ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2024, Volume 19, Issue 2

Small Area Estimation using Multilevel Regression and Poststratification to Estimate Cannabis Use in the State of Montana

Chase Walker(1), *Kristal Jones*(1), *Brandn Green*(1), *Frances L. Kim*(1)

(1) JG Research and Evaluation; Bozeman, Montana, United States.

CORRESPONDING AUTHOR: Chase Walker, JG Research and Evaluation, Bozeman, Montana, United States.
E-mail: chase@jgresearch.org. Tel: +1.406.360.5044

SUMMARY

Background: Small area substance use prevalence estimates at the county, city, or congressional district level are generally unavailable. In this study, we design a cannabis use survey for the state of Montana and use multilevel regression and poststratification (MRP) to generate county-level population prevalence estimates for past year cannabis use.
Methods: We developed a survey that asks questions about cannabis perceptions and use patterns. We analyzed the survey data specifically for the outcome variable of past year cannabis use using MRP to generate population level prevalence estimates at the county level for the state of Montana.
Results: We received 1,958 responses from our survey. We generated county level estimates by age group for cannabis use over the past year and found that MRP estimates were consistent with prior estimations of cannabis use at the state level and provided the ability to use additional data and validated assumptions to refine and downscale estimations of cannabis use, particularly in counties with low response rates.
Conclusion: Multi-modal survey dissemination was cost effective, but future surveys should intend to recruit a larger and more representative sample to minimize selection bias and improve estimation for demographic sub-groups. Overall, MRP provided a promising methodology for generating small-area cannabis use prevalence estimates, adjusting as much as possible for non-representativeness and non-response.

Keywords: MRP; multilevel regression; poststratification; surveillance; estimates; social media; substance use; cannabis; public health.

INTRODUCTION

Changes in the legal status of cannabis in the United States have created unknowns for public health practitioners seeking to understand trends in use patterns and the possible need for expanded educational or prevention interventions. National and state-level public health surveillance methods (National Survey on Drug Use and Health (NSDUH), Youth Risk Behavior Survey (YRBS)) have a necessary time lag between data collection and reporting, and they often do not estimate substance use at local levels due to small sample sizes, which can create a gap in understanding for local policymakers and agencies. To address these limitations, a broad set of efforts have been underway to use innovative survey recruitment methods for rapid surveillance in combination with new statistical methods for small-area estimation and analysis of non-representative surveys [1-4].

This study applies Bayesian multilevel regression and poststratification (MRP) to generate small area estimates for past year cannabis use from a non-randomized survey distributed through social media in Montana. The application of MRP within the context of drug use surveillance is distinctive. Gelman and Little [5], along with other political scientists [6; 7; 8; 9; 10; 11], have established the use of MRP with data from US pre-election polls to estimate election outcomes for a variety of subnational demographic-geographic groups. Consistently these studies find that MRP adjustments yield estimates consistent to other leading election poll analyses, thus demonstrating that non-representative polling can be used for measuring

DOI: 10.54103/2282-0930/26905
Accepted: 24th October 2024

public opinion. A similar application of MRP generate "dynamic" estimates of changing public opinion over time through its analysis of same sex marriage [12]. Most of these applications focus on state-level estimates from national surveys. However, there are examples of using MRP to generate smaller area estimates of political characteristics [13].

The majority of the use of MRP is centered on political science and forecasting election outcomes. However, other studies have demonstrated that MRP can be used more broadly across disciplines, including public health and epidemiology. Zhang et al [14] used MRP to generate small area estimates for chronic obstructive pulmonary disease; Eke et al [15] used it to predict periodontitis at state and local levels; Christofoletti et al [16] used it to estimate population-level leisure time compared to physical activity levels from large-scale health surveys in Brazil; and Downes et al [17] applied MRP to a large national health study to address analytical biases related to non-participation. Most recently, MRP was used to generate estimates of the proportion of people who identify as transgender for youth and adults in the United States [18]. The broad application to data across disciplines, particularly in public health studies examining socially critiqued health behaviors, suggests that MRP is a potentially useful methodology for generating prevalence estimates of substance use.

