USAID
FROM THE AMERICAN PEOPLE

PREGNANCY HISTORY MODEL ANALYSIS PLAN: TIME TO LIVE BIRTH AFTER NON-LIVE BIRTH PREGNANCY OUTCOME

DHS METHODOLOGICAL REPORTS 36

OCTOBER 2024

This publication was produced for review by the United States Agency for International Development. It was prepared by Sara Riese and Rachael Church.

DHS Methodological Reports No. 36

Pregnancy History Model Analysis Plan: Time to Live Birth after Non-live Birth Pregnancy Outcome

Sara Riese[1,2]
Rachael Church[1,3]

ICF
Rockville, Maryland, USA

October 2024

[1] The DHS Program
[2] ICF
[3] Avenir Health

Corresponding author: Sara Riese, Global Health and Development, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: 301-407-6500; fax: 301-407-6501; email: sara.riese@icf.com

Acknowledgments: The authors wish to thank Dr. Elma Laguna for her thoughtful review and comments and Trevor Croft for guidance on the new full pregnancy recode and recent pregnancy recode datasets.

This study was carried out with support provided by the United States Agency for International Development (USAID) through The DHS Program (#720-OAA-18C-00083). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

Editors: Cindy Young-Turner, Greg Edmondson
Document Production: Natalie Shattuck

Recommended citation:

Riese, Sara, and Rachael Church. 2024. *Pregnancy History Model Analysis Plan: Time to Live Birth after Non-live Birth Pregnancy Outcome*. DHS Methodological Reports No. 36. Rockville, Maryland, USA: ICF.

CONTENTS

TABLES AND FIGURES

PREFACE

The Demographic and Health Surveys (DHS) Program is one of the principal sources of international data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services.

One of the objectives of The DHS Program is to continually assess and improve the methodology and procedures used to carry out national-level surveys as well as to offer additional tools for analysis. Improvements in methods used will enhance the accuracy and depth of information collected by The DHS Program and relied on by policymakers and program managers in low- and middle-income countries.

While data quality is a main topic of the DHS Methodological Reports series, the reports also examine issues of sampling, questionnaire comparability, survey procedures, and methodological approaches. The topics explored in this series are selected by The DHS Program in consultation with the United States Agency for International Development.

It is hoped that the DHS Methodological Reports will be useful to researchers, policymakers, and survey specialists, particularly those engaged in work in low- and middle-income countries, and will be used to enhance the quality and analysis of survey data.

Sunita Kishor
Director, The DHS Program

ACRONYMS AND ABBREVIATIONS

BR	birth recode dataset
CI	confidence interval
DHS	Demographic and Health Survey
GR	full pregnancy recode dataset
HR	hazard ratio
IR	individual recode dataset
NR	recent pregnancy recode dataset

1 INTRODUCTION

This document is a model analysis plan for the further analysis of data from The Demographic and Health Surveys (DHS) Program. This plan is designed for use by students, researchers, and other data users for developing country-specific further analysis plans. It is the second model analysis plan developed by The DHS Program following the release of data from the eighth round of the DHS, which switched from using birth history to pregnancy history as the basis for the reproductive calendar part of the Woman's Questionnaire.

The decision to switch to pregnancy history was made primarily to enable better capture of information on stillbirth and other non-live birth pregnancy outcomes. This model analysis therefore focuses on using these new data on non-live birth pregnancy outcomes to understand more about the time interval between non-live birth outcomes and subsequent live births.

This model analysis plan is organized to show researchers each step of the further analysis process.

2 REQUIREMENTS

Stata data analysis software is required because this model analysis plan is designed for users of Stata. In addition, the DHS datasets are required for this analysis. For most introductory training or model purposes, The DHS Program recommends using the model datasets, which have been created for practice and do not represent any country's actual data. However, the model datasets are based on the DHS-6 questionnaire and recode dataset. Those datasets do not include the full pregnancy history and cannot be used for this model analysis plan. This analysis can only be run with a DHS-8 or later recode dataset. After registering, users can download the relevant data files from The DHS Program website.

3 SETTING UP THE DATA

Description of Relevant Data Files and Variables in DHS-8

With the introduction of pregnancy history in DHS-8, new data files were created to include information on each pregnancy. The key dataset in this analysis is the full pregnancy recode dataset (GR) file. In previous DHS dataset releases, information on women was found in the individual recode dataset (IR) file, and information on births was found in the birth recode dataset (BR) file. The IR and BR files are still available, and now two additional data files are available based on pregnancy history: the GR file and the recent pregnancy recode dataset (NR) file.

The GR file in DHS-8 is the new dataset that contains information on all pregnancies of interviewed women. In the GR file, pregnancy is the unit of analysis, as compared with the previous IR file, in which the interviewed woman was the unit of analysis. Most of the variable numbers in the GR file are the same as the variable numbers in the IR and BR files. New variables related to each pregnancy, which begin with the letter P, have been added. Some of these new variables link the pregnancy history to the birth history. For example, PIDXB is the index from the pregnancy history to the birth history. PIDXB is blank if the pregnancy outcome is not a live birth. Some of these new variables provide information about the pregnancy and its outcome. PIDX is the pregnancy history index variable, and PORD is the pregnancy order. The duration of pregnancy is given in variables p20 and p21. The pregnancy outcome declared by the respondent is given in variable p30, and the pregnancy outcome is reclassified in variable p32 based on subsequent questions (p31) to clarify the length of pregnancy and whether a birth was a live birth or a stillbirth. Variable p3 provides the date of the end of the pregnancy.

Pregnancy is also the unit of analysis in the NR file, but this file is limited to pregnancies in the previous 3 years. The NR file also includes the pregnancy and postnatal care variables from previous DHS phases, such as number of antenatal care visits, receipt of specific aspects of antenatal care, and location of delivery. More details about the NR file can be found in Methodological Report 34.

Table 1 summarizes these new pregnancy history variables in the GR dataset.

