Research Letter

Sequential Surveying Within Existing Studies to Advance Research Expeditiously

Karilynn M Rockhill[1,2], MPH, PhD; Alison G Abraham[1], MS, MHS, PhD; Joshua C Black[2], MS, PhD

[1]Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[2]Rocky Mountain Poison & Drug Safety, Denver Health Medical Center, Denver, CO, United States

Corresponding Author:

Karilynn M Rockhill, MPH, PhD
Rocky Mountain Poison & Drug Safety, Denver Health Medical Center
777 Bannock Street, M/C 0180
Denver, CO 80204
United States
Phone: 1 3016413317
Email: karilynn.rockhill@rmpds.org

Abstract

This study outlines an approach through sampling design and post hoc weighting adjustments to execute a polysubstance use–focused survey with applied bias reduction techniques at lower cost.

J Med Internet Res 2026;28:e95300; doi: 10.2196/95300

Keywords: survey methodology; polysubstance use; bias; sampling; surveillance systems; online surveys

Introduction

Opt-in online panel-based surveys have become a mainstay in epidemiology since they can generate large sample sizes at reduced costs and increased speed relative to probability-based surveys [1]. The need for quick, cost-effective surveys was palpable during the COVID-19 pandemic and applicable to other quickly evolving topics [2]. Opt-in panels offer expedited data collection but can have limited generalizability [3]. Through creative sampling and refined analytical methods, their representation can be improved [4]. We explored the utility of specialized surveys administered sequentially with data linkage to answer new research questions in an expedited, low-cost manner. We developed a custom polysubstance use–focused sequential survey with design-based bias-corrected estimates reflecting the overall population.

Methods

Data Source and Sampling

We used the third-quarter 2024 wave of an ongoing cross-sectional study, the Survey of Non-Medical Use of Prescription Drugs (NMURx), to invite participants to a polysubstance survey, an important topic for the fourth wave of the overdose epidemic [5]. Both were self-administered, confidential online surveys where answers were linked through individual identifiers; therefore, questions were not repeated (eg, demographics), creating space for new content. NMURx recruits from opt-in commercial panels but produces externally valid US representative drug estimates using individual calibration weights by raking across state, demographics, and health metrics ($w_{c,i}$) [6].

The sequential invite was sent within 2 weeks of NMURx and was a retrospective assessment of lifetime polysubstance use. We used stratified random sampling of NMURx participants conditional on 3 domains of last-year drug use: no drug use (0.15 sampling rate), single substance use (0.4), and polysubstance use (0.6) [7]. This safeguarded the sample, ensuring diversity in drug use trajectories while maximizing participation among a polysubstance group.

Sampling Weights

The stratified sampling and subsequent nonresponse introduced selection effects. Individual adjustment weights ($w_{a,i}$) were designed so sequential survey respondents (herein called "respondents") would have similar distributions to the sampling frame (NMURx). The following adjustment weights were calculated as:

$$w_{a,i} = D_i * \frac{1}{s_{d,i}} * f(NR)_i, where\ f(NR)_i = \frac{1}{P(R_i = 1 \mid \mathbf{X_i})}$$

where D_i was the known domain proportion, $s_{d,i}$ was the known sampling fraction by domain, and $f(NR)_i$ was modeled to account for differential nonresponse, described below. The final applied weight was generated to bias-correct all selection effects as the product of the second survey adjustment weight $(w_{a,i})$ and the respondent's NMURx calibration weight $(w_{c,i})$ [6].

Modeling Nonresponse

Respondents and nonrespondents both completed NMURx, providing rich covariate information for models including demographics (age, sex, region, language, race, ethnicity, income, education, insurance), physical and mental health (self-assessed health, limited in daily activities, mental health diagnosis, current depression/anxiety/well-being validated scores), drug use (tobacco, Mahalanobis distance of drug use), and metadata (device type, survey timing). We tested 8 covariate combinations in a propensity score modeling response $(P(R_i = 1 \mid \boldsymbol{X_i}))$ informing the term $f(NR)_i$. The selected model minimized the average Target Absolute Standardized Mean Difference (TASMD) across demographics, health, mental health, and drug use based on 2 equations adapted from the literature [8,9] for continuous and categorical variables, respectively:

$$\text{TASMD} = \frac{\left|\overline{X}_{wg} - \overline{X}^*\right|}{\sqrt{S_g^2}}, \frac{\left|\hat{p}_{wg} - \hat{p}^*\right|}{\sqrt{\hat{p}_g(1-\hat{p}_g)}}$$

where $\bar{X}_{wg}$ and $\hat{p}_{wg}$ and S_g^2 and $\hat{p}_g(1-\hat{p}_g)$ were the weighted means and unweighted variances of the respondent group and $\bar{X}^*$ and $\hat{p}^*$ were the means of the invited (target group being emulated). The "minimally corrected" model included only variables already known to correct for selection effects in NMURx, which was compared to more comprehensive models designed for nonresponse. The relative change was compared to the unweighted TASMD.

Final Estimates

Adjusted (fully weighted) and unadjusted (unweighted) percentage of lifetime substance use (out of entire sample) and proportion of initiation before 18 years of age (out of lifetime use) were calculated.

Ethical Considerations

The Colorado Multiple Institution Review Board (number 16-0922) approved this study. Data have a Certificate of Confidentiality and are deidentified. Participants were consented and compensated for each survey through redeemable points (approximately US $4-6).

Results

There were 29,040 NMURx participants, with no drug use reported by 42.8% (n=12,434), single drug use by 27.5% (n=7998), and polysubstance use by 29.6% (n=8608). After stratified random sampling across domains, we observed similar response rates (responses/invitations): 36.5% (546/1494), 35.4% (932/2634), and 34.6% (1459/4220), respectively. There were 15 respondents who were excluded for careless responses. The median survey time was 19.65 minutes. The sequential survey cost 18% of the NMURx launch total while linking answers across approximately 35 minutes of total survey time.

The unweighted data performed below a common balancing threshold of <0.1 (Figure 1) [9]. For the fully corrected model, there was distribution overlap of responders and nonresponders in propensity score (>99.9%) with no extreme weights. The model reduced the average TASMD by 70.9% for demographics, 26.2% for drug use, 93.7% for health metrics, and 73.7% for mental health metrics, indicating movement toward the target population averages, outperforming the minimally corrected model.

The unweighted estimates and the fully bias-adjusted weighted estimates for lifetime drug use showed dramatic movement in the point estimate, as expected given the differential domain sampling (Figure 2). The subgroup analysis for initiation <18 years old showed a less dramatic shift in estimates after weighting.

Figure 1. Reduction of nonresponse bias through propensity score weighting. There was minimal nonresponse bias among sequential survey respondents compared to the invited sample (target). The fully corrected model (nonresponse designed) outperformed the minimally corrected (selection bias designed) model and lowered the target absolute standardized mean difference (TASMD) across almost all metrics. The fully corrected propensity score model included the following covariates: age, sex, region, race, ethnicity, income, education, limited activities, general health, mental health diagnosis, tobacco smoking, Mahalanobis distance of drug use, and survey taking device type. AI/AN: American Indian/Alaska Native; DAST-10: Drug Abuse Screening Test-10; Dx: diagnosis; GAD-7: Generalized Anxiety Disorder-7; HS: high school; PHQ-9: Patient Health Questionnaire-9; PI: Pacific Islander; SUD: Substance Use Disorder; S-WEMWBS: Short Warwick-Edinburgh Mental Health Wellbeing Scale.

