MRP as a tool in the population sciences: potential benefits and challenges

Marija Pejcinovska[1]
Monica Alexander[1]
Jessie Yeung[1]
Alison Gemmill[2]

[1] University of Toronto
[2] Johns Hopkins University

1 Introduction

The foundation of estimation and inference in sociology and demography is high-quality, representative data sources. 'Traditional' data sources in the population sciences are centered on censuses and large-scale surveys, which are designed to capture either the whole population or a decent representation of it. Additionally, population scientists also often rely on data from civil registration and vital systems to study levels and changes in key demographic outcomes such as fertility and mortality. In high-quality data contexts, counts of the event of interest and the population at risk of exposure are readily available and reliable, such that rates of interest for various population subgroups can easily be calculated and interpreted.

However, the nature in which data are collected and the types of data that are available in the population sciences has changed vastly over the past several decades. It is becoming increasingly difficult and expensive to gather representative survey data, due in part to the decline in usability of landline telephone registers as population sampling frames and general resistance to answering surveys, leading to poorer response rates (Stern, Bilgen, and Dillman 2014). For example, response rates for the General Social Survey in Canada declined from over 83% in the 1980s to just over 52% in 2017 (StatCan 2019). The collection of high-quality census data worldwide faces challenges related to budget cuts, privacy concerns, and delays and lower response rates due to the COVID-19 pandemic. For example, the 2020 Brazilian census was delayed until 2022, and the budget to run the census was cut by 88% (SBPC 2021). In the United States, in response to concerns about respondent confidentiality, the Census Bureau plans to only release 'differentially private' data from the 2020 Census (with intentional errors added to nearly all statistics) (IPUMS 2022). This has implications for how microlevel data from the census can be used to create accurate and meaningful descriptive summaries by subpopulations of interest.

While there is a declining ability to rely on censuses and representative surveys as a primary means of obtaining estimates of interest, there has been an emergence of 'non-traditional' data sources as a resource in population sciences. The increase and widespread use of the internet, and in particular, the emergence of social media websites, has driven a change in how population scientists view and use data. Salganik (2019) illustrates how data from social media, smartphones, and other digital sources can be a powerful tool in combination with more traditional data sources. In the population sciences, social media and other digital data has been used to estimate migration (Alexander, Polimis, and Zagheni 2019) and fertility (Rampazzo et al. 2018), to understand differential preferences by country of origin (Stewart et al. 2019), and to estimate gender inequalities (Kashyap et al. 2022). The network structure of social media data, and Twitter data in particular, has proven to be a useful new resource to study non-familial social ties and the spread of information (Bail 2021). Beyond using digital data in and of itself, digital mediums can be used to collect social data in a more traditional sense, through the use of online surveys. In particular, social media websites can be used as a vehicle to contact targeted populations to ask about behaviors and experiences (Kashyap et al. 2022). For example, Facebook allows any user to post advertisements for a business or a survey. These advertisements can be targeted to reach particular population subgroups of interest. This technique was used to study behaviors and attitudes in response to the COVID-19 pandemic across multiple countries (Perrotta et al. 2021).

While there are many advantages of supplementing more traditional forms of data with these new digital sources, there are notable drawbacks to using these data to make inferences about the broader population. Perhaps the most persistent is one of representativeness — the sample on which data are collected is very unlikely to be a true representation of the underling population of interest. The population of users of a social media website are likely to have different characteristics; for example Facebook users have traditionally been younger on average (Alexander, Polimis, and Zagheni 2020). Even beyond representation of the user base in general, there is likely to be differences in who answers surveys through websites, or what types of people keep their information updated. Issues of a bias user base and a bias set of responses compound, and can potentially lead to extremely small counts of observations by key subpopulations of interest. As such, we need to consider statistical and other estimations to adjust non-representative data in order to make more valid population inferences.

Multilevel regression and post-stratification (MRP) is a promising estimation method to make better use of new data sources in the population sciences, but is as yet under-utilized in applications specific to sociology and demography. MRP allows the relationship between outcomes of interest and key demographic and socioeconomic characteristics, adjusting for the fact that we may have very few observations in some subgroups of interest. Importantly, MRP provides a direct link between non-traditional forms of data and traditional large-scale censuses and surveys, combining and utilizing the strengths of each source. In this chapter we illustrate the potential utility of

MRP in population science applications, focusing on two examples. The first example is estimated fertility intentions based on a simulated dataset where we know the "truth"; the second example is an application to estimating the prevalence of marital name changes in the US using two non-representative surveys. In both applications, we demonstrate the strengths of MRP in overcoming some traditional issues with survey data, while also highlighting the limitations of the technique in correcting for other known issues, such as omitted variable bias.

The remainder of the chapter is structured as follows. We first review existing uses of MRP in the population sciences. We then demonstrate the potential of MRP using simulated fertility intentions data, and then on an application to marital name changes. We end with a discussion of the strengths and limitations of MRP in the population sciences, and suggest areas for future research.

2 MRP in the population sciences

The use of multilevel regression and post-stratification (MRP) as a technique to adjust non-representative surveys and produce small area estimates gained popularity in areas of political science, particularly in forecasting elections based on measuring voter intention, and estimating public opinion (Gelman, Little, and Witter, n.d.; Warshaw and Rodden 2012). For instance, an early application by Park et al. (2004) applied MRP to pre-election polls in the US and compared these estimates to state-level election results. In the somewhat famous 'XBox' paper, Wang et al. (2015) created forecasts for the 2012 US election from a series of daily voter intention polls conducted on the Xbox gaming platform. Beyond academic papers, MRP is now widely used in commercial applications of political forecasting and estimating public opinion, most notably by YouGov (see, for example, (English 2021), which discusses predictions of election outcomes in the UK).

While MRP has gained traction and become relatively widespread in its use in political science, uptake in the other social sciences, and in particular, population sciences, has been relatively slow. The majority of work in other areas has been done in epidemiology and population health. In these contexts, there is demand for population-level estimates of health and well-being outcomes at the small scale, such as country or census-tract levels, in order to best inform public and health policy. However, data on these outcomes is usually only available from surveys that, while may be designed to be representative at the national level, often contain only a few (or no) observations at the small-area level. As such, MRP is useful to obtain reasonable estimates with a reduced level of uncertainty. For instance, Downes et al. (2018) used a large national health survey of Australian males to examine the benefit of MRP for addressing participation bias in the estimation of population descriptive quantities compared to more conventional techniques reliant on sampling weights. Eke et al. (2016) use MRP to generate estimates of the prevalence of adult periodontitis at the state and local levels in the United States using the National Health and Nutrition Examination Survey. More recently, Breen et al. (2021) investigated the use of MRP in estimating age-specific contact patterns at the subnational level, which is valuable information to better understand the spread of COVID-19 and other directly transmitted airborne pathogens.

