Suplementary Code 5: Code to Respond to Reviewers

Richard Meitern

2024-01-04

Introduction

The aim of this document is to respond to the reviewers comments. The comments are:

1) I would like to see whether indeed mothers are statistically more likely to stop having children after a twin birth (perhaps I missed that analysis and it was indeed presented)?
2) Directly test the odds that a mother terminate reproduction after a twin birth? In other words, use statistics to assess post-twinning reproductive outcomes compared to post-singleton reproductive outcomes.
3) Plot the real data as a cloud of points behind these lines- it is important to see the actual dataset in analyses like these.
4) Look at the analyses when limiting to opposite-sex offspring?
5) Based on analyses of Finnish data from 1953-1964 Eriksson and Fellman show that illegitimate maternities have a higher twinning rate than legitimate ones when maternal age and birth order are taken into account. They speculate that unmarried mothers are women who tend to become pregnant more easily than others owing to the effect of certain factors, e.g. a higher ovulation rate and to have relatively better physical qualifications for going through a twin pregnancy than mothers in general. It seems not unlikely that such mothers may remain single of the twin pregnancy, especially in earlier times?

```
#cleanup memory
gcstuff <- gc(verbose=FALSE); rm(gcstuff);
```

```
#get last birth adding function
source("./R/last_birth.R")

#simplified twinR summary tables
source("./R/twinR_summary.R")

#fix twinR compute predictions to do prediction with no lambda  as well
source("./R/twinR_predictions.R")

#simple convenience functions
source("./R/utils.R")
```

```
## Identify number of CPU cores available for parallel computing,
## note: using a large number may lead RAM to max out, so you may have to adjust
## that according to your infrastructure:
```

```r
nb_cores <- min(c(50L, parallel::detectCores() - 1))

## Set option in spaMM:
spaMM::spaMM.options(nb_cores = nb_cores)
```

```
## Registered S3 methods overwritten by 'registry':
##   method               from
##   print.registry_field proxy
##   print.registry_entry proxy
```

Data Import

The Estonian dataset has been formatted to include the same columns as *the data_births_all* dataset from the **twinR** package. The only difference is that the columns *pop* and *monthly* are excluded as these are constant.

```r
#some labels
estLab <- "Estonian"
westLab <- "9 other European"
```

```r
#Import and preproccess Estonian Data

data_births_monthly_EE <- readRDS("./data/data_births_all_EE.rds")

#the twinR package expects population to be present
data_births_monthly_EE$pop <- "Estonia"

## Expand the birth level data for the fit of statistical models:
data_births_monthly_EE <- twinR::expand_data(data_births_monthly_EE)

data_births_monthly_EE <- add_last_birth(data_births_monthly_EE)

data_births_monthly_EE_not_last <- data_births_monthly_EE[!data_births_monthly_EE$last,]

#make the aggregates
dmm_EE <- twinR::aggregate_data(data_births_monthly_EE)
dmm_EE$prob_twin <- dmm_EE$twin_total / dmm_EE$births_total

dmm_EE_nl <- twinR::aggregate_data(data_births_monthly_EE_not_last)
dmm_EE_nl$prob_twin <- dmm_EE_nl$twin_total / dmm_EE_nl$births_total
```

```r
##Import and pre-proccess twinR package data

## Filter the raw data to only keep data with monthly resolution:
data_births_monthly <- twinR::filter_data(twinR::data_births_all)

## Expand the birth level data for the fit of statistical models:
data_births_monthly <- twinR::expand_data(data_births_monthly)

data_births_monthly <- add_last_birth(data_births_monthly)
data_births_monthly_not_last <- data_births_monthly[!data_births_monthly$last,]
```

```
dmm_orig <- twinR::aggregate_data(data_births_monthly)
dmm_orig$prob_twin <- dmm_orig$twin_total /dmm_orig$births_total

dmm_orig_nl <- twinR::aggregate_data(data_births_monthly_not_last)
dmm_orig_nl$prob_twin <- dmm_orig_nl$twin_total /dmm_orig_nl$births_total
```

```
#compared to twinR original add data on :
#-age at first birth (AFB)
#-quantiles range mother birth year
#-quantiles range offspring birth year
#-mean + SE total births
all_tbls <- rbind(build_data_summary.table(data_births_monthly),
                  build_data_summary.table(data_births_monthly_not_last),
                  build_data_summary.table(data_births_monthly_EE),
                  build_data_summary.table(data_births_monthly_EE_not_last))
```

Fitting models

Comment 1: Logistic regression

We can test the hypothesis that mothers are statistically more likely to stop having children after a twin birth by fitting a logistic regression model with the response variable being the binary variable *last* and the predictor variable being the binary variable *twin*.

```
#fit logistic regression Estonia
logit_fit_EE <- glm(twin ~  last,
                 data = data_births_monthly_EE,
                 family = binomial(link = "logit"))
summary_logit_fit_EE <-  summary(logit_fit_EE)
summary_logit_fit_EE
```

```
##
## Call:
## glm(formula = twin ~ last, family = binomial(link = "logit"),
##     data = data_births_monthly_EE)
##
## Coefficients:
##               Estimate Std. Error   z value   Pr(>|z|)
## (Intercept) -4.2840087  0.0159821 -268.0512 < 2.22e-16 ***
## lastTRUE     0.3555958  0.0260050   13.6741 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 66110.47  on 417417  degrees of freedom
## Residual deviance: 65929.38  on 417416  degrees of freedom
## AIC: 65933.38
##
## Number of Fisher Scoring iterations: 7
```

```r
#fit logistic regression Other European
logit_fit <- glm(twin ~  last,
                 data = data_births_monthly,
                 family = binomial(link = "logit"))
summary_logit_fit_W <-  summary(logit_fit)
summary_logit_fit_W
```

```
##
## Call:
## glm(formula = twin ~ last, family = binomial(link = "logit"),
##     data = data_births_monthly)
##
## Coefficients:
##               Estimate Std. Error    z value   Pr(>|z|)
## (Intercept) -4.1855085  0.0283068 -147.86238 < 2.22e-16 ***
## lastTRUE     0.4517268  0.0534924    8.44469 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 17943.05  on 105832  degrees of freedom
## Residual deviance: 17876.42  on 105831  degrees of freedom
## AIC: 17880.42
##
## Number of Fisher Scoring iterations: 7
```

Comment 2: Proportion of twins

The same test gives us the odds of mothers having more children after a twin birth.

```r
# Extract coefficients
coef <- coef(logit_fit_EE)

# Calculate odds ratios
odds_ratios <- exp(coef)
```

For the Estonian dataset the odds of a birth event being the last one for a mother are 1.427 times higher for a twin birth than for a singleton birth.

```r
# Extract coefficients
coef <- coef(logit_fit)

# Calculate odds ratios
odds_ratios <- exp(coef)
```