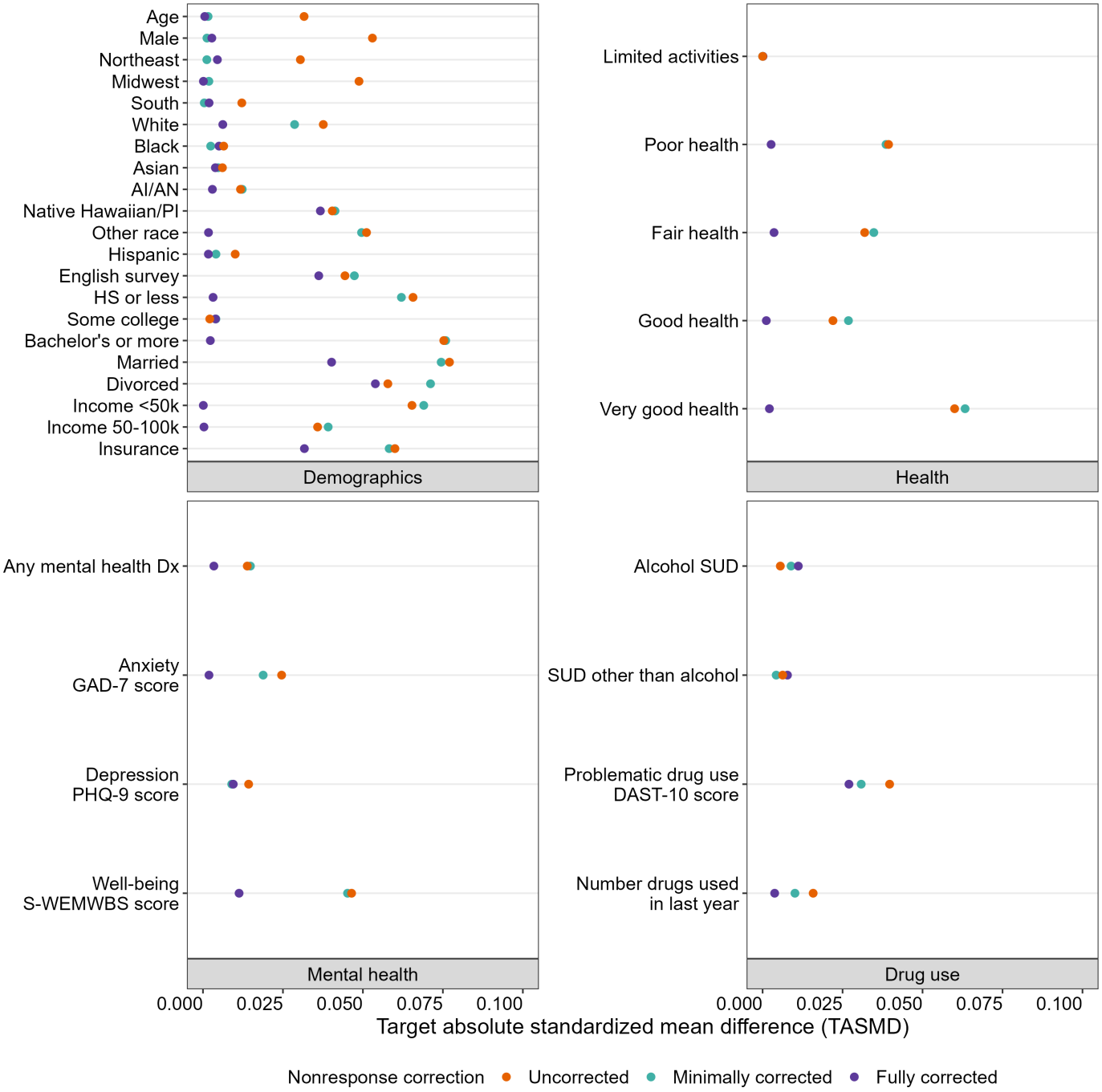
