

Supplementary Figures

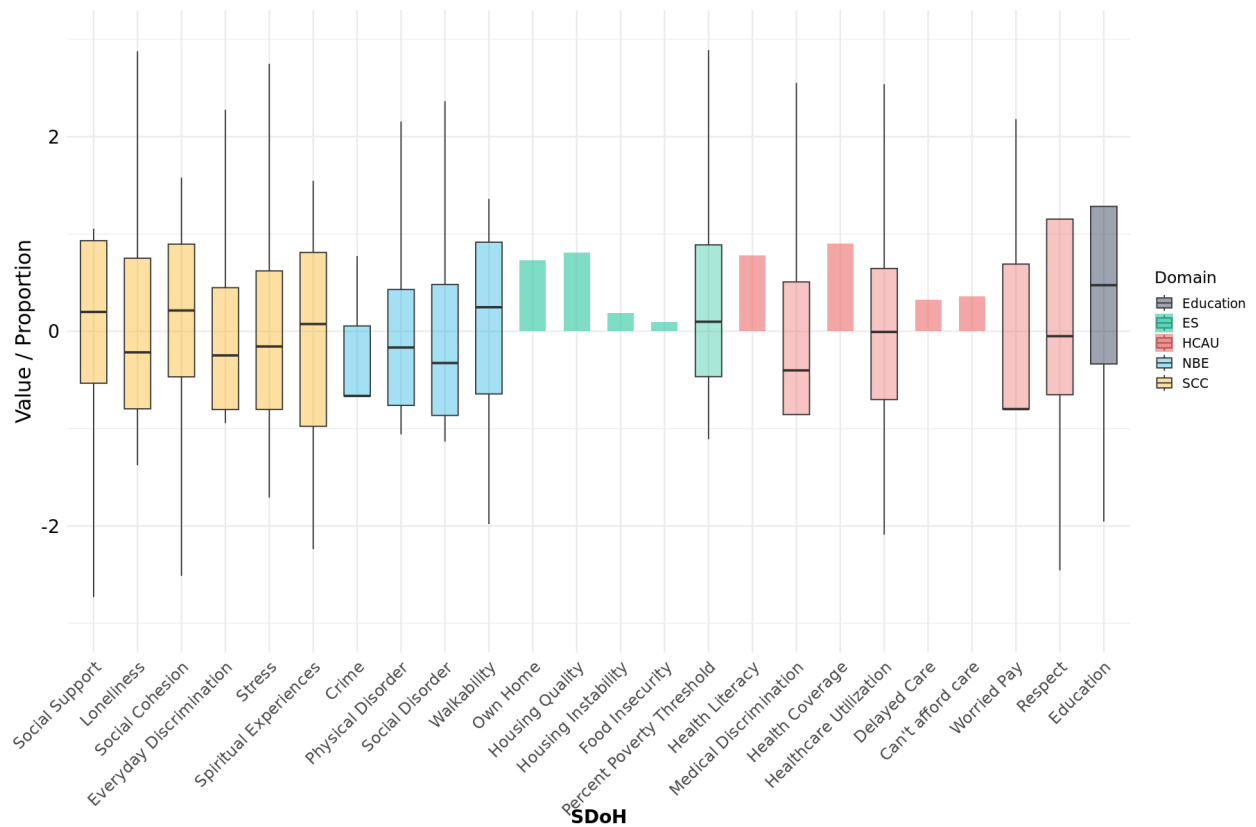


Figure 1: Distribution of Individual-level SDoH

Box plots depicting the distribution of individual-level SDoH in the Individual Cohort. Binary variables are represented as bar chart proportions. SDoH are color coded by HP2030 domain. All continuous SDoH variables were z-score normalized.