For the other dataset the odds of a birth event being the last one for a mother are 1.571 times higher for a twin birth than for a singleton birth.

```r
#compare the precentage of twin births between last births and not last births
cat("Estonian dataset:\n")
```

```
## Estonian dataset:
```

```
tble <- table(data_births_monthly_EE[c("twin", "last")])
tble
```

```
##        last
## twin     FALSE   TRUE
##   FALSE 287874 123152
##   TRUE    3969   2423
```

```
cat("\n")
```

```
#now calcualte the % by hand
cat("Last twins: ", round(100 * tble[2,2] / sum(tble[,2]), 2), "%\n")
```

```
## Last twins:  1.93 %
```

```
cat("Prior twins: ",round(100 * tble[2,1] / sum(tble[,1]), 2), "%\n")
```

```
## Prior twins:  1.36 %
```

```
#compare the precentage of twin births between last births and not last births
cat("Western dataset:\n")
```

```
## Western dataset:
```

```
tblw <- table(data_births_monthly[c("twin", "last")])
tblw
```

```
##        last
## twin    FALSE  TRUE
##   FALSE 83276 20793
##   TRUE   1267   497
```

```
cat("\n")
```

```
#now calcualte the % by hand
cat("Last twins: ", round(100 * tblw[2,2] / sum(tblw[,2]), 2), "%\n")
```

```
## Last twins:  2.33 %
```

```
cat("Prior twins: ",round(100 * tblw[2,1] / sum(tblw[,1]), 2), "%\n")
```

```
## Prior twins:  1.5 %
```

```r
data_births_monthly$monthly <- NULL
#data_births_monthly$pop <- "Western"
dbmFull <- rbind(data_births_monthly_EE[colnames(data_births_monthly)],
                 data_births_monthly)
dbmFull$popEE <- dbmFull$pop == "Estonia"
```

```r
#fit logistic regression
logit_fit_full <- glm(twin ~  last + popEE + popEE:last,
                 data = dbmFull,
                 family = binomial(link = "logit"))
summary_logit_fit_full <-  summary(logit_fit_full)
summary_logit_fit_full
```

Test the Interaction

```
##
## Call:
## glm(formula = twin ~ last + popEE + popEE:last, family = binomial(link = "logit"),
##     data = dbmFull)
##
## Coefficients:
##                       Estimate Std. Error    z value   Pr(>|z|)
## (Intercept)        -4.1855085  0.0283068 -147.86238 < 2.22e-16 ***
## lastTRUE            0.4517268  0.0534924    8.44469 < 2.22e-16 ***
## popEETRUE          -0.0985002  0.0325069   -3.03013  0.0024445 **
## lastTRUE:popEETRUE -0.0961311  0.0594785   -1.61623  0.1060444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 84063.45  on 523250  degrees of freedom
## Residual deviance: 83805.81  on 523247  degrees of freedom
## AIC: 83813.81
##
## Number of Fisher Scoring iterations: 7
```

```r
#' Get Coefficient Table
#'
#' This function returns a table with odds ratios and their confidence intervals.
#'
#' @param summary A summary object from a glm model.
#' @param datasetname A character string specifying the name of the dataset.
#'
#' @return A data frame with odds ratios and their confidence intervals.
get_coef_tbl <- function(summary, datasetname){
  tbl <- as.data.frame(summary$coefficients)
  tbl$dataset <- datasetname
  tbl$variable <- rownames(tbl)
  tbl$`Odds Ratio` <- exp(tbl$Estimate)
```

```
  tbl$`CI low` <- exp(tbl$Estimate - 1.96 * tbl$`Std. Error`)
  tbl$`CI high` <- exp(tbl$Estimate + 1.96 * tbl$`Std. Error`)
  rownames(tbl) <- NULL
  tbl <- tbl[,c(5,6,7:9)]
  return(tbl)
}


#print all coeficient tables in one table
#add dataset name to each table
coefEEtbl <- get_coef_tbl(summary_logit_fit_EE, "Estonia")
ceofWtbl <- get_coef_tbl(summary_logit_fit_W, "Western")
ceofFulltbl <- get_coef_tbl(summary_logit_fit_full, "Combined")



tbl <- rbind(coefEEtbl, ceofWtbl, ceofFulltbl)
knitr::kable(tbl, digits = 3, caption = "Odds ratios and confidence intervals for the logistic regressi
```

Table 1: Odds ratios and confidence intervals for the logistic regression model with the response variable being the binary variable *twin* and the predictor variable being the binary variable *last*.

dataset	variable	Odds Ratio	CI low	CI high
Estonia	(Intercept)	0.014	0.013	0.014
Estonia	lastTRUE	1.427	1.356	1.502
Western	(Intercept)	0.015	0.014	0.016
Western	lastTRUE	1.571	1.415	1.745
Combined	(Intercept)	0.015	0.014	0.016
Combined	lastTRUE	1.571	1.415	1.745
Combined	popEETRUE	0.906	0.850	0.966
Combined	lastTRUE:popEETRUE	0.908	0.808	1.021

```
#print the frequency table
frqTbl <- aggregate(maternal_id ~ twin + last + popEE,
                    data = dbmFull, FUN = length)


# Rename the variables inside the columns
frqTbl$twin <- ifelse(frqTbl$twin == TRUE, "twins", "singleton")
frqTbl$last <- ifelse(frqTbl$last == TRUE, "Yes", "No")
frqTbl$popEE <- ifelse(frqTbl$popEE == TRUE, "Estonia", "Western")

# Rename the column names
colnames(frqTbl) <- c("Birth", "Last Breeding", "Population", "Count")
knitr::kable(frqTbl, digits = 0, caption = "Frequency table of the data used for the logistic regressio
```

Table 2: Frequency table of the data used for the logistic regression model with the response variable being the binary variable *twin* and the predictor variable being the binary variable *last*.

