Abortion MRP slides

Monica Alexander

06/04/2021

Motivating example

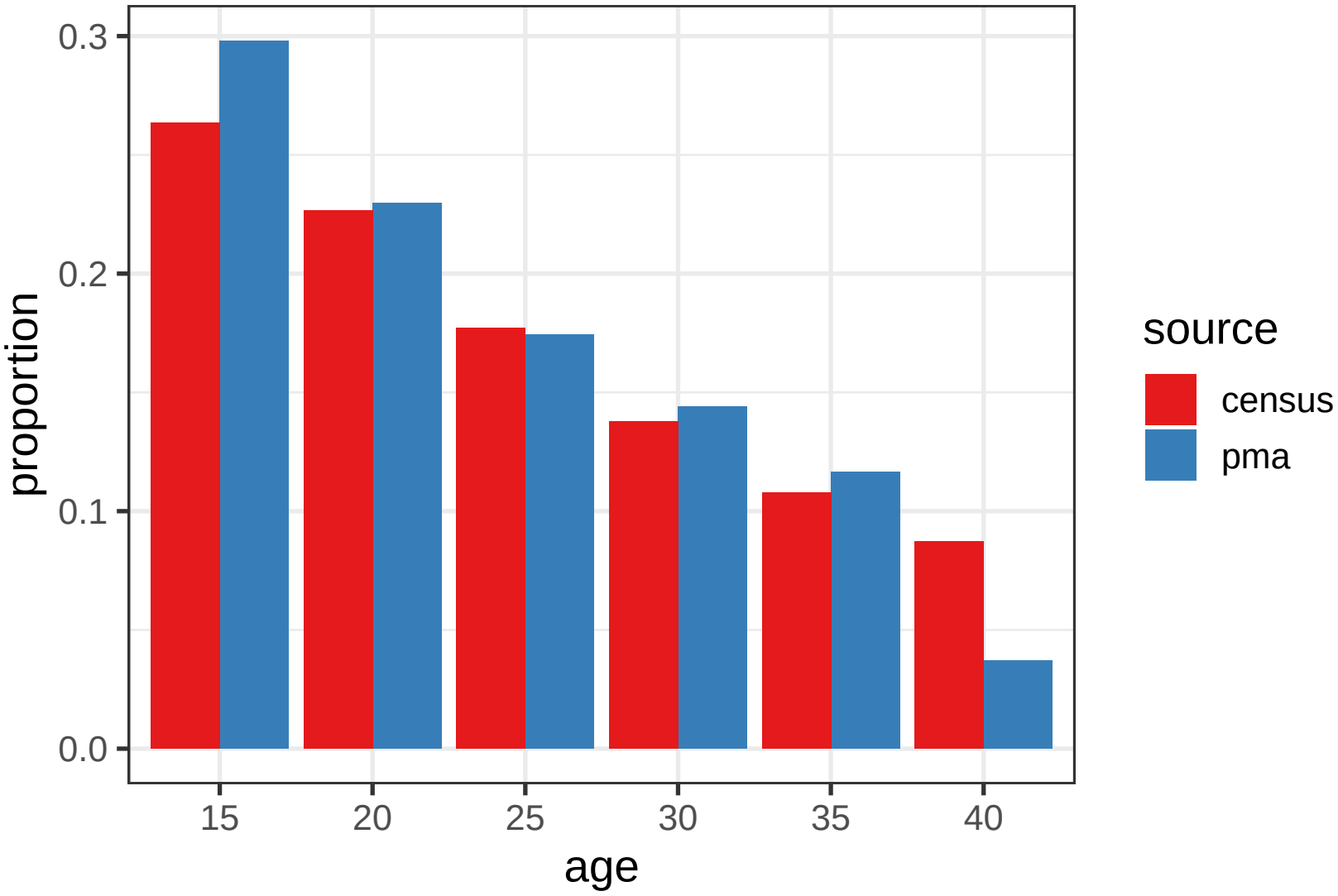
Abortion outcomes in Uganda

- Data from 2018 PMA survey (via IPUMS)
- Interested in factors associated with women ever having an abortion
- Outcome of interest: 'ever had abortion (yes/no)'
- Notes: dropping don't knows, including 'unsuccessful abortions' in 'yes'.
- More notes: Self-reported abortion is very likely to be under-reported

