Social Determinants of Health Enable Early Identification of Premature Mortality Risk

Chintan Alakh Dalal[1], James Daniel McMichael[1], Leah Broder[1], Kyle William McLean[2]

[1]Socially Determined, Inc., Washington, DC, USA · [2]Veritas Data Research, Inc., Delaware, USA

About This Handout

This document extends the ISPOR poster into a structured scientific narrative. It provides full methodological detail, complete results, and operational implications for population health and payer decision-making.

Abstract

OBJECTIVES: Social Determinants of Health (SDOH) are known drivers of health disparities. However, their role in influencing the timing of premature mortality, distinct from cause-of-death (COD) attribution, remains insufficiently characterized in large-scale mortality data. This study evaluates the association between SDOH burden and age at death and the value of SDOH-based risk stratification for early prioritization.

METHODS: We analyzed a national, de-identified mortality dataset of approximately one million individuals, linking person-level SDOH risk scores with age at death and COD. Individuals were ranked by SDOH risk using multivariable stratification. Premature mortality was defined as death before age 75 and quantified as years of life lost (YLL). Regression-based analyses were used to assess differences in YLL across SDOH risk strata and COD cohorts. Model performance was evaluated based on the concentration of YLL among higher SDOH risk rankings relative to uniform prioritization. Mean age-at-death differences between SDOH burden groups were estimated across COD categories, with uncertainty summarized using confidence intervals.

RESULTS: The SDOH-based risk ranking identified individuals with higher YLL. Individuals ranked at the highest SDOH risk accounted for approximately 2.7 times more premature mortality than expected under uniform prioritization. Higher SDOH burden was associated with earlier ages at death across multiple COD categories, with mean age-at-death differences between high- and low-SDOH burden groups ranging from 10 to 20 years for leading causes.

CONCLUSION: Large-scale analyses indicate that SDOH burden is associated with higher YLL and is reflected in the timing of premature mortality rather than recorded COD alone. An SDOH-based risk ranking identifies individuals at elevated risk of YLL and provides value for early prioritization. Integrating person-level SDOH risk data with mortality and cause-of-death information enables precise characterization of premature mortality patterns and informs population health management and prevention-focused investment.

1. Introduction

Social Determinants of Health (SDOH), including financial strain, food insecurity, housing instability, transportation barriers, and health literacy challenges, shape how long and how well people live.[1–4] Differences in SDOH can lead two individuals with the same underlying disease to experience very different ages at death. Understanding these differences is important for studying premature mortality and Years of Life Lost across populations.

Prior studies have consistently linked SDOH with substantial differences in life expectancy and premature mortality. Galea et al. estimated that several social factors contribute meaningfully to mortality burden in the United States.[3] McGinnis and Foege reframed mortality through "actual causes of death," highlighting the role of upstream behavioral and social factors beyond the conventional cause-of-death framework.[4] Chetty et al. reported a 14.6-year life-expectancy gap between the richest and poorest 1% of US men.[5] Stringhini et al. found that low socioeconomic status was associated with mortality effects comparable to several established clinical risk factors.[6] More recently, Bundy et al. observed progressively higher premature mortality risk with increasing cumulative SDOH burden in a national US cohort.[7]

Population-level analyses have traditionally focused on cause-of-death attribution and disease-specific mortality patterns.[8,9] These approaches are essential for disease surveillance and epidemiologic monitoring, but they are less directly oriented toward measuring differences in the timing of death or Years of Life Lost (YLL) across populations. Cause-of-death classification also summarizes complex and often multimorbid clinical trajectories into a single underlying cause.[10] In parallel, many existing SDOH studies evaluate isolated socioeconomic variables rather than multidomain, person-level SDOH burden, which can limit their utility for population stratification and intervention design. Structured SDOH documentation in administrative healthcare data also remains relatively limited, with ICD-10 Z-codes appearing in only a small fraction of US healthcare records.[11]

This study addresses these gaps by linking multidomain person-level SDOH risk scores to a large national mortality dataset from 2025 spanning populations across Florida, Virginia, West Virginia, Maryland, and Washington, DC. The analysis evaluates premature mortality through two complementary perspectives: concentration of premature mortality within high SDOH risk strata, and shifts in age at death across major cause-of-death groups. To support practical applications in population health prioritization and prevention strategy, the study also evaluates whether SDOH-based risk stratification can support operational decision-making in population health and health economics and outcomes research (HEOR).

The remainder of this handout presents the data, methods, results, and operational implications of the analysis.

2. Data

The analysis drew on approximately one million de-identified mortality records from calendar year 2025. After linkage of person-level Social Determinants of Health, or SDOH, risk scores to age at death and cause of death records, the analytic cohort comprised 594,663 individuals with complete information across all three layers. The cohort included decedents from Florida, Maryland, Virginia, West Virginia, and Washington, DC, and was constructed under de-identified, privacy-compliant protocols.

2.1 SDOH risk scores

Person-level SDOH risk scores were generated by Socially Determined. The scoring framework integrates more than 150 indicators across public-sector, commercial, and specialty data streams. These indicators are organized into 15 drivers spanning five SDOH domains: Financial Strain, Food Insecurity, Housing Instability, Transportation Barriers, and Health Literacy Challenges. Risk estimation is performed using a probabilistic machine learning framework based on Bayesian Networks. Each domain is scored on a 1-to-5 ordinal scale ranging from low to high risk. Additional construction details are provided in Appendix A.1.

2.2 Mortality data

Mortality outcomes were sourced from Veritas Data Research. The Veritas Fact of Death index consolidates more than 45,000 mortality sources, including government registries, memorial records, interment records, and other public sources, into a composite mortality record.[12] The consolidation process applies deterministic deduplication logic to merge death events across sources into a single record per individual.

The Veritas Cause of Death solution combines Fact of Death records with clinical real-world data to derive a likely primary cause of death, together with contributing comorbidities, using a deterministic classification framework. Additional processing details are provided in Appendix A.2.

2.3 Record linkage

SDOH risk scores and mortality records were linked at the person level using a privacy-preserving tokenization framework. Records from both sources were resolved through a shared identity graph using multiple independent linkage tokens to support both precision and recall in entity resolution. Linkage quality was verified by confirming one-to-one collapse of traceability identifiers within the merged dataset, reducing the likelihood of duplicate identities during integration.

The resulting linked dataset provided, for each individual, person-level SDOH risk scores together with age at death and primary and contributing causes of death.

2.4 Cohort composition

The merged dataset was prepared in four stages. First, duplicate rows were removed while retaining one record per individual per cause-of-death designation. Second, the primary cause-of-death record was selected for cohort-level analyses, while contributing comorbidities were retained for parallel comorbidity analyses. Third, age at death was calculated from date of birth and date of death, and the cohort was restricted to adults aged 20 years and older. Fourth, records with missing values in analytic fields were excluded, yielding a complete-case cohort of 594,663 individuals.

