UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

Malaria in pregnancy (MiP)

Leonard Eluma Nwogu-Ikojo

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :
dr. Yannick VANDENDIJCK

SUPERVISOR :
Prof. Koen PEETERS

2017
2018

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics

Masterthesis

Malaria in pregnancy (MiP)

Leonard Eluma Nwogu-Ikojo

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

dr. Yannick VANDENDIJCK

SUPERVISOR :

Prof. Koen PEETERS

Malaria in Pregnancy (MiP): Assessing Communities' Response to Community SST (CSST) carried out by CHW

Universiteit Hasselt

Leonard Eluma Nwogu-Ikojo

June 15, 2018

Acknowledgement

Many Thanks to the Almighty God for His Protection and Grace.

Abstract

Malaria is an infection of the red blood cells by any of 4 species of *plasmodium* observed to infect humans: *vivax, ovale, malariae, and falciparum* which is the most prominent. This parasite is transmitted from its host, the female anopheline mosquito, into the human system when the individual receives a bite from a feeding mosquito. Global maternal mortality figures are put at over 10,000 deaths and in sub-Saharan Africa, infant death is estimated between 75,000 to 200,000 every year.

In this study we review the outcomes of the community-scheduled screening and treatment (CSST) program using community health workers (CHW) in three West African Countries, that is, Benin, Burkina Faso and The Gambia, between between November 2013 and November 2015. The study uses Survival analysis techniques, logistics and beta-binomial regression models and generalized estimating equations (GEE) to estimate model parameters in the afore mentioned models in order to determine the effect of the CSST program on the intervention community when compared to randomized control community. The study investigates community differences by adopting appropriate statistical methods vis-à-vis the study research question and the nature of the data being investigated.

From the results observed, we see that the data modeled using different statistical methods and approaches were consistent in their results. We see the effect of the CSST program on health seeking behavior of women in the intervention community in Burkina Faso was significantly different from the control community. Conditioned on the health Center, we also see a reduction in the gestation time to the first ANC in Burkina Faso.These effects were not observed in Benin and The Gambia.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background of the study

Malaria is an infection of the red blood cells by any of 4 species of *plasmodium* observed to infect humans: *vivax, ovale, malariae, and falciparum* which is the most prominent [1, 2]. This parasite is transmitted from its host, the female anopheline mosquito, into the human system when the individual receives a bite from a feeding mosquito[1]. Tropical climates are suitable for mosquitoes to thrive hence it is a common infection in Africa and other areas with similar climates around the world. Commonly observed symptoms upon the acquisition of malaria infection include include fever, headaches and nausea or vomiting[2].

During pregnancy, women are more susceptible to malaria infection due to reduced immunity and hormonal changes when compared with women who are not pregnant[6]. Malaria infection during pregnancy is associated with adverse effects on fetal growth, low infant birth weight, maternal anemia, pre-term delivery, infant mortality and maternal mortality[1, 2, 3]. Global maternal mortality figures are put at over 10,000 deaths and in sub-Saharan Africa, infant death is estimated between 75,000 to 200,000 every year[3, 5]. An effective control of malaria in pregnancy through simple, targeted and sustainable programs could save several lives of mothers and babies. Integrating an effective program with antenatal care (ANC) for pregnant women is a sustainable platform on which these programs can be implemented especially in malaria endemic areas[4].

Several intervention packages to control malaria and its negative effects during pregnancy have been implemented by the World Health Organization,(WHO). Currently, a three pronged approach is in place which includes the use of insecticide treated mosquito nets (ITN), the administration of intermittent preventive treatment with sulfadoxine-pyrimethamine (IPTp-SP), and appropriate case management through prompt and effective treatment of malaria in pregnant women[4, 5]. Observed drawbacks to this approach include health system policy differences and inaccessibility of some communities to antenatal health care facilities or services for prompt case management of Malaria in Pregnancy (MiP) , especially in rural sub-Saharan Africa[3, 6, 7].

A direct approach to the observed challenges is the community-scheduled screening and treatment (CSST) program using community health workers (CHW). It eliminates the challenge of prompt case management of MiP at rural areas especially for women who have difficulty in accessing health care services by providing pregnant women with monthly free malaria testing and treatment close to their homes. The ingenuity of this approach lies in its implicit increase in the likelihood of pregnant women in the program to receive IPTp-SP prevention treatment at every ANC visit as pregnant women are expected to generally be malaria free during such visits to the health center. An increase in the number of IPTp-SP doses during pregnancy would increase the chances of pregnant mothers having a malaria free experience, a position advocated by the WHO in 2015,

encouraging an increase in IPTp-SP intake by pregnant women to at least three times during pregnancy in malaria endemic areas [1, 5].

In this study we review the outcomes of the community-scheduled screening and treatment (CSST) program using community health workers (CHW) in three West African Countries, that is, Benin, Burkina Faso and The Gambia, between between November 2013 and November 2015 [6]. The study uses Survival analysis techniques, logistics and beta-binomial regression models and generalized estimating equations (GEE) to estimate model parameters in the afore mentioned models in order to determine the effect of the CSST program on the intervention community when compared to randomized control community. The study investigates community differences by adopting appropriate statistical methods vis-à-vis the study research question and the nature of the data being investigated.

1.2 Study Objectives

The communal effect of the CSST program on ANC clinic attendance and the administration of intermittent preventive treatment with sulfadoxine- pyrimethamine (IPTp-SP) during pregnancy was investigated in this study. The study tests the hypothesis that there is a difference in these indices between the intervention and control community. Community acceptance of the CSST program is defined by a favorable significant community effect of the CSST program on the intervention community when compared to control community where the program was not implemented. Specific research questions addressed are:

1. Is there a difference in the gestation time to first antenatal care (ANC) visit for pregnant women under the community-scheduled screening and treatment (CSST) program when compared to the control community?

2. Is there a difference in the number of women who attended at least 4 ANC visits during pregnancy in the intervention community when compared to the control community?

3. Is there a difference in the number of women who attended at least three scheduled (IPTp-SP related) ANC visits during pregnancy in the intervention community compared to the control community?

2 Study design and Data Collection

2.1 Study Design

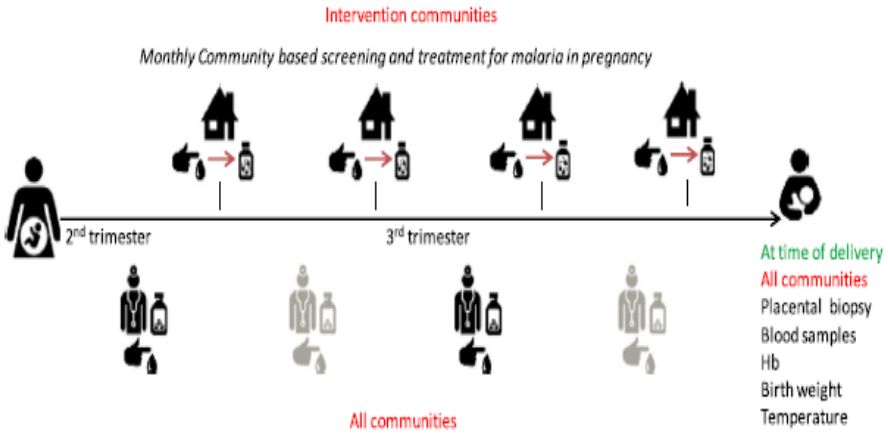
The study protocol (ISRCTN37259296 and NCT01941264) defined the study as "a multi-center, cluster-randomized, two-arm trial"[3]. The trial was conducted at the same time period in Burkina Faso (Nanoro health district), The Gambia (Upper River Region) and Benin (Glo-Djigbe, Zinvie and Ze districts) between November 2013 and November 2015 [3,6]. The countries are located in West Africa.Villages qualified for inclusion in the trial on the basis that their population size was within the range of 1000 and 2000 residents. The villages were also expected to have CHW in the working in the area.Thirty villages were randomly selected from each country, stratified by their distances to the nearest health center and randomized to either control or intervention trial arms (community). Randomization was implemented separately for each country using *Stata Statistical Software* [3].

2.2 CSST Strategy

The CSST program tackles the challenge of prompt and effective treatment of MiP. The CSST provides monthly home-visits by CHW to pregnant women in the intervention communities to test and treat for malaria [3]. This approach increases the likelihood of pregnant women's intake of IPTp-SP during scheduled visits to the health centers. Based on the design, we see the intent for at least 4 visits to the health center during pregnancy. With the CSST program in place, the expectation was that at least 3 of the four visits would be visits where malaria prevention IPTp-SP would be administered.

2.2.1 Latent process flow of the CSST program

Community health workers in the intervention villages followed specific trainings on community-based case management of malaria. The objectives of the training were to increase CHW knowledge about challenges related to MiP, the importance for monthly malaria testing of pregnant women using the rapid diagnostic test (RDT), benefits of early enrollment for ANC services (as early as possible in the second trimester), for effective coverage and administration of IPTp-SP during pregnancy[3,6]. During the trial period, CHW in intervention villages continuously approached pregnant women in the villages and encouraged the women to register as early as possible in their second trimester for antenatal care services at the health center[3,6]. CHW made monthly visits to the homes of ANC enrolled women and administered the RDT malaria test on each visit. This was compulsory. Women who tested positive for malaria based on the RDT test results were given artemether-lumefantrine (AL), a malaria treatment drug [3, 6].



Coutersy Study Protocol (ISRCTN37259296 and NCT01941264)

Figure 1: Study Design for CSST Program

CHW encouraged the women not to miss scheduled appointments for malaria prevention (IPTp-SP) at the health center during these visits. The CHW returned later and assessed uptake compliance of the malaria treatment by administering a questionnaire and checking the empty package of the treatment at the end of the course. CHW referred all other health cases to the health center for attention. There was no intervention in the control villages. CHW in the control villages were not activated for the study[3].

2.3 Data Collection and Handling

Recruitment for the two arms of the trial was done at the health center on the first ANC visit. Gestational age was determined using the fundal height. Women observed to be in their second or third trimester received the first IPTp-SP dose with a scheduled appointment for the next dose[6]. Women who were not up to the second trimester were requested to return for the IPTp-SP as early as possible during the second trimester.Before IPTp-SP administration, all women performed a physical examination,and their blood sample collected. Study staff also collected relevant socio-economic, demographic and other health related information in case report forms[6]. During the study period, participants who returned to the health centers for scheduled or unscheduled visits underwent health assessment. This was done on every repeat visit to the health center. Questionnaires were filled on completion of the assessment. Health assessment tests which prompted the suspicions of Malaria were investigated via RDT and malaria treatment AL administered if positive[6].

