Measuring progress in population health: Evaluating trends in recovery rates from the NLTCS/Medicare data

Anatoli I. Yashin, Igor Akushevich, Svetlana V. Ukraintseva, Lucy Akushevich, Konstantin G. Arbeev, Alexandr Kulminski

Center for Population Health and Aging, Duke University, USA

Traditionally, studying dynamic aspects of disease prevalence includes analyses of trends in incidence and survival rates. In case of terminal chronic conditions such analyses are completely justified. This is because disease prevalence will result from a balance between incidence and survival rates for the disease. In cases when recovery from a disease is possible, traditional analyses may lead to erroneous conclusions. For example, ignoring the presence of recovery may lead to erroneous conclusions about "decompression of morbidity" when it does not take place (in case of recovery improvement). This is because each recovered person will automatically be counted as staying in the unhealthy state. Other major issue, which affects the quality of analyses of time trends, is possible dependence between diseases. Because of such dependence, the prevalence of one disease might substantially be influenced by the time trends in incidence, recovery, and survival rates of other diseases. For instance, improving survival from cancer might favor increase in prevalence of some other disease, if the risk of the latter appears higher among cancer survivors.

Researchers, working in the area of population health, developed predictions of future burden associated with major chronic conditions, including cancer, coronary heart disease, stroke, asthma, and others, by considering such diseases separately. The limited use of multivariate analyses in demography was partly because information about connections among diseases was not well systematized, and the methodology capturing disease dependence and capable of working with large data sets containing such information (e.g., Medicare Service Use Files) was not fully elaborated. At the same time, the methods of multivariate analyses of time series data, which are well elaborated and tested in numerous non-demographic applications nowadays, provide compelling evidence that taking dependence among predicting components into account results in a much better quality of forecasting. One more reason for the limited use of time series analyses of multivariate health status data was that cohort dynamics, typical of demographic calculations, was largely ignored in direct applications of these methods. All these connections become especially important in evaluating time trends in the effects of health care and technological innovations on population health.

The progress in medical technology and health care has three major effects on the dynamics of the individual health trajectories. First, it can affect ages at onset of diseases shifting them toward older ages. This can lead to compression of morbidity provided that this shift is larger than increase of the life span. Second, it can reduce mortality rate for individuals in the disease state. As a result, the duration of unhealthy life span will increase. Third, it can increase the recovery rate. In this case, the duration of healthy life span increases even if age at onset and mortality are not affected. Note that time trends in the recovery rate were not actively studied before, because of difficulties related to the definition of "recovery", which has to be specified separately for each disease used in the study. In the absence of the established notions of recovery rate, we use operational definitions, based on our experience in working with data from Medicare service use files and analysis of available literature. An important property of such newly introduced recovery rates is that they have to be responsive to medical progress associated with treatment of respective disease. In this case, they can measure influence of medical progress on population health. The problem of the recovery definition in our analysis is partly offset by major focus on trends in this measure which is a relatively robust characteristic, i.e., its presence can be detected using different definitions.

The age at onset of a disease is completely characterized by the age-specific incidence rate calculated as the number of new disease cases counted among people survived the given age per fixed number of people (e.g., per 100,000) of a given age living in a population in a particular year, or time interval. The age-specific mortality rate from a disease stands for the number of deaths related to the disease per 100,000 people in a population of a given age in a particular year or time interval. The disease case fatality rate is expressed as the proportion of the patients dying within the follow-up period out of all patients under observation. Case fatality indicates how serious a disease condition is in causing death to the patients, usually within a relatively short (30 days to one year) period of time. The relative survival from a disease refers to the ratio of the observed survival rate for the patient group to the expected survival rate for persons in the general population similar to the patient group with respect to age, sex, race, and year of observation. For example, the 5-year relative survival rate is commonly used to estimate the proportion of cancer patients potentially curable. Disease prevalence represents the proportion of individuals with a given diagnosis living in a general population, no matter when the diagnosis has been made. It broadly characterizes the disease burden. Disease prevalence is determined by the incidence, patients' survival, and recovery rates. The long-term forecasting prevalence in our study will therefore include separate forecasting trends in these rates.

In this study, we selected diseases that satisfy one or more of the following criteria. The diseases should be: (i) common among the elderly and substantially contribute to the burden of geriatric diseases in the U.S.; (ii) among the leading causes of death in the U.S., particularly at ages 65 and older, and (iii) plausible candidate for studying dependent risks (i.e., there should be indications from epidemiology and/or biology that these diseases may be connected and/or influence patients' survival). Based on these criteria, we selected acute coronary heart disease (ACHD), stroke, lung cancer, breast cancer, prostate cancer, colon cancer, melanoma, and asthma. Some of the selected diseases may manifest synergistic or antagonistic interactions that make them plausible candidates for considering the effect of dependency on future trends in disease prevalence. We used National Long Term Care Survey (NLTCS) data merged with Medicare service used files to investigate trends in respective recovery and case fatality rates.

The National Long Term Care Survey (1982, 1984, 1989, 1994, 1999, and 2004/5) contains longitudinal and cross-sectional data on a nationally-representative sample of about 49,000 U.S. elderly persons aged 65+ years, with 17,000-20,000 age-eligible survivors in each of six rounds. The NLTCS provides data on demographic factors, residence type, income, assets, height, weight, alcohol and cigarette use, exercise, 28 major medical conditions, 6 ADLs, 8 IADLs (in screener), 7 functional limitation items, subjective health status, cognitive status (Short Portable Mental Status Questionnaire [SPMSQ] 1982-1994, 2004/5; Mini-mental Status Examination [MMSE] 1999; and information on aberrant behaviors). All NLTCS records are linked to Medicare claims data for 1982–2005 to allow tracking of mortality, Medicare claims, and HMO/MCO enrollment/disenrollment.

Medicare data. Medicare is the primary health insurer for 97% of the US population aged 65+ years. All Medicare beneficiaries receive Part A benefits, which cover inpatient care in short- and long-stay hospitals, skilled nursing facilities, home health, and hospice care. 95% of beneficiaries also subscribe to Medicare Part B to obtain benefits that cover physician services, outpatient care, durable medical equipment, and home health in some cases. The Medicare claims data records contain information on dates and costs of service, types of providers, ICD-9-CM (International Classification of Diseases-9th Revision-Clinical Modification) diagnoses responsible for services, and auxiliary diagnostic codes and procedure codes.

The problem of the consistency of the Medicare data over time is largely offset when diagnoses in the Medicare Part B were added (since 1991). This also offsets the problem of underestimating disease prevalence and provides a reliable basis for evaluating possible bias in the estimates of the disease trends. For instance, we estimated the changes in disease risks from the period from 1991 to 2004 using all seven Medicare sources and selected source(s) (e.g., hospital). Comparing these estimates, we evaluated the role of possible source-specific bias in the disease trends.

In this paper we present the results of analyses of time trends in recovery and case-fatality rates for selected diseases using the NLTCS/Medicare data. These results indicate importance of taking recovery/remission process into account in evaluating trends in population health. They also suggest the need for evaluating possible dependence between health and survival events associated with recovery and case-fatality rates.