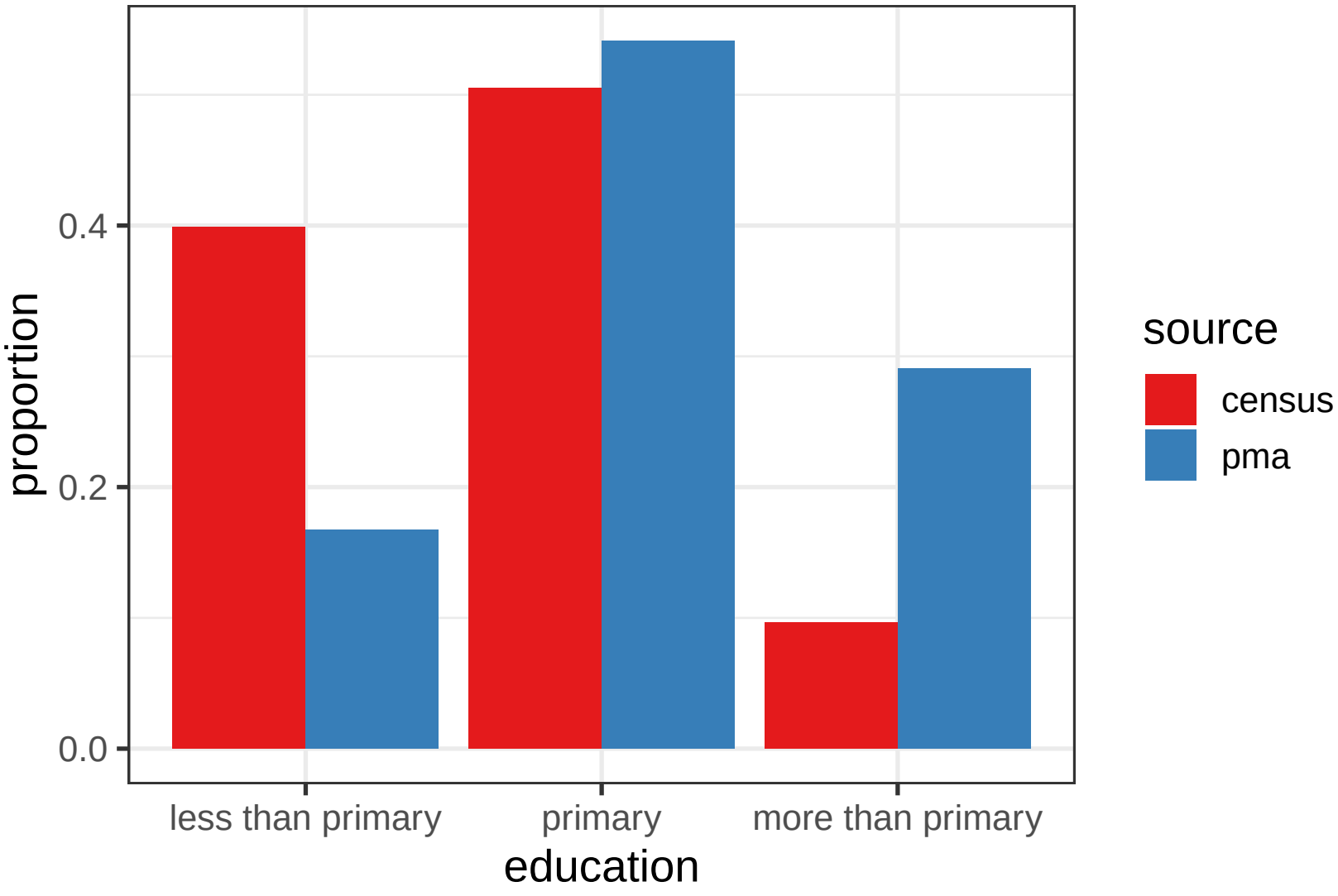
Sample versus census

How representative is the survey? Let's compare to 2014 census.