Years of Life Lost, or YLL, was defined as the difference between age at death and the premature mortality threshold of 75 years.[8,9] The 75-year threshold is widely used in population health and epidemiologic studies because it approximates expected lifespan while preserving comparability across populations and time periods. This framing emphasizes deaths occurring substantially earlier than expected life expectancy and provides an interpretable measure of premature mortality across SDOH burden groups and cause-of-death categories.

Domain-level risk scores were translated into operational burden indicators. A domain was considered elevated when its score reached 4 or 5 on the ordinal risk scale. Individual SDOH burden was defined as the count of elevated domains across the five-domain framework, ranging from 0 to 5. For descriptive analyses, burden was grouped into four strata: SDOH 0 for no elevated domains, SDOH 1 for one elevated domain, SDOH 2 for two elevated domains, and SDOH 3+ for three or more elevated domains. Corresponding cohort distributions are shown in Figure A.4a.

After age adjustment to the Year 2000 US Standard Population, prevalence of elevated risk across all five SDOH domains ranged from 25% to 31%. Age adjustment was applied to account for differences in age structure across mortality subgroups and to support comparability of burden estimates across populations.[13]

To evaluate support for cause-of-death stratified analyses, the proportion of individuals with high SDOH burden, defined as two or more elevated domains, was examined across seven cause-of-death groups and three age-at-death bands. Cause-of-death groupings were derived from the 27 most prevalent ICD-coded primary causes of death in the cohort and mapped into seven mutually exclusive clinical categories, as summarized in Table A3. All cells retained substantial sample sizes, supporting both group-level and subgroup-level inference across the analytic cohort. The corresponding distributions are shown in Figure A.4b.

Comorbidity burden was also evaluated across SDOH strata using counts of contributing conditions recorded at death. Both mean and median comorbidity counts increased progressively with higher SDOH burden, supporting a cumulative disease burden pattern across the cohort. The corresponding distributions are shown in Figure A.4c.

3. Methods

The analysis was designed as a conditional-on-death enrichment study using a national cohort of US decedents from calendar year 2025. The objective was to evaluate how multidomain person-level SDOH burden relates to differences in age at death and concentration of premature mortality within higher-risk SDOH strata. Because the analytic cohort was restricted to decedents observed during the study period, reported estimates should be interpreted as enrichment and mortality-timing differences within decedents rather than as population-level mortality risks.

3.1 Outcome Definitions and Risk Stratification

Premature mortality was defined as death before age 75, consistent with common population-health and epidemiologic convention.[8,9] Two related outcomes were evaluated throughout the analysis. The first was a binary indicator of premature mortality, used for predictive risk stratification. The second was age at death, analyzed as a continuous outcome for mortality timing and Years of Life Lost analyses.

Two representations of SDOH were used throughout the analysis. The first was cumulative SDOH burden, defined as the number of elevated SDOH domains per individual. For comparative analyses, high burden was defined as two or more elevated domains, while low burden was defined as no elevated domains.

The second representation was a predictive risk-ranking framework. Individuals were ranked according to predicted probability of premature mortality and divided into ten equally sized risk deciles ranging from highest predicted risk in Decile 1 to lowest predicted risk in Decile 10.

3.2 Predictive Modeling Framework

A logistic regression model was fit using premature mortality as the binary outcome. Predictors included the five continuous person-level SDOH domain scores together with cumulative SDOH burden. Gender was included as a demographic adjustment variable. Age was excluded from the predictive model because premature mortality was directly defined using age at death.

Logistic regression was selected because it provides a transparent and interpretable framework for evaluating associations between multidomain SDOH burden and premature mortality risk. The resulting probability estimates also support operational deployment within existing risk-stratification workflows. The modeling framework was designed to evaluate association and enrichment within the decedent cohort rather than causal effects of SDOH on mortality outcomes.

3.3 Mortality Timing and Years of Life Lost Analyses

To evaluate how SDOH burden relates to mortality timing across causes of death, mean age-at-death differences and corresponding Years of Life Lost differences were estimated between high-burden and low-burden groups within each cause-of-death stratum.

Analyses were conducted at two levels of resolution: seven mutually exclusive clinical cause-of-death groups and the 27 most prevalent ICD-coded primary diagnoses within those groups. Cause-of-death groupings are summarized in Table A2.

Uncertainty was quantified using bootstrap percentile 95% confidence intervals within each stratum. Bootstrap inference was selected because it does not require distributional assumptions for age-at-death outcomes and provides stable interval estimation across heterogeneous subgroup sizes.

3.4 Model Evaluation and Validation

Model performance was evaluated using both concentration and discrimination metrics. Concentration of premature mortality across risk deciles was quantified using lift, defined as the observed share of premature deaths within a given risk decile divided by the share expected under uniform random ordering of the decedent cohort. A lift value greater than 1.0 indicates enrichment of premature mortality within that risk stratum relative to random allocation.

Discrimination was evaluated using the area under the receiver operating characteristic curve, or AUC, estimated using stratified 5-fold cross-validation. Cross-validation was stratified on the binary premature mortality outcome to preserve outcome prevalence across folds. Mean AUC together with fold-level variability was used to assess overall model stability and generalizability.

3.5 Driver direction methodology used in persona analysis

In addition to the primary predictive and timing analyses, an auxiliary directional analysis was conducted to characterize the SDOH driver profiles associated with each decile-level persona reported in Section 4.4. The directional analysis evaluates whether each persona's exposure profile is shifted toward more-adverse or more-protective values relative to the rest of the cohort, using a set of supplementary features per SDOH domain (Appendix A.5) drawn from the full Socially Determined data dictionary beyond the five primary domain scores used as model predictors.

Each supplementary feature is annotated with a direction interpretation indicating whether higher or lower values correspond to more-adverse exposure. For each persona, feature-level differences between the target decile and the comparator deciles are computed and sign-aligned so that positive differences always reflect more-adverse exposure. Feature-level differences exceeding a pre-specified small-effect threshold are classified as adverse or protective, and differences below the threshold are treated as neutral. The principal driver for each domain in each decile is then assigned a final directionality of Adverse (▲), Protective (▼), or Mixed (●) based on the aggregated pattern of signals across its supplementary features. The classification is descriptive and decile-specific rather than causal, and the Mixed designation indicates that the available signals did not reach the threshold required for a confident directional assignment.

4. Results

Across the analytic cohort, higher multidomain SDOH burden was consistently associated with greater concentration of premature mortality, earlier age at death, and higher comorbidity burden across major causes of death. The sections below evaluate the performance of the SDOH-based risk stratification framework, characterize mortality-timing differences across cause-of-death groups and ICD-coded diagnoses, and present representative decile-level personas illustrating how driver profiles and mortality patterns vary across the SDOH risk distribution.

