Annals of Epidemiology 39 (2019) 15–20

Contents lists available at ScienceDirect

Annals of Epidemiology

Original article

Using multilevel regression with poststratification to obtain regional health estimates from a Facebook-recruited sample

Travis Loux, PhD [a, *], Erik J. Nelson, PhD [b], Lauren D. Arnold, PhD [a], Enbal Shacham, PhD [c], Mario Schootman, PhD [a, 1]

[a] Department of Epidemiology and Biostatistics, Saint Louis University College for Public Health and Social Justice, St. Louis, MO
[b] Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, IN
[c] Department of Behavioral Science and Health Education, Saint Louis University College for Public Health and Social Justice, St. Louis, MO

ARTICLE INFO

Article history:
Received 8 April 2019
Accepted 16 September 2019
Available online 25 September 2019

Keywords:
Prevalence
Public health surveillance
Selection bias
Social media
Statistical models

ABSTRACT

Purpose: We assess the effectiveness of multilevel regression with poststratification (MRP) as a tool to mitigate selection bias from online surveys of small geographical regions.
Methods: We collected self-reported health information from an Internet-based sample of adults residing within the St. Louis, MO, metropolitan area in 2017. We created Bayesian hierarchical models with three sets of predictor variables for each of six common health behaviors and outcomes, with results poststratified using the American Community Survey to estimate region and ZIP Code Tabulation Area –level prevalence.
Results: When comparing MRP estimates with a population-based sample as a reference, we found that adjustment using MRP can reduce bias in prevalence estimates and provide estimates for local area prevalence. 14 of 18 adjusted estimates were closer to the benchmark than the unadjusted estimates and MRP using all three covariate sets resulted in better overall agreement with the benchmark compared with the unadjusted estimates.
Conclusions: MRP can improve prevalence estimates from self-selected Internet-based samples, although a nonnegligible amount of bias may remain. Illustrating the utility and limitations of this method will help researchers develop relevant estimates of the local public health burden, helping local health officials better understand and reduce poor health outcomes.

Introduction

Obtaining survey data for small geographic areas is of primary importance in epidemiological research that focuses on neighborhood and place. However, such data are not readily available via pre-existing sources. National data sets such as the Behavioral Risk Factor Surveillance System (BRFSS) do not provide local area information because of potential lack of confidentiality and low reliability of estimates based on small samples [1,2]. With concerns over data privacy and emphasis on improving data privacy regulation [3], granular geographic data are not likely to be made available to the public by existing surveillance programs. This may impede the implementation of local public health programs and geographically targeted decision-making.

To fill the gaps in availability of secondary data, researchers may collect primary data within a targeted geographic area; however, they often encounter difficulties obtaining a representative sample from their target population. Traditional recruitment strategies such as telephone and mail have shown decreasing response rates over recent years [4,5], resulting in bias in even well-designed studies [6]. Thus, researchers have turned to the Internet as a recruitment platform [7]. While self-selection into an Internet-based survey does not solve the problem of nonrepresentativeness, the resulting sample can have similar quality to those from designed studies with high nonresponse [8–11].

Multilevel regression with poststratification (MRP) is one approach to make results from Internet samples more representative of the target population [12]. The use of multilevel models shrinks covariate-adjusted estimates toward the overall mean,

Conflicts of interest: No potential conflicts of interest relevant to this article were reported.

* Corresponding author. Department of Epidemiology and Biostatistics, Saint Louis University College for Public Health and Social Justice, 3545 Lafayette Ave, St. Louis, MO, 63110. Tel.: 314-977-2003.
E-mail address: travis.loux@slu.edu (T. Loux).
[1] Present address: Department of Clinical Analytics and Insights, Center for Clinical Excellence, SSM Health, St. Louis MO USA

https://doi.org/10.1016/j.annepidem.2019.09.005
1047-2797/

reducing the risk of extreme model predictions, especially for covariate combinations with small sample sizes. In MRP, an outcome model is built from a sample then applied to population-based data such as a government census. An outcome is predicted for each stratum defined by the variables in the regression model. The stratum-specific estimates are then averaged, using the census stratum size as a weight. Assuming the researcher has controlled for all covariates associated with both selection into the sample and the outcome, the resulting estimate will be approximately unbiased in the target population [12]. Originally used in the context of public opinion polling [13–15], MRP has been used in recent population health studies to obtain state and county level estimates for various health outcomes using BRFSS data [16] and to estimate health outcomes for Australian territories [17].

In public health, MRP has predominantly been used for studies that use traditional survey methods with large sample sizes recruited from large, highly populated geographic areas, with poststratification based on the most recent national census estimates. However, the utility of these methods for Internet-based studies of moderate size that focus on small geographic areas is unknown.

We determined the feasibility of MRP for reducing bias in local estimates when the source data come from a moderate-sized, nonrandom, Internet-based sample of a small geographic area by comparing MRP-adjusted estimates from this sample to population-based estimates developed by the Missouri County-Level Study (CLS) [18].

Materials and methods

Survey data

Respondents were recruited to take an online survey via Facebook advertisements between February and December 2017. Inclusion criteria were residence in one of the 122 ZIP codes within the St. Louis metropolitan area, age 18–65 years, English-speaking, and not institutionalized or residents of assisted living facilities. We matched the self-reported residential ZIP code with the corresponding ZIP Code Tabulation Areas (ZCTAs) to delineate geographic areas, as this facilitates the implementation of locally based interventions and inferences [19]. ZCTAs are generalized areal representations of U.S. Postal Service ZIP code service areas and, in most instances, are equivalent to the ZIP code [20]. Health outcomes included self-reported physician-diagnosed arthritis, asthma, depression, and diabetes; physical inactivity in the month proceeding the survey; and current smoking status. These health outcomes were chosen because they are important health problems across many populations and are measured in the CLS [18], providing a reference against which to compare our results. Wording of the survey items were drawn from the CLS, allowing for direct comparisons. All participants provided written consent before taking the online survey. The university's Institutional Review Board approved of this study.

Population data

ZCTA-level population data for the St. Louis metropolitan area were obtained from the American Community Survey (ACS) 5-year estimates (2012–2016) through Integrated Public Use Microdata Series National Historical Geographic Information System (IPUMS NHGIS) [21].

The Census Bureau limits the number of variables cross-classified at the ZCTA level. To compare adjustment models, three separate ZCTA population data files were used: (1) age by sex by education level, (2) age by sex by race, and (3) sex by race by education level. These characteristics have been used in previous MRP studies of health outcomes [16,17]. The ACS does not allow stratification by age, education, and race at the ZCTA level, making analysis based on this variable set not possible.