In the United States more broadly, a number of studies have applied MRP to data from the Behavioral Risk Factor Surveillance System (BRFSS), a state-based random digit dial telephone survey that is conducted annually, and collects state data about US residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services (BFRSS, 2022). For example, Zhang et al. (2014) apply MRP to produce census-block-level estimates of chronic obstructive pulmonary disease prevalence, demonstrating that the MRP framework can produce reliable small area health estimates. Work from the same team further validated MRP as applied to BRFSS data to produce small area estimates on a range of different health outcomes in Missouri (Xingyou Zhang et al. 2015). Validation work has also been carried out for Massachusetts (Y. Wang et al. 2018) and Connecticut (Zheng et al. 2014). Additionally, the 'County Health Rankings & Roadmaps' program of the University of Wisconsin Population Health Institute uses BRFSS and MRP to produce estimates of health outcomes such as the average number of physically unhealthy days reported in past 30 days (University of Wisconsin Population Health Institute 2022).

Beyond health outcomes, work has also been done utilizing MRP to estimate traditionally hard-to-measure populations and beliefs. For example, using the BRFSS in combination with MRP and data from the American Community Survey (ACS), Flores et al. estimated the total adult population who identify as transgender in the US, as well as various characteristics of the transgender population (Flores, Brown, and Herman 2016; Herman et al. 2017). MRP has been also used to more accurately measure religious groups' sizes and demographics in the UK (Claassen and Traunmüller 2020), and a dynamic version has been employed to estimate US state-level religious trends (Wiertz and Lim 2021). Building on the success of previous efforts in population sciences, we believe there is substantial potential for MRP to improve estimate and understanding of social phenomena. The remainder of this paper illustrates the application of MRP to two important social indicators: fertility intentions, and marital name changes.

3 Estimating fertility intentions of women in the United States

The number of children a person or family expects or intends to have in future is an interesting social indicator, and patterns over time and across subgroups are increasingly being studied in demography and sociology. Fertility intentions have been shown to be highly correlated with eventual fertility outcomes (Schoen et al. 1999), and so measuring intentions during a recession or pandemic, for example, may give insights into the effect of such an even on likely future fertility. Additionally, any discrepancies in intentions and realized fertility are important to understand in order to gain insight into why fertility patterns may be changing over time (Morgan and Rackin 2010).

In this section we use MRP to estimate fertility intentions in the United States. For the purposes of illustration, we obtain estimates of the proportion of women who want more children in future from a large national survey, and treat these as the 'truth.' We then generate a small non-representative survey from this dataset, and estimate fertility intentions based on this smaller dataset using several methods, to see how closely we can recover the truth.

3.1 Data

Data on fertility intentions come from the National Survey of Family Growth (NSFG), a national survey which began in 1973 and gathers information on pregnancy and births, marriage and cohabitation, infertility, use of contraception, family life, and general and reproductive health (National Center for Health Statistics 1987). The survey is designed such that representative national-level estimates can be generated. We used the public use data file for the 2015-2017 NFSG, which contained 5,554 women respondents aged 15-49. We will focus on the proportion of women who would like more children by age group for women aged 20-45.[1] The estimates of these proportions based on the full NSFG dataset are shown in Table 1. In general, the proportion of women wanting more children declines with age.

Table 1: Proportion of women who want more children by age group

Age group	Proportion
20-24	0.833
25-29	0.714
30-34	0.471
35-39	0.256
40-44	0.059

Data on population counts by sex, age, education, and marital status in the United States come from the 2016 5-year American Community Survey (ACS). Microdata for the ACS was accessed through IPUMS-USA (Ruggles et al. 2015).

For the purposes of illustration, we drew a smaller sample of the NSFG consisting of just 300 women. Figure 1 compares the distribution of women by age, education, and marital status in the smaller survey to the broader US population, as captured in the ACS. As illustrated, the smaller sample vastly over-represents married women aged 35-39 with a Bachelor degree, while many other cells are empty. For example, there are no women aged 20-24 with a high school degree who have never been married, even though this group is almost 5% of the US population.

[1]This is just one measure of fertility intention and probably the most simplistic; other measures focus on the number of children desired and how that interplays with a person's current parity.

