Suplementary Material 1: Code to Reproduce Analysis and Plots

Richard Meitern

2023-03-23

Introduction

The aim of this document is to display the R code needed the reproduce the findings of the main text. This documents needs the twinR package to be installed to run.

```
#cleanup memory
gcstuff <- gc(verbose=FALSE); rm(gcstuff);</pre>
#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:
nb_cores <- min(c(50L, parallel::detectCores() - 1))</pre>
## 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 Estonia

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.

```
#Import and preprocess 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</pre>
```

Data Import 8 European Populations

```
##Import and pre-process twinR package data
## Filter the raw data to only keep data with monthly resolution:
data births monthly <- twinR::filter data(twinR::data births all)</pre>
## Expand the birth level data for the fit of statistical models:
data_births_monthly <- twinR::expand_data(data_births_monthly)</pre>
data_births_monthly <- add_last_birth(data_births_monthly)</pre>
data births monthly not last <- data births monthly[!data births monthly$last,]
dmm_orig <- twinR::aggregate_data(data_births_monthly)</pre>
dmm_orig$prob_twin <- dmm_orig$twin_total /dmm_orig$births_total</pre>
dmm_orig_nl <- twinR::aggregate_data(data_births_monthly_not_last)</pre>
dmm_orig_nl$prob_twin <- dmm_orig_nl$twin_total /dmm_orig_nl$births_total</pre>
#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),</pre>
                  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))
```

Table 1: Summary data on the Estonian and other European populations for both all births and wihout last birth for same mothers

	8 European populations all	8 European populations without	Estonia populations all	Estonia populations
	births	last birth	births	without last birth
Populations	8 European	8 European	Estonia	Estonia
Dataset	All births	Without last birth	All births	Without last birth
Maternal birth period	1700-1899	1700-1899	1850-1899	1850-1899
1/4 and $3/4$ quantiles	1786-1847	1785-1847	1873-1890	1872-1890
for maternal birth period				
Age at first birth (mean-SE)	24.68-0.013	24.36-0.014	25.06-0.007	24.5-0.0075
Mothers	21290	18520	125575	98183
Non-twinners	19656	17344	119511	94386
Twinners	1634	1176	6064	3797
Twinner rate $(\%)$	77	64	48	39
Offspring birth period	1720-1945	1720-1941	1868-1948	1868-1943
1/4 and $3/4$ quantiles	1817-1879	1815-1877	1903-1921	1901-1918
for offspring birth period				
Births	105833	84543	417418	291843
Singleton births	104069	83276	411026	287874
Twin births	1764	1267	6392	3969
Twinning rate $(\%)$	17	15	15	14
Total births	1-7-18	1-6-17	1-4-16	1-4-15
$(\min\text{-median-max})$				
Total births (mean-SE)	6.71-0.0089	6.15-0.0093	4.75-0.0039	4.36-0.0044

Fitting models

```
# 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)</pre>
```

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 return

./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!

knitr::kable(build_fit_summary.table(dmm_EE_fit\$fit))

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.2	0.03	-154.49
	births_total	0.0	0.01	-0.36
response family	binomial with logit link			
fit info	number of model parameters	2.0		
	marginal log Likelihood	-24446.1		
	marginal AIC	48896.3		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	125575.0		

```
## Estonia - birth level data
formula <- "twin ~ 1 + poly(cbind(age, parity), 3) + (1|maternal_id)"
dbm_EE_fit <- fitPredictions(data_births_monthly_EE, formula, nb_boot = 0)</pre>
```

```
## Warning in fitPredictions(data_births_monthly_EE, formula, nb_boot = 0): Pre-computed fit returned f
## ./data/predictions/twin~1+poly(cbind(age,parity),3)+(1_x_maternal_id)data_births_monthly_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 return
./data/predictions/twin~1+poly(cbind(age,parity),3)+(1_x_maternal_id)data_births_monthly_EEdata_fig.
If you want to re-run this step delete the file or change the saveDir!