Population-based outcome estimates

Prevalence estimates for health outcomes of interest in the St. Louis metropolitan area were obtained from the 2016 CLS [18]. The CLS used random digit dialing to take a county-stratified random sample of noninstitutionalized English-speaking Missouri residents. The 2016 study sampled 1200 respondents from the City of St. Louis, 2000 from St. Louis County, and 400–800 from the other eight counties in the St. Louis metropolitan area, along with similar-sized samples for every other county in the state. Responses were weighted to be representative of the Missouri adult population at the county level following the raking approach used by BRFSS [2]. We obtained metropolitan area estimates using population-weighted averages of the relevant county estimates.

Statistical methods

The Facebook sample contained 425 unique respondents. Demographic information from the respondents was compared with metropolitan area population data via the Cramer's V effect size measure [22] to indicate magnitude of nonrepresentativeness of the sample. By rule of thumb, Cramer's V between 0.3 and 0.5 are considered “moderate” differences, whereas Cramer's V greater than 0.5 are considered “large” differences. In line with the ACS data, Facebook respondent ages were categorized as 18–24, 25–34, 34–44, and $\geq$ 45 years. Respondent race was reported as non-Hispanic white, non-Hispanic Asian, non-Hispanic black, Hispanic, and other. Education level was reported as less than high school diploma, high school diploma or equivalent, some college/associate's degree, bachelor's degree, and graduate/professional degree. Participants' home address was recorded as one of St. Louis City, St. Louis County, or another county in the metropolitan area.

Missing data for any relevant variable was multiply imputed using chained equations [23] to create five complete data sets. Details regarding missing data are provided in the Supplementary Materials. Within each imputed data set, three Bayesian hierarchical binary logistic regressions were performed for each outcome using individual-level and ZCTA-level predictors. For MRP model 1, the ACS variables were age, sex, and education level; for MRP model 2, the ACS variables were age, sex, and race; and for MRP model 3, the ACS variables were sex, race, and education level (Table 1). In

Table 1
Comparison of model predictors for multilevel regression with poststratification (MRP) adjustment

Predictor	MRP model 1	MRP model 2	MRP model 3
ZCTA % white	X	X	X
ZCTA % Bachelor's degree	X	X	X
ZCTA median income	X	X	X
County of residence	X	X	X
Sex	2	2	2
Age	4	4	
Education	5		4*
Race		5	5
Total number of strata	4880	4880	4880
Empty population strata	1365	1542	1639

Numbers in cells of individual-level characteristics represent strata used in MRP analysis. Total number of strata determined by product of 122 ZCTAs and numbers of strata of individual-level characteristics sex, age, education, and race as used in the model.

* Bachelor's degree and graduate degree combined to match ACS stratification.

Table 2
A comparison of the Facebook sample with the St. Louis (STL) metropolitan area

Characteristic	FB sample	STL metro area	Cramer's V
	n = 425	n = 2,108,538	
Age (%)			0.83
18–24	150 (35.3)	192,189 (12.0)	
25–34	120 (28.2)	281.147 (17.5)	
35–44	36 (8.5)	271,235 (16.9)	
Over 44	119 (28.0)	861,081 (53.6)	
Race/ethnicity (%)			0.81
White non-Hispanic	261 (65.4)	1,219,202 (75.9)	
Asian non-Hispanic	58 (14.5)	40,790 (2.5)	
Black non-Hispanic	51 (12.8)	289,681 (18.0)	
Hispanic	14 (3.5)	34,540 (2.2)	
Other	15 (3.8)	21,439 (1.3)	
County (%)			0.79
St. Louis City	167 (39.3)	316,030 (15.0)	
St. Louis County	221 (52.0)	1,000,560 (47.5)	
Other metropolitan area counties	37 (8.7)	791,948 (37.6)	
Education (%)			0.72
Less than high school	8 (1.9)	132,446 (9.2)	
High school	28 (6.6)	359,616 (25.0)	
Some college	82 (19.3)	440,604 (30.6)	
Associate's degree	28 (6.6)		
Bachelor's degree	184 (43.4)	305,039 (21.2)	
Graduate/professional degree	94 (22.2)	201,438 (14.0)	
Income (%)			0.32
Less than $15,000	65 (17.8)	90,410 (10.8)	
$15,000-$25,000	38 (10.4)	79,627 (9.5)	
$25,000k-$35,000	48 (13.1)	79,875 (9.6)	
$35,000-$50,000	50 (13.7)	113,054 (13.5)	
$50,000-$75,000	71 (19.4)	153,327 (18.4)	
$75,000-$100,000	38 (10.4)	105,821 (12.7)	
More than $100,000	56 (15.3)	213,125 (25.5)	
Male (%)	157 (39.3)	1,019,301 (48.3)	0.18

St. Louis metropolitan area figures come from 5-year estimates (2012–2016) of the American Community Survey.
FB = Facebook.

model 3, bachelor's degree and graduate/professional degree were combined to match ACS stratification, which is limited by the number and sizes of strata. Each model also included census-obtained ZCTA-level percentage white, percentage with bachelor's degree or higher, and median income, respondent's county of residence, and a random effect for ZCTA. The coefficients for each predictor used a normal prior with mean 0 and standard deviation 2.5. With a sample of this size, results may be dependent on the prior distributions chosen; therefore, a sensitivity analysis using normal priors with standard deviation 1, causing more shrinkage to the overall mean, was performed. Each model was run on four chains with a 1000 iteration burn in per chain. Model fit was assessed through Bayesian R^2 [24] and Watanabe–Akaike information criterion. For each stratum defined by cross-classifying ZCTA with the ACS variables, 4000 draws from the posterior distribution of the linear predictor were taken. For ZCTAs which were not observed in the Facebook data set, the posterior distribution was marginalized across the other predictor variables. For each of the 4000 draws, the stratum-specific predictions were aggregated to the metropolitan area and ZCTA levels using weighted averages with stratum population counts as weights. The linear predictors were transformed to the proportion scale via inverse logit. The 4000 draws from each of the five imputed data sets were combined for a total of 20,000 draws from the posterior distribution accounting for uncertainty from the imputed data. The point estimate, standard error, and 95% credible intervals were obtained by the median, standard deviation, and middle 95% of the 20,000 draws, respectively.

Metropolitan area MRP results were compared with population-based estimates from the CLS directly and by computing 2-sample z-scores, $z = (p_{MRP} - p_{CLS}) / \sqrt{\sigma^2_{MRP} + \sigma^2_{CLS}}$, which can be interpreted as the distance between estimates accounting for estimation uncertainty. The absolute value of the raw difference and z-scores were summed across all assessed health outcomes for each model to create two measures of overall model closeness. All analyses were performed in R version 3.5.0 [25]. The packages VIM [26] and mice [27] were used for exploring and imputing missing data, respectively, and rstanarm [28] was used for hierarchical regression modeling.