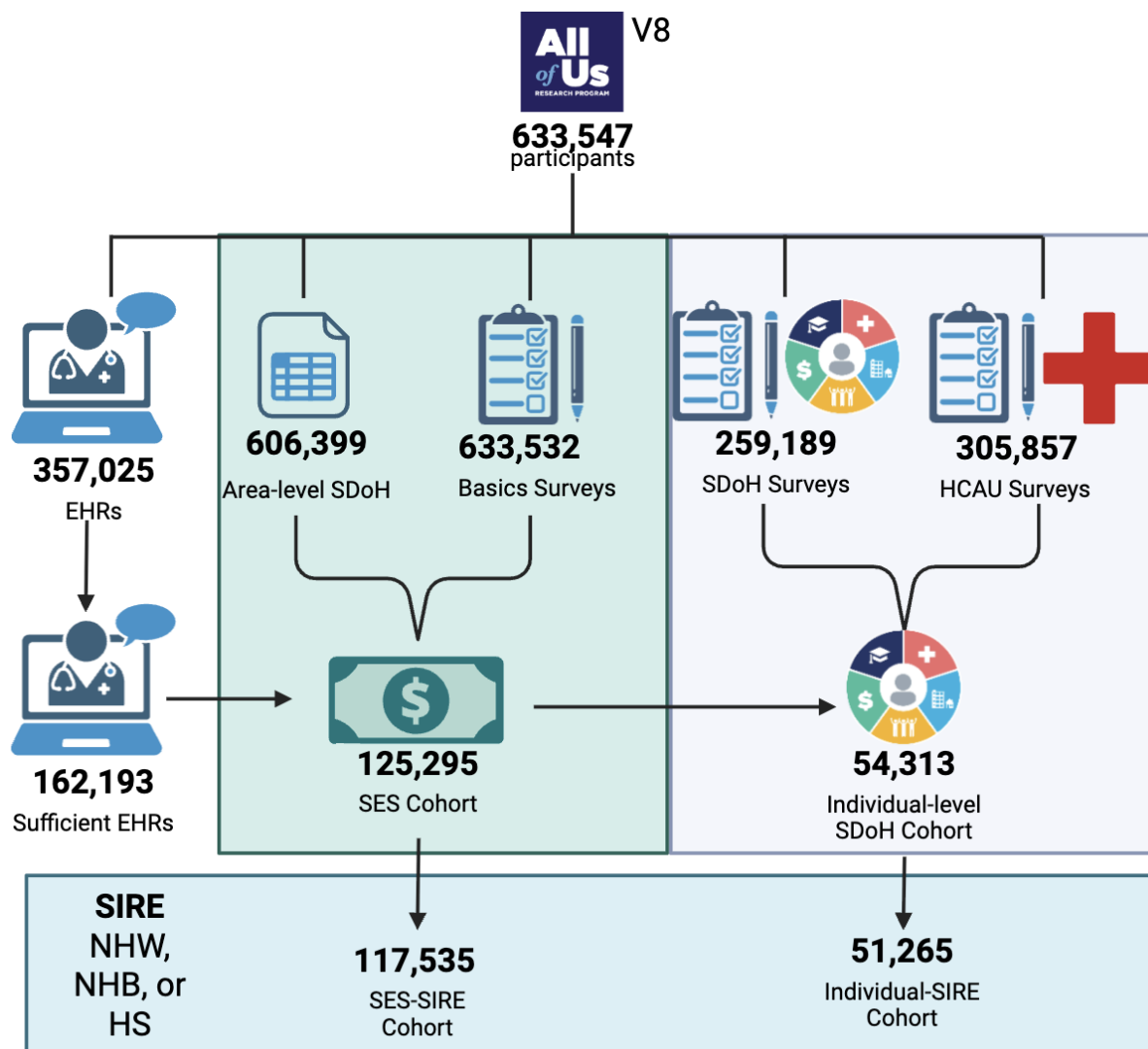


Figure 2: Study Participant Flow Chart

Version Release 8 (V8) of the All of Us Research Program contains data on 633,547 participants. Of those, 357,025 have linked Electronic Health Records (EHRs) and 162,193 meet the criteria for being “sufficient” based on having at least three entries and being at least three years long to meet “medical home” definition. 633,532, 259,189, and 305,857 participants filled out the basics, SDoH, and HCAU surveys respectively. Income data was divided by the corresponding poverty threshold based on location, size of household, and month/year of survey to get the percentage of the poverty threshold. 125,295 individuals had this data in addition to education to make up the SES cohort. Of these, 54,697 filled out the Individual-Level SDoH survey with at least 60% completeness for each domain. Of the Individual and SES Cohorts, 51,265 and 117,535 self-identified as non-Hispanic White (NHW), non-Hispanic Black (NHB), or Hispanic (HS), respectively, to make up the Individual-SIRE and SES-SIRE subcohorts.

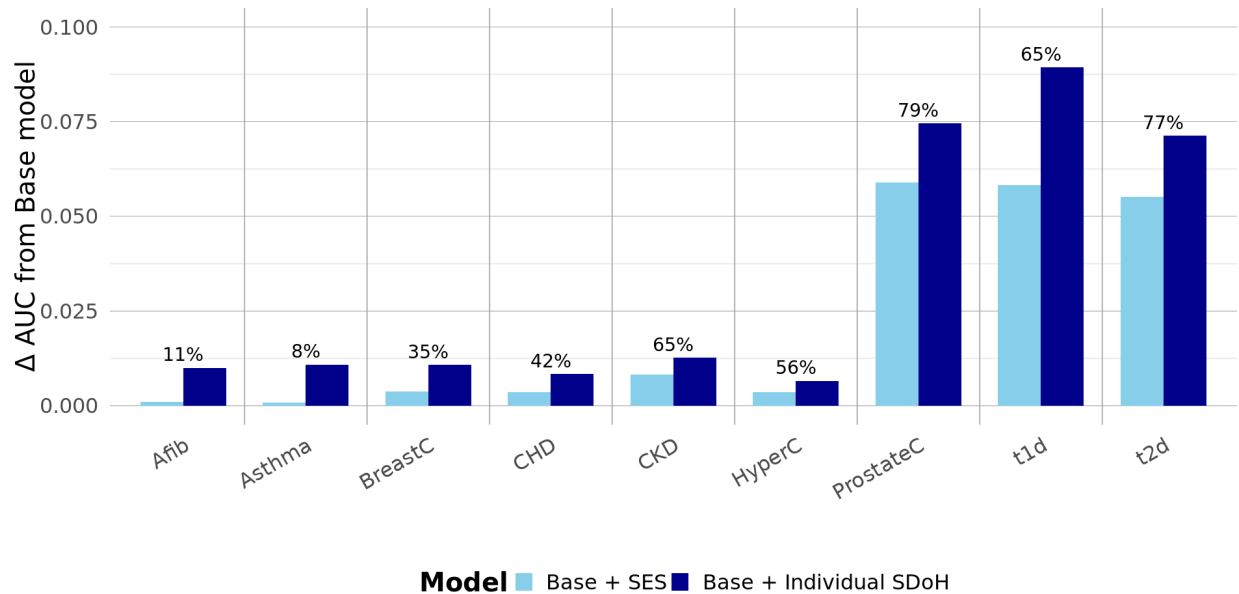


Figure 3: Comparison in change in AUC from the Base model for SES and Individual SDoH models in the Individual SDoH Cohort.

Bar chart of change in AUC from the Base model for SES (light blue) and Individual SDoH models (dark blue). Percent values above the bars indicate the proportion of the predictive improvement from the Individual SDoH model that is captured by the SES model in the Individual SDoH Cohort.