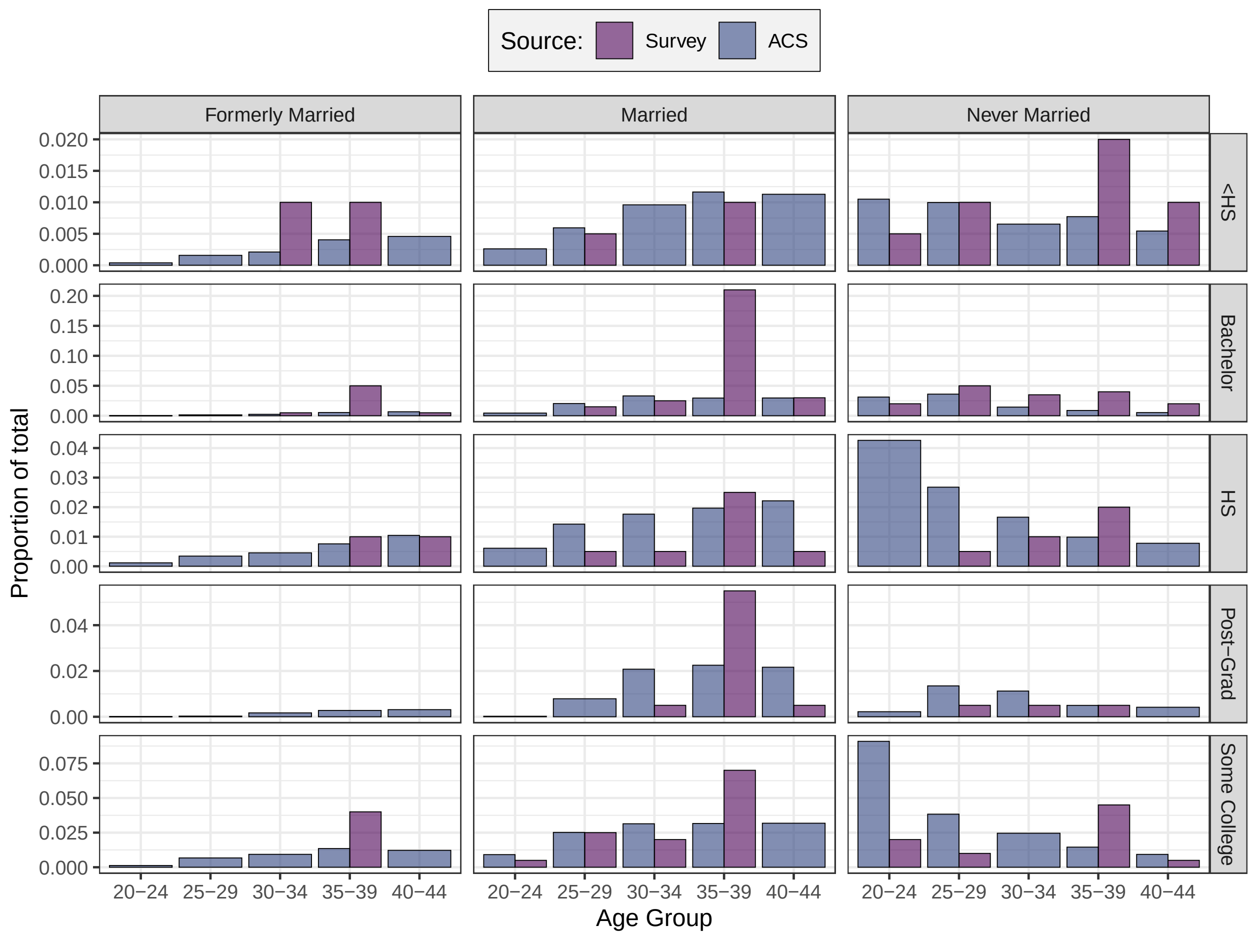

Figure 1: Comparing the distribution of the survey to US population

3.2 Model

For the multilevel regression part, we fit the following model:

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) \\
\pi_i &= \text{logit}^{-1}\left(\beta_0 + \beta_1 \text{formerly married}_i + \beta_2 \text{married}_i + \alpha^{\text{age}}_{j[i]} + \alpha^{\text{edu}}_{k[i]}\right) \\
\alpha^{\text{age}}_j|\alpha^{\text{age}}_{j-1}, \sigma_{\text{age}} &\sim \text{N}\left(\alpha^{\text{age}}_{j-1}, \sigma^2_{\text{age}}\right), \text{ for } j = 2, \dots, 5 \\
\alpha^{\text{edu}}_k|\sigma_{\text{edu}} &\sim \text{N}\left(0, \sigma^2_{\text{edu}}\right), \text{ for } k = 1, \dots, 5
\end{aligned}
$$

where $y_i = 1$ if respondent i wants more children and 0 otherwise, and the formerly married_i and married_i variables are indicator variables. Standard normal priors were placed on $\beta_0, \beta_1, \beta_2, \alpha^{\text{age}}_1$ and standard half-normals were placed on the variance parameters.

It is worth highlighting the assumption about age patterns in the model in particular. Rather than a standard unstructured model on age, we are taking advantage of the fact that we know fertility behaviors and intentions to be strongly correlated across age. By placing a random walk structure over age, we are assuming that a particular age group a is most influenced by adjacent age groups.

3.3 Results

Here we illustrate the results of the MRP estimation of fertility intentions compared to the truth and also standard estimates based on post-stratification. Post-stratification is a relatively standard practice in sociology and population sciences to reweigh survey data based on population proportions to obtain more representative estimates. The post-stratified estimates are

$$
\hat{p}^{\text{ps}}_a = \frac{\sum_{g=1}^{G} \hat{p}^{\text{raw}}_{g[a]} \times N_{g[a]}}{\sum_{g=1}^{G} N_{g[a]}}
$$

where g refers to a particular education/marital status group (e.g. people who are married and have less than a high school degree). There are a total of $G = 5 \times 3 = 15$ groups within each age group. Note that $\hat{p}^{\text{raw}}_{g[a]}$ refers to the observed proportion of women in group g who are aged a who want more children, and $N_{g[a]}$ refers to the size of that particular population group who are aged a in the US population.

In comparison, the multilevel-regression-with-post-stratification (MRP) estimates are

$$
\hat{p}^{\text{MRP}}_a = \frac{\sum_{g=1}^{G} \hat{p}^{\text{MR}}_{g[a]} \times N_{g[a]}}{\sum_{g=1}^{G} N_{g[a]}}
$$

where $\hat{p}^{\text{MR}}_{g[a]}$ is the proportion of women in group g who are aged a who want more children estimated the model presented in the previous section. Compared to traditional post-stratification, MRP replaces the raw survey proportions with estimated proportions based on the multilevel model described above.

Table 2 compares the estimated proportion of women wanting more children for the MRP and post-stratification (PS) methods, as well as the 'true' proportions from the NSFG survey, by age group. We can see that across all ages, the MRP estimates are closer to the truth compared to the PS estimates. The MRP gain is particularly evident in the first two age groups, where cell sizes are relatively small. Additionally, smaller survey had no women aged 40-44 that indicated they want more children, and so the PS estimate is zero; in contrast, the MRP estimate is 'pulled up' by the pattern in the previous age group, thanks to the random walk prior over age.

Table 2: Comparison of MRP and post-stratification (PS) results by age group

Age Group	True	MRP	PS	MRP difference	PS difference
20-24	0.83	0.82	0.46	-0.01	-0.37
25-29	0.71	0.76	0.50	0.05	-0.21
30-34	0.47	0.46	0.52	-0.01	0.05
35-39	0.26	0.25	0.23	-0.01	-0.03
40-44	0.06	0.10	0.00	0.04	-0.06

4 Estimating prevalence of marital name changes based on two surveys

In this section we discuss an example of applying MRP to the estimation of the proportion of US women keeping their last names upon marriage. We use data from two surveys that have very different characteristics to illustrate the strengths and weaknesses of this technique.

4.1 Background

The extent to which women choose to either retain or change their last name in the advent of marriage is an interesting indicator of broad social and cultural change over time. While there is a general belief that the proportion of women choosing to keep their last name has broadly increased over time, obtaining reliable estimates of the population-level prevalence of marital name changes is somewhat challenging from available data. Questions about name changes are generally not asked in large-scale representative surveys; and administrative data is usually not publicly accessible. Previous studies on marital name changes do exist, but various substantially in the surveyed population and types of questions asked. As such the estimated proportions of women who have kept their last name varies substantially across surveys. However, keeping one's last name does appear to be correlated positively with higher educational attainment, higher age at first marriage, and has been more common among career-oriented women and those in higher paying jobs (Scheuble and Johnson 1993; Twenge 1997; Hoffnung 2006). Keeping one's last name has also been associated with religious affiliation, with women marrying in Catholic ceremonies found to be least likely to retain their names (Abel and Kruger 2011). Despite the fact that the fraction of women with higher education and those in higher paying jobs has steadily been increasing in the last few decades, the practice of retaining one's last name after marriage has not exhibited a similar pattern. Historically, there seems to have been a notable increase in the rates of women who kept their last names in the 1970s and early 1980s (Seheuble, Klingemann, and Johnson 2000). Up until the mid-1970s in many states in the US, women still had to fight legal battles to be allowed to vote, obtain passports, or open bank account without assuming their husbands' names upon marriage (Gorence 1976). Increases in the prevalence of women keeping their last name with marriage relates to changes in reversal of these laws, and women's desire to assert their identity and independence and seen it as an indicator of their feminist values. Previous research suggests that rates plateaued around the late 1980s to mid-1990s and started declining some time in the mid to late 1990s, even among women with higher education (Goldin and Shim 2004). In recent years, however, there seems to have been an uptick in the rates of women who have opted to keep their last names upon marriage, especially among those in the younger age groups (The Upshot 2015). It is worth noting that a number of these findings have been based on data that may not have been an entirely good representation of the population of American married women — for instance, a few of these studies have relied on wedding announcements from the New York Times (see, for instance, Hoffnung (2006), Schuster (1997), Hoffnung (2006), Kopelman et al. (2009)). This nonrepresentativeness of the survey data will play an important role in our example.