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

[one-time computation of covariance matrix, which may be slow]

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.23	0.03	-146.78
	poly(cbind(age, parity), 3)1.0	124.69	23.34	5.34
	poly(cbind(age, parity), 3)2.0	-115.54	14.20	-8.14
	poly(cbind(age, parity), 3)3.0	-67.80	14.91	-4.55
	poly(cbind(age, parity), 3)0.1	44.06	35.86	1.23
	poly(cbind(age, parity), 3)1.1	6070.97	20377.08	0.30

Type	Variable	Value	Cond. SE	t-value
	poly(cbind(age, parity), 3)2.1	13475.80	14544.20	0.93
	poly(cbind(age, parity), 3)0.2	-6.92	26.95	-0.26
	poly(cbind(age, parity), 3)1.2	-11971.11	14297.44	-0.84
	poly(cbind(age, parity), 3)0.3	-12.04	12.98	-0.93
random effects	variance between name	0.52		
response family	binomial with logit link			
fit info	number of model parameters	11.00		
	marginal log Likelihood	-32714.88		
	marginal AIC	65451.75		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	417418.00		

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

```
## 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 return
./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!

knitr::kable(build_fit_summary.table(dmm_orig_fit\$fit))

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-3.83	0.10	-36.7
	births_total	-0.03	0.01	-3.9
random effects	variance between name	0.06		
response family	binomial with logit link			
fit info	number of model parameters	3.00		
	marginal log Likelihood	-5993.12		
	marginal AIC	11992.24		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	21290.00		

```
## TwinR - birth level data
formula <- "twin ~ 1 + poly(cbind(age, parity), 3) + (1|maternal_id) + (1|pop)"
dbm_orig_fit <- fitPredictions(data_births_monthly, formula, nb_boot = 0)</pre>
```

```
## Warning in fitPredictions(data_births_monthly, formula, nb_boot = 0): Pre-computed fit returned from
## ./data/predictions/twin~1+poly(cbind(age,parity),3)+(1_x_maternal_id)+(1_x_pop)data_births_monthly_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 return
./data/predictions/twin~1+poly(cbind(age,parity),3)+(1_x_maternal_id)+(1_x_pop)data_births_monthlyda
If you want to re-run this step delete the file or change the saveDir!

knitr::kable(build_fit_summary.table(dbm_orig_fit\$fit))

[one-time computation of covariance matrix, which may be slow]

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.10	0.11	-36.38
	poly(cbind(age, parity), 3)1.0	73.70	31.07	2.37
	poly(cbind(age, parity), 3)2.0	-61.18	17.22	-3.55
	poly(cbind(age, parity), 3)3.0	-47.11	16.42	-2.87
	poly(cbind(age, parity), 3)0.1	-0.94	40.14	-0.02
	poly(cbind(age, parity), 3)1.1	-4005.74	11889.93	-0.34
poly(cbind(age, parity), 3)2.1	7709.69	8669.07	0.89	
poly(cbind(age, par	poly(cbind(age, parity), 3)0.2	18.98	27.35	0.69
	poly(cbind(age, parity), 3)1.2	-4380.22	8120.86	-0.54
	poly(cbind(age, parity), 3)0.3	-18.63	14.09	-1.32
random effects	variance between name	0.48		
	variance between name	0.06		
response family	binomial with logit link			
fit info	number of model parameters	12.00		
	marginal log Likelihood	-8828.38		
	marginal AIC	17680.76		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	105833.00		

No Last Births Data

```
## Estonia
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total"
dmm_EE_nl_fit <- fitPredictions(dmm_EE_nl, formula, predict = T)</pre>
```

- ## Warning in fitPredictions(dmm_EE_nl, formula, predict = T): Pre-computed fit returned from file:
- ## ./data/predictions/cbind(twin_total,singleton_total)~1+births_totaldmm_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 return
- ## ./data/predictions/cbind(twin_total,singleton_total)~1+births_totaldmm_EE_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_nl_fit\$fit))

Value	Cond. SE	t-value
-4.36	0.03	-131.0
0.02	0.01	2.7
2.00		
-15892.99		
31789.98		
	-4.36 0.02 2.00 -15892.99	-4.36 0.03 0.02 0.01 2.00 -15892.99

