Contraceptive Methods and Age

Jessica Spencer

September 2024

Age on Contraceptive Prevalence

This data set is from a 1987 National Indonesia Contraceptive Prevalence Survey. All observations are married women who were definitely not pregnant or did not know yet. Questions on the survey covered topics regarding socio-economic status and general demographics.

1. Exploratoration of Dataset
2. Distribution and Correlation Visualization
3. Relative Odds of Contraceptive Method
4. Predictive Strength of Age on Contraceptive Method

```
library('ggvis')
library('tidyverse')
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()       masks stats::filter()
## x dplyr::lag()          masks stats::lag()
## x ggplot2::resolution() masks ggvis::resolution()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library('ggplot2')

df = read.csv('data/1987 Indonesia Contraception Prevalence Study.csv')
head(df)
```

```
##   Age Education Partner.Education Number.of.Children Religion...Islam
## 1  24         2                 3                  3                1
## 2  45         1                 3                 10                1
## 3  43         2                 3                  7                1
## 4  42         3                 2                  9                1
## 5  36         3                 3                  8                1
## 6  19         4                 4                  0                1
##   Currently.working Husband.Occupation Standard.of.Living Media.Exposure
## 1                 1                  2                  3              0
```

```
## 2                 1                 3                 4              0
## 3                 1                 3                 4              0
## 4                 1                 3                 3              0
## 5                 1                 3                 2              0
## 6                 1                 3                 3              0
##   Contraceptive.Method.Used
## 1                         1
## 2                         1
## 3                         1
## 4                         1
## 5                         1
## 6                         1
```

Variable Information

Variable Information:

```
Age - age of the woman
Education - level of education woman has received (1=low, 4=high)
Partner Education - level of education partner has received (1=low, 4=high)
Number of Children - number of kids mothered by woman
Religion=Islam - woman that identify as Muslim (0=No, 1=Yes)
Currently Working - woman is currently employed (0=Yes, 1=No)
Husbands Occupation - Not specified (categorical 1-4)
Standard of Living - based on the standard of living index (1=low, 4=high)
Media exposure - quality of media exposure (0=Good, 1=Not good)
Contraceptive Method Used - 1=No-use, 2=Long-term, 3=Short-term
```

Exploratory Data Analysis

- Description of Dataframe
- Missingness Check
- Distribution of Variables

```
print("Rows x Columns:")
```

```
## [1] "Rows x Columns:"
```

```
print(dim(df))
```

```
## [1] 1473   10
```

```
print("Feature Type:")# Data types of the variables
```

```
## [1] "Feature Type:"
```

```
print(sapply(df, class))
```

```
##                         Age                   Education         Partner.Education
##                   "integer"                   "integer"                 "integer"
##         Number.of.Children            Religion...Islam         Currently.working
##                   "integer"                   "integer"                 "integer"
##         Husband.Occupation          Standard.of.Living            Media.Exposure
##                   "integer"                   "integer"                 "integer"
## Contraceptive.Method.Used
##                   "integer"
```

```
print("Missing Values Per Feature:") #Checking for missing values in the columns
```

```
## [1] "Missing Values Per Feature:"
```

```
print(colSums(is.na(df)))
```