Table 1 Summary of select new key pregnancy history variables

Variable name	Definition and examples
All pregnancies	
PIDX	Pregnancy history index variable 1 = most recent pregnancy 2 = pregnancy just before the most recent pregnancy 3 = pregnancy before pregnancy with PIDX = 2 Continues for all pregnancies
PIDXB	Index variable from the pregnancy history to the birth history Lists the number where the pregnancy can be found in the birth history. Equal to the value of PIDX ONLY IF all pregnancies ended in a live birth. 0 = termination (not in birth history)
PORD	Pregnancy order variable 1 = first pregnancy 2 = second pregnancy 3 = third pregnancy 4 = fourth pregnancy Continues for all pregnancies
p3	Date of end of pregnancy (CMC)
p11	Preceding pregnancy interval (in months)
p20	Duration of pregnancy in months (1–10)
p21	Duration of pregnancy in days
p30	Pregnancy outcome declared by respondent 1 = born alive 2 = born dead 3 = miscarriage 4 = abortion
p32	Pregnancy outcome reclassified 1 = born alive 2 = born dead 3 = miscarriage 4 = abortion Pregnancy outcome may be reclassified if, for example, a respondent says she had a stillbirth at 5 months of gestation. This outcome would be reclassified as a miscarriage.

In this analysis, we will show analyses using the GR dataset.

4 MODEL ANALYSIS

4.1 Description of Analysis

This analysis estimates the time to a live birth after a prior miscarriage or stillbirth in a population and then seeks to identify factors that are associated with differences in the time to live birth in that population.

4.2 Application and Significance

Findings from analyses such as these can be used by program managers and policymakers to identify the factors that contribute to longer time to live birth after a non-live birth outcome so that these women who have these factors can be prioritized for intervention, if they are seeking to have another child.

4.3 Background Theory

Before moving to the data analysis, it is important to develop a logical framework for the research question. In this case, we look to the previous literature to understand how non-live birth outcomes influence time to subsequent pregnancy and what other factors are associated with time to pregnancy.

Studies looking into the time to pregnancy following a non-live birth have found that women with prior non-live birth outcomes have lower fecundability than women without a prior non-live birth.[1] This fecundability reduction can be compounding, with a stronger effect with additional prior miscarriages.[1]

The literature on time to pregnancy following a non-live birth outcome is limited in quantity, but we were able to identify many characteristics from other studies looking at time to pregnancy that can influence the amount of time it takes women to get pregnant, regardless of prior birth outcome, which may also be important for us to control for in this study. Some of these factors include mother's age, education level, smoking status and history, relationship status, body mass index, use of alcohol and caffeine, menstrual and reproductive history, prior contraceptive use history, environmental and occupational exposures, income, abortion history, and fertility preferences.[1–6]

All of these variables may not be available in the dataset. Most DHS datasets have a standard set of variables, but each country can include country-specific questions. Thus, it is important to review the specific dataset you are working with before proceeding with the analysis. We generally do not have data on maternal chronic disease or tobacco and alcohol use on all mothers, for example.

4.4 Analytic Details

This analysis is presented with the GR file. Code for these analyses is available in Appendix 2 of this report and at The DHS Program's GitHub page. It is important to consider your specific research questions when determining which dataset to use. Since non-live birth pregnancy outcomes are relatively rare, the sample size using one country's NR file, which includes only pregnancies from the past 3 years, is likely to be small. Using the GR file, which includes all pregnancies, allows for a larger sample size. Additionally, since our research question is looking at time between the non-live birth pregnancy outcome and a subsequent live birth, we want to maximize the period of follow-up after the non-live birth pregnancy outcome to be

able to include as many women as possible. The NR file only includes pregnancies from the past 3 years, which limits the period of follow-up.

4.4.1 Coding of Covariates

Many of the variables in this analysis already exist in the DHS-8 GR file, although some will need to be either created or recoded to create the appropriate response categories. These include pregnancy outcome, wealth tertile, education, and current marital status. For guidance on how to categorize any variables or indicators to match The DHS Program, please refer to the DHS indicators section of The DHS Program's GitHub page.

Variables for certain covariates from the DHS-8 GR file needed to be created for this analysis. These included binary variables for most recent pregnancy ending in a live birth and most recent pregnancy ending in a non-live birth. We also created a variable for age at non-live birth. One variable that we wanted to include in the regression model can only be calculated from the contraceptive calendar, which is only available in the DHS-8 IR file. We constructed a variable to identify women with a non-live birth pregnancy outcome and a subsequent live birth who did not use any form of contraception in the interval between the two events from the contraceptive calendar and then merged this variable onto the GR file for the remaining analysis.

After you have coded your outcomes and covariates, it is important to compare your estimates to the findings in the DHS final report from the same survey to ensure that the coding has been done correctly. Any variables that are reported in the final report with the same categorizations should be checked to confirm that the coding is correct.

4.4.2 Creating the Outcome Variable

Although we were most interested in time to *pregnancy* after a non-live birth pregnancy outcome, the point in a pregnancy when women recognize that they are pregnant can vary, and women may not remember the exact date on which they learned they are pregnant. In addition, the date of conception/pregnancy is not collected in the DHS. The date of the live birth is therefore a more reliable date. Thus, our outcome variable is time in months from the non-live birth pregnancy outcome to a subsequent live birth.

In order to calculate the time to live birth following a non-live birth pregnancy outcome, we first identified eligible sequences of non-live birth pregnancy outcomes followed by live births. For these sequences, we used the p11 variable of those eligible live births, which provides the preceding pregnancy interval, the time between the live birth and the end of the previous non-live birth pregnancy outcome.

4.4.3 Analytic Approach

Since we are examining the time to an event, we needed to use survival analysis. The analytic approach described here follows the standard steps of first describing the time to event outcome and then describing the multivariate associations of the covariates with the time to the event.