Surveillance surveys for producing prevalence estimates of marijuana use are primarily conducted at national and state levels. These surveys (e.g., NSDUH) are often conducted face-to-face or via the telephone, and the time intensity often leads to a number of individual responses that are too small to provide sufficient samples to generate small area estimates, especially in largely rural states like Montana with low population density. Web surveys, by contrast, are less expensive and resource intensive but can lead to lower response rates and other issues [19]. These barriers result in a general lack of reliable and accurate information about substance use patterns within many counties. It is also important to consider how willing survey respondents are to report on personal, private, or sensitive matters, as a substantial amount of prior research has found the mode of survey administration can impact the data quality for this type of information [20-25]. Several comparative studies suggest that having a web-based survey for asking respondents about topics like cannabis use, which is potentially sensitive given the ambiguous legal status within the United States, may improve reporting accuracy as compared to in-person or phone-based surveys [25, 26].

This research was originally completed for the state of Montana, which legalized marijuana for recreational use with the passage of ballot initiative 190 (I-190) in the 2020 Montana general election. Public health concerns about legalizing cannabis have included a focus on how use patterns for both adults and youth will change with increased access. County-specific information is especially important in the case of cannabis licensing within Montana, where the state implemented a licensing approach that allowed each county to determine whether or not they would allow for retail sales based on whether the percentage of yes votes on I-190 was greater than 50% or not. With between-county variation (28 of the 56 Montana counties voted in favor of I-190), it became essential to understand how use patterns within counties may be impacted by the new policy environment.

METHODS

MRP provides opportunities for generating small area prevalence estimates from survey data that can address sample bias from online survey sampling without some of the limitations associated with traditional weighting approaches. We developed and administered a cannabis use survey with multi-modal survey distribution and the use of MRP for generating population level small area prevalence estimates to address non-response and selection bias, and non-representativeness within survey samples.

Survey Design and Dissemination

The survey was designed using previously validated survey questions from YRBS, Behavioral Health Risk Factor Surveillance System (BRFSS), and the Canadian Cannabis Survey [27-29]. Survey items included questions on demographics (e.g. age, gender, race, county of residence, education, etc.), cannabis use characteristics for those who used cannabis, and knowledge and perceptions of cannabis and retail use in Montana. A total of 57 questions were asked, with 34 questions only being applicable for those who ever tried cannabis.

The survey was advertised on social media using ads targeted as specific geographies and age groups, providing a cost-effective approach to survey dissemination [30-34], as well as through posters and postcards in public health and social services offices. All modes of recruitment led to a web-based survey on the Alchemer platform. Survey consent questions were asked to ensure participants were at least 15 years old and if they were Montana residents. An IRB was submitted to Western IRB for this project under study number 1319497, however Western IRB found the project to be exempt because it was not collecting personal or identifiable information from subjects.

Multilevel regression

As this research is primarily focused on the potential of MRP methodology for producing SME for substance use behaviors, we present methods and results focused on one outcome, which is past year cannabis use.

We began by developing a multilevel regression model to predict the outcome measure, based on the demographic information from survey respondents, as well as county level predictors. We applied a multilevel logistic regression model to obtain estimates from the individual level survey responses, which allowed for poststratification, or weighting, using these estimates in the second stage of MRP. We started with the model below, generally following the notation of Lopez-Martin et al. [35] and Gelman and Hill [36]:

$$Y_i \sim Bernoulli(\pi_i) \qquad \pi_i = \Pr(past\_year\_use_i = 1)$$

$$logit(\pi_i) = \alpha^{age}_{a[i]} + \beta^{female} Female_i + \beta^{tried} Tried_i + \alpha^{female:age}_{g[i],a[i]} + \alpha^{county}_{c[i]}$$

Where:

$$\alpha^{county}_{c(i)} \sim normal(\mu^{county}_c, \sigma^{county}) \quad \text{for c = 1,…,56}$$

And where:

$$\alpha^{county}_{c(i)} = \delta^0 + \delta^{NW} NW_c + \delta^W W_c + \delta^{SW} SW_c + \delta^{SC} SC_c + \delta^{NC} NC_c + \delta^E E_c + \delta^{I190} I190_c$$