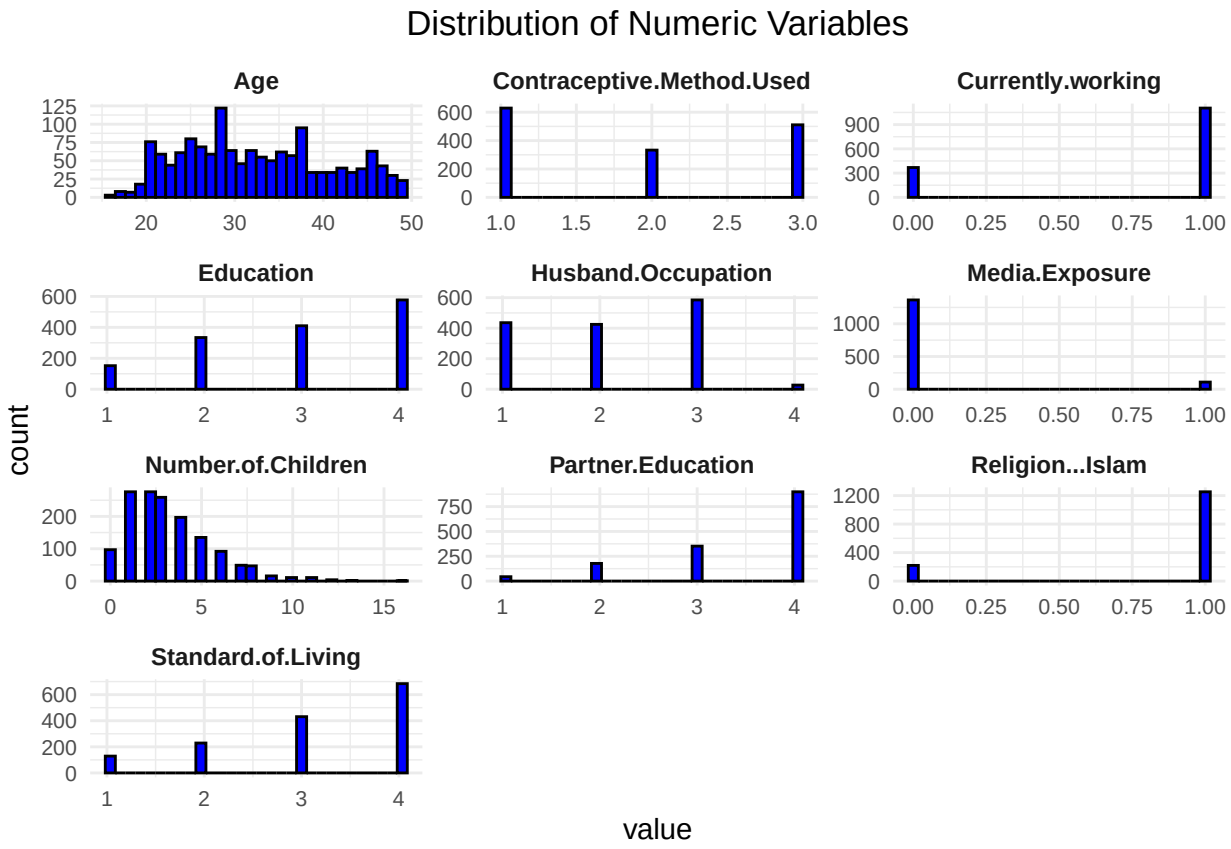
```
##                         Age                   Education         Partner.Education
##                           0                           0                         0
##         Number.of.Children            Religion...Islam         Currently.working
##                           0                           0                         0
##         Husband.Occupation          Standard.of.Living            Media.Exposure
##                           0                           0                         0
## Contraceptive.Method.Used
##                           0
```

```
library(tidyverse)
library(patchwork)
```

```
## Warning: package 'patchwork' was built under R version 4.3.3
```

```
create_numeric_histograms <- function(df, ncol = 3) {
 df %>%
   select(where(is.numeric)) %>%
   pivot_longer(cols = everything()) %>%
   ggplot(aes(x = value)) +
   geom_histogram(bins = 30, fill = "blue", color = "black") +
   facet_wrap(~ name, scales = "free", ncol = ncol) +
   theme_minimal() +
   theme(
     strip.text = element_text(face = "bold"),
     axis.text = element_text(size = 8),
     plot.title = element_text(hjust = 0.5)
   ) +
   labs(title = "Distribution of Numeric Variables")
}

create_numeric_histograms(df)
```