In this study, blinding of the participant's study arm identity was not possible. Some health centers were used by more than one village and in most cases more than one community. However, the identity and study arm of all the participants were masked to

all laboratory staff and supervisors. This was done as to mitigate against observer bias[6].

2.4 Ethical Considerations

Participation in the trials and dropout in both communities was voluntary and resident pregnant women in the communities were encouraged to participate in the program. An informed consent form was required for recruitment into the study[6]. Assent forms were required from individuals with special circumstances namely pregnant women less than 16 years from the Gambia and unmarried pregnant women in Burkina Faso and Benin[6]. Individuals who were illiterate were recruited on the strength of their thumb print accompanied by the signature of a personal witness. Individuals sho reported about sensitivity to sulphonamides and individuals judged to be with mentally deficiency were not recruited for the trial[6].

3 Exploratory Data Analysis

3.1 Summary of dataset collected

The data for the research received via questionnaires were recorded in excel files and forwarded for consolidation and analysis. A total of 4,731 pregnant women were recruited into the trials between November 2013 and November 2015.Table 1 shows the distribution of participants (pregnant women recruited) per country. The number of recruited participants in Benin were the least (about 50% of the other countries). 49 participants were dropped from the analysis due to inconsistency of information on the enrollment list.

	Benin	**B.Faso**	**Gambia**	**Total**
Recruited Participants	971	1800	1960	4731
Paticipants (Analysis)	971	1795	1916	4682
Control	429	894	958	2281
Intervention	542	901	958	2401
Dropped	0	5	44	49

Table 1: Participants by country

3.1.1 Baseline Characteristics

Table 2 provides a summary of selected baseline characteristics per country. We observe that in most countries, baseline characteristics between the intervention arm and the control arm are similar except for religious affiliation, marital status and distance from the nearest health facility in the villages. Village distances to the nearest health center in The Gambia are observed to be different from distances in other countries. Inconsistency in some baseline characteristic totals are due to missing values.

	Benin		Burkina Faso		Gambia	
Variable	**Intervention**	**Control**	**Intervention**	**Control**	**Intervention**	**Control**
Participants	**542 (0.56)**	**429 (0.44)**	**901 (0.50)**	**894 (0.50)**	**958(0.50)**	**958(0.50)**
			Individual Characteristics			
Age						
less than 20	76 (0.14)	56 (0.13)	163 (0.18)	171 (0.19)	180 (0.19)	162 (0.17)
20 - 25	236 (0.43)	185 (0.43)	281 (0.31)	302 (0.34)	366 (0.38)	333 (0.35)
26 and above	230 (0.43)	188 (0.44)	457 (0.51)	419 (0.47)	412 (0.43)	463 (0.48)
Marital Status						
Not Married	297 (0.55)	223 (0.52)	49 (0.05)*	122 (0.14)	18 (0.02)	23 (0.02)
Married	245 (0.45)	206 (0.48)	827 (0.92)	752 (0.84)	939 (0.98)	935 (0.98)
Religion						
Christian	499 (0.92)	386 (0.90)	468 (0.52)*	514 (0.58)	20 (0.02)	8 (0.01)
Muslim	10 (0.02)	13 (0.03)	302 (0.34)	214 (0.24)	936 (0.98)	941 (0.98)
Others	33 (0.06)	30(0.07)	122 (0.14)	160 (0.18)	1 (0.00)	9 (0.01)
HH Mosquito Nets						
Yes	509 (0.94)	396 (0.92)	820 (0.91)	801 (0.89)	814 (0.85)	878 (0.92)
No/ Not Sure	33 (0.06)	33 (0.08)	61 (0.07)	73 (0.08)	139 (0.15)	73 (0.08)
			ANC indicators at enrollment			
Gravidity						
1	106 (0.20)	75 (0.18)	171 (0.19)	180 (0.20)	211 (0.22)	160 (0.17)
2-4	268 (0.49)	215 (0.50)	408 (0.45)	436 (0.49)	429 (0.45)	453 (0.47)
5 and above	168 (0.31)	139 (0.32)	322 (0.36)	278 (0.31)	317 (0.33)	344 (0.36)
Parity						
0	117 (0.22)	90 (0.21)	183 (0.20)	192 (0.21)	218 (0.23)*	163 (0.17)
1	104 (0.19)	84 (0.20)	136 (0.15)	153 (0.17)	165 (0.17)	174 (0.18)
2 and above	321 (0.59)	255 (0.59)	582 (0.65)	548 (0.61)	574 (0.60)	619 (0.65)
Gestation (weeks)						
20 and below	286 (0.53)	229 (0.53)	335 (0.37)*	291 (0.33)	620 (0.65)*	661 (0.69)
21 - 26	164 (0.30)	123 (0.29)	419 (0.47)	481 (0.53)	308 (0.32)	274 (0.29)
27 and above	92 (0.17)	77 (0.18)	147 (0.16)	122 (0.14)	30 (0.03)	23 (0.02)
			Village Characteristics			
Number of Villages	15	15	15	15	15	15
Number of HC	7		13		5	
Distance to HC$^+$						
5km and below	4 (0.27)	5 (0.33)	9 (0.60)	8 (0.53)	0 (0.00)	3 (0.20)
Above 5km	11 (0.73)	10 (0.67)	6 (0.40)	7 (0.47)	15(1.00)	12 (0.80)

Table 2: Baseline Characteristics by Country

$^+$ The reported Gambian distance from the nearest health center to a study village is based on DSS fieldworker estimates.

* control and intervention baseline characteristics are significantly different. **HC** Health Centers **HH** Household Proportion of participants in brackets

3.1.2 Assessing CHW Visits

CHW monthly visits were only implemented in the intervention villages. We observe that approximately 66% of all participants were visited at least 3 times. The CHW coverage was 87% , 93% and 97% in Benin, Burkina Faso and The Gambia. Table 3 gives a summary of CHW activities.

	Benin	**Burkina Faso**	**Gambia**	**Consolidated**
CHW Visits	1293	2395	3318	7006
HH Visited	474	844	932	2250
Participants	542	901	958	2401
Coverage (%)	87	93	97	94
CHW visits/household	2.72	2.84	3.56	3.11
CHW visits (Participants)				
At least 1 visit	474 (0.87)	844 (0.93)	932 (0.97)	2250 (0.93)
At least 2 visits	397 (0.73)	754 (0.84)	887 (0.92)	2038 (0.84)
At least 3 visits	279 (0.51)	549 (0.61)	767 (0.80)	1595 (0.66)
4 visits and above	123 (0.23)	209 (0.23)	522 (0.54)	854 (0.35)

Table 3: Coverage by CHW (Intervention Arm)
HH Household;Proportion of participants in bracket.

3.2 Effect of the CSST program on ANC Attendance

This study assesses the impact of the CSST program in the intervention community on gestation time to first ANC visit and ANC attendance recorded when compared to the control community. For comparison, we categorize ANC visits where IPTp-SP was administered as scheduled visits and ANC visits where IPTp-SP was not administered as unscheduled visits. Table 4 presents a summary of ANC visits by country.

3.2.1 Gestation Time to First Antenatal Care

In the prevention of MiP, weeks 13 to 20 are regarded as critical. Prevention of malaria between this period is considered as a first step to a malaria free pregnancy experience[5]. The event of interest is the gestation time to first ANC attendance within the second trimester of pregnancy. Figures 2 shows the Kaplan Meier (KM) curves by trial arms in Benin, Burkina Faso and The Gambia. The KM curves are a non-parametric approach to the estimation of survival times for time-to-event-data. For convenience in plotting, we shift the time scale to the left by 8 weeks. The survival plots suggest that there seems to be no difference in the gestation time to first ANC between the the intervention and control community in all countries.

3.2.2 Number of ANC visits

Figures 3, 4 and 5 show the country distribution of scheduled and unscheduled ANC visits by the proportion of pregnant women in each community.The graphs suggest no difference in the proportion of pregnant women and their pattern of ANC attendance for scheduled and unscheduled visits in each country. However, when we consider the total number of ANC visits during pregnancy, we see a difference in Burkina Faso suggesting that a higher proportion of pregnant women from the intervention community attended at least four ANC sessions when compared to the control community.

	Benin		**Burkina Faso**		**Gambia**	
	Intervention	**Control**	**Intervention**	**Control**	**Intervention**	**Control**
ANC visits (%)						
Scheduled	838 (0.91)	680 (0.92)	2394 (0.66)	2299 (0.67)	1732 (0.70)	1705 (0.73)
Unscheduled	85 (0.09)	60 (0.08)	1223 (0.34)	1127 (0.33)	738 (0.30)	625 (0.27)
Total	923	740	3617	3426	2470	2330
Scheduled visits(% of participants)						
Participants	542	429	901	894	958	958
At least 1	542 (100)	429 (100)	898 (99)	892 (99)	952 (99)	954 (99)
At least 2	296 (54)	249 (58)	806 (89)	773 (86)	734 (77)	721 (75)
At least 3	0 (0)	2 (0.04)	453 (50)	408 (46)	44 (5)	29 (3)
4 and above	0 (0)	0 (0)	200 (22)	182 (20)	2 (0.2)	1 (0.1)
Unscheduled visits(% of participants)						
At least 1	72 (13)	52 (12)	560 (62)	540 (60)	413 (43)	385 (40)
At least 2	10 (2)	7 (2)	330 (37)	296 (33)	181 (19)	137 (14)
At least 3	3 (0.6)	1 (0.2)	183 (20)	154 (17)	75 (8)	58 (6)
4 and above	0 (0)	0(0)	95(11)	77(9)	37 (4)	27 (3)
Total visits (% of participants)						
At least 3	53(10)	49 (11)	763 (85)	705 (78)	373 (39)	345 (36)
At least 4	10 (2)	7 (2)	553 (61)	490(55)	165 (17)	129 (13)
At least 8	0 (0)	0 (0)	23 (3)	24 (3)	8 (1)	5 (1)