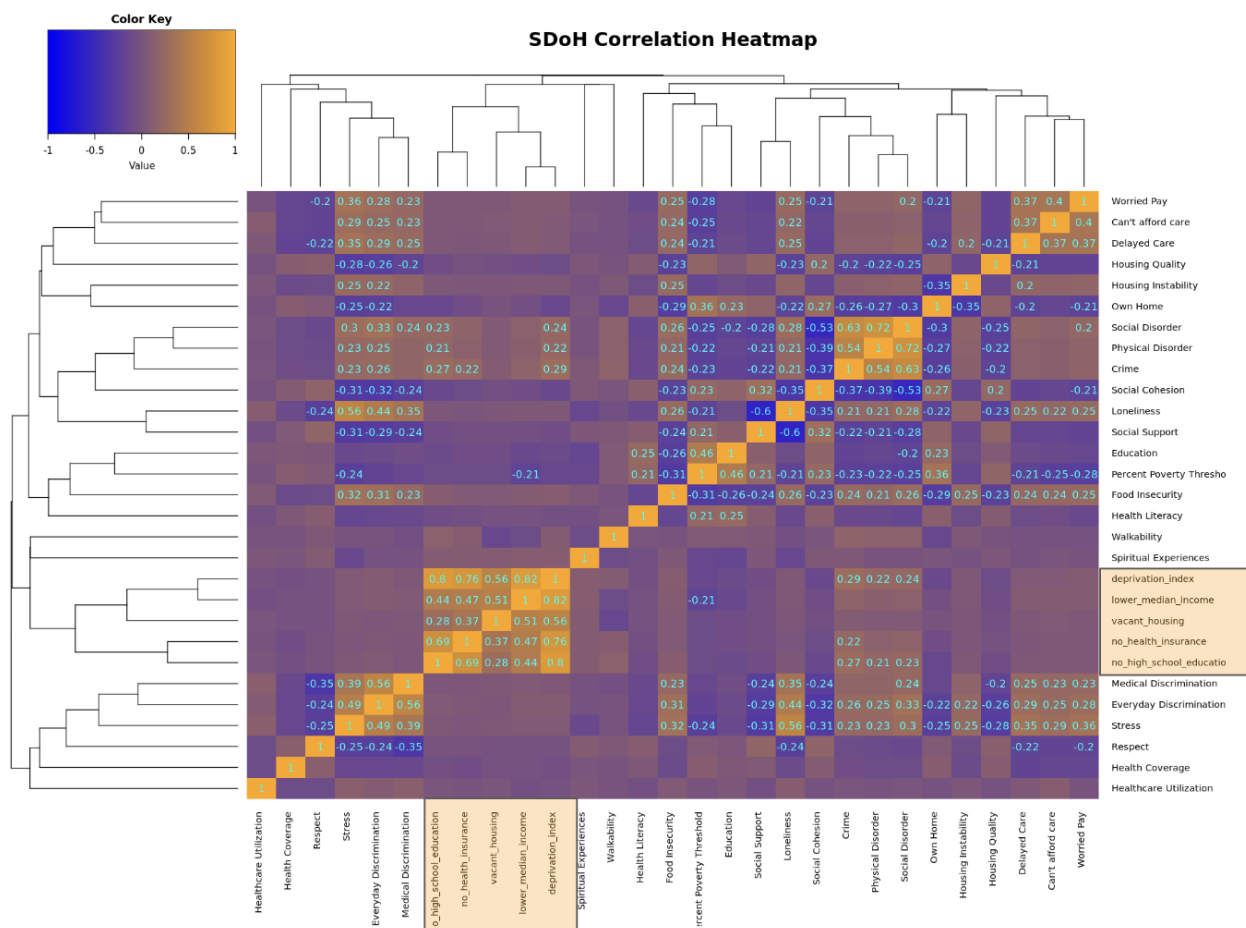


Figure 4: Correlation Heatmap of Individual-level and Area-level Social Determinants of Health in Individual Cohort.

Each cell represents the Pearson correlation coefficient between two SDoH variables. Orange indicates positive correlations, while blue indicates negative correlations. Only correlations exceeding 0.2 are annotated with their respective values. Area-level metric indicators are highlighted in orange.