Results

Consistent with CHERRIES definitions for online survey reporting [29], the Facebook survey had a participation rate of 85% (664 initiated surveys/774 unique survey page visits) and a completion rate of 64% (425 completed surveys/664 initiated surveys), respectively. Age, race/ethnicity, county of residence, and education appear to be highly unrepresentative within the Facebook sample (Table 2). Compared with the metropolitan area population, respondents in the Facebook sample were notably younger, with over 60% of Facebook respondents less than 35 years old. Individuals identifying as Asian were over-represented, whereas individuals identifying as white and black were under-represented. St. Louis City was over-represented in the sample at the expense of other metropolitan area counties. The Facebook sample was more highly educated than the general population, with 65% having obtained a bachelor's degree or more compared with 35% in the metro area population. Given these variables were strongly associated with sample selection, they are especially important to account for in adjustment. Figure 1 shows the geographic distribution of Facebook respondents. Respondents represented 75 of the 122 ZCTAs in the study region. Sample sizes within the 75 ZCTAs ranged from 1 to 33 respondents with a median of 3 respondents per represented ZCTA.

Model MRP 1 included age, sex, and education level as predictors, as well as the predictors used in all models. The means of the Bayesian R^2 posterior distributions ranged from 0.077 to 0.234 across the six outcome models (Supplementary Table 2). This model adjusted estimates closer to CLS estimates in four of the six outcomes by raw difference and five of the six outcomes by z-score

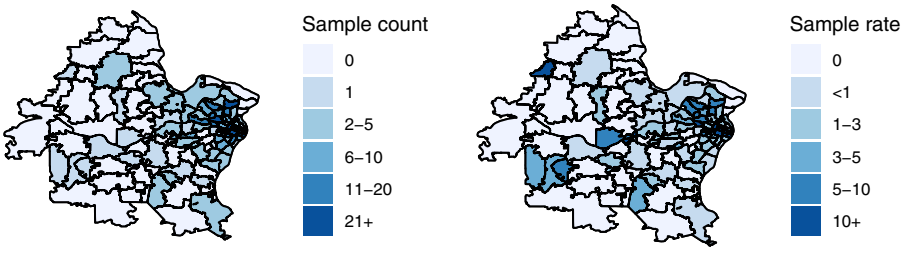


Fig. 1. Geographic distribution of sample. Left panel shows number of respondents per ZCTA. Right panel shows number of respondents per 10,000 ZCTA residents.

Table 3
Point estimates from unadjusted Facebook data and MRP-adjusted results with raw differences (d) 2-sample z-scores (z) comparing estimates to Missouri County-Level Study (CLS) estimates

Outcome	Raw FB data			MRP 1			MRP 2			MRP 3			CLS
	Est	d	z	Est	d	z	Est	d	z	Est	d	z	Est
Arthritis	0.141	−0.128	−4.96	0.208	**−0.061**	**−1.39**	0.187	−0.083	−2.06	0.137	−0.132	−3.89	0.269
Asthma (current)	0.089	**−0.006**	**−0.37**	0.075	−0.02	−0.99	0.084	−0.011	−0.48	0.078	−0.016	−0.79	0.095
Depression	0.284	0.067	2.41	0.315	0.098	2.02	0.269	0.053	1.25	0.254	**0.037**	**0.96**	0.217
Diabetes	0.064	−0.042	−2.76	0.120	0.013	0.36	0.109	**0.002**	**0.07**	0.074	−0.032	−1.04	0.106
Physical inactivity	0.173	−0.057	−2.23	0.280	0.05	0.97	0.210	**−0.02**	**−0.44**	0.187	−0.042	−1.08	0.230
Smoker (current)	0.118	−0.077	−3.46	0.265	0.069	1.51	0.167	**−0.028**	**−0.74**	0.162	−0.034	−0.95	0.196
Abs sum		0.378	16.19		0.312	7.23		**0.196**	**5.04**		0.294	8.69	

The final row gives the sum of the absolute differences across health outcome models. See Table 1 for the set of predictors used in each MRP model. Smallest d and z values for each outcome are in boldface font.
FB = Facebook.

(Table 3). All estimates were within 7% of the CLS estimates with the largest z-score being 2.02. Across all outcomes, there was an 18% reduction in the sum of the absolute raw differences and a 55% reduction in the sum of the absolute z-scores.

Model MRP 2 included age, sex, and race, along with the predictors used in all models. The means of the Bayesian R^2 posterior distributions ranged from 0.069 to 0.230 across the six outcome models (Supplementary Table 2). This model adjusted estimates closer to the CLS estimates in five of the six outcomes (Table 3) and produced the estimates closest to the CLS estimates in three of the six outcomes by both raw difference and z-score. This model reduced the sum of the absolute raw differences by 47% and the sum of the absolute z-scores by 69%, the greatest reduction in both measures across the three adjustment models.

Model MRP 3 included sex, race, education level, and the predictors used in all models. The means of the Bayesian R^2 posterior distributions ranged from 0.058 to 0.157 across the six outcome models (Supplementary Table 2). Compared with the unadjusted estimates, it produced estimates closer to the CLS values in four of the six outcomes by raw difference and five of the six outcomes by z-score (Table 3). The sum of absolute differences was reduced by 24% and the sum of absolute z-scores was reduced by 46%.

Figure 2 provides a visual display of all estimates. The only outcome for which the raw data provided the best estimate was asthma, where all four estimates were within 2% of the CLS estimate and the absolute values all z-scores were less than 1, suggesting very good fit across all estimates with all being easily within sampling variation. MRP 3 produced the smallest adjustments from the raw estimates and MRP 1 produced the largest adjustments, although MRP 1 over-corrected in some outcomes. The raw Facebook estimates farthest from the CLS estimates were arthritis and smoking, for which the Facebook confidence intervals had no overlap with the CLS intervals. In both cases, the MRP adjustments significantly reduced bias and resulted in credible intervals that overlapped with the CLS confidence interval, with the one exception of the MRP 3 arthritis estimate.

Across all MRP 2 results, the largest relative adjustment was to the Facebook diabetes estimate, reducing the absolute z-score from 2.76 to 0.07. The most interesting ZCTA-level plot displays the geographical distribution of physical inactivity across the metropolitan area (Fig. 3). The map shows the highest levels of physical inactivity in the western ZCTAs, which are the rural areas of the metropolitan area, as well as northern St. Louis County in the eastern part of the metropolitan area. The lowest levels of physical inactivity are in the east-central region, mostly making up more affluent areas of St. Louis County. There are two ZCTAs in the central and south central parts of the metropolitan area for which median income was not available in the ACS data and therefore prevalence