Table 4: ANC Attendance by Country

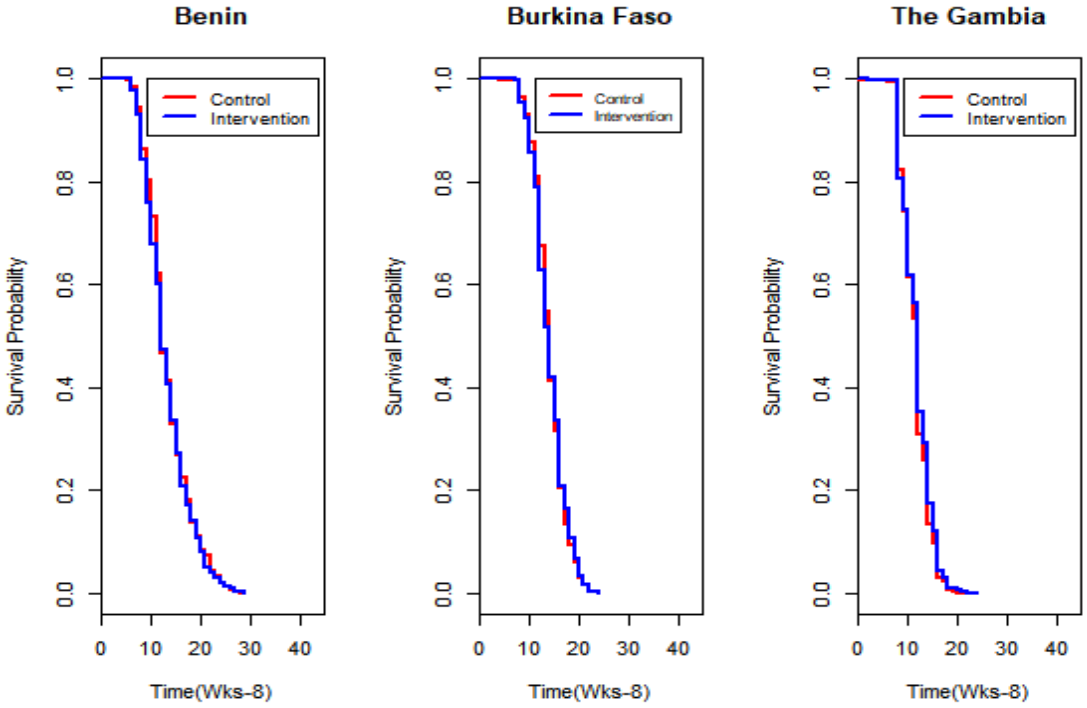


Figure 2: Gestation Time to First ANC (KM Estimates)

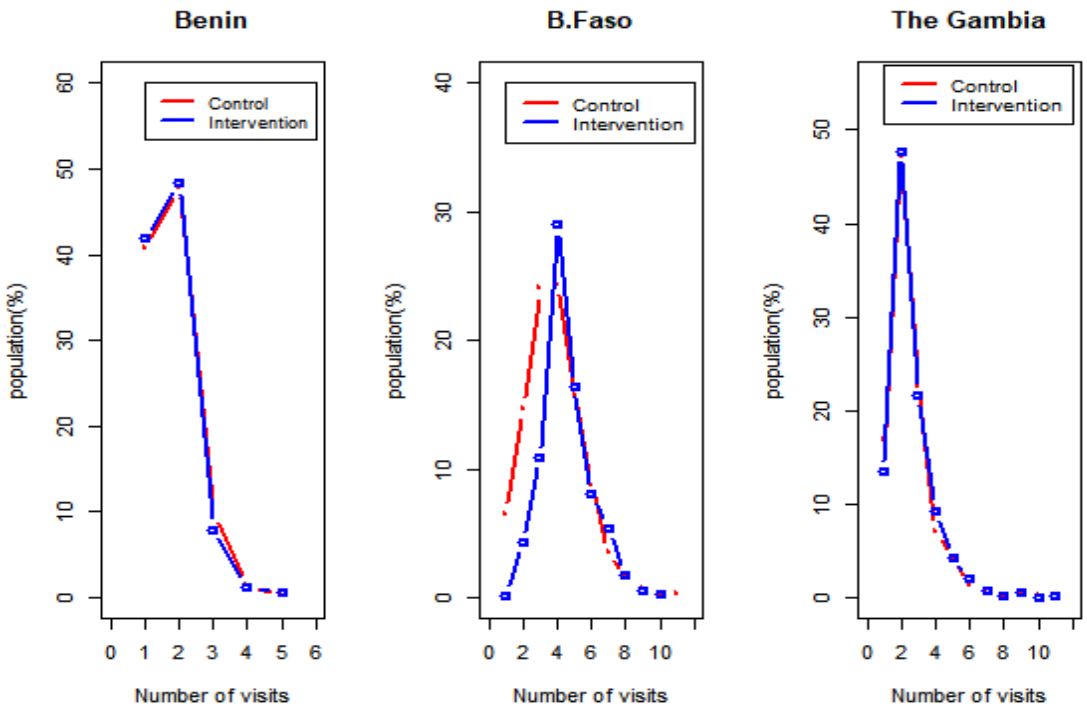


Figure 3: Total ANC Attendance

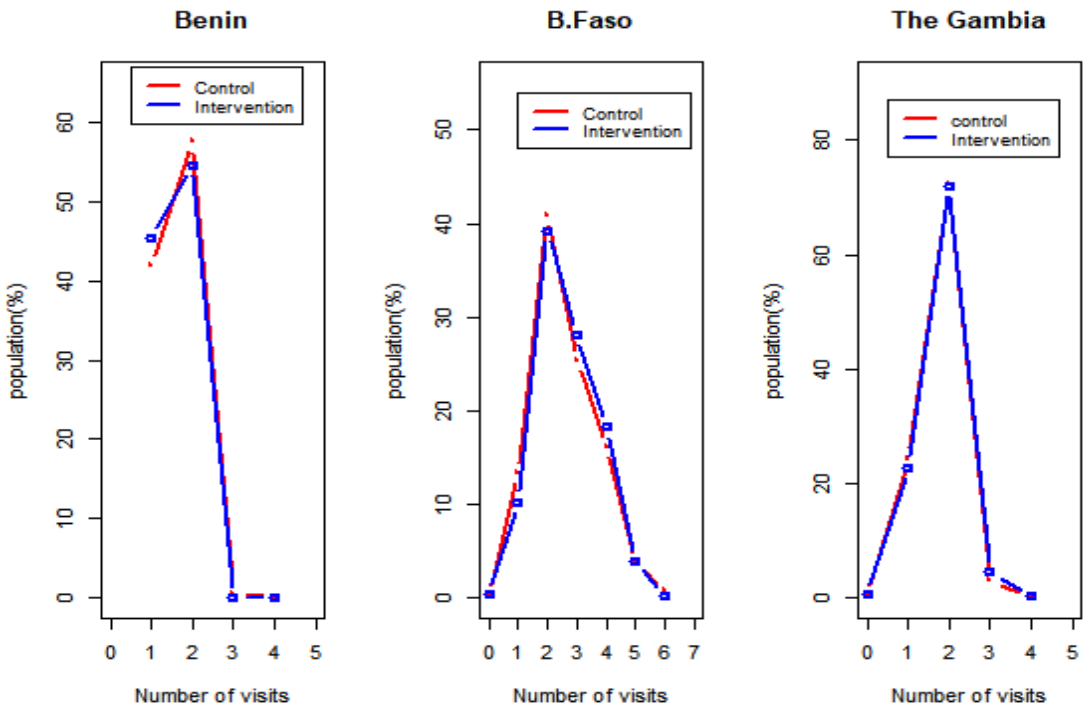


Figure 4: Scheduled ANC Attendance

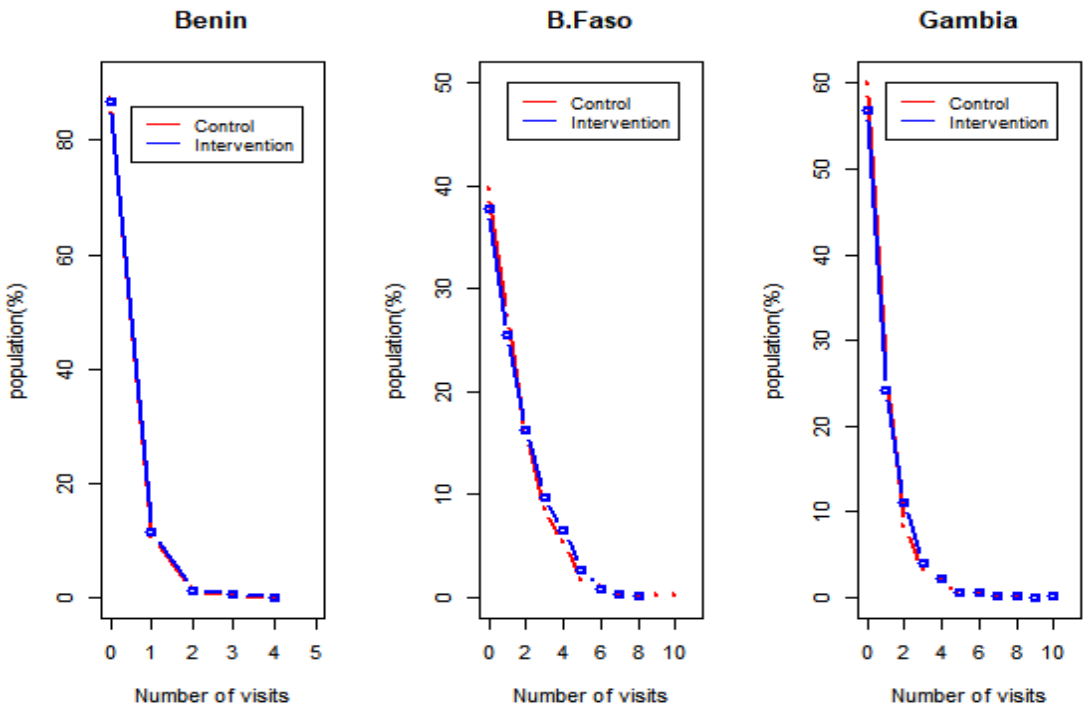


Figure 5: Unscheduled ANC Attendance

3.3 Variable Selection

The KM survival estimates assume independence of observations and do not take into account other factors which influence the gestation time to the first ANC. Literature review shows variables such as age of pregnant mother, religion, marital status, distance from the health facility, educational status, gravidity, parity, number of children alive, number of children dead and, socio-economic status (SES) as important covariates when modeling gestation time to first ANC visit or ANC attendance[7, 8, 9, 10]. Similar covariates were captured for participants during recruitment into the study with the exception of the educational status. Proxies were captured for the socio-economic status (SES) of the participants.

3.3.1 Health Center Effect as a variable

Table 2 shows the number of health facilities used in the study for each country. Health centers were either stand-alone (attended by only one village) or combined (attended to more than one village). This implied that some health centers attended to both trial communities. Only Burkina Faso had both stand-alone and combined health centers. We modeled the center structure in Burkina Faso by categorizing health centers based

on the trial communities from which individuals were attending and taking the centers with attendees from both arms as a single group. Figure 5 show the different KM curves corresponding to health centers in Burkina Faso.