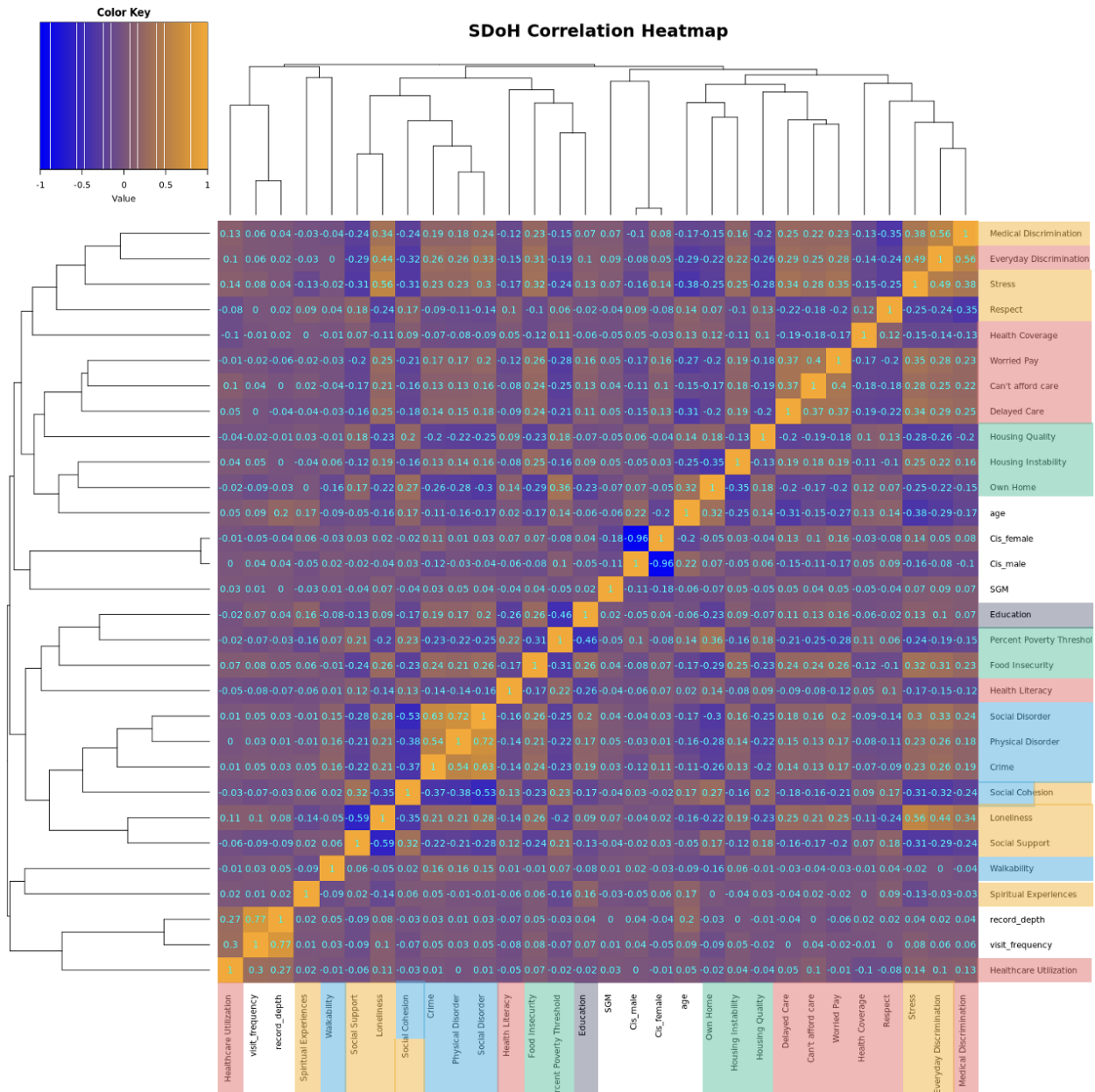


Figure 5: Correlation Heatmap of Social Determinants of Health and Covariates Included in Study

Each cell represents the Pearson correlation coefficient between two SDoH variables derived from the All of Us survey data. Orange indicates positive correlations, while blue indicates negative correlations. Items are colored by HP2030 domain. Component items are color-coded by domain, while latent constructs are depicted in black.

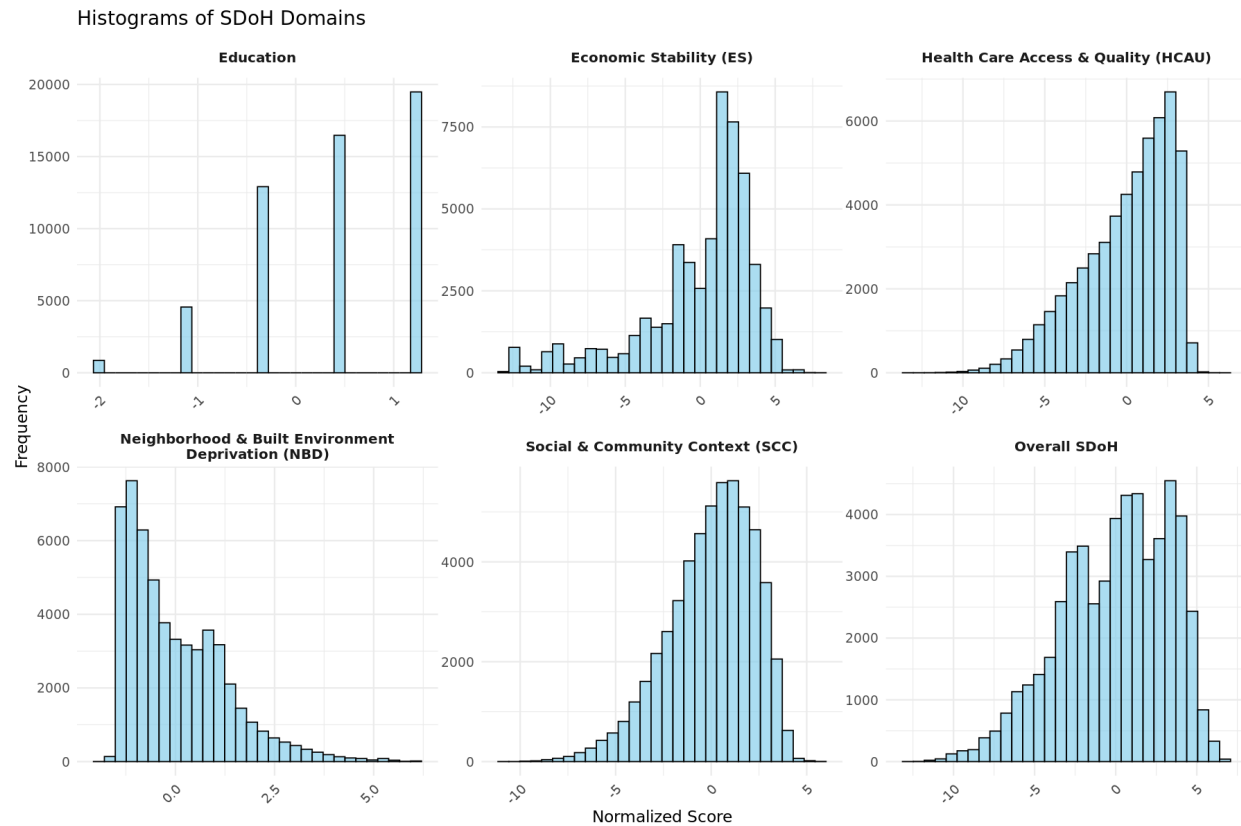


Figure 6: Histograms of SDoH Domains

Normalized domain scores derived from CFA for each latent SDoH domain, as well as an overall SDoH score. Scores have been z-score normalized for direct comparison across domains.