After creating the covariates of interest and the outcome variable of time to live birth in the GR file, we defined our population of interest by identifying women who had a live birth following their most recent non-live birth. Using the information from the GR file in which the unit of analysis is the pregnancy, we

needed to reshape the data to wide so that the unit of analysis is the woman and all of our variables of interest are included for each individual woman. Using this reshaped dataset, we examined the mean time to live birth following a non-live birth pregnancy outcome among all women with a non-live birth pregnancy outcome and by type of non-live birth pregnancy outcome. We calculated the mean and median time to live birth among all women with a non-live birth pregnancy outcome and according to the type of non-live birth pregnancy outcome. We estimated the lower and upper bounds of the 95% confidence interval (CI) of the mean according to a Poisson distribution.

In addition, we conducted multivariable survival analyses to identify factors associated with time to live birth following a non-live birth pregnancy outcome. Model specification was determined after assessing the proportional hazard assumption based on Schoenfeld residuals, which revealed that survival (time to live birth following a non-live birth pregnancy outcome) between covariate groups did not maintain proportionality over time and that relationships between covariates and the outcome were not consistently linear.[7] Therefore, we selected a parametric model rather than the semi-parametric model (Cox proportional hazard model). We estimated the baseline hazard function and hazard ratios for our covariates using a Weibull regression model. The model was restricted to women who had a live birth following a non-live birth pregnancy outcome.

The models included socioeconomic, demographic, and previous reproductive history characteristics of the woman. Some covariates that were originally planned for the analysis were not included due to high correlation. For example, age at non-live birth outcome and gravida at non-live birth outcome were highly correlated ($r = 0.6$), so only age was included in the model.

4.5 Analytic Details

Table 2 shows the distribution of covariates for all women who had a non-live birth outcome with a subsequent live birth by type of non-live birth. Among these women, nearly two-thirds (65%) had a miscarriage, with 24% having an abortion and 12% having a stillbirth. These non-live birth pregnancy outcomes generally took place when the women were between age 20 and age 29. Women who had stillbirths were more likely to have no education or a primary level of education, and women who had other types of non-live birth pregnancy outcomes were more evenly distributed by education level. More than half of women who had each type of non-live birth pregnancy outcome used contraception in the period between their non-live birth and their subsequent live birth. Most women had experienced only one non-live birth outcome.

Table 3 shows the mean and median time to a subsequent live birth in months for all women who had a live birth after a non-live birth pregnancy and by type of non-live birth pregnancy outcome. For all women, the mean time was 31 months. The median time was 23 months. Among women who had a prior stillbirth, the mean time to live birth was 33 months, and the median time was 24 months. Women who had a prior miscarriage had a mean time to live birth of 29 months and median time of 22 months. Women who had an abortion had a longer mean time (36 months) and median time (27 months), as compared with women who had other non-live birth pregnancy outcomes.

Table 2 **Background characteristics (N and %) of women included in the survival analysis, by type of non-live birth outcome**

	Stillbirth	Miscarriage	Abortion	*p* value
Total	597 (11.5%)	3,357 (64.7%)	1,232 (23.8%)	
Age at non-live birth				
10–19	115 (19.2%)	447 (13.3%)	281 (22.8%)	<.001
20–29	290 (48.5%)	1,823 (54.3%)	717 (58.2%)	
30–39	178 (29.8%)	1,025 (30.5%)	224 (18.2%)	
40–49	15 (2.6%)	62 (1.9%)	10 (0.8%)	
Current place of residence				
Urban	244 (40.8%)	1,769 (52.7%)	774 (62.8%)	<.001
Rural	354 (59.2%)	1,588 (47.3%)	458 (37.2%)	
Current wealth tertile				
Poor	161 (26.9%)	608 (18.1%)	185 (15.0%)	<.001
Middle	180 (30.1%)	917 (27.3%)	282 (22.9%)	
Rich	257 (43.0%)	1,832 (54.6%)	765 (62.1%)	
Current level of education				
None or primary	392 (65.5%)	1,663 (49.5%)	572 (46.4%)	<.001
Secondary or higher	206 (34.5%)	1,695 (50.5%)	660 (53.6%)	
Current marital status				
Not married	87 (14.6%)	396 (11.8%)	266 (21.6%)	<.001
Married	510 (85.4%)	2,961 (88.2%)	966 (78.4%)	
Contraceptive use in pregnancy interval				
Yes	369 (61.8%)	1,947 (58.0%)	801 (65.0%)	.002
No	228 (38.2%)	1,410 (42.0%)	431 (35.0%)	
Repeat non-live birth				
No	479 (80.1%)	2,475 (73.7%)	849 (68.9%)	<.001
Yes	119 (19.9%)	882 (26.3%)	383 (31.1%)	

Table 3 **Mean and median times (in months) to live birth after a non-live birth pregnancy outcome among all women who had a birth after a non-live birth pregnancy outcome**

Overall (N = 5,055)	
Mean (95% CI)	Median
30.8 (29.9, 31.8)	23
Stillbirth (n = 654)	
Mean (95% CI)	Median
32.7 (30.3, 35.4)	24
Miscarriage (n = 3,286)	
Mean (95% CI)	Median
28.5 (27.5, 29.6)	22
Abortion (n = 1,115)	
Mean (95% CI)	Median
36.2 (34.0, 38.6)	27

Table 4 shows the hazard ratios (HRs) of the variables included in the survival analysis. Figure A1 in Appendix 1 shows the Kaplan-Meier survival curves for all variables included in the regression. HRs greater than 1 indicate a shorter time period as compared with the reference group, and HRs lower than 1 indicate a longer time period as compared with the reference group. Compared with women who did not have repeat non-live birth pregnancy outcomes, the hazard of time to live birth is significantly lower (HR = 0.9) among women who did have repeat non-live birth pregnancy outcomes. Stated differently, women who had repeat non-live births have a longer time to a live birth outcome from their most recent non-live birth as compared with women who have only had one non-live birth. Compared with women who had a previous stillbirth, the hazard of time to subsequent live birth is significantly higher (HR = 1.2) for women

who had a previous miscarriage. Women who did not use any contraceptive method in the time between their non-live birth pregnancy outcome and the subsequent live birth had an earlier live birth as compared with women who did use a contraceptive method (HR = 2.7).