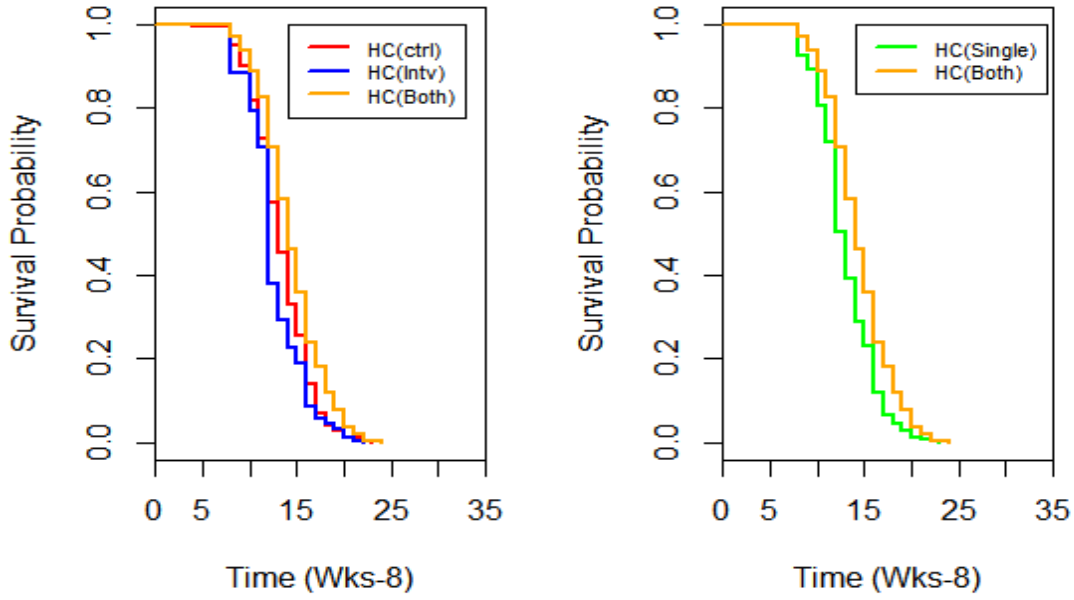


Figure 6: Time to ANC by Health Center(Burkina Faso)

It suggests that centers which are attended by both communities have longer gestation times to first ANC visit compared with the single health centers. This was modeled in the study as an explanatory variable.

3.4 Check for Multicollinearity

The available covariates were also checked for multicollinearity using correlation matrix which are graphically depicted in Figures 7, 8 and 9. The variables age, gravidity, parity, number of children alive and number of children dead were observed to be highly correlated.The age variable was categorized into subgroups of interest as shown in Table 2. The variables parity, the number of children alive and the number of children dead were dropped as model variables due to high multicollinearity with the variable gravidity.

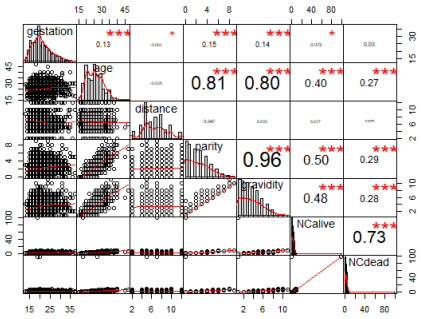


Figure 7: Variable Selection: Correlation Matrix (Benin)

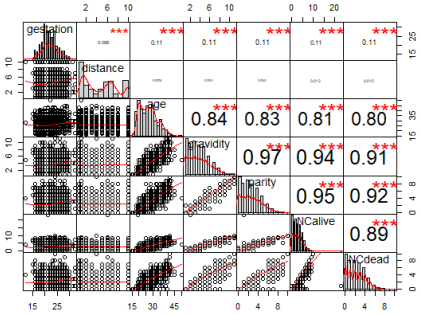


Figure 8: Variable Selection: Correlation Matrix (B.Faso)

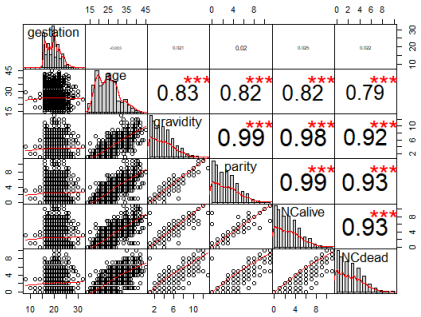


Figure 9: Variable Selection: Correlation Matrix (The Gambia)

3.4.1 Explanatory Variables

Explanatory variables used in this study are summarized as follows;

- **CSSTgrp**: Intervention community (1) or control community (0), *binary*.

- **HCenter**: Health center effect, Health Center Single (0), Health Center both(1), *binary.*
- **distance**: Distance to nearest health center, continuous.
- **malSP**: Malaria SP resistance status of the country , moderate (0), high (1), *binary.*
- **malEND**: Malaria endemicity status of the country, low(0), high(1), *binary.*
- **age**: Age of pregnant women at time to first ANC visit, *continuous.*
- **agecat**: Below 20 years (0), between 20 and 25 years (1), above 25 years (2), *categorical.*
- **mstatus**: Marital status at time of recruitment into the study married(1), unmarried (0), *binary.*
- **gravidity** : Number of times participant has been pregnant, *count.*
- **religion**: Religious affiliation of participant, Christian (0), Muslim (1), others (2), *categorical.*
- **npatients**: Number of patients in each village, *count.*

3.5 Data Structure

In this study we have a 3 level hierarchically ordered data: pregnant women or individuals (level-1), village (level-2) and country (level-3). Pregnant women in the same village are expected to behave more similarly as compared with pregnant women from other villages. This implies some correlation within pregnant women from each the same village and can be referred to as clustering[11].An implication of clustering is that the similarity within the villages reduces response variability which could affect the power to detect true differences between the intervention and control arms[12].

4 Methodology

We review methods of analysis used to investigate the research questions which answer the question and adjust for the data structure.Two levels of analysis are done in this study. We first investigate the research questions within the three countries (Individual Country Analysis) and then investigate the research questions across the countries, (a review with consolidated data from the countries), where appropriate. For uniformity, we specify Individual Country Models and later specify across country models with their results in Section 5.

In section 4.1 we review two models fit to investigate the first research question, Gestation Time to the first ANC visit. Section 4.2 discusses the models fit for research questions 2 and 3. We fit three models for each question. We give an overview of the model and discuss clustering and how the models account for it.The variable 'HCenter' was considered only for Burkina Faso as the variable was not applicable to other countries. It is highlighted in blue.

4.1 Gestation Time to First ANC

Survival analysis refers to statistical techniques which are used in the analysis of time to event data[13]. Survival Analysis techniques make inference about the time to event of interest from a properly defined start point[13]. Data used for survival analysis are characterized by a non-negative response time and a censoring indicator. For our study, we define our response T_{ij} as the gestation time to the first ANC visit of the i^{th} pregnant woman from the j^{th} village. Censoring in survival analysis occurs when the time to an event of interest is not observed for an individual. This could be because the event did not occur, the event occurred at a later date after the study or the individual dropped out from the study[13, 14]. There was no censored data in this study as the time to the event of interest,first ANC visit, was observed occurred for all pregnant women in the study.

Survival analysis methods depend on the assumption of the survival time distribution. The distribution is usually expressed in terms of the survival function, *S(t)* . The survival function is defined as the probability of surviving up to a time point t and is expressed as

$$S(t) = P(T \geq t), 0 < t < \infty \tag{1}$$

The value of this function lies between zero and one. This can be seen as 1-*F(t)* where F(.) is the cumulative density function of the time variable T[14]. The survival function *S(t)* can be expressed in terms of the hazard function *h(t)* or the Cumulative Hazard Function *H(t)*.

$$h(t) = \frac{f(t)}{S(t)} \tag{2}$$

$$H(t) = -\log S(t) \tag{3}$$

where *f(t)* is the probability density function of T and *h(t)* is the probability that an event occurs within a small time interval $(t + \delta)$ given that the event is yet to occur at time t [14].

4.1.1 Cox Proportional Hazards Model

The choice of the model for analysis of survival data is dependent on assumptions made about the distribution of the survival (gestation) times [13, 14].In this study we fit the Cox Proportional Hazards (PH) model which makes no assumptions about the distribution of the survival times and is also able to model the effects of the covariates on the hazards for the event of interest.The explanatory variables are assumed to act multiplicatively on the baseline hazard function $h_0(t)$ inducing the hazards to be proportional[15,16]. We specify the hazards for the i^{th} individual in the j^{th} village (cluster) as

$$h_{ij}(t|\mathbf{Z}_{ij}) = h_0(t)exp\mathbf{Z}'_{ij}\boldsymbol{\beta} = h_o(t) * exp(\tag{4}$$

$$\beta_1 CSSTgrp1_j + \beta_2 distance_j + \beta_3 Gravidity_{ij} + \beta_4 mstatus1_{ij}$$

$$+\beta_5 agecat1_{ij} + \beta_6 agecat2_{ij} + +\beta_7 religion1_{ij} + \beta_8 religion2_{ij}$$

$$+\beta_9 HCenter1_j + \beta_{10} CSSTgrp1_j * HCenter1_j)$$

where i = 1, 2, ..., n_j , j = 1, 2, ..., 30. h_0(t) is the baseline hazard which corresponds to the hazards for a participant with all zero model covariates and n_j is the number of pregnant women in village j. $\boldsymbol{\beta}$ is the vector of parameter estimates. The dummy variables in the vector of covariates $\mathbf{Z}$ are defined below

$$CSSTgrp1_j = \begin{cases} 1, & \text{if village j is an intervention village} \\ 0, & \text{otherwise} \end{cases}$$

$$HCenter1_j = \begin{cases} 1, & \text{if the health center in village j attends to both trial communities} \\ 0, & \text{otherwise} \end{cases}$$

$$mstatus1_j = \begin{cases} 1, & \text{if individual i in village j is unmarried} \\ 0, & \text{otherwise} \end{cases}$$

$$agecat1_{ij} = \begin{cases} 1, & \text{if individual i in village j is between 20 - 25 years} \\ 0, & \text{otherwise} \end{cases}$$

$$agecat2_{ij} = \begin{cases} 1, & \text{if individual i in village j is 26 years and above} \\ 0, & \text{otherwise} \end{cases}$$

$$religion1_{ij} = \begin{cases} 1, & \text{if individual i in village j is a Muslim} \\ 0, & \text{otherwise} \end{cases}$$

$$religion2_{ij} = \begin{cases} 1, & \text{if individual i in village j is other faith} \\ 0, & \text{otherwise} \end{cases}$$

