Studying Health Effects based on Medicare Data: from Empirical Estimates to Causal Inference

I. Akushevich[1], J. Kravchenko[2]

[1]Center for Population Health and Aging, Duke University, Durham, NC

[2]Department of Surgery, Duke University Medical Center, Durham, NC

INTRODUCTION

Determining the national trends in health and vital status of a growing sector of the United States (U.S.) elderly population is a major issue for public health specialists, policymakers, and governmental institutions. To better address the health demands in the elderly and to reduce economic burdens on society, it is important to understand the key factors driving the onset and progression of aging-related diseases. An identification of age patterns of disease incidence, their time trends, and mortality in general population and patients cohorts with sufficient precision requires large population-based databases that are costly to collect. This is why there are few studies on age patterns of diseases, diseases interrelations, and potentially impacting factors in the U.S. in elderly population. As a reasonable alternative, the studies based on observational data could be suggested. In observational studies, the assignment of subjects into a treated group versus a control group is not controlled by investigator; therefore, a special attention has to be paid to statistical methodologies and interpretation of the findings. Information collected in the Medicare Files of Service Use (MFSU) for entire Medicare-eligible population of older adults is an example of observational administrative data. These data, as any administrative health data, are generated through the routine health care programs. Thus, the development of approaches for analyses of Medicare data and their applications for discovery of substantive bio-demographic and clinical results for the US elderly population is a well timed activity that is largely motivated by the lack of such comprehensive and representative analyses at a national level. Uncovering the patterns of disease incidence, risk factors, multimorbidity, and recovery among the patients at advanced ages is very valuable both in theoretical aspect of understanding the interaction of disease incidence and senescence as well as for the practical implementations of evaluated health trends in a forecasting of Medicare expenditures. In this context, the tremendous research potential of the Big Data on health—such as the data from the MFSU—for studying the current and forecasting future health patterns of the US older adults need to be explored in its full.

In this report, we focus on a series of epidemiologic and bio-demographic measures that can be studied using MFSU. In these analyses, we do not make a pre-selection of individuals in Medicare-based datasets to have an opportunity to investigate the demographic and epidemiologic properties of the general elderly population and cohorts of survivors after certain chronic diseases were diagnosed. Also, we discuss how the methods of causal inference can be applied for analysis of time dependent treatment for patient with cancer diagnosis.

Data: SEER-Medicare and NLTCS-Medicare

Two datasets that provide information on health of the older adults at the national level are used in this study: the Surveillance, Epidemiology, and End Results (SEER) Registry data linked to MFSU (SEER-M), and the National Long Term Care Survey (NLTCS) that is also linked to the MFSU (NLTCS-M).

The SEER-M data is the primary dataset that was analyzed in this study. The expanded SEER registry covers approximately 26% of the U.S. population. In total, the Medicare records for 2,154,598 individuals are available in SEER-M; they include individuals i) with common diagnosed cancers such as breast (n=353,285), colon (n=222,659), lung (n=342,961), prostate (n=448,410), and skin melanoma (n=101,123); and ii) from a random 5% sample of Medicare beneficiaries residing in the SEER areas who had none of the abovementioned cancers. For the majority of persons, we have continuous records of Medicare services use from 1991 (or from the time the person has passed the age of 65 after 1990) until the time of his/her death. Also, a small fraction of individuals (e.g., new patients diagnosed with cancer in 2003-2005) has Medicare records from 1998.

The NLTCS-M data contains two of the six NLTCS waves–namely, cohorts of 1994 and 1999. These two waves were chosen primarily because of the high-quality Medicare follow-up data available from 1991, and also because the complete 5-year follow-up after the NLTCS interview is accessible only for these two waves since 1991. In total, 34,077 individuals were followed-up between 1994 and 1999.

In both datasets (i.e., SEER-M and NLTCS-M) Medicare records are available for each institutional (inpatient, outpatient, hospice, or home health agency) and non-institutional (carrier-physician-supplier and durable medical equipment providers) claim type. So-called screener weights released with the NLTCS allowed for the production of the national population estimates.

The NLTCS-Medicare represents the entire US elderly population while the SEER-Medicare represents the population of SEER areas only; therefore, SEER-M represents US general population only approximately. The age and sex distribution of the total SEER population is similar to non-SEER areas, though SEER areas have less whites, more urban residents, and less poor areas compared to non-SEER areas.

Results

Using these extensive sources of information allows for identification of disease incidence and recovery (or long-term remission) events through elaboration and validation of specific computational algorithms of these events from administrative data. We discuss the spectrum of the results obtained using the MFSU and the data linked with Medicare files.