Birth	Last Breeding	Population	Count
singleton	No	Western	83276
twins	No	Western	1267
singleton	Yes	Western	20793
twins	Yes	Western	497
singleton	No	Estonia	287874
twins	No	Estonia	3969
singleton	Yes	Estonia	123152
twins	Yes	Estonia	2423

Comment 3: Original Data on Plots

The reviewers also asked to put the data on the Figure 1.

```
# import the function to do model fit and predictions
source("./R/fit_models.R")
```

```
## full data

## Estonia - mother level data
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total"
dmm_EE_fit <- fitPredictions(dmm_EE, formula, predict = T)
```

```
## Warning in fitPredictions(dmm_EE, formula, predict = T): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_totaldmm_EE_fit.rds
##  If you want to re-run this step delete the file or change the saveDir!
```

```
## Warning in get_predictions(predDataFname, fit, dataset, args, save): Pre-computed predictions returne
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_totaldmm_EEdata_fig.rds
##  If you want to re-run this step delete the file or change the saveDir!
```

```
## TwinR - mother level data
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total + (1|pop)"
dmm_orig_fit <- fitPredictions(dmm_orig, formula, predict=T)
```

```
## Warning in fitPredictions(dmm_orig, formula, predict = T): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+(1_x_pop)dmm_orig_fit.rds
##  If you want to re-run this step delete the file or change the saveDir!
```

```
## Warning in get_predictions(predDataFname, fit, dataset, args, save): Pre-computed predictions returne
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+(1_x_pop)dmm_origdata_fig.rds
##  If you want to re-run this step delete the file or change the saveDir!
```

```
library(ggplot2)
#some nice colors
bc <- c("azure4","purple4", "black", "navy", "darkgoldenrod2", "springgreen3", "gray")
```

```r
#use new base theme that displays also grid lines
source("./R/twinR_theme.R")
```

Fig 1a: Estonian vs TwinR Full Data

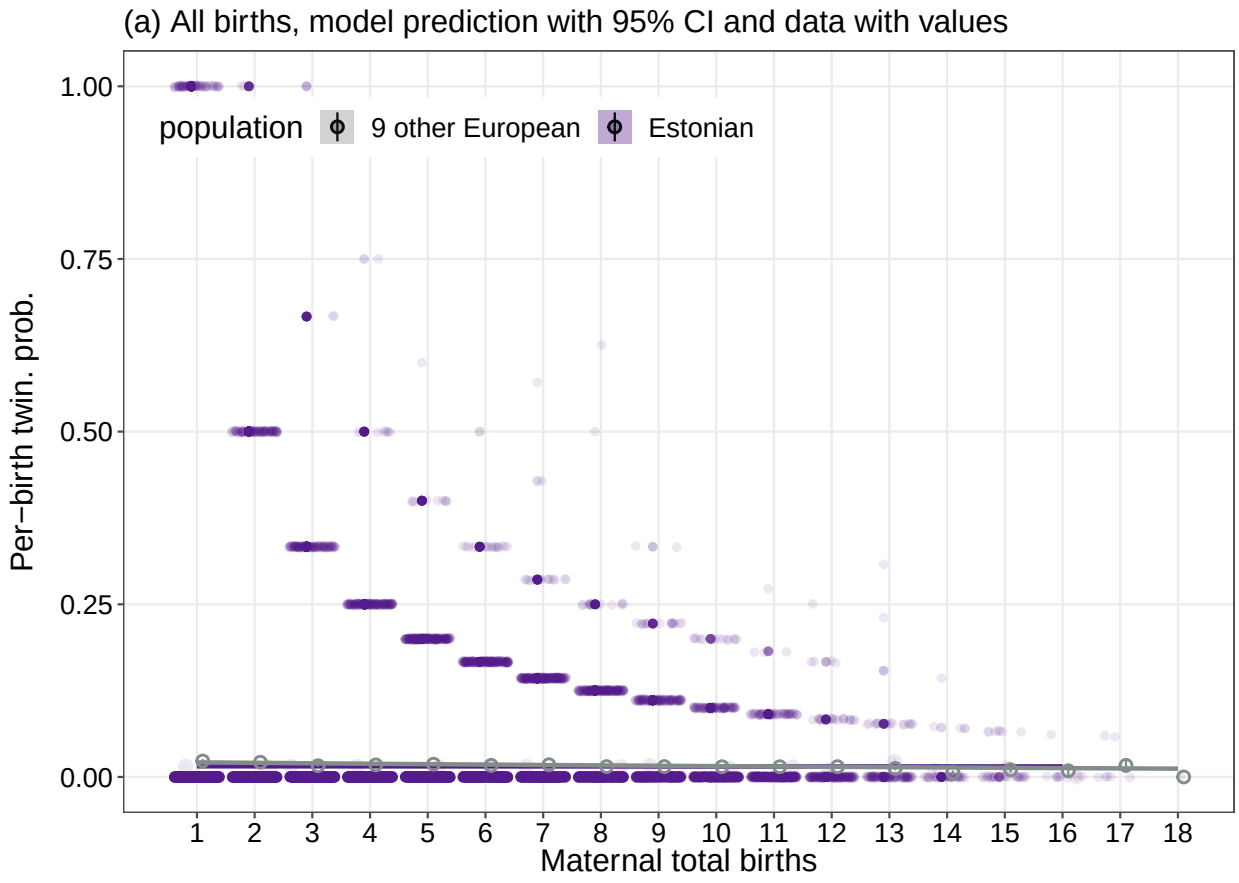
```r
fig2_EE_plot_data <- dmm_EE_fit$results
fig2_orig_plot_data <- dmm_orig_fit$results
```

```r
infoTxt <- "model prediction with 95% CI and data with values"

fig2_ext_orig <- ggplot() +
  geom_line(data=fig2_EE_plot_data,
              aes(y = estimate,  x=births_total, color=estLab),  size = 1) +
  geom_point(data=dmm_EE[dmm_EE$births_total <17, ],
               aes(x=births_total, y=prob_twin, color=estLab, fill = estLab),
               alpha=0.1,
               position = position_nudge(x = -0.1)) +
    stat_summary(data=dmm_EE[dmm_EE$births_total <17, ],
                 aes(x=births_total, y=prob_twin, color=estLab, fill = estLab),
                 alpha=0.1,
                 position = position_jitter(),
                 fun.data=mean_se) +

    geom_ribbon(data=fig2_EE_plot_data,
                aes(y = estimate,  x=births_total, ymin = lwr, ymax = upr,
                    color=estLab, fill = estLab),
                alpha = 0.3) +
    geom_line(data=fig2_orig_plot_data,
              aes(y = estimate,  x=births_total, color=westLab),  size = 1) +
  geom_point(data=dmm_orig[dmm_orig$births_total <19, ],
               aes(x=births_total, y=prob_twin, color=estLab, fill = estLab),
               alpha=0.1,
               position = position_jitter()) +
    stat_summary(data=dmm_orig[dmm_orig$births_total <19, ],
                 aes(x=births_total, y=prob_twin,
                     color=westLab, fill=westLab),
                 alpha=1,
                 shape = 1,
                 position = position_nudge(x = 0.1),
                 fun.data=mean_se) +
    geom_ribbon(data=fig2_orig_plot_data,
                aes(y = estimate,  x=births_total, ymin = lwr, ymax = upr,
                    fill=westLab),
                alpha = 0.1) +
    ggplot2::scale_x_continuous(breaks = 1:18) +

    ggplot2::coord_cartesian() +
    labs(subtitle = paste0("(a) All births, ", infoTxt),
         y="Per-birth twin. prob.",
         x="Maternal total births")
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
p2 <- fig2_ext_orig + base_theme(larger=8) + scale_color_manual(values=bc) +
  scale_fill_manual(values=bc) + guides(color="none") + labs(fill = "population")

p2
```

```
## Warning: Removed 1 rows containing missing values (‘geom_segment()‘).
```



The first plot clearly illustrates that we cannot plot the whole data itself on the plot, as there are women who only gave birth to twins. We can plot a subset (e.g. those with at least 10 children) but this would lose the point of the plot. There is just so much data that is semi-categorical that we cannot think of a way put into a plot so that it gives additional information.