The model is reparametrized for analysis as:

$$\begin{aligned} logit(\pi_i) = {} & \mu^0 + \mu^{county}_{c[i]} + \alpha^{age}_{a[i]} + \beta^{female} Female_i + \beta^{tried} Tried_i + \alpha^{female:age}_{g[i],a[i]} \\ & + \delta^{NW} NW_{c(i)} + \delta^W W_{c(i)} + \delta^{SW} SW_{c(i)} + \delta^{SC} SC_{c(i)} + \delta^{NC} NC_{c(i)} + \delta^E E_{c(i)} \\ & + \delta^{I190} I190_{c(i)} \end{aligned}$$

Where:

$$\alpha^{age}_{a(i,c)} \sim normal(0, \sigma^{age}) \text{ for a = 1,…,4}$$

$$\alpha^{gen.age}_{a(i,c),g(i,c)} \sim normal(0, \sigma^{gen.age}) \text{ for g = 1,2 and a = 1,…,4}$$

$$Female_i = \begin{cases} 1\ if\ individual\ i\ is\ female \\ 0\ if\ individual\ i\ is\ male \end{cases}$$

$$Tried_i = \begin{cases} 1\ if\ individual\ i\ has\ tried\ cannabis\ at\ least\ once\ in\ their\ life \\ 0\ if\ individual\ i\ has\ never\ tried\ cannabis\ before \end{cases}$$

$$NW_{c(i)} = \begin{cases} 1\ if\ individual\ i\ and\ county\ c\ is\ in\ the\ NW\ region \\ 0\ otherwise \end{cases}$$

$$W_{c(i)} = \begin{cases} 1\ if\ individual\ i\ and\ county\ c\ is\ in\ the\ W\ region \\ 0\ otherwise \end{cases}$$

$$SW_{c(i)} = \begin{cases} 1\ if\ individual\ i\ and\ county\ c\ is\ in\ the\ SW\ region \\ 0\ otherwise \end{cases}$$

$$SC_{c(i)} = \begin{cases} 1\ if\ individual\ i\ and\ county\ c\ is\ in\ the\ SC\ region \\ 0\ otherwise \end{cases}$$

$$NC_{c(i)} = \begin{cases} 1\ if\ individual\ i\ and\ county\ c\ is\ in\ the\ NC\ region \\ 0\ otherwise \end{cases}$$

$$E_{c(i)} = \begin{cases} 1\ if\ individual\ i\ and\ county\ c\ is\ in\ the\ E\ region \\ 0\ otherwise \end{cases}$$

$$I190_{c(i)} = \begin{cases} 1\ I190\ vote\ was\ yes\ in\ county\ c\ where\ individual\ i\ lives \\ 0\ I190\ vote\ was\ no\ in\ county\ c\ where\ individual\ i\ lives \end{cases}$$

$$Model\ Priors = normal(0,1)$$

The intercept term represents the individual-level intercept now instead of the county-level intercept, and the term $\mu_{c(i)}^{county}$ represents the county-level adjustments after accounting for differences in regions and I190 votes. The model includes varying intercepts for age, and the interaction term of gender x age can be defined as the adjustments of individual *i's* age or gender x age on the probability of having used cannabis in the past year. $Female_i$ is an indicator variable that takes on a value of 1 if the individual is female and a value of 0 if the individual is male. $Tried_i$ is an indicator variable for if the respondent has tried marijuana at least once in their life, and it takes on a value of 1 if they have and a value of 0 if they have not.

The regional variables are indicator variables, accounting for unexplained variation among regions of the state that may be unaccounted for elsewhere. The county-level predictors represent group variables that account for structural differences among counties, so as to reduce unexplained county level variation. I-190 vote is a variable that represents if a county had a majority vote yes on ballot initiative 190. Including information on I-190 vote accounts for additional county-level variation that can be attributed to general perceptions of marijuana that are not captured by the demographic variables or broad regions.