4.1 Premature mortality concentrates within higher SDOH risk strata

The SDOH-based risk-ranking framework concentrated premature mortality disproportionately within the highest predicted-risk strata of the decedent cohort and demonstrated a monotonic decline in premature mortality across subsequent deciles, as shown in Figure 4.1.

Individuals in Decile 1, representing the highest predicted SDOH risk group, experienced approximately 2.7 times more premature deaths than expected under random allocation, while deciles below Decile 5 fell below the random baseline. The top 20% of the SDOH-ranked cohort accounted for approximately half of all premature deaths in the analytic sample.

Model discrimination, evaluated using stratified 5-fold cross-validation, remained stable across folds, with minimal variability in performance estimates. Together, these findings indicate that a parsimonious SDOH-only logistic framework provides meaningful separation between earlier and later mortality timing within the decedent cohort, consistent with cumulative SDOH burden gradients previously reported in national mortality literature.[7]

4.2 Earlier age at death with higher SDOH burden across causes of death

Higher SDOH burden was associated with substantially earlier age at death across every major cause-of-death group, and the same pattern remained consistent across the 27 ICD-coded primary diagnoses comprising those groups (Figures 4.2a and 4.2b). The magnitude of the observed age-at-death differences was comparable to the broad socioeconomic mortality studies previously described in US population studies.[5,6]

The largest mortality-timing differences were observed in neurologic and cognitive conditions, while cancer-related mortality demonstrated the smallest, though still substantial, age-at-death shift. Importantly, every cause-of-death group and every individual diagnosis demonstrated earlier mortality timing in the high-SDOH-burden population relative to the low-burden population.

The consistency of this pattern indicates that higher multidomain SDOH burden is associated not only with premature mortality overall, but also with substantially earlier mortality timing within specific disease trajectories and causes of death.

4.3 Risk-profile personas across the SDOH risk distribution

Three representative personas were constructed from contrasting locations within the SDOH risk-ranking distribution: Decile 1 representing the highest predicted-risk population, Decile 5 representing intermediate risk, and Decile 10 representing the lowest predicted-risk population (see Table 4.3).

The personas demonstrate that the SDOH risk distribution varies not only in magnitude but also in the composition of social-risk exposure. The Decile 1 Premature Multidomain Risk persona demonstrated the youngest age at death, the highest comorbidity burden, and consistently adverse directional signals across housing-quality and food-accessibility drivers. In contrast, the Decile 10 Later-Life Residual Barriers persona demonstrated the latest age at death and lowest comorbidity burden, together with more protective housing-quality and food-accessibility signals. However, selected adverse signals persisted within the lowest-risk decile, particularly across financial strain and demographic drivers, suggesting that residual social risk remains present even among later-life mortality profiles.

Together, these personas illustrate how multidomain SDOH burden, driver composition, comorbidity accumulation, and mortality timing vary systematically across the SDOH risk distribution. The profiles also demonstrate how the risk-ranking framework can support interpretable population segmentation for operational prioritization and intervention planning without reducing multidomain SDOH structure to a single composite score.

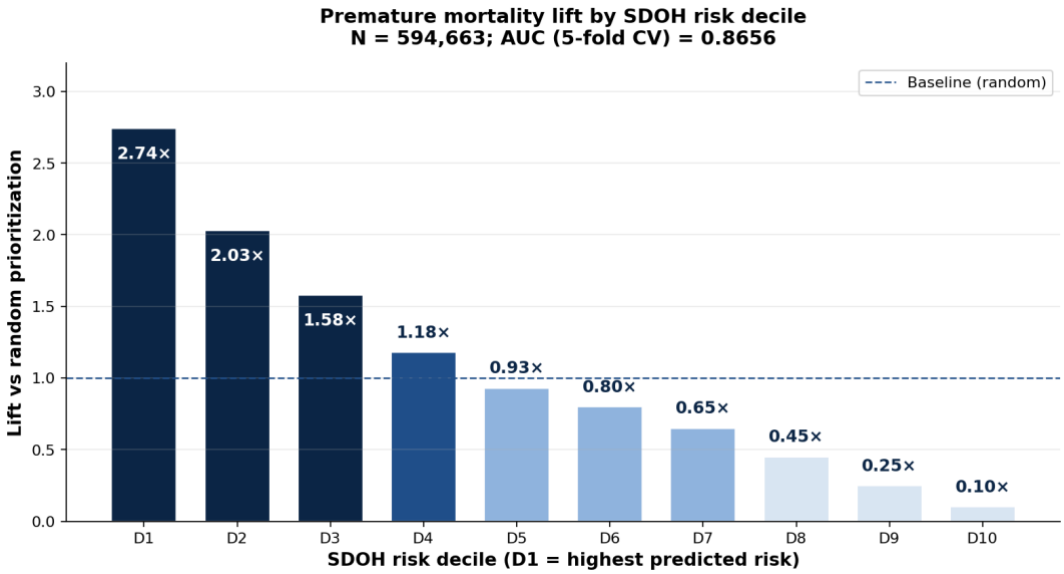


Figure 4.1. Premature mortality lift by SDOH risk decile. D1 = highest predicted risk; D10 = lowest. The dashed line is the random-prioritization baseline (lift = 1.0).

