UT Southwestern Medical Center

Perelman School of Medicine, University of Pennsylvania

Informing Medicare with Evidence on Social Risk Adjustment

Technical Specifications

Prepared for:

The Patrick and Catherine Weldon Donaghue Medical Research Foundation

18 North Main Street

West Hartford, Connecticut 06107

Prepared by:

Danielle Browne, BS, Kano Amagai, MPH, Austin Kilaru, MD, MSHP,
Aidan Crowley, BS, Amol Navathe, MD, PhD, Joshua M. Liao, MD, MSc

March 2026

PROPEL

Penn | The Parity Center

Health Equity through Payment Reform

Purpose

This technical specification document outlines the proposed objectives and methodology of research work supported by the Patrick and Catherine Weldon Donaghue Medical Research Foundation. The specifications below will guide manuscript, conference presentation, and issue brief deliverables.

Background and Objectives

Study Context

The goal of this research is to fill critical knowledge gaps in value-based care and health equity with evidence about social risk adjustment, i.e., risk adjustment that accounts for social needs that may contribute to health outcomes. As the largest payer in the US, the Centers for Medicare & Medicaid Services (CMS) is leading such efforts by incorporating social risk adjustment into large nationwide payment programs using the Area Deprivation Index (ADI) – a composite, neighborhood-level measure of socioeconomic disadvantage that encompasses income, employment, education, housing, transportation, and other social determinants of health. Our work aims to fill Medicare-specific knowledge gaps on social risk adjustment to promote policy and practice change with empirical evidence.

Analytic Objectives

In a prior phase of our work, we demonstrated that after accounting for clinical factors, greater neighborhood socioeconomic disadvantage (measured by ADI) was associated with greater excess mortality, excess hospital admissions, and excess total health care spending. In this next phase of work, we will examine which specific dimensions of neighborhood socioeconomic disadvantage (i.e., individual components of ADI, such as median home value) underlie the observed associations between overall neighborhood social risk and patient outcomes (mortality or hospital admissions) and cost outcomes (per beneficiary per month spending) after adjustment for clinical factors. This approach will establish whether a composite approach, as opposed to component-based approach, to social risk adjustment is appropriate to augment traditional clinical risk adjustment strategies.

As part of this work, we will also evaluate recent modifications of the ADI, including standardization of the underlying component variables measuring neighborhood socioeconomic disadvantage.

Data Sources

Primary Datasets

The following sources are to be used: Medicare Fee-for-Service claims (inpatient, outpatient, carrier, hospice, durable medical equipment, home health agency); Medicare Beneficiary Summary File; 2023 Medicare Enrollment DataBase (EDB) file; 2018-2022 American

Community Survey 5-year Estimates (US Census Bureau); 2020 Neighborhood Atlas version 4.0.1. (University of Wisconsin); 2020 Rural-Urban Commuting Area (RUCA) codes (U.S. Department of Agriculture). The population and respective timeframe will include a 100% sample of Medicare Fee-for-Service beneficiaries continuously enrolled in Medicare Part A and Part B from January 2022 through January 2023.

Data Linkage

The methodology to complete the proposed work will entail combining and linking various datasets and variables. Procedures may include the following:

- Linking Medicare Beneficiary Summary File and Medicare Fee-for-Service claims using beneficiary ID
- Linking 9 digit-zip level ADI to the 9-digit zip-code of beneficiaries' residence listed on the EDB file
- Linking estimates from the American Community Survey based on Census Block Group to 9-digit zip using the 9-digit zip level ADI file, which is then linked to the 9-digit zip of the beneficiaries' residence listed on the EDB file
- Rural-Urban Commuting Area codes, based on 5-digit zip-code, to be linked to the first 5 digits of the beneficiaries' residence listed on the EDB file

Data Access & Permissions

Data permissions are granted through a Data Use Agreement between the University of Pennsylvania and the CMS Research Data Assistance Center (ResDAC). Data is accessible through ResDAC's Chronic Conditions Warehouse Virtual Research Data Center (VRDC). The VRDC satisfies all CMS privacy and security requirements and allows contracted users to access Research Identifiable Files through a secure cloud. To ensure the privacy and confidentiality of all project data accessed outside of the VRDC (e.g., ADI information), we store, use, and analyze identifiable data only on a secure server located in the Health Services Research Data Center, a high-security server center shared by the University of Pennsylvania School of Medicine and the University of Pennsylvania Health System. For more information, please visit: https://resdac.org/cms-virtual-research-data-center-vrdc.

Cohort Construction

Inclusion and Exclusion Criteria

Inclusion criteria are Medicare fee-for-service beneficiaries continuously enrolled in Medicare Part A and Part B from January 2022 through January 2023 (approximately 55M beneficiaries).

Exclusion criteria are Medicare beneficiaries enrolled in Medicare Advantage; beneficiaries with a gap in fee-for-service coverage in 2022 or with disenrollment reported not due to death; beneficiaries with end-stage-renal-disease or age greater than 110 years; beneficiaries who could not be distinguished as either fully or partially dual-eligible for Medicare and Medicaid

beneficiary claims with non-Medicare primary payer; beneficiaries missing geographic information.

For the spending outcomes cohort, in addition to the exclusion criteria listed above, we will remove Maryland residents (due to a state-specific payment model) and those with unavailable standardized cost information.

Index Dates

Outcomes are to be measured starting from January 2023. All baseline information is based on information collected in 2022. Observation time ends when the beneficiary leaves Medicare fee-for-service enrollment or the end of 2023 (whichever occurs first).

Key Variables and Measures

Exposure(s)

The primary exposure variables are the 17 individual components based on American Community Survey (ACS) Census data that make up the Area Deprivation Index (ADI) assigned to each beneficiary based on the 9-digit zip code of their residence listed in the Enrollment Data Base prior to January 2023.