4.2 Data

Data for this example come from two separate online surveys. The first is a short Google Consumer Survey conducted in 2015 by the staff of the Upshot — a New York Times affiliated data journalism site (Google Surveys 2015). The survey targeted married women in the US over the age of 18 and collected information on the decade of their first marriage and their choice of marital name at the time of marriage. Age and geographic location are assumed to have been inferred based on respondents browsing history. 1,187 women completed the survey. We will refer to these data as the NYT survey.

The second survey was conducted by Philip Cohen at the University of Maryland (Cohen 2019). The survey was shared online (primarily through social media) in 2019 and targeted married women who are US residents. In contrast to the NYT survey, Cohen asked respondents about marital name choices relating to their most recent marriage. 4,520 respondents completed the survey and available information ranges from basic demographics, such as age and education, to questions regarding the role social norms and familial contexts play in women's decision to change their last name upon marriage. We refer to these data as the UMD survey.

For the post-stratification dataset, population counts by gender, age, education, marital status, and decade of marriage in the Unites States were obtained from the 2017 American Community

Survey (ACS) 5-year estimates. Data were accessed through IPUMS-USA (Ruggles et al. 2015).

Figure 2 illustrates the composition of both surveys in terms of age of respondents and decade married compared to the overall US population as reported in the ACS. In general, just considering these two dimensions, the NYT survey appears much more representative than the UMD survey. Both the NYT and UMD surveys women over the age of 55 are underrepresented, but most strikingly, in the UMD survey women between the ages of 35 and 44 seem to make up a sizable portion, nearly twice the proportion in the US population. This is a consequence of the this survey collecting a convenience survey through social media avenues. In terms of decade of marriage the NYT survey seems to be quite representative of the overall population. In contrast, and somewhat unsurprising given the age distribution, the vast majority of participants in the UMD survey were women that reported getting married in the last decade or two.

4.3 Model

The outcome of interest for this analysis is a binary variable y_i which is equal to 1 if the respondent kept her last name when married and 0 otherwise. For the NYT survey, we fit the following multilevel model:

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}\left(\pi_i\right) \\
\pi_i &= \text{logit}^{-1}\left(\beta_0 + \alpha^{\text{age}}_{a[i]} + \alpha^{\text{decade}}_{d[i]} + \alpha^{\text{state}}_{s[i]}\right) \\
\alpha^{\text{age}}_a|\ \sigma_{\text{age}} &\sim \text{N}\left(0, \sigma^2_{\text{age}}\right),\ \text{for } a = 1, \dots, 6 \\
\alpha^{\text{decade}}_d|\ \sigma_{\text{decade}} &\sim \text{N}\left(0, \sigma^2_{\text{decade}}\right),\ \text{for } d = 1, \dots, 5 \\
\alpha^{\text{state}}_s|\ \sigma_{\text{state}} &\sim \text{N}\left(0, \sigma^2_{\text{state}}\right),\ \text{for } s = 1, \dots, 50 \\
\sigma_{\text{age}},\ \sigma_{\text{decade}},\ \sigma_{\text{state}} &\sim \text{N}_+\left(0, 1\right) \\
\beta_0 &\sim \text{N}(0, 1)
\end{aligned}
$$

For the UMD survey, we fit a similar model, but with an additional education effect:

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}\left(\pi_i\right) \\
\pi_i &= \text{logit}^{-1}\left(\beta_0 + \beta_1 \text{edu}^{<\text{BA}}_i + \beta_2 \text{edu}^{\text{BA}}_i + \alpha^{\text{age}}_{a[i]} + \alpha^{\text{decade}}_{d[i]} + \alpha^{\text{state}}_{s[i]}\right) \\
\alpha^{\text{age}}_a|\ \sigma_{\text{age}} &\sim \text{N}\left(0, \sigma^2_{\text{age}}\right),\ \text{for } a = 1, \dots, 6 \\
\alpha^{\text{decade}}_d|\ \sigma_{\text{decade}} &\sim \text{N}\left(0, \sigma^2_{\text{decade}}\right),\ \text{for } d = 1, \dots, 5 \\
\alpha^{\text{state}}_s|\ \sigma_{\text{state}} &\sim \text{N}\left(0, \sigma^2_{\text{state}}\right),\ \text{for } s = 1, \dots, 50 \\
\sigma_{\text{age}},\ \sigma_{\text{decade}},\ \sigma_{\text{state}} &\sim \text{N}_+\left(0, 1\right) \\
\beta_0,\ \beta_1,\ \beta_2 &\sim \text{N}(0, 1)
\end{aligned}
$$

Note that a number of the post-stratification cells are empty in the population mainly because they are simply not possible, based on age and decade married. For instance, there are no individuals 18-24 year of age who have married in the 1970s. The α's in the above model denote the varying coefficients, which are assumed exchangeable with independent Normal priors.