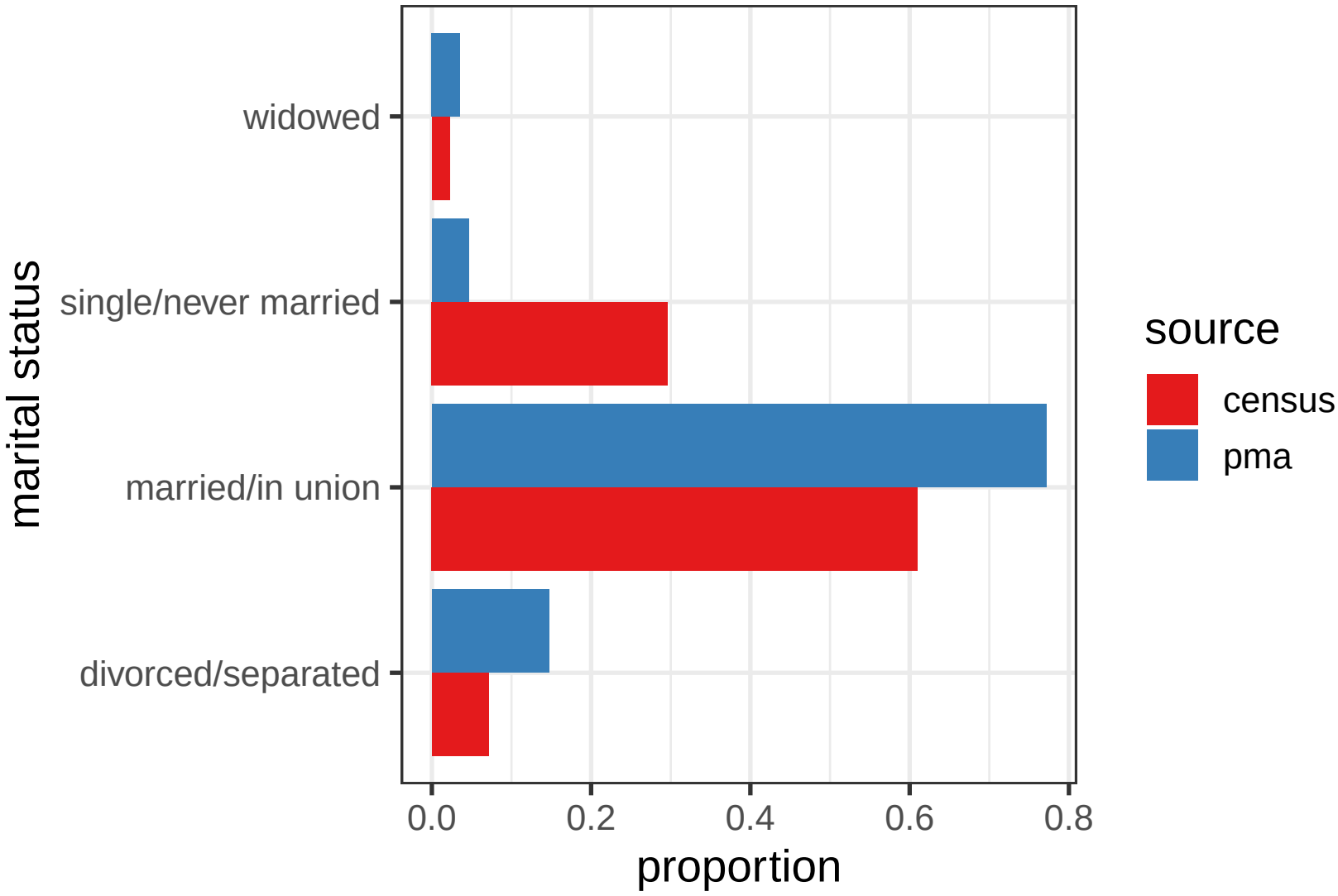
Proportion by age



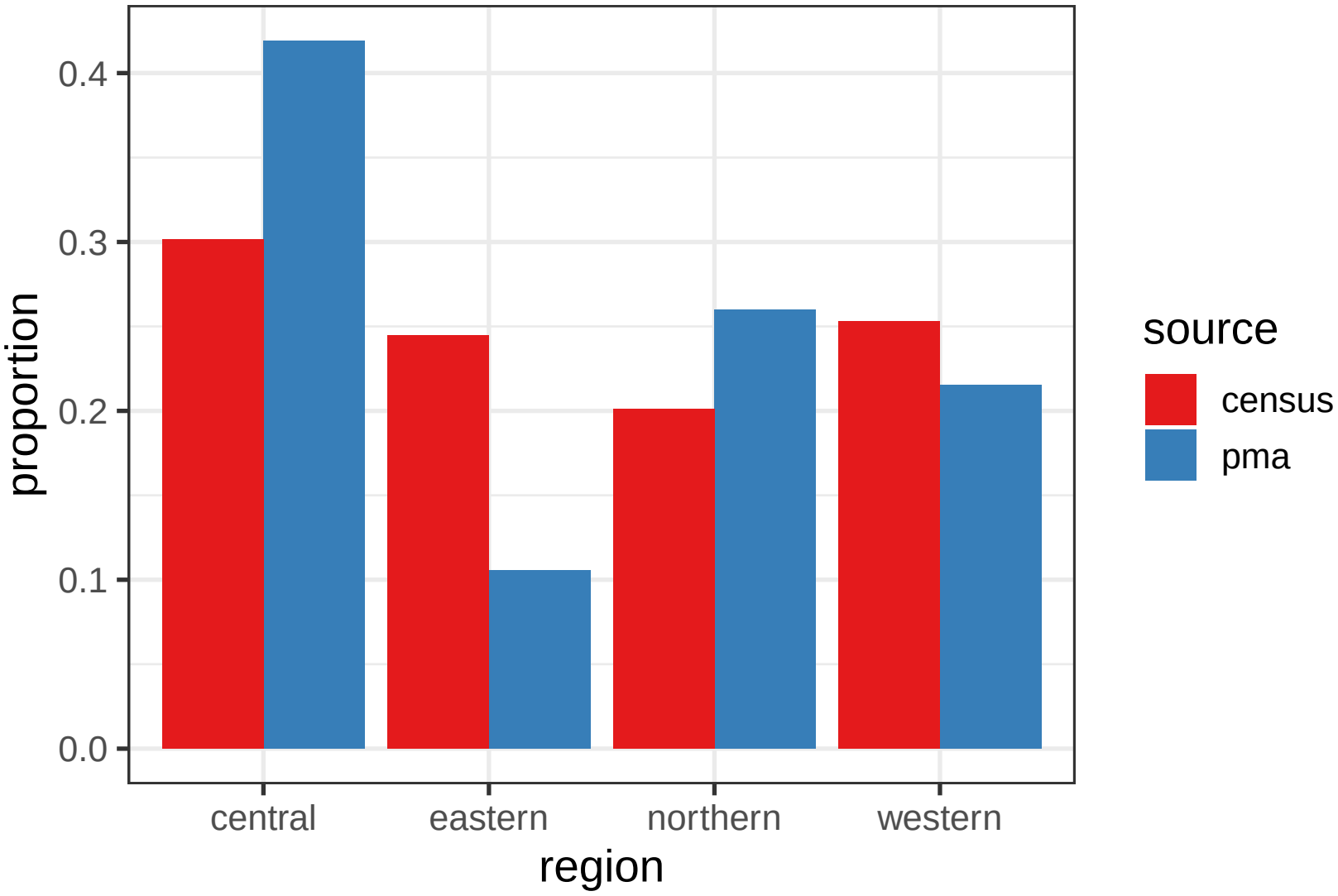
Proportion by education



Proportion by marital status



Proportion by region



Multilevel model

$$Y_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit} p_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha_{j[i]} + \gamma_{k[i]}$$

$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

$$\gamma_k \sim N(0, \sigma_\gamma^2)$$

where

- Y_i refers to whether or not individual i has had an abortion
- The index j refers to region
- The index k refers to age group
- X_1 refers to marital status
- X_2 refers to education

Predicted probabilities

- Our model gives us a way of calculating the predicted probability of reporting abortion for a women in each age/educ/region/marital status group, call it $\hat{p}_g$ for groups $1, \ldots G$ where $G = 6 \times 3 \times 4 \times 4$
- We can use these predicted probabilities to simulate individual outcomes based on the Bernoulli likelihood
- This gives us an estimate of the number of women in each group reporting abortion, with uncertainty