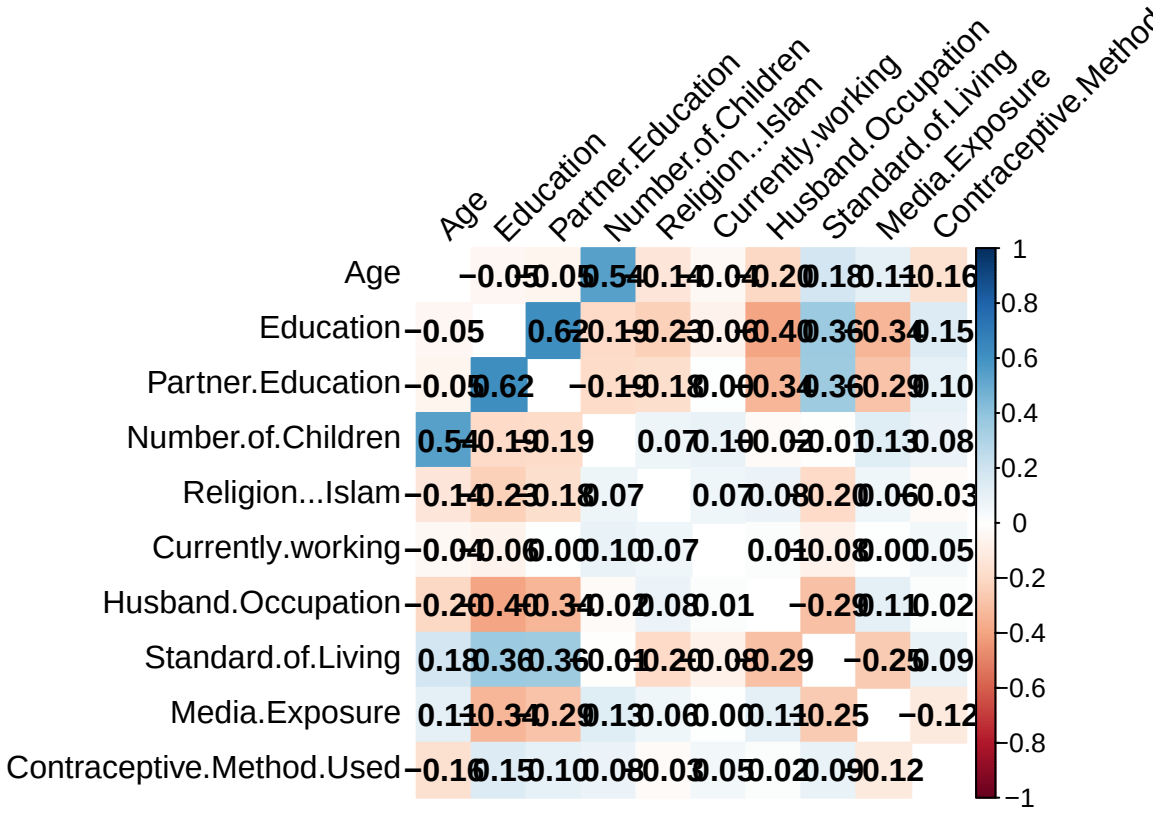

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.95 loaded
```

```
cor_matrix <- df %>% cor()