First, we analyzed morbidity and mortality patterns. We discussed the methods of the calculation of the age patterns of age-related disease incidence and also present the results that were estimated using NLTCS-Medicare and SEER-Medicare data. The performed analyses of chronic diseases in elderly using NLTCS-M and SEER-M data suggested that, the national age-specific incidence patterns can be adequately evaluated from the Medicare data. The majority of studied diseases (e.g., prostate cancer, asthma, and diabetes mellitus) had a monotonic decline (or decline after a short period of increase) in incidence with age. A monotonic increase in incidence with age with a subsequent leveling off and decline was observed for myocardial infarction, stroke, heart failure, ulcer, and Alzheimer's disease. An inverted U-shaped age pattern was detected for lung and colon carcinomas, Parkinson's disease, and renal failure. Among individuals with severe disabilities there were higher rates of stroke, heart failure (males), diabetes, asthma, and Parkinson's disease, while rates of breast and prostate cancers were higher for nondisabled or moderately disabled individuals

Second, we investigated the phenomenon of recovery or long-term remission for geriatric diseases with specific focus on the cohorts of patients after disease onsets. The main research question was whether the patients who stopped visiting doctors are really healthier than the rest

of the patients in the cohort and whether they could be considered as recovered patients. We found that the rates of a long-term remission and recovery from ageing-associated diseases and their time trends are detectable using Medicare data, and the patients who had periods lacking in ICD-9 records are the healthier subcohort.

Third, time trends of incidence of diseases common in the US elderly population were evaluated. These trends are associated with changes in socioeconomic status and demographic structure of population, risk factors prevalence (e.g., smoking, obesity, etc.), as well as with the changes in prevention, screening, and diagnostic strategies. The estimates of the time trends become especially important in the populations with a growing proportion of the elderly for whom to maintain a good health is an important issue. We evaluated time trends of incidence of diseases common in the US elderly population using MFSU (Figure 1). The results showed dramatic increase of incidence rates of melanoma, goiter, chronic renal, and Alzheimer's disease in 1992-2005. Besides specifying widely recognized time trends on age-associated diseases, new information was obtained for trends of asthma, ulcer, and goiter among the older adults in the US

Fourth, we investigated the phenomenon of multimorbidity in the US elderly population by analyzing the mutual dependence in disease risks, i.e., calculate the disease risks for individuals with specific pre-existing conditions. Although multimorbidity is common among the older adults, however, for many chronic diseases no information is available for U.S. elderly population on how earlier-manifested disease can affect the risk of another disease that manifests later in life. Synergistic and antagonistic dependences in geriatric disease risks were observed among US elderly confirming the known and detecting new associations of a wide spectrum of age-associated diseases (Figure 2). The existence of inverse associations for the later-in-life diagnosed disease risk may provide important insights into disease mechanisms and also new opportunities for disease prevention and therapy, allowing to focus on increase in healthy lifespan of the patient rather than concentrating the efforts on reduction of risk for each particular disease separately.

Fifth, we analyzed the group of lung cancer patients. We demonstrated that the observational data are particularly valuable when practical considerations limit the ability to perform randomized clinical trials to address the questions about the impact of different lung cancer treatments on survival. We applied the Marginal Structural Model (MSM)—currently, the most successful known statistical technology that is capable to address the problem. This method uses the inverse probability weight approach to evaluate individual (stabilized) weights and then evaluates the effects of time-dependent treatments within a weighted repeated measure approach. We apply MSM to SEER-Medicare data to study causal effects of treatments (surgery, radiation, and/or chemotherapy or absence of treatment) on lung cancer survival given individual patient tumor characteristics, comorbidities, and demographic and socioeconomic factors. Specific attention was paid to dynamic interrelations of treatment and comorbidities in that comorbidity impacts both treatment choice and effectiveness, while cancer therapy can aggravate co-existing conditions.

The results of all five directions of analyses will be presented and discussed at the Annual Meeting of Population Association of America (PAA) 2014. Figure 1 and 2 demonstrate the example of the results obtained for time-trends and mutual disease risks.

Discussion and Conclusion

In this report, we presented a spectrum of Medicare-based analyses to clarify multiple aspects in biomedical demography of US population of older adults. All these results demonstrate the usefulness of Medicare in filling gaps in quantitative knowledge about health effects at

advanced ages in the nationally representative population. For example, heart disease and stroke account for more than 40% of all deaths among persons aged 65 to 74 years, and almost 60% of those aged 85 years and older; however, there are no nationally representative data available on incidence, severity, or recurrence of acute coronary or stroke events in either the inpatient or outpatient settings, with the performance measures which are not consistent across databases. Therefore, Medicare based datasets could be very useful in estimating epidemiology and bio-demography of aging-related diseases and associated medical costs in the U.S. elderly population.