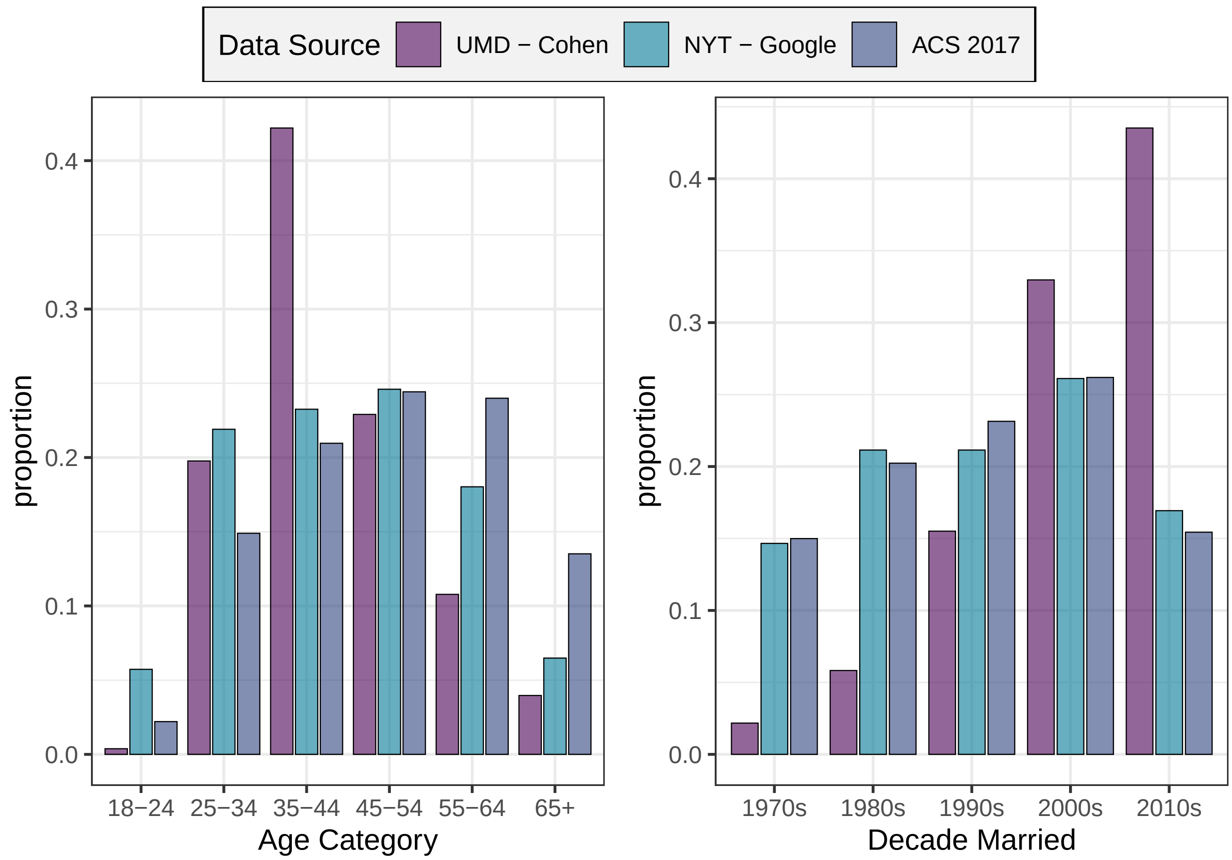

Figure 2: Proportion of sample by age and decade married, for NYT, UMD and ACS.

4.4 Results

Figure 3 shows the raw and estimated national-level prevalence of women keeping their last name on marriage across both surveys. As shown in this graph, the raw estimate from the NYT survey is much lower than the raw UMD survey (around 17% versus 46%), which is most likely a consequence of the UMD survey vastly over-representing relatively young and highly educated women. Focusing on the MRP estimates, note that the MRP point estimate for the NYT survey is very close to the original raw estimate, while the MRP estimate for the UMD is substantially different. This is related to the differing representativeness of the surveys — along the demographic lines where information was collected (i.e., age and decade married), the NYT survey appears to be fairly representative, and so the post-stratification aspect adds very little. In contrast, the UMD MRP estimates are pulled downwards substantially. The result is that while the MRP UMD estimate is still higher than the NYT survey, they are now much closer, and the uncertainty intervals over lap (UMD: 0.27 (0.196, 0.364) versus NYT: 0.16 (0.087, 0.269)).

Figure 4 shows raw and MRP estimates by survey across age groups and decade of marriage. Again we see that NYT estimates are generally lower, and the MRP does not change the point estimates substantially. In comparison, the MRP estimates for the UMD are substantially lower than the raw point estimates, but still higher than the NYT survey. Looking at the estimates by age group, the NYT results suggest the proportion of women who kept their last name declines over age, from about 20% (95% CI: (0.117,0.363)) in the youngest age group to 15% (95% CI: (0.077, 0.257)) for ages 65+. Looking at estimates by age and education presented in Figure 5 in the Appendix, the increasing trend over age is being driven by trends in those with a bachelor or above.

In contrast, in the UMD results the trend over age is reversed, with the estimate for the 65+ age group being the highest at around 40% (95% CI: (0.281,0.523)). Turning to the estimates by decade married, we see that the trends over decade are largely consistent across the two surveys, with lowest proportions in the 1970s and highest in the 2010s.

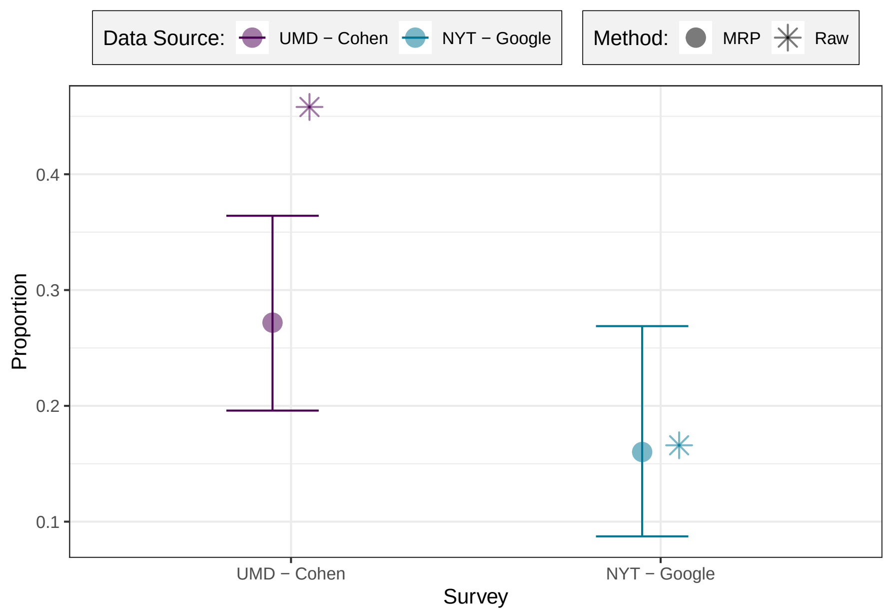

Figure 3: Raw and MRP national-level estimates of the proportion of married women who kept their last name.