Comment 4: Dizygotic Twins

We can do the same analysis for dizygotic twins.

```r
#re-import the data
dbm_EE <- readRDS("./data/data_births_all_EE.rds")
```

```r
#take out mothers that had twins of single sex
single_sex_twin_mothers <- as.character(unlist(unique(
  dbm_EE[dbm_EE$twin & dbm_EE$cf_sex != "mixed", "maternal_id"])
  ))

dbmEE_wo_single_sex_twins <- dbm_EE[!(dbm_EE$maternal_id %in%
                                        single_sex_twin_mothers),]
```

```r
#the twinR package expects population to be present
dbmEE_wo_single_sex_twins$pop <- "Estonia"

## Expand the birth level data for the fit of statistical models:
dbmEE_wo_single_sex_twins <- twinR::expand_data(dbmEE_wo_single_sex_twins)

dbmEE_wo_single_sex_twins <- add_last_birth(dbmEE_wo_single_sex_twins)

dbmEE_wo_single_sex_twins_not_last <- dbmEE_wo_single_sex_twins[!dbmEE_wo_single_sex_twins$last,]

#make the aggregates
dmm_EE_mixed <- twinR::aggregate_data(dbmEE_wo_single_sex_twins)
dmm_EE_mixed$prob_twin <- dmm_EE_mixed$twin_total / dmm_EE_mixed$births_total

dmm_EE_mixed_nl <- twinR::aggregate_data(dbmEE_wo_single_sex_twins_not_last)
dmm_EE_mixed_nl$prob_twin <- dmm_EE_mixed_nl$twin_total / dmm_EE_mixed_nl$births_total
```

Full Data

```r
options(scipen = 99)
```

```r
## Estonia
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total"
dmm_EE_mixed_fit <- fitPredictions(dmm_EE_mixed, formula)
```

```
## Warning in fitPredictions(dmm_EE_mixed, formula): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_totaldmm_EE_mixed_fit.rds
##  If you want to re-run this step delete the file or change the saveDir!

## Warning in get_predictions(predDataFname, fit, dataset, args, save): Pre-computed predictions returne
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_totaldmm_EE_mixeddata_fig.rds
##  If you want to re-run this step delete the file or change the saveDir!
```

```r
knitr::kable(build_fit_summary.table(dmm_EE_mixed_fit$fit))
```

```
## Warning: Use of .data in tidyselect expressions was deprecated in tidyselect 1.2.0.
## i Please use ‘"Estimate"‘ instead of ‘.data$Estimate‘
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
## Warning: Use of .data in tidyselect expressions was deprecated in tidyselect 1.2.0.
## i Please use '"object"' instead of '.data$object'
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

## Warning: Use of .data in tidyselect expressions was deprecated in tidyselect 1.2.0.
## i Please use '"name"' instead of '.data$name'
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

## Warning: Use of .data in tidyselect expressions was deprecated in tidyselect 1.2.0.
## i Please use '"value"' instead of '.data$value'
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-5.07	0.044	-115
	births_total	-0.0146	0.00843	-1.73
response family	binomial with logit link			
fit info	number of model parameters	2		
	marginal log Likelihood	-11192		
	marginal AIC	22388		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	121779		

No Last Births Data

```
## Estonia
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total"
dmm_EE_mixed_nl_fit <- fitPredictions(dmm_EE_mixed_nl, formula)
```

```
## Warning in fitPredictions(dmm_EE_mixed_nl, formula): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_totaldmm_EE_mixed_nl_fit.rds
##  If you want to re-run this step delete the file or change the saveDir!

## Warning in get_predictions(predDataFname, fit, dataset, args, save): Pre-computed predictions returne
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_totaldmm_EE_mixed_nldata_fig.rds
##  If you want to re-run this step delete the file or change the saveDir!
```

```
knitr::kable(build_fit_summary.table(dmm_EE_mixed_nl_fit$fit))
```

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-5.27	0.0549	-96
	births_total	-0.00123	0.0112	-0.11

Type	Variable	Value	Cond. SE	t-value
response family	binomial with logit link			
fit info	number of model parameters	2		
	marginal log Likelihood	-7104		
	marginal AIC	14212		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	94662		

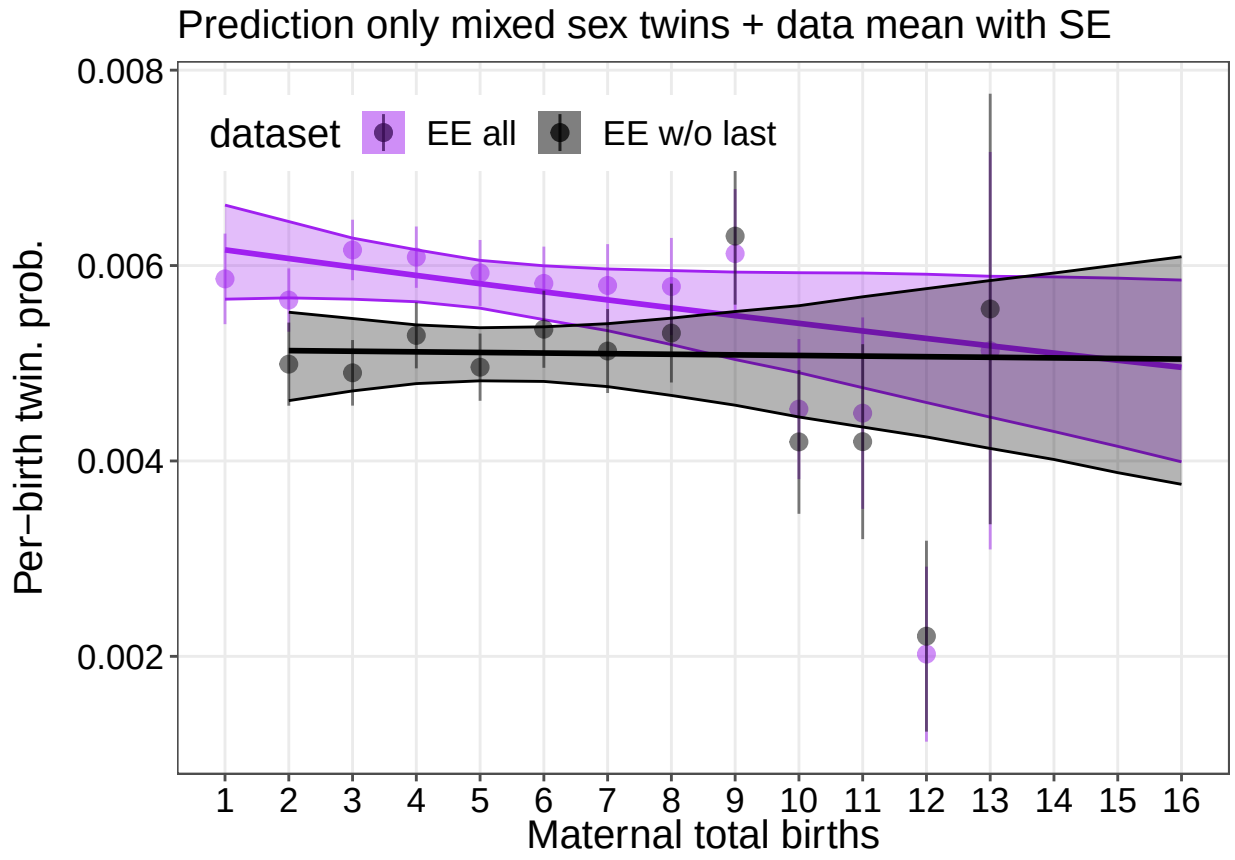
Plot the Predictions