Specific attention was paid to the application of MSM to SEER-Medicare lung cancer cases for counterfactual time-dependent treatments allowed us to estimate causal treatment effects on overall and cancer-specific survival. This model considers both time-dependent treatment and comorbidity—two important factors affecting patients' survival that are commonly interrelated. The obtained results are clinically important, because clinicians frequently have to make difficult therapeutic decisions regarding lung cancer, especially among older adults, where the benefits of curative and life-prolonging therapy are balanced with potential mortality due to the presence or development of other comorbid conditions. Unfortunately, the existing guidelines do not provide detailed information that helps to make these difficult decisions. The presence and severity of comorbid conditions are generally known to increase the risk of treatment toxicity, but the data that more specifically guides the therapy is severely lacking and the treatment is essentially guided by a subjective clinical judgment, i.e., on an individual case basis. With the proposed study we addressed this knowledge gap by developing a novel method to assess a comparative effectiveness of the treatments. In particular, we focused on the impact of different treatment modalities on lung cancer specific and overall survival by targeting different groups of the elderly patients based on dynamic characteristics of comorbidities and treatment using a comprehensive analysis of SEER-Medicare data. The results should assist clinicians in choosing optimal lung cancer treatments that will balance treatment risks and benefits based on personalized patient characteristics.

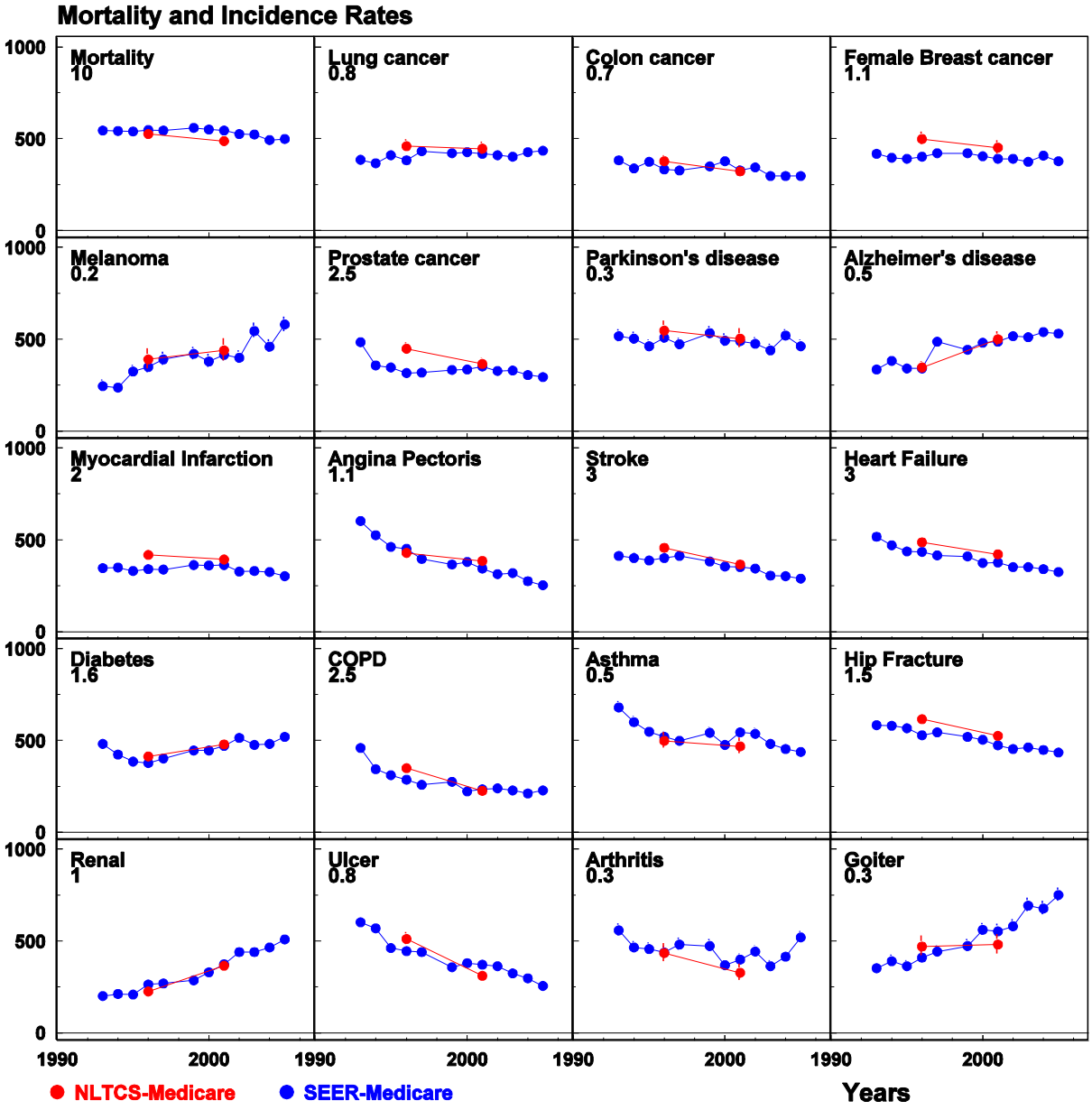

Figure 1. Time patterns of age-adjusted rates of total mortality and disease incidence calculated using NLTCS-Medicare (red dots) and SEER-Medicare (blue dots). Values on plots are rescaled factors. Rates for different diseases are rescaled to use the same scale on all plots to compare rates for different diseases: the original rate can be calculated by dividing the values obtained from plot by the rescaled factor.