Looking at the sociodemographic characteristics, women who were age 30–39 or age 40–49 at the time of the non-live birth pregnancy outcome have a significantly higher hazard of time to live birth (HR = 1.2 for age 30–39, HR = 1.3 for age 40–49) as compared with women who were age 20–29 at the time of the non-live birth. Women who completed secondary or higher levels of education at the time of the survey have a significantly higher hazard of time (HR = 1.2) to live birth as compared with women who completed lower levels of education. Neither wealth nor marital status at the time of the survey was significantly associated with time to live birth after a non-live birth pregnancy outcome.

Table 4 Hazard ratio and 95% confidence interval according to multivariable Weibull regression of time to live birth outcome after a non-live birth pregnancy outcome

		HR	95% CI
Repeat non-live birth outcome (ref: No)			
	Yes	0.9*	0.8–1.0
Non-live birth pregnancy outcome (ref: Stillbirth)			
	Miscarriage	1.2*	1.0–1.3
	Abortion	0.9	0.8–1.0
Contraceptive use in pregnancy interval (ref: Used contraception)			
	No	2.7***	2.5–2.9
Age group at time of non-live birth (ref: 20–29 years)			
	10–19	0.9	0.8–1.1
	30–39	1.2***	1.1–1.3
	40–49	1.3*	1.0–1.6
Current marital status (ref: Not married)			
	Married	1.1	0.9–1.2
Wealth tertile (ref: Poor)			
	Middle	1.0	0.9–1.1
	Rich	0.9	0.8–1.0
Current highest level of education (ref: None or primary)			
	Secondary or higher	1.2***	1.1–1.3

***p < .001; *p < .05.

4.6 Conclusions

The full pregnancy history allows researchers to conduct many new kinds of analyses to better understand a woman's reproductive history. This model analysis presents one of these new types of analysis, which focused on the duration of the interval between a non-live birth outcome and a subsequent live birth. We also aimed to identify characteristics of women that are associated with differences in the time to live birth. This type of analysis is relevant for many policy and programming decisions. This model analysis could help countries understand which groups of women who have a non-live birth outcome may take longer to have a subsequent live birth and design policies or programs to support women who want to have a child after they have experienced a non-live birth. This analysis can be adapted to answer a variety of questions about pregnancy outcomes. Results from this type of analysis should be interpreted carefully, however. When using the pregnancy history and controlling for current characteristics, it is important to consider the length of time since the pregnancy and whether the current characteristics are time-varying and may have changed from the time of the pregnancy to the time of data collection.

REFERENCES

1. Arge LA, Håberg SE, Wilcox AJ, Næss Ø, Basso O, Magnus MC. The association between miscarriage and fecundability: the Norwegian Mother, Father and Child Cohort Study. *Hum Reprod.* 2022;37:322–332. https://doi.org/10.1093/humrep/deab252

2. Barden-O'Fallon J, Speizer IS, Calhoun LM, Moumouni NA. Return to pregnancy after contraceptive discontinuation to become pregnant: a pooled analysis of West and East African populations. *Reprod Health.* 2021;18:141. https://doi.org/10.1186/s12978–021–01193-w

3. Gayatri M, Utomo B, Budiharsana M, Dasvarma G. Pregnancy resumption following contraceptive discontinuation: hazard survival analysis of the Indonesia Demographic and Health Survey data 2007, 2012 and 2017. *PLoS One.* 2022;17:e0264318. https://doi.org/10.1371/journal.pone.0264318

4. Polis CB, Cox CM, Tunçalp Ö, Mclain AC, Thoma ME. Estimating infertility prevalence in low-to-middle-income countries: an application of a current duration approach to Demographic and Health Survey data. *Hum Reprod.* 2017;32:1064–1074. https://doi.org/10.1093/humrep/dex025

5. Sapra KJ, Mclain AC, Maisog JM, Sundaram R, Buck Louis GM. Successive time to pregnancy among women experiencing pregnancy loss. *Hum Reprod.* 2014;29:2553–2559. https://doi.org/10.1093/humrep/deu216

6. Wang X, Chen C, Wang L, Chen D, Guang W, French J. Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertil Steril.* 2003;79:577–584. https://doi.org/10.1016/s0015–0282(02)04694–0

7. Johnson LL, Shih JH. An introduction to survival analysis. In: Gallin JI, Ognibene FP, eds. *Principles and Practice of Clinical Research*. Cambridge, MA: Academic Press; 2007:273–282.

APPENDIX 1: KAPLAN-MEIER SURVIVAL CURVES OF TIME TO LIVE BIRTH AFTER A NON-LIVE BIRTH OUTCOME, BY WOMAN'S CHARACTERISTICS

