rural areas were used and the individuals accessing HIV testing in such health facilities were allocated to the study clusters where the respective health facilities were located.

6.5.2.1 Impact on TB and HIV testing access rates in the first 12 months of intervention

In this section, results for the analysis using data for the first 12 months are presented for the TB and HIV testing uptake rates.

6.5.2.2 Impact on TB testing uptake rates in the first 12 months of the intervention

As some components of the intervention involved increasing awareness about TB in the community, including the recognition and referral of TB cases, an investigation was made to see if these activities generally increased the number of presumptive TB cases accessing TB testing services. Using the first 12 months' data with a marginal model with baseline presumptive TB testing uptake rates, intervention status, distribution of testing sites, gender distribution, time, and an interaction term between intervention status and time included as cluster-level covariates, all the covariates were significantly associated with TB testing uptake rates (Table 24).

The effect of the intervention on increasing TB testing uptake rates among presumptive TB cases was estimated at 1.152 (IRR=1.215*0.948=1.152) after adjusting for all covariates considered in the model when a marginal model was fitted. This suggests that there was a 15.2% increase in the number of presumptive TB cases accessing TB testing services in the Early intervention arm compared to the Delayed arm. Only baseline TB testing uptake was significantly associated with increased uptake rates following the intervention with an estimated 0.4% increase. The results further suggest that being female was more associated

with reduced TB testing uptake rate by 97% after controlling for all covariates considered in the model.

When the random effects models (random intercepts and random coefficients models) were fitted, all the variables except the baseline outcome measurements were significantly associated with testing uptake rates after controlling for the other covariates (see Table 24). The interaction effect between intervention status and time was significantly associated with TB testing uptake when all the models were fitted, suggesting that the effect of the intervention depended on the duration of the intervention. When the random intercepts model with all covariates was fitted, the estimated intracluster correlation coefficient was 0.0192 suggesting that there was minimal clustering effects in the TB testing uptake rates. The corresponding standard deviation for the random intercepts parameter $\sqrt{\sigma_{11}^2}$ and the overdispersion scale parameter ϕ used to estimate the ICC were 0.0386 and 5.12 respectively (Table 24). Fitting the random effects models did not improve much compared to the marginal effects model (see model fit indices in Table 24).

A graphical presentation of the effect of the intervention in increasing TB testing uptake, monthly and cumulative uptake rates suggested that the numbers of TB cases accessing testing services in the first 12 months showed slightly more presumptive TB cases in the Early intervention arm accessing TB testing services compared to the Delayed arm (Figure 14 shown in section 6.6.2). When the intervention was rolled out to the Delayed intervention arm, there was no difference in cumulative rates towards the end of the second phase (11 month) of the intervention period described in section 6.6.2.

Of the presumptive TB cases which accessed testing sites in the first 12 months, 8.6% (375/4384) were identified as smear-positive 7.5% (165/2189) in the Early intervention arm and 9.6% (210/2,195) in the Delayed intervention arm. The effect of the intervention in increasing smear-positive cases was then assessed by fitting marginal, random intercepts and coefficients models adjusted for intervention status, baseline smear-positive cases in the testing sites, gender distribution, time and distribution of TB microscopy sites and an interaction term for intervention status and time.

Table 24: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring access rates to testing services for presumptive TB cases in the first 12 months

	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	1.215	1.071, 1.379	0.003	1.215	1.001, 1.475	0.049	1.215	1.001, 1.475	0.049
Base[5]	1.004	1.004, 1.004	<0.001	1.004	0.998, 1.010	0.187	1.004	0.998, 1.010	0.187
Occasion[6]	0.940	0.906, 0.975	0.001	0.940	0.908, 0.973	<0.001	0.940	0.908, 0.973	<0.001
Gender	0.028	0.028, 0.029	<0.001	0.028	0.003, 0.233	0.001	0.028	0.003, 0.233	0.001
Treat*Month[7]	0.948	0.913, 0.986	0.013	0.948	0.902, 0.998	0.040	0.948	0.902, 0.998	0.040
testing sites[8]	0.920	0.919, 0.921	<0.001	0.920	0.848, 0.999	0.045	0.920	0.848, 0.999	0.045
Random part[9]									
$\sqrt{\sigma_{11}^2}$				0.0386				0.000	
$\sqrt{\sigma_{22}^2}$								0.004	
ICC[10]								0.0192	
Model fit indices									
Log likelihood		-165.317		-165.63				-165.82	
AIC		340.635		347.268				349.642	
BIC		348.553		359.936				363.893	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline TB testing uptake rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of TB testing sites; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

The results for the marginal and random intercepts models were similar. Fitting the random intercepts or coefficients models did not greatly improve the model fit (similar log likelihoods, AIC and BIC) (Table 25). However, there was an attenuation of the effect size when the random coefficients model was used (Table 25). The intervention resulted in an increase in TB smear-positive cases detected at the testing sites of 11.7% (IRR=1.123*0.995=1.117) when the marginal and random intercepts models were fitted; however, this effect was not statistically significant at the 5% level for either the marginal or random intercepts models

owing to the small sample sizes available for this analysis as demonstrated in the simulations presented in Chapter 5. None of the covariates included in the model were significantly associated with the number of smear positive TB cases detected in all the three models(Table 25). When the model with all covariates was fitted using the random intercepts model, the estimated intracluster correlation coefficient was 0.0 and the corresponding random intercept parameter $\sqrt{\sigma_{11}^2}$ and the overdispersion scale parameter ϕ used to estimate the ICC were respectively 0.0 and 3.09, suggesting no between cluster variability in presumptive TB testing uptake rates (Table 25).

Table 25: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for smear -positive TB cases identified at testing sites during the first 12 months

	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	1.123	0.700, 1.803	0.631	1.123	0.686, 1.840	0.645	1.073	0.608, 1.891	0.809
Base[5]	1.176	0.963, 1.436	0.113	1.176	0.947, 1.459	0.142	1.149	0.879, 1.503	0.309
Occasion[6]	0.982	0.961, 1.004	0.107	0.982	0.945, 1.021	0.365	0.982	0.945, 1.021	0.365
Gender	0.421	0.045, 3.915	0.447	0.421	0.014, 13.10	0.622	0.490	0.013, 18.213	0.699
Treat*Month[7]	0.995	0.963, 1.027	0.738	0.995	0.938, 1.054	0.854	0.995	0.938, 1.054	0.854
Testing sites[8]	1.048	0.793, 1.385	0.741	1.048	0.805, 1.364	0.727	0.996	0.681, 1.458	0.985
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.000			0.001	
$\sqrt{\sigma_{22}^2}$								0.106	
ICC[10]								0.000	
Model fit indices									
Log likelihood		-161.624			-161.624			-161.587	
AIC		333.247			339.247			341.174	
BIC		344.631			357.461			361.664	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline smear positive TB testing uptake rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of TB testing sites; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

6.5.2.3 Impact on HIV testing uptake rates

To assess whether the engagement of informal health care providers in sensitising communities to the need for voluntary HIV testing improved uptake rates, the HIV testing uptake rates between the Early and Delayed intervention arms were compared.

At baseline, 43% of HIV tests (28,792/66,612) were registered in HIV testing sites in the Delayed intervention areas while the remaining 57% (37820/66612) were registered in the Early intervention areas. After the first year of the study, 54.5% (29,087) of HIV testing cases were reported in the Early arm, a slight decline in the relative proportion of HIV tests before adjusting for the potential confounders.

To assess the effect of the intervention on HIV testing uptake rates in the first 12 months of the intervention, marginal, random intercepts and random coefficients models were fitted. When only the variable for intervention status was fitted, the intervention increased HIV testing uptake by 22.1% (IRR=1.221; 95% CI: 0.810 to 1.839; p=0.340), 19.8% (IRR=1.198; 95% CI: 0.755 to 1.900; p=0.443) and 19.8% (IRR=1.198; 95% CI: 0.755 to 1.899; p=0.444) for the marginal, random intercepts and random coefficients models respectively. Using intervention status, baseline HIV testing uptake rates, gender distribution, HIV testing sites, measurement occasion and an interaction term between intervention status and measurement occasion as covariates, the HIV testing uptake rate significantly increased by 61.0% (IRR=1.626*0.990=1.61, p=0.001) when the marginal and random effects models were used (Table 26).

Comparison of the log likelihood values between the marginal and random effects models indicated that the model with a random effects specification better fitted the data (deviance=181.352, chi-square distribution with 1 degree of freedom, p<0.001). This suggests that the intercepts of the random effects varied significantly between the clusters.

The intracluster correlation coefficient estimated after fitting the random intercepts model adjusting for all covariates was 0.1287, suggesting a slightly higher between cluster variability in HIV testing uptake rates than the other outcome measures (Table 26). The corresponding standard deviation for the random intercept parameter $\sqrt{\sigma_{11}^2}$ and the overdispersion scale parameter ϕ used to estimate the ICC were 0.0987 and 115.326 respectively, suggesting a very high overdispersion in HIV testing uptake rates which was accounted for by use of robust estimates. The estimated ICC was slightly higher than the ICC of 0.081 from the baseline data

from the Triage Plus study used in the simulation studies presented in Chapter 5 though the statistical power was still maintained due to the high effect size of 61%.

All covariates in the models, except baseline HIV testing measurements, were significantly associated with HIV testing uptake rates in the marginal model. However, for the two random effects models, only the baseline HIV testing uptake was not significantly associated with HIV testing uptake rates.

In all the models, the distribution of HIV testing sites (a proxy measure for geographical proximity) was significantly associated with increasing HIV testing uptake rates (estimated increase of 13.4% in the marginal model). A unit increase in the number of health facilities offering HIV testing was associated with an estimated 11.8% increase in the HIV testing uptake per cluster while controlling for the other covariates when the random intercepts and random coefficients models were fitted. The results also suggest that female gender distribution was associated with a decreased service access for HIV testing by over 98% in all the three models fitted, suggesting that females were less likely to access HIV testing compared to males.

With the interaction term between intervention status and time being significantly associated with HIV testing uptake when the random effects models were fitted suggests that the effect of the intervention varied with time of the intervention. When plots for the unadjusted monthly and cumulative HIV testing uptake rates over the first 12 months were plotted, the results showed a consistent increase in the number of new clients testing for HIV in the Early intervention arm compared to the Delayed intervention arm (Figure 11). This suggests that the intervention was most effective in increasing the numbers of persons being tested for HIV in the community. However, due to challenges noted in the data after 12 months where some sites registered very few cases, which was atypical for these sites and that some registers could not be traced in some sites for specific months especially in sites with outreach services (see section 2.6.1), the data from 12 months onward was not considered.

Table 26: Maximum likelihood estimates for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring access rates to HIV testing in the first 12 months

	Marginal effects Poisson	Random intercepts Poisson	Random coefficients Poisson

Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]			
Treat[4]	1.626	1.443,1.832	<0.001	1.622	1.221,2.155	0.001	1.622	1.221,2.155	0.001
Base[5]	0.999	0.999,0.9998	0.016	1.000	0.999,1.000	0.279	1.000	0.999,1.000	0.279
Occasion[6]	0.949	0.934,0.964	<0.001	0.949	0.945,0.953	<0.001	0.949	0.945,0.953	<0.001
Gender	0.011	0.003,0.039	<0.001	0.018	0.002,0.183	0.001	0.018	0.002,0.183	0.001
Treat*Month[7]	0.990	0.960,1.020	0.489	0.990	0.984,0.995	<0.001	0.990	0.984,0.995	<0.001
H.facility[8]	1.134	1.079,1.193	<0.001	1.118	1.046,1.195	0.001	1.118	1.046,1.195	0.001
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.0987			0.098	
$\sqrt{\sigma_{22}^2}$								0.000	
ICC[10]								0.1287	
Model fit indices									
Log likelihood		-1933.180			-1751.828			-1752.943	
AIC		3876.360			3519.656			3523.885	
BIC		3887.308			3537.173			3543.592	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline HIV testing uptake rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of HIV testing sites in rural areas; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

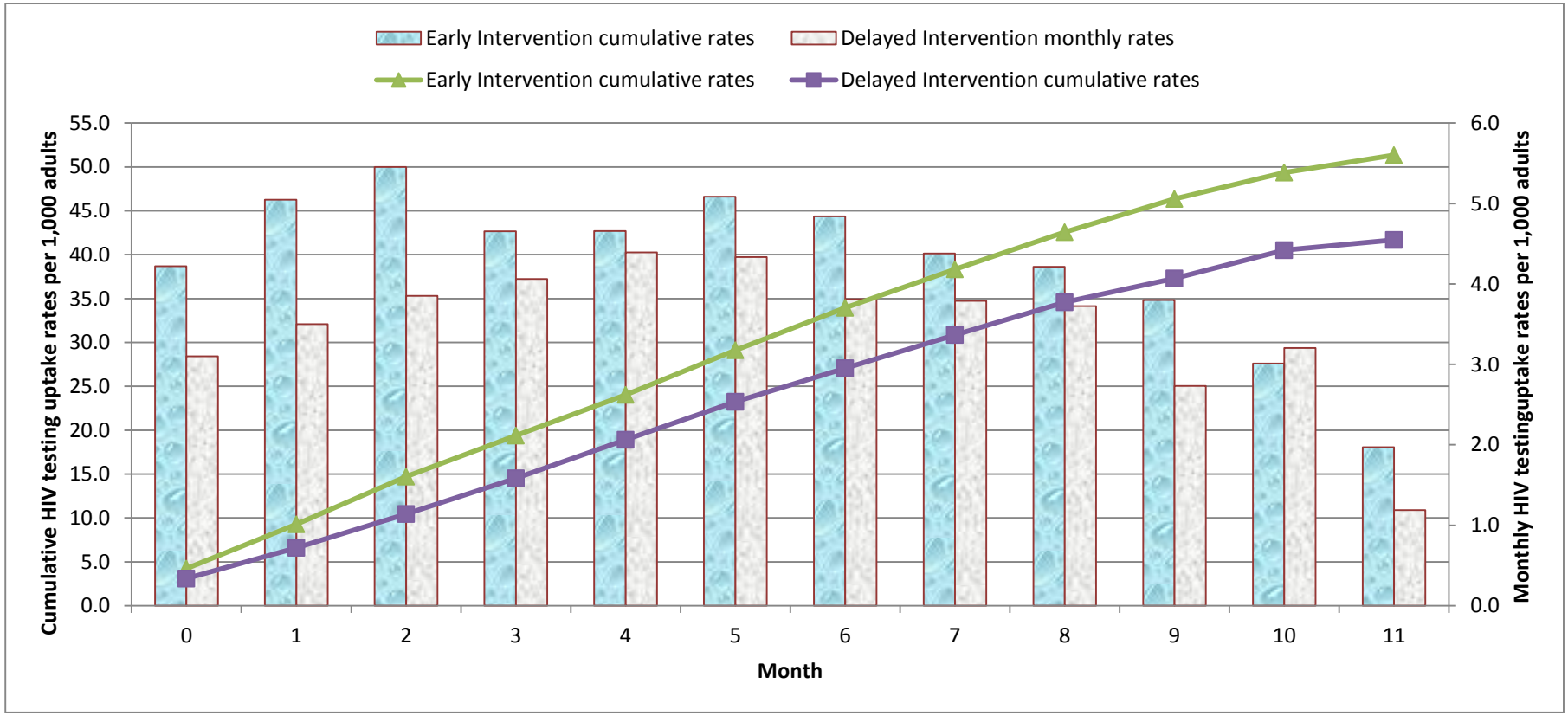


Figure 11: Monthly and cumulative HIV testing uptake rates per 1,000 adults.

Results in individuals aged 12 years and above over the first 11 months of the intervention before the scale up of the Delayed intervention arm are shown. The line graphs represent the cumulative HIV testing uptake rates. The purple cubed line graph indicates the Delayed intervention arm and the green triangle line graph the Early intervention arm. The solid bar graphs represent the monthly HIV testing uptake rates per 1,000 adults for every month over the 11 months of the intervention. The blue bars represent the Early Intervention Arm while the grey bars represent the Delayed Intervention Arm.

6.6 Impact on TB and ART treatment initiation and testing uptake rates in the second phase (11 months) of the intervention

In this section, results based on the second half (11 months) of the study period are presented, during which time both the Delayed and Early arms received the intervention (see Figure 3 in section 2.3.3.5). Thus, the incidence rate ratios determined here compare the incidence rates in the Early arm against those in the Delayed intervention group when both of these groups were receiving the intervention (but for different lengths of time). In all the analyses, same baseline covariates were used when assessing the effectiveness of the intervention.

6.6.1 Impact on TB and ART treatment initiation rates in the second phase (11 months) of the intervention

In this section, results for the analysis using data for the next 11 months after scaling up the intervention to the Delayed arm are presented for the TB and HIV treatment initiation rates.

6.6.1.1 Impact on TB treatment initiation rates in the second phase (11 months) of the intervention

The effect size estimate based on the marginal effects model with a single covariate contrasting intervention status was IRR = 0.758 (95% CI: 0.445 to 1.290). This is slightly lower than the IRR of 0.824 (95% CI: 0.513 to 1.24)obtained in the first 12 months. When baseline TB treatment initiations were adjusted for, the incidence rate ratio increased to 0.909 (95% CI: 0.645 to 1.280), which was similar to the IRR of 0.938 (95% CI: 0.650 to 1.354), obtained during the evaluation of the first 12 months of the study period.

Using the random intercepts and random coefficients models, the effect size estimate with a single covariate contrasting intervention status was IRR = 0.821 (95% CI: 0.516 to 1.306) for the random intercepts model and IRR = 0.824 (95% CI: 0.520 to 1.307) for the random coefficients model. These results were slightly lower than the IRR of 0.876 (95% CI: 0.583 to 1.313) for the random intercepts and 0.878 (95% CI: 0.585 to 1.316) for the random coefficients models obtained in the first 12 months. When baseline TB treatment initiations were adjusted for, the incidence rate ratios were 0.931 (95% CI: 0.686 to 1.264) and 0.924 (95% CI: 0.669 to 1.274) for the random intercepts and random coefficients models respectively, which were slightly lower than the IRR of 0.963 (95% CI: 0.695 to 1.333) for the random intercepts and 0.990 (95% CI: 0.723 t0 1.357) for the random coefficients model obtained during the first 12 months of the study.

When the most improved model was fitted, with gender, distribution of health facilities, time, same baseline TB treatment initiation rates used in the first 12 months and interaction terms

between time and intervention status (as implemented in the first 12 months' data), the incidence rate ratio comparing the Early intervention and Delayed intervention were 0.552 (0.529 * 1.044), 0.555 (0.532*1.044) and 0.562 (0.538*1.044) for the marginal, random intercepts and random coefficients models which were lower than the estimates obtained in the first 12 months (see Table 21), suggesting that the intervention was effective in initiating more TB treatment in the Delayed intervention arm when the intervention was scaled up in this arm.

The intracluster correlation coefficient estimated after fitting the random intercepts model adjusting for all covariates was 0.0073, suggesting low clustering in TB treatment initiation rates in the next 11 months of the intervention (Table 27). The corresponding standard deviation for the random intercept parameter $\sqrt{\sigma_{11}^2}$ and the overdispersion scale parameter ϕ used to estimate the ICC were 0.046 and 3.497 respectively. Fitting the random coefficients model did not greatly improve the model fit when compared to the random intercepts model (see the log likelihood, AIC and BIC in Table 27).

All covariates considered in the marginal model were significantly associated with TB treatment initiation rates (Table 27). Using the random intercepts and random coefficients models, no significant improvements in the effect of the intervention were observed [IRR=0.555 (0.532*1.044) for the random intercepts model, IRR=0.562 (0.538*1.044) for the random coefficients model, neither of which were statistically significant] (Table 27). Thus, (numerically) the effect of scaling up the intervention to the Delayed arm compared to the Early arm was associated with a 45% reduction in the number of new TB patients starting treatment in the Early arm when the marginal model was used and about 44% reduction when the random intercepts model was used.

In the random effects models only baseline TB treatment initiations, time, interaction effect and distribution of health facilities initiating treatment were significantly associated with TB treatment initiation rates. Using the random intercepts model, only the distribution of health facilities, baseline TB treatment initiations and interaction effect were associated with increasing number of TB cases starting treatment with an estimated magnitude of 27%, 3.8% and 4.4% respectively. The results also indicate that being female was associated with reduced TB treatment initiation rates by 69% as observed in the first 12 months (Table 27).

Table 27: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring overall access rates to TB treatment initiation rates using the data from the second phase of the intervention (11 months)

	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	0.529	0.387, 0.722	<0.001	0.532	0.247, 1.145	0.106	0.538	0.249, 1.161	0.114
Base[5]	1.038	1.021, 1.054	<0.001	1.038	1.020, 1.057	<0.001	1.038	1.017, 1.059	<0.001
Occasion[6]	0.974	0.959, 0.989	0.001	0.974	0.948, 0.999	0.049	0.974	0.948, 0.999	0.049
Gender	0.320	0.117, 0.874	0.026	0.314	0.046, 2.139	0.237	0.274	0.039, 1.931	0.194
Treat*Month[7]	1.044	1.027, 1.062	<0.001	1.044	1.003, 1.088	0.036	1.044	1.003, 1.088	0.036
H.facility[8]	1.269	1.155, 1.395	<0.001	1.268	1.123, 1.431	<0.001	1.268	1.111, 1.447	<0.001
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.046			0.041	
$\sqrt{\sigma_{22}^2}$								0.114	
ICC[10]								0.0073	
Model fit indices									
Log likelihood		-190.639			-190.541			-189.999	
AIC		391.279			397.083			397.980	
BIC		402.227			414.600			417.687	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline TB treatment initiation rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of health facilities offering TB treatment initiations; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

When TB patients starting treatment were assessed in terms of monthly and cumulative TB treatment initiation rates (computed per 10,000 adults aged 12 years and above), both the unadjusted monthly treatment initiation rates and the cumulative treatment initiation rates were high in the Delayed intervention group compared to the Early intervention group. When the intervention was added to the Delayed intervention arm in the 13th month, a widening gap was observed between the two arms in terms of cumulative TB treatment initiation rates (Figure 12) potentially due to the high number of health facilities in the Delayed arm

compared to the Early arm (2 health facilities in Early arm and 4 health facilities in the Delayed arm as shown in Figure 8 in section 6.4.1). At baseline (i.e. the preceding 12 months before the intervention) similar trends were observed with more cases in the Delayed arm than in the Early arm (Table 17). Thus, with more health facilities initiating TB treatment in the Delayed arm compared to the Early arm coupled with a widening gap in the cumulative TB treatment initiation rates after scaling the intervention to the Delayed arm (increased number of TB cases starting treatment in Delayed arm), it would appear that the effect of the intervention depended on the distribution of health facilities between the two intervention arms.

6.6.1.2 Impact on smear- positive TB treatment initiation rates in the second phase (11 months) of the intervention

When the smear-positive TB patients starting treatment were analysed over the final 11 months of the intervention, the effect size based on the marginal effects model with a single covariate contrasting intervention status was estimated as IRR = 0.781 (95% CI: 0.551 to 1.107), which is slightly higher than the IRR = 0.622 (95% CI: 0.427 to 0.905) obtained when the first 12 months' data were used. Using the random intercepts and random coefficients models, the incidence rate ratios for the next 11 months were 0.799 (95% CI: 0.575 to 1.110) and 0.804 (95% CI: 0.585 to 1.105) respectively. These estimates were slightly higher than the 0.647 (95% CI: 0.450 to 0.930) for the random intercepts model and 0.646 (95% CI: 0.448 to 0.931) for the random coefficients model obtained when the first 12 months were used (see section 6.5.1.2), suggesting that the number of smear-positive cases starting TB treatment slightly increased in the Early arm compared to the Delayed arm before controlling for other covariates.

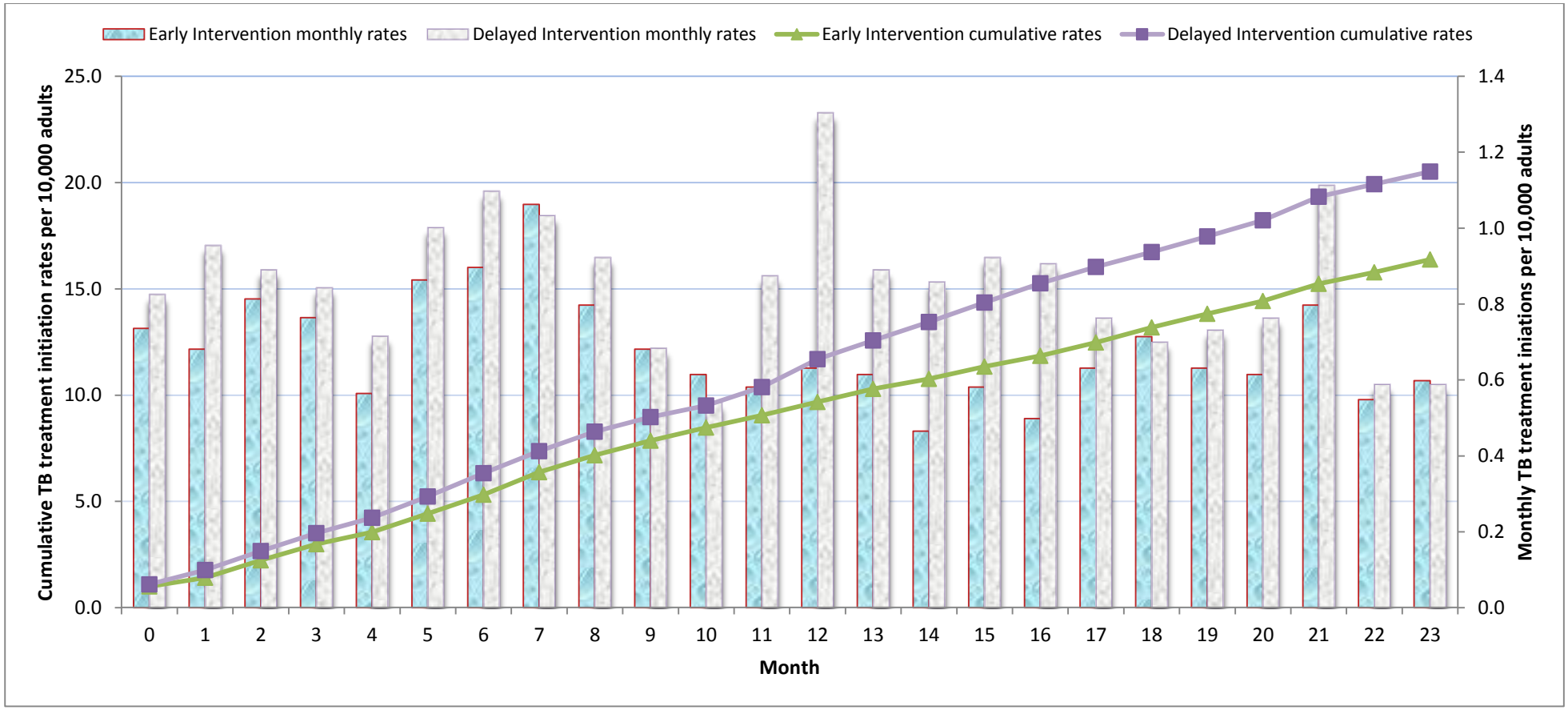


Figure 12: Monthly and cumulative TB treatment initiations rates per 10,000 adults.