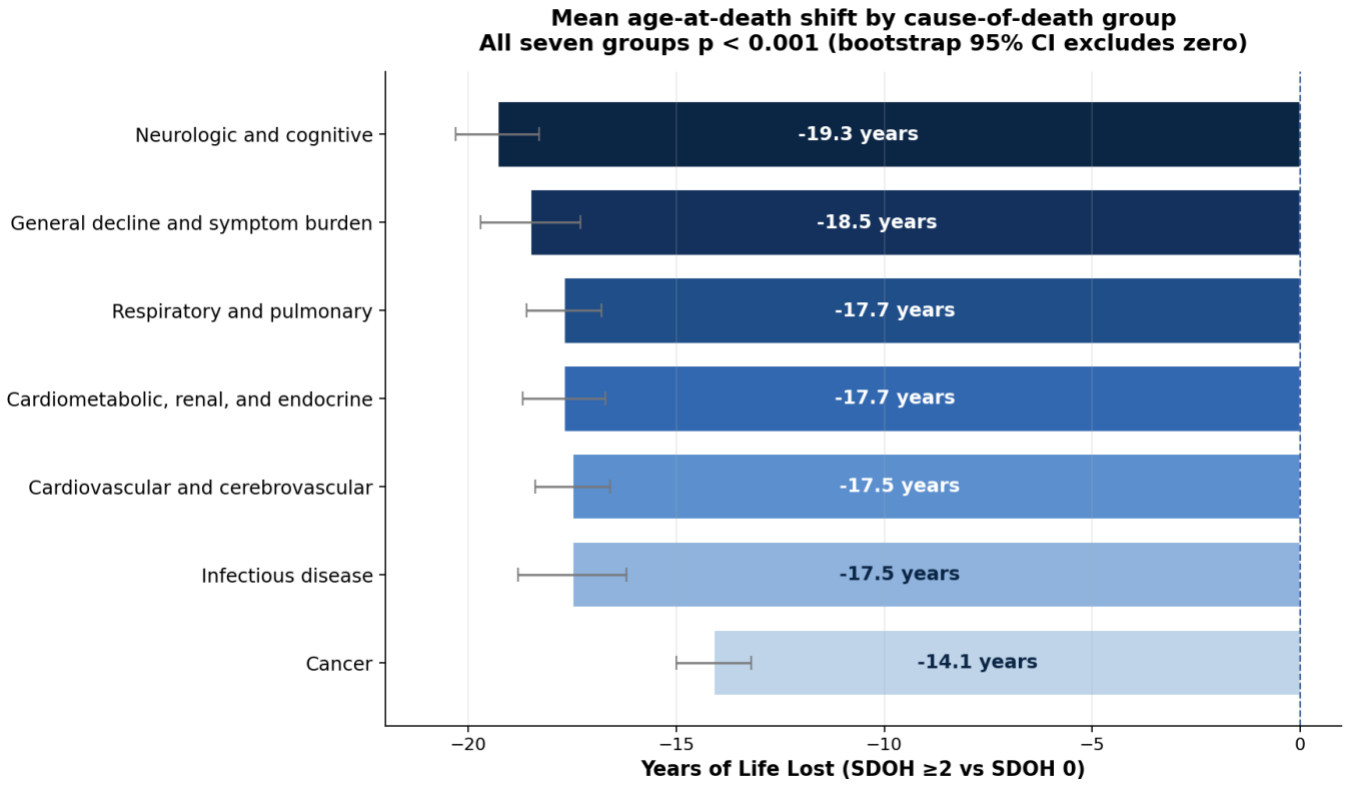


Figure 4.2a. Mean age-at-death shift between high (≥2 elevated domains) and low (0 elevated domains) SDOH burden groups, across the seven cause-of-death groups. Error bars are bootstrap 95% confidence intervals; all seven groups p < 0.001.

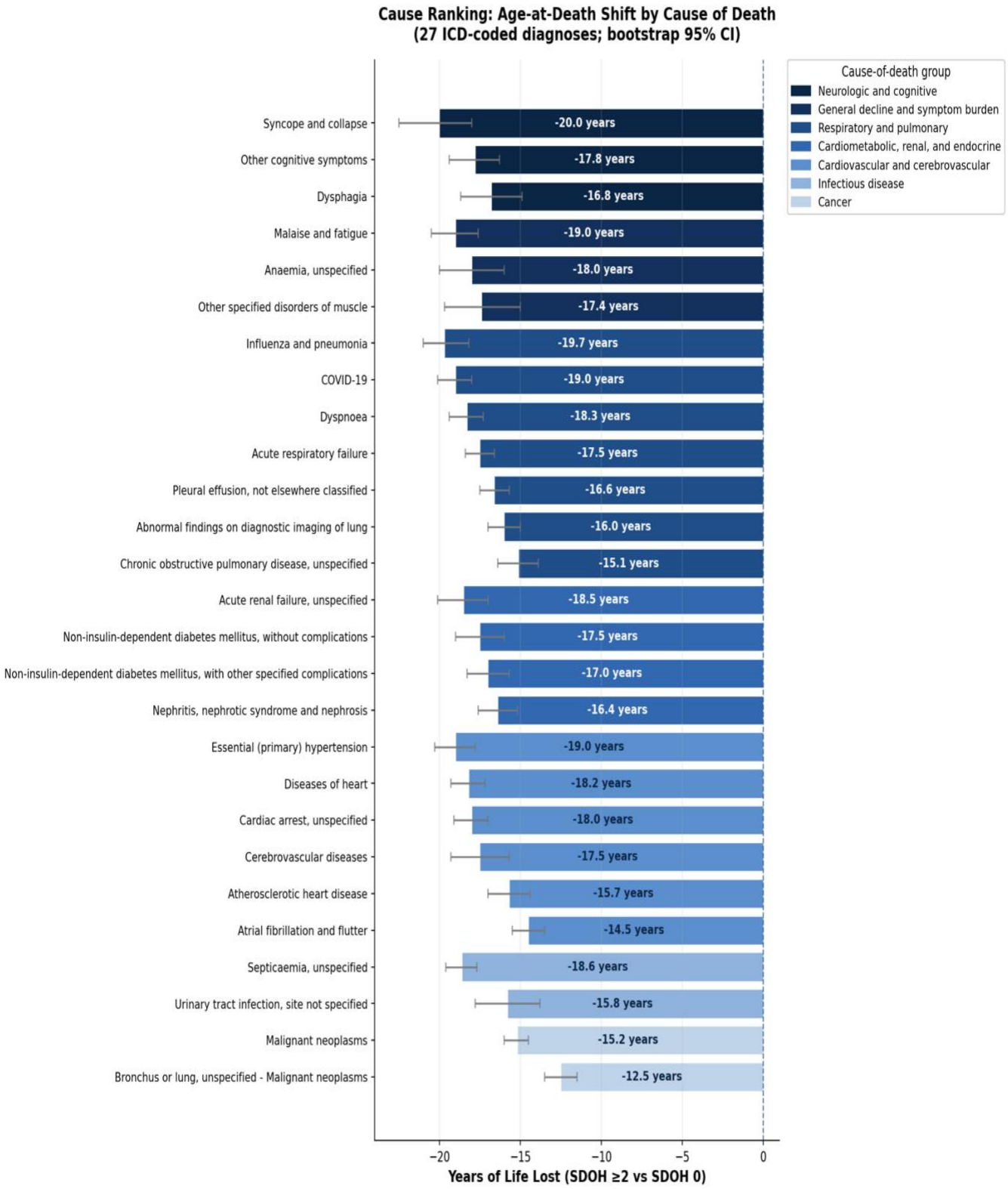


Figure 4.2b. Bars show mean age-at-death difference (years earlier) between high SDOH burden (≥2 elevated domains) and low SDOH burden (0 elevated domains). Error bars are bootstrap 95% CIs. Color indicates cause-of-death group.