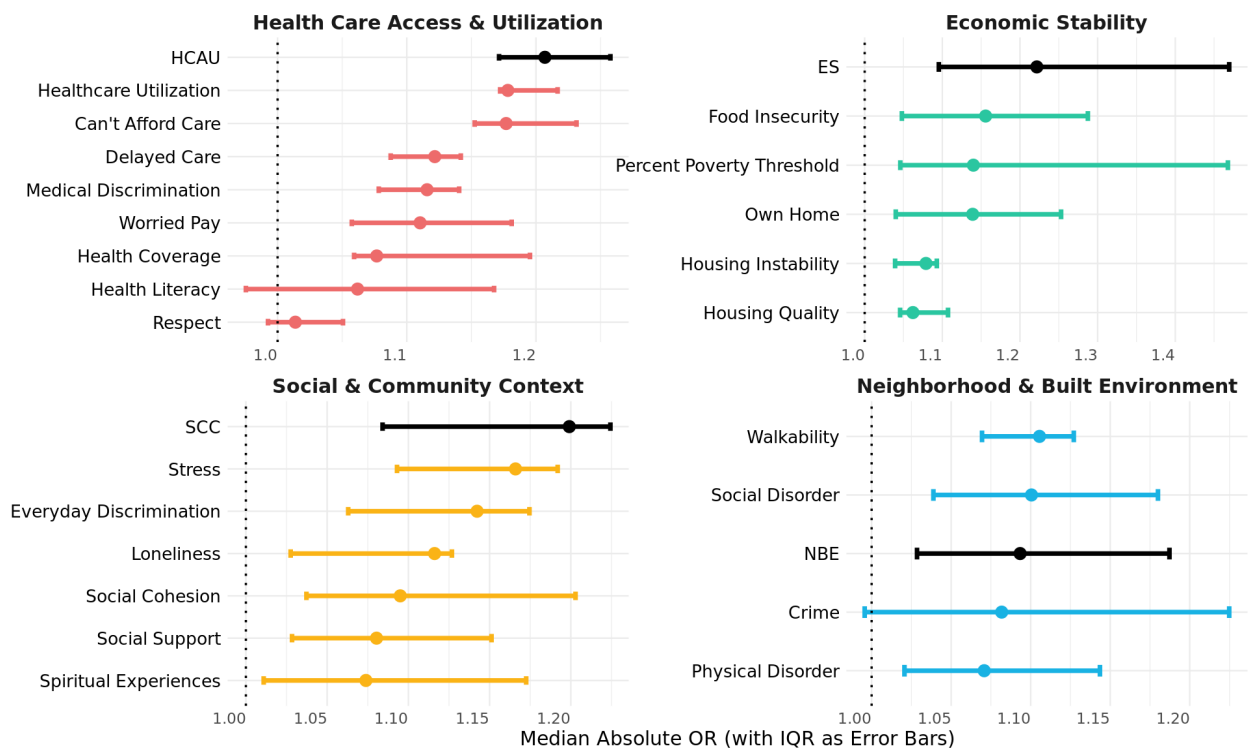


Figure 7: Latent SDoH Domain Constructs Perform Similarly to Components Across Nine Chronic Conditions

Estimates from logistic regressions assess the association between each SDoH domain and their components with nine chronic conditions in the Individual SDoH Cohort. The x-axis is ordered by the median odds ratio across the conditions, with interquartile ranges (IQRs) represented as error bars. Component items are color-coded by domain, while latent constructs are depicted in black. Each model was adjusted for age, age2, Sex/Gender, record depth, visit frequency.

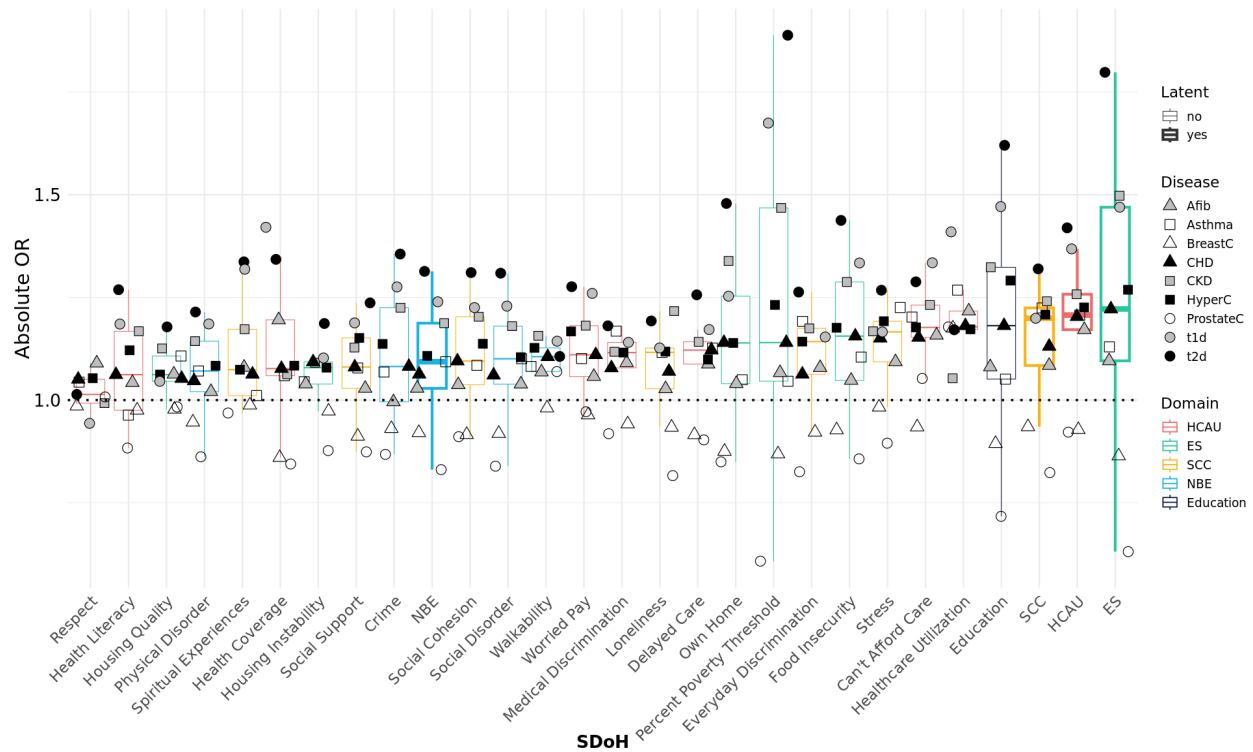


Figure 8: Distribution of Odds Ratios for 9 chronic conditions across 24 individual-level predictors and 4 latent predictors.