Although it is more common to use indicator variables or fixed effects for demographic information such as age, representing this information as varying intercepts allows information to essentially be shared or to co-vary between levels of each of these variables, therefore preventing groups with less data from being overly sensitive to having fewer observed values, which is frequently an occurrence in some of our survey groups from lower population areas of the state [35].

We perform a Bayesian analysis using the stan_glmer function from the rstanarm [37] package in R to obtain a vector of 1,000 draws from joint posterior distribution of the model parameters. The MRP estimate for the given outcome in a Bayesian setting is a posterior distribution, and the estimate is displayed as the mean of that distribution. In the Bayesian specification, priors are a necessary part of the model to account for existing knowledge and information about relationships in the model. The rstanarm package by default provides weakly informative priors, but more information can be added to the priors if there is a known prior distribution for a variable, if the default priors make the posterior distribution difficult to explore, or if the default priors lead to computational issues, which is often times the case in modeling applications. To avoid issues, following Lopez-Martin et al. [35], we introduced stronger priors on the scaled coefficients at normal distribution (0, 1) and adjusted the adapt_delta to equal 0.95.

Poststratification

The second step in the MRP modeling process is poststratification: weighting the model estimates for subgroups with more representative population data to correct for some of the known differences between the sample and population of interest. The poststratification table was created by generating a cell with every possible combination of demographic and descriptive attributes and then weighting each corresponding model estimate for each cell by the relative proportion in the population estimates. The demographic subgroups specifically were used for every possible combination of the demographic and geographic variables in the multilevel model. For example, one poststratification cell may include the total number of females, aged 21 to 30, in a given county, meaning there will be a cell with the total count of individuals for every other combination of gender, age, and county.

Once again following the notation of Lopez-Martin et al. [35], the poststratification estimate can be defined as:

$$\theta^{MRP} = \frac{\sum N_j \theta_j}{\sum N_j}$$

Where θ^{MRP} is the final MRP estimate, θ_j is the estimate generated from the multilevel model (using the survey data) for demographic subgroup j in the poststratification table, with corresponding entry of N_j representing the number of people in that subgroup from the population.

In most applications of MRP, population level data come from comprehensive population level surveys, ACS 5-year estimates conducted by the U.S. Census Bureau or the Decennial Census; however, these data are not publicly available at the county level. To address this limitation, we used an alternative source of demographic county-level data (the Survey, Epidemiology, and End Results (SEER) data from the U.S. National Cancer Institute) for the post-stratification table [38]. SEER receives individual response data from the U.S. Census Bureau bridged population estimates and reports county and census tract estimates that include several demographic characteristics (sex, age, race).

The county data included demographic information (proportions) for gender and age by county, as well as the group predictors included in the modeling that vary by county only: the percentage of people by county who voted yes on I-190 and region of the state. Additionally, we needed information on the percentage of people who had ever tried cannabis in the poststratification table. Similar to the group predictors (region and I-190 vote), the poststratification data for ever having tried cannabis would not add new cells to the poststratification table, but rather than varying by county, the ever-tried variable would vary by the age groups. The data for ever tried was gathered from multiple sources, including NSDUH [28], Montana YRBS [27], and Gallup [39]. The county-level poststratification variables were defined as follows:

Gender: Female, Male (G = 2 categories)

Age Group: 15–20, 21–30, 31–40, and 41+ (A = 4 categories)

County: (C = 56)

The poststratification table has a cell for every possible combination of these levels, which in this case was 2 x 4 x 56 = 448 rows.

RESULTS

While this survey covered a wide range of cannabis related questions, this paper focuses only on prevalence of cannabis use in the last year and the methodology used to generate these estimates. The survey relied on self-selected responses; therefore, we do not expect the sample to be representative of the Montana population. Overall, the cannabis survey contained a slightly larger proportion of responses from females than males compared to the population. The largest proportion of responses came from the oldest age group, which is also the highest population age group in Montana. Most significantly, the survey was not representative of overall historical cannabis use. This is likely due to non-response or selection bias, evidenced by a substantial skew within the survey data towards individuals who have used cannabis responding to the survey at a much higher rate than non-users, as displayed in Table 1 below. In addition, there are low response rates for many counties across the state, particularly rural counties with already low populations. These biases are a primary motivation for utilizing MRP to generate estimates.