Attribute	D1 — Highest Risk	D5 — Mid Risk	D10 — Lowest Risk
Persona label	Premature Multidomain Risk	Mixed Chronic Burden	Later-Life Residual Barriers
Years of Life Lost (YLL)	−20 years	+4 years	+10 years
Mean (median) age at death, years	51.8 (55)	≈ 79	84.9 (87)
Gender prevalence	Male (60%)	Male (majority)	Female (76%)
Region prevalence	Rural	Suburb	Urban
Median comorbidities at death	≈ 7	≈ 5	≈ 3
Top causes of death	Alcohol & liver related; Respiratory; Cancer	Cardiovascular; Respiratory; Neurologic	Alzheimer & dementia; Essential hypertension; General decline
Top Housing driver	Quality 49% ▲	Quality 55% ●	Quality 60% ▼
Top Food driver	Accessibility 48% ▲	Accessibility 54% ●	Accessibility 72% ▼
Top Transportation driver	Health Resource Density 41% ●	Mobility Restrictions 44% ●	Mobility Restrictions 47% ●
Top Financial driver	Assets 49% ●	Assets 61% ●	Assets 43% ▲
Top Health-literacy driver	Education 67% ●	Education 39% ●	Demographics 74% ▲

Table 4.3. Decile Personas from the SDOH Risk Prioritization Model: *D1 (highest predicted SDOH risk), D5 (mid risk), and D10 (lowest predicted SDOH risk). YLL values are relative to the 75-year premature mortality threshold: negative values indicate age at death below 75 (years of life lost); positive values indicate age at death above 75 (years gained beyond the threshold). Driver direction symbols: ▲ adverse, ▼ protective, ● mixed signal. Driver direction describes the observed relationship between a given driver and premature mortality, based on driver-level analysis. A driver classified as Adverse (▲) is associated with earlier age at death; a driver classified as Protective (▼) is associated with later age at death; a driver classified as ● shows a mixed directional signal across the cohort.*

5. Discussion

In this national decedent cohort of more than half a million individuals linked to person-level multidomain SDOH risk scores, premature mortality concentrated disproportionately within the highest predicted-SDOH-risk strata. Decedents with high SDOH burden also reached substantially earlier ages at death across every major cause-of-death group and across the 27 most prevalent ICD-coded primary diagnoses. In parallel, comorbidity burden at death increased

progressively with SDOH burden. Together, these findings indicate that SDOH not only shape what contributes to death, but also when death occurs and how many years of life precede it.

The magnitude and consistency of these findings align with prior US population-level literature reporting large life-expectancy differences across socioeconomic strata,[5] mortality effects of low socioeconomic position comparable to established clinical risk factors,[6] and progressively higher premature mortality associated with cumulative unfavorable SDOH exposure.[7] The present study extends that literature by operationalizing SDOH at the person level across five interacting domains, linking multidomain SDOH profiles to a national mortality dataset, and evaluating SDOH within a practical risk-stratification framework.

For payer, provider, and HEOR audiences, the findings suggest that multidomain SDOH risk can support earlier identification of individuals at elevated risk for premature mortality, potentially before advanced clinical disease becomes fully apparent within traditional healthcare workflows. The framework also supports upstream prevention prioritization across financial, food, housing, transportation, and health-literacy domains, together with more precise estimation of avoidable mortality and prevention-related return on investment.

Several limitations should be considered when interpreting these findings. The analysis is conditional on death and therefore evaluates mortality timing and premature mortality concentration within a decedent cohort rather than population-level mortality risk. The study is observational and does not establish causal effects between SDOH exposure and mortality outcomes. The cohort was drawn from five US jurisdictions, and generalization beyond these geographies should therefore be made cautiously. Cause-of-death classifications were derived using deterministic assignment methods with limited secondary-cause variation. In addition, SDOH risk scores reflect the most recently observable exposure profile rather than cumulative lifetime exposure history.

Overall, the findings support the use of multidomain person-level SDOH risk as an operational framework for prevention investment, population-health prioritization, and SDOH-adjusted health-economics modeling. Integration of SDOH risk with mortality and cause-of-death information may help identify individuals with a greater opportunity for life-years potentially recoverable through earlier social-risk identification and intervention.

References

1. Closing the gap in a generation: Health equity through action on the Social Determinants of health - final report of the Commission on Social Determinants of Health. World Health Organization. Accessed May 2026. https://www.who.int/publications/i/item/WHO-IER-CSDH-08.1.
2. Social Determinants of Health. Social Determinants of Health - Healthy People 2030. Accessed May 2026. https://odphp.health.gov/healthypeople/priority-areas/social-determinants-health
3. Galea S, Tracy M, Hoggatt KJ, DiMaggio C, Karpati A. Estimated deaths attributable to social factors in the United States. *Am J Public Health*. 2011;101(8):1456-1465. doi:10.2105/AJPH.2010.300086
4. McGinnis JM, Foege WH. Actual causes of death in the United States. *JAMA*. 1993;270(18):2207-2212. doi:10.1001/jama.1993.03510180077038
5. Chetty R, Stepner M, Abraham S, et al. The association between income and life expectancy in the United States, 2001-2014. *JAMA*. 2016;315(16):1750-1766. doi:10.1001/jama.2016.4226
6. Stringhini S, Carmeli C, Jokela M, et al. Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women. *Lancet*. 2017;389(10075):1229-1237. doi:10.1016/S0140-6736(16)32380-7
7. Bundy JD, Mills KT, He H, et al. Social determinants of health and premature death among adults in the USA from 1999 to 2018: a national cohort study. *Lancet Public Health*. 2023;8(6):e422-e431. doi:10.1016/S2468-2667(23)00081-6
8. Gardner JW, Sanborn JS. Years of potential life lost (YPLL): what does it measure? *Epidemiology*. 1990;1(4):322-329.
9. Martinez R, Soliz P, Caixeta R, Ordunez P. Reflection on modern methods: years of life lost due to premature mortality, a versatile and comprehensive measure for monitoring non-communicable disease mortality. *Int J Epidemiol*. 2019;48(4):1367-1376. doi:10.1093/ije/dyy254
10. Désesquelles A, Demuru E, Salvatore MA, et al. Analysis of multiple causes of death: a review of methods and practices. *Epidemiology*. 2023;34(3):333-344. doi:10.1097/EDE.0000000000001585
11. Guo Y, Chen Z, Xu K, et al. Utilization of social determinants of health ICD-10 Z-codes among hospitalized patients in the United States, 2016-2017. *Med Care*. 2020;58(12):1037-1043. doi:10.1097/MLR.0000000000001411
12. Kapilivsky, J. et al. Validation of a composite mortality end point in a large clinicogenomic real-world database of patients with advanced cancer, *JCO Clinical Cancer Informatics;* 2026;10(2). doi:10.1200/cci-25-00291.
13. Anderson RN, Rosenberg HM. *Age Standardization of Death Rates: Implementation of the Year 2000 Standard*. National Vital Statistics Reports. 1998;47(3):1-16. National Center for Health Statistics, Centers for Disease Control and Prevention.