Box plots of OR for associations between SDoH on the x-axis and 9 different chronic diseases ordered by OR on the y-axis. SDoH are ordered from left to right by increasing median OR. SDoH are color coded by domain with latent predictors having thicker line width. Different diseases are indicated by shape/fill.

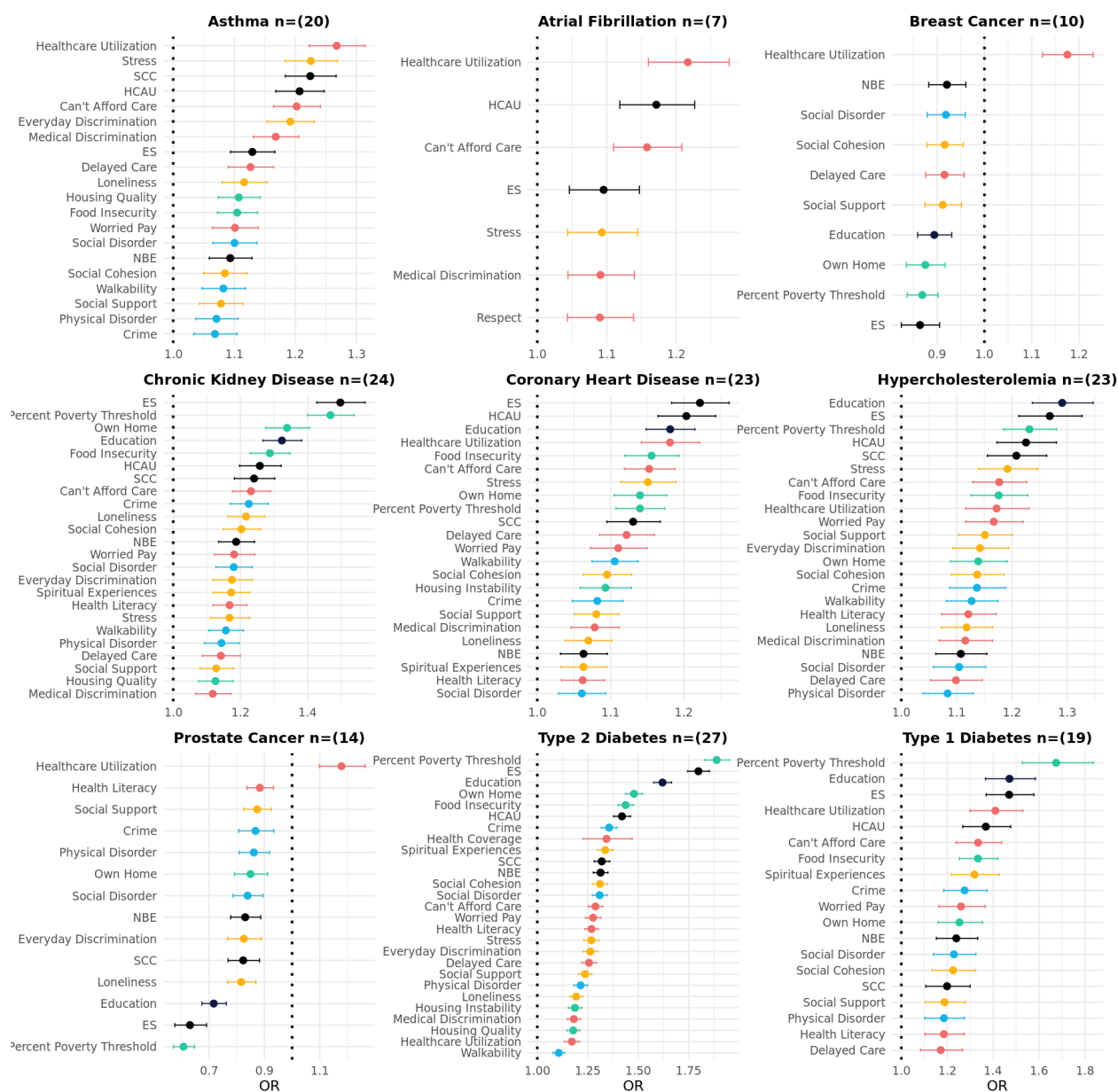


Figure 9: Significant Associations Between Individual-level SDoH and Disease Outcomes

Each panel displays the estimate and standard error for each SDoH significantly associated with the disease (Bonferroni corrected p-value $< (.05/(29 \text{ tested variables} \times 9 \text{ disease outcomes}))$). The number of significant associations is in parenthesis in the title of each disease. Each panel is ordered independently by strength of association. Domains are colored by type in accordance with the HP2030 coloring in Figure 1, and with black indicating domain-level factor scores derived from CFA. Domains are in the direction of risk. Imputed individual component predictors were used to maintain a consistent sample.

Within each disease, domains are ordered from highest to lowest estimated effect size. Points indicate the magnitude and direction of association; horizontal lines represent 95% confidence intervals. The dotted vertical line at zero marks the null effect.