Results in individuals aged 12 years and above over the 23 months of the intervention are shown. The line graphs represent the cumulative TB treatment initiation rates. The green triangle line graph indicates the Early Intervention arm and the purple cubed line graph indicates the Delayed Intervention arm. The solid bar graphs represent the monthly TB treatment initiation rates per 1,000 adults for each month of the intervention. The blue bars represent the Early Intervention Arm while the grey bars represent the Delayed Intervention Arm. The Delayed intervention arm started receiving the intervention in month 13.

After adjusting for baseline smear-positive TB treatment initiation rates, the IRR increased to 1.016 (95% CI: 0.896 - 1.152), which was a marked increase compared to the IRR of 0.778 (95% CI: 0.591 - 1.023) obtained when the first 12 months were used, although neither estimate was statistically significant.

After fitting the most improved model with gender, time, distribution of large health facilities, baseline smear- positive TB treatment initiation rates and interaction terms between time and intervention status as covariates, there was an increased attenuation of the effect of the intervention. Using the marginal model, the effect of the intervention was estimated as IRR = 0.384 (0.362 * 1.061), which is almost half the IRR value of 0.723 obtained when the first 12 months' data was used. Thus, using data from the second 11 month period, the Early intervention areas experienced a reduction in TB smear-positive cases of 61.6%, although this was not significant at the conventional 5% level due to a small number of smear- positive TB cases detected later in the intervention (Table 28). Only baseline smear positive TB treatment initiation rates and the distribution of health facilities were significantly associated with the rate at which smear-positive TB patients started treatment (Table 28).

When the random effects models were fitted, the effect of the intervention was estimated as IRR = 0.384 (0.362 * 1.061) for both the random intercepts and random coefficients models and these were similar to that obtained using the marginal model, suggesting a 61.6% reduction in TB treatment initiations per cluster in the Early arm when the random effects models were fitted. Thus, when the intervention was scaled up in the Delayed arm, more smear-positive TB cases were detected than in the Early arm (i.e. with more health facilities in the Delayed arm as shown in Figure 8 in section 6.4.1 coupled with the scaling up of the intervention in this arm led to increased treatment initiations of smear-positive TB patients). Only baseline smear positive TB treatment initiation rates was significantly associated with TB treatment initiations among smear-positive TB patients in both the random intercepts and random coefficients models. There was no clustering in the smear positive TB treatment initiation rates (ICC=0.0) and the random intercepts parameter $\sqrt{\sigma_{11}^2}$ was approximately 0.0 and the overdispersion scale parameter ϕ used to estimate the ICC was 5.024. Fitting of the random coefficients model did not improve the model fit when compared to the random intercepts model (Table 28).

In both the random effects and marginal effects models, baseline smear positive TB treatment initiations, distribution of health facilities initiating TB treatment and the interaction effect

were associated with increased TB treatment initiations (Table 28). Being female was associated with decreased TB treatment initiations among the smear-positive TB cases as noted when the first 12 months data was used.

Table 28: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring overall access rates in smear positive TB treatment initiation rates using the second phase (11 months) of the intervention

	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	0.362	0.102, 1.288	0.117	0.362	0.089, 1.483	0.158	0.362	0.089, 1.483	0.158
Base[5]	1.191	1.162, 1.220	<0.001	1.191	1.053, 1.346	0.005	1.191	1.053, 1.346	0.005
Occasion[6]	0.987	0.933, 1.045	0.663	0.987	0.940, 1.037	0.616	0.987	0.940, 1.037	0.616
Gender	0.762	0.576, 1.009	0.058	0.762	0.147, 3.958	0.747	0.762	0.147, 3.958	0.747
Treat*Month[7]	1.061	0.988, 1.139	0.102	1.061	0.984, 1.144	0.125	1.061	0.984, 1.144	0.125
H. facility[8]	1.070	1.049, 1.091	<0.001	1.070	0.897, 1.276	0.452	1.070	0.897, 1.276	0.452
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.000			0.000	
$\sqrt{\sigma_{22}^2}$								0.006	
ICC[10]								0.000	
Model fit indices									
Log likelihood		-140.457			-140.457			-140.457	
AIC		290.915			296.915			298.915	
BIC		301.863			314.432			318.621	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline smear positive TB treatment initiation rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of health facilities offering TB treatment initiations; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

6.6.1.3 Impact on ART treatment initiation rates in second phase (11 months) of the intervention period

When ART treatment initiation rates, based on the second phase (the next 11 months) of the intervention period were assessed, the marginal effects model with a single variable contrasting intervention status showed that ART treatment initiation rates significantly reduced by 20.1% (IRR=0.799; 95% CI:0.664 to 0.960, p=0.016) in the Early arm compared to the Delayed arm. This is in contrast to the IRR of 1.128 obtained when the first 12 months' data was used. No improvement in effect size was noted after adjusting for baseline ART treatment initiation rates (IRR=0.798; 95% CI:0.665 - 0.958, p=0.015).

However, using the random effects models with only a variable contrasting intervention status, ART treatment initiation rates increased 20.6% (IRR=0.794; 95% CI: 0.670 to 0.940, p=0.007) in the Early arm compared to the Delayed arm using the random intercepts model and 20.6% (IRR=0.794; 95% CI: 0.670 to 0.940, p=0.008) using the random coefficients model. These results are lower than the 11.2% (IRR=1.112; 95% CI: 0.890 to 1.389) for both random intercepts and random coefficients models (see section 6.5.1.3), suggesting an increased ART treatment initiations in the Delayed arm after rolling out the intervention.

When an improved model with intervention, baseline ART initiation rates, gender, and distribution of large health facilities but without an interaction term between intervention status and time was fitted, the IRR was 0.873 (0.794 - 0.959, p=0.005). However, when an interaction between intervention and time was allowed for, the overall effect size estimate increased to IRR= 0.924 (0.927*0.997), which was not statistically significant. This estimate was considerably lower than the IRR=1.347 obtained when first 12 months' data was used.

Similar reductions in ART initiation rates were noted when the random intercepts and coefficients models were fitted (Table 29). The estimated intracluster correlation coefficient was 0.0 and the corresponding random intercepts parameter $\sqrt{\sigma_{11}^2}$ and overdispersion scale parameter ϕ used to estimate the ICC were respectively 0.0 and 6.843, suggesting no between cluster variability in ART initiation rates in the next 11 months of the intervention. Fitting of the random coefficients model was not superior to the random intercepts model based on the model fit indices (Table 29), suggesting that the results from the random intercepts model were adequate.

Only ART baseline initiation rates, time and distribution of health facilities were significantly associated with ART initiation rates when both marginal and random effects models were fitted.

Using the 11 months' data from the second phase of the intervention, ART treatment initiation rates in the Early intervention areas reduced by 8% compared to the areas that received the Delayed intervention which was not significant at 5% level after using both marginal and random effects models. This suggests that the addition of the intervention in the Delayed arm narrowed the effect size of the intervention observed in the first 12 months. Hence, the fact that the difference between Early and Delayed intervention arms diminished after rolling out the intervention clearly demonstrates that the intervention was effective in increasing ART treatment initiation rates in the community. In fact, the intervention was effective in increasing ART treatment initiation rates even within a shorter time period.

When the monthly and cumulative ART treatment initiation rates over the whole study period were computed and plotted, the monthly and cumulative rates initially showed an increase in number of ART cases starting treatment in the Early intervention arm in the first 12 months compared to the Delayed intervention arm (Figure 13). As the intervention was rolled out to the Delayed arm, there was a narrowing gap between the two intervention arms.

Table 29: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and random coefficients models for measuring overall ART initiation rates using the second phase (11 months) of the intervention

	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	0.927	0.389, 2.205	0.863	0.927	0.588 , 1.459	0.742	0.927	0.588, 1.459	0.742
Base[5]	0.994	0.990, 0.999	0.011	0.994	0.989, 0.999	0.025	0.994	0.989, 0.999	0.025
Occasion[6]	1.08	1.062, 1.100	<0.001	1.08	1.064, 1.098	<0.001	1.081	1.064, 1.098	<0.001
Gender	1.476	0.128,16.960	0.755	1.476	0.063, 34.817	0.809	1.476	0.063, 34.763	0.809
Treat*Month[7]	0.997	0.949, 1.047	0.898	0.997	0.974, 1.020	0.789	0.997	0.974, 1.020	0.789
H. facility[8]	1.146	1.093, 1.201	<0.001	1.146	1.050 , 1.250	0.002	1.146	1.050, 1.250	0.002
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.000			0.000	
$\sqrt{\sigma_{22}^2}$								0.0191	
ICC[10]								0.000	
Model fit indices									
Log likelihood		-285.478			-286.111			-286.485	
AIC		580.956			588.223			590.969	
BIC		591.904			605.740			610.676	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline ART initiation rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of health facilities offering ART treatment initiations; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

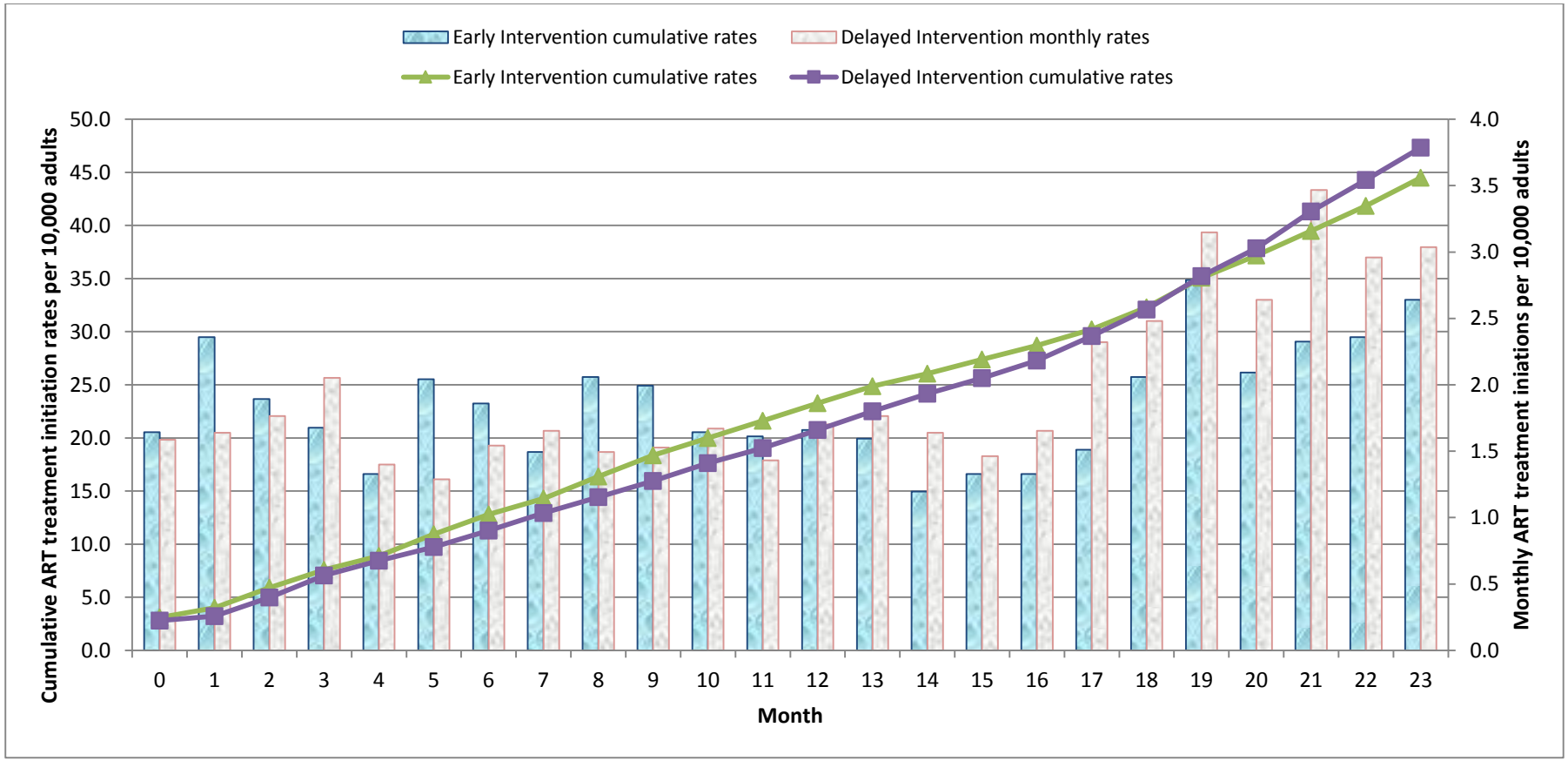


Figure 13: Monthly and cumulative ART treatment initiations rates per 10,000 adults.

Results in individuals aged 12 years and above over the 23 months of the intervention are shown. The line graphs represent the cumulative ART initiation rates. The green triangle line graph indicates the Early Intervention arm and the purple cubed line graph indicates the Delayed Intervention arm. The solid bar graphs represent the monthly ART initiation rates per 1,000 adults for every month of the intervention. The blue bars represent the Early Intervention Arm while the grey bars represent the Delayed Intervention Arm. The Delayed intervention arm started receiving the intervention in month 13.

6.6.2 Impact on TB testing uptake rates in the second phase (11 months) of the intervention

To investigate the effect of scaling up the intervention to the Delayed intervention arm on the number of people accessing TB testing services, marginal and random effects models were fitted using data from the second phase of the trial (11 months). The results from the marginal effects models showed that there were reductions in the number of presumptive TB cases accessing testing services when compared to the first 12 months' data (IRR=0.659=0.648*1.017 i.e. an 34% reduction in the 11 months' data vs IRR=1.215*0.948=1.152 i.e. a 15.2% increase in the preceding12 months' data; $p<0.05$). Using the random effects models, the number of presumptive TB cases accessing TB testing reduced by 34% (IRR=0.659=0.648*1.017) for the random intercepts model and 34% (IRR=0.660=0.649*1.017) for the random coefficients model. When the random intercepts model was fitted, the estimated intracluster correlation coefficient was 0.004 and the corresponding random intercepts parameter $\sqrt{\sigma_{11}^2}$ and the overdispersion scale parameter ϕ used to estimate the ICC were respectively 0.021 and 4.406 (Table 30). Fitting the random coefficients model did not improve the model fit when compared to the random intercepts model (Table 30).

The 34% reduction in uptake rates in the marginal, random intercepts and random coefficients models were all significant after adjusting for baseline access to TB testing services, distribution of testing sites, gender, time and an interaction term between intervention status and time as cluster-level covariates (Table 30).

In the second phase (11 months) of the trial, all the covariates except the interaction effect were significant at the 5% level in the marginal, random intercepts and random coefficients models. The distribution of health facilities, gender and the interaction effect were the only covariates associated with increasing the number of presumptive cases accessing TB testing sites in all the models fitted. In general, approximately 7 times more females than males were accessing TB testing in the next 11 months of the intervention compared to treatment initiation rates reported in both the first 12 months and the next 11 months of the intervention potentially due to embedding of the intervention. The reduction in the effect size observed when the TB testing uptake rates were compared between Early and Delayed intervention arms suggests that the intervention also improved the TB testing uptake rate in the Delayed intervention arm. Similar observations are noted in monthly and cumulative rates shown in Figure 14.

Table 30: Adjusted incidence rate ratios for intervention effects using the marginal effects, random intercepts, and coefficients models for measuring presumptive TB testing uptake rates in the second phase (11 months) of the intervention.

	Marginal effects Poisson			Random intercepts Poisson			Random coefficients Poisson		
Fixed part	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]	IRR[1]	95% CI[2]	P[3]
Treat[4]	0.648	0.445,0.943	0.023	0.648	0.438,0.961	0.031	0.649	0.436,0.966	0.033
Base[5]	0.993	0.988,0.998	0.008	0.992	0.987, 0.998	0.003	0.993	0.986,0.999	0.030
Occasion[6]	0.963	0.959,0.967	0.002	0.963	0.949, 0.978	<0.001	0.963	0.949,0.978	<0.001
Gender	6.997	2.421,20.231	<0.001	7.092	0.946,53.17	0.057	8.213	1.104,61.11	0.040
Treat*Month[7]	1.017	0.992, 1.043	0.180	1.017	0.996, 1.039	0.118	1.017	0.996,1.039	0.118
H.facility[8]	1.119	0.757,0.876	<0.001	0.814	0.755, 0.877	<0.001	0.816	0.745,0.892	<0.001
Random part[9]									
$\sqrt{\sigma_{11}^2}$					0.021			0.020	
$\sqrt{\sigma_{22}^2}$								0.0197	
ICC[10]								0.004	
Model fit indices									
Log likelihood		-325.555			-327.068			-327.592	
AIC		661.110			670.136			673.185	
BIC		672.058			687.653			692.892	

1 Incidence Rate Ratio (IRR) = rate in Early intervention/rate in delayed intervention; 2 Confidence interval (CI); 3 p-value; 4 Intervention status (Early/Delayed); 5 Baseline TB testing uptake rates; 6 Repeated measurement time; 7 interaction between intervention status and measurement time; 8 Distribution of TB testing sites; 9 σ_{11}^2 is the between cluster variance and σ_{22}^2 is the slope variance; 10 ICC=Intracluster correlation.

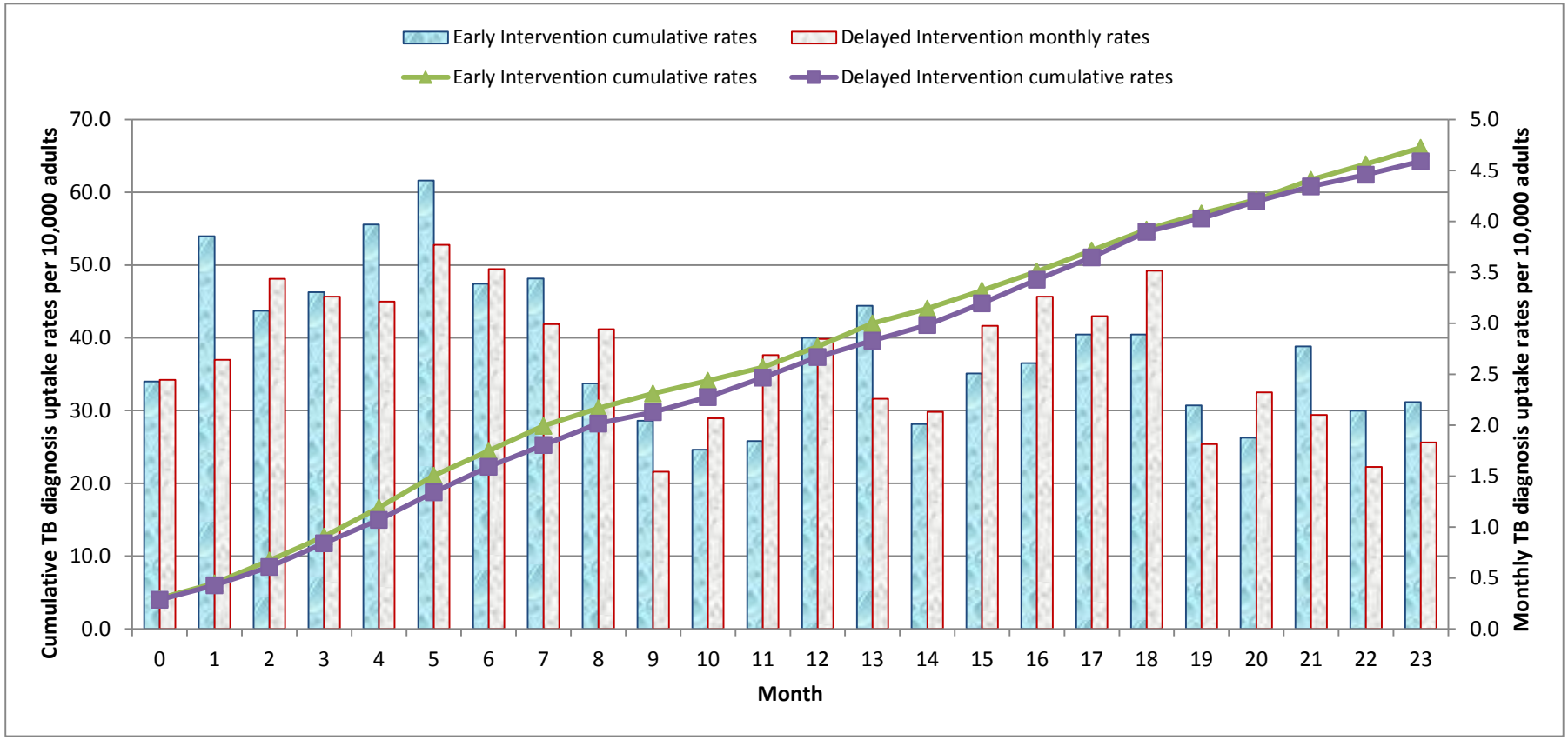


Figure 14: Monthly and cumulative TB testing uptake rates per 10,000 adults.

Results in individuals aged 12 years and above over the 23 months of the intervention are shown. The line graphs represent the cumulative TB testing uptake rates. The green triangle line graph indicates the Early Intervention arm and the purple cubed line graph indicates the Delayed Intervention arm. The solid bar graphs represent the monthly TB testing uptake rates per 1,000 adults for each month of the intervention. The blue bars represent the Early Intervention Arm while the grey bars represent the Delayed Intervention Arm. The Delayed Intervention arm started receiving the intervention in month 13.

6.7 Summary and conclusion

In this chapter, results from the analysis of the Triage Plus study that employed informal health care providers to implement TB and HIV community interventions were presented. The results from this chapter can be summarised as:

- The adjusted analysis showed that involvement of the informal providers significantly increased the number of presumptive TB cases accessing testing sites by 15.2% (p=0.003) and the number of TB patients starting treatment by 18%, 19.6% and 21.5% for the marginal, random intercepts and random coefficients models in the first 12 months of the intervention but not significant at 5% level. In the next 11 months of the intervention, the number of presumptive cases accessing testing services were similar between the two intervention arms. In the second phase (11 months) of the intervention, the number of TB treatment initiations were significantly reduced by 45% (p<0.001) in the Early intervention compared to the Delayed arm, suggesting that rolling out of the intervention to the Delayed arm resulted in increased TB treatment initiations in the Delayed arm.

- For smear- positive TB, the number of patients starting TB treatment in the Early intervention areas in the first 12 months were 28% (p=0.02) less than the Delayed arm. In the second phase (11 months) of the intervention, the number of smear-positive TB patients starting treatment in the Early arm fell by 61%, though this was not statistically significant at the 5% level.

- In regards to the adjusted analysis for the HIV services access rates, the involvement of informal healthcare providers significantly increased the number of patients starting ART in the Early arm by 34.7% (p=0.048) during the first 12 months of the intervention. In the next 11 months of the intervention, the number of patients starting ART were similar between the two intervention arms (IRR = 0.924 with 95% CI: 0.389, 2.205, p=0.863). The HIV testing uptake rates in the Early arm increased by 61% (RR =1.61; 95% CI:1.44 to1.83, p<0.001) compared to the Delayed arm in the first 12 months of intervention.

Thus, the results of the study clearly demonstrate that engaging informal health care providers was effective in improving TB and HIV testing uptake and also increased ART initiation. This reinforces the need for community participation in integrated TB and HIV interventions to combat the two diseases. However, for these providers to be effective in promoting TB treatment initiation, numbers of sites offering TB testing and treatment initiation in rural areas should be increased.

Furthermore, the results of the study demonstrate that the effectiveness of engaging informal health care providers in improving access to TB and HIV services was dependent on the gender distribution. In general, access to TB and HIV services was lower in females than in males due to gender related structural barriers such as poverty (Storla et al., 2008). It is worth noting that TB testing uptake increased by approximately 7 times more in females compared to males in the next phase of the intervention potentially due to more women being exposed to the intervention over time although this did not lead to improved TB treatment initiations. However, the assessment of the effectiveness of the intervention over time (based on the analysis of time treatment interactions) showed no significant effect on time treatment interactions overall. This absence of statistical significance is mainly due to the low statistical power of the significance test for the interaction terms (a common problem in statistical analysis).

6.8 Statistical implications of the findings of the Triage Plus study

Due to the fact that the 3 clusters per arm study design was used in the Triage Plus study, the results, especially in regards to the TB outcome measures, needed to be interpreted cautiously. For example, the adjusted effect size in increasing TB treatment initiations was 18% but was not statistically significant at the 5% significance level. This supports the findings from the simulation studies that clearly showed that with a true effect size of < 20% in low incidence disease conditions, such as TB in our case, no adequate power would be achieved even after increasing the number of repeated measurement times to 12 points. However, with effect sizes ≥ 20%, such as that found in smear-positive TB patients starting treatment, adequate power was available to detect significant differences in TB initiation rates between the two arms. The findings from the analysis of the Triage Plus study therefore confirm the conclusions made in the simulations studies presented in Chapter 5 that no adequate power would be achieved in low incidence disease conditions with an effect size less than 20% even after 12 measurement time points.

For the high incidence outcomes of ART initiation rates and testing uptake rates for TB and HIV, achieving adequate statistical power in assessing the differences between the two arms was not difficult. For example, the simulation studies presented in Chapter 5 showed that in high incidence outcomes adequate power could be achieved even with the number of clusters as low as 3 clusters per arm but with at least 3 measurement time points when the ICC was 0.00154. With an ICC of 0.081, adequate statistical power was achieved with effect sizes >40% in a 3 cluster per arm design.

Overall, the ICCs obtained in the final analysis of the Triage Plus study were comparable to those used in the simulation studies presented in Chapter 5. This, therefore, suggests that the interpretation of the main outcome measures and the statistical power for the final analysis of the Triage Plus study data using the repeated monthly counts over the intervention period were not affected by changes in the ICCs for the main outcome measures.

CHAPTER 7

GENERAL DISCUSSION, POLICY IMPLICATIONS AND CONCLUSIONS

7.1 Introduction

This final Chapter provides a discussion of the research findings, strengths and possible limitations of both the simulation and Triage Plus intervention studies. To begin with, the true statistical power of the Triage Plus study is explored based on the findings of the simulation study. The limitations of both are also reviewed because of the power issues within each study. The methodological conduct of the Triage Plus study is then considered, including a review of the implications of the study's key findings on the policy implications for community engagement intervention in the promotion of TB and HIV testing and treatment.

7.2 Statistical power and parameter estimation in a cluster randomised trial with a limited number of clusters but adopting a repeat observation design

In simulation studies of a cluster randomised intervention design consistent with that of Triage Plus, it was clear that any correlations arising from the study design had to be accounted for in the analysis. Mixed effects models, such as the Poisson mixed effects model, help to capture such correlations, but they also limit the power and the precision of parameter estimates as the analysis is done at the cluster level, a major problem when there are only limited numbers of clusters available for analysis. The extensive Monte Carlo simulations implemented in Chapter 5 using baseline data sets from the Triage Plus study (i.e. using baseline mean TB treatment initiations, actual estimates of intracluster correlations from the Triage Plus study and the 3 clusters per arm) indicated that adequate power would have been difficult to achieve if a basic pre- and post- evaluation design was used.