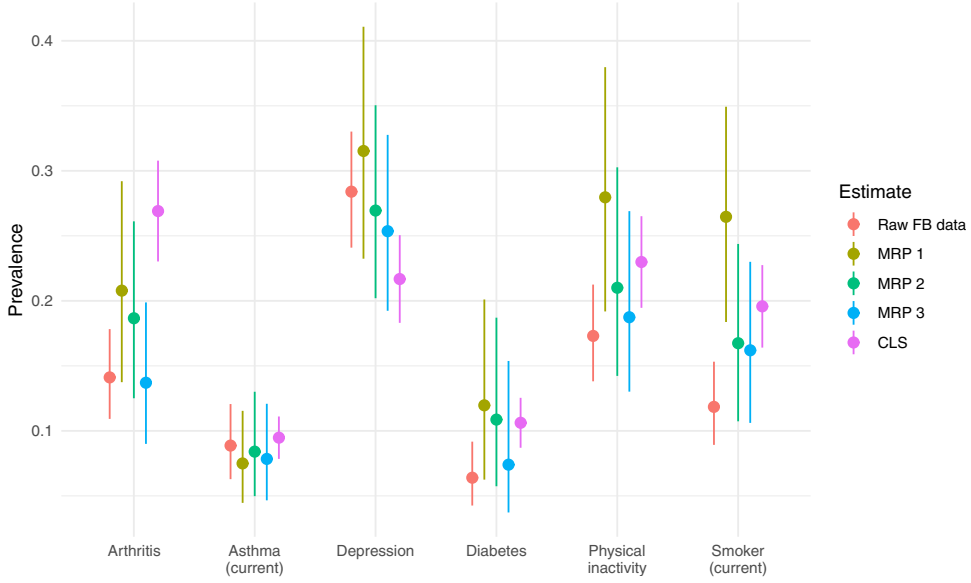


Fig. 2. Estimates of health outcomes with 95% credible intervals. FB: Facebook.

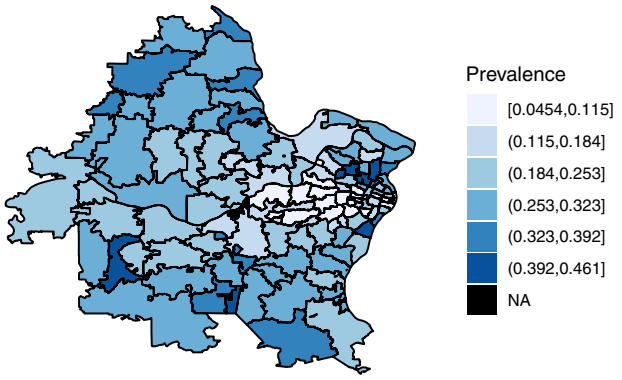


Fig. 3. ZCTA estimates of 30-day physical inactivity using Facebook data with MRP 2 model adjustment. NA: ZCTAs with missing population information could not be predicted.

cannot be predicted. A deeper investigation into the ZCTA-level adjustments made by MRP 2 is provided in the Supplementary Material, along with maps for all outcomes after adjustment by MRP 2.

The results from the sensitivity analysis follow those presented here and are available in the Supplementary Material.

Discussion

Our results suggest MRP can improve estimates of health behaviors and outcomes from self-selected online samples, although bias remains. MRP 2 provided estimates closest to the CLS estimates because of its combination of census variables. The variables included in an MRP model should be strongly associated with the likelihood of selection into the survey as well as the outcome of interest. Age and race have the two least representative distributions among the demographic data (county of residence was included in all MRP models), making them important factors to include in adjustment. In addition, the St. Louis region has substantial racial social and health disparities, suggesting race is doubly important. Thus, it should be no surprise that adjustment by age and race provided the estimates closest to the population-based CLS estimates.

There is much recent work to analyze selection bias using tools similar to causal inference methodologies [30]. Hernan et al used causal diagrams to formalize selection bias [31], which allows selection bias to be addressed using tools from that approach [32]. Others have explored probability-of-selection weights to generalize to a broader population [33,34]. Such approaches will directly produce a marginal estimate, whereas MRP produces conditional estimates which are then averaged to yield a marginal estimate. The availability of conditional estimates through MRP is beneficial both as estimates in their own right and as a model checking tool.

Choice of adjustment variables is extremely important for the quality of estimates MRP and other adjustment strategies produce [35]. In fields such as public opinion, very strong predictors of future action exist. For example, political party registration and recent voting history are strong predictors of future voting behavior [14,15]. In public health, such strong predictors are difficult to obtain, making MRP—or any alternative adjustment method—potentially less powerful in these cases. In addition, the lack of population data with local geographic identifiers (e.g., ZCTA) limits the number of variables that can be used for small area estimation. As we see in our results, the ability to adjust for only a few variables can potentially leave substantial bias remaining if there are multiple strong predictors of selection and outcome. To the extent that neighborhood characteristics predict individual level outcomes, some of these shortfalls can be addressed using aggregated data, such as ZCTA level education or income data, as we have done here. Still, the median and mean Bayesian R^2 across all models were less than 0.25, indicating that additional covariates could improve estimates.

Any comparison of estimates in this analysis relies on the quality of the CLS estimates to which all others are being compared. While the CLS is a population-based study following a rigorous methodology, its estimates have their own sampling variability and potential biases. Even with these caveats, the CLS likely provides the best estimates for the St. Louis metropolitan area. In addition, the relatively small sample size makes cross-model comparisons difficult. While the small samples did not keep MRP from improving point estimates, wide and overlapping credible intervals make it difficult to claim that certain models provide systematically better estimates.

Conclusions

MRP has been shown to be a valuable tool to adjust nonrepresentative survey estimates [13–17]. In the present study, we developed and compared MRP estimates using data from a sample of participants recruited through Facebook in a single metropolitan area. Previous public health studies using MRP [16,17] have estimated outcomes at larger geographical levels and have built models from large, relatively representative data sets (both greater than 10,000 respondents). Our study extends the investigation of MRP adjustment using a smaller, less-representative sample and estimating health behaviors and outcomes for small geographic areas. The results are positive, although there are clear indications that MRP is not a silver bullet for eliminating bias from highly self-selected samples in public health research. Careful application of MRP is a feasible strategy for reducing bias in local estimates when the source data come from a nonrandom, Internet-based sample, although lack of highly stratified population data at such a granular geographic level is a major limitation.

Acknowledgment

No grant funding or financial support was reported.

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.annepidem.2019.09.005.

References

[1] Center for Disease Control and Prevention. Comparability of data BRFSS 2013. Atlanta, GA: Center for Disease Control and Prevention; 2014.
[2] Center for Disease Control and Prevention. Behavioral risk factor surveillance system overview: BRFSS 2016. Atlanta, GA: Center for Disease Control and Prevention; 2017.
[3] European Commission. Statement by vice-president ansip and commissioner jourová ahead of the entry into application of the general data protection regulation. Brussels: European Commission; 2018.
[4] Kohut A, Keeter S, Doherty C, Dimock M, Directors A, Christian L, et al. Assessing the representativeness of public opinion surveys. Washington, DC: Pew Research Centeer for The People & The Press; 2012.
[5] Tourangeau R, Plewes TJ. Nonresponse in social science surveys: a research agenda. Washington, DC: The National Academies Press; 2013.
[6] Lim S, Immerwahr S, Lee S, Harris TG. Estimating nonresponse bias in a telephone-based health surveillance survey in New York city. Am J Epidemiol 2013;178:1337–41.

