Igor Akushevich, Alexander Kulminski, Lucy Akushevich, and Kenneth G. Manton

Age Patterns of Disease Incidences in the U.S. Elderly: Population-Based Analysis
Age Patterns of Disease Incidences in the U.S. Elderly:
Population-Based Analysis*

Igor Akushevich, Alexander Kulminski, Lucy Akushevich and Kenneth G. Manton
Center for Demographic Studies, Duke University
Durham, North Carolina 27708-0408, USA

TRENDS Working Paper 06-06

Igor Akushevich, Research Scientist, Sr., igor.akushevich@duke.edu, (919)668-2715; (corresponding author)
Alexander Kulminski, Research Scientist, Sr., kam@cds.duke.edu, (919)684-4962
Lucy Akushevich, Analyst Programmer II, lucy@cds.duke.edu, (919)668-2717
Kenneth G. Manton, Scientific Director and Research Professor, kgm@cds.duke.edu, (919)668-2719

* This research was supported by grant P01 AG17937-05 from National Institute on Aging. A.K. also acknowledges support from K12-AG-000982-05 NIA grant.
Age Patterns of Disease Incidences in the U.S. Elderly: Population-Based Analysis

Abstract

Age patterns of incidence of major age-associated diseases (cancer, diabetes, cardio- and cerebrovascular, and neurodegenerative diseases) are calculated in the U.S. elderly population (aged 66-107) using ICD-9 diagnosis reported in Medicare records linked to 1982 to 1999 NLTCS survey respondents. Detailed sensitivity analysis, including effects of latent censoring, incorrect age reporting, and other procedural uncertainties showed their stability. Found agreement between the U.S. age-specific incidence rates estimated from the linked NLTCS-Medicare files and results from other epidemiological datasets suggests the utility of Medicare administrative records for characterizing national disease incidence rates — data not available for many diseases from other sources.
INTRODUCTION

Determining national trends in health and vital status in the population with growing proportions of elderly is a major concern of government as well as health care providers such as Medicare and Medicaid. To better address the health demands of the elderly, and to reduce economic burdens on society, it is important to understand the forces governing the health of aging persons. A key component of national health care associated with aging is the age specific incidence of major age-associated diseases. Unfortunately, identification of their age patterns with sufficient precision requires large population-based databases that are costly to collect. This is a reason why studies of age patterns of specific diseases in the U.S., along with investigation of factors affecting them, are rather rare, especially in the elderly population. The diseases whose incidence is best studied at a national level are cancers, which were followed for a long period of time in the SEER program (Ries et al., 2004). Age patterns of cardiovascular, cerebrovascular (Arnold et al., 2005) and neurodegenerative (Fitzpatrick et al., 2004) disease incidences were recently presented using the Cardiovascular Health Study (CHS). The CHS, however, is not nationally representative. Various chronic diseases important for elder well-being such as Alzheimer’s disease and related dementias (Gao, Hendrie, Hal, & Hui, 1998), and diabetes mellitus (McBean, Li, Gilbertson, & Collins, 2004), have been analyzed in detail, but it is not clear what the age patterns would be in the entire U.S. elderly population. A number of studies of diseases of the elderly were conducted in local area populations (e.g., the Rochester, Minnesota and Cache County studies of Dementia and Alzheimer's disease epidemiology (Kokmen, Beard, O'Brien, & Kurland, 1996; Miech et al., 2002) which are also not representative of the national population. The lack of comprehensive and representative analyses of age patterns of incidence of major age-associated diseases at a national level motivates this analysis.

Incidence rates can be calculated using data from national Registers or from surveys representative of the U.S. population. In the latter case, special procedures are needed to generalize survey respondent results to the national level. One way is to use a weight function (possibly time dependent), assigned to each individual in the sample such that “weighted” sums (means) over individuals in the sample give quantities at the national level. A national survey with such a design is the 1982 to 1999 National Long Term Care Survey (NLTCS), which focuses on the U.S. elderly (65+) population. Since NLTCS is linked with Medicare records and Vital Statistics files, this is a unique dataset following the elderly for up to 23 years (1982 to 2005) – with continuous recording of health service use, age of death and detailed reassessment of health status by survey every five years (except
for the first two waves in 1982 and 1984). NLTCS participants were drawn from enrollment lists for Medicare. In the 1982 to 1999 NLTCS, there were about 400,000 person-years of exposure over age 65, including over 100,000 person-years of exposure over age 85. Sample weights are provided by U.S. Census Bureau and are available for each year of follow up. Thus, the NLTCS design provides an excellent opportunity to study age-associated disease incidence patterns for the U.S. elderly population.