Figure A1 Survival curve by non-live birth pregnancy outcome

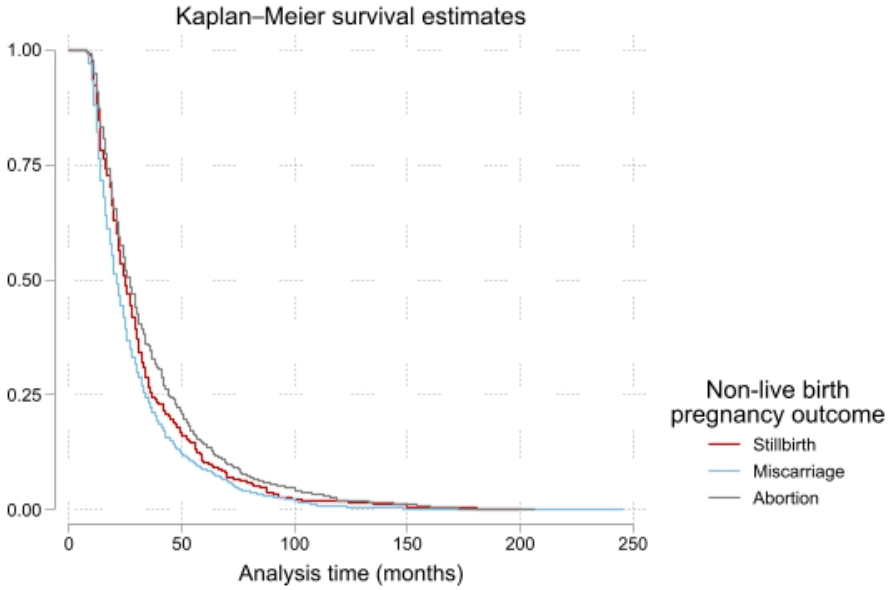


Figure A2 Survival curve by repeat non-live birth status

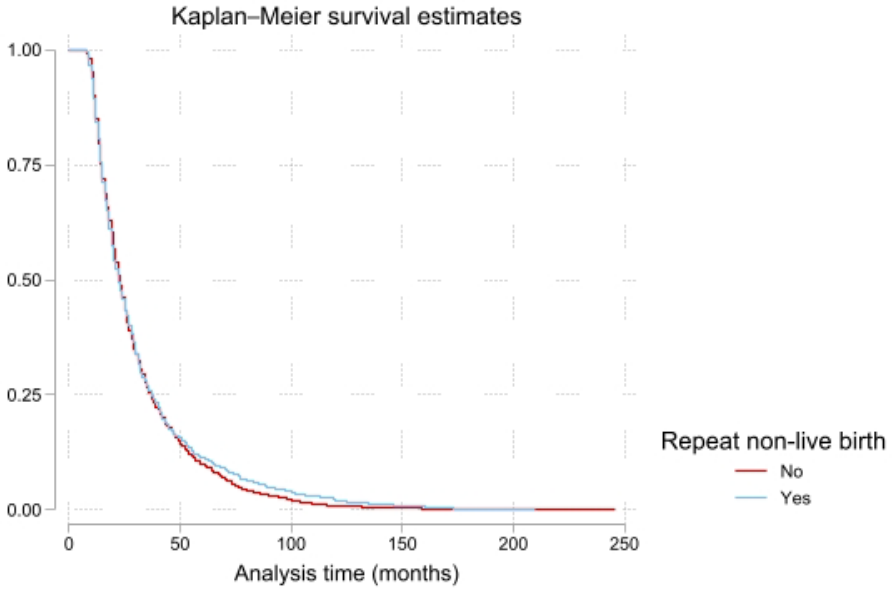


Figure A3 Survival curve by age at non-live birth

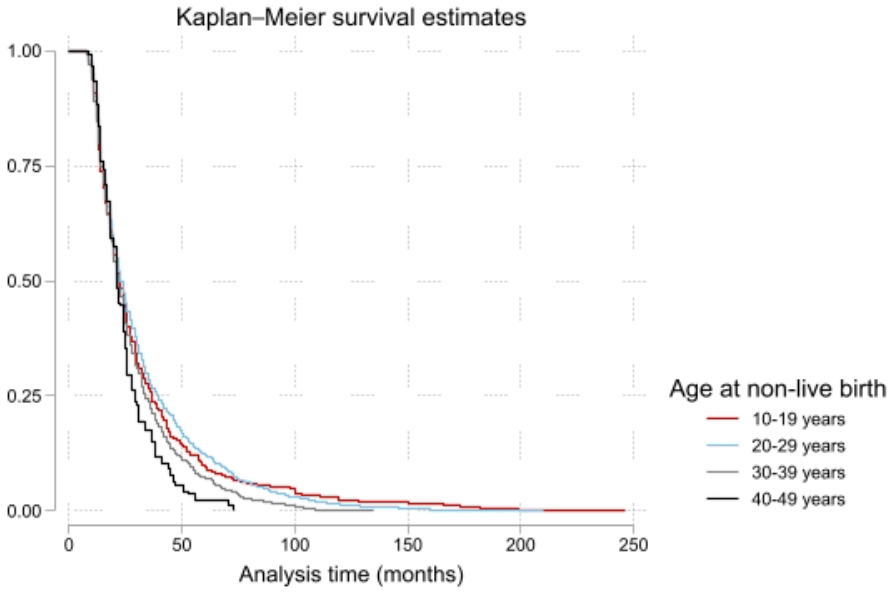