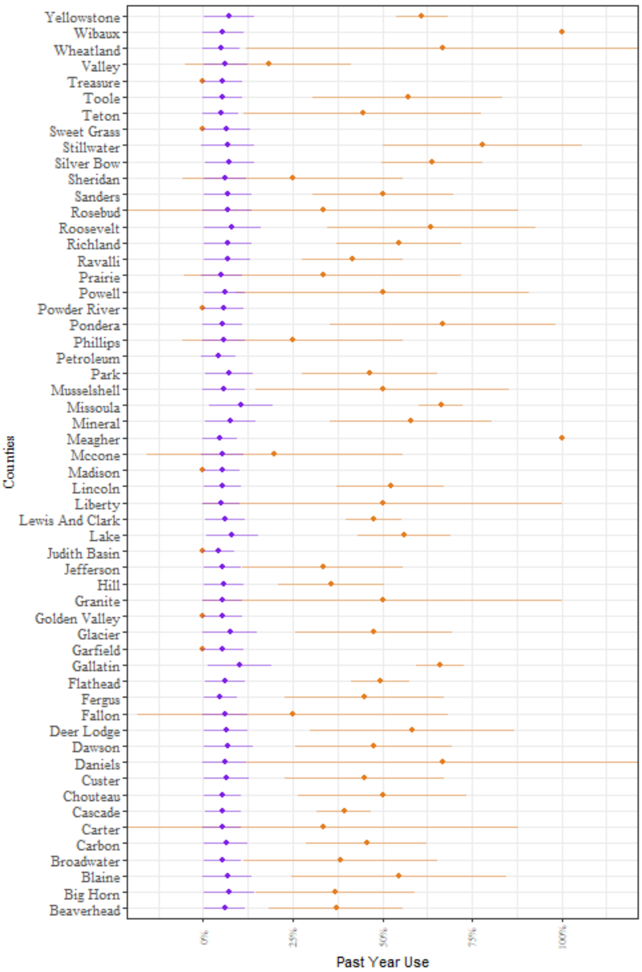
Table 1. Survey respondents who have ever used marijuana

Response	N	%
Tried Marijuana	1,569	*80.46*
Never Tried Marijuana	381	*19.54*

Estimates for past-year cannabis use, measured as the mean of the vector of posterior distributions from MRP by county are displayed below. We find that, in general, counties with a higher population in the state have slightly higher estimated cannabis use patterns than more rural counties. Missoula and Gallatin counties have the largest estimated percentages of cannabis use within the state. These results are unsurprising, as we found that the age group from 21 to 30 exhibit the highest cannabis use of any age group, and this age group is found in higher density in the higher population areas. Additionally, Missoula and Gallatin counties contain the two large universities in the state, which concentrate a large portion of young adults, who are consequently more likely to use cannabis according to our results. Figure 1 displays MRP estimates (purple), and 95% confidence intervals compared to the raw and unadjusted survey responses (orange). As can be seen from the figure, the higher-than-expected use rates across all counties are evident in the raw data, consistent with a greater response rate for users than non-users. The use of MRP provides a clear advantage over using the survey results as the only source of information and not attempting to correct for bias. Centered around a much lower range of use across counties, the MRP estimates match expectations that marijuana use is not likely to display such a wide range of use across counties.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2024, Volume 19, Issue 2

Figure 1. Cannabis use by county: MRP vs. raw survey results



DISCUSSION

One of the primary and well-documented advantages to using MRP is the ability to adjust for highly non-representative data to generate reliable population estimates at national and state levels [11]. In this study, we found that there are key differences and potential barriers between national and state-level applications compared to smaller area applications such as ours that are important for researchers to consider.

One of the advantages of the Bayesian specification and multilevel model is that estimates for demographic or geographic groups with relatively sparse information can be improved through "borrowing strength" from demographically similar cells [35, 11]. Having few survey responses from a demographic or geographic group is likely to be a common issue in surveys that target small areas such as counties, congressional districts, or small towns, where the number of responses may be substantially lower than larger population areas. If, for instance, all of the individuals responding to the survey within a certain age group indicate they were marijuana users, then under a traditional point estimate approach the model would predict that all people within that age group are marijuana users. In the approach used in this study, following Lopez-Martin et. al [35], the model partially pulls the varying intercept for the age group in our example towards the average across all of the other age groups.

