Life Tables by Nativity for the 65 and Older Population Using Linked Medicare Enrollment Data: A Feasibility Study

Shannon Sabo[1], Esther Miller[1], Elizabeth Arias[2], Heather King[1]

[1]Population Division
U.S. Census Bureau

[2]National Center for Health Statistics
Centers for Disease Control and Prevention

Abstract for submission to the annual meeting of the Population Association of America, Austin, TX, April 11-13th.

This abstract is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau or the National Center for Health Statistics.

Short Abstract

The U.S. Census Bureau's Population Estimates Program (PEP) and the National Center for Health Statistics (NCHS) are collaborating to develop nativity-specific life tables for the 65 and older population. This paper documents a feasibility study of using Medicare enrollment data linked to Social Security Administration data to produce the tables. We illustrate data challenges, present descriptive statistics, and show a series of tables and figures that describe enrollment status, retirement coverage, and other characteristics in the linked Medicare data. This exploratory analysis should lead to a refinement of the population at risk implied by the Medicare enrollment data, with a particular focus on the 90 and older ages. Using the results from this study, PEP and NCHS will develop nativity-specific life tables for the 65 and older population in future work. These tables will improve population estimates and provide a more accurate story of population health within the U.S.

Extended Abstract

Research has established that the foreign born have an advantage over native born residents in health and mortality in the United States (Dupre, Gu, Vaupel, 2012; Elo, Turra, Kestenbaum, 2004; Fenelon, Chinn, Anderson, 2017; Lariscy, Hummer, and Hayward, 2014; Mehta, Elo, Engelman, Lauderdale, Kestenbaum, 2016; Palloni and Arias, 2004; Preston and Elo, 2014; Singh and Hiatt, 2006; Turra and Elo, 2008; Turra and Goldman, 2007). The precise extent of the mortality advantage is unknown because official life tables by nativity are not produced by the National Center for Health Statistics (NCHS), the agency that publishes official vital statistics in the United States.

The U.S. Census Bureau's Population Estimates Program (PEP) and NCHS are collaborating to develop nativity-specific life tables. NCHS currently publishes life tables by sex for the total population, by selected race groups, and by Hispanic origin, where death rates in the oldest ages are modeled using information from aggregate Medicare enrollment data (Arias, Heron, Xu 2014). The life tables by nativity proposed here are based directly on linked person level Medicare data for the 65 and older population. To obtain nativity status, we link the Medicare enrollment data to data from the Social Security data Administration (SSA). The SSA data provide country of birth detail, from which we can infer nativity status. The final product will be a series of life tables by sex and nativity for the 65 and older population.

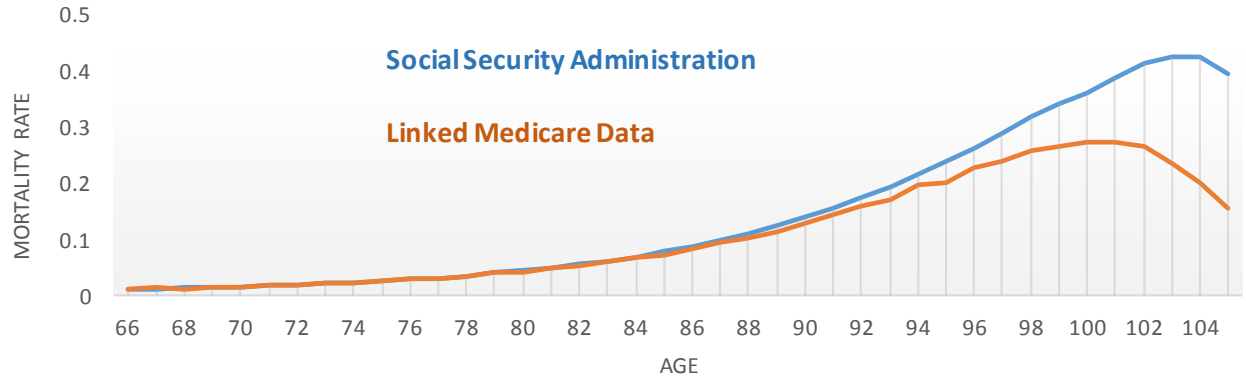
The aim of this paper is to document preliminary findings and conduct a feasibility study of using person level Medicare data, combined with Social Security Administration data, as a basis for producing nativity-specific life tables. We illustrate data challenges, present descriptive statistics, and show a series of tables and figures that describe enrollment status, retirement coverage, and other characteristics in the linked Medicare data. This exploratory analysis should lead to a refinement of the population at risk implied by the Medicare enrollment data, with a particular focus on the 90 and older ages.