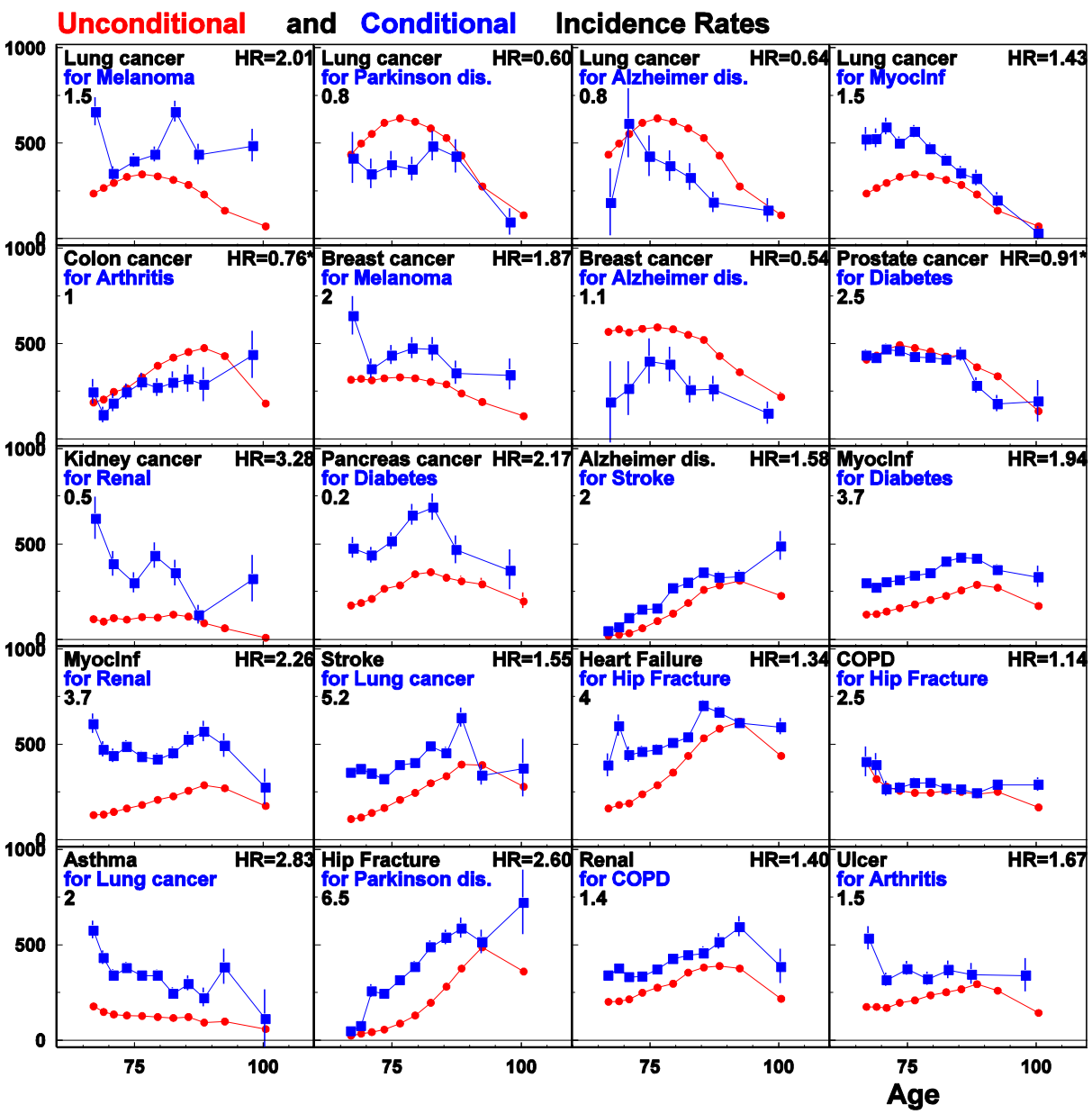

Figure 2. Age-specific rates of disease incidence conditional on the onset of another disease calculated using SEER-Medicare (blue dots). Respective unconditional rates (red dots). Values on plots are rescaled factors. Rates for different diseases are rescaled to use the same scale on all plots to compare rates for different diseases: the original rate can be calculated by dividing the values obtained from plot by the rescaled factor.