```
fig2_EE_plot_data <- dmm_EE_mixed_fit$results
fig2_EE_plot_data_nl <- dmm_EE_mixed_nl_fit$results
fig2_EE_plot_data_nl$births_total <- fig2_EE_plot_data_nl$births_total + 1
```

```
dmm_EE_mixed_nl_plot <- dmm_EE_mixed_nl
dmm_EE_mixed_nl_plot$births_total <- dmm_EE_mixed_nl_plot$births_total + 1
```

```
#colors
bc <- c("purple", "black")
fig2_ext_orig <- ggplot() +
  geom_line(data=fig2_EE_plot_data,
              aes(y = estimate,  x=births_total, color="EE all"),
             linewidth = 1) +
    stat_summary(data=dmm_EE_mixed[dmm_EE_mixed$births_total <14, ],
                 aes(x=births_total, y=prob_twin, color="EE all",
                     fill = "EE all"), alpha=0.5,
                 fun.data=mean_se) +
    geom_ribbon(data=fig2_EE_plot_data,
                aes(y = estimate,  x=births_total,
                    ymin = lwr, ymax = upr, color="EE all", fill = "EE all"),
                alpha = 0.3) +
    geom_line(data=fig2_EE_plot_data_nl,
              aes(y = estimate,  x=births_total, color="EE w/o last"),
              linewidth = 1) +
    stat_summary(data=dmm_EE_mixed_nl_plot[dmm_EE_mixed_nl_plot$births_total <14,],
                 aes(x=births_total, y=prob_twin, color="EE w/o last",
                     fill = "EE w/o last"), alpha=0.5,
                 fun.data=mean_se) +
    geom_ribbon(data=fig2_EE_plot_data_nl,
                aes(y = estimate,  x=births_total, ymin = lwr,
                    ymax = upr, color="EE w/o last", fill = "EE w/o last"),
                alpha = 0.3) +
    ggplot2::scale_x_continuous(breaks = 1:18) +
    ggplot2::scale_y_continuous(breaks = seq(0,0.03, by=0.002)) +
    ggplot2::coord_cartesian() +
    labs(subtitle = "Prediction only mixed sex twins + data mean with SE",
         y="Per-birth twin. prob.",
         x="Maternal total births")
p2 <- fig2_ext_orig + base_theme(larger=8) + scale_color_manual(values=bc) +
  scale_fill_manual(values=bc) + guides(color="none") + labs(fill = "dataset")
```

```
p2
```



The mixed sex twinning rate seem to remain constant if last births are excluded from the data and slightly decline if last births are included. The decline seems not to be significant.

Comment 5: Illegitimate Maternities

We can add a predictor to the model describing if the women had first birth or first conception before being married. We will do both to investigate if illegitimate maternities affect twinning rate.

```
#add the illigimate children columns
illigimate_birth <- unique(data_births_monthly_EE[c( "illigimate_birth",
                                                     "maternal_id")])
illigimate_birth <- setNames(illigimate_birth$illigimate_birth,
                             illigimate_birth$maternal_id)

illigimate_conception <- unique(data_births_monthly_EE[c( "illigimate_conception","maternal_id")])
illigimate_conception <- setNames(illigimate_conception$illigimate_conception, illigimate_conception$ma
```

```
dmm_EE$illigimate_conception <- illigimate_conception[dmm_EE$maternal_id]

dmm_EE$illigimate_birth <- illigimate_birth[dmm_EE$maternal_id]
```

```
dmm_EE_nl$illigimate_conception <- illigimate_conception[dmm_EE_nl$maternal_id]

dmm_EE_nl$illigimate_birth <- illigimate_birth[dmm_EE_nl$maternal_id]
```

```
cat("First birth before marriage:\n")
```

```
## First birth before marriage:
```

```
table(dmm_EE$illigimate_birth)
```

```
##
##  FALSE   TRUE
## 108483  17092
```

```
cat("First conception before marriage:\n")
```

```
## First conception before marriage:
```

```
table(dmm_EE$illigimate_conception)
```

```
##
## FALSE  TRUE
## 89896 35679
```

With All Data

```
## Estonia
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total + illigimate_birth"
dmm_EE_fit_birth <- fitPredictions(dmm_EE, formula)
```

```
## Warning in fitPredictions(dmm_EE, formula): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+illigimate_birthdmm_EE_fit.rds
##  If you want to re-run this step delete the file or change the saveDir!

## Warning in get_predictions(predDataFname, fit, dataset, args, save): Pre-computed predictions returne
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+illigimate_birthdmm_EEdata_fig.rd
##  If you want to re-run this step delete the file or change the saveDir!
```

```
knitr::kable(build_fit_summary.table(dmm_EE_fit_birth$fit))
```

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.14	0.0276	-150
	births_total	-0.00255	0.00504	-0.506
	illigimate_birthTRUE	-0.0759	0.0425	-1.79
response family	binomial with logit link			

Type	Variable	Value	Cond. SE	t-value
fit info	number of model parameters	3		
	marginal log Likelihood	-24445		
	marginal AIC	48895		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	125575		

```
## Estonia
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total + illigimate_conception"
dmm_EE_fit_conception <- fitPredictions(dmm_EE, formula)
```

```
## Warning in fitPredictions(dmm_EE, formula): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+illigimate_conceptiondmm_EE_fit.r
##  If you want to re-run this step delete the file or change the saveDir!
```

```
## Warning in get_predictions(predDataFname, fit, dataset, args, save): Pre-computed predictions returne
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+illigimate_conceptiondmm_EEdata_r
##  If you want to re-run this step delete the file or change the saveDir!
```