However, the varying intercepts approach, which assumes that each group shares a common distribution, does not always generate model predictions that are useful or accurate when there is little information available for certain groups. The use of varying intercepts for demographic variables partially pools information from each level of a variable towards the average across all levels of that variable. For example, while we found that we were able to effectively generate estimates for demographic subgroups such as age and gender, we found that estimates for race were more challenging. Although survey responses were proportionally similar to the racial distribution in Montana, all minority races except for Native American had responses in the single digits. Estimates that would have been generated from the models, particularly in the small area estimation from the county-level model, for any given minority racial group would be very close to the average across all races, which is primarily white in Montana (92% of the survey respondents were white, which is very similar to the percentage of the population in Montana that is white). Therefore, the average across all races is heavily weighted by responses from white individuals, meaning the estimates for minorities would be pulled closer to the estimated rates of use for white individuals and therefore likely not provide accurate estimations of use for other races.

Another key consideration in the use of MRP to quantify responses from a multi-modal survey as used in this study is the need to account for non-response and selection bias in the results. MRP is capable of adjusting for these biases and is a key reason that it is used in forecasting election outcomes, where researchers include variables to control for political factors such as prior votes in the respondent's geographic area, respondents party affiliation, respondents' religious ideology, and respondents' income [6, 10, 11]. This potential informed our selection of MRP for the project, and the research team included controls that we believed to influence whether someone had used marijuana in the past year such as whether a person had ever tried marijuana previously in their life, and the I-190 vote in their county.

A key aspect of MRP is that all control variables in the first-stage model must also be present in the poststratification data at the population level in order to generate estimates. These control factors do not add any rows to the poststratification but are typically expanded to match with geographic factors such as state or region. In our case, when our estimates are by county, the gaps in public health surveillance data on cannabis use made it challenging for our team to identify reliable estimates of some control variables, (i.e., the proportion of individuals who have ever tried marijuana) at the county level. Specifically, our control variable accounting for whether someone had ever tried marijuana in their life did not have population-level estimates by county for Montana, thus introducing a substantial and unanticipated challenge in our modeling that had not appeared to be a previous issue based on our reading of the literature.

To address this challenge, we utilized estimates from other surveys that had attempted to generate estimates at the state and national levels on marijuana use, which included the NSDUH [29], the Montana YRBS [28], and Gallup polls [39]. Each of these surveys contains estimates by age group for the percentage of the population who had ever tried marijuana, so we utilized a combination of these data to incorporate the information into the poststratification table as an alternate work-around to not having county data on lifetime cannabis use available.

Future researchers seeking to utilize MRP will need to consider data availability at the population level for poststratification during the research design phase, and plan survey design and questions for control factors they intend to include with this in mind to ensure that estimation is possible and limit modeling setbacks.

ABBREVIATIONS

I-190: Initiative 190
MRP: Multilevel regression and poststratification
NSDUH: National Survey on Drug Use and Health
YRBS: Youth Risky Behavior Survey

DISCLAIMERS

The results presented here are the work of the authors alone and do not reflect the views of the Department of Public Health and Human Services of the State of Montana.

FUNDING

The original data collection for this work was supported by a contract from the Department of Public Health and Human Services of the State of Montana.

ACKNOWLEDGEMENTS

We would like to thank Megan Higgs of Critical Inference, LLC for providing a review of our statistical methodology. This report results from work originally done for the Montana Public Health Institute (MPHI) and Montana Department of Health and Human Services (DPHHS). MPHI and DPHSS played no role in the writing of this report, the methodology used, or the decision to submit this paper for publication.

CONFLICTS OF INTEREST

None declared.