An alternative to the Cox PH regression model is the Accelerated Failure Time (AFT) models. AFT models assume a distribution of the survival times and also model the effect of covariates on the survival time. The log linear presentation of the model is used more often and has been adopted by most computer software packages[15,16]. It is specified as follows

$$\log(T_{ij}) = \mu + \mathbf{Z}'_{ij}.\boldsymbol{\beta} + \sigma\epsilon_{ij} \tag{5}$$

where μ and σ are parameters known as the intercept and scale respectively. ϵ_{ij} is a random variable (error term) which is independent and identically distributed. ϵ_{ij} used to model deviations from $\log(T_{ij})$. For T_{ij} following a Weibull distribution,ϵ_{ij} is an extreme distribution known as the gumbel distribution and for T_{ij} following a log-normal distribution, ϵ_{ij} follows a standard normal distribution [15, 16].
In this study we define our AFT model as

$$\log(T_{ij}) = \beta_0 + \beta_1 CSSTgrp1_j + \beta_2 distance_j + \beta_3 Gravidity_{ij} + \beta_4 mstatus1_{ij} \tag{6}$$

$$+\beta_5 agecat1_{ij} + \beta_6 agecat2_{ij} + +\beta_7 religion1_{ij} + \beta_8 religion2_{ij}$$

$$+\beta_9 HCenter1_j + \beta_{10} CSSTgrp1_j * HCenter1_j + \sigma\epsilon_{ij}$$

where i = 1, 2, ..., n_j , j = 1, 2, ..., 30.
Both models have different motivation, parameter interpretation and assumptions. Parameters from the Cox PH model describe the effect of the explanatory variables on the hazard function while parameters from the AFT model give absolute effects of changes in the explanatory variables on the survival time[13]. The AFT models allow for a higher precision in estimation and they are helpful in predicting survival times beyond

the observed times[14].The drawback is that it requires the correct specification of the distribution.These assumptions are sometimes difficult to confirm hence the continued popularity of the Cox PH model[15, 16]

This does not rule the cox proportional model from drawbacks as a key assumption for the model is the proportional hazards assumption. It assumes the hazard ratio across levels of a group is the same over time. This assumption is central to the acceptance of estimates from the model.If this assumption is not met, AFT models can be used[15,16].

Both models assume independence of the survival times. In the Cox PH model, the effect of clustering is corrected for by calculating the robust variance estimators using the jacknife variance estimators[16,21].Similar to the sandwich estimator, the jacknife estimator provides true variance estimates for correlated data when an independent observation is left out[16]. We believe the same situation applies in the AFT model as in *RStudio* robust standard errors for both models are achieved by using the term *cluster()* on the village variable in the model statement[16,21].

4.2 Number of ANC visits

We investigate research questions 2 and 3 by fitting a logistic regression model with GEE and a beta binomial model.

4.2.1 At least 4 ANC visits

Logistic Regression Model

The logistic regression model estimates the probability that an event is successful for binary data given explanatory variables.The model estimates the linear relationship which exists between the logit of the mean response and the explanatory variables in the model. A key assumption with this model is that observations are independent[11,17]. When applied to clustered or correlated data , a logistic regression model provides parameter estimates which have invalid standard errors. This is due to possible overdispersion. Overdispersion occurs when there is evidence that the model variance is higher than the expected variance of the distribution[11,17]. If overdispersion is present- due to clustering or omitted covariates- adjustments need to be made to the standard errors as the test statistics and overall goodness-of-fit will be distorted[18]. We fit the logistic model using Generalized Estimating Equations to adjust for the clustering. Our model is expressed below as

$$Y_j \sim B(n_j, \pi_j) \tag{7}$$

$$logit(\pi_j) = \beta_0 + \beta_1 CSSTgrp1_j + \beta_2 distance_j + \beta_3 * HCenter1_j \tag{8}$$

where Y_j = the number of pregnant women with at least 4 ANC visits during pregnancy

in village j, n_j is the number of pregnant women in village j and π_j is the probability of success to observe pregnant women in village j who had at least 4 ANC visits).

We fit a logistic regression model using generalized estimation equations (GEE). Proposed by Liang and Zeger[11,18,19], the basic idea of this approach is to model the mean response without modeling the covariance structure in the response.The approach requires correct specification of the univariate marginal distribution of the response and assuming a 'working' assumption about the the associative covariance structure observed.[11, 19]. To make inferences, the asymptotic normality of the estimators together with the sandwich-estimated covariance matrix are used.

An advantage of the GEE estimates is that model parameter estimates are valid even with misspecification of the correlation structure and more appropriate standard errors (robust standard errors) result from adjustments by the GEE method[11,19]. Various 'working' correlation structures can be assumed in GEE. Commonly used working covariance are the independent, exchangeable and unstructured. In this study, the exchangeable and independent working correlation seem appropriate for our data. However issues with convergence of the model limited our choice of working correlation to the independent working correlation.

Beta Binomial Regression Model

The Beta-Binomial regression model has become an alternative to model over-dispersed binary data[18]. Based on the work of Skellam , Kleinman and Williams[18], we assume that the probability π_j of a pregnant women in the j^{th} village having our event of interest is random and follows a beta distribution.

$$\pi_j \sim beta(\alpha_j, \beta_j) \tag{9}$$

having mean

$$E(\pi_j) = \frac{\alpha_j}{\alpha_j + \beta_j} \tag{10}$$

Conditional on π_j the number of women who the desired event of interest in the j^{th} village follows a binomial distribution. This can be formally expressed as

$$y_j|\pi_j \sim B(n_j, \pi_j) \tag{11}$$

where n_j is the total number of pregnant women in the j^{th} village.
The intra cluster correlation ρ_i has been shown to be equal to

$$\rho_j = \frac{1}{\alpha_j + \beta_j + 1} \tag{12}$$

The mean structure of the beta binomial model is similar to that of the logistic model with the logit function linking π_j to covariates[20].

4.2.2 At least 3 Scheduled Visits

The methods used are consistent with section 4.2.1. We change our definition of Y_j to be the number of women in village j who had at least 3 scheduled visits during pregnancy.

4.3 Model Selection

We adopt basic statistical tests in our analysis. Model fit checks using the likelihood ratio test (LRT) and the Akaike Information Criterion (AIC) are performed in this study. LRT is used to compare the significance of having an extra explanatory variable in the model. It tests the hypothesis that the value of the coefficient of the term is not statistically different from Zero and should not be added to the model . Akaike Information Criterion (AIC) is also used to check the fit of a model. The model with a lower AIC value is considered to have a better fit. The AIC is obtained by

$$AIC = -2\log L + 2K \tag{13}$$

where K = number of parameters in the model and $\log L$ is the log likelihood value. LRT is used for comparing nested models while AIC is used for comparing non-nested model fits[14].
All explanatory variables which were available and relevant to the study based on literature review were modeled. Interaction terms were obtained firstly by stepwise variable selection approach in `R` and reviewed for plausibility. Non-significant interaction terms during the analysis were dropped.

4.4 Statistical Analysis

All analysis were done using *RStudio Software*. Statistical inference were taken at two sided $\alpha = 0.05$. Complete case analysis was performed.

5 Results and Interpretation

We present the results and interpretation to the analysis below.

5.1 Gestation Time to First ANC

5.1.1 Exploratory Univariate Analysis

Figure 10 shows the distribution of the gestation time for the three countries, Benin, Burkina Faso and The Gambia. The responses from Benin and The Gambia look skewed to the right. The median time to first ANC visit is about 20 weeks for Benin and The Gambia and 22 weeks for Burkina Faso.

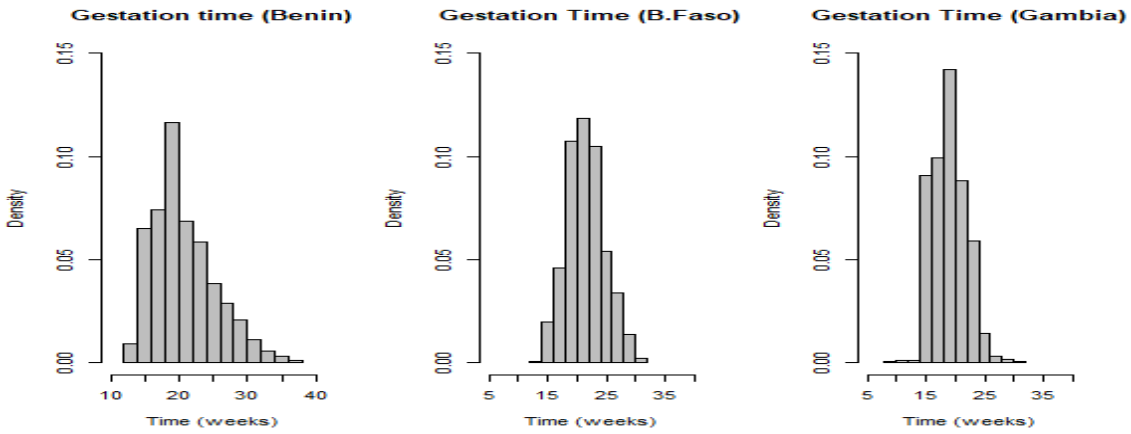


Figure 10: Distribution of Gestation Time to First ANC

5.1.2 Cox Proportional Hazards Model

Model Selection

Our objective is to model the time to the first ANC visit for pregnant women in the study. Our interest is to see if pregnant women in the intervention community have shorter time to their first ANC visits compared to pregnant women from the control community. Variables used in the model were obtained as specified in section 3.3 and 3.4.

Model Specification

A model was fit for each country to investigate the time to the first ANC visit. Except for the interaction terms which could differ among the different countries, the same variables were included in all country specific models. The model specifications are specified as in equation 4.

Interpretation of Individual Country Results

The parameter estimates and robust standard errors of the Cox PH model are summarized in Table 5. The overall effect of categorical variables with a significant level coefficient is first ascertained using the Likelihood Ratio Test (LRT) before interpretation can be made. Categorical variables with overall effects which are not significant are not interpreted. In general positive coefficients ($\beta > 0$) are associated with an increased hazard, in our case, a shorter gestation time to first ANC visit and negative coefficients indicate longer times to first ANC visit.

In Benin, we observe no significant difference between the intervention community and the control community in terms of the gestation time to first ANC visit. We observe that the variable gravidity is a significant determinant of the time to the first ANC visit. A unit increase in gravidity decreases the hazard associated with gestation time to first ANC visit by 5.4% ($e^{-0.0505}$). We associate participants with higher gravidity with a lower chance of having an early time to their first ANC visit. The test for the overall effect of the categorical variable religion in the model was not significant(p-value = 0.1026) meaning that religion does not significantly influence the time gestation time to first ANC visit.

In the Gambia, distance to the health center is a significant determinant of the gestation time to the first ANC visit. We observe that for a unit increase in the distance to the health center, the hazards associated with gestation time to first ANC visits is increased by 1.5% ($e^{0.0149}$). This position is counter-intuitive.There was no significant difference between the intervention and control community.