Thus, by using a repeated measurement design with a small number of clusters, a similar level of statistical power was achieved when compared to the cross-sectional design with more clusters enrolled (Hedeker and Gibbons, 2006). This is because the repeated measurements in the same cluster are not perfectly correlated (which is the case in community interventions which usually have low ICC (see Gulliford, Ukoumunne, and Chinn, 1999). As a result, the repeated measurements in the same clusters provide more independent data than in cross-sectional designs such as pre- and post- evaluations (Hedeker and Gibbons, 2006).

Thus, the findings of the simulation studies reported in Chapter 5, relating to the power and accuracy of parameter estimates and to the variance components in cluster randomised studies with small numbers of clusters, are considered to be critical in several respects.

The population incidence of the studied disease was found to be critical in determining statistical efficiency— a finding noted also by Moineddin, Matheson and Glazier (2007). The simulation study showed that in low incidence disease conditions such as TB treatment initiation rates, achieving adequate statistical power and precise estimates required more clusters and repeated measurement times. With effect sizes of less than 10%, no adequate power could be achieved regardless of the number of clusters and measurement times (see Table 6, 8 & Figure 6).

In contrast, in high incidence disease conditions such as HIV testing / treatment initiation and TB testing, the simulations showed that at an effect size of 10%, adequate power could be achieved even in a 3 cluster per arm design but only by having more repeat observations (in our case 12 time points for the 3 clusters per arm), when the ICC was low (ICC=0.00154). With an ICC of 0.081, adequate statistical power could only be achieved with an effect size of at least 40% in a 3 cluster per arm design. The required number of repeat observations to achieve adequate power decreased with the increasing number of clusters per arm (see Table 7 & 9). However, when the effect size was at least 20%, the simulation studies showed that in the low incidence disease conditions adequate statistical power could be achieved after more repeat observations. At effect size of 20%, about 12 repeat observations were needed to achieve adequate power of at least 80% in a 3 cluster per arm design when the ICC was 0.00154. However, with an ICC of 0.081, at least 9 clusters with 6 measurement times were needed. In general, the number of repeat observations decreased with the increasing number of clusters and effect sizes in both low and high incidence disease conditions (see Tables 6 - 9, Figures 6 & 7). Increasing the ICC resulted in decreased statistical power as also observed in other simulation studies (Amatya et al., 2013).

The findings from the simulations in general showed that statistical power is affected by several factors such as disease incidence, number of clusters, effect size, and repeated measurements and ICC (see Tables 6 - 9, Figure 6 & 7). These simulations, therefore, allowed exploration of complex designs that often arise in practice but for which conventional power calculation formulae are inadequate.

Although previous simulations have recommended a minimum of 30 groups per arm (for example Maas and Hox, 2005; and Moineddin et al., 2007), our simulations clearly showed that the statistical power of cluster randomised trials with only a limited number of clusters available per arm (such as the Triage Plus study) can be improved considerably by increasing the number of measurement times and can be applicable in high prevalence or incidence disease conditions with moderate ICCs (Murray et al., 1998b; Heo and Leon, 2009; Moineddin et al., 2007; Amatya et al., 2013).

The simulation methods used in this dissertation are universally applicable to determine the optimal number of clusters and the design conditions needed to obtain valid estimates (Murray et al., 1998b; Arnold et al., 2011; Burton et al., 2006; Bennett et al., 2002; Moineddin et al., 2007).

However, the simulation studies presented in Chapter 5 were limited to cluster randomised designs with repeated measurements nested within clusters when assessing statistical efficiencies in terms of power and parameter estimate precision. The design adopted a situation where repeated measurements were taken from the same clusters and not necessarily from the same individuals in the group (i.e. cross-sectional design). Thus, the findings of this study are most applicable to similar study designs—but as pointed out by Murray et al. (1998b), the findings may also be applicable to cohort studies collecting repeated measurements in the same clusters as well as for individuals within the clusters.

7.3 Statistical power of the Triage Plus study

The positive effects of the Triage Plus intervention were intended to persist in the communities and increase as the intervention progressed. In this way, more patients in the community would be more aware of the signs and symptoms of TB and HIV which would necessitate them seeking care from health centres (Storla et al., 2008). With the limited number of clusters used in the Triage Plus study, coupled with the effects of the intervention persisting over a longer duration, a parallel cluster randomised approach was therefore appropriate. Thus, all 3 clusters in each arm were simultaneously allocated either to an Early intervention arm or to a Delayed intervention arm - consistent with other parallel designs (see Corbett et al., 2010).

The Triage Plus intervention used what could be termed a 'semi stepped wedged design' in that the scheduling of the implementation of the intervention within the Early arm clusters

was randomly drawn, and, based on the randomisation process, the implementation was sequential. In addition, the Triage Plus intervention also adopted a phased intervention: after implementing the intervention in the Early intervention arm for the first 12 months, the intervention package was also implemented in the Delayed arm for the following 11 months, which acted as both control and intervention clusters as in a stepped wedged design (The Gambia Hepatitis Study Group, 1987; De Allegri et al., 2008). Though analysis based on the pure stepped wedge design has the advantage of allowing internal comparisons (the clusters all experience both the control and intervention conditions to act as their own controls) which can lead to improved power, this design is only valid when there is a relatively large number of clusters and the clusters receive the intervention sequentially; hence, the pure stepped wedge design is not applicable to our study (Hussey and Hughes, 2007), and the design was not considered in any further detail.

The sequential and phased cluster randomised design used in the Triage Plus study allowed efficient project implementation and also removed the ethical concern of withholding the intervention from the control clusters (Hussey and Hughes, 2007). The study design further allowed us to assess how the effects of the intervention changed over the intervention period when the Delayed arm clusters received the intervention. When the intervention is scaled to the Delayed intervention clusters, the direction of the effect size provides valuable information regarding the effectiveness of the intervention. If the intervention is effective, an effect size and direction similar to that found in the first phase of the study would be anticipated, if this does not happen, a more thoughtful interpretation of effect may be needed.

However, such analyses may require more statistical power than is available when there are only a small number of clusters, as evidenced in the lack of statistical significance when the effectiveness of the intervention was assessed after the Delayed arm received the intervention: a reduction in smear positive TB treatment initiation rates by 61% was not statistically significant (see Table 28). This may be due to the earlier intervention arm depleting much of the smear-positive TB cases during the first phase of the intervention (in the first 12 months), but more cases are detected in the newly scaled interventions (as a result of the intervention), leading to huge differences between the two arms in terms of the estimated incidence rate ratio. Very few smear-positive TB cases detected in the second phase of the intervention in the Early intervention areas resulted in having small mean counts per month but with an increased variance around this estimate that affected the statistical power (see also in Moineddin et al., 2007).

Thus, the analytical design used in Triage Plus was efficient in attaining adequate statistical power for HIV outcomes (ART treatment initiation rates and HIV testing uptake rates) because they were very common. In addition, outcome measures related to testing uptake rates among presumptive TB cases were also adequately powered. However, TB treatment initiation outcome measures were inadequately powered for small effective sizes of less than 20% even after using the repeated measurements design (see Table 21). Findings from the simulation study in Table 6 & 8 also indicated that no adequate power could be achieved at an effect size less than 20%. Ideally, therefore, more clusters were needed in the Triage Plus intervention to achieve adequate statistical power across all important outcome measures as well as to derive valid and precise estimates of effect size for the intervention on the TB treatment initiation rates.

7.4 How the Triage Plus study was conducted and its limitations

7.4.1 Research design

The use of multiple measurements in a cluster randomised intervention trial substantially improves statistical power (Murray et al., 1998b; Hedeker and Gibbons, 2006), as also demonstrated in Chapter 5. Using multiple measurements also allows for concurrent comparison of outcome measures over time. Together, these provide a clearer measure of intervention effectiveness than could be obtained through before-after evaluation methods that may fail to account for indirect effects such as secular trends. In general, there is a downward trend in TB in Malawi (Glynn**,** 2004) and an upward trends in ART uptake rates (Cook et al., 2010). Our approach was, therefore, robust in accounting for these secular trends.

Applying a repeated measurement design (i.e. collecting data from the same clusters over several time points) allowed us to compute better estimates of the effectiveness of the intervention. With repeat observations, the general trend in the effectiveness of the intervention over time was clearly demonstrated by using cumulative rates in treatment initiations rates for TB and ART as well as uptake rates in TB and HIV testing (see Figures 11 to 14 in Chapter 6).

However, using the repeated measurement design necessitated sophisticated statistical methods requiring numerical integration approaches to evaluate the likelihood of the generalised linear mixed regression models for the mixed effects Poisson regression (Breslow and Clayton, 1993).

In the cluster randomised design, it is acknowledged that the randomisation process could have resulted in two alternative scenarios. In an extreme scenario, the randomisation process may have achieved the complete balancing of known and unknown confounding variables, in which case the assessment of the effect of the intervention can be implemented without adjusting for covariates and without necessarily affecting the trial conclusions (Bennett et al., 2002; Corbett et al., 2010). In a more moderate scenario, our randomisation process may have resulted in the moderate imbalance of some confounding factors due to the limited number of clusters available. Our matching of clusters at the randomisation stage was intended to ensure an even distribution of confounding factors between the two arms. However, the statistical comparisons made showed clear differences between the two study arms in terms of baseline TB and HIV services access rates (see Tables 17 & 18). In addition, there was uneven distribution of health facilities initiating TB and ART treatment between the two arms (see Figure 8). Thus, model adjustment for confounding variables when assessing the effectiveness of an intervention was imperative.

Bennett et al. (2002), in their simulation studies, observed that the adjustment for a confounder with moderate imbalance but having a large effect on the intervention effect was robust even with a very small number of clusters. Thus, in the analysis of the Triage Plus study, statistical modelling, adjusting for covariates, was undertaken. This permitted controlling for effects resulting from an uneven distribution of factors, such as the number of large health facilities between the two intervention arms. Furthermore, the effects of clustering and the repeated measurements from the same clusters were adjusted for by our use of random effects modelling approaches (Breslow and Clayton, 1993).

By using mixed effects Poisson regression models and repeated measurements from the same clusters in the analysis of the Triage Plus study, the challenges in type I error rates usually encountered in cluster randomised trials were averted, referred to as 'Cornfield penalties' (Cornfield, 1978). By incorporating random effects in the models, the design addressed the extra-variation problem due to clustering and overdispersion. Given the small number of clusters available in the Triage Plus study, the use of repeated measurement design provided the additional degrees of freedom necessary for a valid estimation of intervention effectiveness (Murray et al., 1998b; Hedeker and Gibbons, 2006).

7.4.2 Study limitations

Despite the strength provided by the multiple measurements within a limited cluster randomisation and the use of random effects modelling, the study had its inherent limitations.

First, the evaluation process depended on routine data recorded in patient registers in health facilities; therefore, potentially inaccurately recorded patient address details were used to locate patients. Use of both active and routine data collection methods for impact evaluation would be necessary. The use of routine data may have contributed to biased estimates of the intervention effect for several reasons:

i) Patients who received treatment in urban areas while temporarily residing with relatives due to their illness may have been less likely to mention their actual place of residence and, therefore, may have been missed in our database, which only collected patients with rural addresses, leading to an underestimation of the number of TB or HIV cases in the rural areas. No specific direction of the bias could be determined as this was likely to occur on all intervention arms.

ii) Failure to accurately identify the actual villages that are in rural or urban areas (especially those bordering the urban areas) may also have led to the misallocation of patients to clusters for analysis. This occurred randomly between the two arms as they all shared boundaries with the urban areas.

iii) As the intervention improved the community's knowledge of TB and HIV and the relationship between the two, a certain proportion of patients may have either avoided seeking health care services for fear of being labelled HIV positive (Moller et al., 2011) or used other names and wrong addresses in the course of seeking health care for fear of being identified and traced. This may have lowered the effect size of the intervention.

iv) As routine health services data were used for this evaluation, some addresses were so poorly recorded that we were unable to properly assign them to either intervention arm in about 1.8% of cases (0.8% for TB and 2.2% for ART). However, the failure to assign the patients occurred randomly, and there was no consistent pattern to the missing residential addresses, so this bias was probably small.

Second, although the use of repeated measurements might have addressed the statistical power issues in terms of having adequate degrees of freedom, the limited number of clusters used in the randomisation might not have evenly spread the unmeasured confounding factors (Donner and Klar, 2000; Murray 1998; Varnell et al., 2001) as observed in the uneven distribution of health facilities initiating both ART and TB treatment between the two arms (see Figure 8). Although other, smaller health facilities offering TB and HIV testing and/ or ART services were, however, evenly distributed between the two arms, there may have been

other unmeasured confounding factors. Because of the limitations of pair-matched design especially when there are few clusters for randomisation and the difficulty in predicting in advance the effectiveness of the matching process, alternative approaches needed to be considered in future such as stratified designs in which there are more than two clusters in each stratum (Klar and Donner, 1997; Hayes and Moulton, 2009).

7.5 Effectiveness of the Triage Plus study

7.5.1 Intervention effect on ART treatment initiation rates and HIV testing uptake rates

The findings of the Triage Plus intervention indicate that the involvement of unpaid informal health care providers in HIV community awareness, disease recognition and referral has contributed significantly in ART treatment initiation rates and HIV testing uptake rates. By the end of the first phase (the first 12 months), HIV testing uptake and ART treatment initiation rates had increased by more than 61% and 34% respectively in the Early intervention arm compared to the Delayed arm (see Tables 23 & 26). In addition, there was no difference in ART treatment initiation rates between the two arms in the second phase (the next 11 months) of the intervention (see in Table 29).

Furthermore, the differences in cumulative ART treatment initiation rates and HIV testing uptake rates increased over time, with higher rates in the Early arm than in the Delayed arm in the first 12 months (for treatment and testing) followed by a reduction in these differences in the second phase for ART treatment (the next 11 months) (see Figures 11 & 13). This major improvement in ART treatment initiation rates and HIV testing uptake following intervention suggests that the intensification of the engagement of informal healthcare providers in the delivery of HIV services in such under-served rural areas could likely lead to improved service access and, therefore, reduce HIV related deaths and transmission (Cohen et al., 2011).

Our results are consistent with other studies done in the region. A community based intervention in a rural district in KwaZulu-Natal, South Africa, involving paid community care workers to provide an integrated TB/HIV/PMTCT, showed improved uptake rates for HIV testing (92% acceptance rates) and screening of sexually transmitted infections (32%) compared to sexually transmitted infections screening rates of 7% in control areas (Uwimana et al., 2012).

7.5.2 Intervention effect on TB treatment initiation rates and testing uptake rates

The results indicate that the involvement of unpaid informal health care providers in HIV community awareness, disease recognition, support in sputum collection and referral

significantly increased TB testing uptake rates. In addition, there was a non-significant increase in TB treatment initiation rates. Results from the first 12 months showed that the engagement of informal healthcare providers increased the number of presumptive TB cases accessing testing sites in the Early arm by 15% (p=0.003), but there was no difference between the intervention arms in the next 11 months when the intervention was rolled out to the Delayed arm (see Tables 24 & 30). Studies conducted within the region involving informal healthcare providers have shown similar results (Datiko and Lindtjorn, 2009; Simwaka et al., 2012). In Ethiopia, the engagement of informal healthcare providers (called extension health workers) in the identification of presumptive TB cases, sputum collection and transportation to smear microscopy sites resulted in an increase in TB case detection by 122.2% in intervention areas compared to 69.4% in control areas (Datiko and Lindtjorn, 2009). In Malawi, the use of this type of provider (in this case storekeepers) in referring presumptive TB cases significantly increased the number of smear-positive TB case notifications in the intervention areas after one year of the intervention (0.6 per 1000 people in the control areas vs. 1.2 per 1000 in the intervention areas) (Simwaka et al., 2012).

On the other hand, TB treatment uptake rates in the Early arm increased by 18% after the first 12 months, which was not significant (possibly because of inadequate statistical power) (see Table 6). At this effect size in low incidence disease conditions like TB, the simulation findings showed that achieving adequate power is a problem (see Table 6).

In contrast, despite the significant increase in uptake rates of TB testing observed in the first 12 months of the intervention, there was a significant reduction in the number of smear-positive TB cases starting treatment in the Early arm compared to the Delayed arm (see Table 22). These results are also consistent with the fact that many of the cases in Malawi are smear-negative because of the high HIV burden that increases the number of smear- negative TB patients (NAC, 2007 & 2010)

In implementing Triage Plus, a number of activities were implemented by informal healthcare providers including patient referrals and sputum collection and delivery to the nearest testing sites, an intervention known to improve TB case detection rates (Xiong et al., 2007; Simwaka et al., 2012). Presumptive TB cases consequently were started on treatment, an intervention known to reduce TB incidence.

However, it is not clear whether the observed significant reduction in smear- positive TB patients starting treatment in areas that received the intervention early could be attributed to

the reduction in infectious TB incidence in the study areas following the intervention. Thus, there are two interpretations of the effect of engaging informal health care providers on smear-positive TB cases commencing treatment:

First, engaging informal health care providers was not effective because there was a significant reduction in smear- positive TB treatment initiation rates following intervention. The number of smear-positive TB cases starting treatment in the Early arm in the first 12 months was 28% less than those starting treatment in the Delayed arm (see Table 22).

If this is the direction one would take, then one would argue that with chronic cough being a cardinal sign for TB, which was emphasised during community awareness interventions conducted by informal health care providers, meant that the intervention was to greatly affect individuals with pulmonary TB, convincing them to start TB treatment. Thus, the reductions in smear-positive TB cases as well as all types of TB cases starting treatment would suggest that the intervention was counterproductive: more individuals with TB did not access the TB services because of the TB and HIV integration (see Moller and Erstad, 2007). The TB and HIV integration might have led to increased stigmatisation because in some instances TB is usually equated with HIV (Skordis-Worrall, Hanson and Mills; 2010). At community level, TB and HIV sensitisation messages meant to address the two diseases were given as a package. The messages highlighted the association between TB and HIV with a view to promote testing and eventual treatment initiations for the two diseases. This association has shown to be a major source for stigmatisation and affected timely presentation for TB treatment in high HIV prevalence areas (Moller and Erstad, 2007). In a community survey exploring attitudes to TB in South Africa showed that the majority of the respondents (89%) cited fear of learning one's HIV-positive at health clinics and the belief that almost all TB patients are HIV positive and that only people who are HIV positive get TB were found to be the most important barrier for timely TB case detection (Moller et al., 2011).

Given that there were improved uptake rates of TB and HIV testing in the Early arm in the first 12 months suggests that the notion that TB and HIV integration might have led to other individuals seeking care from other sources because of their fear of an HIV positive result (Moller et al., 2011) or the stigma associated with HIV (Moller and Erstad, 2007) is not feasible in this case. If individuals were going elsewhere for testing, there should have been a corresponding significant reduction in the uptake rates of TB and HIV testing because these cases would be seeking care from alternative sources.

A second possible explanation was a falling TB incidence over the time period of the study. Community interventions using participatory approaches have led to major reductions in latent TB incidence rates within a few years (O'Donnell et al., 2012). O'Donnell and colleagues observed a sustained reduction in the incidence of latent TB infection among all racial groups in a community based participatory intervention. Incidence of latent TB infection reduced by 50 % from the baseline levels over a 3 year period. Using active case finding strategies in high-density residential suburbs in Harare, Corbett and colleagues (2010) observed a 40% reduction in smear- positive TB cases from the baseline at the end of 3 years. However, TB incidence in our study could only have been reduced as a consequence of the intervention in the Early arm if TB treatment initiation had been significantly increased. If TB incidence was already falling in both the Early and Delayed arms as a result of background effectiveness of the TB programme, then the intervention should still have resulted in a detectable increase in TB treatment initiation.

Overall, therefore, the most likely explanation relates to the extent to which TB testing and HIV testing are available in the rural areas. HIV testing was available at 20 sites in the Early intervention arm and 18 sites in the Delayed arm. This compares with TB testing which was only available at 7 sites in the Early arm and 9 sites in the Delayed arm (see Figure 1). Thus, a more likely explanation for the lack of increase in TB treatment initiation is that people in this poor, rural area will have used substantial resources (including for travel) to access the infrequently available TB testing sites (Kemp et al., 2007). Kemp and colleagues found that TB patients in peri-urban Lilongwe use between 2 and 5 times their monthly income in reaching a TB diagnosis and starting treatment. By contrast, patients accessing the more frequently and locally available HIV testing facilities will have used a smaller percentage of their available resources. In support of this, work in Malawi suggests that patients accessing ART spend less than their monthly income in care-seeking (Namakhoma et al submitted). Furthermore, there is evidence that linkage between HIV testing and ART initiation is hindered when treatment initiation is less accessible (Mac Pherson et al, 2012). Overall, therefore, we hypothesise that people who have accessed TB testing will frequently have depleted their resources to the extent that going on to access TB treatment initiation will have been beyond their means. By contrast, people who have accessed HIV testing will not have depleted their reserves to the same extent and will have found it easier to travel to, and access the HIV treatment initiation sites. Similar numbers of HIV and TB treatment initiation sites were available in both arms of the study: 3 and 2 respectively in the Early arm and 4 and 6 respectively in the Delayed arm.

Our findings reinforce recent policy recommendations on the integration of TB and HIV services in order to reduce the burden of TB and HIV in co-infected patients (WHO, 2004). Integration of TB and HIV services in health facilities has greatly improved treatment initiation and success rates (Hermans et al., 2012).

7.5.3 Flexibility of the intervention

To ensure the scalability of the intervention by the national programmes, implementation of the intervention was flexible, in terms of types of informal health care providers engaged and as well as the nature of activities implemented by them. Implementing the intervention in such a way that it could be undertaken by national programmes enhances its chances of being adopted and scaled up nationally. The Triage Plus study used a variety of informal healthcare providers to implement the various components of the intervention, thereby enabling national programmes to use any of these informal health care providers in the national level programmes without restricting the program to a specific type of informal provider. In addition, by allowing the informal health care providers to implement specific intervention activities, which they were capable of doing or were currently doing as part of their existing community work, meant that concerns regarding the sustainability of the interventions would be minimised. For instance, store keepers engaged in recognising and referring presumptive TB cases during their encounters with clients when buying medication are more likely to continue identification and referral even after the intervention study because of the knowledge gained through training.

Follow up visits by the research team to check on the progress of informal health care providers participating in the intervention project was minimised to only two visits in each year to be consistent with the pragmatic nature of the intervention and to reflect what the study team thought would be sustainable in the long term. To reduce costs, supervisory visits were done only for providers where challenges were observed.

However, the fact that several players are engaged in this intervention presents difficulties in isolating the effectiveness of the intervention: it is difficult to determine what is attributed to the specific activities implemented by the informal health care providers and other stakeholders sensitised in the intervention areas. Thus, the effects of the different components of the intervention, such as community awareness meetings and support in sputum collection by informal health care providers on one hand, and engagement coupled with orientation of health surveillance assistants to support project implementation on the other, cannot be

isolated in terms of the effectiveness of the intervention, and this was not the intention of the project. It is worth noting that it would be difficult practically in the long term to implement targeted interventions with only one type of informal care provider or one type of engagement (e.g. sputum submission without any additional raising of general awareness or training on infection control). Thus, the holistic approach adopted was realistic in real field conditions.

7.6 Policy implications

The findings from the Triage Plus study have several policy implications in improving TB and HIV services access and disease control. They are as follows:

7.6.1 Policy implications on TB control

The observed improvements in TB testing uptake rates due to engaging informal healthcare providers means that the scaling up of such interventions to many districts through national programmes would increase TB case detection and timely treatment initiation, thereby reducing transmissions in the general population, as most transmissions occur outside households (Verver et al., 2004).

Furthermore, the rapid reduction of smear- positive TB cases initiating treatment due to the depletion of infectious cases in the community following the intervention, as shown elsewhere (Corbett et al., 2010), suggests that a decline in the rates of new TB infections can be achieved within a reasonable number of years if the intervention is successfully implemented in the general population nationally.

7.6.2 Policy implications on HIV transmission prevention

The fact that the intervention was able to improve ART and HIV services access in the general community within one year, suggests that by engaging informal health care providers in community interventions, national HIV programmes are able to reach even those who normally would have not accessed the services. The increase in number of people tested for HIV and in starting ART coupled with timely ART initiation means that HIV transmission rates can greatly be reduced over a shorter period of time because of the protective effect of the ART (Cohen et al., 2011).

As more people accessed HIV testing following the community sensitisation meetings conducted by the informal healthcare providers as part of the intervention, the increase, in turn, leads to improved changes in risky behaviour, thereby reducing HIV incidence in the general population (Bello et al., 2011; Hallet et al., 2006). This means scaling the intervention to other areas would result in a reduction in HIV transmission in the general population.

7.7 Conclusions

In this section, conclusions regarding the simulation studies and the actual analysis of Triage Plus are given.

Findings of the simulation studies

1. The simulations show that there are many factors that need to be considered when designing and powering a cluster randomised trial where the outcome measure is a count of events: minimum detectable effect size, ICC - between cluster variability and within cluster correlation), incidence of outcome measure (this will be related to the unit length of time used for each observation point), and number of observation times for which data are collected. These will primarily influence the optimal number of clusters needed, but prevalence or incidence will directly affect the size of each cluster (to ensure sufficient events per cluster).

2. Previous recommendations may be overly pessimistic as to the minimum number of clusters needed, which could be as small as 3-4 per intervention group, provided the incidence of the outcome event is reasonably high and ICC is moderate. Adding more assessment times may compensate for a small incidence and/or limitation in the number of available clusters.

3. Pairing clusters for randomisation will help to balance confounding factors, but will only be partially successful when the number of clusters available per group is small – for this reason, it may be optimal to use unconditional statistical methods that ignore the pairing structure with the inclusion of an adjustment to the effect size for all known confounding factors. This may be as good as or probably even better, in terms of statistical efficiency, than conditional analyses as there are more degrees of freedom available in the analysis.

4. In complex study designs, sample size calculations cannot be computed using the existing formulae, and extended formulae may prove to be too complex to be of practical value. It may be necessary to use simulation methods to overcome these issues.

5. In conclusion, cluster randomised trials with a count outcome measure ideally need a minimum of 3 clusters per group with at least 12measurement times for effect size of 20% (or higher) in low incidence diseases with low ICCs. However, for high incidence outcomes, as few as 3 measurements times with 3 clusters per arm may be adequate to achieve statistical power of at least 80%. With moderate ICCs (in our case ICC of 0.081), at least 9 clusters were needed to achieve adequate statistical power of 80% with an effect size of 20% and using 6 and 12 measurement time points respectively for high and low incidence disease conditions. For an effect size of 40%, at least 3 clusters per arm were needed to achieve adequate statistical power with 4 repeated measurement times in low incidence diseases and 3 measurement times for high incidence diseases. Thus, having more clusters is not necessarily that advantageous and adequate statistical power can still be obtained with fewer clusters per group provided the incidence of the outcome of interest is not too small, and it is possible to take multiple measurements (replicates) over time within each intervention phase

Findings from the Triage Plus study

- The Triage Plus intervention showed that the engagement of the informal health care providers at the community level in sputum collection, TB and HIV sensitisation and awareness campaigns was effective in improving HIV services access rates (both testing uptake and treatment initiation) and TB testing uptake rates. But more clusters were needed to more accurately evaluate the impact of the intervention in improving TB treatment initiations.
- Our conclusions on the effectiveness of engaging informal healthcare providers in TB and HIV services access are made more likely based on the following reasons:
    - First, because the data used for evaluation, which showed improvement in services access for HIV and TB testing, is facility based and followed standardised management guidelines for patients in testing for TB and HIV, but similar improvements were not seen in the Delayed arm in the first 12 months. In addition, baseline ART treatment initiation rates and testing uptake rates, especially for presumptive TB cases, were similar between the intervention arms.
    - Second, the fact that TB testing uptake rates increased in areas that received the Early intervention and then were followed by a reduction in smear-

positive TB treatment initiation rates, a trend which was not seen in the Delayed intervention areas.