[7] Van Gelder M, Pijpe A. E-epidemiology: a comprehensive update. Cancer 2013;15:1–7.
[8] Graham AL, Milner P, Saul JE, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. J Med Internet Res 2008;10:e50.
[9] Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14:e20.
[10] McCabe SE, Boyd CJ, Couper MP, Crawford S, d'Arcy H. Mode effects for collecting alcohol and other drug use data: web and US mail. J Stud Alcohol 2002;63:755–61.
[11] Rhodes SD. Collecting behavioural data using the world wide web: considerations for researchers. J Epidemiol Community Health 2003;57:68–73.
[12] Gelman A, Little TC. Poststratification into many categories using hierarchical logistic regression. Surv Methodol 1997;23:127–35.
[13] Park DK, Gelman A, Bafumi J. Bayesian multilevel estimation with post-stratification: state-level estimates from national polls. Polit Anal 2004;12:375–85.
[14] Ghitza Y, Gelman A. Deep interactions with MRP: election turnout and voting patterns among small electoral subgroups. Am J Pol Sci 2013;57:762–76.
[15] Wang W, Rothschild D, Goel S, Gelman A. Forecasting elections with non-representative polls. Int J Forecast 2014;3:980–91.
[16] Zhang X, Holt JB, Yun S, Lu H, Greenlund KJ, Croft JB. Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system. Am J Epidemiol 2015;182:127–37.
[17] Downes M, Gurrin LC, English DR, Pirkis J, Currier D, Spittal MJ, et al. Multilevel regression and poststratification: a modelling approach to estimating population quantities from highly selected survey samples. Am J Epidemiol 2018;187:1780–90.
[18] Missouri Department of Health and Senior S. Missouri County-Level Study. Missouri Department of Health and Senior Services; 2017.
[19] Diez Roux AV. Investigating neighborhood and area effects on health. Am J Public Health 2001;91:1783–9.
[20] U.S. Census Bureau. Zip code tabulation areas. ZCTAs; 2018. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html. [Accessed 13 August 2019].
[21] Manson S, Schroeder J, Van Ripper D, Ruggles S. IPUMS National Historical Geographic Information System: Version 12.0 [Database]. Minneapolis: University of Minnesota; 2017.
[22] Cramer H. Chapter 21. The two-dimensional case. In: Mathematical Methods of Statistics. Princeton University Press; 1946.
[23] White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30:377–99.
[24] Gelman A, Goodrich B, Gabry J, Vehtari A. R-squared for Bayesian regression models. Am Stat 2019;73:307–9.
[25] R Core Team. R. A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
[26] Kowarik A, Templ M. Imputation with the R Package VIM. J Stat Softw 2016;74:1–16.
[27] Buuren S van, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45:1–67.
[28] Stan Development Team. Rstanarm: Bayesian applied regression modeling via Stan. 2016. http://mc-stan.org/. [Accessed 17 July 2018].
[29] Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6:e34.
[30] Mercer AW, Kreuter F, Keeter S, Stuart EA. Theory and practice in non-probability surveys: parallels between causal inference and survey inference. Public Opin Q 2017;81:250–71.
[31] Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15:615–25.
[32] Bareinboim E, Pearl J. Causal inference and the data-fusion problem. Proc Natl Acad Sci 2016;113:7345–52.
[33] Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. Am J Epidemiol 2010;172:107–15.
[34] Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. J R Stat Soc Ser A 2011;174:369–86.
[35] Pew Research Center. For weighting online opt-in samples, what matters most? Washington, DC: Pew Research Center; 2018.

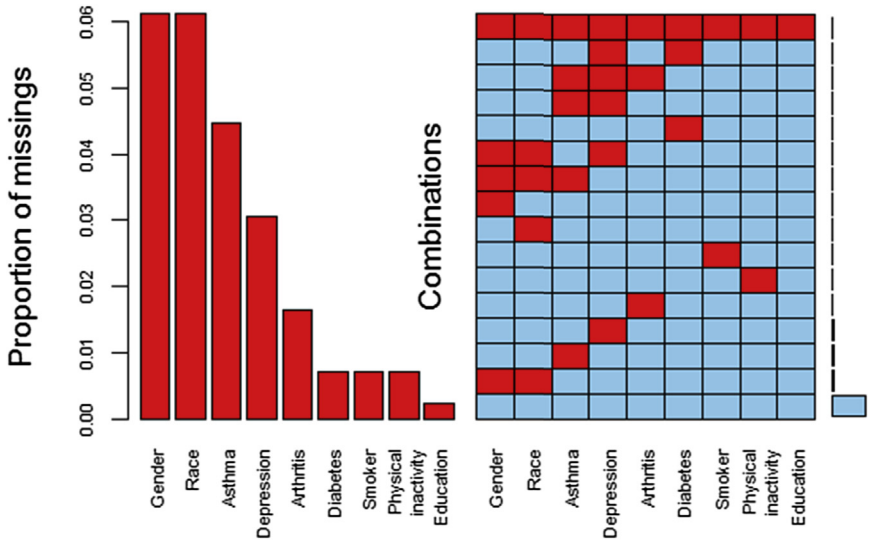
Appendix

Using multilevel regression with poststratification to obtain regional health estimates from a Facebook-recruited sample: Supplementary material

Exploration of missing data

Supplementary Figure 1 displays the proportions and patterns of missing data. Gender and race/ethnicity had the most missing data at about 6% a piece, and were often missing together—both were missing in 26 cases, with 24 respondents missing both responses. All other variables had less than 5% missing, with physician-diagnosed asthma, depression, and arthritis being the only other variables with more than 1% missing.

Complete cases were compared with respondents with any missing data via χ^2 tests. Complete cases were significantly different from cases with missing data in most of the assessed variables. To highlight some differences, complete cases tended to be younger than those with missing data, more likely to be residents of St. Louis County compared with St. Louis City, and living in slightly less-educated ZIP codes. Counts and percents within each group and test results for all variables are provided in Supplementary Table 1.



Supplementary Figure 1. Missing data counts and patterns. Variables zip code, age, county of residence, percentage white in zip code, percentage with bachelor's degree in zip code, and median income within zip code were completely observed and excluded from this figure.