corrplot(cor_matrix,
         method = "color",
         addCoef.col = "black",
         tl.col = "black",
         tl.srt = 45,
         diag = FALSE)
```



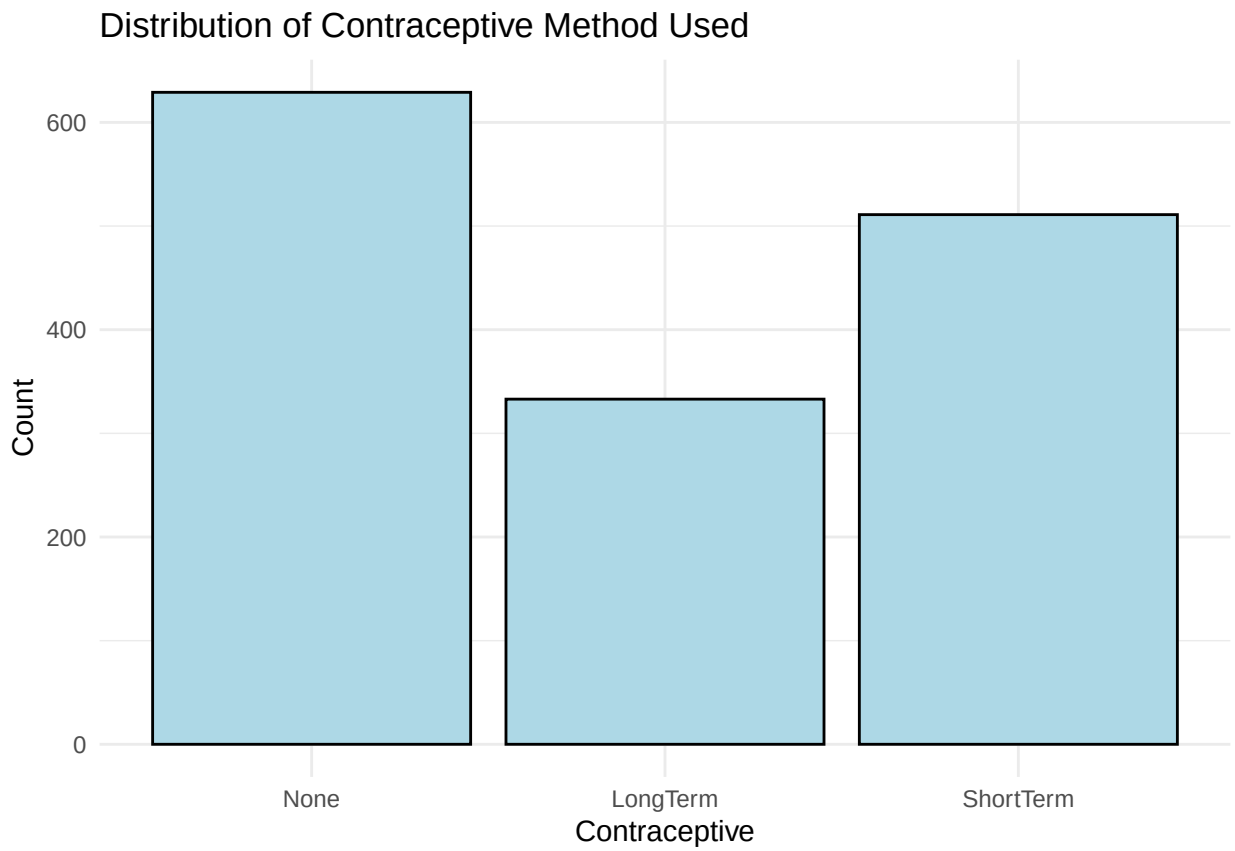
Impacts of Age on Contraceptive Method Used

Contraceptive Method Used: * 1 = No-use * 2 = Long-term * 3 = Short-term

```
df$contraceptive.method <- as.factor(df$Contraceptive.Method.Used)
#levels(df$contraceptive.method)
levels(df$contraceptive.method) <- c("None", "LongTerm", "ShortTerm")

ggplot(df, aes(x = contraceptive.method)) +
 geom_histogram(stat = "count", fill = "lightblue", color = "black") +
 theme_minimal() +
 labs(
   title = "Distribution of Contraceptive Method Used",
   x = "Contraceptive",
   y = "Count"
 )
```

```
## Warning in geom_histogram(stat = "count", fill = "lightblue", color = "black"):
## Ignoring unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad‘
```




```
dplyr::count(df, contraceptive.method, sort = TRUE)
```

```
##   contraceptive.method   n
## 1                 None 629
## 2            ShortTerm 511
## 3             LongTerm 333
```

There are 3 unequal classes of contraceptive use. Most women (629 participants) in this sample do not use contraceptives, followed by short term contraceptive use (511 participants), and finally, about twenty-two percent of these women use long term contraceptives (333 participants).

Relative Odds of Contraception Method

```
library(nnet)
model1 = multinom(df$Contraceptive.Method.Used~1)
```

```
## # weights:  6 (2 variable)
## initial  value 1618.255901
## final  value 1571.363231
## converged
```