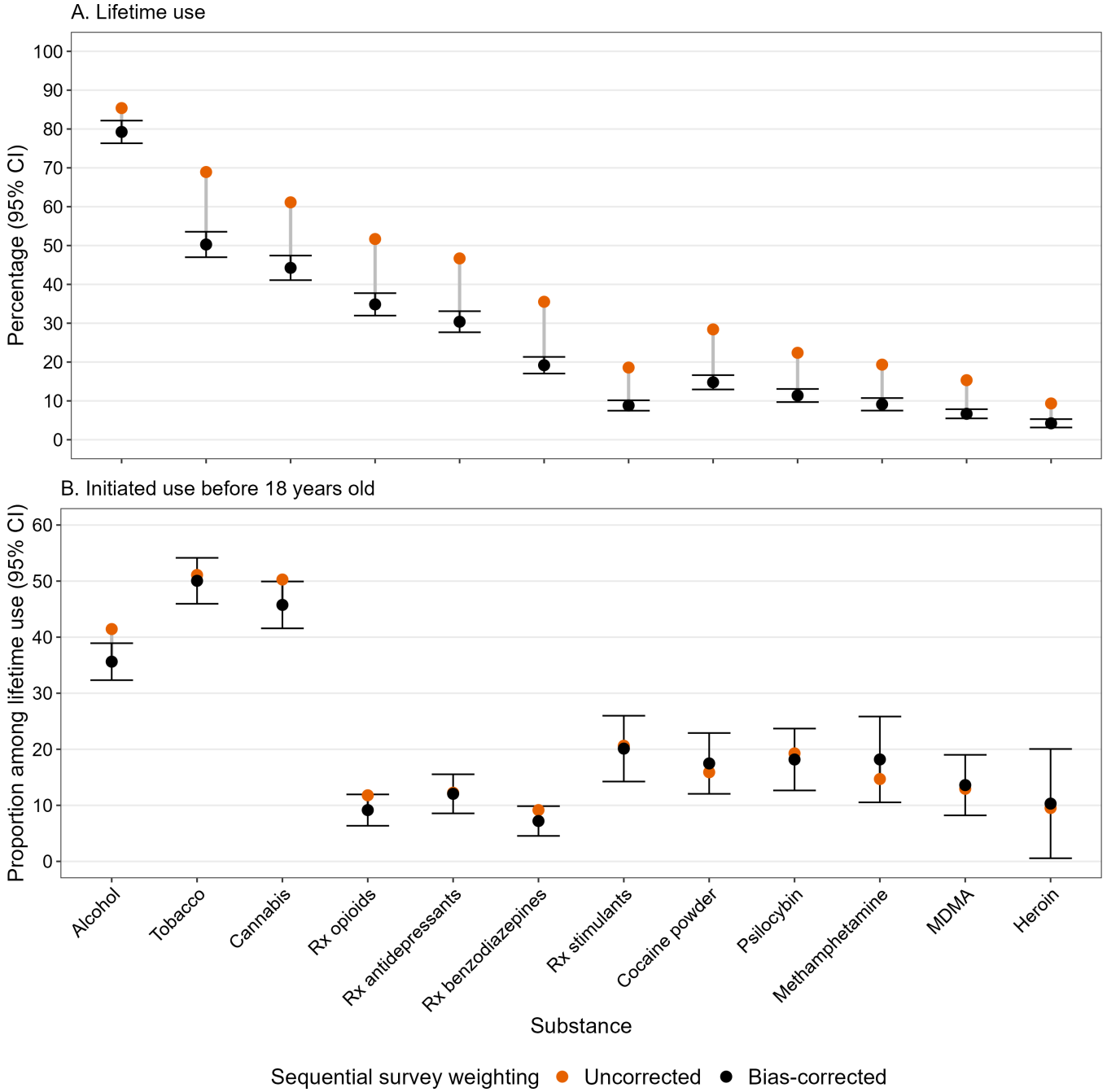