Type	Variable	Value	Cond. SE	t-value
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	98183.00		

```
## Warning in fitPredictions(data_births_monthly_EE_not_last, formula, predict = T, : Pre-computed fit :
## ./data/predictions/twin~1+poly(cbind(age,parity),3)+(1_x_maternal_id)data_births_monthly_EE_not_last
## 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 return
./data/predictions/twin~1+poly(cbind(age,parity),3)+(1_x_maternal_id)data_births_monthly_EE_not_last
If you want to re-run this step delete the file or change the saveDir!

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

[one-time computation of covariance matrix, which may be slow]

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.31	0.04	-118.27
	poly(cbind(age, parity), 3)1.0	118.86	22.47	5.29
	poly(cbind(age, parity), 3)2.0	-33.64	14.46	-2.33
	poly(cbind(age, parity), 3)3.0	-21.94	13.98	-1.57
	poly(cbind(age, parity), 3)0.1	54.44	37.75	1.44
	poly(cbind(age, parity), 3)1.1	-11865.19	17925.73	-0.66
	poly(cbind(age, parity), 3)2.1	9736.49	11627.66	0.84
	poly(cbind(age, parity), 3)0.2	19.59	29.17	0.67
	poly(cbind(age, parity), 3)1.2	-11242.39	11738.46	-0.96
	poly(cbind(age, parity), 3)0.3	-3.45	12.90	-0.27
random effects	variance between name	0.49		
response family	binomial with logit link			
fit info	number of model parameters	11.00		
	marginal log Likelihood	-20819.80		
	marginal AIC	41661.59		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	291843.00		

```
## TwinR
formula <- "cbind(twin_total, singleton_total) ~ 1 + births_total + (1|pop)"
dmm_orig_nl_fit <- fitPredictions(dmm_orig_nl, formula, predict = T)</pre>
```

```
## Warning in fitPredictions(dmm_orig_nl, formula, predict = T): Pre-computed fit returned from file:
## ./data/predictions/cbind(twin_total,singleton_total)~1+births_total+(1_x_pop)dmm_orig_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 return

./data/predictions/cbind(twin_total,singleton_total)~1+births_total+(1_x_pop)dmm_orig_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_orig_nl_fit\$fit))

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.14	0.12	-34.85
	births_total	0.00	0.01	-0.09
random effects	variance between name	0.07		
response family	binomial with logit link			
fit info	number of model parameters	3.00		
	marginal log Likelihood	-4490.84		
	marginal AIC	8987.68		
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	18520.00		

```
## TwinR - birth level data
formula <- "twin ~ 1 + poly(cbind(age, parity), 3) + (1|maternal_id) + (1|pop)"
dbm_orig_nl_fit <- fitPredictions(data_births_monthly_not_last, formula, nb_boot = 0)</pre>
```

```
## Warning in fitPredictions(data_births_monthly_not_last, formula, nb_boot = 0): Pre-computed fit retu
## ./data/predictions/twin~1+poly(cbind(age,parity),3)+(1_x_maternal_id)+(1_x_pop)data_births_monthly_n
## 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 return
./data/predictions/twin~1+poly(cbind(age,parity),3)+(1_x_maternal_id)+(1_x_pop)data_births_monthly_n
If you want to re-run this step delete the file or change the saveDir!

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

[one-time computation of covariance matrix, which may be slow]

Type	Variable	Value	Cond. SE	t-value
fixed effects	(Intercept)	-4.24	0.13	-33.43
	poly(cbind(age, parity), 3)1.0	85.90	31.19	2.75
	poly(cbind(age, parity), 3)2.0	-34.85	17.95	-1.94
	poly(cbind(age, parity), 3)3.0	-22.85	15.92	-1.43
	poly(cbind(age, parity), 3)0.1	-23.14	42.20	-0.55
	poly(cbind(age, parity), 3)1.1	4849.99	11442.24	0.42
	poly(cbind(age, parity), 3)2.1	2439.65	7666.86	0.32
	poly(cbind(age, parity), 3)0.2	7.32	29.51	0.25
	poly(cbind(age, parity), 3)1.2	1548.37	7323.01	0.21
	poly(cbind(age, parity), 3)0.3	-11.86	13.04	-0.91
random effects	variance between name	0.54		
	variance between name	0.07		
response family	binomial with logit link			
fit info	number of model parameters	12.00		
	marginal log Likelihood	-6480.49		
	marginal AIC	12984.99		

Type	Variable	Value	Cond. SE	t-value
	conditional AIC (cAIC)			
data info	number of fitted observations (N)	84543.00		

Plots

```
library(ggplot2)
#some nice colors
bc <- c("azure3","purple4", "black", "navy", "darkgoldenrod2", "springgreen3", "gray")
#use new base theme that displays also grid lines
source("./R/twinR_theme.R")</pre>
```

Fig 1A: Estonian vs TwinR Full Data

