Comparative Evaluation of Reweighting Methods to Address Selection Bias in the All of Us Research Program

Authors:

Elizabeth Mitchell, PhD Candidate[*,†]; Jieun Lee, PhD; Zhiyan Li, MPH[*]; Piaopiao Li, MS[*,¶]; Qiaochu Xue[*]; Carl Yang, PhD[*,§,||]; Jithin Varghese, PhD[*,‡]; Young-Rock Hong, MPH, PhD[†,¶]; Ilana Graetz, PhD [†]; Hui Shao, MD, PhD[*,†,‡,¶]

Author Affiliations:

[*] Emory Global Diabetes Research Center, Woodruff Health Sciences Center and Emory University; Atlanta, Georgia

[†] Department of Health Policy and Management, Rollins School of Public Health, Emory University; Atlanta, Georgia

[‡] Hubert Department of Global Health, Rollins School of Public Health, Emory University; Atlanta, Georgia

[§] Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University; Atlanta, Georgia

[||] Department of Computer Science, Emory University; Atlanta, Georgia

[¶] Department of Family and Preventive Medicine, School of Medicine, Emory University; Atlanta, Georgia

Abstract

Background:

The All of Us Research Program (AoU) is a large-scale health research cohort, but volunteer-based recruitment limits population-level inference. Evidence comparing methods to address its sampling bias is limited.

Objective:

To evaluate reweighting methods for approximating national population estimates from the AoU cohort.

Design:

Comparative evaluation of inverse probability weighting (IPW) and clustering reweighting methods for the AoU cohort using the National Health and Nutrition Examination Survey (NHANES) as a benchmark.

Setting:

Secondary data analysis of national survey and cohort data.

Participants:

Adults from NHANES (n=8,947) and AoU (n=296,617), 2017-2020.

Measurements:

Bias reduction assessed by covariate balance (standardized mean difference [SMD]), weight distribution stability, and effective sample size (ESS). Secondary validation assessed balance in covariates excluded from the reweighting algorithms and within an analytic subset.

Results:

All methods reduced bias relative to unweighted AoU (mean SMD=0.14). IPW with raking nearly eliminated imbalances (mean SMD<0.001), whereas IPW without post-calibration and IPW with post-stratification achieved substantial bias reduction (mean SMD=0.03) with larger ESS. Clustering methods showed more modest reduction (mean SMD=0.08 direct covariate, 0.09 graph-based) but produced the most stable weights and largest ESS. In validation, probability-based methods outperformed clustering methods on bias reduction across sociodemographic covariates, though all approaches improved over unweighted AoU: mean SMD reductions of 73%-77% for IPW and 42%-44% for clustering methods.

Limitations:

Results conditional on the covariates and analytic sample evaluated. Validation highlighted reweighting of exam and laboratory measures is constrained by data collection differences.

Conclusion:

IPW without post-calibration achieved strong bias reduction while preserving weight stability, offering the best trade-off for most applications. IPW with post-calibration achieved the greatest bias reduction but at the cost of statistical power and inflated variance.

Primary Funding Source:

National Institutes of Health.