- Percent of the block group's population aged ≥ 25 years with < 9 years of education
- Percent aged ≥ 25 years with greater than or equal to a high school diploma
- Percent of employed persons ≥16 years of age in white-collar occupations
- Median family income
- Income disparity
- Median home value
- Median gross rent
- Median monthly mortgage
- Percent owner-occupied housing units (home ownership rate)
- Percent of civilian labor force population ≥ 16 years of age unemployed (unemployment rate)
- Percent of families below the poverty level
- Percent of population below 150% of the poverty threshold
- Percent of single-parent households with children < 18 years of age
- Percent of occupied housing units without a motor vehicle
- Percent of occupied housing units without a telephone
- Percent of occupied housing units without complete plumbing
- Percent of occupied housing units with more than one person per room (crowding)

Outcome(s)

The primary outcome is individual-level mortality in 2023, using the Medicare Beneficiary Summary File to identify beneficiary death dates. Secondary outcomes include hospital admissions and total spending in 2023 using claims for inpatient, outpatient, physician, post-

acute care, home health, hospice, and durable medical equipment services. For total spending, we will adjust for the duration of individual Medicare fee-for-service coverage during 2023 to derive total spending per beneficiary per month. All spending outcomes are to be standardized to account for regional differences.

Covariates

Planned confounders and their operational definitions are listed below.

- Age: report continuously in years as measured in 2022
- Sex: Male or Female
- US region of residence: Northeast, South, Midwest, or West
- Rural-urban community area codes: Metropolitan, Micropolitan, Small town, Rural
- Dual eligibility for Medicaid and Medicare in 2022: partial dual, full dual, or non-dual
- Current reason for Medicare entitlement: old age, disability
- Original reason for Medicare entitlement: old age, disability
- CMS' Hierarchical Condition Categories (CMS-HCC) v28: 115 variables that group together ICD-10 diagnosis codes that share similar characteristics in terms of predicting healthcare costs
- 9-digit zip code: most recently reported beneficiary residence zip code as of 12/31/2022

Variable Construction Logic

CMS-HCC to be constructed using inpatient, outpatient, carrier, and skilled nursing facility claim files from 2022 and created using the logic listed in this website: https://www.cms.gov/medicare/payment/medicare-advantage-rates-statistics/risk-adjustment. Construction of standardized ADI using downloads of 5-year American Community Survey data from 2022 for Census block group, Census tract, and county. We will follow methodologies published in this article: https://academic.oup.com/healthaffairsscholar/article/1/5/qxad063/7342005.

Construction of the 17 ADI components using 5-year American Community Survey data from 2022 for census block group. We examined the components listed in this article: https://pmc.ncbi.nlm.nih.gov/articles/PMC1447923/

Analytic Approach

Study Design

The proposed work entails a series of cross-sectional analyses to compare outcomes among Medicare beneficiaries across specific components of ADI.

Primary Analysis Methods

Beneficiary-level ordinary least squares regression models will be used to estimate the association between each individual component of ADI and beneficiary mortality in 2023. There will be two models run for each individual component of ADI: one in which the exposure variable is standardized and one non-standardized (actual value). The models will be adjusted for CMS-HCC score (clinical risk).

Beneficiary-level Least Absolute Shrinkage and Selection Operator (LASSO) regression models with 5-fold cross-validation will be used to identify components most relevant to the relationship between ADI and mortality after accounting for HCC score.

Johnson's relative weight analysis will be used to quantify the contribution of each of the 17 ADI components to the variance of mortality.

To reduce dimensionality and identify socioeconomic domains represented by the 17 ADI components, we will perform variable clustering. All ADI components will be standardized prior to the analysis. Derived domain scores will be incorporate into an ordinary least squares regression model to examine the association between broader neighborhood socioeconomic domains and mortality after adjusting for CMS-HCC score.

Handling of Missing Data

Medicare beneficiaries missing geographic information will be excluded from our analysis as noted under exclusion criteria.

Sensitivity/Secondary Analyses

There are two secondary analyses planned: hospitalizations and spending. First, beneficiary-level ordinary least squares regression models may be used to estimate hospital admissions in 2023 and determine the relationship between individual components of ADI and hospitalization after adjusting for CMS-HCC score. Second, beneficiary-level ordinary least squares regression models may be used to estimate standardized spending per beneficiary per month in 2023 and determine the relationship between individual components of ADI and standardized spending after adjusting for CMS-HCC score.

For robustness, sensitivity analyses may include the following procedures:

- Repeat primary analysis using logistic regression models
- Stratify main analysis by dual-eligibility status
- Stratify main analysis by rurality (separate beneficiaries located in metropolitan 5-digit zip codes from those living in non-metropolitan areas)

Statistical Tools and Software

Analyses are to be performed with SAS version 9.4 and R (4.5.2).

Dissemination

Output and Deliverables

Planned outputs for the proposed bodies of work (e.g., manuscript(s), issue brief(s), conference presentation(s)) may include tables and figures that capture the following characteristics or information:

- Patient characteristics, overall
- Unadjusted and/or adjusted mortality, hospital admissions, and per beneficiary per month spending, by ADI component
- Domain score associations with mortality, hospitalization, and per beneficiary per month spending
- Table showing the relative weight for each ADI component and whether it was selected by LASSO or not

Public Sharing Plans

Platforms for dissemination may include the Leonard Davis Institute of Health Economics; University of Pennsylvania and University of Texas Southwestern Medical Center digital platforms. Additional dissemination activities may include real-time discussions, webinars, and/or policy briefings to disseminate findings with relevant stakeholders.