Supplementary modeling and MRP results

Supplementary Table 1
Comparison of complete records to records with any missing data

Characteristic	Incomplete cases	Complete cases	*P*
	n = 115	n = 310	
Male (%)	44 (49.4)	113 (36.5)	.037
Age (%)			<.001
18–24	68 (59.1)	82 (26.5)	
25–34	27 (23.5)	93 (30.0)	
35–44	3 (2.6)	33 (10.6)	
Over 44	17 (14.8)	102 (32.9)	
Race/ethnicity (%)			.302
White non-Hispanic	54 (60.7)	207 (66.8)	
Asian non-Hispanic	19 (21.3)	39 (12.6)	
Black non-Hispanic	9 (10.1)	42 (13.5)	
Hispanic	3 (3.4)	11 (3.5)	
Other	4 (4.5)	11 (3.5)	
Education (%)			.047
Less than high school	2 (1.8)	6 (1.9)	
High school	9 (7.9)	19 (6.1)	
Some college	23 (20.2)	59 (19.0)	
Associate's degree	1 (0.9)	27 (8.7)	
Bachelor's degree	58 (50.9)	126 (40.6)	
Graduate/professional degree	21 (18.4)	73 (23.5)	
Income (%)			.145
Less than $15,000	8 (14.3)	30 (9.7)	
$15,000-$25,000	4 (7.1)	44 (14.2)	
$25,000k-$35,000	12 (21.4)	38 (12.3)	
$35,000-$50,000	9 (16.1)	62 (20.0)	
$50,000-$75,000	2 (3.6)	36 (11.6)	
$75,000-$100,000	12 (21.4)	53 (17.1)	
More than $100,000	9 (16.1)	47 (15.2)	
County (%)			.009
St. Louis City	55 (47.8)	112 (36.1)	
St. Louis County	46 (40.0)	175 (56.5)	
Other	14 (12.2)	23 (7.4)	
Zip percent white (mean [SD])	60.86 (22.76)	56.47 (26.26)	.114
Zip percent with bachelor's degree (mean [SD])	41.61 (17.21)	35.29 (17.55)	.001
Zip median income (mean [SD])	51,801.64 (18,925.13)	49,419.94 (17,469.97)	.224
Depression	21 (20.6)	96 (31.0)	.059
Diabetes	6 (5.4)	21 (6.8)	.764
Asthma (current)	4 (4.2)	32 (10.3)	.099
Arthritis	7 (6.5)	52 (16.8)	.013
Smoker (current)	10 (8.9)	40 (12.9)	.345
Physical inactivity	13 (11.6)	60 (19.4)	.087

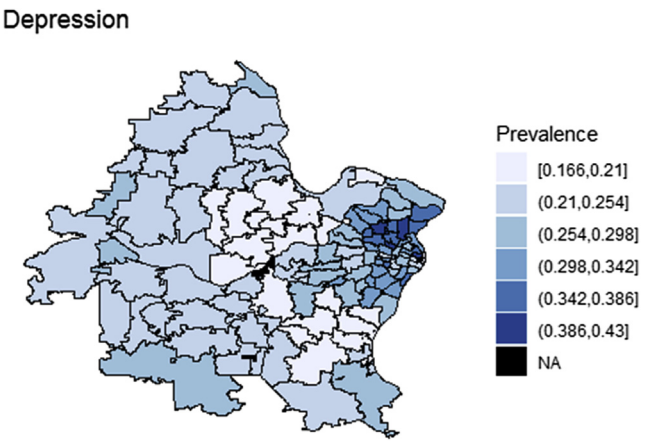
Supplementary Table 2
Bayesian R^2 and WAIC for each outcome model

Outcome	Measure	MRP 1	MRP 2	MRP 3
Arthritis	Bayesian R^2	0.234/0.234/0.023	0.23/0.229/0.025	0.157/0.157/0.026
	WAIC	256.65/26.24	252.8/26.65	301.46/27.92
Asthma	Bayesian R^2	0.077/0.08/0.025	0.069/0.072/0.023	0.058/0.061/0.022
	WAIC	272.04/29.27	275.37/29.87	280.85/30.06
Depression	Bayesian R^2	0.098/0.099/0.02	0.102/0.102/0.019	0.098/0.099/0.018
	WAIC	490.29/21.65	478.64/21.7	480.57/22.03
Diabetes	Bayesian R^2	0.172/0.173/0.037	0.195/0.195/0.038	0.153/0.153/0.044
	WAIC	180.14/26.26	170/25.73	199.51/28.1
Physical inactivity	Bayesian R^2	0.152/0.152/0.026	0.135/0.135/0.027	0.125/0.125/0.027
	WAIC	365.19/26.56	370.89/26.49	376.94/26.96
Smoke (current)	Bayesian R^2	0.172/0.172/0.027	0.119/0.12/0.026	0.124/0.125/0.026
	WAIC	270.43/26.4	293.76/27.3	286.58/27.49

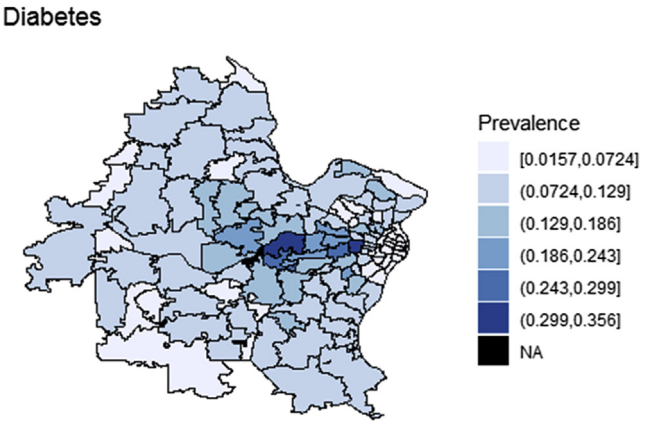
Values reported as median/mean/standard deviation and mean/standard deviation of posterior distribution, respectively.

Maps of estimated zip code prevalence (model MRP 2 only)

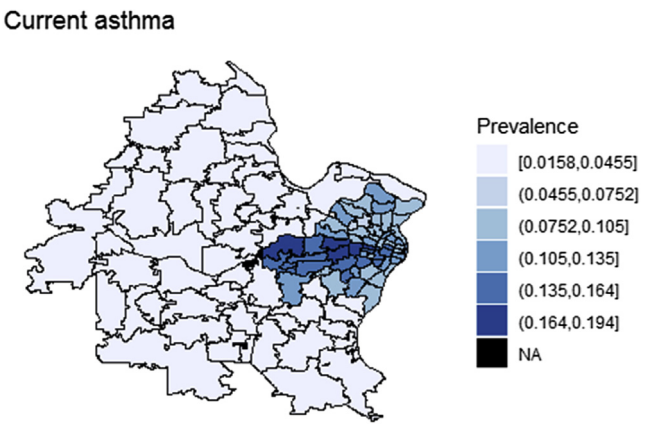
Supplementary Figures 2–7 provide maps of MRP estimates of ZCTA-level prevalence for all health behaviors and outcomes assessed.