- Assessment using both TB and HIV testing uptake rates and treatment initiation rates in our study provided a clearer measure of intervention effectiveness than could be obtained from treatment initiation rates alone. The use of TB testing uptake rates, for instance, allowed a better interpretation of the observed reductions in smear- positive TB initiating treatment, a finding that would be difficult to discern if treatment initiations rates were the sole outcome measure.
- The intracluster correlation coefficients estimated after fitting the random intercepts models adjusting for all covariates obtained in the final analysis of the Triage Plus study were comparable to the ICCs used in the simulation studies presented in Chapter 5. However, the estimated ICC for HIV testing uptake was slightly higher than the ICC of 0.081 from the baseline data from the Triage Plus study. With the high effect size of 61% and the fact that the mean count for HIV testing uptake per month was high, the slightly higher ICC levels did not lead to reduced statistical power. This, therefore, suggests that the interpretation of the main outcome measures and the statistical power for the final analysis of the Triage Plus study data using the repeated monthly counts over the intervention period were not affected by changes in the ICCs for the main outcome measures.
- We, therefore, finally conclude that an effective approach to TB and HIV control is to engage informal health care providers in order to reach those failing to access TB and HIV services due to a number of factors including poverty, poor TB and HIV knowledge, geographical barriers and perceived attitudes towards disease prevention and control.

7.8 Future work

7.8.1 Introduction

Although the Triage Plus study data were monthly counts measured over the duration of the intervention and therefore Poisson models presented in 4.2.3 were the ideal models. However, an alternative approach for assessing the effectiveness of the intervention would have been to conduct multiple household surveys to obtain data required for the evaluation of the Triage Plus. In this way, individuals would be interviewed to assess if they have had an outcome of interest in a specified time period (e.g. access to ART or TB treatment). These surveys, therefore would generate binomial data. Section 7.8.2 briefly presents the log-binomial regression methods to model of such data using the same notations used in Poisson model formulations (where i indexes repeated measurements in a cluster j) presented in Chapter 4.

In addition to the likelihood based estimation approaches using numerical integration presented in section 4.3, Bayesian estimation methods presented in section 7.8.3 can be used to estimate the generalised linear mixed models presented in Chapters 4-6. They are presented here as an alternative approach to the analysis of the Triage Plus study.

7.8.2 Hierarchical model for binomial data

The log-binomial random intercepts model

Given y_{ij} is the observed number of events (e.g. TB treatment initiations) out of the n_{ij} (subjects in cluster j and time point i); p_{ij} are the true probabilities of success at repeated measurement time point i in cluster j, x_{ij} is a vector of explanatory variables including the intervention status variable, β is a vector of regression coefficients, and u_j is the random intercept for cluster j; then the random intercepts log-binomial model for clustered data is defined as (Thompson, Warner and Turner, 2004, p393)

$$y_{ij}|x_{ij}, u_j \sim Binomial(n_{ij}, p_{ij})$$

$$\log(p_{ij}) = \beta x_{ij} + u_j \leq 0 \qquad (7.1)$$

$$u_j \sim N(0, \sigma^2)$$

The random effects u_j at time i for cluster $j$are random intercepts and follow the truncated half-Normal distribution, with mean zero and variance σ^2 (Turner, Omar and Thompson,

2001), in order to be consistent with the probability laws of within [0, 1) for model convergence.

The parameter space is defined as

$$\begin{cases} \beta x_{ij} \leq 0 \\ \beta x_{ij} + u_j \leq 0 \end{cases} \qquad (7.2)$$

where u_j is the cluster level random effect, assumed to be independent, and follows a Normal distribution with mean zero and variance σ^2 (variance of cluster-specific probabilities).

Since the random effect u_j is drawn from a Normal distribution with mean zero and variance σ^2, it can take any value in the range $(-\infty, \infty)$. Thus, it is symmetrically distributed around a defined point. To ensure that the probability p is within [0, 1), two approaches are possible.

First, the log-binomial method may be used, and if convergence is not achieved, the COPY Method as described in section 3.6.2.3 in Chapter 3 can be employed when using likelihood based estimation approaches.

Second, the Bayesian analysis approach may be used, where u_j is confined to the interval $(\beta x_{ij} \leq u_j \leq -(\beta x_{ij}))$ as suggested by Thompson, Warn and Turner (2004), so that the parameter space can be defined as

$$\begin{cases} \beta x_{ij} \leq 0 \\ \beta x_{ij} \leq u_j \leq -(\beta x_{ij}) \end{cases} \qquad (7.3)$$

Using the notation in model (4.13) for the random intercepts variance parameters, the extent of the dependence of binary observations within a cluster is given by

$$ICC = \frac{\sigma_{11}^2}{\sigma_{11}^2 + \pi^2/3} \qquad (7.4)$$

where σ_{11}^2 is the between cluster variance and $\pi^2/3$ is the residual variance on logit scale within clusters (Nakagawa and Schielzeth, 2010).

The log-binomial random coefficients model

To investigate whether the effect of the intervention varies across clusters over the duration of the intervention, random coefficients log-binomial model is fitted by including a random coefficients as follows:

$$y_{ij}|x_{ij}, u_j, u_{ij} \sim Binomial\ (n_{ij}, p_{ij})$$

$$\log(p_{ij}) = \beta x_{ij} + u_{ij} + u_j x_{ij} < 0 \qquad (7.5)$$

and the parameters' space defined in (7.2) becomes

$$\begin{cases} \beta x_{ij} \leq 0 \\ \beta x_{ij} + u_j x_{ij} \leq 0 \end{cases} \qquad (7.6)$$

As usual, random effects $u_j|x_{ij}$ are normally distributed with mean 0 and covariance matrix:

$$D = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}, \sigma_{21}^2 = \sigma_{12}^2$$

with σ_{11}^2 and σ_{22}^2 as variances for random intercepts and random coefficients respectively, σ_{21}^2 and σ_{12}^2 are the covariances between the random intercepts and random coefficients.

In a two arm intervention study like the Triage Plus, the random intercepts u_{ij} represents between-cluster variability in control areas, and the random intercepts plus random coefficients effects ($u_{ij} + u_j$) represent between-cluster variability in intervention areas. They respectively, therefore, represent deviations of a specified cluster from the mean intercept and the average slope for the intervention variable coefficient in all the scenarios discussed.

7.8.3 Bayesian estimation of incidence rate ratios

7.8.3.1 Introduction:

Bayesian methods were introduced by the Reverend Thomas Bayes (1702-1761), and Bayesian estimation, first used by La Place in 1786, is increasingly used in the analysis of different models because of its distinctive advantages. Bayesian approaches are currently growing in popularity for the following reasons:

- Uncertainty in variance components is easily accounted for, and the use of prior information in the parameter estimation process is allowed (Spiegelhalter, 2001; Turner, Omar & Thompson, 2001).
- Flexibility in estimating prevalence ratios or risk ratios when log-binomial models are used, and the adjustment of the domain of parameter estimation is only required to ensure non permissible probabilities greater than 1 do not occur (Thompson, Warner and Turner, 2004). In addition, they are flexible in modelling individual or cluster level covariates and variance structure as pointed out by Thompson, Warner and Turner, (2004).
- Marginal likelihood obtained by integrating out random effects is generally intractable, but it is easily implemented using Bayesian estimation methods such as the Markov chain Monte Carlo methods (MCMC) (see Zhao et al., 2006).

7.8.3.2 The Bayesian paradigm

Given y_{ij} are the observed data at measurement time i and in cluster j, θ are the parameters of the probability model, and the prior probability is denoted by π(θ). The posterior distribution in Bayesian analysis is given by

$$p(\theta|y_{ij}) = \frac{p(y_{ij}|\theta)\pi(\theta)}{p(y_{ij})} \qquad (7.7)$$

where

$p(y_{ij}) = p(y_{ij}|\theta)\pi(\theta)d\theta$ is the normalising constant of $\pi(\theta)$ which iscalled the'marginal distribution'; $p(y_{ij}|\theta)$is the likelihood function of θ (the probability of observing the data y_{ij}given the model with parameter θ). In this case the posterior distribution $p(\theta|y_{ij})$provides the parameter estimates of interest (e.g. regression coefficients of covariates).

Thus, in a Bayesian framework, the key elements are the likelihood function $p(y_{ij}|\theta)$ and the prior distribution $\pi(\theta)$. The likelihood function reflects information about the parameters contained in the data while the prior distribution quantifies what is known about the parameters before observing the data. Therefore, the larger the values of the likelihood, the more the parameters are supported by the data (Glickman and Dyk, 2007).

7.8.3.3 Computation of posterior distributions

In regards to more general problems, the joint posterior distribution $p(\theta|y_{ij})$ in (7.7) is usually complex, high dimensional (or not available in closed form) and posterior quantities cannot be obtained analytically. Therefore, numerical integration techniques using Markov chain simulations are widely used to obtain reliable posterior estimates. A discussion of the major methods for approximating integrals, with special emphasis on Bayesian integration problems, is provided by Evans and Swartz (1995).

Among the Markov chain techniques, the Markov chain Monte Carlo (MCMC), first proposed by (Metropolis et al., 1953), is among the best known simulation techniques that have been used in numerical integration. Of the MCMC techniques, the Metropolis-Hastings algorithm (M-H) (Metropolis et al., 1953, Hastings, 1970)and the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990)are the widely used MCMC methods. The development of modern statistical software, including WinBugs (Lunn et al., 2000), made the computation of complex models using MCMC simulation methods very easy. Below is a brief description of the common MCMC methods used in the simulation of posterior distributions.

Metropolis-Hastings algorithm

The Metropolis-Hastings (M-H) algorithm, named after Nicholas Metropolis and W. Keith Hastings, is a Markov chain Monte Carlo method that generates observations from a posterior distribution without computing the normalisation factor, which is a major aspect of the algorithm (Metropolis et al., 1953; Hastings, 1970).

The algorithm starts with the objective distribution $(\theta^{(1)}, \ldots., \theta^{(i-1)})$ in order to generate observations from a proposed distribution $q(\theta, \theta^{(i-1)})$ that depends on the current state to generate the new proposed sample. If a sample of size n with a posterior distribution$p(\theta|y_{ij})$, and θ^{i}, a vector of generated values in i iterations is to be generated, then the algorithm is summarised as follows:

1. Set initial values $\theta^{(0)}$

2. For $i = 1, \ldots, n$repeat the following steps
   a. Set $\theta = \theta^{(i-1)}$
   b. Generate a new parameter value θ' from a proposal distribution $q(\theta', \theta)$

   c. This generated value is accepted with acceptance rate

$$h = \min\left\{1, \quad \frac{p(\theta'|y_{ij})}{p(\theta|y_{ij})} \text{x} \frac{q(\theta|\theta')}{q(\theta'|\theta)}\right\} = \min\left\{1, \quad \frac{p(y_{ij}|\theta')}{p(y_{ij}|\theta)} \text{x} \frac{\pi(\theta')}{\pi(\theta)} \text{x} \frac{q(\theta|\theta')}{q(\theta'|\theta)}\right\}$$

   d. Obtain the new value $\theta^t = \theta'$ with probability h ; otherwise set $\theta^i = \theta$ with probability $(1 - h)$.

where the $\frac{p(\theta'|y_{ij})}{p(\theta|y_{ij})}$ represents the likelihood ratio between the proposed new value θ' and the previous value θ; $\frac{q(\theta|\theta')}{q(\theta'|\theta)}$ is the ratio of the proposed density function for θ' given the previous value θ. The superscript i denotes the iterations. The desired acceptance rate h (the fraction of the proposed samples accepted among the previous samples) depends on the target distribution for efficient estimation. The algorithm requires starting values $\theta^{(0)}$ to start simulations. The choice of starting values affects the rate of convergence if the value chosen is outside the range covered by the posterior distributions (Lunn et al., 2000).

The Gibbs sampling

The Gibbs sampling algorithm, first introduced by Geman and Geman (1984), is a special case of the Metropolis-Hastings algorithm which generates a Markov chain by sampling from posterior distributions that condition upon all other model parameters and the data (full conditional distributions) with 100% acceptance rate. With the development of adaptive rejection sampling (Gilks and Wild, 1992), the Gibbs sampling algorithm efficiently samples from any conditional distribution with log concave density functions (e.g. the log-binomial and Poisson models, and most common priors and likelihood functions).

Given that there are three regression coefficients β_0, β_1 and β_2 for a vector of explanatoryvariables x_{ij} that are to be estimated using Gibbs sampling, then the following steps are used:

1) Choose starting values for the parameters $\beta_0^{(0)}, \beta_1^{(0)}$ and $\beta_2^{(0)}$.
2) Generate a new value $\beta_0^{(i)}$ for β_0 from its posterior distribution conditioning on the data y_{ij} and the current values for β_1 and β_2.
3) Generate a new value $\beta_1^{(i)}$ for β_1 from its posterior distribution conditioning on the data y_{ij} and the current values for β_0 and β_2.
4) Generate a new value $\beta_2^{(i)}$ for β_2 from its posterior distribution conditioning on the data y_{ij} and the current values for β_0 and β_1.
5) Repeat steps 1- 4 several times while increasing i by 1 each time until convergence is achieved.

Although the development of adaptive rejection sampling to effectively sample from conditional distributions whose density functions are not log-concave (Gilks and Wild, 1992), routine use is limited. Therefore, Neal (2003) proposed a sampling technique called 'slice sampling' that is implemented for both univariate and multivariate distributions and can be used to sample from any continuous distribution.

7.8.3.4 Diagnosing the convergence of posterior distributions

Due to the iterative and Monte Carlo nature of the MCMC techniques, there is a need to assess the convergence of the MCMC methods. Procedures for diagnosing convergence in MCMC methods have been proposed (Brooks and Gelman, 1998). The main issues considered in MCMC methods include how quickly the simulated chains converge to a target posterior distribution (i.e. the required number of 'burn-in' iterations that are excluded before making necessary inferences) and how efficiently the posterior quantities are estimated from the sample chains.

Built-in statistical and graphical tools can be used for checking convergence (Brooks and Gelman, 1998). Convergence can also be ascertained by monitoring the Monte Carlo errors for estimates with small values of this error, indicating the parameter estimate of interest is estimated with precision (Ntzoufras, 2009).

When convergence is achieved, plots of samples of parameters show a random scatter around a stable mean value in no specific pattern, and there is much overlap of the trace plots chains, which indicate independence in the samples (as an example see Figure 15 showing convergence after 10,000 iterations with 1000 iterations as 'burn-in' for the coefficient for the intervention variable measuring effect size using TB treatment initiations data from the Triage plus study. The two chains show overlapping patterns indicating convergence is achieved). However, if the samples are not largely independent, then there may be autocorrelation between samples, which can be addressed by generating additional samples (Spiegelhalter et al., 2003). More iterations should be made after convergence to obtain posterior inference after discarding the burn-in iterations. To reduce burn-in period, different starting values for model parameters and variance components should be tried to ensure convergence.

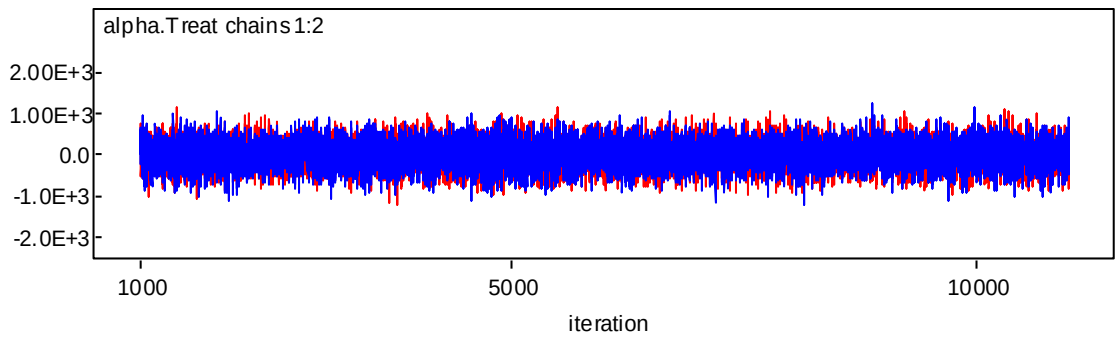


Figure 15: History plots for the converged chains.

The history plots for the coefficient for the intervention variable measuring effect size using TB treatment initiations data from the Triage plus study. The two chains show overlapping patterns indicating convergence is achieved.

The Monte Carlo standard error (MC error), an estimate of the difference between the mean of sampled values and the true posterior mean of a parameter, is used to evaluate the efficiency of the MCMC methods (MC error measures the extent simulation error contributes to uncertainty in mean estimation) and is usually reduced by generating more samples. The efficiency of the posterior estimates is guaranteed if the MC error is less than 5% of the posterior standard deviation of the parameter estimate (Spiegelhalter et al., 2003).

7.8.3.5 Prior distributions

To carry out Bayesian analysis, prior distributions of the model parameters need to be specified. Best knowledge about the parameters is used to make a selection; otherwise, non-informative priors are usually adopted. The correct choice of the priors can be ascertained easily if the posterior distributions are relatively stable (see Figure 16), an indication that the data contains sufficient information. When the correct prior for the model and inference has been correctly determined, it can then be used to carry out analysis for the model and inference.

Different priors are available and have been used in the literature for Bayesian analysis. The choice of priors might be guided by two approaches. Firstly, 'informative priors' are selected based on a strong prior belief regarding the distribution of the parameters of interest in which subjective prior information is available**.** Informative priors tend to show impact on the posterior distributions and are dominated by the likelihood. Secondly, when no prior belief is available, objective priors are used and analysis is conducted without the influence of prior information (Spiegelhalter, 2001; Turner, Omar and Thompson, 2001).

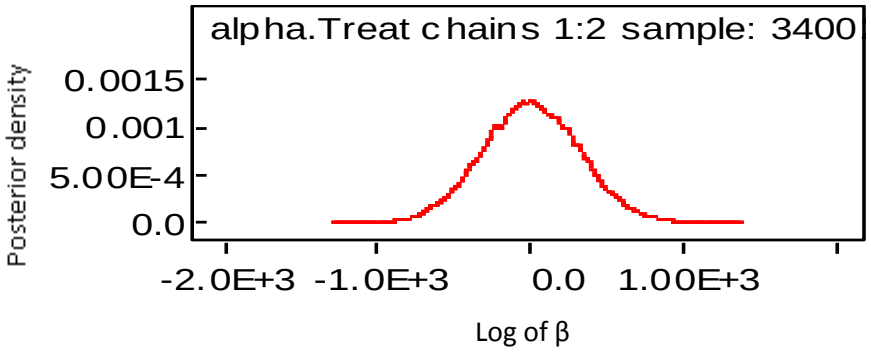


Figure 16: Density plot for the posterior distribution for the coefficient of the intervention using TB treatment initiations data from the Triage plus study. The x-axis is the log of the coefficient of variable *treat* (i.e. log of incidence rate ratio for variable *treat*) and the y-axis is the posterior density of the log of the coefficient of variable *treat*.

Prior assumptions

In the Bayesian paradigm, model formulation is complete when priors for model parameters are specified, which accounts for all available knowledge about the parameters. In the random effects models considered here, the prior distributions will often depend on higher level

parameters called 'hyperparameters', in which case additional hyper priors have to be specified.

Prior distributions of fixed effects parameters

Guided by the need for priors to have minimal impact on the posterior distribution and due to the lack of previous knowledge on the size of intervention effects (in our case effect in regards to engaging informal health care providers), prior distribution for fixed effects vector β takes the form

$$\beta \sim N(0, D(\theta))$$

for some covariance matrix F as given by Zhao et al. (2006). By making the covariance matrix $D(\theta)$ diagonal with large variance of 100,000 coupled with appropriate choices of variance components, a proper joint posterior distribution of parameters can be achieved (Zhao et al., 2006). The commonly used prior distributions for analysis of fixed effects parameters are $\beta \sim N(0, 100000)$. For Bayesian analysis the precision is expressed as an inverse of the variance and therefore the fixed effects parameters are instead expressed as $\beta \sim N(0, 0.00001)$.

Prior distributions for variance parameters:

Inverse-gamma

The common prior density for variance parameters often used in generalised linear mixed models for random effects precision is the inverse-gamma $(0.001,\ 0.001)$ proper prior distribution (Spiegelhalter et al., 2003) for single variance components and the inverse Wishart distribution for a variance-covariance matrix because of the conjugacy properties with the Normal distribution for random effects (Zhao et al., 2006).

Uniform prior

When prior information is not available or cannot subjectively be specified for model parameters, uniform priors, also known as 'flat priors', are the common choice for non-informative priors. Generally, relatively uniform proper prior distributions that integrates to 1, (for binomial distributions, the uniform prior is given as $\pi\ (p) = 1$) are desired. Among the set of relatively flat distributions, the choice of prior makes no difference if the data contains adequate information about the parameters of interest (Glickman ME, and Dyk DA, 2007).

The uniform prior, with large standard deviation and 0 lower bound (i.e. U(0,1000), has been suggested in the literature as the prior for random effects variance parameters (Gelman, 2006). According to Gelman (2006), the uniform $(0, \infty)$ prior usually yields limiting proper posterior distributions. However, it leads to a heavy-tailed posterior distribution that result in overestimating σ, especially in small numbers of clusters (Gelman, 2006). This results in less than optimal shrinkage for estimating group level random effects (Gelman, 2006).

Posterior quantities:

In Bayesian model estimation, posterior mean or median and the variance of the estimated posterior means for the model parameters are the primary quantities of interest. The median has been preferred at times because of its stability (Thompson, Warn and Tanner, 2004).

Given $G(\theta)$ is a function of a parameter of interest θ estimated after generating a sample size (iterations) i iterations, the maximum number of iterations excluding burn-in iterations, the posterior mean of the estimate is given by

$$\hat{E}\left(G(\theta)\middle|y_{ij}\right) = \overline{G(\theta)} = \frac{1}{n}\sum_{i=1}^{n} G(\theta^i) \qquad (7.8)$$

and the posterior standard deviation is given by

$$\widehat{SD}\left(G(\theta)\middle|y_{ij}\right) = \frac{1}{n-1}\sum_{i=1}^{n}[G(\theta^i) - \hat{E}\left(G(\theta)\middle|y_{ij}\right)]^2 \qquad (7.9)$$

The MC error is given as a measure of the efficiency of posterior estimation. Hence, to calculate the parameter of interest with increased precision, the MC error has to be low as it measures the variability of the estimate due to the simulation process and is achieved by having a sufficient number of iterations. To estimate the MC error, two methods have commonly been used: the *batch mean* method and the *window estimator* method. Although the window estimator method is more precise, the batch mean method is more popular and easier to implement (Ntzoufras, 2009).

Confidence intervals

In Bayesian analysis, credible intervals and the highest posterior density (HPD) are derived as part of the posterior quantities (Ntzoufras, 2009). The credible interval estimates are derived as the interval of the ordered sample as follows:

Given an MCMC sample$\{\theta_i, i = 1,2, \ldots . n\}$,

i. Obtain the ordered values of $\theta_{(1)} \leq \theta_{(2)} \leq \cdots \leq \theta_{(n)}$ by sorting$\{\theta_i, i = 1,2, \ldots . n\}$.
ii. Compute the 100(1-α)% credible intervals for each of the MCMC sample.
iii. The smallest interval of the credible intervals in (ii) is the 100(1-α)% highest posterior density (HPD) interval given the same density for the two boundaries of unimodal distributions.

7.8.3.6 Bayesian model selection

Just as in the likelihood based approaches, identifying an appropriate model that fits the given data may be challenging. This is especially true when there is a given set of potential statistical models that can be used for the given data. In Bayesian analysis, approaches such as the Akaike Information Criterion (AIC) (Akaike, 1981), Bayesian Information Criterion (BIC) (Schwarz, 1978), Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) and Bayes Factor (Kass and Raftery, 1995) have been used in determining the best model. However, when none of the models show clear superiority, Bayesian model averaging (Hoeting et al. 1999) may be used to find the appropriate model. A brief description of the model selection methods is given.

7.8.3.6.1 Model selection with Bayes factors

Bayes factors have dominated the model selection process because not only are they easy to calculate and do not require models to be nested, but also because they enable external information to be incorporated when assessing a hypothesis. Kass and Raftery, (1995) provide a detailed review of the Bayes factors and other model selection approaches.

Definition of a Bayes factor:

Given two models, M_0 and M_1, arising from data y_{ij}, with probability densities (marginal likelihood) $p(y_{ij} | M_0)$ and $p(y_{ij} | M_1)$ respectively, and assuming posterior probabilities for models M_0 and M_1 are $p(M_0 | y_{ij})$ and $p(M_1 | y_{ij})$ respectively, the model that maximises $p(M_k|y_{ij})$ is normally chosen (where k=0,1) according to Wasserman (2000).

Given prior probabilities being $p(M_0)$ and $p(M_1)$ for model M_0 and M_1 respectively, then according to Bayes' theorem as presented in Kass and Raftery (1995 page 776), the posterior probabilities are obtained as:

$$p(M_1|y_{ij}) = \frac{p(y_{ij}|M_1)p(M_1)}{p(y_{ij}|M_0)p(M_0) + p(y_{ij}|M_1)p(M_1)}, \qquad (7.10)$$

So that

$$\frac{p(M_1|y_{ij})}{p(M_0|y_{ij})} = \frac{p(y_{ij}|M_1)p(M_1)}{p(y_{ij}|M_0)p(M_0)} \qquad (7.11)$$

Rearranging (7.11), the Bayes Factor in favour of model M_1 against model M_0isgiven by:

$$BF_{10} = \frac{p(y_{ij}|M_1)}{p(y_{ij}|M_0)} = \frac{p(M_1|y_{ij})}{p(M_0|y_{ij})} \div \frac{p(M_1)}{p(M_0)} \qquad (7.12)$$

The Bayes factor is, therefore, the posterior odds in favour of M_1 divided by the prior odds in favour of M_0, and it measures the relative agreement of the data for the two models being compared (M_1 versus M_0) over the parameter space and summarises the evidence in support of M_1 against M_0 provided by the data (Raftery, 1996). If the prior odds for the two models are equal, which is usually the case, then the posterior odds of the two models is equal to the Bayes factor (Wasserman, 2000), in which case $p(M_1) = p(M_0)=0.5$ and $p(M_1) + p(M_0)=1$.