Figure A4 Survival curve by contraceptive use status

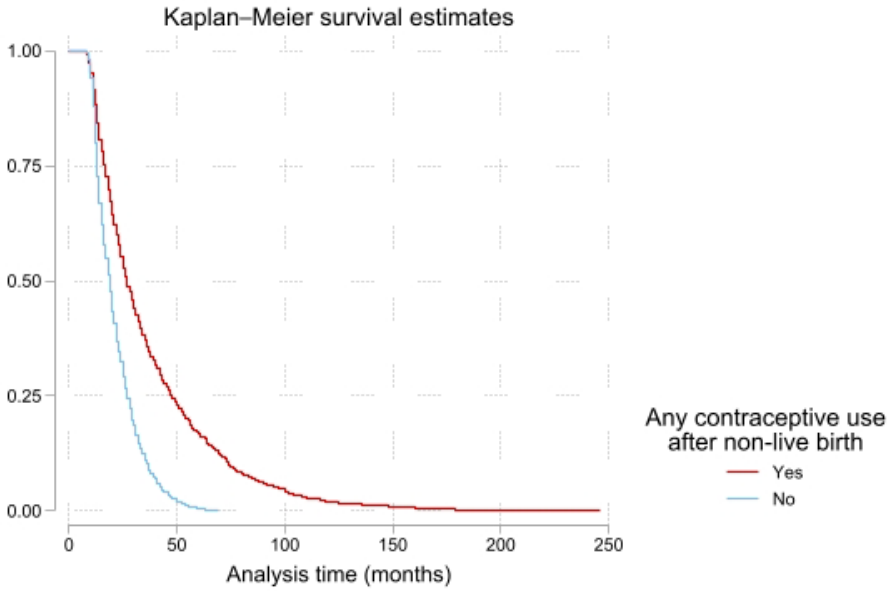


Figure A5 Survival curve by marital status

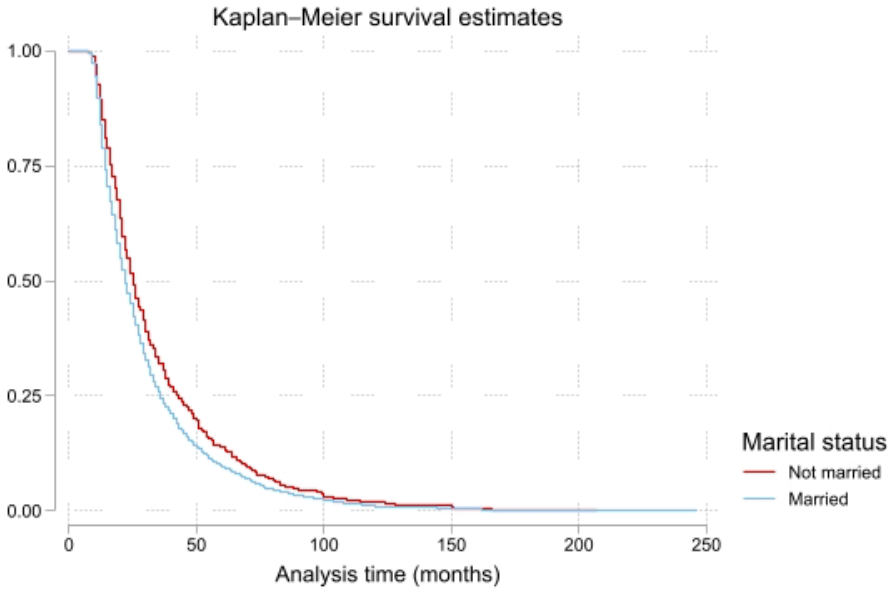


Figure A6 Survival curve by wealth status

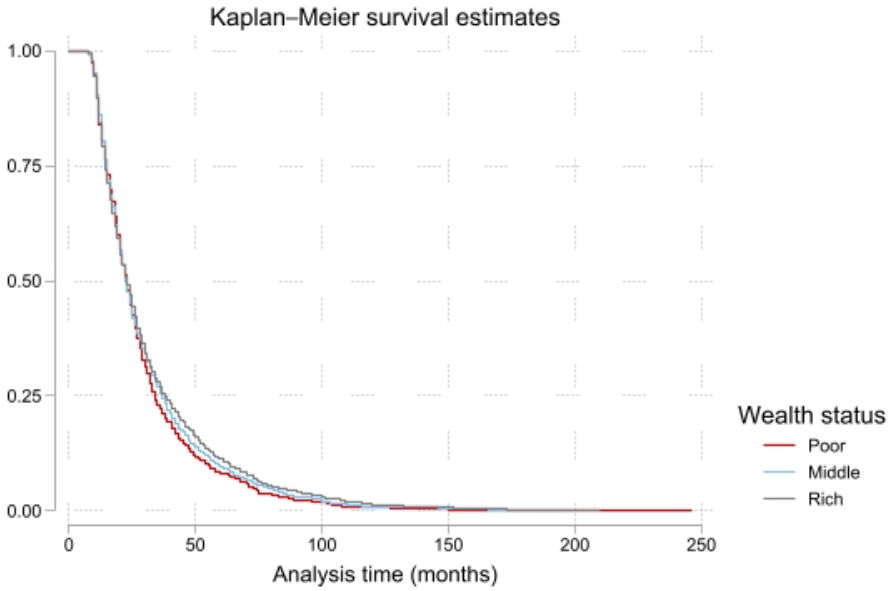
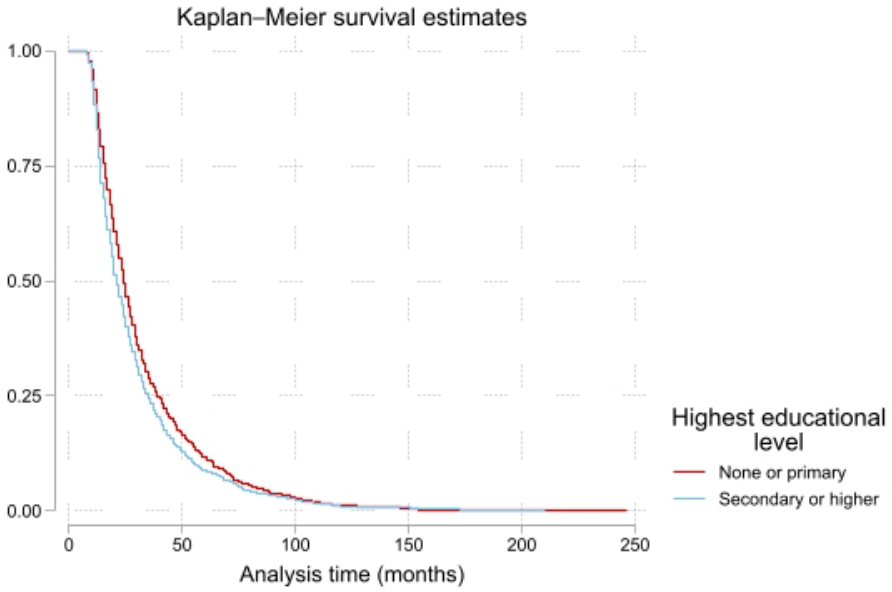