The joint effort between PEP and NCHS to produce nativity-specific life tables has revealed issues in the Medicare enrollment data that are particularly pronounced for the population age 90 and older. Namely, the population at risk appears to contain an excessive number of censored records, where death status is

unknown or unrecorded. The Social Security Administration (SSA) refers to these records in their files as "phantom records" in their tabulations of life expectancy (Bell and Miller 2005).

In this paper, we seek to similarly remove "phantom records" from the linked Medicare data using inclusion criteria based on Medicare variables that describe benefit status and retirement insurance. Including "phantom records" inflates the population at risk and distorts and dampens the age-specific mortality at the oldest ages. See Figure 1 for this effect on female age-specific mortality rates in 2014 using preliminary tabulations of the linked Medicare enrollment data compared to the SSA's mortality rates for the same population. Using the SSA rates as a benchmark, the linked Medicare-based mortality rates seen here require additional refinement and study.

Figure 1. Age-specific Mortality Rates in 2014 by Data Source: Females, 65 and older



As shown in Figure 1, our initial attempts at producing nativity-specific life tables have revealed issues in the quality of the linked Medicare data for the population age 90 and older. Namely, the population at risk appears to contain an excessive number of censored records, or "phantom records," as SSA refers to them. According to the SSA, plausible Medicare-based death rates in the oldest ages are obtained by limiting the universe of Medicare enrollees who have current retirement insurance, either through Social Security, the Railroad Retirement Board, or government pensions. This eliminates approximately 3% of the Medicare records (Bell and Murphy 2005). We aim to replicate SSA's procedure of removing "phantom records" by using some combination of retirement insurance/benefits fields in the linked Medicare data.

Data and Methods

Before we can confidently produce life tables by nativity using the Medicare linked data, we start by examining the plausibility of death rates for the total population using these data. After counts of deaths and age-specific death rates by sex for the total population are validated, we can examine these same indicators by nativity. We use annual death counts from NCHS and age-specific rates produced by the SSA from calendar year 2014 to benchmark our results for the total population 65 and older. We test differing sets of inclusion criteria in the linked Medicare data using variables on Medicare benefits status and retirement insurance status. We will select the final inclusion criteria based on the plausibility of the resulting age-specific death rates (i.e. improvements on Figure 1).

Death rates in this linked study are derived from two data sources. The population at risk is comprised data from two administrative record data sources: the Medicare Enrollment Database Full File (MEDB) linked to the Numerical Identification System (Numident) from the SSA. Death counts are derived from reported deaths in the MEDB.

Medicare Enrollment Database Full File (MEDB). This file is obtained from the Centers for Medicare and Medicaid Services (CMS) and used by PEP for other purposes.[1] The data include date of birth, date of death, sex, race, mailing address, and information on Medicare benefits and retirement insurance. Of importance in this study are 8 variables on Medicare status and retirement insurance.

Numident. The Numident is a comprehensive administrative record file maintained by the Social Security Administration. The file contains information for every Social Security Number (SSN) ever issued beginning in November 1936. Data items include an individual's name, social security number, sex, date of birth, date of death, and most importantly for this research, place of birth, including foreign countries used to assign nativity. Census replaces enrollee personally identifiable information with a Protected Identification Key

[1] PEP uses the MEDB to estimate migration patterns for the 65 and older population in the annual population estimates and to examine coverage by the decennial census in the Demographic Analysis program in the oldest ages.

