Influences of morbidity, lifestyle, and functional status on health risks in the U.S. elderly population.

Igor Akushevich, Lucy Akushevich, Julia Kravchenko, Svetlana V. Ukraintseva, Konstantin G. Arbeev, Anatoliy I. Yashin

Duke University, Trent Hall, Room 002, Box 90408, Durham, NC, 27708-0408, USA.

Extended Abstract

Introduction

Information collected in a large scale of population-based studies can be used to quantify possible effects of lifestyle, functional status of elderly, and self-reported morbidities on the risk of aging-associated conditions/events representing by the risks of mortality, developing chronic diseases (cancer and non-cancer), and survival after acute conditions like myocardial infarction and stroke.

Epidemiologic and demographic studies performed during the last decades demonstrated that multiple factors of distinct nature contribute to mortality risk. However, many important questions still remain unanswered: how these factors interact to predict the mortality risk, in what extent the results of these studies can be used for population forecasting, what has to be developed further to use the established relationships for providing individualized prognoses, and how these findings can be joint with needs of evaluating time trends of the characteristics and predicting future economic costs of agencies such as Medicare.

Health status of elderly population is associated with subclinical and clinical chronic diseases and subsequent disability. Most of the chronic diseases are considered to be associated with multiple risk factors, and many of the determining factors are measurable. If known, many of these risk factors can be potentially preventable, e.g., about 80% of cancer risk factors, including behavioral/lifestyle-associated are considered preventable. Combining information in several datasets, such as demographic surveys and Medicare Service use files, allows for testing of the hypotheses about the contributions of specific risk factors, their groups, and aggregated indices to risk of aging-related chronic diseases.

Besides, lifestyle and other measurable characteristics can determine the risk of death in early and late post-acute periods of myocardial infarction and stroke. A lot of information about trends and determinants in circulatory diseases was provided by the World Health Organization MONICA Project. However, there was also formulated a set of research questions that awaits analysis by the WHO MONICA Project and maybe other studies in the future. Which component of declining mortality will be carried forward into subsequent decades? Can lifestyle influence case fatality? Can medical care influence coronary-event rates? These questions can also be analyzed using population-based studies collected in the U.S.

In this report we present an approach to estimate the contributions of a spectrum of various measurable risk factors, including behavioral/lifestyle risk factors, to mortality, aging-related chronic diseases incidence risk, and mortality after acute circulatory conditions in the U.S. elderly population.

Data and Methods.

The four sources of data used are 1) nationally representative National Long Term Care Survey (NLTCS) - for measuring the functional status, behavioral factors, and other characteristics potentially determining the health state in the elderly, 2) Medicare Claim files linked to each person from the NLTCS – for estimating the date of onset for chronic diseases, individual survival after acute conditions, and date of death (if applicable) in the U.S. population, 3) SEER data used as a "gold standard" for comparison of SEER age patterns of selected cancers with age patterns predicted through using NLTCS-Medicare data; and 4) SEER-Medicare data: Medicare Claim files linked to each case in SEER data and additionally included Medicare files for a random sample of 5% of Medicare beneficiaries residing in the SEER areas who do not have cancer. The latter will be used for specifying algorithms for evaluating the cancer date of onset. The 1994 and 1999 NLTCS were analyzed and, in total, more than 200 variables for each survey were selected and grouped in fourteen groups, including daily living activities, physical activities, smoking, alcohol consumption, social activities, self-reported comorbidity, health insurance, medical providers, and others. Elaborated characteristics were tested for association with selected cancer and non-cancer aging-associated chronic disease incidences during the following 5 years since the date of interview.

Twelve chronic conditions/diseases were selected for the investigation of the association of their incidence with the measured characteristics. They include acute coronary heart disease (ACHD), (ICD-9-CM: 410, 411, 413), stroke (431,436, 433.*1, 434.*1), ulcer (531-534), breast

cancer (174), prostate cancer (185), lung cancer (162), and colon cancer (153), melanoma (172), diabetes (250), asthma (493), Parkinson's disease (332), and Alzheimer's disease (331.0, 290.1). Cancers listed above were selected for analyses due to their high incidence rates and because their incidence rates can be reconstructed from Medicare data relatively well.