Post stratification

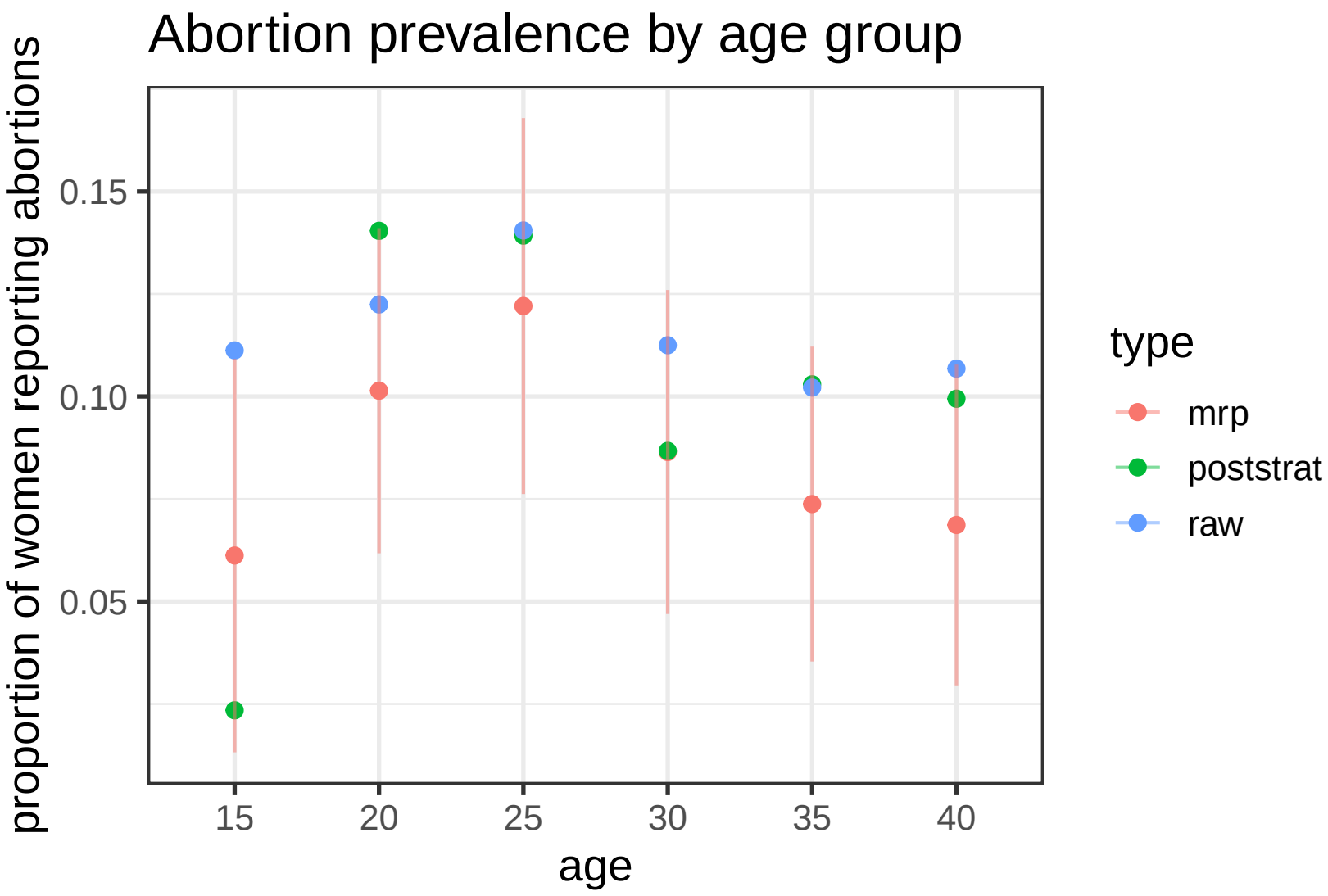
- Once we have an estimate of the probability of reporting abortion in each age/educ/region/marital status group, we can multiply these by the total number of women in each group based on the census, to get an estimate of the number of women reporting abortion in the population
- Can then use these estimates to calculate abortion incidence nationally and by different groupings
- This is called post-stratification
- E.g. national estimate of abortion incidence would be