	Cox PH Model		
Parameter	**Benin**	**The Gambia**	**Burkina Faso**
CSSTgrp	0.0335 (0.1018)	-0.0582 (0.0659)	0.2986 (0.1536)·
distance	0.0170 (0.0251)	0.0149 (0.0050)**	-0.0161 (0.0146)
gravidity	-0.0550 (0.0170)**	-0.0205 (0.0126)	-0.0387 (0.0162)*
mstatus_unmarried	0.0649 (0.0706)	0.1687 (0.1324)	0.0861 (0.1395)
agecat_20-25	0.0158 (0.0827)	0.0006 (0.0629)	-0.1536 (0.0880)·
agecat_26 above	-0.0401 (0.1088)	0.0692 (0.0795)	-0.1492 (0.0954)
religion_muslim	0.4849 (0.1989)*	0.2341 (0.1945)	-0.1137 (0.0647)·
religion_others	-0.0186 (0.1590)	-0.2428 (0.3357)	-0.0413 (0.0726)
HCenter(Both)	-	-	-0.3003 (0.0983)**
CSSTgrp:HCenter(Both)	-	-	-0.2815 (0.1803)

Significance codes :p-value 0.1 ‘ ’ 0.05 ‘.’ 0.01 ‘*’ 0.001 ‘’ 0 ‘***’**

Table 5: Cox PH Parameter Estimates for Benin, Burkina Faso and Gambia

In Burkina Faso, we observe a borderline significant difference (p-value = 0.0518) between the intervention community and the control community. Since we have an inter-

action term between the community and the health center status, we first ascertain if the overall effect of the interaction term is contributing to the observed position. The interaction term is significant (p-value = 0.0150). Gravidity is also seen to be a significant predictor for the gestation time to the first ANC visit. It would be inappropriate to interpret the significance of the HCenter variable as the effect is influenced by the interaction term therefore we interpret the main effect of the HCenter conditional on the CSST group membership.

In summary, the expected effect of the CSST program on earlier time to the first ANC visit for pregnant women in the intervention community was not observed in Benin and The Gambia. In Burkina Faso, an intervention effect on gestation time to first ANC was observed when only villages with standalone health centers in both communities were considered.

5.1.3 Accelerated Failure Time (AFT) Models

Figures 11, 12 and 13 are KM survival curves fr the gestation time to first ANC. The graphs suggest that either the Weibull or log-normal distributions would fit the survival distributions for the three countries. These graphs are strictly exploratory and should not be over interpreted as the KM curves do not consider the effects of covariates on the time variable of interest.

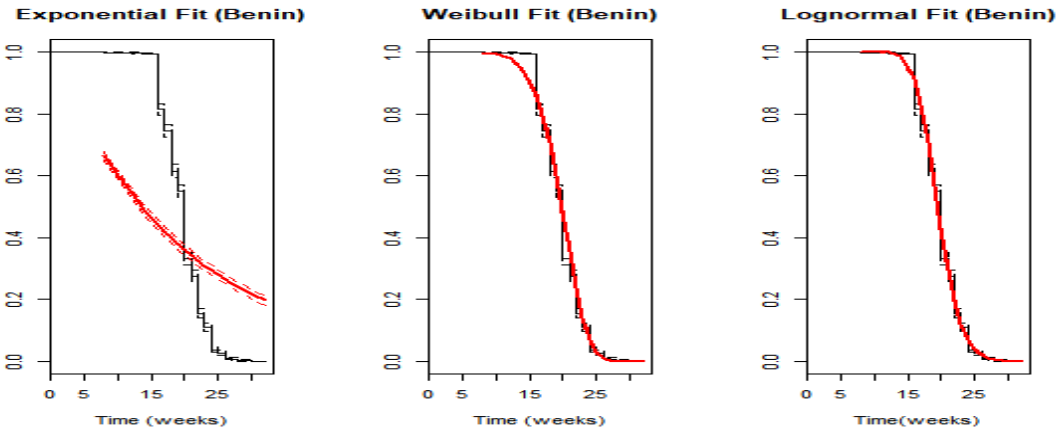


Figure 11: KM Curve for Gestation Time to First ANC (Benin)

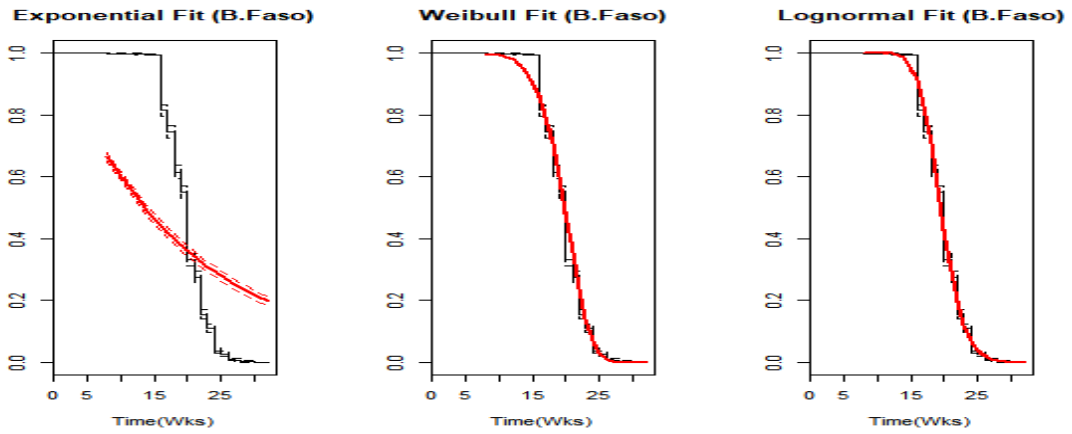


Figure 12: KM Curve for Gestation Time to First ANC(B.Faso)

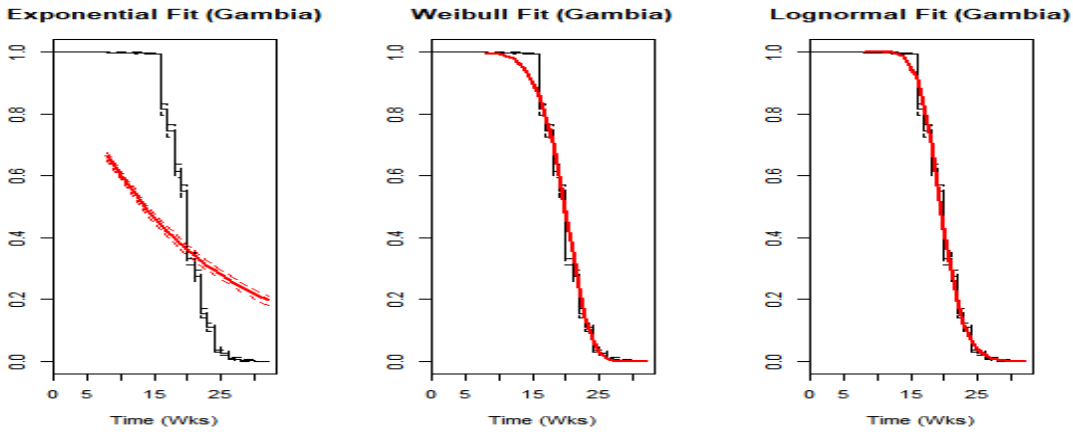


Figure 13: KM Curve for Gestation Time to First ANC (The Gambia)

Model Selection

A more formal approach to choose the appropriate distribution for T would be to use a richer family of models, (which should include our models of interest as sub-models), and carry out some formal test for fit which would enable us ascertain the suitability of the sub-model fit for the survival data[21]. The LRT tests would be used to test the model fit . We attempted to use the Generalized F and Generalized Gamma distribution to our data but we experienced convergence challenges. The AIC from fitting the log-normal and Weibull models correcting for covariates was used instead as a formal approach to show the difference between both distributional assumptions. Table 6 shows the

corresponding AIC values for all the countries. The Lower AIC is better hence we fit log-normal AFT models for the countries.

	AIC	
Country	**Weibull**	**Log-normal**
Benin	5780.87	5564.33
Gambia	9600.51	9335.55
Burkina Faso	9088.84	8906.01

Table 6: AIC for AFT Models Fit

Model Specification

A log-normal AFT model was fit for each country to investigate the time to the first ANC visit. The model specifications are specified in equation 6.

Interpretation of Individual Country Results

Table 8 is a summary of parameter estimates with robust standard errors corresponding to Benin, The Gambia and Burkina Faso. We review the results form the AFT model vis-à-vis the results from the Cox PH model results. Positive coefficients $\beta > 0$ imply an increase in the time to first ANC visit while negative coefficients, $\beta < 0$ indicate a decrease in the time. For this study, a negative coefficient for the intervention arm coefficient (CSSTgrp) implies an effect of the CSST program.

	Parameter Estimates AFT Models		
Parameter	**(Benin)**	**(The Gambia)**	**(B.Faso)**
CSSTgrp	-0.0132 (0.0211)	0.0057 (0.0093)	-0.0421 (0.0234)
distance	-0.0069 (0.0058)	-0.0020 (0.0007)**	0.0017 (0.0020)
gravidity	0.0127 (0.0037)***	0.0021 (0.0017)	0.0051 (0.0022)*
mstatus_unmarried	-0.0143 (0.0150)	-0.0095 (0.0175)	-0.0189 (0.0185)
agecat_20-25	-0.0057 (0.0177)	0.0076 (0.0090)	0.0152 (0.0127)
agecat_26 above	0.0069 (0.0222)	0.0005 (0.0116)	0.0162 (0.0137)
religion_muslim	-0.0762 (0.0363)*	-0.0201 (0.0282)	0.0046 (0.0100)
religion_others	0.0125 (0.0352)	0.0603 (0.0540)	0.0087 (0.0105)
HCenter_both	None	None	0.0467 (0.0171)**
CSSTgrp:HCenter	None	None	0.0342 (0.0273)
Intercept	3.0509 (0.0426)***	2.9989 (0.0324)***	3.0152 (0.0181)***
scale	0.203	0.142	0.143

Significance codes :p-value 0.1 ‘ ’ 0.05 ‘.’ 0.01 ‘*’ 0.001 ‘’ 0 ‘***’**

Table 7: AFT Parameter Estimates for Benin, Burkina Faso and Gambia

In Benin we see that gravidity is a significant predictor for the gestation time to the first

ANC visit.Conditional on all other factors being fixed, with a unit increase in gravidity, the gestational time to first ANC visit is lengthened by a factor 1.012 ($e^{0.0127}$). In The Gambia, we observe that distance is a significant predictor for gestational time to the first ANC visit. For significance of categorical variables in Burkina Faso we look at the Table 8. We first investigate the interaction term between the treatment community and the health center status.The overall effect is significant.