```
knitr::kable(build_fit_summary.table(dmm_EE_fit_conception$fit))
```

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.15	0.0278	-149
	births_total	-0.00178	0.00502	-0.356
	illigimate_conceptionTRUE	-0.00944	0.0283	-0.334
response family	binomial with logit link			
fit info	number of model parameters	3		
	marginal log Likelihood	-24446		
	marginal AIC	48898		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	125575		

No Last Births Data

```
## Estonia
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total + illigimate_birth"
dmm_EE_nl_fit_birth <- fitPredictions(dmm_EE_nl, formula)
```

```
## Warning in fitPredictions(dmm_EE_nl, formula): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+illigimate_birthdmm_EE_nl_fit.rds
##  If you want to re-run this step delete the file or change the saveDir!
```

```
## Warning in get_predictions(predDataFname, fit, dataset, args, save): Pre-computed predictions returne
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+illigimate_birthdmm_EE_nldata_fig
##  If you want to re-run this step delete the file or change the saveDir!
```

```
knitr::kable(build_fit_summary.table(dmm_EE_nl_fit_birth$fit))
```

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.37	0.0338	-129
	births_total	0.018	0.00655	2.74
	illigimate_birthTRUE	0.0364	0.0547	0.665
response family	binomial with logit link			
fit info	number of model parameters	3		
	marginal log Likelihood	-15893		
	marginal AIC	31792		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	98183		

```
## Estonia
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total + illigimate_conception"
dmm_EE_fit_nl_conception <- fitPredictions(dmm_EE_nl, formula)
```

```
## Warning in fitPredictions(dmm_EE_nl, formula): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+illigimate_conceptiondmm_EE_nl_f
##  If you want to re-run this step delete the file or change the saveDir!

## Warning in get_predictions(predDataFname, fit, dataset, args, save): Pre-computed predictions returne
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+illigimate_conceptiondmm_EE_nlda
##  If you want to re-run this step delete the file or change the saveDir!
```

```
knitr::kable(build_fit_summary.table(dmm_EE_fit_nl_conception$fit))
```

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.37	0.0344	-127
	births_total	0.0174	0.00656	2.65
	illigimate_conceptionTRUE	0.0443	0.0355	1.25
response family	binomial with logit link			
fit info	number of model parameters	3		
	marginal log Likelihood	-15892		
	marginal AIC	31790		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	98183		

Illigimate Birth Figures for Estonian mothers With and Without Last Birth

```
#' Plot Binary Predictions
#'
#' This function generates a ggplot based on the provided parameters.
#'
#' @param plot_data The data to be plotted.
#' @param summary_data The data to be summarized.
#' @param color_var The color variable for the plot.
```

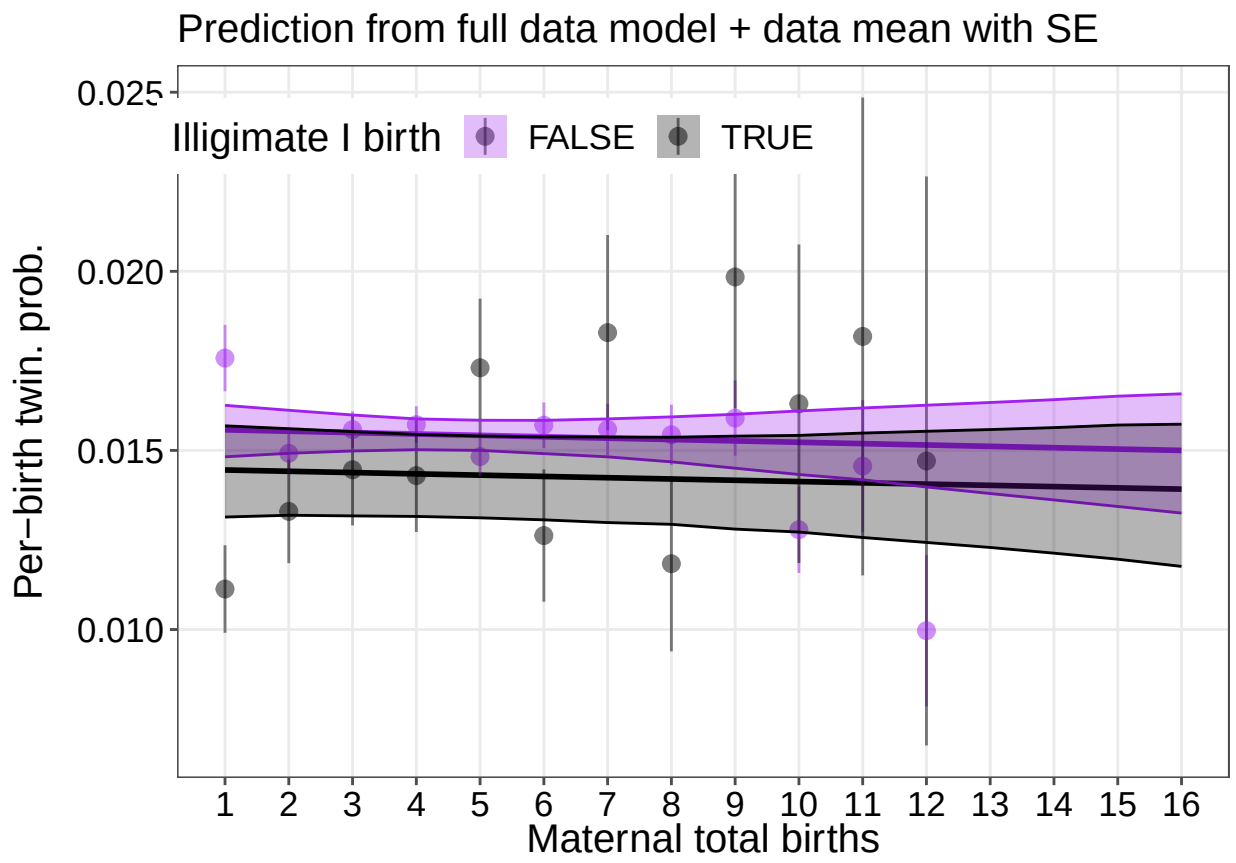
```r
#' @param threshold The threshold for births_total mean data to display.
#'
#' @return A ggplot object.
#'
plot_predictions <- function(plot_data, summary_data, color_var, threshold = 13) {
  fig2_ext_orig <- ggplot() +
    geom_line(data=plot_data,
              aes_string(y = "estimate",  x="births_total", color=color_var),
              linewidth = 1) +
    stat_summary(data=summary_data[summary_data$births_total < threshold, ],
                 aes_string(x="births_total", y="prob_twin", color=color_var,
                            fill = color_var), alpha=0.5,
                 fun.data=mean_se) +
    geom_ribbon(data=plot_data,
                aes_string(y = "estimate",  x="births_total", ymin = "lwr",
                           ymax = "upr",
                           color=color_var, fill = color_var),
                alpha = 0.3) +
    ggplot2::scale_x_continuous(breaks = 1:18) +
    ggplot2::scale_y_continuous(breaks = seq(0,0.03, by=0.005)) +
    ggplot2::coord_cartesian()
  return(fig2_ext_orig)
}
```