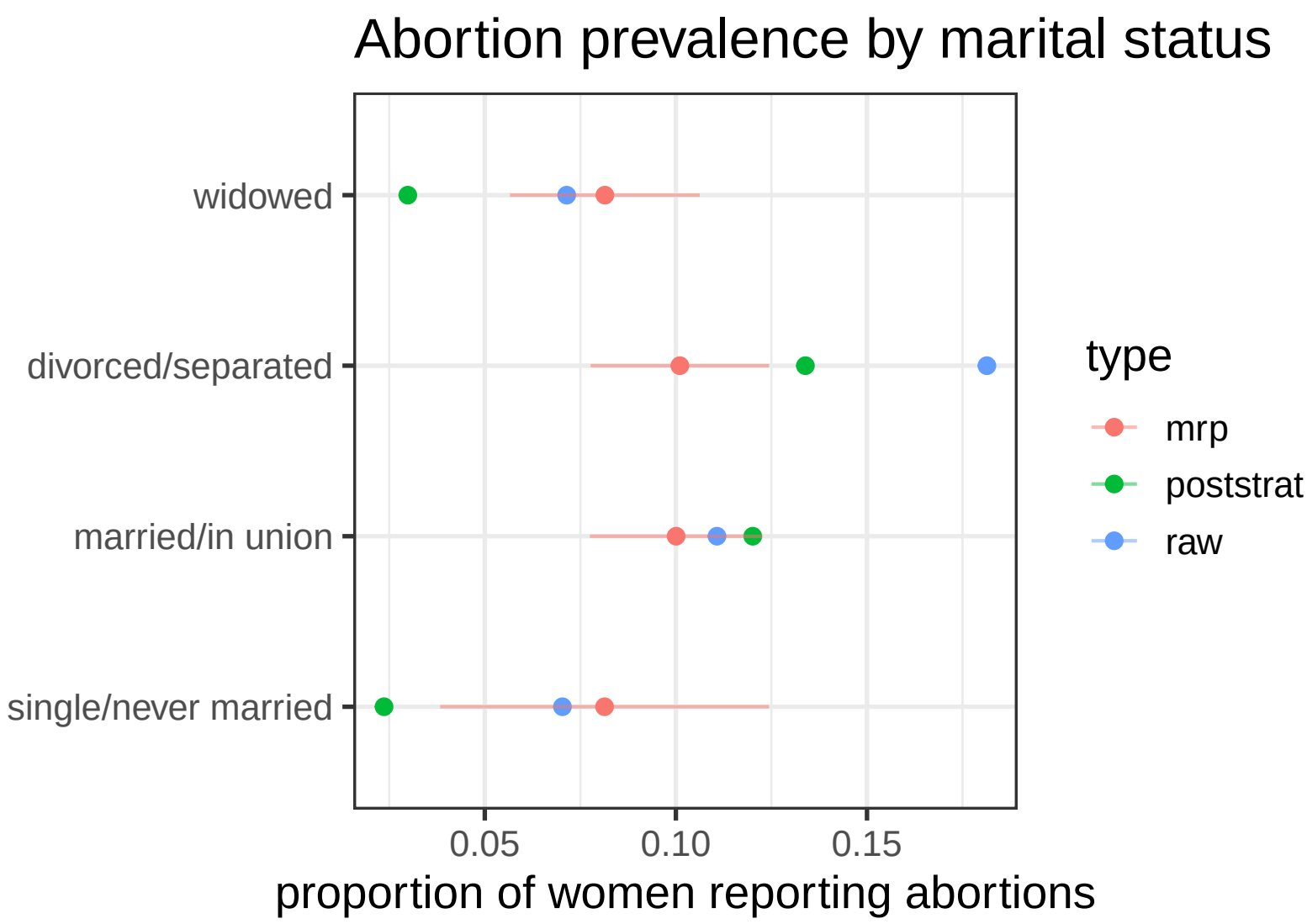
$$\frac{\sum_g \hat{p}_g \cdot N_g}{\sum_g N_g}$$