Test for fixed Effects (Burkina Faso)			
Parameter	**DF**	Deviance	P-Value
CSSTgrp	1	0.7462	0.3876
HCenter	1	68.4412	0.0000****
distance	1	1.4610	0.2267
agecat	2	19.0279	0.0000****
gravidity	1	4.5386	0.03313
mstatus	1	1.5670	0.2106
religion	2	0.6249	0.7316
CSSTgrp:HCenter	1	4.3771	0.0363*

Significance codes :p-value 0.1 ‘ ’ 0.05 ‘.’ 0.01 ‘*’ 0.001 ‘’ 0 ‘***’**

Table 8: Test for fixed Effects

In summary, conditional on the health center,we see a significant reduction in gestational time to the first ANC visit in participants in intervention villages with stand alone health centers when compared to individuals in the control. This result is consistent with the Cox PH result indicating a positive impact of the CSST program in intervention villages with stand-alone health centers.

5.2 Number of ANC visits

5.2.1 At least 4 ANC visits to the Health Center

We present the Individual country and consolidated country results in this section.

Model Specification

The three models fit for each country are specified in equation 8.

Interpretation of Individual Country Results

The data was modeled first with the binomial logistic model to assess over-dispersion. This was done by inspection of the deviance residual divided by the degrees of freedom. In all countries, we observed that there was evidence of over-dispersion.

	Model Type (Benin)		
Parameter	**Binomial ML**	**GEE(Independent)**	**Beta-Binomial**
Intercept	-6.4833 (1.0732)***	-6.4833 (1.2924)***	-6.549 (1.3410)***
CSSTgrp	0.4484 (0.5272)	0.4484 (0.6356)	0.4646 (0.6492)
Distance	0.2978 (0.1117)*	0.2977 (0.1123)*	0.3082 (0.1363)*
Overdispersion	-	$\hat{\phi}$= 1.557	$\hat{\rho}$=0.0163
AIC	62.27	-	62.03

Significance codes :p-value 0.1 ' ' 0.05 '.' 0.01 '*' 0.001 '' 0 '***'**

Table 9: Parameter Estimates (Robust Standard Errors) for Benin

Table 9 is a presentation of model parameters estimates for Benin. We observe that there is no significant effect of the CSST program on villages in the intervention community. However we see that for the three models distance is a significant determinant for pregnant women to achieve at least 4 ANC visits during pregnancy.

	Model Type (Burkina Faso)		
Parameter	**Binomial ML**	**GEE(Independent)**	**Beta-Binomial**
Intercept	0.8071 (0.1202)***	0.8071 (0.2335)***	0.8140 (0.2050)***
CSSTgrp	0.3976 (0.1001)*	0.3976 (0.1520)*	0.3892 (0.1673)*
Distance	-0.0034 (0.0160)	-0.0034 (0.0290)	-0.0037 (0.0263)
HCenter_Both	-0.8868 (0.1223)***	-0.8868 (0.2122)***	-0.8846 (0.2019)***
Overdispersion	-	$\hat{\phi}$= 3.2470	$\hat{\rho}$=0.0307
AIC	226.86	-	204.70

Significance codes :p-value 0.1 ' ' 0.05 '.' 0.01 '*' 0.001 '' 0 '***'**

Table 10: Parameter Estimates (Standard Errors) for Burkina Faso

Table 10 is a presentation of model parameters estimates for Burkina Faso. We observe that there is a significant effect of the CSST program on villages in the intervention community when compared to the control. An interaction term between the health center and the CCST group was fit for the beta-binomial model and it was not significant and therefore dropped from the model(p-value = 0.984).

	Model Type (The Gambia)		
Parameter	**Binomial ML**	**GEE(Independent)**	**Beta-Binomial**
Intercept	-1.1262 (0.1529)***	-1.1262 (0.4149)***	-1.5580 (0.3839)***
CSSTgrp	0.1252 (0.1329)	0.1252 (0.3333)	0.0288 (0.3197)
Distance	-0.0605 (0.0104)**	-0.0605 (0.0292)*	-0.0199 (0.0254)
Overdispersion	None	$\hat{\phi}$= 6.4757	$\hat{\rho}$=0.0916
AIC	296.18	None	194.1

Significance codes :p-value 0.1 ' ' 0.05 '.' 0.01 '*' 0.001 '' 0 '***'**

Table 11: Parameter Estimates (Standard Errors)
for The Gambia

Table 11 is a presentation of model parameters estimates for The Gambia.We observe that there is no significant effect of the CSST program on villages in the intervention community when compared to the control. The GEE model shows that distance is a significant predictor of our response but distance is not a significant predictor in the beta-binomial model.

Consolidated Analysis

$$logit(\pi_{jk}) = \beta_0 + \beta_1 CSSTgrp1_{jk} + \beta_2 distance_{jk} + \beta_3 * HCenter1_{jk} + \beta_4 * malEND_k + \beta_5 * malSP_k \quad (14)$$

where π_{jk} = probability of success to observe women who had at least 4 ANC visits during pregnancy in the j^{th} village from the k^{th} country) , j = number of villages(1-30) , k=country index(1,2,3), malEND :Malaria Endemicity status of the country, low(0), high(1) , malSP : Malaria SP resistance status of the country , moderate (0), high (1).

	Model Type (Consolidated)		
Parameter	**Binomial ML**	**GEE(Independent)**	**Beta-Binomial**
Intercept	-0.6143 (0.1558)*	-0.6143 (0.4068)	-0.9983 (0.3480)*
CSSTgrp	0.2940 (0.0786)*	0.2940 (0.1597)	0.2354 (0.1618)
Distance	-0.0409 (0.0083)*	-0.0409 (0.0229)	-0.0136 (0.0181)
malSP_high	-4.097 (0.2514)*	-4.0975 (0.3478)*	-3.8640 (0.3438)*
HCenter_Both	-0.8012 (0.1181)*	-0.8012 (0.2479)*	-0.8188 (0.2424)*
malEND_high	1.578 (0.0955)*	1.5780 (0.2733)*	1.8770 (0.2238)*
Overdispersion	None	$\hat{\phi}$= 3.7428	$\hat{\rho}$=0.0574
AIC	597.5	None	462.4

Significance codes :p-value 0.1 ‘ ’ 0.05 ‘.’ 0.01 ‘*’ 0.001 ‘’ 0 ‘***’**

Table 12: Parameter Estimates (Standard Errors) for the consolidated data set

Interpretation of Consolidated data results

At the consolidated level, we observed that the status of endemicity and SP resistance of the countries were significant predictors of the response. Villages from highly endemic countries were 6 times($e^{1.8770}$) more likely to have women who had at least 4 ANC visits when compared to villages from low endemic countries. The odds of a village from a country with a high malaria SP resistance status having women who had at least 4 ANC visits was 98% ($e^{-3.8640}$) less than that for villages in countries with low malaria SP resistance status.

5.2.2 At least 3 ANC IPTp-SP visits to the Health Center

Analysis by country

Three models fit for Burkina Faso and Gambia are expressed as

$$logit(\pi_j) = \beta_0 + \beta_1 CSSTgrp1_j + \beta_2 distance_j + \beta_3 * HCenter1_j + \beta 4 CSSTgrp1_j * HCenter1_j \quad (15)$$

where π_j = probability of success to observe women who had at least 3 ANC IPTp-SP visits during pregnancy in the j^{th} village) , j = number of villages(1-30) .

	Model Type (Burkina Faso)		
Parameter	**Binomial ML**	**GEE(Independent)**	**Beta-Binomial**
Intercept	-0.2276 (0.1236)	-0.2276 (0.0973)*	-0.2212(0.2127)
CSSTgrp	0.3509 (0.1899)	0.3509 (0.1934)	0.3424 (0.3283)
Distance	0.0171 (0.0155)	-0.0171 (0.0225)	0.0175 (0.02688)
HCenter_Both	-0.0435 (0.1505)	-0.0435 (0.1630)	-0.05648 (0.2584)
CSSTgrp:HCenter_Both	-0.1868 (0.2229)	-0.1868 (0.2755)	-0.1797 (0.3852)
Overdispersion	None	$\hat{\phi}$= 3.4774	$\hat{\rho}$=0.0339
AIC	236.05	None	214.90

Significance codes :p-value 0.1 ‘ ’ 0.05 ‘.’ 0.01 ‘*’ 0.001 ‘’ 0 ‘***’**

Table 13: Parameter Estimates (Standard Errors) for Burkina Faso

	Model Type (Gambia)		
Parameter	**Binomial ML**	**GEE(Independent)**	**Beta-Binomial**
Intercept	-2.7794 (0.2905)*	-2.7794 (0.3399)*	-2.906 (0.3802)*
CSSTgrp	0.2670 (0.2535)	0.2670 (0.3133)	0.2928 (0.3239)
Distance	-0.0572 (0.0196)*	-0.0572 (0.0186)*	-0.0420 (0.0264)*
Overdispersion	None	$\hat{\phi}$= 1.8098	$\hat{\rho}$=0.0120
AIC	123.82	None	124.4

Significance codes :p-value 0.1 ‘ ’ 0.05 ‘.’ 0.01 ‘*’ 0.001 ‘’ 0 ‘***’**

Table 14: Parameter Estimates (Standard Errors) for The Gambia

Consolidated Analysis

$$logit(\pi_{jk}) = \beta_0 + \beta_1 CSSTgrp1_{jk} + \beta_2 distance_{jk} + \beta_3 * HCenter1_{jk} + \beta_4 * malEND_k \quad (16)$$

where π_{jk} = probability of success to observe women who had at least 3 scheduled ANC

visits during pregnancy in the j^{th} village from the k^{th} country) , j = number of villages(1-30) , k=country index(1,2), malEND :Malaria Endemicity status of the country, low(0), high(1) .

	Model Type (Consolidated)		
Parameter	**Binomial ML**	**GEE(Independent)**	**Beta-Binomial**
Intercept	-3.1323 (0.1976)*	-3.1323 (0.2891)*	-3.1160 (0.3317)*
CSSTgrp	0.2184 (0.0898)*	0.2184 (0.1433)	0.2134 (0.1448)
Distance	-0.0125 (0.0115)*	-0.0125 (0.0178)	-0.0034 (0.0184)
HCenter_Both	-0.0696 (0.1104)*	-0.0696 (0.1454)	-0.0919 (0.1801)
malEND_high	3.0470 (0.1491)*	3.0470 (0.2390)*	3.0070 (0.2527)*
Overdispersion	None	$\hat{\phi}$= 2.6776	$\hat{\rho}$=0.0286
AIC	365.27	None	331.7

Significance codes :p-value 0.1 ‘ ’ 0.05 ‘.’ 0.01 ‘*’ 0.001 ‘’ 0 ‘***’**

Table 15: Parameter Estimates (Standard Errors) for consolidated dataset

Interpretation of results

Tables 13, 14 and 15 present a summary of the model parameter estimates.The models for at least 3 scheduled ANC visits were fit for only Burkina Faso and the Gambia.This was because only two persons in Benin had more than 3 scheduled visits. The effect of the CSST program was not observed at the country level and in the consolidated analysis. The predictors distance and Health center were not significant in explaining the response. In the consolidated analysis, the malaria endemic status of the country was a significant predictor.