REFERENCES

1. Fazzino TL, Rose G, Pollack S, Helzer J. Recruiting U.S. and Canadian College Students via Social Media for Participation in a Web-Based Brief Intervention Study. Journal of Studies on Alcohol and Drugs. 2015;76(1):127-132. doi:10.15288/jsad.76.1.127
2. Park DK, Gelman A, Bafumi J. Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls. Polit anal. 2004;12(4):375-385. doi:10.1093/pan/mph024
3. Rait MA, Prochaska JJ, Rubinstein ML. Recruitment of adolescents for a smoking study: use of traditional strategies and social media. Behav Med Pract Policy Res. 2015;5(3):254-259. doi:10.1007/s13142-015-0312-5
4. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res. 2016;18(11):e286. doi:10.2196/jmir.5698
5. Gelman A, Little TC. Poststratification Into Many Categories Using Hierarchical Logistic Regression. Survey Methodology. Published online 1997:1-17.
6. Ghitza Y, Gelman A. Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups: DEEP INTERACTIONS WITH MRP. American Journal of Political Science. 2013;57(3):762-776. doi:10.1111/ajps.12004
7. Lauderdale BE, Bailey D, Blumenau J, Rivers D. Model-based pre-election polling for national and sub-national outcomes in the US and UK. International Journal of Forecasting. 2020;36(2):399-413. doi:10.1016/j.ijforecast.2019.05.012
8. Trangucci R, Ali I, Gelman A, Rivers D. Voting Patterns in 2016: Exploration Using Multilevel Regression and Poststratification (MRP) on Pre-Election Polls.; 2018.
9. Lax JR, Phillips JH. Gay Rights in the States: Public Opinion and Policy Responsiveness. Am Polit Sci Rev. 2009;103(3):367-386. doi:10.1017/S0003055409990050
10. Park DK, Gelman A, Bafumi J. Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls. Polit anal. 2004;12(4):375-385. doi:10.1093/pan/mph024
11. Wang W, Rothschild D, Goel S, Gelman A. Forecasting elections with non-representative polls. International Journal of Forecasting. 2015;31(3):980-991.