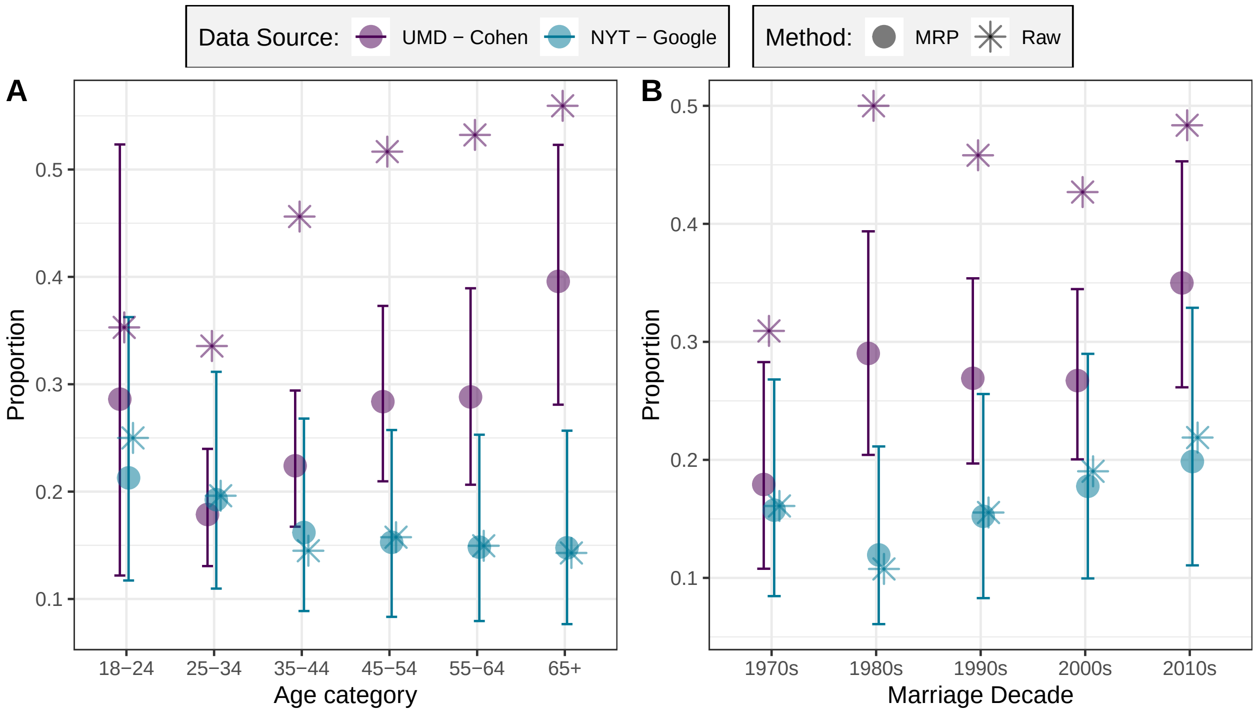

Figure 4: Raw and MRP estimates of the proportion of married women who kept their last name. Estimates are given for both surveys by (A) age category, (B) decade of marriage.

5 Discussion

In this chapter we discussed the use of multilevel regression and post-stratification (MRP) in the context of estimation in the population sciences. While MRP is widely used in the adjacent field of political science, the technique is somewhat under-utilized in the sociological and population sciences. However, as estimation of population-level indicators relies increasingly on 'non-traditional' data sources, such as data sourced from social media websites or online sources, MRP presents a method of adjusting for non-representative samples, producing more reasonable estimates for small or missing cells, and quantifying the level of uncertainty around these estimates.

We illustrated the application of MRP on two examples that were distinct both in terms of subject matter and type of data. In the first example, we considered data on measuring US women's intentions to have children in future. The original data came from a large, nationally-representative survey. For illustrative purposes, we took estimates of future fertility intentions as the 'truth' and sampled a smaller, less representative sample of women. Estimates obtained from MRP and post-stratification were then compared to the truth, with the former out-performing the latter. Of particular note was the ability of MRP to produce a non-zero estimate for the oldest age group (40-44 years) when no women in this age group in the smaller sample were observed to want more children. This shows the advantage of pooling information across age groups, with the resulting estimates being a weighted average of the information from the data and the information from the multilevel structure. In this case, we place a random walk structure on the age random effects to account for the fact that fertility intentions are likely to be correlated over age. As such the effect in each age group is conditionally independent of all age groups apart from the previous one. Allowing for patterns over age is likely to be appropriate to many estimation contexts across the population sciences.

In the second example, we applied MRP to two different surveys measuring the extent to which women chose to keep their name upon marriage. The two surveys — a Google Consumer Survey run by NYT, and an online survey run by sociologist Phil Cohen (the UMD survey) — are very different in terms of the information they collect and how information was collected. While the details of the NYT survey collection are not entirely known, it appears that respondents were targeted or collected to match the distribution of the national population based on age (and potentially other demographic characteristics). Information collected in the NYT survey was relatively minimal and it seems that age and state information was inferred rather than self-reported. In contrast, the UMD survey is a convenience sample, capturing a highly selected subgroup of women, mostly through social media channels (Twitter). However, the information collected in this survey is much more detailed: importantly, education attainment is recorded, but there is also information on the spouse, and a more detailed description of the name change; for example, who changed their name, if any one, or whether the names were combined. Due to wanting to compare results to the NYT survey, this level of detail was not used here, but could be useful for future analyses. The effect of the MRP method on the two surveys was quite different, with MRP estimates deviating only marginally from the original raw estimates for the NYT survey, while estimates were decreased substantially as a consequence of MRP for the UMD survey, adjusting for the survey sample being skewed towards the highly educated. The UMD MRP estimates, however, were generally higher than NYT estimates, with the gap at the national level being about 10 percentage points. Additionally, the two surveys suggested diverging trends over age.

In the marital name change example, the question remains: which estimates do we believe more? It's difficult to say for certain in the absence of a gold standard set of estimates, but the results in this case study nicely highlight what MRP can and can't do in terms of adjusting imperfect survey data. In the UMD survey case, MRP was quite effective in adjusting for the skewed representation, particularly along education lines, and for obtaining a set of estimates for stratification cells that had very few or no observations. However, MRP models used in this way are unable to adjust for other issues related to response bias. For the UMD survey, most women opted to answer after seeing the survey advertised on Twitter. Note that this 'opt-in' style of data collection is different to the classic Xbox paper for example, where users were made to take a survey before continuing on with the video game. This presents different data issues. Firstly, it is reasonable to think that those women answering the survey are more likely to have changed their name and want that fact to be recorded. Secondly, the group of women who use Twitter — and would have seen this tweet, made by an academic sociologist — are likely to be quite different to the general population, even after controlling for age, education, and decade married. There are many unobserved confounders that could help to explain the surprising trend over age, for example.

The analysis presented in this chapter highlights the strengths of MRP in population applications in its ability to produce estimates and uncertainty that account for issues of small populations and non-representativeness. However, the analysis also highlighted the limitations of MRP in dealing with non-response or omitted variable bias. In essence, MRP is founded in a generalized linear model framework, and if the assumptions behind that framework are invalid, then so are the estimates. In our view, an interesting avenue of future research in MRP as applied to population estimation problems is to focus on extending the multilevel model framework in order to better account for issues related to under- or over-reporting. For example, in the marital name change case, if we had additional information on name changes from another source (like newspaper announcements or administrative records), this could be used in combination with survey data to get a more holistic view of patterns by demographic characteristics and potentially be able to make adjustments to survey responses in a probabilistic way. When viewed in a Bayesian framework, standard multilevel models can naturally be extended in this way, accounting for various data types and reported populations. This flexible modeling strategy, in combination with post-stratification, presents a useful tool as the data used in population research continues to evolve.

6 References

BFRSS, 2022. https://www.cdc.gov/brfss/index.html.

IPUMS 2022. https://www.ipums.org/changes-to-census-bureau-data-products.

SBPC 2021. http://portal.sbpcnet.org.br/noticias/orcamento-de-2021-deixa-de-fora-o-censo-demografico/.

StatCan 2019. https://www150.statcan.gc.ca/n1/pub/89f0115x/89f0115x2019001-eng.htm#appb.