Corresponding Authors

Chintan A. Dalal, PhD · Socially Determined, Inc. · chintandalal@sociallydetermined.com

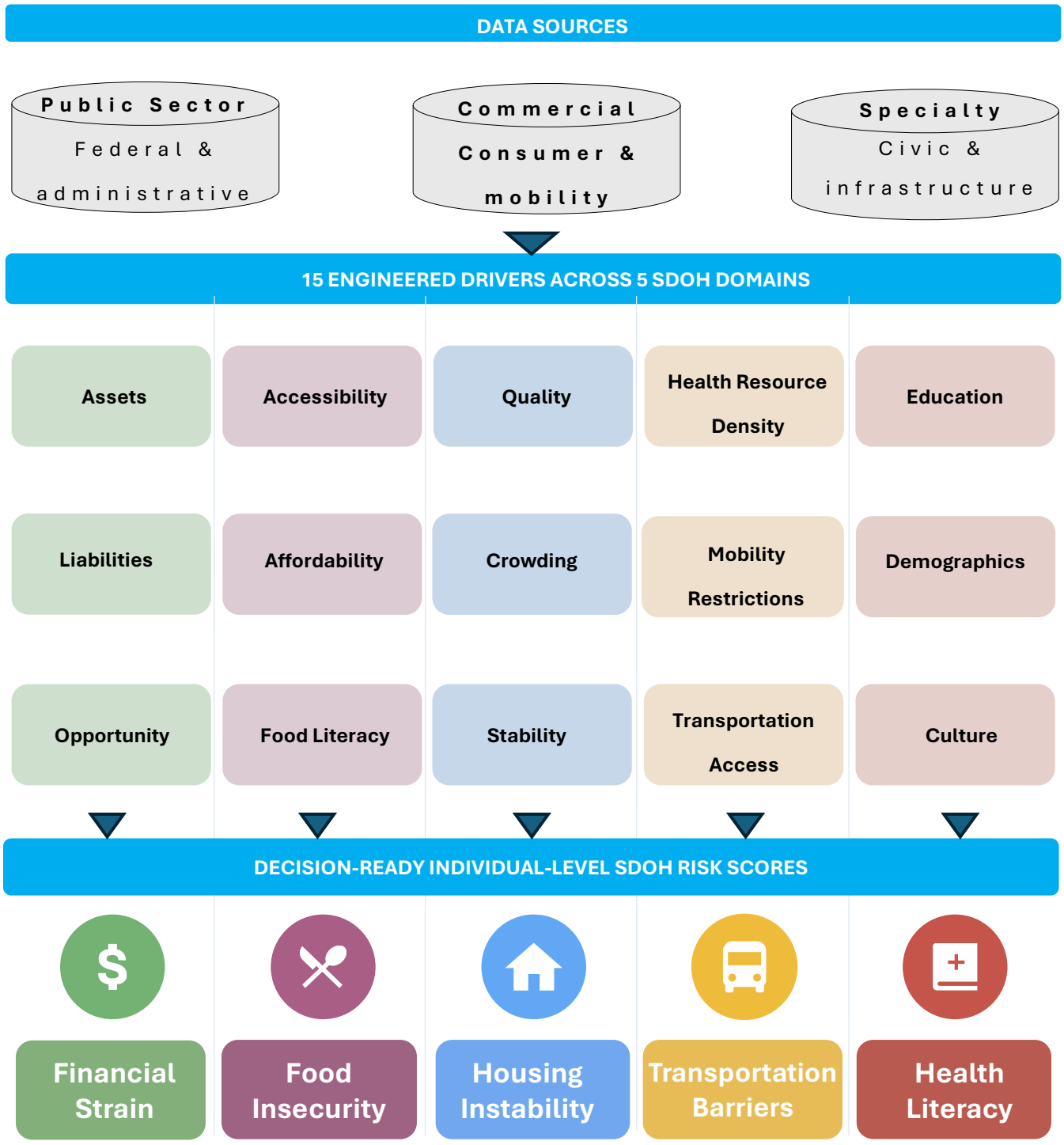
James Daniel McMichael · Socially Determined, Inc. · danmcmichael@sociallydetermined.com

Leah Broder, MPH · Socially Determined, Inc. · leahbroder@sociallydetermined.com

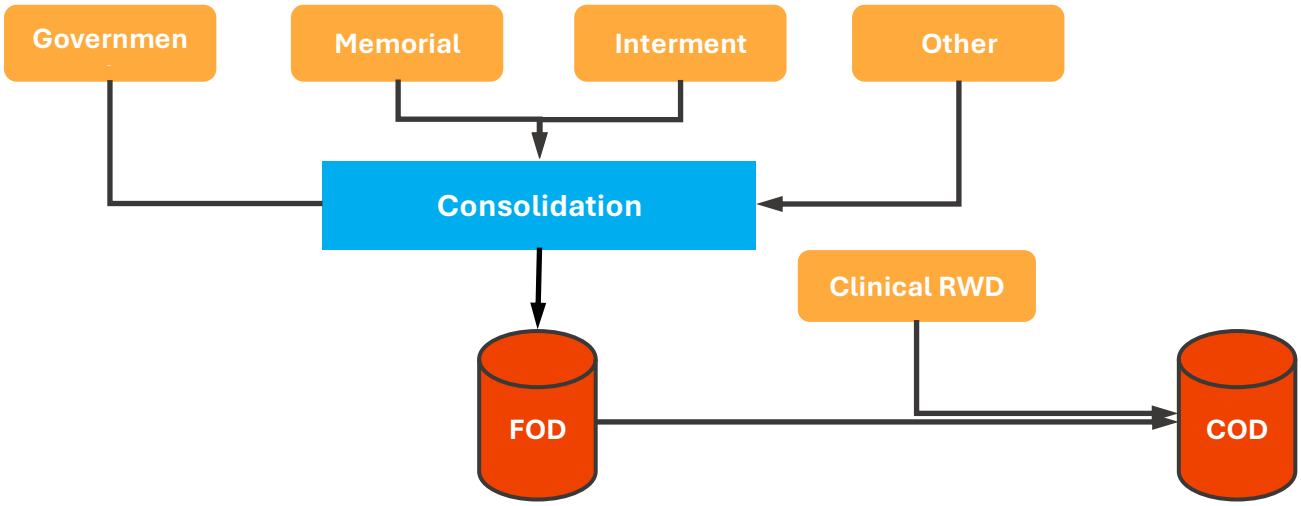
Kyle William McLean, PhD · Veritas Data Research, Inc. · kyle.mclean@veritasdataresearch.com