Given the marginal likelihood of data y_{ij} for model M_k as $p(y_{ij}|M_k)$, which is obtained by integration (Kass and Raftery, 1995) as shown by:

$$p(y_{ij}|M_k) = \int p(y_{ij}|\theta_k, M_k)p(\theta_k|M_k)\, d\theta_k \qquad (7.13)$$

where θ_k is the parameter to be estimated under model M_k for $k = 0$ and 1, $p(\theta_k|M_k)$ is a prior distribution for θ_k under M_k and $p(y_{ij}|\theta_k, M_k)$ is the likelihood function for θ_k .

Then the Bayes factor, BF, can be rewritten as the ratio of the marginal probabilities of M_1 and M_0 as:

$$BF_{10} = \frac{p(M_1|y_{ij})}{p(M_0|y_{ij})} \div \frac{p(M_1)}{p(M_0)} = \frac{p(y_{ij}|M_1)}{p(y_{ij}|M_0)} = \frac{\int p(y_{ij}|\theta_1, M_1)p(\theta_1|M_1)\, d\theta_1}{\int p(y_{ij}|\theta_0, M_0)p(\theta_0|M_0)\, d\theta_0} \qquad (7.14)$$

However, when more than two models are under consideration, according to Raftery (1996 page 253), then each model M_1, .. M_k is compared against model M_0 in turn, yielding corresponding Bayes factors BF_{10}, .. BF_{k0}. Using the notation by Raftery (1996 page 253), the posterior probability that M_k is true and takes into account model uncertainty is given by

$$p(M_k|y_{ij}) = \frac{\alpha_k BF_{k0}}{\sum_{k=0}^{K} \alpha_k BF_{k0}} \qquad (7.15)$$

with $\alpha_k = p(M_k)/p(M_0)$ being the prior odds for M_k against M_0, and with equal prior odds assumed therefore $\alpha_k = 1$.

For the purposes of model selection, if $2ln(BF_{10})$ >5, then M_1 is substantially in better agreement with the data (Kass and Raftery, 1995).

The integration of the posterior normalising constant $p(y_{ij}|\theta_k, M_k)p(\theta_k|M_k)$ requires a numerical evaluation that is difficult for models with high-dimensional parameters and the fact that the estimates of the marginal densities from the MCMC samples may be unstable (Albert and Chib 1993 cited by Carlin and Chib, 1995). However, according to Diciccio et al (1997), by combining simulation and asymptotic approximations, simulated versions of methods such as Laplace, Bartlett correction or importance sampling can be used to calculate Bayes factors efficiently, provided the posterior distribution has a single dominant mode and direct calculation of the posterior normalising constant is avoided.

Calculation of Bayes Factors

Given that there are two models M_1 and M_0, to be compared and recall that the normalising constant $p(y_{ij_1})$ for model M_1 is given as $p(y_{ij_1}) = f_1(y_{ij}) = \int h\left(\theta|y_{ij_1}\right) d\theta$, where $h(\theta|y_{ij}) = p(y_{ij}|\theta)\pi(\theta)$ with parameter $\theta \in \Omega$, the Ω denotes the parameter space for θ . Another constant is $p(y_{ij_0}) = f_0(y_{ij}) = \int q\left(\theta|y_{ij_0}\right) d\theta$, with y_{ij_0} being part of $y_{ij}, \theta \in \Omega_0$. The subscripts 1 and 0 stand for model M_1 and model M_0.The Bayes factor in support of M_1 against M_0 is given as the ratio of the constants:

$$BF_{10} = \frac{f_1(y_{ij})}{f_0(y_{ij})} = E_0\left[\frac{h\left(\theta|y_{ij_1}\right)}{q\left(\theta|y_{ij_0}\right)}\right] \qquad (7.16)$$

where E_0isthe expectation with respect to the posterior distribution of model M_0. To estimate the Bayes factor using importance sampling, suppose $\{\theta^{(i)}, i = 1, \ldots, n\}$ is a random sample

drawn from a simulated distribution G with a probability density q, then the importance sampling estimate of the Bayes factor BF_{10} is given by

$$\widehat{BF}_{10} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{h(\theta^i|y_{ij_1})}{q(\theta^i|y_{ij_0})}\right] \qquad (7.17)$$

When density q is properly chosen such that it is similar to h in order to reduce the ratio $h(\theta^i|y_{ij_1}/q(\theta^i|y_{ij_0})$, the $\widehat{BF}_{10}$ is an unbiased and consistent estimator of BF_{10}.

By using the sampling-importance sampling algorithm, the posterior sample from the posterior distribution for model M_1 may be obtained as follows. A new sample with probability proportional to the ratio $h(\theta^i|y_{ij_1}/q(\theta^i|y_{ij_0})$ should be drawn from a simulated sample θ^i so that it is approximately distributed according to the model M_1 posterior (Albert, 1996).

To avert problems in computing integrals for the Bayes factors, Carlin and Chib (1995) proposed the inclusion of model M in the sampling process to obtain a sample from the marginal posterior distribution for M and regard the prior distributions $p(\theta_k), k = 1, \dots K,$ as part of model specification. Given model M_k with a likelihood $p(y_{ij}|\theta_k, M_k)$ and a prior$p(\theta_k|M_k)$ and data $y_{ij}$with $\theta_k's$ conditionally independent given the model M, Carlin and Chib (1995) completed the model specification by choosing 'pseudopriors' $p(\theta_k|M \neq k)$ such that:

$$p(y_{ij}|M_k) = \int p(y_{ij}|\theta, M_k)p(\theta|M_k)d\theta = \int p(y|\theta_{k,}M_k)p(\theta_k|M_k)d\theta_k\,, \qquad (7.18)$$

from the conditional independence assumptions. According to Carlin and Chib (1995), the joint distribution of y_{ij} and θ when $M = k$ given the model prior probability $p(M_k)$ is

$$p(y_{ij}, \theta, M_k) = p(y_{ij}|\theta_k, M = k)\left\{\prod_{k=1}^{K} p(\theta_k|M_k)\right\}p(M_k)\,. \qquad (7.19)$$

and the full conditional distributions for each $\theta_k$is given as

$$p(\theta_k|\theta, M, y_{ij}) \propto \begin{cases} p(y_{ij}|\theta_k, M = k)p(\theta_k|M = k), & M = k, \\ p(\theta_k|M \neq k), & M \neq k, \end{cases} \qquad (7.20)$$

where k denotes the candidate models. In this specification, M=k then the Gibbs sampler generates the usual full conditional distributions for the model k. However, if the $M \neq k$ Carlin and Chib (1995) proposed to generate from the linking density he called pseudopriors.Using the MCMC methods, the routine generation of model M is given by

$$p(M = k|\theta, y_{ij}) = \frac{p(y_{ij}|\theta_k, M = k)\{\prod_{k=1}^{K} p(\theta_k|M = k)\}p(M_k),}{\sum_{k=1}^{K} p(y_{ij}|\theta_k, M = k)\{\prod_{k=1}^{K} p(\theta_k|M = k)\}p(M_k)} \quad (7.21)$$

The MCMC algorithm will produce samples from correct posterior distributions because all the full conditional distributions are well defined. According to Carlin and Chib (1995), the ratio $\hat{p}(M = k|y_{ij})$ provides a ratio that could be used to estimate Bayes factor between any two models.

However, if there is an extreme imbalance in one of the $p(M = k|y_{ij})$, then the model probability $p(M_k)$ is adjusted before retaining samples for estimation of Bayes factors (Carlin and Chib, 1995). Corresponding standard errors are obtained from the MCMC outputs, and it is assumed independent samples and batching techniques can be used in cases of autocorrelation (Carlin and Chib, 1995, p 478).

7.8.3.6.2 Deviance Information Criterion

The Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002) is a generalisation of the AIC that is used as a measure of model fit and complexity and as an approach for model selection. Classical deviance for model assessment is based on the difference in the log-likelihoods between the fitted and saturated model as defined in (4.31) for the likelihood ratio test. Using the Bayesian approach, the model in (4.32) is given as:

$$D(\theta) = -2\log\left(p(y_{ij}|\theta)\right) + 2\log(p(y_{ij}|\hat{\theta}) \quad (7.22)$$

where $D(\theta)$ is the Bayesian deviance, $\log\left(p(y_{ij}|\theta)\right)$ is the log likelihood for the saturated model, $\log p(y_{ij}|\hat{\theta})$ is the log likelihood for the fitted model. Using this specification, the L_1 and L_0 in model (4.31) is represented by the $\log\left(p(y_{ij}|\theta)\right)$ and $\log p(y_{ij}|\hat{\theta})$ respectively.

The DIC as a model selection criterion was based on the posterior distribution of the Bayesian deviance and is given as:

$$\text{DIC} = \hat{D} + PD \quad (7.23)$$

where

$\widehat{D} = E[D(\theta)]$ is the Bayesian measure of goodness of model fit;

PD is the effective number of parameters that measures the complexity of the model and is defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean $\hat{\theta}$ of the posterior distribution as:

$$PD = \widehat{D} - D(\hat{\theta}) = E[D(\theta)] - D(E(\theta)) \qquad (7.24)$$

Thus by rearranging (7.24) to $\widehat{D} = D(\hat{\theta}) + PD$, gives

$$DIC = D(\hat{\theta}) + 2PD \qquad (7.25)$$

The smaller the DIC for a given model, the better the model fits the data and the better the model will be for making predictions. More complex models fit the data better, and thus, have smaller values of $\widehat{D}$. Differences of DIC>10 are regarded as substantial and differences of DIC < 5 may be negligible (Best, Mason and Li, 2011).

7.8.3.6.3 Bayesian model averaging (BMA)

However, when none of the models show clear superiority, Bayesian model averaging (Hoeting et al., 1999) is superior to the single models in terms of its predictive performance. The predictions are a result of weighted single model predictions that provide posterior mean expectations (Hoeting et al., 1999) where model uncertainty is present.

Given M=$\{ M_1, \ldots, M_k \}$ is a set of all models under consideration and θ representsthe parameters of interest (Hoeting, Raftery and Madigan, 2002; Hoeting et al. 1999) for data y_{ij}, the Bayesian model averaging is given as

$$p(\theta|y_{ij}) = \sum_{k=1}^{K} p(\theta|M_k, y_{ij}) p(M_k|y_{ij}) \qquad (7.26)$$

that averages posterior model distributions under each model weighted by corresponding model probabilities to account for model uncertainty (Madigan and York, 1995; Hoeting et al., 1999; Hoeting, Raftery and Madigan, 2002). While the posterior model probability for model M_k is given by

$$p(M_k|y_{ij}) = \frac{p(y_{ij}|M_k)p(M_k)}{\sum_{k=1}^{K} p(y_{ij}|M_k)p(M_k)} \qquad (7.27)$$

where

$p(y_{ij}|M_k) = \int p(y_{ij}|\theta_k, M_k)p(\theta_k|M_k)d\theta_k$ is the marginal likelihood for model M_k,

$p(M_k)$is the prior probability for model M_k.

$\theta_k$is a vector of parameters (Hoeting et. 1999).

Therefore, the posterior mean and variance of θ given the data y_{ij}(as described by Hoeting et al. 1999 page 383) is given by

$$E[\theta|y_{ij}] = \sum_{k=1}^{K} \hat{\theta}_k p(M_k|y_{ij}) \qquad (7.28)$$

and

$$Var[\theta|y_{ij}] = \sum_{k=1}^{K} var[\theta|y_{ij}, M_k] + \hat{\theta}_k^2 p(M_k|y_{ij}) - E[\theta|y_{ij}]^2) \qquad (7.29)$$

Implementing Bayesian Model Averaging (BMA)

However, because there may be a great number of candidate models, including the BMA, resulting from an increased number of covariates in (7.26), computation of posterior probabilities for each model might not be practical (Hoeting, Raftery and Madigan, 2002). Therefore, Markov chain Monte Carlo model composition proposed by Madigan and York (1995)is used to directly approximate the equation in(7.26)with considerable flexibility and provides improved predictive performance against any single best model.

If M is a class of models under consideration, then in constructing a simulated Markov chain $\{M_{(i)}, i= 1,...,n\}$ with a given state space and equilibrium distribution $p(M_k \mid y_{ij})$ a sequence of model observations M_k must be obtained. For any function $g(M_k)$, according to Madigan and York (1995) the average

$$\hat{G} = \frac{1}{n}\sum_{i=1}^{n} g\{M(i)\} \qquad (7.30)$$

converges to give an estimate of E(g(M) as N→∞ (Smith and Roberts 1993 as cited by Hoeting, Raftery and Madigan, 2002). Madigan and York recommended setting g(M) = p(θ |M, y_{ij}) in order to compute equation in (7.26).

*7.8.4 F*uture simulation work

The simulation study findings reported in Chapter 5 showed that statistical efficiency and power varied in the different design conditions investigated including the 3 clusters per arm design. However, the simulations did not assess the potential impact on statistical power and the accuracy of the parameter / effect size estimates of varying the group level explanatory variables. The ICCs considered in this dissertation represented situations of low to high correlations between clusters. The simulation studies presented in Chapter 5 showed that the size of the intracluster correlation affected both statistical power and the accuracy of parameter estimates (see Tables 8 & 9) when intervention status was used as the only group level explanatory variable.

Thus, future simulation studies are needed as a matter of priority to assess the effect of varying intracluster correlation and group level explanatory variables on statistical power and accuracy of parameter estimates. Less urgently, but still importantly, further simulation studies are needed to more fully evaluate the optimum numbers clusters and repeated measurement times needed in different study design situations.

8.0 Appendix

Appendix 8.1: A statement of declaration of the contributions made by me and other people in the whole study

I write to declare that my key contributions in the whole study were contributing to refining the study conceptual framework of engaging informal healthcare providers in TB and HIV community interventions to improve TB and HIV service uptake rates as well as designing a proper cluster randomised trial with more than one cluster per arm (i.e. I changed the original study design of assigning one cluster per study arm to a design with multiple clusters). However, because the study was limited to one district due to cost and the administrative logistics of having the study done in multiple districts, only a small number of clusters could be defined in order to limit the possible contamination between study arms (in this case 3 clusters per arm). To ensure that adequate statistical power was achieved in the final analysis of the study given the limited number of clusters, I conducted simulation studies that informed the appropriate design conditions that were then pursued in the evaluation of the actual Triage Plus intervention. In addition, implementation of the study design, interventions and quality data management procedures was ensured in order to align to standard cluster randomised trials (e.g. blinding patient allocation to clusters) and this work was accepted for oral presentation at the 2013 Union Conference (see Appendix 8.7).

My supervisors namely Prof Bertel Squire, Dr Brian Faragher and Rachael Thomson from the Liverpool School of Tropical Medicine were key in the initial conceptualisation of the intervention as well as obtaining funding for the implementation of the study, whose final results forms part of this thesis. They provided technical support throughout the implementation of the Triage Plus study. The technical support included in the definition of the clusters and randomisation process, support in the definitions of the study objectives and outcome measures, presentation of the results as well as the structure of the thesis and continuously reviewing each stage of the thesis development until it was submitted.

Appendix 8.2: Data recording/collection forms

Appendix 8.2.1: TB treatment registration form-sample

Facility Name	Reg. Date	TBNO	Sex	Age	TA	Village	Treat unit	TB Classific	Patient Category	Initial Smear Results	Smear Date	Out-come	Date Outcome	HIV status
Area 25	19/01/2012	12	2	79	Chitukula		Area 25	P	New	Negative				0
Area 25	21/02/2012	22	2	20	Chimutu	Kamlembo	Area 25	P	New	Negative	24/04/2012			1
Area 25	2/3/2012	25	1	56	Mtema	Mzungu	Ngoni	P	New	Negative				0
Area 25	2/3/2012	29	1	62	Chitukula		Area 25	P	New	Negative	16/05/2012			0
Area 25	8/3/2012	30	2	27	Chitukula		Area 25	EP	New	Negative				0
Area 25	9/3/2012	31	2	31	Kabudula	Mngwangwa	Area 25	P	New	Negative				1
Area 25	15/03/2012	33	2	75	Njewa	Chikhuthe	Area 25	P	New	Positive				0
Area 25	16/03/2012	38	1	29	Mtema	Mtsukwa	Area 25	EP	New	Negative				0
Bwaila	5/1/2012	14	2	19	Chimutu	Mseche	Chiwamba	EP	New	Negative				0
Bwaila	6/1/2012	17	2	33	Kabudula	Malovu	Chizu	EP	New	Negative				0
Bwaila	9/1/2012	26	1	60	Mtema	Chule	Ngoni	P	New	Negative				0
Bwaila	9/1/2012	28	1	57	Kabudula	Chiswentche	Chikowa	P	New	Negative	4/1/2012			0
Bwaila	9/1/2012	46	2	46	Chadza	Sanjiko	Bwaila	P	New	Negative	5/1/2012			1
Bwaila	12/1/2012	60	1	45	Tsabango	Chiuzira	Bwaila	P	New	Negative	10/1/2012			1
Bwaila	17/01/2012	75	1	49	Chimutu	Kaomba	Area 18	P	New	Negative				1
Bwaila	17/01/2012	78	2	26	Malili		Chitedze	P	New	Negative				0

Appendix 8.2.2: Anti-retroviral treatment registration form

Health facility	ARV No.	Registration Date	Sex	Age	TA	Village	ART Status	Reason for starting ART	KS	E	e	ART Starting Date
Kabudula	626	06.01.2009	F	25	Khongoni	Kaluzi	TI	3				01.02.2008
Kabudula	627	06.01.2009	F	27	Kalolo	Sani	FT	3				06.01.2009
Kabudula	628	06.01.2009	M	40	Kalolo	Sani	FT	3				06.01.2009
Kabudula	629	13.01.2009	F	30	Kabudula	Kawiya	TI	4				18.04.2008
Kabudula	630	13.01.2009	M	39	Kabudula		FT	3				13.01.2009
Kabudula	631	13.01.2009	M	47	Khongoni	Chilowa	FT	3		Default	01.06.2009	13.01.2009
Kabudula	632	13.01.2009	M	68	Khongoni	Chilowa	FT	3				13.01.2009
Kabudula	633	13.01.2009	F	54	Kabudula	Katutula	FT	3				13.01.2009
Kabudula	634	13.01.2009	F	33	Kabudula	Kadala	FT	3		Default	30.08.2009	13.01.2009
Kabudula	635	13.01.2009	F	30	Khongoni	Chimbayo	FT	3			01.12.2009	13.01.2009
Kabudula	636	13.01.2009	F	40	Kalolo	Phandula	FT	3		Transfer	04.03.2009	13.01.2009
Kabudula	637	13.01.2009	F	33	Kabudula	Lundu	FT	3				13.01.2009
Kabudula	638	13.01.2009	M	56	Kabudula	Kambudzi	FT	3				13.01.2009
Kabudula	639	13.01.2009	F	54	Kabudula	Kambudzi	FT	3				13.01.2009
Kabudula	640	13.01.2009	F	40	Khongoni		FT	3				13.01.2009
Kabudula	641	14.01.2009	M	41	Kabudula	Mavele	FT	3		Transfer		14.01.2009
Kabudula	642	20.01.2009	F	32	Kabudula	Chipeni	FT	3				20.01.2009
Kabudula	643	20.01.2009	F	39	Kabudula	Kadiya	FT	4				20.01.2009
Kabudula	644	20.01.2009	F	30	Khongoni	Chizu	FT	3				20.01.2009

Appendix 8.2.3: Presumptive TB testing registration form

Facility Name	Lab SN	Date	Sex	Age	Treatment Unit	TA	Village	Sputum	Specimen results
ABC	3	06.01.2009	F	36	ABC	Njewa		New	Negative
ABC	7	09.01.2009	M	10	ABC			New	Negative
ABC	18	29.01.2009	M	39	Mtsiliza Clinic	Njewa		New	Negative
ABC	38	17.02.2009	F	21	Mtsiliza Clinic	Njewa	Muzu	New	Negative
ABC	39	17.02.2009	M	43	Mtsiliza Clinic	Chimutu	Salambula	New	Negative
ABC	48	24.02.2009	F	69	Mtsiliza Clinic	Njewa	Lusi	New	Negative
ABC	50	24.02.2009	F	23	Mtsiliza Clinic	Njewa	Nsanje	New	Negative
ABC	55	05.03.2009	M	40	Mtsiliza Clinic	Chitukula	Chisusu	New	Negative
ABC	61	16.03.2009	M	39	Mtsiliza Clinic	Njewa	Chilota	New	Negative
ABC	63	19.03.2009	M	36	ABC	Chitukula	Chatata	New	Positive
ABC	74	01.04.2009	M	52	ABC	Njewa	Chagonga	New	Negative
ABC	75	01.04.2009	M	23	Mtsiliza Clinic	Chitukula	Mwenela	New	Negative
ABC	82	20.04.2009	M	40	ABC	Njewa	Chimombo	New	Negative
ABC	86	22.04.2009	F	70	Mtsiliza Clinic	Njewa	Kanthache	New	Negative
ABC	93	06.05.2009	F	27	ABC	Njewa	Chikwawo	New	Negative
ABC	94	06.05.2009	M	50	ABC	Kalolo	Nsundwe	New	Negative
ABC	121	12.06.2009	F	59	Mtsiliza Clinic	Njewa	Chimombo 2	New	Negative
ABC	122	12.06.2009	F	51	Mtsiliza Clinic	Njewa	Nsanje	New	Negative
ABC	134	27.06.2009	M	76	ABC	Masumbankhunda		New	Negative
ABC	174	31.08.2009	M	49	Mtsiliza Clinic	Chitukula	Mwenela	New	Negative
ABC	176	01.09.2009	M		ABC	Kabudula	Mdzibwa	New	Negative
ABC	182	02.09.2009	F	39	Mtsiliza Clinic	Chitukula	Mwenela	New	Negative

Appendix 8.2.4: HIV testing registration form

Facility Name	Date	Patient ID Number	Sex	Age	Ever had HIV test	Name of Testing Unit	Final result
Chimbalanga	02.01.2009	1	FNP	39	Y	Chimbalanga	Negative
Chimbalanga	02.01.2009	2	M	49	Y	Chimbalanga	Negative
Chimbalanga	02.01.2009	3	FP	16	N	Chimbalanga	Negative
Chimbalanga	02.01.2009	4	FP	20	Y	Chimbalanga	Negative
Chimbalanga	02.01.2009	5	M	27	Y	Chimbalanga	Negative
Chimbalanga	02.01.2009	6	M	47	Y	Chimbalanga	Negative
Chimbalanga	02.01.2009	7	FNP	20	Y	Chimbalanga	Negative
Chimbalanga	02.01.2009	8	FNP	20	Y	Chimbalanga	Negative
Chimbalanga	05.01.2009	9	FP	22	Y	Chimbalanga	Negative
Chimbalanga	05.01.2009	10	FNP	33	Y	Chimbalanga	Negative
Chimbalanga	05.01.2009	11	M	25	Y	Chimbalanga	Negative
Chimbalanga	05.01.2009	12	FP	22	Y	Chimbalanga	Negative
Chimbalanga	05.01.2009	13	FP	19	N	Chimbalanga	Negative
Chimbalanga	05.01.2009	14	FNP	58	N	Chimbalanga	Negative
Chimbalanga	05.01.2009	15	FP	23	Y	Chimbalanga	Negative
Chimbalanga	05.01.2009	16	FP	27	N	Chimbalanga	Negative
Chimbalanga	05.01.2009	17	FNP	22	N	Chimbalanga	Negative
Chimbalanga	05.01.2009	18	M	26	Y	Chimbalanga	Negative
Chimbalanga	05.01.2009	19	M	23	Y	Chimbalanga	Negative
Chimbalanga	05.01.2009	20	M	25	N	Chimbalanga	Negative
Chimbalanga	12.01.2009	61	FP	26	Y	Chimbalanga	Negative
Chimbalanga	12.01.2009	62	FP	32	Y	Chimbalanga	Positive
Chimbalanga	12.01.2009	63	FNP	45	Y	Chimbalanga	Negative

Appendix 8.3 Cluster level population sizes

Cluster-pair (cluster IDs)*	Early arm			Delayed arm		
	Total	**Male**	**Female**	**Total**	**Male**	**Female**
1 (3 & 5)	191615	91,962	99,653	229184	111,387	117797
2 (1 & 6)	243826	120,388	123,438	232433	114,152	118,281
3 (4 & 2)	166702	80,414	86,288	167074	82,023	85051
Total	**602143**	**292764**	**309,379**	**628691**	**307562**	**321129**
***Cluster IDs 3, 1, 4 are for the Early arm and Cluster IDs 5,6,2 are for the Delayed arm**						

Appendix 8.4: Sample of dataset layout for TB treatment initiations (similar layout used for ART initiations, HIV testing and TB testing)

Clustid	time	y	treat	Base	Sites	ptfemale	clpop	timecenter	treat_time	y-center
1	1	13	1	16.9	1	0.465	243826	-10	1	-1.25
1	2	16	1	16.9	1	0.465	243826	-8	2	1.75
1	3	24	1	16.9	1	0.465	243826	-6	3	9.75
1	4	13	1	16.9	1	0.465	243826	-4	4	-1.25
1	5	19	1	16.9	1	0.465	243826	-2	5	4.75
1	6	18	1	16.9	1	0.465	243826	0	6	3.75
1	7	13	1	16.9	1	0.465	243826	0	7	-1.25
1	8	11	1	16.9	1	0.465	243826	2	8	-3.25
1	9	11	1	16.9	1	0.465	243826	4	9	-3.25
1	10	10	1	16.9	1	0.465	243826	6	10	-4.25
1	11	12	1	16.9	1	0.465	243826	8	11	-2.25
1	12	11	1	16.9	1	0.465	243826	10	12	-3.25
2	1	16	0	13.7	2	0.347	167074	-10	0	2.58
2	2	12	0	13.7	2	0.347	167074	-8	0	-1.42
2	3	10	0	13.7	2	0.347	167074	-6	0	-3.42
2	4	11	0	13.7	2	0.347	167074	-4	0	-2.42
2	5	17	0	13.7	2	0.347	167074	-2	0	3.58
2	6	13	0	13.7	2	0.347	167074	0	0	-0.42
2	7	18	0	13.7	2	0.347	167074	0	0	4.58
2	8	11	0	13.7	2	0.347	167074	2	0	-2.42
2	9	7	0	13.7	2	0.347	167074	4	0	-6.42
2	10	12	0	13.7	2	0.347	167074	6	0	-1.42
2	11	18	0	13.7	2	0.347	167074	8	0	4.58
2	12	16	0	13.7	2	0.347	167074	10	0	2.58
3	1	16	1	9.4	3	0.531	191615	-10	1	1.33
3	2	11	1	9.4	3	0.531	191615	-8	2	-3.67
3	3	7	1	9.4	3	0.531	191615	-6	3	-7.67
3	4	12	1	9.4	3	0.531	191615	-4	4	-2.67
3	5	13	1	9.4	3	0.531	191615	-2	5	-1.67
3	6	20	1	9.4	3	0.531	191615	0	6	5.33
3	7	30	1	9.4	3	0.531	191615	0	7	15.33
3	8	20	1	9.4	3	0.531	191615	2	8	5.33
3	9	10	1	9.4	3	0.531	191615	4	9	-4.67
3	10	15	1	9.4	3	0.531	191615	6	10	0.33
3	11	7	1	9.4	3	0.531	191615	8	11	-7.67
3	12	15	1	9.4	3	0.531	191615	10	12	0.33
4	1	12	1	18	0	0.483	166702	-10	1	-4
4	2	22	1	18	0	0.483	166702	-8	2	6