Figure A7 Survival curve by highest educational level



APPENDIX 2: STATA DO FILE FOR TIME TO LIVE BIRTH AFTER A NON-LIVE BIRTH OUTCOME

```
/*******************************************************************************
Program:                MR36_dofile
Purpose:                Code datasets and run analysis for MR36 - Time to
                        live birth after non-live birth outcome
Data inputs:            IR and GR files
Data outputs:           coded data files
Author:                 Sara Riese and Rachael Church
*******************************************************************************/

*Set directory where your data is found and where you want to output tables
cd "Folder"

use "XXIR8XFL.DTA", clear

      ***Country***
      gen country =substr(v000,1,2)
      scalar country=country[1]
      local country1=country


*Create a variable for if a woman used contraceptives between her non-live
birth and the subsequent live birth
gen trim_cal = trim(vcal_1)
label variable trim_cal "trimmed calendar"
gen recentT = strpos(trim_cal,"T")
label variable recentT "Position of most recent termination"
gen firstT = strpos(reverse(trim_cal),"T")
label variable firstT "Position of first termination"

*NOTE: If you do not have the -egenmore- package installed, you will need to
install it here in order to run the next few lines
*ssc install egenmore
egen count_T=noccur(trim_cal), string("T")
label variable count_T "Number of terminations in calendar"

gen recentB = strpos(trim_cal,"B")
label variable recentB "Position of most recent birth"
gen baftert=1 if recentB<recentT & recentB!=0
label variable baftert "any birth after most recent termination"
gen piece = substr(trim_cal, recentB, recentT) if baftert==1
label variable piece "Piece of calendar between most recent birth and most
recent termination if there was any birth"

gen pos2 = strpos(piece,"T")
gen piece2 = substr(piece, 1, pos2) if baftert==1
gen pos3 = strpos(reverse(piece2),"B")
gen piece3=substr(reverse(piece2), 1, pos3)
```

```
gen noFP = ustrpos(piece3, "0") > 0 & ustrpos(piece3, "P") > 0 &
ustrpos(piece3, "B") > 0 & ustrpos(piece3, "T") > 0

keep caseid v000 vcal_1 trim_cal-noFP

save "IR_FP.dta", replace



use "XXGR8XFL.DTA", clear

*survey set & weighting
gen wt = v005/1000000
gen strata=.
cap replace strata=v022
svyset v021 [pw=wt], strata(strata) singleunit(centered)

      ***Country***
      gen country =substr(v000,1,2)
      scalar country=country[1]
      local country1=country

**************************************************
**COVARIATES
**************************************************

*Urban/Rural
gen residence=v025
label var residence "Place of Residence"
label values residence v025

*Wealth tertile
xtile wealth = v191, nq(3)
label define wealth 1 "Poor" 2 "Middle" 3 "Rich"
label var wealth "Wealth Level"
label values wealth wealth

*Education
recode v106 (0 1=1 "None or primary") (2/3=2 "Secondary or higher"),
gen(edu)
label var edu "Education"

*Marital status
gen married = 0
replace married =1 if v502==1
label var married "Currently married"
label def mar 0 "Not married" 1 "Married"
label val married mar

*Outcome of last non-live birth
*Create a variable that includes all nlb
gen nlb=0
replace nlb=1 if p32!=1
```

```
*Create nonlive birth order variable
gen nlbord=0
replace nlbord=pord if p32!=1

*Pregnancy outcome
recode p32 (1=.) (2 = 1 "stillbirth") (3 = 2 "miscarriage") (4 = 3
"abortion"), gen(pregoutcome)
label var pregoutcome "Outcome of nonlive birth"


*Create variable with pregnancy order number of when 1st non-live birth event
happened for a woman
egen firstnlb=min(pord/(nlb==1)), by(caseid)

egen totalnlb=total(nlb), by (caseid)

*Create variable for most recent live birth
bysort caseid: egen recentlb = max(pord) if nlb==0
gen mostrecentlb=1 if recentlb!=. & pord==recentlb
replace mostrecentlb=0 if recentlb!=. & pord!=recentlb
label var mostrecentlb "Most recent live birth"
label define mostrecentlb 0 "Not most recent" 1 "Most recent"
label values mostrecentlb mostrecentlb

*Create variable for most recent nonlive birth among women who have more than
1 nonlive birth
bysort caseid: egen recentnlb = max(pord) if nlb==1 & totalnlb>=2
gen mostrecentnlb=1 if recentnlb!=. & pord==recentnlb
replace mostrecentnlb=0 if recentnlb!=. & pord!=recentnlb
label var mostrecentnlb "Most recent nonlive birth"
label define mostrecentnlb 0 "Not most recent" 1 "Most recent"
label values mostrecentnlb mostrecentnlb

*Age at non-live birth
gen agenlb=.
replace agenlb=(p3-v011)/12 if nlb==1
recode agenlb (10/19.999=1 "10-19") (20/29.999=2 "20-29") (30/39.999=3 "30-
39") (40/max=5 "40-49"), gen (agegroupnlb)
label var agegroupnlb "Age at nonlive birth"

/*Repeat non-live birth (is the most recent nlb a repeat nlb, not has the
woman had multiple nlbs)
*create variable for number of nlb a woman has had (need to use egen to
create a variable within a case)
*/
gen repeatnlb=.
replace repeatnlb=1 if pord>firstnlb & totalnlb>1 & nlb==1
replace repeatnlb=0 if pord==firstnlb & totalnlb>1 & nlb==1
replace repeatnlb=0 if nlb==1 & totalnlb==1
label var repeatnlb "Repeat nonlive birth"
label define repeatnlb 0 "nonrepeat" 1 "repeat"
label values repeatnlb repeatnlb
```

```
*************************
*********OUTCOME**********
*************************

*time to pregnancy following nonlive birth

*restricting to women who had a birth following most recent non live birth
bysort caseid (pord): gen eligible=1 if nlb==0 & nlb[_n-1]==1 &
mostrecentlb==1 //pregnancy that followed the nonlive birth