Figure 2. Comparison of final bias-corrected estimates for drug use behaviors. After applying the full weighting scheme, among the polysubstance use survey sample, the estimates for lifetime use (A) had a more dramatic shift in estimates as compared to the subgroup analysis among those who report lifetime use of each substance for initiation of use before 18 years of age (B). MDMA: 3,4-Methylenedioxymethamphetamine; Rx: prescription.



Discussion

We successfully collected enriched data on polysubstance use through sequential surveying at reduced cost, generating design-based corrections at multiple stages. We improved estimates of population parameters beyond those already accounted for by adjusting for differences in nonresponse through a modeling approach. Careful consideration of independence between nonresponse and the sampling domains is necessary as this may not hold true for all samples as we assumed here. Though low response rates for sequential surveys would also be expected, our findings confirmed nonresponse bias was not driven by low response rates [10]. One limitation is that the retrospective recall of lifetime drug use may have residual measurement error (eg, recall or social desirability bias).

We demonstrated efficiency in primary data collection by supplementing a well-established, robust surveillance system to produce novel, timely, and adjusted estimates at a fraction of the cost. This approach could be adapted to partner with studies of any type with well-understood sampling frames as the foundation for online, expedited survey tools where multiple research questions could be quickly examined. Overall, this approach overcame potential limitations to

inference and shows an applied method to address research questions in an expedited and bias-reduced manner.

Acknowledgments

We would like to acknowledge Lori Crane (Colorado School of Public Health, Department of Community & Behavioral Health) on her survey expertise and contributions to the follow-up survey questionnaire and Evelyn Fox (Rocky Mountain Poison & Drug Safety) for the management of the execution of the follow-up survey.

Funding

Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health under award number R36DA057413. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to certificate of confidentiality and privacy protections of data around sensitive topics such as drug use. These data are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: KMR, AGA, JCB
Funding acquisition: KMR, AGA
Methodology: KMR, JCB
Data curation: KMR
Formal analysis: KMR
Validation: JCB
Writing – original draft: KMR
Writing – review & editing: AGA, JCB

Conflicts of Interest

None declared.

References

1. Sammut R, Griscti O, Norman IJ. Strategies to improve response rates to web surveys: a literature review. Int J Nurs Stud. Nov 2021;123:104058. [doi: 10.1016/j.ijnurstu.2021.104058] [Medline: 34454334]
2. De Man J, Campbell L, Tabana H, Wouters E. The pandemic of online research in times of COVID-19. BMJ Open. Feb 23, 2021;11(2):e043866. [doi: 10.1136/bmjopen-2020-043866] [Medline: 33622948]
3. Si Y, Wagner JR, Kessler RC. Probability, probability-based, and nonprobability surveys in psychiatric epidemiological research. JAMA Psychiatry. Mar 1, 2026;83(3):298-305. [doi: 10.1001/jamapsychiatry.2025.3652] [Medline: 41335451]
4. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. Epidemiology. Jul 2017;28(4):553-561. [doi: 10.1097/EDE.0000000000000664] [Medline: 28346267]
5. Jenkins RA. The fourth wave of the US opioid epidemic and its implications for the rural US: a federal perspective. Prev Med. Nov 2021;152(Pt 2):106541. [doi: 10.1016/j.ypmed.2021.106541] [Medline: 34482994]
6. Black JC, Rockhill K, Forber A, et al. An online survey for pharmacoepidemiological investigation (Survey of Non-Medical Use of Prescription Drugs Program): validation study. J Med Internet Res. Oct 25, 2019;21(10). [doi: 10.2196/15830] [Medline: 31654568]
7. Rockhill KM, Black JC, Iwanicki J, Abraham A. Polysubstance use profiles among the general adult population, United States, 2022. Am J Public Health. May 2025;115(5):747-757. [doi: 10.2105/AJPH.2024.307979] [Medline: 40112266]
8. Chattopadhyay A, Hase CH, Zubizarreta JR. Balancing vs modeling approaches to weighting in practice. Stat Med. Oct 30, 2020;39(24):3227-3254. [doi: 10.1002/sim.8659] [Medline: 32882755]
9. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. Nov 10, 2009;28(25):3083-3107. [doi: 10.1002/sim.3697] [Medline: 19757444]
10. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: a meta-analysis. Public Opin Q. 2008;72(2):167-189. [doi: 10.1093/poq/nfn011]

Abbreviations

NMURx: Survey of Non-Medical Use of Prescription Drugs

TASMD: target absolute standardized mean difference

Edited by Amaryllis Mavragani; peer-reviewed by Brian Ward, Katherine Irimata; submitted 13.Mar.2026; initial comments to author 09.Sep.2026; final revised version received 23.Oct.2026; accepted 01.Nov.2026; published 20.Nov.2026

Please cite as:
Rockhill KM, Abraham AG, Black JC
Sequential Surveying Within Existing Studies to Advance Research Expeditiously
J Med Internet Res 2026;28:e95300
URL: https://www.jmir.org/2026/1/e95300
doi: 10.2196/95300

© Karilynn M Rockhill, Alison G Abraham, Joshua C Black. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.