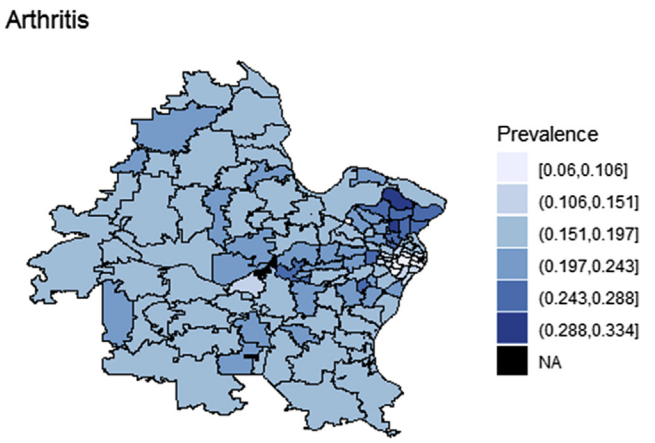
Supplementary Figure 2. Zip code estimates of diagnosed depression using Facebook data with MRP 2 model adjustment.



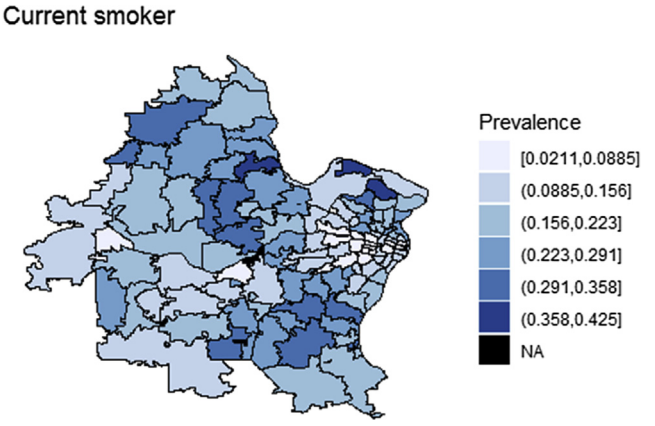
Supplementary Figure 3. Zip code estimates of diagnosed diabetes using Facebook data with MRP 2 model adjustment.



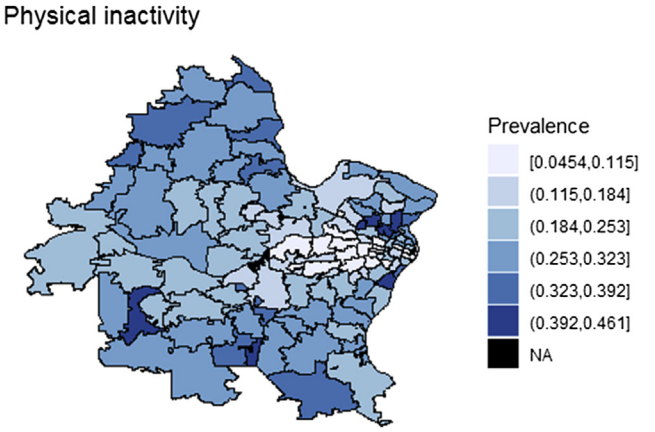
Supplementary Figure 4. Zip code estimates of current asthma diagnosis using Facebook data with MRP 2 model adjustment.



Supplementary Figure 5. Zip code estimates of diagnosed arthritis using Facebook data with MRP 2 model adjustment.



Supplementary Figure 6. Zip code estimates of current smoking prevalence using Facebook data with MRP 2 model adjustment.



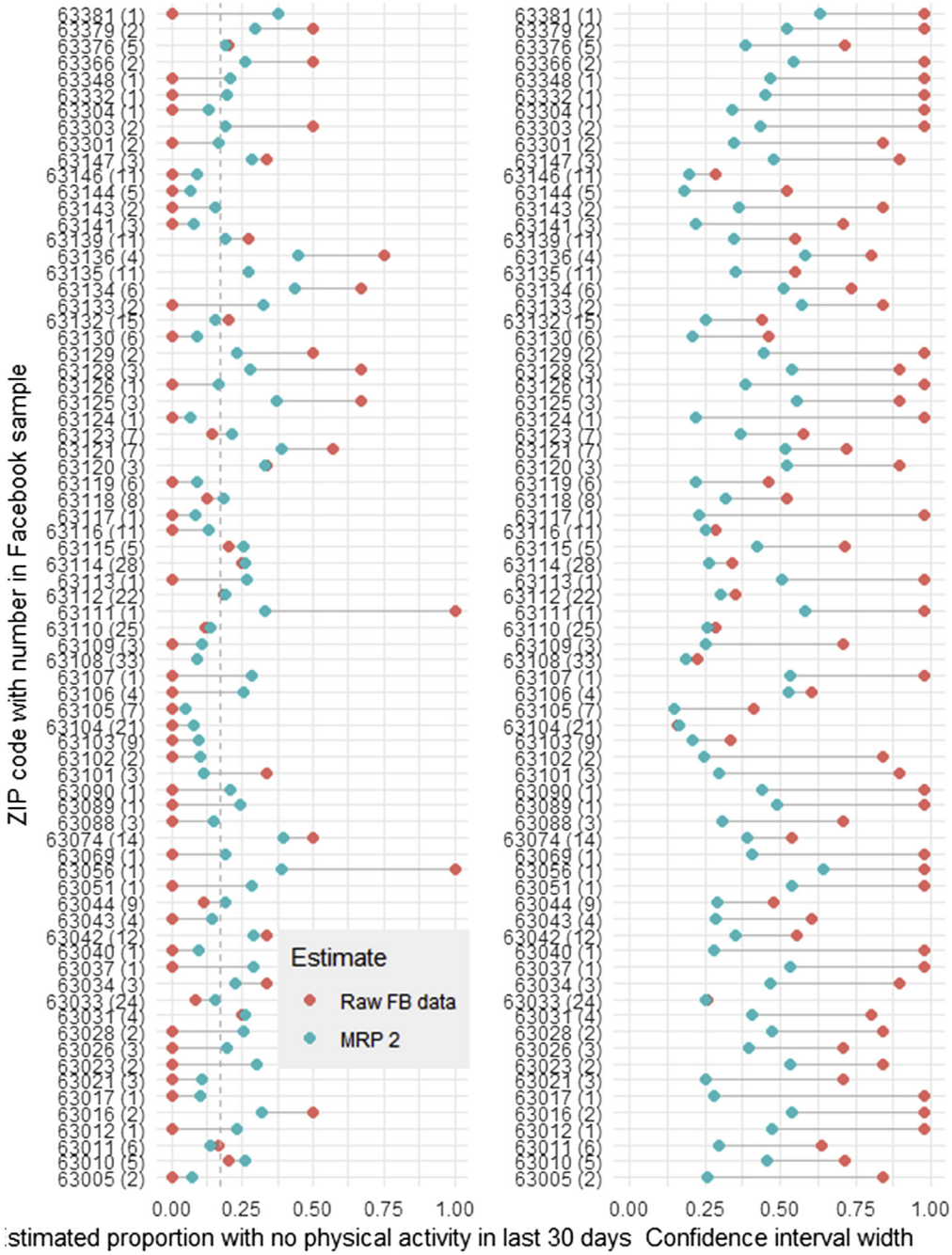
Supplementary Figure 7. Zip code estimates of 30-day physical inactivity using Facebook data with MRP 2 model adjustment. Same as article Fig. 2.