```r
fig2_EE_plot_data <- dmm_EE_fit_birth$results
fig2_EE_plot_data_nl <- dmm_EE_nl_fit_birth$results
fig2_EE_plot_data_nl$births_total <- fig2_EE_plot_data_nl$births_total + 1
```

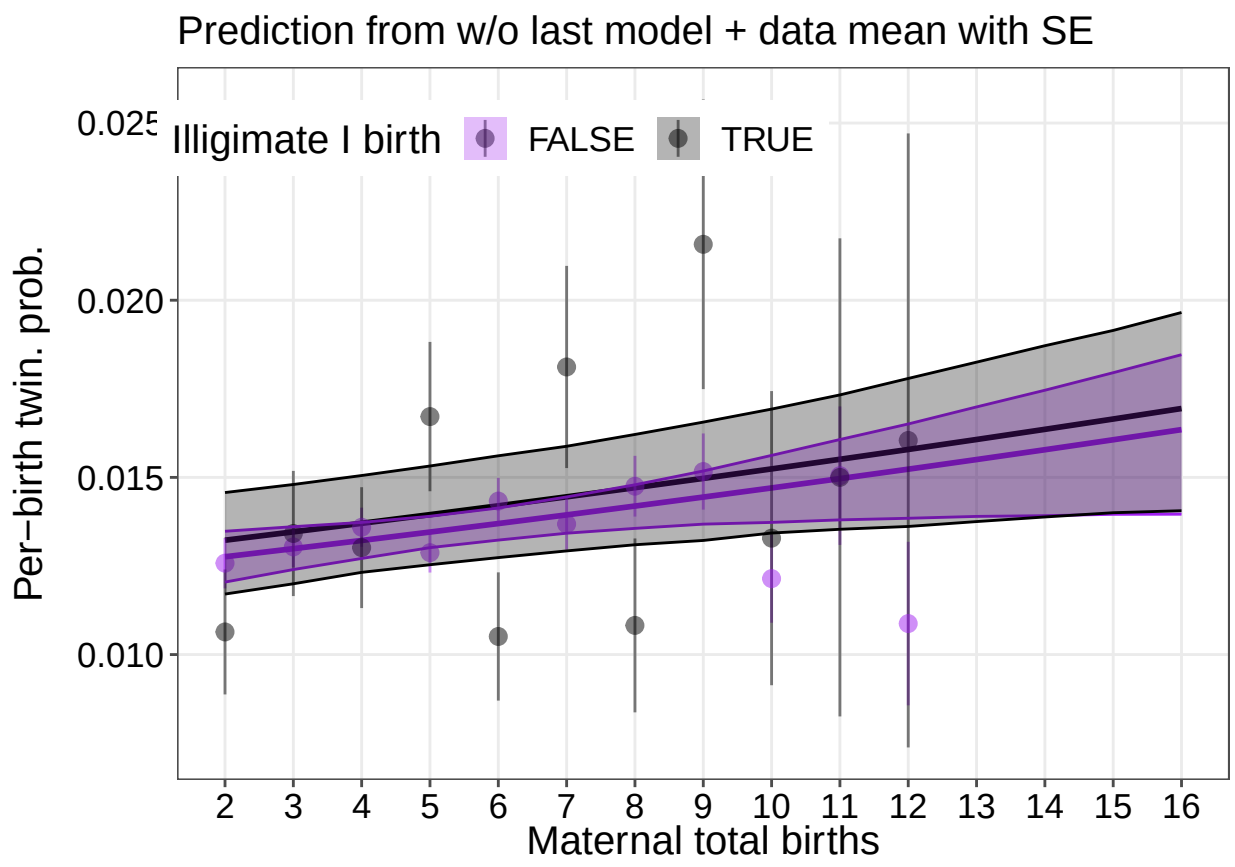
```r
dmm_EE_nl_plot <- dmm_EE_nl
dmm_EE_nl_plot$births_total <- dmm_EE_nl_plot$births_total + 1
```

```r
plot_predictions(plot_data = fig2_EE_plot_data,
                 summary_data = dmm_EE,
                 color_var = "illigimate_birth") +
  labs(subtitle = "Prediction from full data model + data mean with SE",
         y="Per-birth twin. prob.",
         x="Maternal total births") +
  base_theme(larger=8) + scale_color_manual(values=bc) +
  scale_fill_manual(values=bc) + guides(color="none") +
  labs(fill = "Illigimate I birth")
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



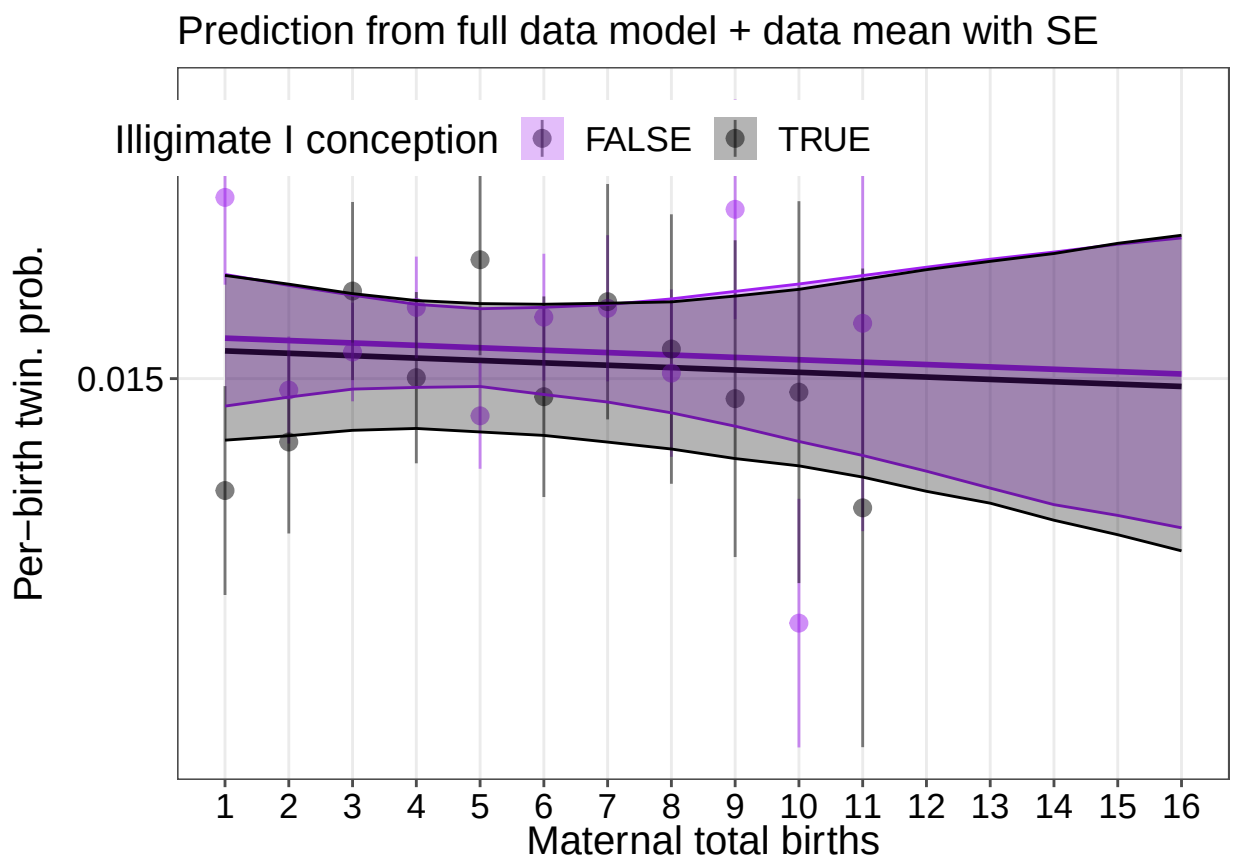

```
plot_predictions(plot_data = fig2_EE_plot_data_nl,
                 summary_data = dmm_EE_nl_plot,
                 color_var = "illigimate_birth") +
    labs(subtitle = "Prediction from w/o last model + data mean with SE",
         y="Per-birth twin. prob.",
         x="Maternal total births")+
  base_theme(larger=8) + scale_color_manual(values=bc) +
  scale_fill_manual(values=bc) + guides(color="none") +
  labs(fill = "Illigimate I birth")
```