```
summary(model1)
```

```
## Call:
## multinom(formula = df$Contraceptive.Method.Used ~ 1)
##
## Coefficients:
##   (Intercept)
## 2  -0.6359864
## 3  -0.2077754
##
## Std. Errors:
##   (Intercept)
## 2  0.06777021
## 3  0.05955488
##
## Residual Deviance: 3142.726
## AIC: 3146.726
```

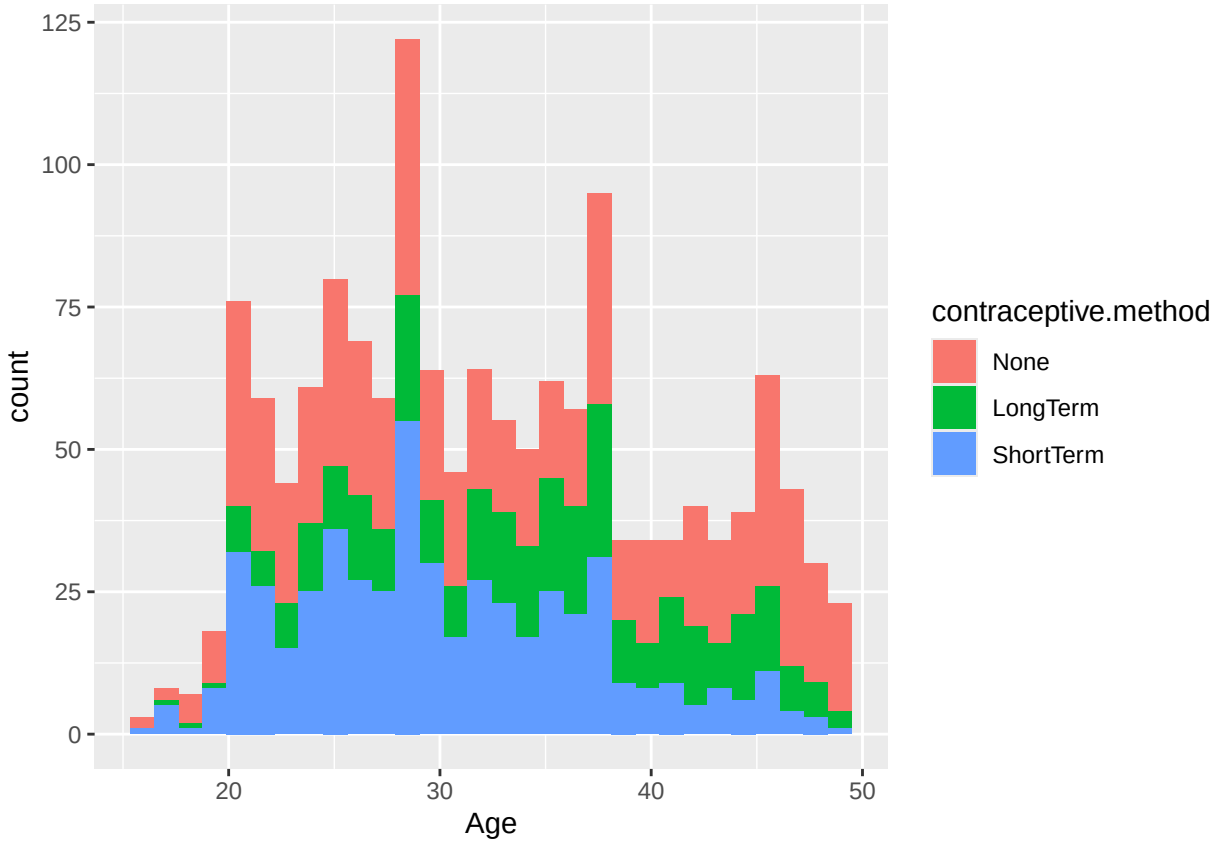
Based on a null "mlogit" model, the relative odds of certain form of Contraceptive Use relative to neither are: * P(*Long-Term Contraception*)/P(No Use)=exp(-0.6359864)= *0.529* * P(*Short-Term Contraception*)/P(No Use)=exp(-0.2077754)= *0.812*

Predictive Strength of Age

```
df$contraceptive.method <- as.factor(df$Contraceptive.Method.Used)
#levels(df$contraceptive.method)
levels(df$contraceptive.method) <- c("None", "LongTerm", "ShortTerm")
#df = df %>%group_by(contraceptive.method) %>%mutate(Frequency = n())

ggplot(df, aes(Age, fill = contraceptive.method)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
model2=multinom(contraceptive.method ~ Age, data = df)
```

```
## # weights:  9 (4 variable)
## initial  value 1618.255901
## final  value 1538.602530
## converged
```

For Readability