Appendix A. Data Sources and Population Summary

A.1 Socially Determined SDOH risk score construction



A.2 Veritas Fact-of-Death (FOD) and Cause-of-Death (COD) data pipeline



A.3 Cause-of-death groupings used in the analysis

Table A3. Cause-of-death groupings used in the analysis. The 27 most prevalent ICD-coded primary causes of death in the cohort were grouped into seven mutually exclusive clinical categories

Cause-of-death group	ICD-coded diagnoses descriptions included
Cancer	• Bronchus or lung, unspecified, malignant neoplasms • Malignant neoplasms
Cardiovascular and cerebrovascular	• Atrial fibrillation and flutter • Atherosclerotic heart disease • Cerebrovascular diseases • Cardiac arrest, unspecified • Diseases of heart • Essential, primary hypertension
Respiratory and pulmonary	• Chronic obstructive pulmonary disease, unspecified • Abnormal findings on diagnostic imaging of lung • Pleural effusion, not elsewhere classified • Acute respiratory failure • Dyspnoea • COVID-197 • Influenza and pneumonia
Cardiometabolic, renal, and endocrine	• Nephritis, nephrotic syndrome and nephrosis • Non-insulin-dependent diabetes mellitus, with other specified complications • Non-insulin-dependent diabetes mellitus, without complications • Acute renal failure, unspecified
Neurologic and cognitive	• Dysphagia • Other and unspecified symptoms and signs involving cognitive functions and awareness • Syncope and collapse
Infectious disease	• Urinary tract infection, site not specified • Septicaemia, unspecified

Cause-of-death group	ICD-coded diagnoses descriptions included
General decline and symptom burden	• Other specified disorders of muscle • Anaemia, unspecified • Malaise and fatigue

A.4 Cohort composition figures

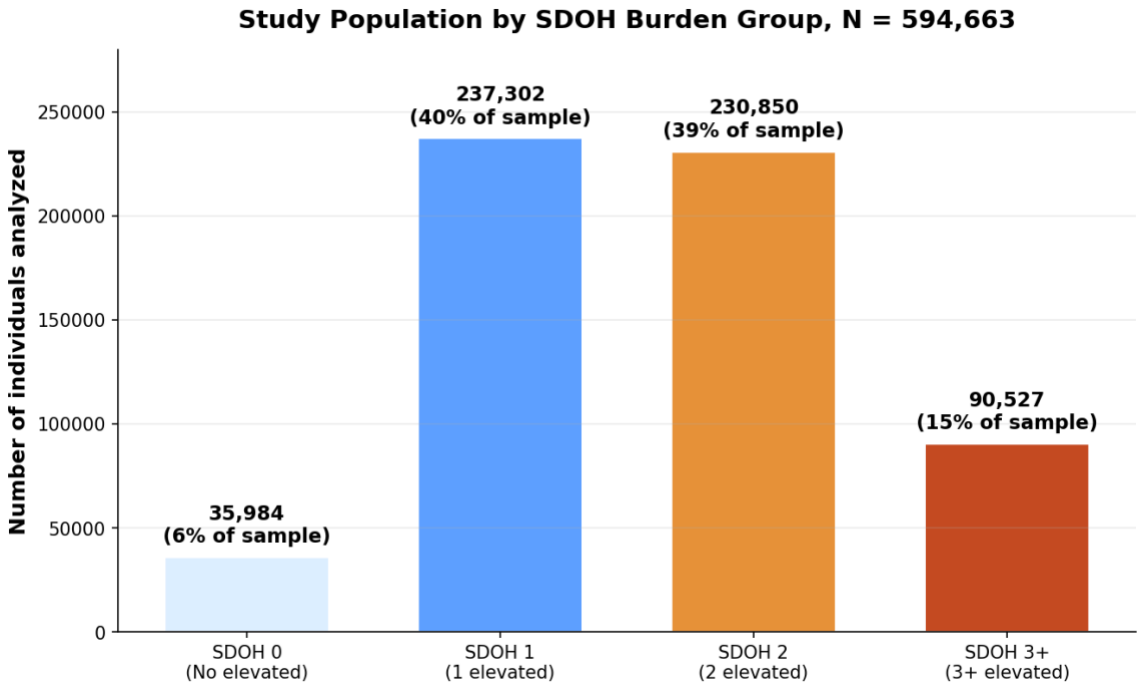


Figure A.4a. Study population by SDOH burden group (N = 594,663). Counts and within-cohort percentages for individuals with 0, 1, 2, or 3+ elevated SDOH domains.