Earlier analyses of the NLTCS provided consistent results on functional disability, active life expectancy, and chronic disease prevalence for the U.S. elderly population. In particular, a decline in chronic disability of 15% (1.1% per annum) from 1982 to 1994 (Manton, Corder, & Stallard, 1997; Manton, Stallard, & Corder, 1998) was documented. Manton and Gu (2001) confirmed this trend up to 1999 and noted its acceleration from 1994 to 1999. Declines in disability and functional limitations were also documented in other nationally representative studies (Freedman, Martin, & Schoeni, 2002; Freedman et al., 2004). Mortality rates of the U.S. elderly population was calculated by Akushevich, Kulminski, & Manton (2005) and Manton, Akushevich, & Kulminski (2005). The results are in agreement with national projections of Social Security Administration (SSA, 2003) and Society of Actuaries (SoA, 2000) and with data in the Human Mortality Database (HMD, 2005) for ages 65-95. The NLTCS shows a plateau effect and declines in the per annum hazard rate among survivors to ages 100+.

The Medicare service use files (linked to the NLTCS) provide detailed information about disease diagnoses and dates of onset. Analysis of incidence rates at the national level using administrative databases requires detailed investigation of stochastic and systematic unobservable effects, which may affect age patterns. Statistical uncertainties in our study are characterized by 95% confidential intervals, whose calculation has to reflect corrections for sample design. There exist several demographic effects which can lead to systematic bias in measured epidemiological quantities, especially at advanced ages. Possible effects of uncertainties in Medicare data were analyzed by Kestenbaum (1992) and by Kestenbaum and Ferguson (2002, 2005). They demonstrated the reliability of Medicare databases for such studies and suggested ways to further reduce uncertainties at advanced ages; e.g., by using only data about Medicare Part B coverage, for which a premium must be paid monthly. Another possible source of uncertainty is errors in age reporting as discussed by Rosenwaike & Stone (2003), Hill, Preston, & Rosenwaike (2000), and Preston, Elo, Rosenwaike, & Hill (1996). They compared Medicare/SSA data with Census data at the end of 19th century and produced an empirical distribution of age corrections (both positive and negative) conditional on sex and race.
In the analyses below we present 1) estimates of age-specific disease incidences for the U.S. elderly population and 2) detailed sensitivity analyses validating findings in age patterns. Reliable disease incidence estimates above age 85 are rare; hence we can compare our estimates only up to age 85 to other studies – many of which are produced in non-nationally representative studies. Thus, the disease incidence estimates for advanced ages presented here should be very valuable both in theoretical aspect of understanding the interaction of disease incidence and senescence and in practical implementations for analyzing U.S. population health trends and forecasting future Medicare expenditures.

DATA and METHODS

NLTCS and Medicare Data and Estimation Procedure

The NLTCS was started in 1982 with roughly 20,000 individuals examined in 1982, 1984, 1989, 1994, and 1999. At the time of each survey – beginning from 1984, a cohort sample of about 5,000 persons was added to the sample of survivors to replace deaths occurring since the prior survey and to ensure that the new sample was representative of the entire elderly population aged 65+. In total, the 1982 to 1999 NLTCS files represent 41,947 different individuals, with roughly 25,000 deaths recorded to 2003 and about 400,000 (100,000) person-years of exposure over age 65 (85). Response rates in the NLTCS are high – roughly 95% in all survey years.

Continuous Medicare history files (linked to the NLTCS) contain information about costs, treatments, and diagnoses on the service dates – as well as the date of death. Medical information (disease diagnoses and service dates) will be used from the following sources: clinical labs, durable medical equipment regional carrier (DMERC), home health agency, hospice, inpatient/SNF, outpatient, carrier (other than DMERC) and Skilled Nursing Facility claim records. For our present calculations we used data from 1984 to 2001.

Because the NLTCS was designed to have both longitudinal and cross-sectional components, respondents entered the sample at different times and ages. Some of them stay healthy and alive during the survey, so different censoring schemes have to be considered. The incidence rate for a specific age group is calculated as the ratio of the number of new disease onsets observed in the group to the number of person-years accumulated by the members of the group. The calculation of the incidence rates has to be performed taking into account right-censoring. For our problem this means we have to calculate the individual duration of observation in the NLTCS and Medicare rather than use life tables. This allows us to perform the calculation to be accurate to within one day. Date
of birth, onset, death, (dis)enrollment (from) into the NLTCS and Medicare Parts A&B, and last record dates are known with one-day accuracy.