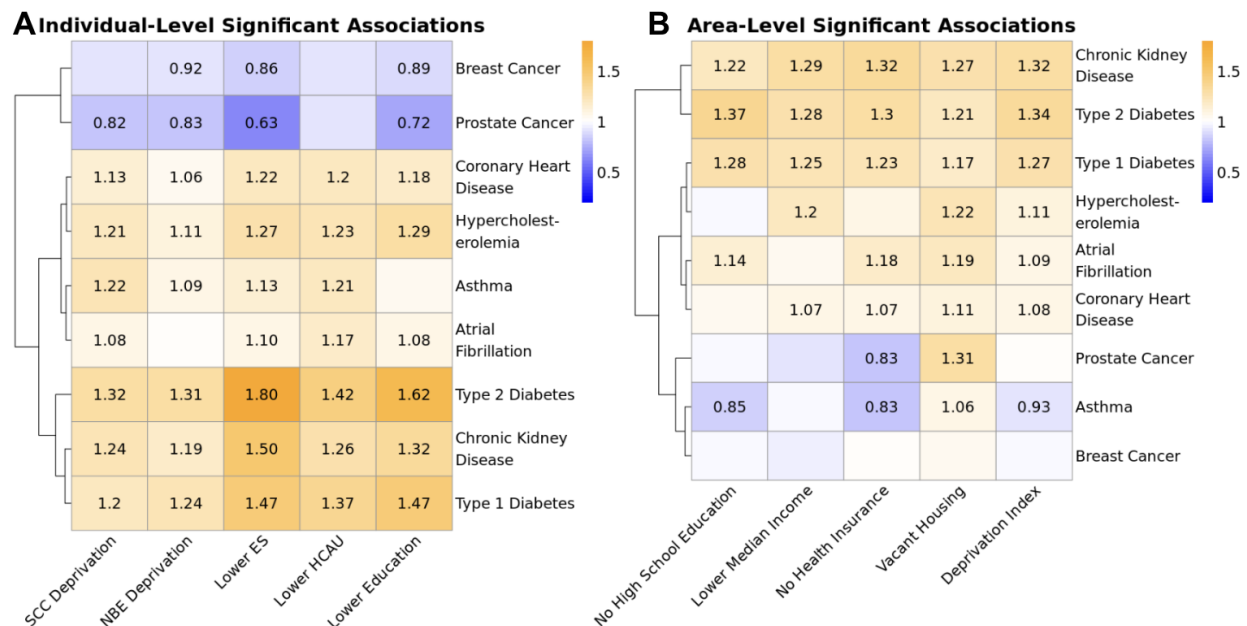


Figure 10: Distinct Patterns of Associations Between SDoH Domains and Nine Chronic Conditions at Individual and Area Levels

A. Heatmap displaying odds ratios from a series of logistic regressions assessing the relationship between individual-level SDoH domains and nine chronic conditions in the Individual SDoH Cohort. Only statistically significant associations are annotated with effect sizes. **B.** Heatmap illustrating the odds ratios from a series of logistic regressions evaluating area-level SDoH domains in the SES Cohort. For both heatmaps, SDoH variables were run in independent models and values below one indicate that higher levels in a specific SDoH risk are associated with reduced probability of disease. Both heatmaps are on the same scale to facilitate direct comparison between individual-level and area-level SDoH associations, though the order of disease outcomes varies.

