Appendix A

Methodology Update: Linear to Logistic Regression Models

Introduction

We changed our statistical modeling approach in this year's report from the prior two reports. The 2025 report includes estimates from logistic regression to estimate the prevalence of each condition or healthcare use metric, instead of linear regression used in the 2022 and 2024 reports. This change means that the specific numeric values from this report are not directly comparable to prior years reports – including that absolute differences in a metric between the 2024 and 2025 reports may be functions of this methodological change rather than a true change in the population health metric. The reports can, however, be compared relatively to illustrate broader trends in whether disparities are improving or worsening over time. This document enumerates in more detail the rationale for this change and illustrates how the reports' findings can be interpreted compared to the prior reports.

Background and rationale for the change in regression model

All of the outcome metrics in this report are binary – yes or no, a person did or did not have the metric of interest (a condition or behavior, had insurance, delayed/avoided care). Linear regression models can be extrapolated such that they may yield practically impossible prevalence estimates, like <0% or >100% prevalence. Our report adjusts for categorical variables – age categories, sex, and income and race categories – which can drive these out-of-bound estimates, especially in combination. When the combinations of age, sex, race and income categories result in small sample sizes, especially in an already rare outcome, the linear model coefficients may lack statistical precision that contribute to these out-of-bounds estimates. In Table A below, after adjusting for age, sex, and income, the estimated prevalence of having 2+ emergency department visits in the prior year was below 0 (-0.37%) for Non-Hispanic Asian adults.

TABLE A **Estimated prevalence from linear regression models of having 2+ emergency department visits in the prior year, by race/ethnicity**

	Non-Hispanic White	Non-Hispanic Black	Non-Hispanic Asian	Hispanic, any race	Other races (includes multi-race)
Unadjusted	3.61	6.27	1.35	5.29	5.74
Age- and Sex-Adjusted	2.47	5.1	0.24	4.14	4.54
Age-, Sex-, and Income -Adjusted	1.6	3.7	**-0.37**	2.8	3.45

Data Source: National Health Interview Survey, 2023. Predicted values are for a 35-44 year old male, who male >$150k per year

Logistic regression inherently bounds these predicted values since probabilities – our approximation for prevalence – cannot mathematically extend beyond 0 and 1 (or 0-100%). There are other regression model approaches commonly used to generate the prevalence difference, especially in academic research, namely Poisson or negative binomial regression. We chose logistic regression and probabilities because their wider use in business and employer settings meant greater familiarity in the audience most likely to read this report.

Interpretation and comparisons between the different approaches

The results from this report are estimated probabilities, which we describe and interpret as prevalence. We acknowledge this interpretation of a probability as a reflection of prevalence is more common in practice than in nuanced academic discussion, but this report has a more practice oriented view and focus. Instead, we can see that probabilities from logistic regression have the same values when unadjusted (Table B). As we adjust for age and sex, and then additionally for income, the comparisons of 'prevalence' from the linear models (Table A) and logistic models (Table B) diverge somewhat numerically.

TABLE B **Estimated probabilities from logistic regression models of having 2+ emergency department visits in the prior year, by race/ethnicity**

	Non-Hispanic White	Non-Hispanic Black	Non-Hispanic Asian	Hispanic, any race	Other races (includes multi-race)
Unadjusted	3.61	6.27	1.35	5.29	5.74
Age- and Sex-Adjusted	2.64	4.6	0.99	3.89	4.17
Age-, Sex-, and Income -Adjusted	1.99	3.06	0.8	2.61	2.97

Data Source: National Health Interview Survey, 2023. Predicted values are for a 35-44 year old male, who male >$150k per year

For example, the age, sex, and income-adjusted model estimates have a 1.99% prevalence vs. 1.6% prevalence for non-Hispanic White adults and a 3.06% prevalence vs 3.7% prevalence for non-Hispanic Black adults between logistic and linear regression. However, they illustrate the same overall racial disparity pattern between groups:

- the ranking from highest to lowest prevalence of using the ED 2+ times across these groups is the same, and
- the relative magnitude of the difference between the most and least common is similar (black-to-white prevalence ratio: 3.06/1.99 = 1.53 times higher in logistic models vs. 3.7/1.6 = 2.3 times higher in linear models)

These differences between linear and logistic model estimates tend to be more pronounced when the sample sizes are small, like racial and sexual orientation based disparities. The differences are smaller for income-based disparities where the income groups are fairly large, which in turn results in more stable regression model coefficients.

TABLE C **Estimated age-, sex-, and race-adjusted probabilities from linear and logistic regression models of having 2+ emergency department visits in the prior year, by income category**

	(1) Under $50,000	(2) $50,000 to $74,999	(3) $75,000 to $99,999	(4) $100,000 to $149,999	(5) $150,000 or more
Linear Model	4.59	4.41	3.02	2.28	1.6
Logistic Model	4.11	4.05	3.07	2.53	1.99

Data Source: National Health Interview Survey, 2023. Predicted values are for a 35-44 year old male, who male >$150k per year

Conclusion

We adjusted our statistical approach in this report to ensure our prevalence estimates used to describe health disparities in the employer-sponsored insurance population stay between 0 and 100%. This change prevents comparing the absolute values of a metric to what we reported in prior years' reports, although the overall trends in disparities reported previously are certainly comparable.