```
fig2_EE_plot_data <- dmm_EE_fit$results
fig2_orig_plot_data <- dmm_orig_fit$results</pre>
```

```
infoTxt <- "model prediction with 95% CI and data mean SE"
fig2_ext_orig <- ggplot() +
  geom_line(data=fig2_EE_plot_data,
              aes(y = estimate, x=births_total, color="Estonia"), size = 1) +
    stat_summary(data=dmm_EE[dmm_EE$births_total <17, ],</pre>
                 aes(x=births_total, y=prob_twin, color="Estonia", fill = "Estonia"),
                 alpha=0.5,
                 fun.data=mean_se) +
    geom_ribbon(data=fig2_EE_plot_data,
                aes(y = estimate, x=births_total, ymin = lwr, ymax = upr,
                    color="Estonia", fill = "Estonia"),
                alpha = 0.3) +
    geom_line(data=fig2_orig_plot_data,
             aes(y = estimate, x=births_total, color="8 European"), size = 1) +
    stat_summary(data=dmm_orig[dmm_orig$births_total <19, ],</pre>
                 aes(x=births_total, y=prob_twin,
                     color="8 European", fill="8 European"),
                 alpha=0.5,
                 fun.data=mean_se) +
    geom_ribbon(data=fig2_orig_plot_data,
                aes(y = estimate, x=births_total, ymin = lwr, ymax = upr,
                    fill="8 European"),
                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() +
    labs(subtitle = pasteO("All births, ", infoTxt),
         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 = "population")
```

p2

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

All births, model prediction with 95% CI and data mean SE

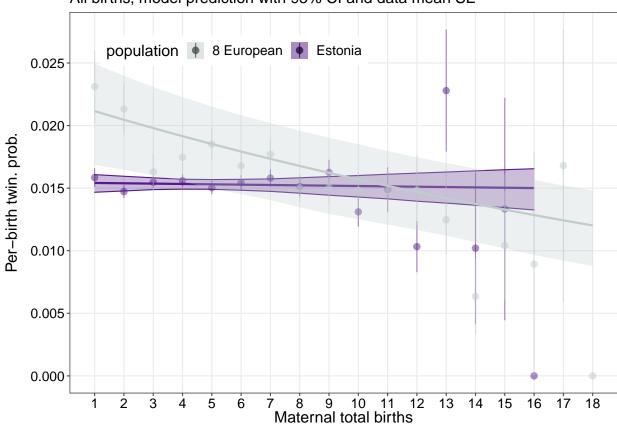
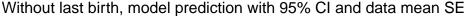


Fig 1B: Estonian vs Others No Last Birth

```
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="Estonia", fill = "Estonia"),
                alpha = 0.3) +
    geom_line(data=fig2_orig_plot_data_nl,
             aes(y = estimate, x=births_total, color="8 European"), size = 1) +
    stat_summary(data=dmm_orig_nl_plot,
                 aes(x=births_total, y=prob_twin,
                     color="8 European", fill="8 European"),
                 alpha=0.5,
                 fun.data=mean_se) +
    geom_ribbon(data=fig2_orig_plot_data_nl,
                aes(y = estimate, x=births_total, ymin = lwr, ymax = upr,
                   fill="8 European"),
                alpha = 0.3) +
    ggplot2::scale_x_continuous(breaks = 1:18, limits = c(1,NA)) +
    ggplot2::scale_y_continuous(breaks = seq(0,0.03, by=0.005)) +
    ggplot2::coord_cartesian(ylim=c(0,0.03)) +
    labs(subtitle = paste0("Without last birth, ", infoTxt),
         y="Per-birth twin. prob.",
         x="Maternal total births")
p3 <- fig2_ext_orig + base_theme(larger=8) + scale_color_manual(values=bc) +
  scale_fill_manual(values=bc) + guides(color="none") + labs(fill = "population")
рЗ
```

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



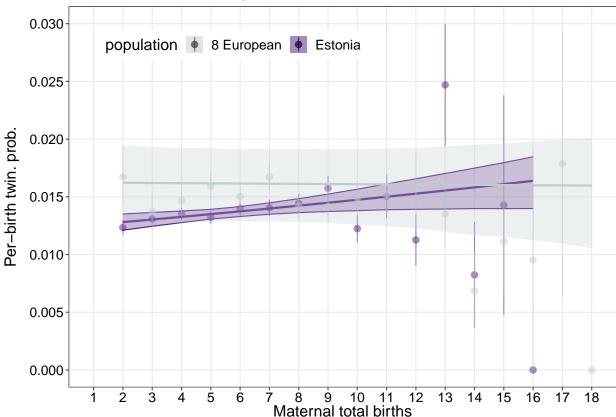


Figure 2: Model Mean Predictions from Birth Level Data