Appendix 8.5: Stata simulation programme

```
/* Simulation programme for the results presented in Chapter 5*/

version 11.2
set more off
quietly log
local logon = r(status)
if "`logon'" == "on" {
log close
      }
*log using "D:\PhD sim\data\dtb_log12.log", text replace

/*****************************************************************/
/* File Name:    mc_pois.do
/* Date      :    August 21, 2012
/* Author    :    George Bello
/* Purpose   :    Monte Carlo simulation for multi-level data.
/* Output sim Data:        Timeeffects1-4
/* Independent vars Data: Timeeffects-indvars1-4
/* Output File:            pois.log
/*****************************************************************/

*****************************************************************/
* parameters:
* tclust : number of treatment clusters
* cclust : number of control clusters
* time   : number of repeated measurements time points per cluster
* mu     : mean estimate of the outcome in the control group
* sdclust : sd of random effect at the cluster level
* sdtime  : sd of random effect at the repeated time level
*****************************************************************/


     /***************************************/
     /* Setting up multi-level data.        */
     /***************************************/

capture program drop _all

capture program define mc_pois, rclass
version 11.2

syntax [, tclust(real 0) cclust(real 0) time(real 0)]

* internal calculations
local nclust = `tclust'+`cclust'      /* total number of clusters */
local time   = `time'                 /* repeated measurements/observations per cluster */
local b0     = log(15)                /* mean intercept of the outcome in the control arm
                                         (15 and 70 for low and high incidence diseaes)*/
local beta   = log(22.5)               /* effect size in treatment arm  0.1=16.5, 0.2=18,
                                          0.3=19.5, 0.4=21, 0.5=22.5, 0.6=24, 0.8=27 */
local matsize = 30

set mat `matsize'
```

```
/* create a time level dataset and time-level random effects. */

clear

set  obs `time'              /*generate data set according to number of repeated time points */

generate temp1= _n

drawnorm u1, mean(0) sd(0.5)  /*generate repeated meaurements random effects(u1)using ICC from Triage Plus*/

expandcl `nclust', generate(temp2) cluster(temp1)

generate n= _n

generate cluster=temp2-((temp1-1)*`nclust')

generate time=temp1

/* Generate the cluster-level random effects. */

drawnorm u2, mean(0) sd(0.005)  /*generate cluster level random effects (u1)using ICC from Triage Plus*/

drop temp1 temp2

sort     cluster time

replace  n= _n

save "D:\PhD sim\data\time12.dta",replace  /*generated data saved acording to number of
                                             repeated time points 2-12*/
```

```
/* create fixed independent variables. */

drop n cluster time u2

clear

set  obs  `nclust'

gen n=_n

generate trt=0

recode   trt (0=1) if n>((`nclust')/2)          /*generate intervention variable with equal
                                                distribution, 0=control, 1=intervention cluster*/

gen temp1=_n

drawnorm u2, mean(0) sd(0.005)                  /*generate cluster level random effects (u1)
                                                  using ICC from Triage Plus*/

expandcl `time', generate(newcl) cluster(temp1)

drop newcl

replace n=_n

gen time=n-((temp1-1)*`time')

gen cluster = temp1

drop temp1


/*now merge 1:1 the two data sets using n time cluster for timeeffects created above*/

merge 1:1 n time cluster using "D:\PhD sim\data\time12.dta"

drop _merge

generate mu=exp(`b0' + u1 + u2)
replace  mu=exp(`beta' + u1 + u2) if trt==1

generat count= rpoisson(mu)

save "D:\PhD sim\data\time12.dta",replace
```

```
xtmepoisson count trt if time<=`time'|| cluster:, cov(ex) iterate(100)

        scalar b_trt   = _b[trt]

        scalar se_trt  = _se[trt]

end

             /*************************/
             /* Run the monte carlo. */
             /*************************/
clear

save "D:\PhD sim\data\sim12.dta",replace emptyok

      foreach time of numlist 2 3 4 6 12 {
           forvalues effect = 1.1(0.1)1.8 {

                simulate b_trt=b_trt se_trt=se_trt  ///
                beta=(`effect') time=(`time') converged=e(converged), ///
                reps(1000) seed(3321): mc_pois, tclust(12) cclust(12) time(`time')

                     append using "D:\PhD sim\data\sim12.dta"

                     save "D:\PhD sim\data\sim12.dta", replace

                }
}



      /*************************************************/
      /* Testing that simulation programme is well set */
      /*under null treatment effect                    */
      /*************************************************/

gen zrat= b_trt/se_trt              /*creating Z ratio to calculate statistical power */

gen pv=2*(1-normprob(abs(zrat)))              /* the pvalue*/

ksmirnov pv=pv

hist pv, xlab="p-value" main="Fig: Distribution of 2-sided p values from 1,000 simulations under null
                                treatment effect"

exit
```

```
      /*********************************/
      /* Interpret the results.       */
      /*********************************/

*calculation of IRR with 95% CI

drop irr lci uci
gen irr = exp(b_trt)
gen lci = exp(b_trt - 2*se_trt)
gen uci = exp(b_trt + 2*se_trt)

bysort beta time:sum irr,detail

drop selength
gen selength = abs(uci-lci)
bysort beta time:sum selength

*bias calculation

drop bias propbias
gen bias  =  irr-beta
gen propbias = bias/beta
bysort beta time:sum bias propbias

*calculation of stndardised bias

gen setrt=exp(se_trt)
gen stdbias = bias/setrt

*accuracy calculation based on mean square error
gen mse = (bias)^2 + (setrt)^2




*calculation of coverage % of times the 95% ie 100(1-alpha)% CI _b[trt]+-1.96*(_se[trt]) include `b1'.

gen cov95 = 2*1.96*(2*se_trt)        /*divide by 1000 simulations*/

/*Calculation of non-coverage of SE.regress the parameter & note the 95% CI and put in formula next*/

bysort beta time:sum lci uci
drop cov95t
gen cov95=0 if  irr>=1.7346 & irr<=1.870862
recode cov95 (0=0) (.=1),gen (cov95t)
drop cov95
by beta time:tab cov95t if  irr!=.

*power calculation
drop p
gen p = 2*normal(-abs(b_trt/se_trt))       /* the pvalue*/
bysort beta time:count if p<0.05

sum irr lci uci bias probias stdbias mse cov95

save "D:\data\sim1_2.dta", replace
log close
```

Appendix 8.6: Stata data processing and analysis programme

```
Version 11.2

set more off
  quietly log
  local logon = r(status)
  if "`logon'" == "on" {
    log close
    }
log using results_tb, text append //

clear

cd "D:\PhD\Data phd\cleaned\"

/*  ****************************************************************/
/*  File Name:      results.do                                     */
/*  Date:           October 20, 2012                               */
/*  Author:         George Bello                                   */
/*  Purpose:        Data cleaning and analysis programme           */
/*  Covariates:     Treat, base, occasion, gender,treat*occasion   */
/*  Output File:    results_tb.log                                 */
/*  ****************************************************************/

/*************************************************************************************************/
 /* Importing excel based data into stata for data cleaning to remove duplicates: similar approach was done all datasets*/
 /* Initial data cleaning done in excel */

 /* TB treatment and testing data*/
 insheet using "D:\PhD\Data phd\cleaned\dtb\dtb_all.csv",clear  //Total TB cases starting treatment
 duplicates  list  ta  nameofpatient village tbno sex age year
 duplicates  drop  ta  nameofpatient village tbno sex age year, force
 save "D:\PhD\Data phd\cleaned\dtb_all_clean",replace
 /*************************************************************************************************/
```

```
/* Summarising the individual based data to aggregated data by cluster and occasion(time point)
  and then add cluster level covariates manually in excel*/

use "D:\PhD\Data phd\cleaned\dtb_all_clean",clear

gen y = 1
keep y age sex clustid occasion treat
collapse (count) y (mean)sex age, by (clustid occasion)
table clustid occasion

bysort cluster:tab sex
/*Enter cluster level covariates manually in excel and when the data is in wide form in csv */

insheet using "D:\PhD\Data phd\cleaned\dtb\dtb_all_clean.csv",clear

*reshaping wide data form to long
reshape long count1, i(clustid) j(occasion)

save "D:\PhD\Data phd\cleaned\dtb_all_clean",replace

/* Exploring data*/
 xtset clustid occasion

/*To explore the participation patterns for the entire period of data collection */
xtdescribe, i(clustid) t(occasion)
table t,contents(count y) col*

bysort treat: summarize y,detail
bysort treat: summarize clpop
bysort treat: summarize ptfemale
bysort treat: summarize base
bysort treat: summarize hf

/* Note variances and means. are variances greater than mean, hence over-dispersion*/

tabstat y, by(treat) stats(mean v n)
tabstat count1, by(clustid) stats(mean v n)
tabstat count1, by(occasion) stats(mean v n)

* To investigate the within cluster, btn cluster and total variability of the covariates */

xtsum y treat occasion, i(cluster)
```

```
/*******************************************************/

/* DATA ANALYSIS*/

/* Analysis of data for the first 12 months */

/*Marginal effects modelling*/
//Used vce to produce standard errors based on the sandwich estimator to take clustering into account//

xtset clustid time
glm y treat if occasion<=12, family(Poisson) offset(lpop) link(log) eform vce(cluster clustid)robust // only intervention variable included
glm y treat base if occasion<=12, family(Poisson) offset(lpop) link(log) eform vce(cluster clustid)robust // baseline added
glm y treat base ptfemale  hf if occasion<=12, family(Poisson) offset(lpop) link(log) eform vce(cluster clustid)robust // gender & sites added
glm y treat base occasion ptfemale  trt_occasion hf if occasion<=12, family(Poisson) offset(lpop) link(log) eform vce(cluster clustid)robust //all
estimates store marginal1

/* Random intercepts modelling to account for dependences of events detected over time, within and between clusters
  and assumes that the effect of intervention is the same across clusters */

xtmepoisson y treat if occasion<=12|| clustid:,  covariance(exchangeable) irr offset(lpop) // only intervention variable included
xtmepoisson y treat base if occasion<=12|| clustid:,  covariance(exchangeable) irr offset(lpop) // baseline added
xtmepoisson y treat base ptfemale hf if occasion<=12|| clustid:,  covariance(exchangeable) irr offset(lpop) // gender & sites added
xtmepoisson y treat base occasion ptfemale trt_occasion hf if occasion<=12|| clustid:,  covariance(exchangeable) irr offset(lpop) //all
estimates store ri1

/* Model selection using likelihood ratio test (pvalue divided by 2: marginal vs random intercepts  */

lrtest marginal1 ri1, force    //use force since the two models fitted using different commands (i.e. glm and xtmepoisson)

/* Random coefficients modelling to account for dependences of events detected over time, within and between clusters
  and allows the effect of the intervention to vary across clusters */

xtmepoisson y treat if occasion<=12|| clustid: treat,  covariance(exchangeable) irr offset(lpop) // only intervention variable included
xtmepoisson y treat base if occasion<=12|| clustid: treat,  covariance(exchangeable) irr offset(lpop) // baseline added
xtmepoisson y treat base ptfemale hf if occasion<=12|| clustid: treat,  covariance(exchangeable) irr offset(lpop) // gender & sites added
xtmepoisson y treat base occasion ptfemale trt_occasion hf if occasion<=12|| clustid: treat,  covariance(exchangeable) irr offset(lpop) //all
estimates store rc1


/* Model selection using likelihood ratio test to assess whether rc moddel is better than ri pvalues divided by 2 */
lrtest ri1 rc1

/* Calculating model fit indices for marginal and random effects models(ll, AIC and BIC) */
estimates table marginal1 ri1 rc1, b(%9.3f) se(%9.3f) p(%9.3f) eform style(columns) stats(ll aic bic) title(TB treatment initiations)
```

```
/* Analysis of data for the next 11 months */

/*Marginal effects modelling*/

glm y treat if occasion>12, family(Poisson) offset(lpop) link(log) eform vce(cluster clustid)robust // only intervention variable included
glm y treat base if occasion>12, family(Poisson) offset(lpop) link(log) eform vce(cluster clustid)robust // baseline added
glm y treat base ptfemale  hf if occasion>12, family(Poisson) offset(lpop) link(log) eform vce(cluster clustid)robust //gender &sites added
glm y treat base occasion ptfemale  trt_occasion hf if occasion>12, family(Poisson) offset(lpop) link(log) eform vce(cluster clustid)robust//all
estimates store marginal2

/* Random intercepts modelling */

xtmepoisson y treat if occasion>12|| clustid:,  covariance(exchangeable) irr offset(lpop) // only intervention variable included
xtmepoisson y treat base if occasion>12|| clustid:,  covariance(exchangeable) irr offset(lpop) //baseline added
xtmepoisson y treat base ptfemale hf if occasion>12|| clustid:,  covariance(exchangeable) irr offset(lpop) //gender & sites added
xtmepoisson y treat base occasion ptfemale trt_occasion hf if occasion>12|| clustid:,  covariance(exchangeable) irr offset(lpop)//all
estimates store ri2

/* Model selection using likelihood ratio test: marginal vs random intercepts  */

lrtest marginal2 ri2, force

/* Random coefficients modelling */

xtmepoisson y treat if occasion>12|| clustid: treat,  covariance(exchangeable) irr offset(lpop) // only intervention variable included
xtmepoisson y treat base if occasion>12|| clustid: treat,  covariance(exchangeable) irr offset(lpop) // baseline added
xtmepoisson y treat base ptfemale hf if occasion>12|| clustid: treat,  covariance(exchangeable) irr offset(lpop)//gender & sites added
xtmepoisson y treat base occasion ptfemale trt_occasion hf if occasion>12||clustid: treat, covariance(exchangeable)irr offset(lpop)//all
estimates store rc2

/* Model selection using likelihood ratio test: random intercepts vs Random coefficients */
lrtest ri2 rc2

/* Calculating model fit indices for marginal and random effects models(ll, AIC and BIC) */
estimates table marginal2 ri2 rc2, b(%9.3f) se(%9.3f) p(%9.3f) eform style(columns) stats(ll aic bic) title(TB treatment initiations)

log close results_tb
```

```
 // Calculating intracluster correlation (ICC)
*********************************************************************************************************************
//TB treatment
glm count1 treat base occasion ptfemale  trt_occasion bighf if occasion<=12, family(Poisson)  link(log) eform
xtmepoisson count1 treat base occasion ptfemale trt_occasion sites if occasion<=12|| clustid:, covariance(uns) irr offset(lpop)
di sqrt(e(chi2)/e(df_m))
predict xb,xb
scalar xb1=exp(xb+.112^2/2)  // mean count (intercept variance=0.112)
di .112^2/(.112^2+1.865^2*ln(1/exp(xb)+1))   // ICC on log scale (ICC=0.067)
di xb1*(exp(.112^2) - 1)/(xb1*(exp(.112^2) - 1)+ 1.865^2) // ICC on original scale  (ICC=0.066)
*********************************************************************************************************************
//ART treatment initiations
glm count1 treat base occasion ptfemale  trt_occasion sites if occasion<=12, family(Poisson)  link(log) eform
xtmepoisson count1 treat base occasion ptfemale trt_occasion bighf if occasion<=12|| clustid:,  covariance(uns) irr offset(lpop)
di sqrt(e(chi2)/e(df_m))
predict xb,xb
scalar xb1=exp(xb+.056^2/2)  // mean count (intercept var=0.056)
di .056^2/(.056^2+1.867^2*ln(1/exp(xb)+1))   // ICC on log scale (ICC=0.04022)
di xb1*(exp(.056^2) - 1)/(xb1*(exp(.056^2) - 1)+ 1.867^2) // ICC on original scale  (ICC=0.039992)
*********************************************************************************************************************
//HIV testing
xtmepoisson count1 treat base occasion ptfemale trt_occasion sites if occasion<=12|| clustid:,  covariance(uns) irr offset(lpop)
di sqrt(e(deviance-p)/e(df))
predict xb,xb
scalar xb1=exp(xb+.0987^2/2)  // mean count (intercept variance=0.0987)
di .0987^2/(.0987^2+10.739^2*ln(1/exp(xb)+1))   // ICC on log scale (ICC=0.1287)
di xb1*(exp(.0987^2) - 1)/(xb1*(exp(.0987^2) - 1)+ 10.739^2) // ICC on original scale  (ICC=0.1297)
 ///
```

Appendix 8.7: Abstract presented at the 2013 Union Conference

ter understand the populations who accessed these services and the benefits of each.

Methods: A model was developed where nine community HCT stand-alone centers were established and maintained in partnership with non-governmental organizations from 2008 to 2012, in high TB-HIV burden communities around Cape Town. Services were provided from either the stand-alone centre or on a mobile 'outreach' basis. Mobile services were provided from tents and caravans that were set up at busy community thoroughfares and transport hubs. All clients were tested for HIV and screened for TB according to national guidelines. Routine data was collected in HCT registers and entered into a database. The chi-square test was used to determine if either the stand-alone or mobile service was associated with any of the HIV or TB indicators collected in the routine data.

Results: 96 137 clients accessed these services. The mobile services accessed a greater proportion of clients (80%) than the stand-alone service (20%) and had a greater proportion of male clients (50%) compared to the stand-alone centers (45%). The stand-alone centers diagnosed a significantly higher proportion of clients with HIV (13%, $P < 0.001$) and a significantly higher proportion of co-infected clients (16%, $P < 0.001$) than the mobile services. The stand-alone centers were also better able to link HIV-infected clients to care ($P < 0.001$) than the mobile services.

Table HIV and TB indicators from routine data collected for clients who accessed either a stand-alone or mobile service in the community

Indicator	Stand alone *n* (%)	Mobile *n* (%)	Total *n* (%)
Total clients	19 398	76 739	96 137
Males	8744 (45)	38 188 (50)	46 932 (49)
Tested	18 087 (93)	72 164 (94)	90 251 (94)
HIV diagnosed	2 331 (13)	4 148 (6)	6 918 (7)
HIV+ve referred	2 140 (92)	4 148 (90)	6 288 (91)
HIV+ve linked to care	1 574 (74)	2 825 (68)	4 399 (70)
Total TB symptoms (100% screened)	2 169 (11)	4 405 (6)	6 574 (7)
HIV+ve with TB symptoms	470 (22)	535 (12)	1 005 (15)
HIV+ve TB tests	388 (82)	387 (72)	775 (77)
HIV+ve TB diagnosed	63 (16)	29 (7)	92 (12)
HIV+ve commenced TB Rx	59 (94)	28 (97)	87 (95)

Conclusion: This model was able to successfully integrate HIV and TB testing for stand-alone and mobile services. The mobile service was able to access a

OP-128-01 Engaging informal health care providers in case detection for tuberculosis and HIV in rural Malawi

G Bello,[1] B Faragher,[2] L Sanudi,[1] I Namakhoma,[1] H Banda,[1] R Malmborg,[3] R Thomson,[2] S B Squire.[2]
[1]REACH Trust, Research for Equity and Community Health (REACH) Trust, Lilongwe, Malawi; [2]CRESTHA, Liverpool School of Tropical Medicine, Liverpool, UK; [3]International Cooperation, HL, Norwegian Heart and Lung Patients Association, Oslo, Norway. e-mail: r.thomson@liv.ac.uk

Introduction: The poor and vulnerable face structural barriers in accessing TB and HIV/AIDS services. We implemented a randomised, controlled health system intervention trial involving informal health care providers (IHP): 'Triage Plus' in rural Malawi. We aimed to determine the overall effect on: a) diagnostic uptake for TB and HIV and b) TB and Anti Retroviral Therapy (ART) treatment initiation.

Methods: The intervention consisted of training non-paid IHP (such as store-keepers) in TB and HIV disease recognition, sputum specimen collection, referral to the public health system, and raising community awareness. Front line public health personnel and community leaders were sensitised to support the intervention. A phased, matched, parallel cluster design was used with a 1:1 randomisation of six clusters (average total population size per cluster = 205 139) to an early intervention arm (received the intervention early in the first 12 months) and a delayed intervention arm (received the intervention after one year). Repeated measurement Poisson models were used for evaluation of incidence rate ratios between the two arms over 23 months. Development, set up, and recurrent cost data were collected and adjusted for inflation.

Preliminary results: The number of presumptive TB cases accessing diagnostic sites was 15.2% higher in the early arm ($P = 0.003$) after the first 12 months, but there was no difference between the two arms after the next 11 months (i.e., after the intervention was rolled out to the delayed arm). The number of TB cases starting treatment was non-significantly higher (18%) in the early arm ($P = 0.112$) after 12 months, but 45% lower ($P < 0.001$) after the next 11 months. HIV testing uptake was 61% higher in the early arm ($P < 0.001$) after the first 12 months but there was no difference between the two arms after the next 11 months. ART initiation was 34.7% higher ($P = 0.048$) in the early arm after the first 12 months and again there was no difference between

the arms after the next 11 months. The incremental cost-effectiveness of the intervention was $16 per additional HIV test and $600 per additional new case starting ART. As duration and coverage increase, the fixed cost proportion will reduce, improving cost-effectiveness.

Conclusions: Engagement of IHP in integrated TB and HIV services at community level was associated with a substantial increase in services access and HIV treatment initiation. There may be several explanations for the decrease in TB treatment initiation over time.

OP-129-01 Intensified tuberculosis case finding in PMTCT settings in Nigeria

M Odo,[1] T Idaboh,[1] T Odusote.[2] [1]Prevention, Care and Treatment, FHI360, Abuja, [2]TB, United States Agency for International Development, Abuja, Nigeria

low yield intervention and may not be cost effective. It however provides an opportunity for TB awareness and education.

OP-130-01 Experience with the Xpert® MTB/RIF assay in routine programme conditions with different HIV prevalence and risk of MDR-tuberculosis

A-L Page,[1] E Ardizzoni,[2,3] M Lassovsky,[4] B Kirubi,[5] D Bichkova,[6] M Bonnet,[1] A Pedrotta,[7] F Varaine.[2] [1]Clinical Research, Epicentre, Paris, [2]Operational Center Paris, Médecins Sans Frontières (MSF), Paris, France; [3]Mycobacteriology Unit, Institute of Tropical Medicine, Antwerp, Belgium; [4]Operational Center Geneva, MSF, Nhlangano, Swaziland; [5]Operational Center Paris, MSF, Nairobi, Kenya; [6]Operational Center Paris, MSF, Sukhumi, Abkhazia, Georgia; [7]Operational Center Paris, MSF, Kampong Cham, Cambodia. e-mail: anne-laure.page@epicentre.msf.org

9.0 References

Akaike H, 1981. Likelihood of a model and information criteria. *Journal of Econometrics*, 16, 3 - 14.

Allen JD, Stoddard AM, Mays J, and Sorensen G, 2001. Promoting breast and cervical cancer screening at the workplace: results from the Woman to Woman Study. *Am J Public Health,* 91, 584–590.

Anglemyer A, Rutherford GW, Baggaley RC, Egger M, *and* Siegfried N, *2011.* Antiretroviral therapy for prevention of HIV transmission in HIV-discordant couples. Cochrane Database Syst Rev., 8, CD009153. doi: 10.1002/14651858.CD009153.pub2.

Arnold BF, Hogan DR, Colford jr JM, and Hubbard AE, 2011. Simulation methods to estimate design power: an overview for applied research [online]. *BMC Medical Research Methodology*, 11:94. Available at Http://www.biomedcentral.com/1471-2288/11/94.

Atienza AA, and King AC, 2002. Community-based Health Intervention Trials: an Overview of Methodological Issues. *Epidemiol Rev.*, 24 (1), 72-79.

Austin PC, 2007. A comparison of the statistical power of different methods for the analysis of cluster randomisation trials with binary outcomes. *Statist. Med.*, 26, 3550-3565.

Aveyard P, Sherratt E, Almond J, Lawrence T, Lancashire R, Griffin C, and Cheng KK, 2001. The change-in-stage and updated smoking status results from a cluster randomized trial of smoking prevention and cessation using the trans-theoretical model among British adolescents. *Prev Med.,* 33, 313–324.

Axelson O, Fredricksson M, and Ekberg K, 1994. Use of the prevalence ratio v the prevalence odds ratio as a measure of risk in cross sectional studies. *Occup Environ Med.*, 51, 574.

Ayles H, Schaap A, Nota A, Sismanidis C, Tembwe R, De Haas P, Muyoyeta M, and Beyers N, 2009. Prevalence of tuberculosis, HIV and respiratory symptoms in two Zambian communities: implications for tuberculosis control in the era of HIV. *PLoS ONE*, 4(5), e5602. doi: 10.1371/journal.pone.0005602.

Ayles H, Muyoyeta M, Du Toit E, Schaap A, Floyd S, Simwinga M, Shanaube K, Chishinga N, Bond V, et al.,2013. Effect of household and community interventions on the burden of tuberculosis in southern Africa: the ZAMSTAR community-randomised trial. *Lancet* , 382: 1183–94.

Bandura A, 1971. *Social Learning Theory*. Morristown, NJ: General Learning Corporation.

Bandura A, 1986. *Social foundations of thought and action: a social-cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Barros AJ, and Hirakata VN, 2003. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology,* 3:21. Available at *http://www.biomedcentral.com/1471-2288/3/21*.

Barter DM, Agboola SO, Murray MB, and Bärnighausen T, 2012. Tuberculosis and poverty: the contribution of patient costs in sub-Saharan Africa – a systematic review. *BMC Public Health*,12, 980. doi: *10.1186/1471-2458-12-980*.

Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, and Fernandez ME, 2006. *Planning health promotion programs: an Intervention Mapping approach.* San Francisco, CA: Jossey-Bass.

Bartholomew LK, Parcel GS, Kok G, Gottlieb NH, and Fernandez ME, 2011. *Planning health promotion programs: an Intervention Mapping approach.* San Francisco, CA: Jossey-Bass.

Beckwith CG, Flanigan TP, del Rio C, Simmons E, Wing EJ, Carpenter, CCJ, and Bartlett JG, 2005. It is time to implement routine, not risk-based, HIV testing. *Clin Infect Dis*., 40(7),1037-1040.

Bello G, Simwaka B, Ndhlovu T, Salaniponi F, and Hallet TB, 2011. Evidence for changes in behaviour leading to reductions in HIV prevalence in urban Malawi. *Sex Transm Infect;* 87(4), 296-300.

Bennett S, Parpia T, Hayes R, and Cousens S, 2002. Methods for the analysis of incidence rates in cluster randomised trials. *International Journal of epidemiology*, 31, 839-846.

Benson T, Kaphuka J, Kanyanda S, and Chinula R, 2002. *Malawi : an atlas of social statistics,* Zomba, Malawi.

Berkhof J, and Snijders TAB, 2001. Variance component testing in multilevel models. *Journal of education and behavioural statistics*, 26(2), 133-152.

Best N, Mason A, and Li P, 2011. *Bayesian Hierarchical Modelling using WinBUGS*. *Short Course, Feb 17–18* [online]. Imperial College. Available at: http://www.bias-project.org.uk.