Abel, Ernest L, and Michael L Kruger. 2011. "Taking Thy Husband's Name: The Role of Religious Affiliation." *Names* 59 (1): 12–24. https://doi.org/10.1179/002777310X12759861710628.

Alexander, Monica, Kivan Polimis, and Emilio Zagheni. 2019. "The Impact of Hurricane Maria on Out-Migration from Puerto Rico: Evidence from Facebook Data." *Population and Development Review*, 617–30.

———. 2020. "Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States." *Population Research and Policy Review*, 1–28.

Bail, Chris. 2021. "Breaking the Social Media Prism." In *Breaking the Social Media Prism*. Princeton University Press.

Breen, Casey, Ayesha Mahmud, Dennis Feehan, and others. 2021. "Estimating Subnational Age-Specific Contact Patterns Using Multilevel Regression with Poststratification." *SocArXiv. November* 10.

Claassen, Christopher, and Richard Traunmüller. 2020. "Improving and Validating Survey Estimates of Religious Demography Using Bayesian Multilevel Models and Poststratification." *Sociological Methods & Research* 49 (3): 603–36. https://doi.org/10.1177/0049124118769086.

Cohen, Philip N. 2019. "Marital Name Change Survey," July. https://doi.org/10.17605/OSF.IO/UZQDN.

Downes, Marnie, Lyle C Gurrin, Dallas R English, Jane Pirkis, Dianne Currier, Matthew J Spittal, and John B Carlin. 2018. "Multilevel Regression and Poststratification: A Modeling Approach to Estimating Population Quantities From Highly Selected Survey Samples." *American Journal of Epidemiology* 187 (8): 1780–90. https://doi.org/10.1093/aje/kwy070.

Eke, P. I., X. Zhang, H. Lu, L. Wei, G. Thornton-Evans, K. J. Greenlund, J. B. Holt, and J. B. Croft. 2016. "Predicting Periodontitis at State and Local Levels in the United States." *Journal of Dental Research* 95 (5): 515–22. https://doi.org/10.1177/0022034516629112.

English, Patrick. 2021. "New YouGov MRP Model Shows Conservatives Losing up to 32 Seats in the Red Wall." https://yougov.co.uk/topics/politics/articles-reports/2021/10/04/new-yougov-mrp-model-shows-conservatives-losing-32.

Flores, Andrew R, Taylor NT Brown, and Jody Herman. 2016. *Race and Ethnicity of Adults Who Identify as Transgender in the United States.* Williams Institute, UCLA School of Law Los Angeles, CA.

Gelman, Andrew, Thomas C Little, and Morgan Stanley Dean Witter. n.d. "Poststrati Cation Into Many Categories Using Hierarchical Logistic Regression," 17.

Goldin, Claudia, and Maria Shim. 2004. "Making a Name: Women's Surnames at Marriage and Beyond." *Journal of Economic Perspectives* 18 (2): 143–60. https://doi.org/10.1257/0895330041371268.

Google Surveys. 2015. "Google Surveys - Survey Report." https://surveys.google.com/reporting/question?hl=en&survey=tpjh7dfv56ff2&question=1&raw=false&transpose=false&tab=chart&synonyms=true. https://surveys.google.com/reporting/question?hl=en&survey=tpjh7dfv56ff2&question=1&raw=false&transpose=false&tab=chart&synonyms=true.

Gorence, Patricia J. 1976. "Women's Name Rights." *Marquette Law Review* 59 (4): 26.

Herman, Jody L, Andrew R Flores, Taylor NT Brown, Bianca DM Wilson, and Kerith J Conron. 2017. *Age of Individuals Who Identify as Transgender in the United States.* eScholarship, University of California.

Hoffnung, Michele. 2006. "What's In a Name? Marital Name Choice Revisited." *Sex Roles* 55 (11-12): 817–25. https://doi.org/10.1007/s11199-006-9133-9.

Kashyap, Ridhi, R Gordon Rinderknecht, Aliakbar Akbaritabar, Diego Alburez-Gutierrez, Sofia Gil-Clavel, André Grow, Jisu Kim, et al. 2022. "Digital and Computational Demography."

Kopelman, Richard E., Rita J. Shea-Van Fossen, Eletherios Paraskevas, Leanna Lawter, and David J. Prottas. 2009. "The Bride Is Keeping Her Name: A 35-Year Retrospective Analysis of Trends and Correlates." *Social Behavior and Personality: An International Journal* 37 (5): 687–700. https://doi.org/10.2224/sbp.2009.37.5.687.

Morgan, S Philip, and Heather Rackin. 2010. "The Correspondence Between Fertility Intentions and Behavior in the United States." *Population and Development Review* 36 (1): 91–118.

National Center for Health Statistics. 1987. *National Survey of Family Growth.*

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12 (4): 375–85. https://doi.org/10.1093/pan/mph024.

Perrotta, Daniela, André Grow, Francesco Rampazzo, Jorge Cimentada, Emanuele Del Fava, Sofia Gil-Clavel, and Emilio Zagheni. 2021. "Behaviours and Attitudes in Response to the COVID-19 Pandemic: Insights from a Cross-National Facebook Survey." *EPJ Data Science* 10 (1): 17.

Rampazzo, Francesco, Emilio Zagheni, Ingmar Weber, Maria Rita Testa, and Francesco Billari.

2018. "Mater Certa Est, Pater Numquam: What Can Facebook Advertising Data Tell Us about Male Fertility Rates?" In *Twelfth International AAAI Conference on Web and Social Media*.

Ruggles, Steven, Robert McCaa, Matthew Sobek, and Lara Cleveland. 2015. "The IPUMS Collaboration: Integrating and Disseminating the World's Population Microdata." *Journal of Demographic Economics* 81 (2): 203–16.

Salganik, Matthew J. 2019. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

Scheuble, Laurie, and David R. Johnson. 1993. "Marital Name Change: Plans and Attitudes of College Students." *Journal of Marriage and the Family* 55 (3): 747. https://doi.org/10.2307/353354.

Schoen, Robert, Nan Marie Astone, Young J Kim, Constance A Nathanson, and Jason M Fields. 1999. "Do Fertility Intentions Affect Fertility Behavior?" *Journal of Marriage and the Family*, 790–99.

Schuster, Sherril Horowitz. 1997. "Here Comes the Bride: Wedding Announcements and Bridal Norms." *Sociological Focus* 30 (3): 279–94. https://doi.org/10.1080/00380237.1997.10571079.

Seheuble, Laurie K., Katherine Klingemann, and David R. Johnson. 2000. "Trends in Women's Marital Name Choices: 1966–1996." *Names* 48 (2): 105–14. https://doi.org/10.1179/nam.2000.48.2.105.

Stern, Michael J, Ipek Bilgen, and Don A Dillman. 2014. "The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design." *Field Methods* 26 (3): 284–301.