5.3 Discussion

Three research questions were investigated in this study. Our objective was to ascertain if there was an influence of the CSST program on the gestational time to the first ANC visit, the number of women who had at least for ANC visits during their pregnancy and the number of women who had at least three scheduled visits. These three outcomes are related as late gestational time to the first ANC visit could affect the chances of the pregnant women to achieve at least 4 ANC visit or at least 3 scheduled visits.

Using consolidated figures(Table 3), we see that the average coverage of CHW was approximately 94%. We also see that at approximately 66% of the participants were visited at least three times by the CHW. This shows high coverage by CHW. While the assessment of the CHW is not an objective in this study, it is important to note that they are integral in the CSST program implementation.

There was no significant difference between the gestation time to first ANC in Benin

and Gambia using the cox PH model and the lognormal AFT model. However in Burkina Faso, conditional on the Health center we see a significant difference between the intervention community and the control community for villages with standalone health centers shown by both the Cox PH model and the lognormal AFT model.The Cox PH model shows that conditioned on the health center being stand-alone and other variables being fixed, the hazard associated with the the gestational time to first ANC visit for villages in the intervention community compared to the control community is $1.34(e^{0.2986})$ times that of the control community.

Interestingly, we see that the hazard associated with the gestational time to first visit for participants from villages in the control community who attend health centers which are not standalone is 26% $(e^{-0.3003})$less than the hazard for participants from villages in the control community who attend standalone health centers. Similar results are observed with the AFT lognormal model.While this is an interesting find, we should be cautious to attribute this observed effect to just the fact that the health centers are standalone.

In reviewing the number of women who had at least 4 ANC visits during pregnancy, we see that the GEE model and the Beta-Binomial model are consistent in their results across the three countries. As both models are marginal, we can compare their outcomes even thought they use the different parameter estimation approaches namely quasi-likelihood and full likelihood approaches respectively. We observed that the effect of the Community scheduled screening and treatment(CSST) program on women who had at least 4 ANC visits during pregnancy is observed only in Burkina Faso (Table 10). We observed that the odds of women in the CSST program (intervention villages) having at least 4 visits is 48% higher than that for pregnant women in the control villages $(e^{0.3976})$ if all other covariates were held constant.

We did not observe any significant difference between the intervention community and the control community in our analysis for women who had at least 3 scheduled visits. All models fit consistently showed no impact of the CSST program on the intervention community when compared to the control community. We see this as a setback as although there was a significant effect on health seeking behavior of the pregnant women, this did not translate into an increase in the uptake of preventive IPTp-SP which is central to the CSST program objectives.

5.4 Summary and Conclusion

From the results observed, we see that the data modeled using different statistical methods and approaches were consistent in their results. We see the effect of the CSST on health seeking behavior of women in the intervention community in Burkina Faso was significantly different from the control community. Conditioned on the health Center, we also see a reduction in the gestation time to the first ANC in Burkina Faso.These effects were not observed in Benin and The Gambia.

It is however important to note that shared health centers implicitly introduce distance challenges to the villages in which the health center is not located. This would translate to pregnant women having to make health center visits not at their convenience like in villages where the health centers are located but by planning to the visits.

References

[1] World Health Organization. (2015).*Guidelines for the Treatment of Malaria*. 3rd edition. Geneva.

[2] chantz-Dunn, J., & Nour, N. M. (2009). Malaria and Pregnancy: A Global Health Perspective. *Reviews in Obstetrics and Gynecology*, 2(3), 186–192.

[3] cott, S., Mens, P. F., Tinto, H., Nahum, A., Ruizendaal, E., Pagnoni, F.,D'Alessandro, U. (2014). Community-based scheduled screening and treatment of malaria in pregnancy for improved maternal and infant health in The Gambia, Burkina Faso and Benin: study protocol for a randomized controlled trial. *Trials*, 15, 340. http://doi.org/10.1186/1745-6215-15-340.

[4] Anchang-Kimbi, J. K., Achidi, E. A., Apinjoh, T. O., Mugri, R. N., Chi, H. F., Tata, R. B.,Troye-Blomberg, M. (2014). Antenatal care visit attendance, intermittent preventive treatment during pregnancy (IPTp) and malaria parasitaemia at delivery. Malaria Journal, 13, 162. http://doi.org/ 10.1186/1475-2875-13-162.

[5] World Health Organization. (2018). Implementing malaria in pregnancy programs in the context of World Health Organization recommendations on antenatal care for a positive pregnancy experience. World Health Organization. http://www.who.int/iris/handle/10665/259954. License: CC BY- NC-SA 3.0 IGO.

[6] COSMIC consortium,(2016). Community-based malaria screening and treatment for pregnant women receiving standard intermittent preventive treatment with sulfadoxine-pyrimethamine: a multicentre (The Gambia, Burkina Faso and Benin) cluster randomised controlled trial,Unpublished Manuscript.

[7] Abosse, Z., Woldie, M., & Ololo, S. (2010). Factors Influencing Antenatal Care Service Utilization in Hadiya Zone. Ethiopian Journal of Health Sciences, 20(2), 75–82.

[8] Hajizadeh, S., Ramezani Tehrani, F., Simbar, M., Farzadfar, F. (2016). Factors Influencing the Use of Prenatal Care: A Systematic Review. Journal of Midwifery and Reproductive Health, 4(1), 544-557. doi: 10.22038/jmrh.2016.6431.

[9] Gulema, H., & Berhane, Y. (2017). Timing of First Antenatal Care Visit and its Associated Factors among Pregnant Women Attending Public Health Facilities in Addis Ababa, Ethiopia. Ethiopian Journal of Health Sciences, 27(2), 139–146.

[10] Fagbamigbe, A. F., Mashabe, B., Lepetu, L., & Abel, C. (2017). Are the timings and risk factors changing? Survival analysis of timing of first antenatal care visit among pregnant women in Nigeria (2003–2013). International Journal of Women's Health, 9, 807–819. http://doi.org/10.2147/IJWH.S138329.

[11] Agresti, A. (2013). Categorical Data Analysis, 3rd Edition. Wiley, New York.

[12] Killip, S., Mahfoud, Z., & Pearce, K. (2004). What Is an Intracluster Correlation Coefficient? Crucial Concepts for Primary Care Researchers. Annals of Family Medicine, 2(3), 204–208. http://doi.org/10.1370/afm.141

[13] Stevenson, M., & EpiCenter: IVABS Massey University. (2009, June 4). An Introduction to Survival Analysis [PDF]. Retrieved from http://www.massey.ac.nz/massey/fms/Colleges/College%20of%20 Sciences/ Epicenter/docs/ASVCS/ *Stevenson/survival/analysis*/195721.pdf.

[14] Moore, D. F. (2016). Applied survival analysis using R. Switzerland: Springer.

[15] Collett, D. (2003). Modelling survival data in medical research (2nd ed.). London: Chapman & Hall / CRC Press.

[16] Therneau, T. M., & Grambsch, P. M. (2011). Modeling survival data: Extending the Cox model. New York: Springer.

[17] Aerts,M.,(2016). Generalized Linear Models II,Unpublished Lecture Notes, 2016-2017 Session: Hasselt University.

[18] Aerts, M., Geys, H., Molenberghs, G. and Ryan, L. M. (eds) (2002). Topics in modelling of clustered data. Chapman and Hall/CRC, London.

[19] Molenberghs, G. and Verbeke, G. (2005). Models for Discrete Longitudinal Data. Springer-Verlag New York.

[20] Najera-Zuloaga, J., Lee, D., & Arostegui, I. (2017). Comparison of beta-binomial regression model approaches to analyze health-related quality of life data. Statistical Methods in Medical Research, 096228021769041. doi:10.1177/0962280217690413 .

[21] Burzykowski,T. (2017). Survival Data Analysis, Unpublished Lecture Notes, 2017-2018 session: Hasselt University.

5.5 Appendix

```
##survival

benin_surv2<-coxph(Surv(FHEIGHT_E2_C2,status)~ group + distance + factor(agecat) +
                     + factor(relcat)+ GRAVIDIT_E2_C2 + factor(mstatuscat)
                   + cluster(VILLCODE_E2_C2),data=benin_comp)
summary(benin_surv2)
cox.zph(benin_surv2)


benin_surv2<-coxph(Surv(FHEIGHT_E2_C2,status)~ group + strata(distcat) + factor(agecat) +
                     + factor(relcat)+ GRAVIDIT_E2_C2 + factor(mstatuscat)
                   + cluster(VILLCODE_E2_C2),data=benin_comp)
summary(benin_surv2)
cox.zph(benin_surv2)



g2.aftl<-survreg(Surv(FHEIGHT,status)~GROUP + factor(agecat)+ GRAVIDIT + factor(mstatuscat)
                 + factor(relcat) + cluster(VILLCODE2)  +DISTANCE
                 ,data=gambia_com_case, dist='lognormal')

summary(g2.aftl)
anova(g2.aftl)

head(beta_new)
###binomial =
bf_work11 <-glm(cbind(abT4,less4)~ group + distance + factor(new_group)
               ,family=binomial(link = "logit"),data=newbfaso)
summary(bf_work11)



###binomial =evidence of overdispersion
bf_work33 <-glm(cbind(abT4,less4)~  group + distance + factor(new_group)
               ,family=quasibinomial(link = "logit"),data=newbfaso)
summary(bf_work33)

#####GEE
require(gee)###
bf_work44 <-gee(cbind(abT4,less4)~ group + distance +
                  factor(new_group) , family=binomial(link = "logit"),
                id=code_vil, corstr="independence", data=newbfaso)
summary(bf_work44)

#####
bf_work55<-betabin(cbind(abT4,less4)~group + distance + factor(new_group)
                   ,~1,data=newbfaso)
summary(bf_work55)

AIC(bf_work55)

bf_work551<-betabin(cbind(abT4,less4)~group + distance + factor(new_group) +  factor(new_group)*group
                    ,~1,data=newbfaso)
summary(bf_work551)
anova(bf_work55,bf_work551)###factor by group not significant.
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Malaria in pregnancy (MiP)

Richting: **Master of Statistics-Biostatistics**
Jaar: **2018**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Nwogu-Ikojo, Leonard Eluma

Datum: **15/06/2018**