Blizzard L, and Hosmer DW, 2006. Parameter estimation and goodness-of-fit in log binomial regression. *Biometrical J.*, 48, 5–22.

Borgdorff MW, Floyd K, and Broekmans JF, 2002. Interventions to reduce tuberculosis mortality and transmission in low- and middle-income countries. *Bulletin of the World Health Organization*, 80 (3), 217-227.

Borm GF, Melis RJF, Teerenstra S, and Peer PG, 2005. Pseudo cluster randomization: a treatment allocation method to minimize contamination and selection bias. *Statist. Med.*, 24, 3535–3547.

Brechtl JR, Breitbart W, Galietta M, Krivo S, and Rosenfeld B, 2001. The use of Highly Active Antiretroviral Therapy (HAART) in patients with advanced HIV-infection: Impact on medical, palliative care, and quality of life outcomes. *Journal of Pain and Symptom Management, 21,* 41-51.

Breslow NE, and Clayton DG, 1993. Approximate inference in generalised linear mixed models. *Journal of the American Statistical Association*, 88 (421), 9-25.

Breslow NE, and Lin X, 1995. Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika,*82, 81-91.

Breslow NE, 1996. Generalised linear models: checking assumptions and strengthening conclusions [online]. Prepared for Congresso Nazionale Societa' Italiana di Biometrica Centro Convegni S.Agostino, Cortona, 16-17 June, 1995.

Broekmans J, 1994. Control strategies and programme management. In: Porter JDH, McAdam KPWJ, editors. Tuberculosis— back to the future. Chichester: John Wiley & Sons.

Brooks SP, and Gelman A, 1998. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics,* 7(4), 434–455.

Brown CH, Wyman PA, Guo J, and Peňa J, 2006. Dynamic wait-listed designs for randomised trials: new designs for prevention of youth suicide. *Clinical Trials*, 3, 259-271.

Burton A, Altman DG, Royston P, and Holder RL, 2006. The design of simulation studies in medical statistics. *Statist. Med.*, 25, 4279-4292.

Cai Z, Fan J, and Li R, 2000. Efficient Estimation and Inferences for Varying-coefficient Models. *Journal of the American Statistical Association,* 95, 888–902.

Campbell MK, Elbourne DR, and Altman DG, 2004. **CONSORT statement: extension to cluster randomised trials.** *BMJ,* 328, 702-708.

Campbell CH Jr, Marum ME, Alwano-Edyegu M, Dillon BA, Moore M, and Gumisiriza E, 1997. The role of HIV counselling and testing in the developing world. *AIDS Educ Prev.*, 9(Suppl B), 92-104.

Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, and Tyrer P, 2000. Framework for design and evaluation of complex interventions to improve health. *BMJ, 321:694.*

Carlin BP, and Chib S, 1995. Bayesian model choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B*, 57, 473 - 484.

Castilla J, Sobrino P, De La Fuente L, Noguer I, Guerra L, Parras F, 2002. Late diagnosis of HIV infection in the era of highly active antiretroviral therapy: consequences for AIDS incidence. *AIDS*, 16(14), 1945-1951.

Central Statistical Agency [Ethiopia] and ICF International, 2012. Ethiopia Demographic and Health Survey 2011. Addis Ababa, Ethiopia and Calverton, Maryland, USA: Central Statistical Agency and ICF International.

Chavasse DC, Shier RP, Murphy OA, Huttly SRA, Cousens SN, and Akhtar T, 1999. Impact of fly control on child diarrhoea in Pakistan: community-randomised trial. *Lancet*, 353, 22 – 25.

Chiang CY, Chang CT, Chang RE, Li CT, and Huang RM, 2005. Patient and health system delays in the diagnosis and treatment of tuberculosis in Southern Taiwan. Int J Tuberc Lung Dis., 9(9), 1006-1012.

Chin HC and Quddus MA, 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis and Prevention,* 35, 253–259.

Clark AB and Bachmann MO, 2009. Bayesian methods of analysis for cluster randomized trials with count outcome data. *Statist. Med.* 29, 199-209.

Clark DR, and Thayer CA, 2004. A Primer on the Exponential Family of Distributions on generalised linear models. *Discussion Paper Program Casualty Actuarial Society - Arlington, Virginia,* 2004, 117-148.

Cohen J, 1992. A power primer. Psychol Bull., 112(1), 155-159.

Cohen C, Revicki DA, Nabulsi A, Sarocco PW, and Jiang P, 1998. A randomized trial of the effect of ritonavir in maintaining quality of life in advanced HIV disease. *AIDS, 12,* 1495-1502.

Cohen MS, and Gay CL, 2010. Treatment to prevent transmission of HIV-1. *Clin Infect Dis.*, 50 **(Supplement 3),** S85-S95.

Cohen MS, Chen YQ, McCauley M, et al, 2011. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med.,* 365:493-505.

Collins LM, Schafer JL, and Kam C, 2001. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4), 330-351.

Colombo M, Mosso C, and De Piccoli N, 2001. Sense of community and participation in urban contexts. *J. Commun. Appl. Soc. Psychol.,* 11, 457–464.

Cook GS, Frank C Tanser FC, Bärnighausen TW, and Newell M, 2010. Population uptake of antiretroviral treatment through primary care in rural South Africa [online]. BMC Public Health, 10, 585 doi:10.1186/1471-2458-10-585.

Corbett EL, Bandason T, Duong T, Dauya E, Makamure B, Churchyard GJ, Williams BG, Munyati SS, Butterworth AE, Mason PR, Mungofa S, and Hayes RJ, 2010. Comparison of two active case-finding strategies for community-based diagnosis of symptomatic smear-positive tuberculosis and control of infectious tuberculosis in Harare, Zimbabwe (DETECTB): a cluster randomised trial. *Lancet*, 376, 1244-1253.

Corbett EL, Churchyard GJ, Charambos S, Samb S, Moloi V, Clayton TC, Grant AD, Murray J, Hayes RJ, and De Cock KM, 2002. Morbidity and mortality in South African gold miners: impact of untreated disease due to human immunodeficiency virus. *Clin Infect Dis*, 34(9), 1251-1258.

Corbett EL, Marston B, Churchyard GJ, and De Cock KM, 2006. Tuberculosis in sub-Saharan Africa: opportunities, challenges and change in the era of antiretroviral treatment. *Lancet,* 367,926–37.

Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, and Dye C, 2003. The Growing Burden of Tuberculosis: global trends and interactions with the HIV epidemic. *Archives of Internal Medicine,*163, 1009–1021.

Corkery E, Palmer C, Foley M E, Schechter C B, Frisher L, and Roman S H, 2007. Effect of a bicultural community health worker on completion of diabetes education in a Hispanic population. *Diabet Care*, 20, 254–257.

Cornfield J, 1978. Randomization by group: a formal analysis. *Am J Epidemiol.*108, 100–102.

Coutinho LMS, Scazufca M, and Menezes PR, 2008. Methods for estimating prevalence ratios in cross sectional studies. *Rev Saúde Pública*, 2008;42(6), 992-998.

Creek TL, Ntumy R, Seipone K, Smith M, Mogodi M, Smit m, Legwaila K, Molokwane I, Tebele G, Mazhani L, Shaffer N, and Kilmarx PH, 2007. Successful Introduction of Routine Opt-Out HIV Testing in Antenatal Care in Botswana. *J. Acquir Immune Defic Syndr.*, 45:102–107.

Cummings KM, Hyland A, Saunders-Martin T, Perla J, Coppola PR, and Pechacek TF, 1998. Evaluation of an enforcement program to reduce tobacco sales to minors. *Am J Public Health,* 88, 932–935.

Curtale F, Siwakoti B, Lagrosa C, Laraja M, and Guerra R, 1995. Improving skills and utilisation of community health volunteersin Nepal. Social Science Medicine, 40 (8), 1117-1125.

Datiko DG, and Lindtjørn B, 2009. Health Extension Workers Improve Tuberculosis Case Detection and Treatment Success in Southern Ethiopia: a community randomized trial. *PLoS ONE*, 4(5), e5443. doi:10.1371/journal.pone.0005443.

Davies R, Hedberg GA, and Fischer M, 1948. A complete community survey for tuberculosis: a second report on effectiveness of the procedure as a method of tuberculosis control. *Am Rev Tuberc.*, 58(1), 77-84.

Dean CB, Ugarte MD, and Militino AF, 2004. Penalized quasi-likelihood with spatially correlated data. *Computation Statistics and Data Analysis*, , 45, 235-248.

Declaration of Alma-Ata, International Conference on Primary Health Care, Alma-Ata, USSR, 6-12 September 1978.

De Allegri M, Pokhrel S, Becher H, Dong H, Mansmann U, Kouyaté B, Kynast-Wolf G, Gbagou A, Sanon M, Bridges J, and Sauerborn R, 2008. Step-wedge cluster-randomised community-based trial: an application of the study of the impact of community health insurance. *Health Research Policy and Systems*, 6, 10-17.

De Cock KM, and Chaisson RE, 1999. Will DOTS do it? A reappraisal of tuberculosis control in countries with high rates of HIV infection. Int J Tuberc Lung Dis., 3(6), 457- 465.

Deddens JA, and Petersen MR, 2004. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am. J. Epidemiol.*, 159, 213–214.

Deddens JA, and Petersen MR, 2008. Approaches for estimating prevalence ratios. *Occup. Environ. Med.* 65, 501-506.

Deddens JA, Petersen MR, and Lei X, 2003. Estimation of prevalence ratios when PROC GENMOD does not converge. In: Proceedings of the 28th Annual SAS Users Group International Conference, March 30–April 2, 2003. Paper 270-28. Cary, NC: SAS Institute Inc. Available at *http://www2.sas.com/proceedings/sugi28/270-28*. pdf.

Diciccio TJ, Kass RE, Raftery A, and Wasserman L, 1997. Computing Bayes factors by combining simulations and asymptotic approximation. *Journal of the Statistical Association*, 92(439), 903 - 915.

Diehr P, Martin DC, Koepsell T, and Cheadle A, 1995. Breaking the matches in a paired t-test for community interventions when the number of pairs is small.*Statist. Med.*, 14(13), 1491-1504.

Donner A, 1999. Some aspects of the design and analysis of cluster randomization trials. *Applied Statistics,* 47, 95–113.

Donner A, Brown KS, and Brahser P, 1990. A methodological review of non-therapeutic intervention trials employing cluster randomisation, 1979-1989. *International Journal of Epidemiology*, 19, 795-800.

Donner A, Eliasziw M, and Klar N, 1994. A comparison of methods for testing homogeneity of proportions in teratological studies. *Statist. Med.*, 13, 1253-1264.

Donner A , and Klar N, 2000. *Design and Analysis of Cluster Randomization Trials in Health Research* .Oxford Press, New York.

Donner A, Klar N, 1996. Statistical considerations in the design and analysis of community intervention trials. *J Clin Epidemiol*, 49, 435–439.

Donner A, Taljaard M, and Klar N, 2007. The merits of breaking the matches: a cautionary tale. *Statist. Med.*, 26, 2036-2051.

Dunson D, 2001. Commentary: Practical advantages of Bayesian analysis in epidemiologic data. *American Journal of Epidemiology*, 153, 1222–1226.

Dye C, 2006. Global epidemiology of tuberculosis. *Lancet*, 367, 938-940.

Eaves L, and Erkanli A, 2003. Markov chain Monte Carlo approaches for the analysis of genetic and environmental components of human developmental change and G X E interaction. *Behavior Genetics,33,* 279-299.

Eubank RL, Huang C, Maldonado YM, Wang N, Wang S, and Buchanan RJ, 2004. Smoothing spline estimation in varying-coefficient models. *Journal of the Royal Statistical Society B,* 66, 653–667.

Evans M, and Swartz T, 1995. Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems. *Statistical Science*, 10 (3), 254-272.

Fabio LC, Paula GA, and de Castro M, 2012. A Poisson mixed model with non-normal random effect distribution. *Computational Statistics and Data Analysis*, 56, 1499–1510.

Faridi Z, Grunbaum JA, Gray BS, Franks A, and Simoes E, 2007. Community-based participatory research: necessary next steps. *Prev. Chronic Dis.*, 4, 1–5.

Feldman HA, and McKinlay SM, 1994. Cohort versus cross-sectional design in large field trials-precision, sample-size, and a unifying model. Stat. Med. 13, 61–78.

Feldman HA, McKinlay SM, and Niknian N, 1996. Batch sampling to improve power in a community trial: experience from the Pawtucket Heart Health Program. *Evaluation Review,* 20 (3), 244-274.

Feng Z, Diehr P, Peterson A, and McLerran D, 2001. Selected statistical issues in group randomized trials. *Annual Review of Public Health,* 22, 167–187.

Gambia Hepatitis Study Group. The Gambia Hepatitis intervention study. *Cancer Research*, 47, 5782-5787.

Gamerman D, 1998. Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika,* 85, 215–227.

Gebrekristos HT, Lurie MN, Mthethwa N, and Karim QA, 2009. Disclosure of HIV status: experiences of patients enrolled in an integrated TB and HAART pilot programme in South Africa. *African Journal of AIDS Research,* 8(1), 1–6.

Gelberg L, Andersen RM, and Leake BD, 2000. Healthcare access and utilization, the Behavioral Model for vulnerable populations: application to medical care use and outcomes for homeless people. *Health Services Research*, 34(6), 1273-1302.

Gelfand AE, and Smith AFM, 1990. Sampling-based approaches to calculating marginal densities. *J Amer Statistic Assoc.*, 85, 398–409.

Gelman A, 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515-533.

Gelman A, and Hill J, 2007. *Data analysis using regression and multilevel/hierarchical models-Analytical methods for social research*.1st ed. New York: Cambridge University Press.

Geman S, and Geman D, 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions On Pattern Analysis And Machine Intelligence, 6(6), 721-741.*

Gibbons RD, Segawa E, Karabatsos G, Amatya AK, Bhaumik DK, Brown CH, Kapur K, Marcus SM, Hur K, and Mann JJ, 2008. Mixed-effects Poisson regression analysis of adverse event reports: the relationship between antidepressants and suicide. *Statist Med.*, 27(11), 1814-1833.

Gilks WR, Thomas A, and Spiegelhalter DJ, 1994. A language and program for complex Bayesian modelling. *Statistician*, 43, 169 –177.

Gilks WR, and Wild P, 1992. Adaptive Rejection Sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.

Girardi E, Antonucci G, Vanacore P, Libanore M, Errante I, Matteelli A, and Ippolito G, 2000. Impact of combination antiretroviral therapy on the risk of tuberculosis among persons with HIV infection. *AIDS*, 14, 1985-1991.

Giuliano M, and Vella S, 2007. Inequalities in health: access to treatment for HIV/AIDS, 43(4), 313-316.

Glickman ME, and Dyk DA, 2007. Basic Bayesian Methods. *In Methods in Molecular Biology, vol. 404: Topics in Biostatistics,* W. T. Ambrosius (eds). Humana Press Inc., Totowa, NJ.

Glynn JR, Crampin AC, Ngwira BMM, et al, 2004. Trends in tuberculosis and the influence of HIV infection in northern Malawi, 1988-2001. *AIDS*, 18 (10), 1459-1463.

Glynn JR, Warndorff DK, Fine PE, Munthali MM, Sichone W, and Pönnighaus JM, 1998. Measurement and determinants of tuberculosis outcome in Karonga District, Malawi. *Bull World Health Organ.*, 76(3), 295–305.

Goldstein H, and Rasbash J, 1996. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society A,* 159, 505-513.

Golub JE, Mohan C, Comstock GW, and Chaisson RE, 2005. Active case finding of tuberculosis: historical perspective and future prospects. *Int J Tuberc Lung Dis.,* 9**,** 1183–1203.

Gopi PG, Subramani R, and Narayanan PR, 2006. Trend in the prevalence of TB infection and ARTI after implementation of DOTS programme in south India. *International Journal of Tuberculosis and Lung Disease*, 10, 346-348.

Graham SM, Holte SE, Peshu NM, et al, 2007. Initiation of antiretroviral therapy leads to a rapid decline in cervical and vaginal HIV-1 shedding. *AIDS,* 21, 501-507.

Greenland S, 2004. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol*, 160, 301–305.

Grosskurth H, Mosha F, Todd J, et al, 1995. Impact of improved treatment of sexually transmitted diseases on hiv infection in rural tanzania: randomised controlled trial. *Lancet,* 346, 530–536.

Gueorguieva R, and Krystal JH, 2004. Move Over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. *Archives of General Psychiatry*, 61, 310-317.

Gulliford MC, Ukoumunne OC, and Chinn S, 1999. Components of variance and intraclass correlations for the design of community-based surveys and intervention studies. *American Journal of Epidemiology,* 149 (9), 876-83.

Guo G, and Zhao H, 2000. Multilevel modelling for binary data. *Annual Review of Sociology*, 26, 441-462.

Hallet TB, Aberle-Grasse J, Bello G, et al, 2006. Declines in HIV prevalence can be associated with changing sexual behaviour in Uganda, urban Kenya, Zimbabwe, and urban Haiti. *Sex Transm Infect*, 82 (Suppl I):i1-i8.

Hannan PJ, Murray DM, Jacobs DR, and McGovern PG, 1994. Parameters to aid in the design and analysis of community trials: intraclass correlations from the Minnesota Heart Health Program. *Epidemiology,* 5(1), 88-95.

Harries AD, Hargreaves NJ, Chimzizi R, and Salaniponi FM, 2002. Highly active antiretroviral therapy and tuberculosis control in Africa: synergies and potential. *Bulletin of the World Health Organization*, 80, 464-469.

Harries AD, Maher D, and Nunn P, 1998. An approach to the problems of diagnosing and treating adult smear-negative pulmonary tuberculosis in high-HIV-prevalence settings in sub-Saharan Africa. *Bulletin of the World Health Organization*, 76, 651-662.

Hartzel J, Agresti A, and Caffo B, 2001. Multinomial logit random effects models [online]. *Statistical Modelling,* 1, 81: Sage. Available at http://smj.sagepub.com/content/1/2/81

Hastie T, and Tibshirani R, 1993. Varying-coefficient models. *Journal of the Royal statistical Society Series B*, 55, 757-796.

Hastings WK, 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.

Hausman J, Hall BH, and Griliches Z, 1984. Econometric models for count data with application to the patents -R & D relationship. *Econometrica, 52 (4), 909-938.*

Hayes RJ, Alexander NDE, Bennett S, and Cousens SN, 2000. Design and analysis issues in cluster randomised trials of interventions against infectious diseases. *Statistical Methods in Medical Research*, 9, 95-116.

Hayes R.J and Moulton L.H, 2009. *Cluster randomised trials*. 1st ed. Florida: Chapman & Hall.

Hedeker D, and Gibbons RD, 2006. *Longitudinal data analysis*. 1st ed. New York: Wiley.

Heo M, and Leon AC, 2005. Performance of a mixed effects logistic regression model for binary outcomes with Unequal cluster size. *journal of Biopharmaceutical Statistics*, 15, 513–526.

Heo M, and Leon AC, 2009. Sample size requirements to detect an intervention by time interaction in longitudinal cluster randomized clinical trials. *Stat Med.*, 28(6), 1017–1027.

Heo M, Xue X, and Kim MY, 2013. Sample size requirements to detect an intervention by time interaction in longitudinal cluster randomized clinical trials with random slopes. *Computational Statistics and Data Analysis*, 60, 169–178.

Hermans SM, Castelnuovo B, Katabira C, Mbidde P, Lange JMA, Hoepelman AIM, Coutinho A, and Manabe YC, 2012. Integration of HIV and TB services results in improved TB treatment outcomes and earlier prioritized art initiation in a large urban HIV clinic in Uganda. *Journal of Acquired Immune Deficiency Syndromes,* 60 (2), e29-e35.

Hoa NB, Sy DN, Nhung NV, Tiemersma EW, Borgdorff MW, Cobelens FGJ, 2010. National survey of tuberculosis prevalence in Viet Nam. *Bulletin of the World Health Organization*, 88, 273–280.

Hodgson T and Burke M, 2000. On simulation and the teaching of statistics. *Teaching statistics*, 22, 91-96.

Hoeting JA, Davis RA, Merton AA, and Thompson SE, 2006. Model Selection for Geostatistical Models. *Ecological Applications*, 16(1), 87-98.

Hoeting JA, Madigan D, Raftery AE, and Volinsky CT, 1999. Bayesian Model Averaging: a tutorial. *Statistical Science,* 14(4), 382–417.

Hoeting JA, Raftery AE, and Madigan D, 2002. Bayesian Variable and Transformation Selection in Linear Regression. *Journal of Computational and Graphical Statistics* , 11 (3), 485-507.

Hogg RS, Heath KV, Yip B, Craib KJ, O'Shaughnessy MV, Schechter MT, Montaner JS, 1998. Improved survival among HIV infected individuals following initiation of antiretrovirals. *JAMA,* 279, 450–454.

Hoover DR, Rice JA, Wu CO, and Yang LP, 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika,* 85, 809–822.

Howard AA, and El-Sadr WM, 2010. Integration of Tuberculosis and HIV Services in Sub-Saharan Africa: Lessons Learned. *Clin Infect Dis.*, 50 **(Supplement 3):** S238-S244.

Hox JJ, 2002. *Multilevel analysis: techniques and applications.* Mahwah, N.J., Lawrence Erlbaum Publishers.

Hu FB, Goldberg J, Hedeker D, Flay BR, Pentz MA, 1998. Comparison of population-averaged and subject specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology* 147, 694-703.

Huber PJ, 1967. The behavior of maximum likelihood estimates under non-standard conditions. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol 1. Berkeley, CA: University of California Press, pp. 221–233.

Hussey MA, and Hughes JP, 2007. Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials,* 28, 182–191.

Isaakidis P, And Ioannidis PA, 2003. Evaluation of cluster randomised controlled trials in Sub-Saharan African. *Am J Epidemiol.,* 158, 921-926.

Janega JB, Murray DM, Varnell SP, Blitstein JL, and Birnbaum AS, 2004. Assessing intervention effects in a school based nutrition intervention trial: which analytic model is most powerful. *Health Education and Behavior,* 31(6), 756-774.

Kass RE, and Raftery AE, 1995. Bayes factors, *Journal of the American Statistical Association*, 90(430), 773-795.

Kemp J, Mann G, Nhlema Simwaka B, Salaniponi FML, and Squire S.B, 2007. Can the poor afford free TB treatment? Patient and household costs associated with a TB diagnosis in Lilongwe. *Bulletin of the World Health Organisation*, 85(8), 580–585.

Kenny DA, Mannetti L, Pierro A, Livi S, and Kashy DA, 2002. The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, 83, 126–137.

Kerry SM, Bland JM, 1998. The intracluster correlation coefficient in cluster randomisation. *BMJ*, 316, 1455.

Kerschberger B, Hilderbrand K, Boulle AM, Coetzee D, Goemaere E, De Azevedo V, and Van Cutsem G, 2012. The effect of complete integration of HIV and TB services on time to initiation of antiretroviral therapy: a before-after study. *PLoS ONE*, 7(10), e46988. doi:10.1371/journal.pone.0046988.

Kim H-Y, Preisser JS, Rozier RG, and Valiyaparambil JV, 2006. Multilevel analysis of group-randomized trials with binary outcomes. *Community Dent Oral Epidemiology*; 34, 241–51.

Kim JY, and Ammann A, 2004. Is the "3 by 5" initiative the best approach to tackling the HIV pandemic? PLoS Med 1: e37.

Kirk O, Gatell JM, Mocroft A, Pedersen C, Proenca R, Brettle RP, Barton SE, Sudre P, and Phillips AN, 2000. Infections with Mycobacterium tuberculosis and Mycobacterium avium among HIV-infected patients after the introduction of highly active antiretroviral therapy. EuroSIDA Study Group. *American Journal of Respiratory and Critical Care Medicine*, 162, 865-872.

Kish L, 1965. *Survey Sampling*. New York : John Wiley & Sons.

Klar N and Donner A, 1997. The merits of matching in community intervention trials: a cautionary tale. *Statist. Med.*, 16, 1753-1764.

Kochi A, 1997. Tuberculosis control— is DOTS the health breakthrough of the 1990s? *World Health Forum,* 18, 225-232.

Koepsell TD, Wagner EH, Cheadle AC, Patrick DL, Martin DC, Diehr PH, Perrin EB, Kristaz AR, Allan-Andrilla CH, and Dey LJ, 1992. Selected methodological issues in evaluating

community-based health promotion and disease prevention programs. *Annu. Rev. Public Health,* 13, 31–57.

Kramer MS, Martin RM, Sterne JAC, Shapiro S, Dahhou M, and Platt RW, 2009. The double jeopardy of clustered measurement and cluster randomisation. *BMJ,* 339,b2900, doi: 10.1136/bmj.b2900.

Kreft IGG, and Leeuw J, 1998. *Introducing multilevel modelling.* Thousand Oaks, CA: Sage.

Kuhn L, Davidson LL, and Durkin MS, 1994. Use of Poisson Regression and Time Series Analysis for Detecting Changes over Time in Rates of Child Injury following a Prevention Program. *Am J Epidemiol,* 140, 943-955.

Larsen K, Petersen JH, Budtz-Jorgensen, and Endahl L, 2000. Interpreting parameters in the logistic regression model with random effects. *Biometrics*, 56, 909-914.

Lawn SD, Afful B, and Acheampong JW, 1998. Pulmonary tuberculosis: diagnostic delay in Ghanaian adults. *Int J Tuberc Lung Dis*, 2(8), 635-640.

Lawn SD, Harries AD, Anglaret X, Myer L, and Wood R, 2008. Early mortality among adults accessing antiretroviral treatment programmes in sub-Saharan Africa. *AIDS,* 22(15), 1897-1908.

Legido-Quigley H, Montgomery CM, Khan P, Atun R, Fakoya A, Getahun H, and Grant AD, 2013. Integrating tuberculosis and HIV services in low- and middle- income countries: a systematic review. *Tropical Medicine and International Health*, 18(2), 199 - 211.

Lee J, and Chia KS, 1993. Estimation of prevalence rate ratios for cross sectional data: an example in occupational epidemiology. *Br J Ind Med,* 50, 861- 864.

Lee D, and Shaddick G, 2005. Modelling the effects of air pollution on health using Bayesian Dynamic Generalised Linear Models, *Technical report*, University of Bath.

Lee D, and Shaddick G, 2007. Time-Varying Coefficient Models for the Analysis of Air Pollution and Health Outcome Data. *Biometrics*, 63, 1253–1261.

Lewin S, Dick J, Pond P, Zwarenstein M, Aja G, van Wyk B, Bosch-Capblanch X, and Patrick M, 2005. Lay health workers in primary and community health care.*Cochrane Database of Systematic Reviews,* CD004015.

Liang KY, and Zeger SL, 1986. Longitudinal data analysis using generalised linear models. *Biometrika*, 73, 13-22.

Lin DY, and Wei LJ, 1989. The robust inference for the Cox proportional hazards model.*Journal of the American Statistical Association,* 84**,** 1074-1078.

Lobue PA, Perry S, and Catanzaro A, 2000. Diagnosis of tuberculosis. In: Reichman LB, Hershfield ES, editors. Tuberculosis— a comprehensive international approach. New York: Marcel Dekker. pp. 341-75.