Stewart, Ian, René D Flores, Timothy Riffe, Ingmar Weber, and Emilio Zagheni. 2019. "Rock, Rap, or Reggaeton?: Assessing Mexican Immigrants' Cultural Assimilation Using Facebook Data." In *The World Wide Web Conference*, 3258–64.

The Upshot, THE UPSHOT. 2015. "About the Maiden-Name Analysis." *New York Times (Online)*. https://www.proquest.com/docview/1713685533/abstract/79AFF06FF0674238PQ/1.

Twenge, Jean M. 1997. "'Mrs. His Name': Women's Preferences For Married Names." *Psychology of Women Quarterly* 21 (3): 417–29. https://doi.org/10.1111/j.1471-6402.1997.tb00122.x.

University of Wisconsin Population Health Institute. 2022. "County Health Rankings & Roadmaps 2022." https://www.countyhealthrankings.org/.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting* 31 (3): 980–91.

Wang, Yan, James B Holt, Fang Xu, Xingyou Zhang, Daniel P Dooley, Hua Lu, and Janet B Croft.

2018. “Using 3 Health Surveys to Compare Multilevel Models for Small Area Estimation for Chronic Diseases and Health Behaviors.” *Preventing Chronic Disease* 15.

Warshaw, Christopher, and Jonathan Rodden. 2012. “How Should We Measure District-Level Public Opinion on Individual Issues?” *The Journal of Politics* 74 (1): 203–19. https://doi.org/10.1017/S0022381611001204.

Wiertz, Dingeman, and Chaeyoon Lim. 2021. “The Rise of the Nones Across the United States, 1973 to 2018: State-Level Trends of Religious Affiliation and Participation in the General Social Survey.” *Sociological Science* 8: 429–54. https://doi.org/10.15195/v8.a21.

Zhang, X., J. B. Holt, H. Lu, A. G. Wheaton, E. S. Ford, K. J. Greenlund, and J. B. Croft. 2014. “Multilevel Regression and Poststratification for Small-Area Estimation of Population Health Outcomes: A Case Study of Chronic Obstructive Pulmonary Disease Prevalence Using the Behavioral Risk Factor Surveillance System.” *American Journal of Epidemiology* 179 (8): 1025–33. https://doi.org/10.1093/aje/kwu018.

Zhang, Xingyou, James B Holt, Shumei Yun, Hua Lu, Kurt J Greenlund, and Janet B Croft. 2015. “Validation of Multilevel Regression and Poststratification Methodology for Small Area Estimation of Health Indicators from the Behavioral Risk Factor Surveillance System.” *American Journal of Epidemiology* 182 (2): 127–37.

Zheng, Xi, Xingyou Zhang, Celeste Jorge, and Diane Aye. 2014. “Model-Based Community Health Surveillance via Multilevel Small Area Estimation Using State Behavioral Risk Factor Surveillance System (Brfss): A Case Study in Connecticut.” *Available at SSRN 4051412*.

7 Appendix

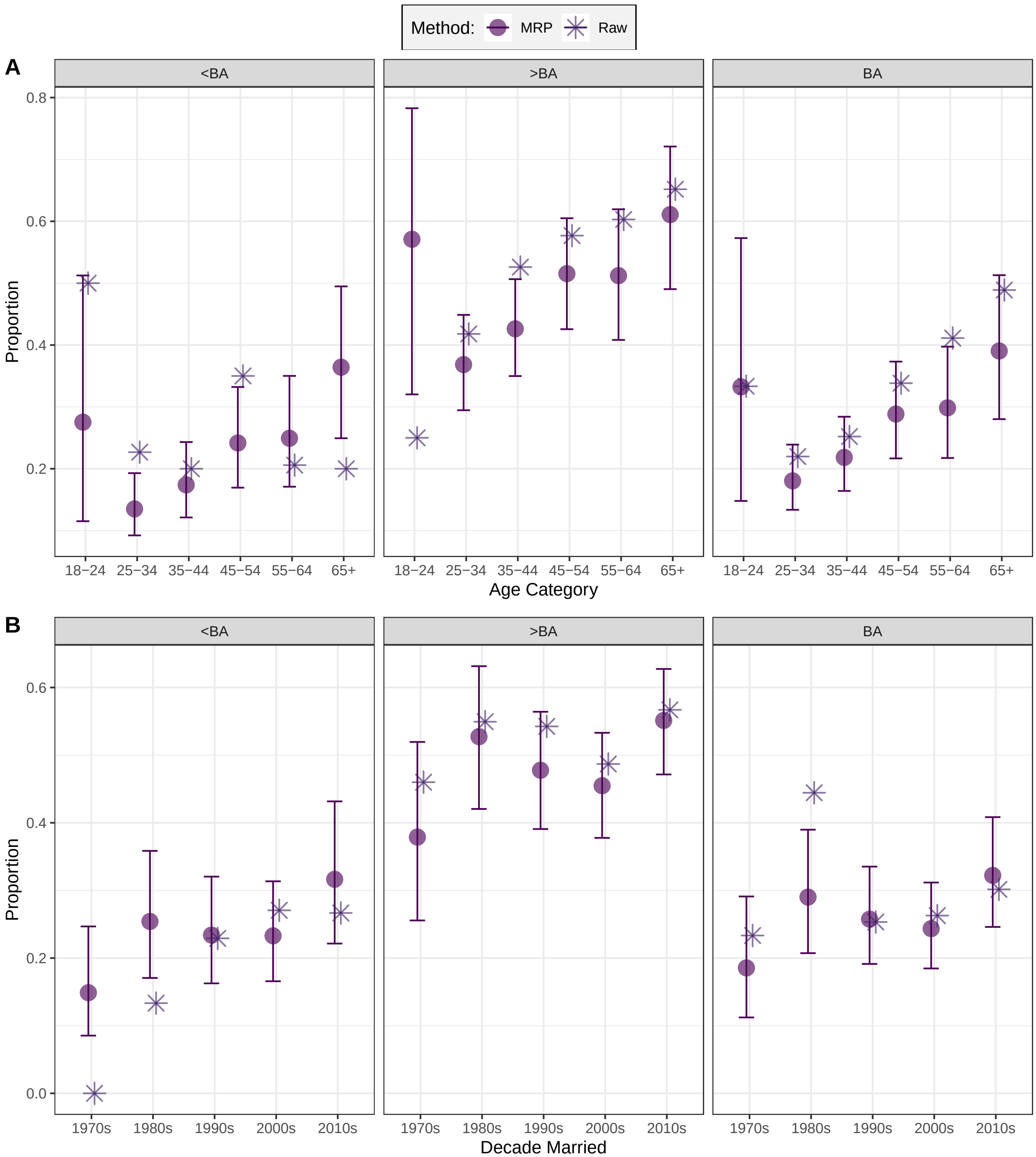

Figure 5: Raw and MRP estimates of the proportion of married women who kept last name. Estimates are based on the UMD survey alone and are shown by (A) age category and (B) decade of marriage further browken down by education level