Projection of estimates from the NLTCS data to the U.S. elderly population (65+) may be biased by the sample design. To have estimates representative of the entire U.S. elderly population, sample design effects are represented using CDS screener weights (Manton and Gu, 2001). The sample-weight function for each individual can “jump” when, for example, a new wave starts. The date when the sample weight function jumps, is known with one-day accuracy. Therefore, each individual in the NLTCS can be associated with the precise time interval under observation with a corresponding weight function and life history extracted from Medicare files. Age patterns of incidence rates will be assessed by stratifying the sample into relevant age categories (a year, or several years). The richness of the NLTCS data allows us to produce statistically significant estimates of incidence rates in three-year intervals. Empirical age-specific risks (λₐ) are calculated as a ratio of weighted numbers of cases to weighted person-years at risk:

\[ \lambda_a = \frac{n(a)}{N(a)}; \quad n(a) = \sum_i w_i(a); \quad N(a) = \sum_i w_i(a), \]

where \( w_i(a) \) is the individual weight at age \( a \); \( n \) runs over all disease onsets detected in the age group, and \( i \) runs over all individuals at risk in \( a^{th} \) age group.

Standard error (SE) and confidence interval (CI) calculations must also be adjusted for sample design effect (Manton et al., 1997, Manton and Gu, 2001). The approach suggested by the Census Bureau for the NLTCS uses generalized variance function methods (Wolter, 1985), in which the SE are adjusted for sample design using,

\[ SE = \sigma_E = \sqrt{\frac{b}{N(a)}} \lambda_a (1 - \lambda_a). \] (1)

Parameter \( b \) is an adjustment factor for study design effects. Numerical values of \( b \) produced by the U.S. Census Bureau are available for each NLTCS wave (U.S. Bureau of Census, 2001). This factor is close to the mean individual weight for a specific wave. To calculate age specific effects, Eq. (1) has to be generalized. First, we have to take into account that individuals contribute to person-years of the same age group from different waves, so the factor \( b \) might be not a constant. Second, Eq. (1) (and the standard Wald’s CI) does not work when \( N(a) \) is small (Brown, Cai, & DasGupta, 2001). A generalization based on Wilson’s approach (Brown et al., 2001) uses,
\[ CI_w = \frac{N_b(a)\lambda_a + \frac{1}{2}z_{\alpha/2}^2}{N_b(a) + \frac{1}{2}z_{\alpha/2}^2} \pm \frac{z_{\alpha/2}\sqrt{N_b(a)}}{N_b(a) + z_{\alpha/2}^2} \left( \frac{\lambda_a(1-\lambda_a) + \frac{1}{4}z_{\alpha/2}^2}{N_b(a)} \right), \quad (2) \]

where \( N_b(a) = \sum_i w_i(a) / b(w_i(a)) \) and \( b(w_i(a)) \) can vary with time (i.e., from wave to wave). For large \( N_b(a) \) and constant \( b(w_i(a)) = b \) (one wave), Eq. (2) recovers the standard Wald’s estimates of CIs, i.e., \( \frac{1}{2}\sigma \pm z_{\alpha/2} \) \( \Phi^{-1}(1-\alpha/2), \) where \( \Phi(x) \) is the standard normal distribution function and \( \alpha \) denotes the confidence level.

An important subtask in our analyses will be an investigation of censoring effects. It is known that risks calculated using Medicare files can be underestimated; e.g., because of death or disease onset occurring outside the survey areas. Kestenbaum (1992) suggests that such effects may materially bias estimates for the oldest-old (85+) and centenarian (100+) population. We pay special attention to this problem because of the focus of our analysis on disease incidence at extreme ages.

Our basic calculation uses the censoring scheme in which the final individual date of observation period is defined from the Medicare vital statistics file (August 6, 2003 for the analysis). We will consider two alternative schemes. One affecting the final day is the last day of Part B coverage, and the second is when this day corresponds to the last record either in the Medicare claims or NLTCS interview files.