Lönnroth K, Castro KG, Chakaya JM, Chauhan LS, Floyd K, Glaziou P, and RaviglioneMC, 2010. Tuberculosis control and elimination 2010–50: cure, care, and social development. *Lancet*, 375, 1814–1829.

Lönnroth K, Jaramillo E, Williams BG, Dye C, and Raviglione M, 2009. Drivers of tuberculosis epidemics: the role of risk factors and social determinants. *Social Science and Medicine*, 68(12), 2240-2246.

Lu Y and Zhang R, 2009. Smoothing spline estimation of generalised varying –coefficient mixed model. *Journal of Nonparametric Statistics*, 21 (7), 815-825.

Lumley T, Kronmal R, and Ma S, 2006. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. UW Biostatistics Working Paper Series, paper 293.

Lunn DJ, Thomas A, Best N, and Spiegelhalter D, 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325--337.

Maas CJM, and Hox JJ, 2005. Sufficient sample sizes for multilevel modelling. *Methodology*, 1(3), 86-92.

Macinko J, Guanais F, de Fatima M, and de Souza M, 2006. Evaluation of the impact of the Family Health Programme in infant mortality in Brazil, 1990 – 2002. *J Epidemiol Community Health*, 60(1), 13-19.

MacPherson P, Corbett EL, Makombe SD, van Oosterhout JJ, Manda E, Choko AT, Thindwa D, *Squire SB*, Mann GH, Lalloo DG,2012. Determinants and consequences of failure of linkage to antiretroviral therapy at primary care level in Blantyre, Malawi: a prospective cohort study.***PLoS One***. 7(9):e44794. doi: 10.1371/journal.pone.0044794. Epub 2012 Sep 11.

Madigan D, York J, and Allard D, 1995. Bayesian graphical models for discrete data. *International Statistical Review*, 63 (2), 215-232.

Maher D, 2010. Re-thinking global health sector efforts for HIV and tuberculosis epidemic control: promoting integration of programme activities within a strengthened health system. *BMC Public Health,* 10, 394. doi:10.1186/1471-2458-10-394.

Mannheim L, 1999. Health services research clinical trials: issues in the evaluation of economic cost and benefits. Controlled Clinical Trials, 19, 149–158.

Mantel N, and Haenszel W, 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *JNCI*, 22, 719-748.

Marsh VM, Mutemi WM, Muturi J, Haaland A, Watkins WM, Otieno G, and Marsh K, 1999. Changing home treatment of childhood fevers by training shop keepers in rural Kenya. *Trop Med Int Health*, 4(5), 383-389.

Marsh VM, Mutemi WM, Willetts A, Bayah K, Were S, Ross A, and Marsh K, 2004. Improving malaria home treatment by training drug retailers in rural Kenya. *Trop Med Int Health*, 9(4), 451-60.

Martin DC, Diehr P, Perrin EB, and Koepsell TD, 1993. The effect of matching on the power of randomised community intervention studies. *Statistics in Medicine*, 12, 329-38.

Matebesi Z, and Booysen F, 2004. Treatment adherence among tuberculosis patients. *Acta Academica,* 36(3), 140–171.

McClatchey MW, Cohen SJ, and Reed FM, 1992. The usefulness of matched pair randomization for medical practice-based research. *Family Practice Research Journal*, 12, 235-243.

McCullagh P, and Nelder JA, 1989. *Generalised Linear Models*. Second edition. London: Chapman and Hall.

McNutt LA, Wu C, Xue X, and Hafner JP, 2003. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol*, 157, 940–943.

Meade TW, Roderick PJ, Brennan PJ, Wilkes HC, Kelleher CC, 1992. Extracranial bleeding and other symptoms due to low dose aspirin and low intensity oral anticoagulation. Thromb Haemostasis, *68, 1-6*.

Metropolis N, Rosenbluth AW, Rosenbluth MN, and Teller AH, 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087-1092

Mickey RM, Goodwin GD, and Costanza MC, 1991. Estimation of the design effect in community intervention studies. *Statistics in Medicine*,10, 53–64.

Mock J, McPhee SJ, Nguyen T, Wong C, Doan H, Lai KQ, Nguyen KH, Nguyen TT, and Bui-Tong N, 2007. Effective lay health worker outreach and media-based education for promoting cervical cancer screening among Vietnamese American women. *Am J Public Health*, 97, 1693–1700.

Moineddin R, Matheson FI, and Glazier RH, 2007. A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology*, 7:34 doi:10.1186/1471-2288-7-34.

Møller V, and Erstad I, 2007. Stigma associated with tuberculosis in a time of HIV/AIDS: narratives from the Eastern Cape, South Africa. *South African Review of Sociology,* 38(2), 103–119.

Møller V, Erstad I, Cramm JM, Nieboer AP, Finkenflügel H, Radloff S, Ndoro T and Kwizera SA, 2011. Delays in presenting for tuberculosis treatment associated with fear of learning one is HIV-positive. *African Journal of AIDS Research, 10(1), 25–36.*

Moore DF and Tsiatis A, 1991. Robust estimation of the variance in moment methods for extra-binomial and extra-Poison variation. *Biometrics*, 47, 383-401.

Mukadi YD, Maher D, and Harries A, 2001. Tuberculosis case fatality rates in high HIV prevalence populations in sub-Saharan Africa. *AIDS*, 15, 143–152.

Mupere E, Schiltz NK, Mulogo E, Katamba A, Nabbuye-Sekandi J, and Singer ME, 2013.Effectiveness of active case-finding strategies in tuberculosis control in Kampala, Uganda. The International Journal of Tuberculosis and Lung Disease, 17(2), 207-213.

Murphy SA and Johnson LC, 2006. Methodological Issues Associated With Group Intervention Research. *Archives of Psychiatric Nursing*, 20 (6), 276–281.

Murray DM, 1998. *Design and Analysis of Group-Randomized Trials.* New York : Oxford University Press.

Murray DM, Hannan PJ, and Baker WL, 1996. A Monte Carlo study of alternative responses to intraclass correlation in community trials: is it ever possible to avoid Cornfield's penalties? *Evaluation Review,* 20(3), 313-337.

Murray DM, Hannan PJ, Wolfinger RD, Baker WL, and Dwyer JH, 1998. Analysis of data from group-randomised trials with repeat observations on the same group. *Statistics in Medicine,* 17, 1581-1600.

Murray DM, Pals SL, Blitstein JL, Alfano CM, and Lehman L, 2008. Design and analysis of group-randomized trials in cancer: a review of current practices. *J Natl Cancer Inst*, 100, 483 – 491.

Murray DM , Varnell SP, and Blitstein JL, 2004. Design and analysis of group randomized trials: a review of recent methodological developments .*Am J.Public Health,* 94 (3), 423 – 432.

Malawi National AIDS Commission, 2007. *Sentinel Surveillance Report*, National AIDS Commission, Lilongwe, Malawi.

Malawi National AIDS Commission, 2010. *Sentinel Surveillance Report*, National AIDS Commission, Lilongwe, Malawi.

Naidoo S, Taylor M, and Jinabhai C, 2007. Critical risk factors driving the tuberculosis epidemic in KwaZulu-Natal, South Africa. *The Southern African Journal of Epidemiology and Infection,* 22, 45–49.

Nakagawa S and Schielzeth H, 2010. Repeatability for gaussian and non-Gaussian data: a practical guide for biologists. *Biol. Rev*, 85,935-956.

National Population Commission (NPC) and ICF Macro. 2009. *Nigeria Demographic and Health Survey 2008: Key Findings*. Calverton, Maryland, USA: NPC and ICF Macro.

National Statistical Office (NSO), 2010. *The Population and Housing census, 2008*. Zomba, Malawi.

National Statistical Office (NSO) and ICF Macro. 2011. *Malawi Demographic and Health Survey 2010*. Zomba, Malawi, and Calverton, Maryland, USA: NSO and ICF Macro.

Neal, RM, 2003. Slice sampling. *Ann. Statist*., 31(3), 705 - 767.

Needham DM, Foster SD, Tomlinson G, and Godfrey-Faussett P, 2001. Socioeconomic, gender and health services factors affecting diagnostic delay for tuberculosis patients in urban Zambia. *Trop Med Int Health,* 6(4), 256-259.

Nelder JA, and Wedderburn RWM, 1972. Generalised linear Models. *Journal of the Royal Statistical Society, Series A,* 135, 370-384.

Nijem K, Kristensen P, Al-Khatib A, and Bjertness E, 2005. Application of different statistical methods to estimate relative risk for self-reported health complaints among shoe factory workers exposed to organic solvents and plastic compounds. *Norsk Epidemiologi,* 15, 111-116

Norris SL, Chowdhury FM, Van Le K, Horsley T, Brownstein JN, Zhang X, Jack L Jr, and Satterfield DW, 2006. Effectiveness of community health workers in the care of persons with diabetes. *Diabet Med.*,23(5), 544–556.

Ntzoufras I, 2008. Bayesian Modeling Using WinBUGS. 2nd ed. John Wiley & Sons, Inc.

Nyirenda T, 2006. Epidemiology of Tuberculosis in Malawi. *Malawi Medical journal*, 18(3), 147-159.

Oakeshott P, Kerry SM, and Williams JE, 1994. Randomised controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referral for radiographic examination. *Br J Gen Pract, 44,197-200.*

O'Campo P, 2003. Invited commentary: Advancing theory and methods for multilevel models of residential neighbourhoods and health. *Am J Epidemiology*, 157, 9-13.

O'Donnell MR, Chamblee S, von Reyn CF, Marsh BJ, Moreland JD, Narita M, Johnson LS, and Horsburgh Jr CR, 2012. Sustained reduction in tuberculosis incidence following a community-based participatory intervention. *Public Health Action*, 2(1): 23–26.

Omar RZ, and Thompson SG, 2000. Analysis of a cluster randomized trial with binary outcome data using a multi-level model .*Statist*. *Med*., 19, 2675-2688.

Osborn J and Cattaruzza MS, 1995. Odds ratio and relative risk for cross-sectional data.*International Journal of Epidemiology,* 24, 464-465.

Ouedraogo M, Kouanda S, Boncoungou K, Dembele M, Zoubga ZA, Ouedraogo SM, and Coulibaly G, 2006. Treatment seeking behaviour of smear-positive tuberculosis patients diagnosed in Burkina Faso. *Int J Tuberc Lung Dis,* 10(2), 184-187.

Palella FJ, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, Aschman D.J, and Holmberg SD, 1998. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *N Engl J Med,* 338, 853–860.

Palmer RH, Louis TA, Hsu LN, Peterson HF, Rothrock JK, Strain R, Thompson MS, and Wright EA, 1985.A randomized controlled trial of quality assurance in sixteen ambulatory care practices. *Med Care*, 23, 751–770.

Pals Sl, Wiegand RE, and Murray DM, 2011. Ignoring the group in group-level HIV/AIDS intervention trials: a review of reported design and analytic methods. *AIDS*, 25, 989-996.

Parker DR, Evangelou TE, and Eaton CB, 2005. Intraclass correlation coefficients for cluster randomized trials in primary care: the cholesterol education and research trial. *Contemporary Clinical Trials*, 26, 260–267.

Petersen MR, and Deddens JA, 2006. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol,* 163**,** 1157-1163.

Petersen MR and Deddens JA, 2008. A comparison of two methods estimating prevalence ratios. BMC Med Res Methodol. 2008; 8: 9. doi: *10.1186/1471-2288-8-9*.

Phillips K A, Morrison KR, Andersen R, and Aday LA, 1998. Understanding the context of health care utilization: assessing environmental and provider-related variables in the Behavioral Model of Utilization. *Health Services Research, 33,* 571–596.

Pocock SJ, 1997. *Clinical trials: a practical approach*. New York. John Wiley & Sons Ltd.

Porter K, Babiker A, and Bhaskaran, 2003. Determinants of survival following HIV-1 seroconversion after the introduction of HAART. *Lancet*, 362, 1267-1274.

Preidis GA, McCollum ED, Kamiyango W, Garbino A, Hosseinipour MC, Kazembe PN, Schutze GE, and Kline MW, 2013. Routine inpatient provider-initiated HIV testing in Malawi, compared with client-initiated community-based testing, identifies younger children at higher risk of early mortality. *J Acquir Immune Defic Syndr.,* 63(1):e16-22.

Preisser JS, Reboussin BA, Song E, and Wolfson M, 2007. The importance and role of intracluster correlations in planning cluster trials. *Epidemiology,* 18 (5), 552-560.

Rabe-Hesketh S, and Skrondal A, 2008. *Multilevel and longitudinal modelling using Stata*. 2nd ed. Texas: Stata Press Publication.

Rabe-Hesketh S, and Skrondal A, 2012. *Multilevel and longitudinal modelling using Stata*. 3rd ed. Texas: Stata Press Publication.

Rabe-Hesketh S, Skrondal A, and Pickles A, 2002. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* 2 (1), 1-21.

Rabe-Hesketh S, Skrondal A, Pickles A, 2005. Maximum likelihood estimation of limited and discrete dependant variable models with nested random effects. *Journal of Econometrics,* 128, 301-323.

Raftery AE, 1996. Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika*, 83, 251-266.

Rajeswari R, Chandrasekaran V, Suhadev M, Sivasubramaniam S, Sudha G, and Renu G, 2002. Factors associated with patient and health system delays in the diagnosis of tuberculosis in South India. The International Journal of Tuberculosis and Lung Disease, 6 (9), 789-795.

Rao VG, Bhat J, Yadav R, Gopalan GP, Nagamiah S, et al., 2012. Prevalence of Pulmonary Tuberculosis - A baseline survey in Central India. *PLoS ONE* 7(8), e43225. doi:10.1371/journal.pone.0043225.

Raudenbush SW, and Bryk AS, 2002. *Hierarchical Linear Models in Social and Behavioral Research: Applications and Data-Analysis Methods*, 2nd Edition, Sage Publications, Thousand Oaks.

Raudenbush SW, and Liu X, 2000. Statistical power and optimal design for multisite randomized trials. *Psychological Methods*,5, 199-213.

Raudenbush SW, and Xiao-Feng L, 2001. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6(4), 387-401.

Ray M, Logan R, Sterne JA, et al, 2010. The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals. *AIDS*, 24, 123-137.

Rodriguez G, and Goldman N, 1995. An assessment of estimation procedures for multilevel models with binary responses. *J R Stat Soc A*, 158, 73–90.

Rodriguez G and Goldman N, 2001. Improved estimation procedures for multilevel models with binary response: a case study. *Journal of the Royal Statistical society*, 158, 73-89.

Rosenstock IM, Strecher VJ, and Becker MH, 1988. Social Learning Theory and the Health Belief model. *Health Educ. Behav., 15(2), 175 - 183.*

Rotheram-Borus MJ, Lee MB, Gwadz M, and Draimin B, 2001. An intervention for parents with AIDS and their adolescent children. *Am J Public Health,* 91, 1294–1302.

Russell S, 2004. The economic burden of illness for households in development countries: a review of the studies focusing on malaria, tuberculosis and HIV/AIDS. *Am J Trop Med Hyg*., 71(Suppl 2), 147–155.

Salaniponi FM, Harries AD, Banda HT, Kang'ombe C, Mphasa N, Mwale A, Upindi B, Nyirenda TE, Banerjee A, and Boeree MJ, 2000. Care seeking behaviour and diagnostic processes in patients with smear-positive pulmonary tuberculosis in Malawi. *Int J Tuberc Lung Dis*, 4(4), 327-332.

Savu A, Liu Q and Yasu Y, 2010. Estimation of relative risk and prevalence ratio. Stat Med., 29(22), 2269-2281.

Scheines R, Hoijtink H, and Boomsma A, 1999. Bayesian estimation and testing of structural equation models. *Psychometrika, 64*, 37–52.

Schouten EG, Dekker JM, and Kok FJ, 1993. Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. *Stat Med.,* 12, 1733-1745.

Schwarz G, 1978. Estimating the dimension of a model. *Ann. Statis.*, *6(2), 461-464.*

Sekandi JN, Neuhauser D, Smyth K, and Whalen CC, 2009. Active case finding of undetected tuberculosis among chronic coughers in a slum setting in Kampala, Uganda. *Int J Tuberc Lung Dis.,* 13,508–513.

Seltzer MH, Wong, WH, and Bryk AS, 1996. Bayesian analysis in applications of hierarchical models: issues and methods. *Journal of Educational and Behavioral Statistics,21*, 131–167.

Simpson JM, Klar N, and Donner A, 1995. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993 .*Am J Public Health,* 85(10), 1378 – 1383.

Simwaka BN, Theobald S, Willets A, Salaniponi FML, Nkhonjera P, Bello G, and Squire SB, 2012. Acceptability and Effectiveness of the Storekeeper-Based TB Referral System for TB Suspects in Sub-Districts of Lilongwe in Malawi. *PLoS ONE,* 7(9), e39746. doi:10.1371/journal.pone.0039746.

Sjögren T, Nissinen KJ, Järvenpää SK, Ojanen MT, Vanharanta H, and Mälkiä EA, 2005. Effects of a workplace physical exercise intervention on the intensity of headache and neck and shoulder symptoms and upper extremity muscular strength of office workers: a cluster randomized controlled crossover trial. *J Int Assoc Stud Pain,* 116, 119–128.

Skordis-Worrall J, Hanson K, and Mills A, 2010. Confusion, caring and tuberculosis diagnostic delay in Cape Town, South Africa. *International Journal of Tuberculosis and Lung Disease,* 14(2), 171–180.

Skov T, Deddens JA, Petersen MR, and Endahl L, 1998. Prevalence proportion ratios: estimation and hypothesis testing. *International Journal of Epidemiology,* 27,91-95.

Smeeth L and Siu-Woon Ng E, 2002. Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC Trial of the Assessment and
Management of Older People in the Community. *Controlled Clinical Trials*, 23, 409–421.

Snijders TAB, 1996. Analysis of longitudinal data using the hierarchical linear model. *Quality and Quantity*, 30, 405-426.

Snijders T, and Bosker R, 1999. *Multilevel Analysis: an Introduction to basic and advanced multilevel modelling*, Sage Publications, Thousand Oaks.

Snijders TAB, and Bosker RJ, 2012. Multilevel analysis: an introduction to basic and advanced multilevel modelling. Thousand Oaks, CA:Sage.

Spiegelhalter DJ, 2001. Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine*, 20, 435–452.

Spiegelhalter DJ, Best NG, Carlin BP, and van der Linde A, 2002. Bayesian measures of model complexity and fit (with discussion). *J Roy Statist Soc B*, 64, 583-639.

Spiegelhalter D, Thomas A, Best N, and Lunn D, 2003. WinBUGS user manual. *Cambridge: MRC Biostatistics Unit*.

Spiegelman D, and Hertzmark E, 2005. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol,* 162, 199–200.

Squire S B, Belaye A K, Kashoti A, Salaniponi F M L, Mundy CJF, Theobald S, Kemp J, 2005. "Lost" smear positive pulmonary tuberculosis cases; where are they and why did we lose them? *Int J Tuberc Dis.,* 9, 25-31.

Stiratelli R, Laird NM, and Ware JH, 1984. Random effects models for serial observations with binary response. Biometrics, 40, 961-971.

Storla DG, Yimer S, and Bjune GA, 2008. A systematic review of delay in the diagnosis and treatment of tuberculosis. *BMC Public Health*, 14, 8:15.

Stromberg U, 1994. Prevalence Odds Ratios v.s. Prevalence Ratio. *Occupational Environmental Medicine,* 51, 143-144.

Stryn H, Sanchez J, Morley P, Booker C, and Dohoo IR, 2006. Interpretation of variance parameters in multilevel Poisson regression models. Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics. Available at www.sciquest.org.nz

Styblo K, and Bumgarner JR, 1991. Tuberculosis can be controlled with existing technologies: evidence. The Hague: Tuberculosis Surveillance Research Unit; 1991. Progress Report 1991. pp. 60-72.

Subramani R, Radhakrishna S, Frieden TR, Kolappan C, Gopi PG, Santha T, Wares F, Selvakumar N, and Narayanan PR, 2008**.** Rapid decline in prevalence of pulmonary tuberculosis after DOTS implementation in a rural area of South, India. *Int J Tuberc Lung Dis*, 12 (8), 916–920.

Thall PF, and Vail SC, 1990. Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46, 657-671.

Thompson B, Coronado G, Snipes SA, and Puschel K, 2003. Methodologic advances and ongoing challenges in designing community-based health promotion programs.*Annu. Rev. Public Health*, 24, 315–340.

Thompson ML, Myers JE, Kriebel D, 1998. Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: what is to be done? *Occup Environ Med*., 55, 272–277.

Thompson SG, Warn DE and Turner RM, 2004. Bayesian methods for analysis of binary outcome data in cluster randomized trials on the absolute risk scale. *Statist. Med.*, 23, 389–410.

Torgerson DJ, 2001. Contamination in trials: is cluster randomisation the answer? *BMJ,* 322, 355–357.

Turner RM, Omar RZ, and Thompson SG, 2001. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine,*20, 453–472.

Tutz G, and Kauermann G, 2003. Generalised linear random effects models with varying coefficients. *Computational Statistics and Data Analysis,*43, 13-28.

Uplekar M, and Raviglione MC, 2007. The "vertical-horizontal" debates: time for the pendulum to rest (in peace)? *Bull World Health Organ*, 85, 413-414.

Uwimana J, Zarowsky C, Hausler H and Jackson D, 2012. Training community care workers to provide comprehensive TB/HIV/PMTCT integrated care in KwaZulu-Natal: lessons learnt. *Tropical Medicine and International Health*, 17 (4), 488–496.

Uyei J, Coetzee D, Macinko J, and Guttmacher S, 2011. Integrated delivery of HIV and tuberculosis services in sub-Saharan Africa: a systematic review. *The LancetInfectious Diseases,* 11, 855-867.

Varnell SP, Murray DM , and Baker WL, 2001. An evaluation of analysis options for the one group per condition design: can any of the alternatives overcome the problems inherent in this design? *Eval Rev.,* 25(4), 440 – 453.

Varnell SP, Murray DM, Janega JB, and Blitstein JL, 2004. Design and analysis of group-randomized trials: a review of recent practices .*Am J Public Health,* 94 (3), 393 – 399.

Vernazza PL, Troiani L, Flepp MJ, et al, 2000. Potent antiretroviral treatment of HIV infection results in suppression of the seminal shedding of HIV. *AIDS*, 14, 117-121.

Verver S, Bwire R, and Borgdorff MW, 2001. Screening for pulmonary tuberculosis among immigrants: estimated effect on severity of disease and duration of infectiousness. *International Journal of Tuberculosis and Lung Disease*, 5, 419-425.

Verver S, Warren RM, Munch Z, Richardson M, van der Spuy GD, Borgdorff MW, Behr MA, Beyers N, and van Helden PD, 2004. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. *Lancet,* 363**,** 212–14.

Wacholder S, 1986. Binomial regression in GLIM, estimating risk ratios and risk differences. *Am J Epidemiol*, 123, 174–184.

Wandwalo ER, and Morkve O, 2000. Delay in tuberculosis case-finding and treatment in Mwanza, Tanzania. *Int J Tuberc Lung Dis,* 4(2), 133-138.

Wasserman L, 2000. Bayesian Model selection and model averaging. *Journal of Mathematical Psychology*, 44, 92-107.

Wedderburn RWM, 1974. Quasi-likelihood functions, generalised linear models and the Gauss-Newton method. *Biometrika*, 61, 439-447.

Weiner B, 1986. An *attributional theory of motivation and emotion*. New York: Springer-Verlag.

World Health Organization, 1989. *Strengthening the performance of community health workers in primary health care. Report of a WHO Study Group*. Geneva, World Health Organization (WHO Technical Report Series, No. 780).

World Health Organization, 2004. Interim policy on collaborative TB/HIV activities. Geneva: World Health Organization.

World Health Organization, 2006. *The Stop TB Strategy: building on and enhancing DOTS to meet the TB-related Millennium Development Goals*. Geneva, World Health Organization, (WHO/HTM/TB/2006.368).

World Health Organization, 2007. Global Tuberculosis Control: Surveillane, Planning, Financing. Geneva, World Health Organisation Report 2007.

World Health Organization, 2009. *Global tuberculosis control: Epidemiology, Strategy, Financing*. World Health Organization, Geneva, World Health report, 2009.

World Health Organization, 2010. *Global Tuberculosis Control: Surveillance, Planning, Financing*: World Health Organisation report, 2010.

World Health Organization, 2011. Early detection of Tuberculosis: an overview of approaches, guidelines and tools. WHO/HTM/STB/PSI/2011.21.

WHO/UNAIDS/UNICEF, 2007. *Towards universal access. scaling-up priorities HIV/AIDS interventions in the health sector*. Progress report, April 2007.

Wood R, Middelkoop K, Myer L, Grant AD, Whitelaw A, Lawn SD, Kaplan G, Huebner R, McIntyre J, and Bekker LG, 2007.Undiagnosed tuberculosis in a community with high HIV prevalence: implications for tuberculosis control. *Am J Respir Crit Care Med.,* 175,87–93.

Xiong CF, Fang Y, Zhou LP, Zhang XF, Ye JJ, Li GM, Liu X, Wang XJ, and Yang CF, 2007. Increasing TB case detection through intensive referral of TB suspects by village doctors to county TB dispensaries. *Int J Tuberc Lung Dis*, 11(9), 1004–1007.

XU B, Diwan VK, and Bogg L, 2007. Access to tuberculosis care: What did chronic cough patients experience in the way of healthcare-seeking? *Scandinavian Journal of Public Health*, 35, 396–402.

Yee JL, and Niemeier D, 1996. Advantages and disadvantages: longitudinal vs. repeated cross-section surveys [online]. FHWA, HPM-40. Accessed 23rd April 2013 at http://ntl.bts.gov/data/letter_am/bat.pdf.

Yimer S, Bjune G, and Alene G, 2005. Diagnostic and treatment delay among pulmonary tuberculosis patients in Ethiopia: a cross sectional study. *BMC Infect Dis*, 5, 112. doi: *10.1186/1471-2334-5-112*.

You Z, Williams OD, Aban I, Kabagambe EK, Tiwari HK, and Cutter G, 2011. Relative efficiency and sample size for cluster randomized trials with variable cluster sizes. Clinical Trials 2011; 8: 27–36.

Yu B, and Wang Z, 2008. Estimating relative risks for common outcome using PROC NLP. *Computer Methods and Programs in Biomedicine*, 90, 179–186.

Zhao Y, Staudenmayer J, Coull B A, and Wand MP, 2006. General design Bayesian generalized linear mixed models. *Statistical Science*, 21(1), 35–51.

Zhang J, and Yu K, 1998. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*, 280, 1690–1691.

Zhang Z, Hamagami F, Wang L, Grimm KJ, and Nesselroade JR, 2007. Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioural Development*, *31*(4), 374-383.

Zou G, 2004. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol*, 159, 702–706.

Zocchetti C, Consonni D and Bertazzi PA, 1995. Estimation of prevalence rate ratios from cross-sectional data [letter; comment]. *International Journal of Epidemiology,* 24**,** 1064-1065.

Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, and Moher D, 2008. CONSORT and Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* , 337, a2390 doi: 10.1136/bmj.a2390.