Investigation of ZIP code level estimates

Looking further into ZIP code level estimates of physical inactivity in the 30 days preceding the survey, we see that in nearly all cases MRP adjusts the raw point estimates closer to the overall sample mean (Supplementary Fig. 8, left panel). As expected, this is especially true if raw point estimates are extreme or are from small samples.

Confidence interval width is also shrunk dramatically by as much as 0.50 (Supplementary Fig. 8, right panel), although none of the intervals suggest a high level of accuracy, with the most narrow still being 0.15 wide (i.e., a margin of error of ±0.075). This is in contrast to the metropolitan area estimates, for which MRP increased the confidence interval widths. This distinction may be due to two forces pulling in opposite directions: MRP partially pools, or "borrows" information from adjacent areas to increase precision from small samples, but also increases the width of the confidence/credible intervals to capture additional uncertainty due to adjustment for self-selection which is not present in the unadjusted estimates. The presence of missing data may also play some role in the widening of intervals.

A larger sample size would likely shrink the widths of the adjusted intervals further. While an overall sample size of 425 may not yield highly precise estimates, the adjustments we see indicate that MRP can not only help to stabilize point estimates, but also shrink confidence intervals so that large samples within each ZIP code may not be necessary to provide highly precise estimates.



Supplementary Figure 8. Comparison of point estimates and confidence interval widths using raw Facebook data and MRP 2 adjustment. Vertical dashed line in left panel represents overall sample proportion.

Sensitivity analysis

A sensitivity analysis was performed in which all coefficients were given a normal prior with mean 0 and standard deviation 1, as opposed to an SD of 2.5 in the primary analysis. This change would potentially affect the results by shrinking coefficients closer to 0. This led to MRP estimates with less adjustment than in the primary analysis, although ultimately the results are similar to those presented in the main text. Supplementary Table 3 provides model fit statistics while Supplementary Table 4 and Supplementary Figure 9 show outcome estimates.

Supplementary Table 3
Sensitivity analysis

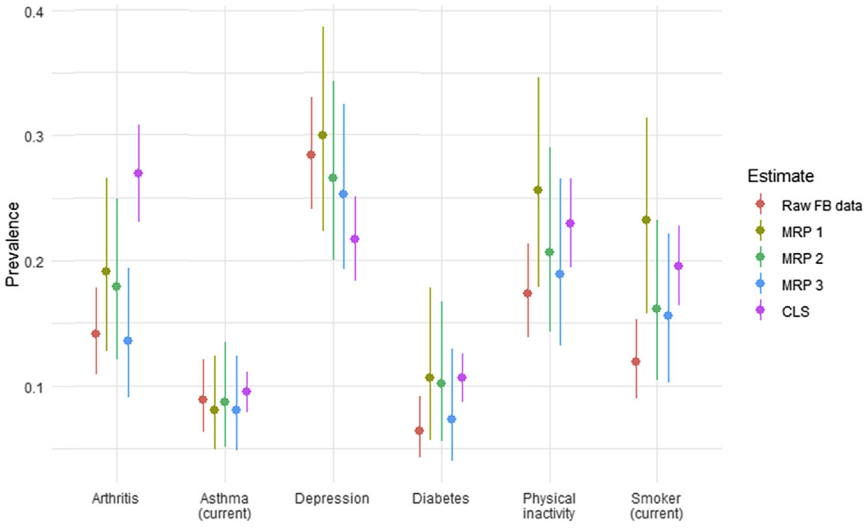
Outcome	Measure	MRP 1	MRP 2	MRP 3
Arthritis	Bayesian R^2	0.208/0.207/0.024	0.207/0.206/0.025	0.142/0.142/0.025
	WAIC	251.24/24.17	248.14/24.41	296.89/26.55
Asthma	Bayesian R^2	0.059/0.062/0.022	0.054/0.057/0.021	0.046/0.049/0.02
	WAIC	269.97/28.46	271.84/28.99	277.55/29.38
Depression	Bayesian R^2	0.086/0.086/0.019	0.09/0.091/0.018	0.088/0.089/0.018
	WAIC	487.66/20.76	476.86/20.86	478.9/21.17
Diabetes	Bayesian R^2	0.123/0.126/0.033	0.132/0.134/0.033	0.098/0.102/0.036
	WAIC	176.17/24.35	168.59/23.36	195.43/27.03
Physical inactivity	Bayesian R^2	0.129/0.129/0.026	0.117/0.118/0.026	0.11/0.111/0.026
	WAIC	363.26/25.71	367.83/25.64	373.57/26.26
Smoke (current)	Bayesian R^2	0.13/0.13/0.025	0.096/0.097/0.024	0.098/0.1/0.023
	WAIC	271.31/25.27	290.44/26.26	284.46/26.28

Bayesian R^2 and WAIC for each outcome model. Values reported as median/mean/standard deviation and mean/standard deviation of posterior distribution, respectively.

Supplementary Table 4
Sensitivity analysis

Outcome	Raw FB data			MRP 1			MRP 2			MRP 3			CLS
	Est	d	z	Est	d	z	Est	d	z	Est	d	z	Est
Arthritis	0.141	−0.128	−4.96	0.190	**−0.079**	**−1.94**	0.179	−0.09	−2.36	0.135	−0.134	−4.08	0.269
Asthma (current)	0.089	**−0.006**	**−0.37**	0.080	−0.015	−0.69	0.086	−0.008	−0.36	0.080	−0.015	−0.69	0.095
Depression	0.284	0.067	2.41	0.299	0.082	1.82	0.265	0.048	1.2	0.253	**0.036**	**0.95**	0.217
Diabetes	0.064	−0.042	−2.76	0.106	**0**	**−0.01**	0.102	−0.004	−0.15	0.073	−0.034	−1.3	0.106
Physical inactivity	0.173	−0.057	−2.23	0.256	0.026	0.56	0.207	**−0.023**	**−0.55**	0.189	−0.041	−1.07	0.230
Smoker (current)	0.118	−0.077	−3.46	0.232	0.036	0.83	0.161	**−0.035**	**−0.95**	0.156	−0.04	−1.16	0.196
Abs sum	NA	0.378	16.19	NA	0.238	5.85	NA	**0.209**	**5.57**	NA	0.299	9.26	NA

Point estimates from unadjusted Facebook data and MRP-adjusted results with raw differences (d) 2-sample z-scores (z) comparing estimate to Missouri County-Level Survey (CLS) estimate. The final row gives the sum of the absolute differences across health outcome models. See Table 1 for the set of predictors used in each MRP model. Smallest d and z values for each outcome are in boldface font.



Supplementary Figure 9. Sensitivity analysis. Estimates of health outcomes with 95% credible intervals. FB: Facebook; MRP: Multilevel regression with poststratification; CLS: Missouri County-Level Study.