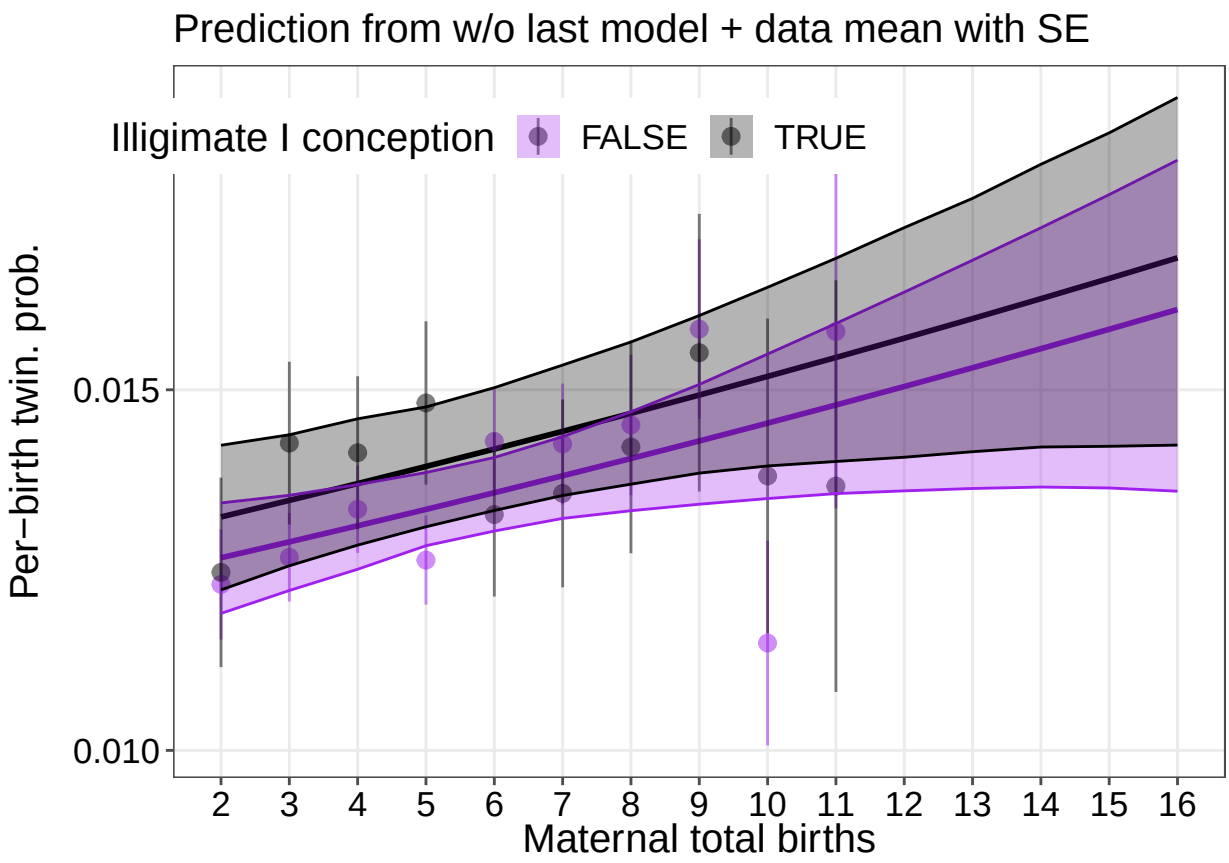
Illigimate Conception Figures for Estonian mothers With and Without Last Birth

```
fig2_EE_plot_data <- dmm_EE_fit_conception$results
fig2_EE_plot_data_nl <- dmm_EE_fit_nl_conception$results
fig2_EE_plot_data_nl$births_total <- fig2_EE_plot_data_nl$births_total + 1
```

```
plot_predictions(plot_data = fig2_EE_plot_data,
                 summary_data = dmm_EE,
                 color_var = "illigimate_conception",
                 threshold = 12) +
  labs(subtitle = "Prediction from full data model + data mean with SE",
         y="Per-birth twin. prob.",
         x="Maternal total births") +
  base_theme(larger=8) + scale_color_manual(values=bc) +
  scale_fill_manual(values=bc) + guides(color="none") +
  labs(fill = "Illigimate I conception")
```




```
plot_predictions(plot_data = fig2_EE_plot_data_nl,
                 summary_data = dmm_EE_nl_plot,
                 color_var = "illigimate_conception",
                 threshold = 12) +
    labs(subtitle = "Prediction from w/o last model + data mean with SE",
         y="Per-birth twin. prob.",
         x="Maternal total births")+
  base_theme(larger=8) + scale_color_manual(values=bc) +
  scale_fill_manual(values=bc) + guides(color="none") +
  labs(fill = "Illigimate I conception")
```


Prediction from w/o last model + data mean with SE
Illigimate I conception
FALSE
TRUE
Per−birth twin. prob.
0.015
0.010
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Maternal total births

```
#END
```