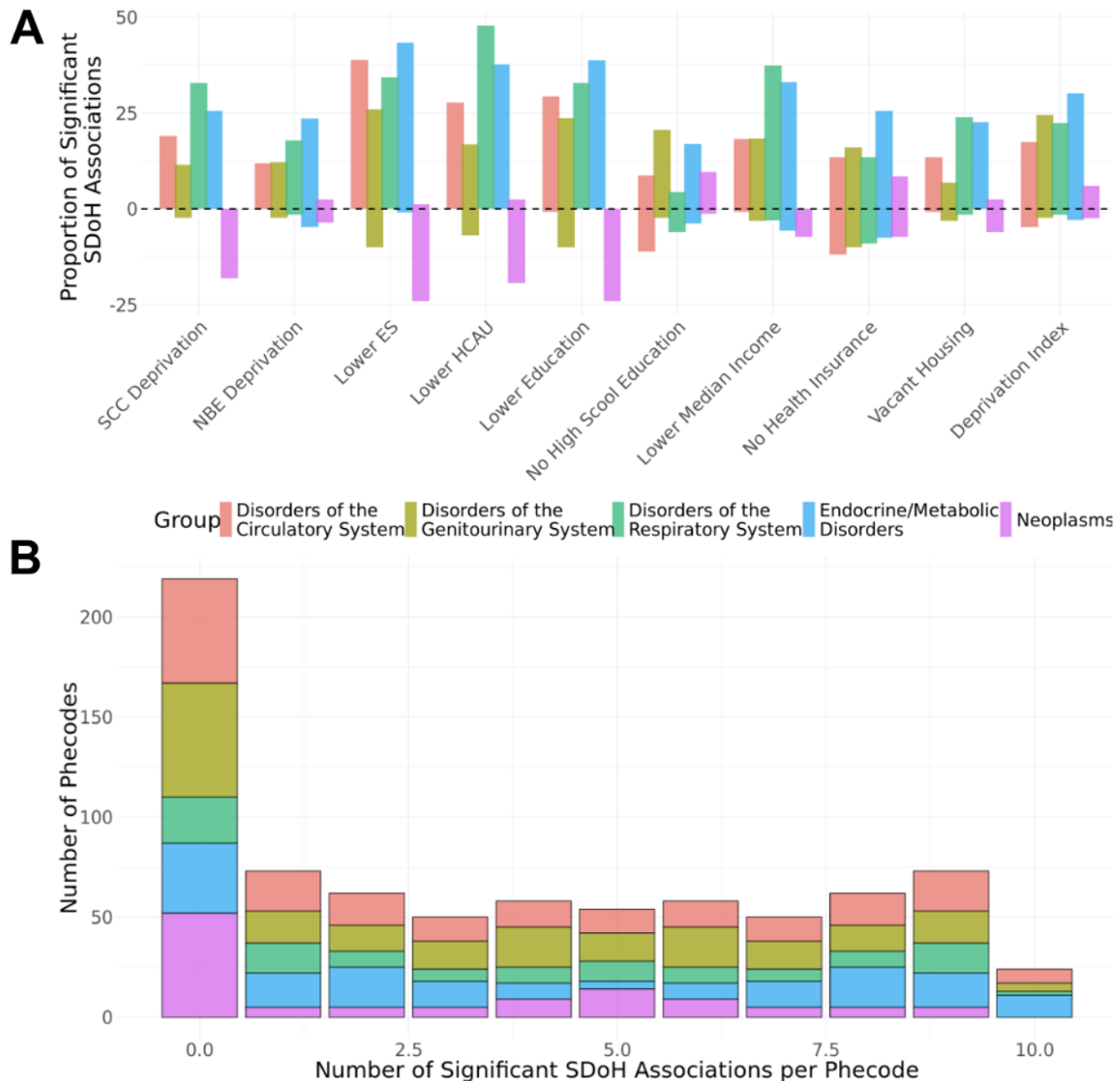


Figure 11: Targeted Phenome-Wide Association Study of SDOH in Individual Cohort

A. Box plot showing the proportion of significant SDOH associations for each disease group. The y-axis represents the proportion of significant associations, split into positive (above the x-axis, associated with increased disease prevalence) and negative (below the x-axis, associated with decreased disease prevalence) effects. Boxes are grouped by SDOH category and color-coded by disease group. **B.** Histogram showing the distribution of significant SDOH associations across phecodes. The x-axis indicates the number of SDOH significantly associated with each phecode (maximum SDOH tested = 10), while the y-axis shows the count of phecodes with that number of associations. The colors correspond to the same group colors in panel A.

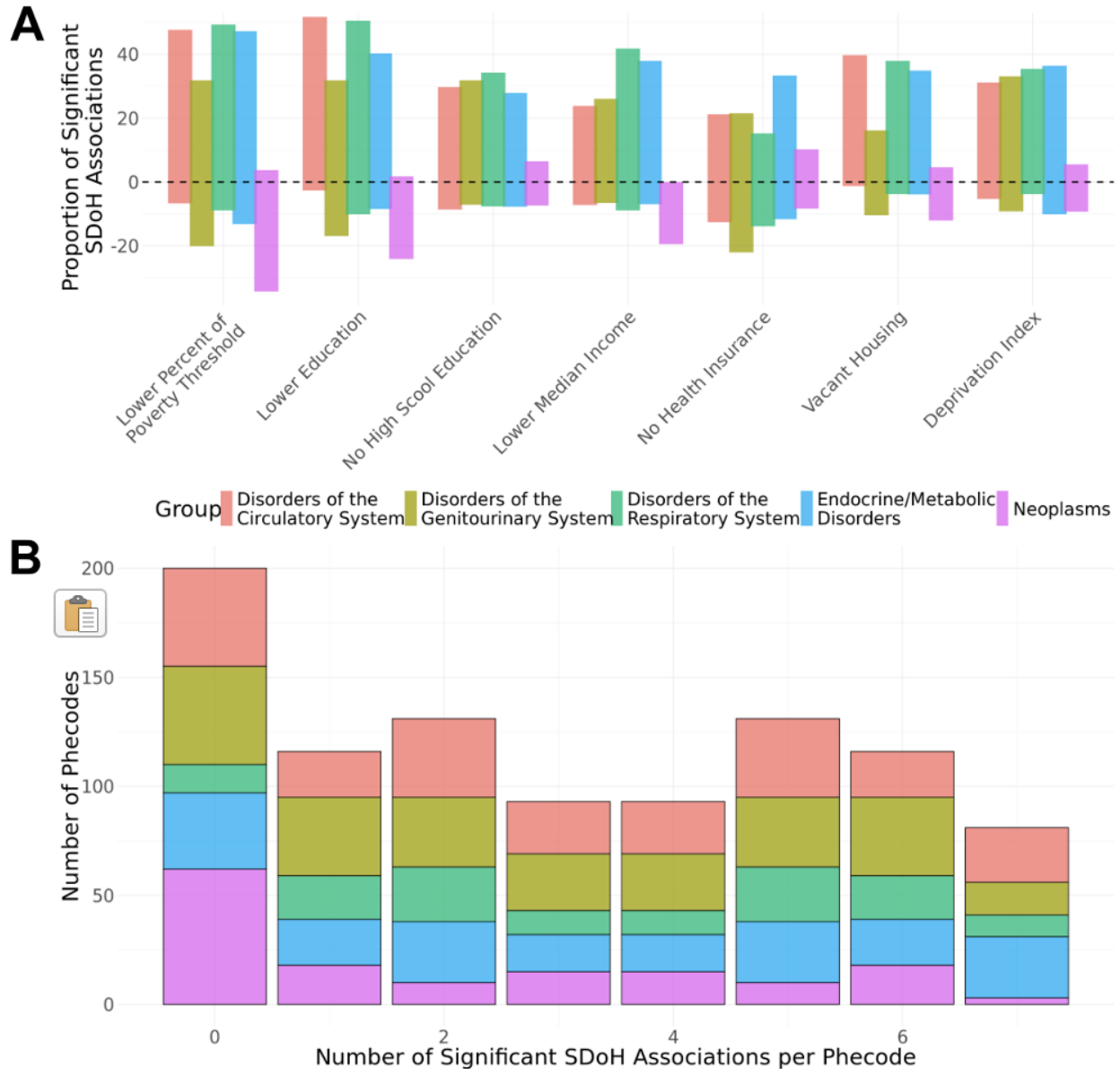


Figure 12: Targeted Phenome-Wide Association Study of SDOH in SES Cohort

A. Box plot showing the proportion of significant SDOH associations for each disease group. The y-axis represents the proportion of significant associations, split into positive (above the x-axis, associated with increased disease prevalence) and negative (below the x-axis, associated with decreased disease prevalence) effects. Boxes are grouped by SDOH category and color-coded by disease group. **B.** Histogram showing the distribution of significant SDOH associations across phecodes. The x-axis indicates the number of SDOH significantly associated with each phecode (maximum SDOH tested = 7), while the y-axis shows the count of phecodes with that number of associations.