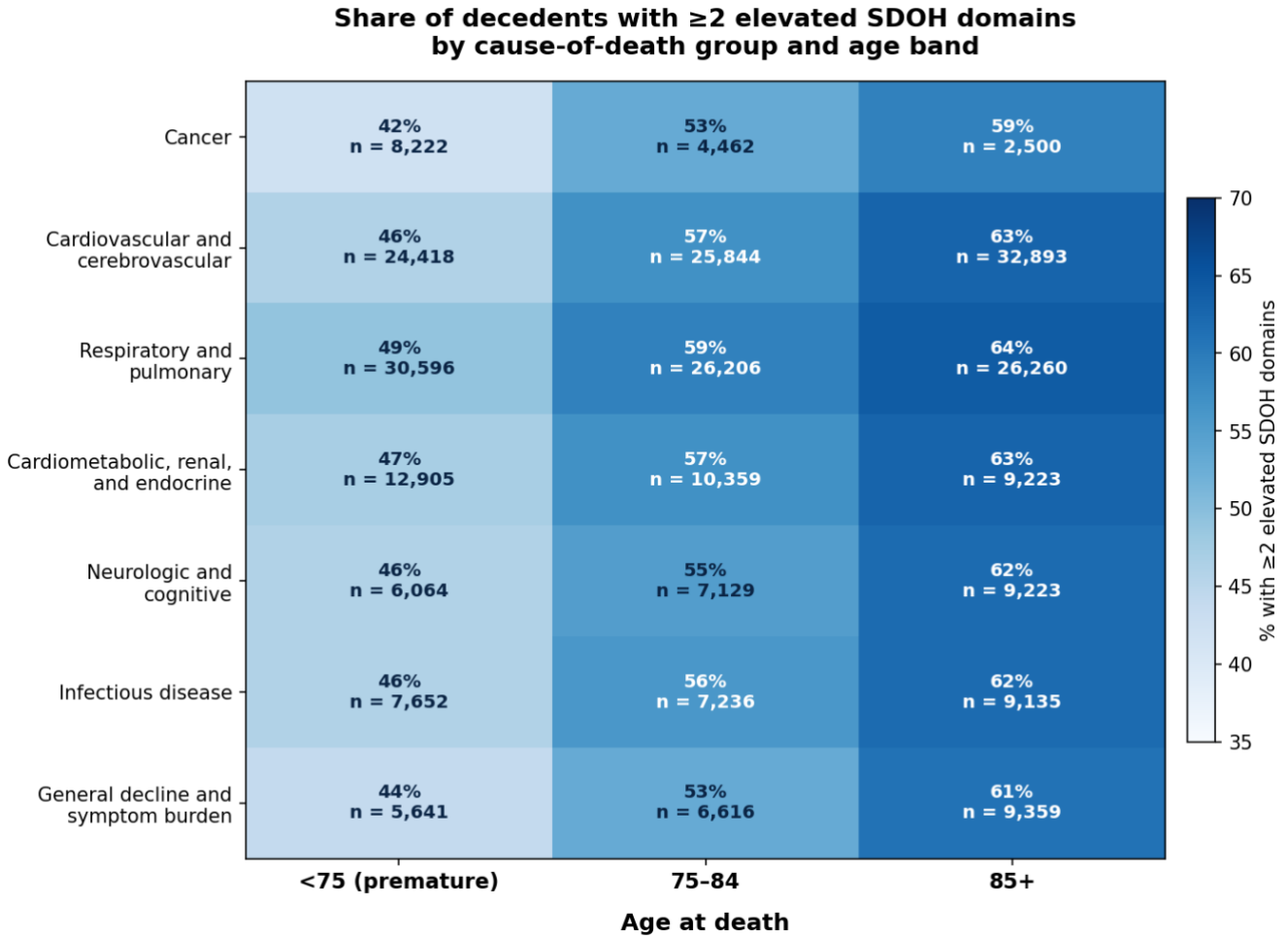


Figure A.4b. Share of decedents with ≥2 elevated SDOH domains, by cause-of-death group and age band. Sample sizes (n) shown in each cell confirm that every group is well-powered for stratified inference. The high-burden share rises monotonically with age within each cause-of-death group.

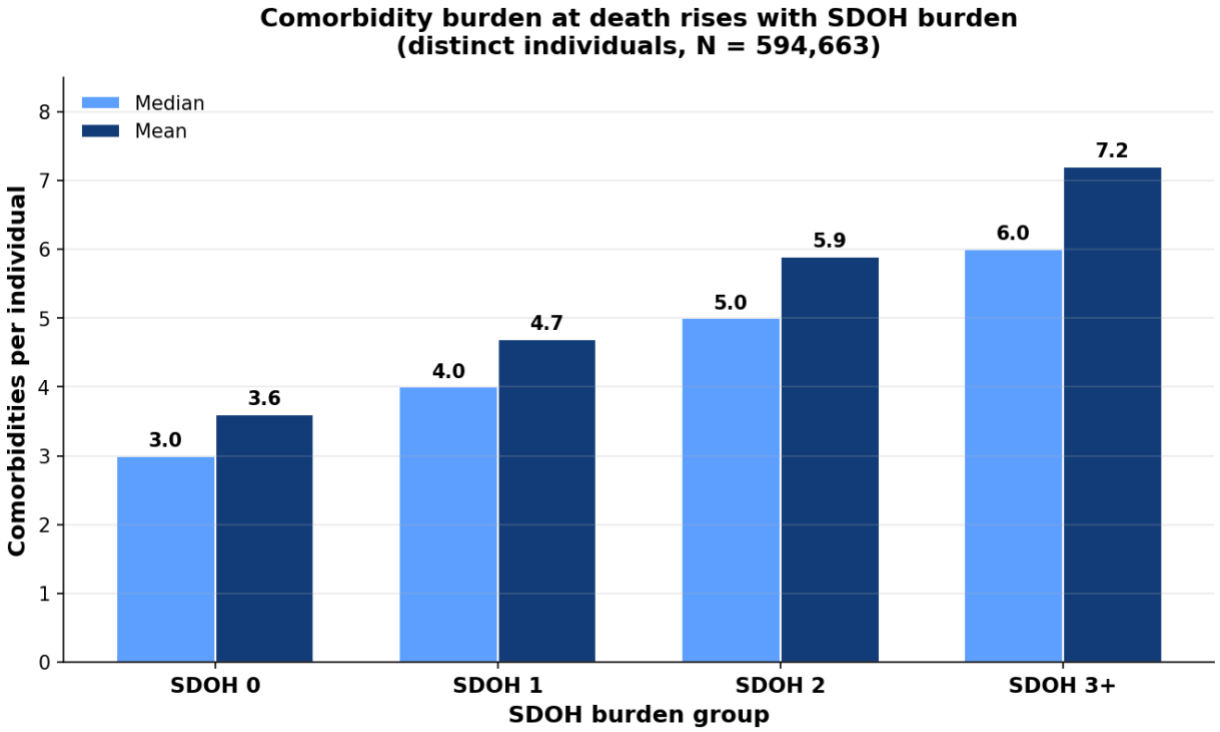


Figure A.4c. Comorbidity burden at death rises with SDOH burden. Median and mean number of comorbidities per individual, by SDOH burden group, in the analytic cohort (N = 594,663).

A.5 Supplementary features used in the driver-direction analysis

Table A.5. Supplementary features used in the directional driver analysis for the Decile 1, Decile 5, and Decile 10 personas (27 features across the five SDOH domains).

Domain	Description	Direction rule
Financial	Financial strain (composite)	Higher = adverse
Financial	Household income and poverty status	Higher = adverse
Financial	Economic opportunity index	Lower = adverse
Financial	Household income relative to poverty threshold (GIPAH)	Higher = adverse
Food	Food insecurity (composite)	Higher = adverse
Food	Healthy food balance	Lower = adverse
Food	Food swamp indicator	Higher = adverse
Food	Food desert indicator	Higher = adverse
Housing	Housing instability (composite)	Higher = adverse
Housing	Address volatility (frequency of address changes)	Higher = adverse
Housing	Household crowding index	Higher = adverse
Housing	Housing quality index	Lower = adverse
Housing	Household size	Higher = adverse
Housing	Occupancy score	Higher = adverse
Transportation	Transportation barriers (composite)	Higher = adverse
Transportation	Public transportation access index	Lower = adverse
Transportation	Health resources access index	Lower = adverse
Transportation	Provider desert indicator (general)	Higher = adverse
Transportation	Pharmacy desert indicator	Higher = adverse
Transportation	Hospital desert indicator	Higher = adverse
Transportation	Primary care provider desert indicator	Higher = adverse
Transportation	High-volume specialist desert indicator	Higher = adverse
Transportation	High-impact specialist desert indicator	Higher = adverse
Transportation	Behavioral-health specialist desert indicator	Higher = adverse
Health literacy	Health literacy challenges (composite)	Higher = adverse
Health literacy	Household primary language other than English	Higher = adverse
Health literacy	Lack of educational attainment	Higher = adverse