```
beta.mat<-coef(model2)
beta.mat
```

```
##           (Intercept)         Age
## LongTerm    -1.114493  0.01411370
## ShortTerm    1.374399 -0.04975711
```

```
se.mat <-summary(model2)$standard.errors
se.mat
```

```
##           (Intercept)         Age
## LongTerm    0.2881547 0.008224668
## ShortTerm   0.2491427 0.007654223
```

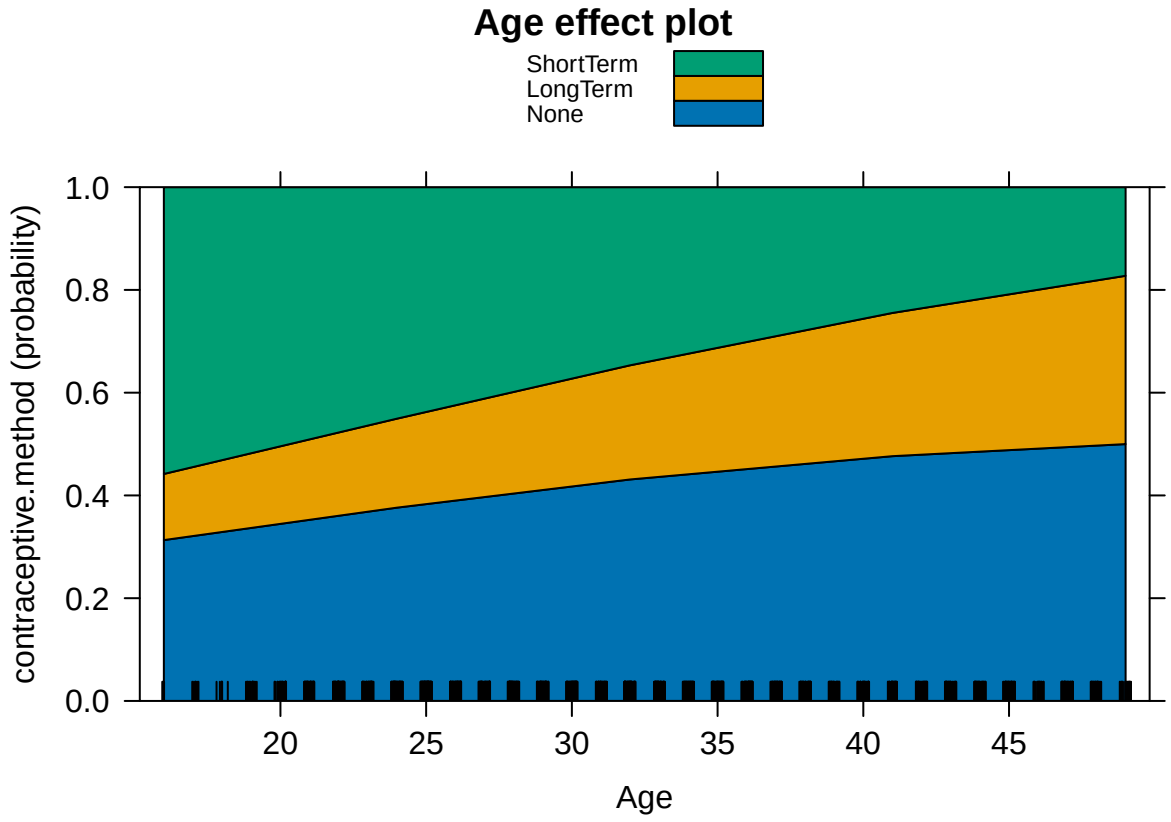
```
library(effects)
```

```
## Loading required package: carData
```

```
## Warning in check_dep_version(): ABI version mismatch:
## lme4 was built with Matrix ABI version 1
## Current Matrix ABI version is 0
## Please re-install lme4 from source or restore original 'Matrix' package
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
plot(Effect("Age",model2),style="stacked")
```




```
# shows a stacked vertical bar chart of pred
```

Interestingly, it is predicted that using long-term contraception is much higher for younger participants. As the age in the sample increases, predicted long-term contraceptive use declines. The other contraceptive uses rise slightly with age.