Detection of Disease Onsets
To calculate age-specific disease incidence rates we need to know the date of disease onset. These dates are extracted from 1984-2001 Medicare service use files linked with the NLTCS. Medicare data provides dates of claims for medical service, which are accompanied by the corresponding ICD codes. We will assume that an individual might experience an onset of a disease during the period of observation if there is at least one record with the ICD-9-CM code corresponding to this disease on a single institutional claim (inpatient, skilled nursing facility, home health care, hospice, and outpatient) or non-institutional claim (carrier/physician supplier/Part B [1991-2001 only], durable medical equipment, clinical labs). Actually, Medicare data do not contain information on whether the appearance of an ICD-9-CM code is a true onset or just a visit to treat a disease with onset registered beyond the observation period. Therefore, to determine a date of onset we will assume that beneficiaries with a chronic condition receive medical care at least once within the first 6 months since his/her enrollment into Medicare. If certain diagnoses appear in Medicare files within an initial 6-month period, such an individual will be considered as
chronically impaired at the time of enrollment in Medicare. Otherwise the date of the first appearance of the corresponding diagnosis will be considered the date of onset.

RESULTS

Our analysis focuses on the following age-associated diseases:

- **Cancer**, which includes all malignant neoplasms (140-208);
- **Diabetes mellitus**, which includes both insulin-dependent and non-insulin-dependent forms (250);
- **Cardiovascular diseases** (CVD), which include acute rheumatic fever (390-392), chronic rheumatic heart disease (393-398), hypertensive heart disease (402), hypertensive heart and renal disease (404), ischemic heart disease (410-414), diseases of pulmonary circulation (415-417), and other forms of heart disease (420-429);
- **Cerebrovascular disease** (CBVD) (430-438);
- **Neurodegenerative disorders** (NDD), which include psychoses (290-299), nonpsychotic mental disorders (300-316), and hereditary and degenerative diseases of the central nervous system (330-337).

We begin our analysis by examining the Medicare sources for ICD codes and their relative contribution to age-specific incidence rates. Our goal is to identify conditions for obtaining the most stable results that can be considered as basic. Then we perform detailed sensitivity analysis on them.

Medicare claim data have certain limitations that are with a matter of the determination of diagnoses. One is associated with lack of ICD codes from the Physician/Supplier/Part B source before 1991. Lack of this information might result in underestimation of the incidence rates and shift onsets to later ages. We tested the significance of this effect by comparing incidence patterns for the period from 1992 to 2001 with, and without, information from this source. We found that physician diagnoses account for about 30-40% of total onsets. Differences between incidence rates calculated for different diseases for 1984-1990 (where Physician/Supplier/Part B information is not available) and 1992-2001 are of this order. The Part-B-associated loss of diagnoses is mostly disease independent. It is slightly larger for NDD and slightly smaller for diabetes. It is also independent of age – with the exception of very advanced ages (100+), where differences essentially vanish. The effect is sex-dependent, with more loss of diagnoses for females for most diseases and ages. This conclusion about the essential role of diagnoses from the Physician/Supplier/Part B source is also confirmed by analysis of the time
distribution of new diagnoses, in which an abnormally large peak appears in 1991 for participants of the NLTCS cohort from the first three (1982, 1984 and 1989) waves. Consequently, we limit our analysis to data from 1992 to 2001.

Estimates of age-specific disease incidence rates are presented in Table 1. Incidence rates for CVD and NDD monotonically increase with age. This is because CVD and NDD groups cover a wide range of specific diseases that are prevalent in the elderly population. Relatively few individuals at advanced ages do not have diagnoses corresponding to diseases from these groups, significantly decreasing the number of unexposed persons. By focusing on specific diseases from these groups, we observe less pronounced increases – or even declines – of incidence rates at advanced ages that result in a peak in the age pattern. In contrast, incidence rates associated with diabetes remain practically independent of age.

Comparison of incidence rates calculated for 1984-1990 and for 1992-2001 (no Physician/Supplier/Part B) gives estimates of time trends, which should be in agreement with those observed over the age ranges examined in previous studies. Our results show i) a decline in CVD incidence in agreement with Ergin, Muntner, Sherwin, & He, (2004), ii) an increase in NDD incidence comparable to that shown by Rocca, Cha, Waring, & Kokmen (1998), and iii) a moderate increase in the diabetes rate (McBean et al., 2004). Detailed analysis of time trends will be presented elsewhere.