Several approaches were tested to elaborate the measures in each group and aggregated scores. The first is the logistic regression for each considered risk (i.e., for mortality, selected chronic disease incidences, and case fatality of acute circulatory conditions) and explanatory variables selected using stepwise technique allowing for extracting specific variables in considered groups or in the whole sample, which are most predictive in respect to the considered risk. This approach provides reasonable results for the risk of mortality. However, reduced statistical power, correlation in explanatory variables, and uncertainties in evaluating the date of onset/manifestation for chronic diseases require some preliminary aggregation of the explanatory variables when other risks are considered. The second approach deals with the aggregation of measured information by statistical methods. In this approach the aggregated scores are obtained by factor analysis methods to extract one or several factors from each group, and the latent linear structure analysis (the newest version of Grade of Membership analysis) approaches which was recently generalized for using high-dimensional datasets of categorical variables. The third approach deals with substantive approaches for aggregation of indices in i) each group and ii) across the groups. The following groups of aggregated indices have been used: self-reported diseases/conditions, demographic characteristics (related and non-related to individual's genotype), everyday living physical activity, exercising/sports physical activities, nutrition, consumption of tobacco/alcohol/drugs, self-estimated satisfaction with life quality and health, emotional/mental status, social activities, types and sources of medical care, types and sources of social support, etc.

Quality of descriptions (e.g., represented by the area under the receiver operating characteristic, i.e., ROC curve) is compared across all methods and all considered risks.

Results and Discussion.

Substantively, the results of analyses showed associations between all considered risks with characteristics representing distinct features of human aging. Risk of mortality was analyzed using the approach where all variables are used without constructing indices. So, it has been shown that physical activity (e.g., swimming, walking exercises, weight lifting), normal range of BMI, absence of majority of chronic diseases, higher social activity, and better coverage of medical insurances were associated with decreased cancer mortality risk; while the male gender,

smoking, disability, mental disorders, and severe obesity were associated with increased cancer morality risk. Apart from these expected effects, several unexpected results were obtained, e.g., frequent headaches and frequent flu are associated with lower mortality risk. The risk of incidence for chronic diseases and the risk of death 28 days after onset of acute circulatory conditions were analyzed using all three approaches, i.e., keeping significant contributions of original variables and constructing generalized indices statistically and substantially.

Methodologically, the results will allow us to evaluate several popular approaches of dealing with high dimensional categorical measurements. Specifically, this uncovers the conditions where substantive approaches to aggregating separate categorical measurements are (not) preferable comparing to approaches where such an aggregation is performed using pure statistical approaches not containing biological, medical, or demographic assumptions.

One result of this analysis is a quantitative description of higher cancer risk groups in terms of variables measured in NLTCS and aggregated indices constructed from these variables. Groups of parameters of physical activity, tobacco consumption, comorbid conditions, demographic characteristics, and health insurance and medical care providers showed significant contributions in increasing or decreasing the risk of mortality, incidence of chronic disease, and case fatality of acute circulatory conditions. Contribution of groups of factors, and of individual factors inside the group varied depending on the considered case, e.g., smoking and pulmonary comorbidity were strongly associated with lung cancer incidence risk, and physical activity was associated with breast cancer risk. The performed sensitivity analyses included the impact of the scheme for filling missing data and the comparison of the results obtained from different algorithms of cancer onset reconstruction.

Measured characteristics representing a spectrum of characteristics of elderly life (behavioral risk factors, physical activities, morbidity, etc) analyzed both individually and grouped in related clusters, significantly affect risks in the U.S. elderly population. The most influential of the potentially preventable risk factors can be detected using approaches discussed in this report and applied to further deeper analyses, including other data sets with detailed risk factors description. Using information about the duration and intensity of influential risk factors, a more individualized forecast of cancer risk can be made and more individualized prevention strategies can be applied.