doi:10.1016/j.ijforecast.2014.06.001
12. Gelman A, Lax J, Phillips J, Gabry J, Trangucci R. Using Multilevel Regression and Poststratification to Estimate Dynamic Public Opinion.; 2018.
13. Tausanovitch C, Warshaw C. Representation in Municipal Government. Am Polit Sci Rev. 2014;108(3):605-641. doi:10.1017/S0003055414000318
14. Zhang X, Holt JB, Lu H, et al. 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. 2014;179(8):1025-1033. doi:10.1093/aje/kwu018
15. Eke PI, Zhang X, Lu H, et al. Predicting Periodontitis at State and Local Levels in the United States. J Dent Res. 2016;95(5):515-522. doi:10.1177/0022034516629112
16. Christofoletti M, Benedetti TRB, Mendes FG, Carvalho HM. Using Multilevel Regression and Poststratification to Estimate Physical Activity Levels from Health Surveys. IJERPH. 2021;18(14):7477. doi:10.3390/ijerph18147477
17. Downes M, Gurrin LC, English DR, et al. Multilevel Regression and Poststratification: A Modeling Approach to Estimating Population Quantities From Highly Selected Survey Samples. American Journal of Epidemiology. 2018;187(8):1780-1790. doi:10.1093/aje/kwy070
18. Herman JL, Flores AR, O'Neill KK. How Many Adults and Youth Identify As Transgender In The United States? UCLA Williams Institute School of Law; 2022.
19. Jones MK, Calzavara L, Allman D, Worthington CA, Tyndall M, Iveniuk J. A Comparison of Web and Telephone Responses From a National HIV and AIDS Survey. JMIR Public Health Surveill. 2016;2(2):e37. doi:10.2196/publichealth.5184
20. Chang L, Krosnick JA. National Surveys Via Rdd Telephone Interviewing Versus the Internet. Public Opinion Quarterly. 2009;73(4):641-678. doi:10.1093/poq/nfp075
21. Fricker S, Galesic M, Tourangeau R, Yan T. An Experimental Comparison of Web and Telephone Surveys. Public Opinion Quarterly. 2005;69(3):370-392. doi:10.1093/poq/nfi027
22. Greene J, Speizer H, Wiitala W. Telephone and Web: Mixed-Mode Challenge: Mixed Mode Surveys. Health Services Research. 2007;43(1p1):230-248. doi:10.1111/j.1475-6773.2007.00747.x
23. Link M, Mokdad A. Effects of survey mode on self-reports of adult alcohol consumption: A comparison of mail, web and telephone approaches. Journal of Studies on Alcohol. 2005;66(2):239-245. doi:10.15288/jsa.2005.66.239
24. Plante C, Jacques L, Chevalier S, Fournier M. Comparability of Internet and Telephone Data in a Survey on the Respiratory Health of Children. Canadian Respiratory Journal. 2012;19(1):13-18. doi:10.1155/2012/318941
25. Tourangeau R. Measurement Properties of Web Surveys. In: Beyond traditional survey taking: adapting to a changing world. In: Proceedings of Statistics Canada Symposium. ; 2014.
26. Burkill S, Copas A, Couper MP, et al. Using the Web to Collect Data on Sensitive Behaviours: A Study Looking at Mode Effects on the British National Survey of Sexual Attitudes and Lifestyles. Cardoso MA, ed. PLoS ONE. 2016;11(2):e0147983. doi:10.1371/journal.pone.0147983
27. Fischer B, Lee A, Robinson T, Hall W. An overview of select cannabis use and supply indicators pre- and post-legalization in Canada. Subst Abuse Treat Prev Policy. 2021;16(1):77. doi:10.1186/s13011-021-00405-7
28. Montana Office of Public Instruction. 2021 Montana Youth Risky Behavior Survey - High School Results.; 2021. https://opi.mt.gov/Portals/182/Page%20Files/YRBS/2021YRBS/2021_MT_YRBS_FullReport_Sept30.pdf?ver=2021-10-01-082901-390
29. SAMHSA. National Survey on Drug Use and Health: Model-Based Prevalence Estimates (50 States and the District of Columbia).; 2018. https://www.samhsa.gov/data/sites/default/files/reports/rpt32805/2019NSDUHsaeExcelPercents/2019NSDUHsaeExcelPercents/2019NSDUHsaePercents.pdf
30. Fazzino TL, Rose G, Pollack S, Helzer J. Recruiting U.S. and Canadian College Students via Social Media for Participation in a Web-Based Brief Intervention Study. Journal of Studies on Alcohol and Drugs. 2015;76(1):127-132. doi:10.15288/jsad.76.1.127
31. Ford KL, Albritton T, Dunn TA, Crawford K, Neuwirth J, Bull S. Youth Study Recruitment Using Paid Advertising on Instagram, Snapchat, and Facebook: Cross-Sectional Survey Study. JMIR Public Health Surveill. 2019;5(4):e14080. doi:10.2196/14080
32. Park B, Calamaro C. A systematic review of social networking sites: innovative platforms for health research targeting adolescents and young adults. Journal of Nursing Scholarship. 2013;45(3). doi:10.1111/jnu.12032
33. Rait MA, Prochaska JJ, Rubinstein ML. Recruitment of adolescents for a smoking study: use of traditional strategies and social media. Behav Med Pract Policy Res. 2015;5(3):254-259. doi:10.1007/s13142-015-0312-5
34. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res. 2016;18(11):e286. doi:10.2196/jmir.5698
35. Lopez-Martin J, Phillips J, Gelman A. Multilevel Regression and Poststratification Case Studies. Published 2021. https://bookdown.org/jl5522/MRP-case-studies/
36. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press; 2007.
37. Goodrich B, Gabry J, Ali I, Brilleman S. rstanarm: Bayesian applied regression modeling via Stan. Published online 2020. https://mc-stan.org/rstanarm

38. National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2020). National Cancer Institute, DCCPS, Surveillance Research Program www.seer.cancer.gov/popdata

39. Jones J. Nearly Half of U.S. Adults Have Tried Marijuana. Gallup. https://news.gallup.com/poll/353645/nearly-half-adults-tried-marijuana.aspx. Published 2021.