```
#use new draw_fig_4C function
source("./R/twinR_theme.R")
```

```
#extract predictions
fig4C_EE_plot_data <- dbm_EE_fit$results
fig4C_orig_plot_data$twin <- fig4C_EE_plot_data$estimates

fig4C_orig_plot_data <- dbm_orig_fit$results
fig4C_orig_plot_data$twin <- fig4C_orig_plot_data$estimates

fig4C_EE_nl_plot_data <- dbm_EE_nl_fit$results
fig4C_EE_nl_plot_data$twin <- fig4C_EE_nl_plot_data$estimates

fig4C_EE_nl_plot_data$twin <- fig4C_EE_nl_plot_data$parity + 1

fig4C_orig_nl_plot_data <- dbm_orig_nl_fit$results
fig4C_orig_nl_plot_data$twin <- fig4C_orig_nl_plot_data$estimates
fig4C_orig_nl_plot_data$twin <- fig4C_orig_nl_plot_data$estimates
fig4C_orig_nl_plot_data$twin <- fig4C_orig_nl_plot_data$parity + 1</pre>
```

```
rescale_cd <- scale_y_continuous(limits = c(NA, 0.035))
fig2a <- draw_fig_4C(fig4C_orig_plot_data) +</pre>
```

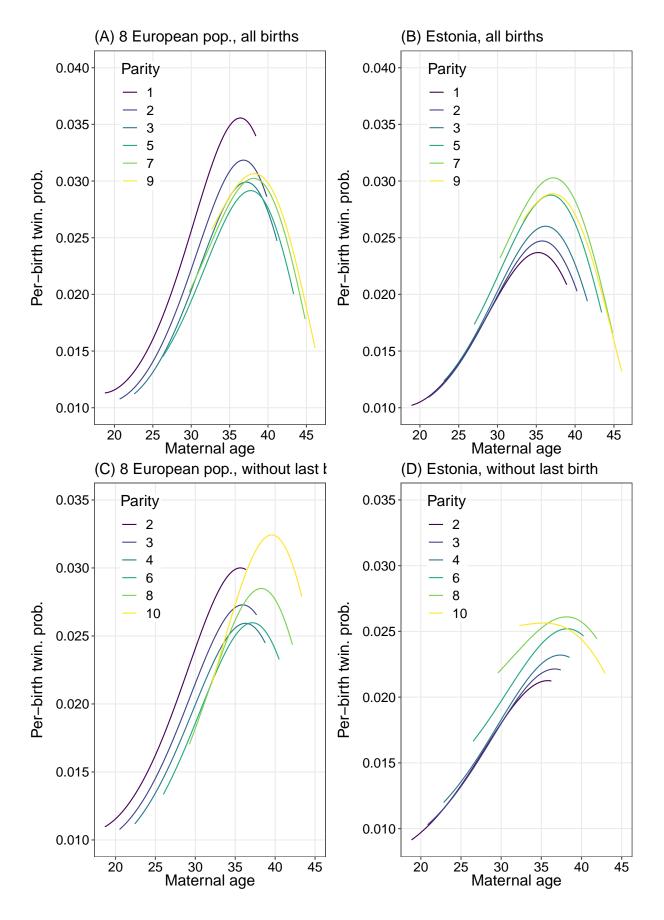
```
labs(subtitle = "(A) 8 European pop., all births")

fig2b <- draw_fig_4C(fig4C_EE_plot_data) +
    labs(subtitle = "(B) Estonia, all births")

fig2c <- draw_fig_4C(fig4C_orig_nl_plot_data, y_lims = c(0.01, 0.035)) +
    labs(subtitle = "(C) 8 European pop., without last birth")

fig2d <- draw_fig_4C(fig4C_EE_nl_plot_data,y_lims = c(0.01, 0.035)) +
    labs(subtitle = "(D) Estonia, without last birth")

gridExtra::grid.arrange(fig2a + guides(), fig2b, fig2c, fig2d, ncol=2)</pre>
```



#END