bysort caseid (pord): gen lbfternlb=1 if nlb==0 & nlb[_n-1]==1 &
mostrecentlb==1
bysort caseid (pord): replace eligible=1 if eligible==. & eligible[_n+1]==1
//nonlive birth prior to the live birth
bysort caseid (pord): gen nlbpriortolb=1 if eligible==. & eligible[_n+1]==1

*months since prior non-live birth minus months since most recent live birth
is variable p11

save "GR_coded.dta", replace

merge m:1 caseid v000 using "IR_FP.dta"

save "TTBcoded.dta", replace


*****************************************************************************
*******************************ANALYSIS**************************************
*****************************************************************************
*Reshape data from pregnancies to women
use "TTBcoded.dta", clear
keep caseid pord residence wealth edu married nlb nlbord pregoutcome
mostrecentnlb mostrecentlb agegroupnlb lbfternlb firstnlb repeatnlb eligible
p11 p32 v000-v101 wt strata vcal_1 noFP

keep if eligible==1

reshape wide pord nlbord mostrecentnlb mostrecentlb agegroupnlb lbfternlb
firstnlb repeatnlb eligible p11 p32 pregoutcome vcal_1 noFP, i(caseid v000)
j(nlb)

save "TTBcoded_wide.dta", replace

*Export sample characteristics table
dtable i.agegroupnlb1 i.residence i.wealth i.edu i.married i.noFP1
i.repeatnlb1 i.eligible0, by(pregoutcome1, nototals tests) svy
column(by(hide) test(p-value)) export("table1_ttb.xlsx", replace)

*****************************************************************************
************************Mean and median time to live birth*****************
*****************************************************************************
```

```
      *survey set and set scalar for country
      svyset v021 [pw=wt], strata(strata) singleunit(centered)

*mean time to live birth-all
      svy: poisson p110
            matrix mnall=r(table)
                  *estimate
                  scalar s4p = exp(mnall[1,1])
                  *confidence intervals
                  scalar s4lb = exp(mnall[5,1])
                  scalar s4ub = exp(mnall[6,1])

*mean time to live birth-after different types of previous non-live births
*NOTE: You may need to modify the number of "forvalues" in the below loop
from 1/3 to 1/2 if there are no abortions in the dataset. If you keep 1/3,
you will get an error due to no observations.

forvalues i = 1/3 {
      svy: poisson p110 if pregoutcome1 == `i'
            matrix mn=r(table)
                  *estimate
                  scalar s`i'p = exp(mn[1,1])
                  *confidence intervals
                  scalar s`i'lb = exp(mn[5,1])
                  scalar s`i'ub = exp(mn[6,1])
            }

      *Combine
      forvalues i = 1/4 {

            foreach x in s`i'p s`i'lb s`i'ub {

            *round
            scalar `x' = round(`x', 0.1)

            *create string
                  scalar `x'_str = string(`x')
                  scalar `x'_str = string(`x')

            *add 0 before numbers <1
                  if `x'<1 {
                  scalar `x'_str="0"+`x'_str
                  }

            *adding .0 after number if it rounded to zero decimal places
                  if `x' == int(`x') {
                  scalar `x'_str =`x'_str + ".0"
                  }
            }
      }

      scalar sallCI=s4p_str + " " + "(" + s4lb_str + "," + s4ub_str + ")"
      scalar sstillCI=s1p_str + " " + "(" + s1lb_str + "," + s1ub_str + ")"
```

```
      scalar smissCI=s2p_str + " " + "(" + s2lb_str + "," + s2ub_str + ")"
      scalar saborCI=s3p_str + " " + "(" + s3lb_str + "," + s3ub_str + ")"



*median time
      _pctile p110 [pw=wt] , p(50)
      scalar all = round(r(r1), 0.1)

      _pctile p110 [pw=wt] if pregoutcome1==1, p(50)
      scalar still = round(r(r1), 0.1)

      _pctile p110 [pw=wt] if pregoutcome1==2, p(50)
      scalar miss = round(r(r1), 0.1)

      _pctile p110 [pw=wt] if pregoutcome1==3, p(50)
      scalar abor = round(r(r1), 0.1)


      putexcel set "meanmed.xlsx", modify sheet ("poissonmeanmedtime")
      putexcel    B2=("Mean all") C2=("Median all") ///
                        D2=("Mean stillbirth") E2=("Median stillbirth") ///
                        F2=("Mean miscarriage") G2=("Median miscarriage") ///
                        H2=("Mean abortion")    I2=("Median abortion")  ///

                        B3=(sallCI)       C3 = (all) ///
                        D3=(sstillCI)     E3=(still) ///
                        F3=(smissCI)      G3=(miss) ///
                        H3=(saborCI)      I3=(abor)


*************************************************************************
*************************Survival analysis*******************************
*************************************************************************
*set survival
      stset p110 [pw=wt]

*survival curves
sts graph, failure
sts graph, by(agegroupnlb1)
sts graph, by(repeatnlb1)
sts graph, by(pregoutcome1)
sts graph, by(married)
sts graph, by(wealth)
sts graph, by(noFP1)

sts graph, by(edu)

*run survival

      *Check proportionality using Schoenfeld's residuals
      stcox i.repeatnlb1 i.pregoutcome1 i.noFP1 ib2.agegroupnlb1 i.married
i.wealth i.edu
```

```
      estat phtest, detail
      *If global value p-value is <0.05, then the proportional hazards
assumption does not hold, and should use a parametric model

svy: streg i.repeatnlb1 i.pregoutcome1 i.noFP1 ib2.agegroupnlb1 i.married
i.wealth i.edu, d(weibull)


****store results
      *mat mat1= r(table)
 cd $results
      outreg2 using "ttb_results.xls", eform stats(coef se ci) alpha(0.001,
0.01, 0.05) sideway dec(1) label(insert) noparen nocons append
```