(PIK). The PIK is a unique identifier and allows us to link information about individuals across multiple sources of administrative data.

Expected Results

We aim to replicate SSA's procedure of removing "phantom records" by using some combination of retirement insurance/benefits fields in the linked Medicare data. However, this issue is more nuanced than it appears due to the many categories in each pertinent variable. For example, there are over 150 categories in the variable that describes entitlement to benefits. An enrollee may qualify through his/her own work experience, through a spouse or some other relative, or through a legal statute.[2] We will determine the appropriate inclusion criteria that removes these "phantom records" from our linked Medicare data by evaluating the plausibility of the resulting deaths rates by age and sex in the 90 and older ages.

Conclusions

This project highlights the issues of working with numerous linked administrative data sources to create plausible demographic rates and indicators. The data challenges shown here can benefit other researchers interested in using administrative data for these purposes. Preliminary results show that our linked Medicare data is overestimating the population at the oldest ages. However, through benchmarking, a scrupulous exploratory analysis of the data items, and interagency collaboration, these data challenges may be more fully understood and eventually mitigated. After this preliminary work is completed, we will produce nativity-specific life tables in future work. The results from this study will help us better understand mortality at the oldest ages and enhances two major programs at the Census Bureau: (1) the 2020 Demographic Analysis program, one of two official ways the Bureau evaluates the coverage of each decennial census; and (2) the annual population estimates published by PEP. And perhaps most importantly, this study may lead to the first full set of nativity-specific life tables produced by NCHS, to the benefit of demographers, social scientists, and the general public alike.

[2] For example, some special transition rules were enacted in 1966 that approved benefits to those who did meet work requirements and were at least 72 years old.

References

Arias E, Heron M, Xu JQ. United States life tables, 2014. National vital statistics reports; vol 66(4). Hyattsville, MD: National Center for Health Statistics. 2017.

Bell FC, and ML Miller. 2005. "Life Tables for the United States Social Security Area 1900-2100." Actuarial Study no. 120. Social Security Administration.

Dupre ME, D Gu, and JW Vaupel. 2012. "Survival Differences among native-born and foreign-born older adults in the United States". PLoS One, 7(5) E37177. Online at http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0037177

Elo I, Turra CM, Kestenbaum B, Ferguson, BR. 2004. Mortality among elderly Hispanics in the United States: Past evidence and new results." *Demography*. 48, 241-65.

Fenelon A, JJ Chinn, and RN Anderson. 2017. "A Comprehensive Analysis of the Mortality Expereince of Hispanic Subgroups in the United States: Variation by Age, Country of Origin, and Nativity." *Social Science and Medicine—Population Health* 3:217-54.

Lariscy JT, Hummer RA, Hayward MD. 2015. "Hispanic older adult mortality in the United States: New Estimates and an assessment of factors shaping the Hispanic Paradox." *Demography* 52: 1-14.

Mehta N, Elo IT, Engelman M, Lauderdale DS, Kestenbaum BM. 2016. "Life expectancy among U.S.-born and foreign-born older adults in the United States: Estimates from linked social security and Medicare data." *Demography* 53: 1109-34.

Palloni A, Arias E. 2004. "Paradox lost: Explaining the Hispanic adult mortality advantage." *Demography* 41(3): 385-415.

Preston SH, IT Elo. 2014. "Anatomy of a municipal triumph: New York's upsurge in life expectancy." *Population and Development Review* 40: 1-29.

Singh GK, and Hiatt RA. 2006. "Trends and disparities in socioeconomic and behavioural characteristics, life expectancy, and cause-specific mortality of native-born and foreign-born populations in the United States, 1979-2003." *International Journal of Epidemiology,* 35, 903-919.

Turra CM, Goldman N. 2007. "Socioeconomic differences in mortality among US adults: insights into the Hispanic paradox." *The Journals of Gerontology*. Series B 62(3): S184-S192.

Turra CM, Elo IT. 2008. "The impact of the salmon-bias on the Hispanic mortality advantage: new evidence from social security data." *Population Research and Policy Review*. 27: 515-30.