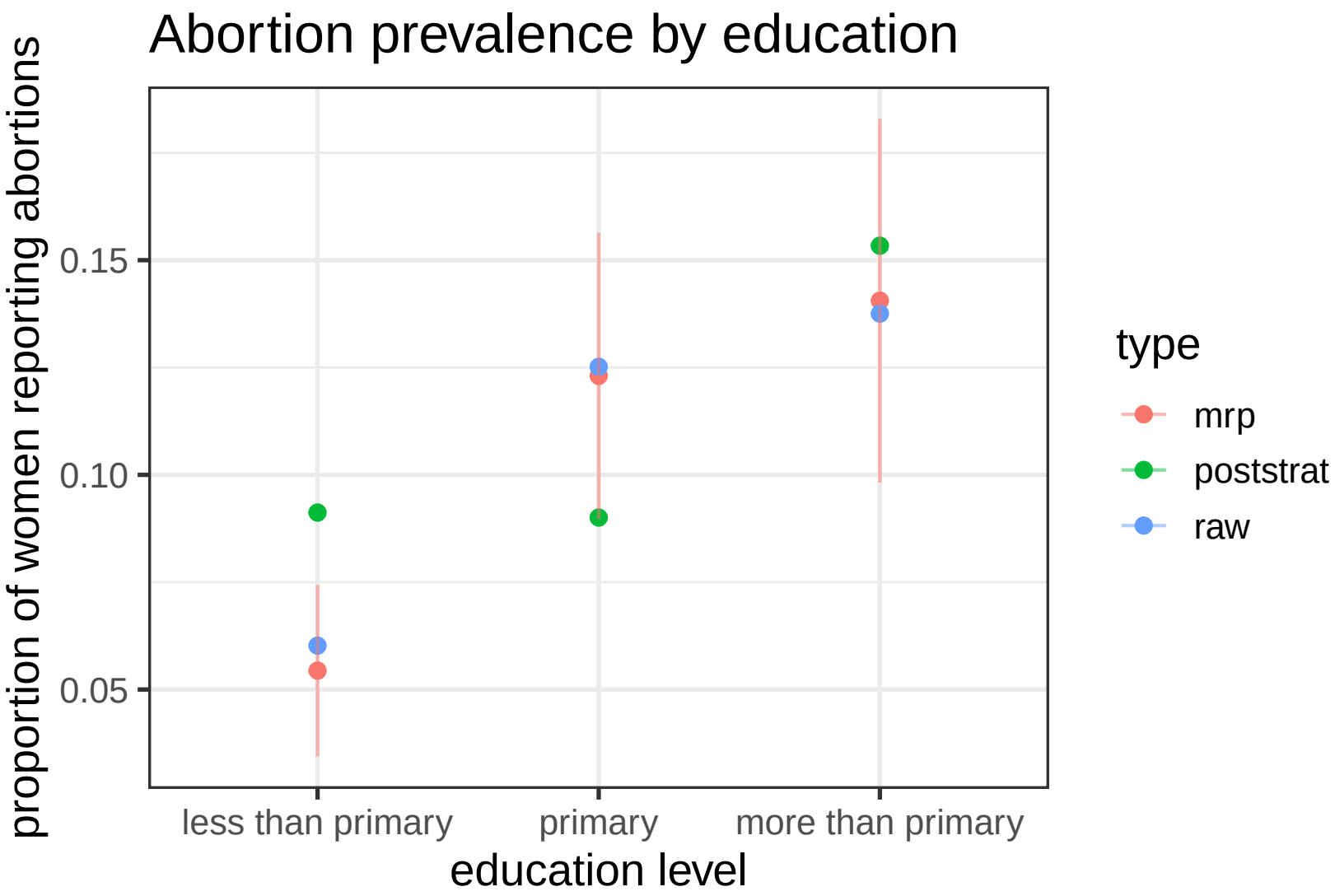
MRP versus other options

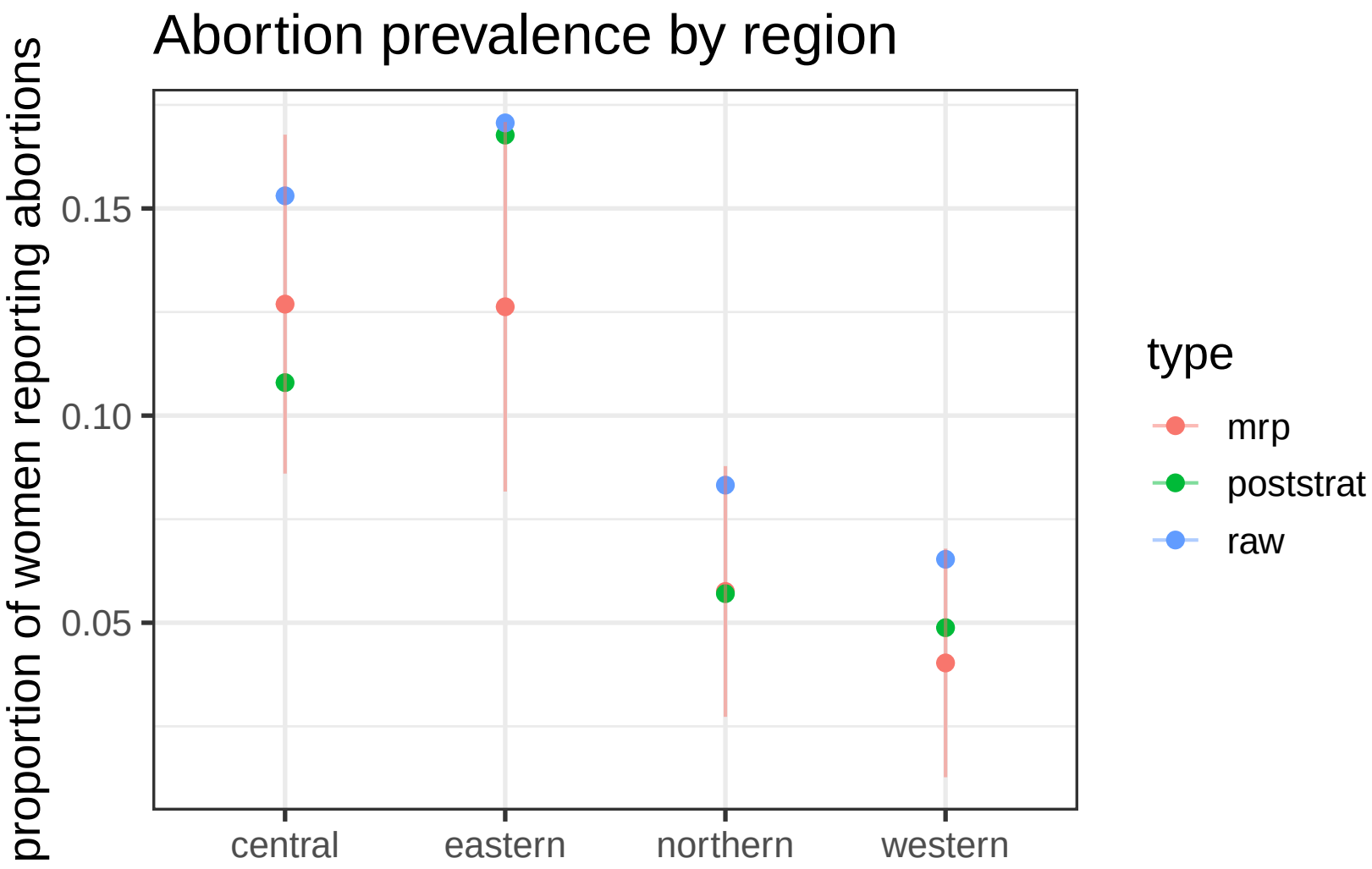
- This is called multilevel regression and post-stratification because we do just that – run a ML regression then post-stratify based on a representative data source
- Other options:
  1. just take raw incidence proportions from survey
  2. calculate raw incidence proportions from survey then post-stratify these
- Compared to 1, 2 and MRP are more representative
- Compared to 2, MRP downweights raw proportion estimates from groups that have small sample sizes (but potentially at the expense of increased variance in estimates)

Results


Abortion prevalence by age group
proportion of women reporting abortions
0.15
0.10
0.05
15
20
25
30
35
40
age
type
mrp
poststrat
raw


Abortion prevalence by marital status
widowed
divorced/separated
married/in union
single/never married
0.05
0.10
0.15
proportion of women reporting abortions
type
mrp
poststrat
raw


Abortion prevalence by education
proportion of women reporting abortions
0.15
0.10
0.05
less than primary
primary
more than primary
education level
type
mrp
poststrat
raw


Abortion prevalence by region
proportion of women reporting abortions
0.15
0.10
0.05
central
eastern
northern
western
type
mrp
poststrat
raw