Sensitivity analysis

A disadvantage of large administrative databases is the presence of factors producing systematic over/underestimation of the number of diagnoses or the age at onset. One reason for such uncertainties concerns incorrect dates of onset. Other sources involve incorrect reporting of date of birth and date of death. While the first affects age at onset, the second tends to reduce the number of person-years at risk. To evaluate the effect of these uncertainties we i) perform calculations with different definitions of disease onset; ii) use censoring schemes employing alternative data to define individual observation periods; and iii) simulate unobservable effects of errors in reporting date of birth. Comparison of the recalculated incidence rates with our basic results provides estimates of uncertainties due to these sources of error.
<table>
<thead>
<tr>
<th></th>
<th>Cancer</th>
<th>Diabetes</th>
<th>CVD</th>
<th>CBVD</th>
<th>NDD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66-68</td>
<td>57 (52,63)</td>
<td>39 (34,43)</td>
<td>111 (102,120)</td>
<td>37 (33,42)</td>
<td>50 (45,56)</td>
</tr>
<tr>
<td>69-71</td>
<td>64 (59,68)</td>
<td>38 (35,41)</td>
<td>117 (111,125)</td>
<td>41 (38,44)*</td>
<td>55 (51,59)</td>
</tr>
<tr>
<td>72-74</td>
<td>66 (61,71)</td>
<td>32 (29,36)</td>
<td>110 (103,119)</td>
<td>46 (43,50)**</td>
<td>56 (52,60)+</td>
</tr>
<tr>
<td>75-77</td>
<td>67 (61,73)</td>
<td>38 (34,42)</td>
<td>129 (120,140)</td>
<td>62 (57,67)</td>
<td>73 (68,79)</td>
</tr>
<tr>
<td>78-80</td>
<td>71 (64,78)*</td>
<td>32 (28,36)</td>
<td>139 (127,152)</td>
<td>68 (62,74)</td>
<td>74 (67,80)+</td>
</tr>
<tr>
<td>81-83</td>
<td>90 (81,99)</td>
<td>37 (32,42)</td>
<td>155 (139,171)</td>
<td>80 (72,88)</td>
<td>97 (89,107)*</td>
</tr>
<tr>
<td>84-86</td>
<td>92 (81,104)</td>
<td>34 (28,41)</td>
<td>152 (133,173)</td>
<td>83 (74,94)</td>
<td>109 (98,121)</td>
</tr>
<tr>
<td>87-89</td>
<td>106 (91,124)</td>
<td>42 (34,51)</td>
<td>229 (197,265)</td>
<td>98 (85,114)</td>
<td>131 (115,150)</td>
</tr>
<tr>
<td>90-92</td>
<td>101 (80,126)*</td>
<td>48 (36,63)</td>
<td>226 (178,282)</td>
<td>116 (95,142)</td>
<td>133 (109,162)</td>
</tr>
<tr>
<td>93-95</td>
<td>116 (85,157)</td>
<td>49 (32,73)</td>
<td>261 (189,348)+</td>
<td>134 (101,176)</td>
<td>143 (109,186)</td>
</tr>
<tr>
<td>96-98</td>
<td>115 (67,192)</td>
<td>59 (32,107)</td>
<td>203 (109,346)+</td>
<td>101 (59,168)</td>
<td>194 (131,279)</td>
</tr>
<tr>
<td>99-101</td>
<td>113 (39,288)</td>
<td>25 (5,118)</td>
<td>324 (126,614)</td>
<td>77 (26,204)+</td>
<td>248 (126,430)+</td>
</tr>
<tr>
<td>102-104</td>
<td>41 (2,492)*</td>
<td>84 (11,434)</td>
<td>112 (5,747)</td>
<td>97 (15,437)+</td>
<td>56 (4,463)+</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66-68</td>
<td>40 (36,44)</td>
<td>30 (27,34)</td>
<td>97 (90,105)</td>
<td>35 (31,39)</td>
<td>63 (57,68)</td>
</tr>
<tr>
<td>69-71</td>
<td>38 (35,41)</td>
<td>32 (29,35)</td>
<td>103 (98,109)</td>
<td>38 (35,41)+</td>
<td>62 (58,66)</td>
</tr>
<tr>
<td>72-74</td>
<td>43 (40,47)</td>
<td>31 (28,34)</td>
<td>109 (103,116)+</td>
<td>46 (43,50)+</td>
<td>64 (60,68)</td>
</tr>
<tr>
<td>75-77</td>
<td>48 (44,52)+</td>
<td>34 (31,37)</td>
<td>121 (113,129)+</td>
<td>58 (54,62)</td>
<td>69 (64,74)+</td>
</tr>
<tr>
<td>78-80</td>
<td>49 (45,53)</td>
<td>33 (30,36)</td>
<td>132 (123,142)</td>
<td>60 (55,65)+</td>
<td>81 (76,87)+</td>
</tr>
<tr>
<td>81-83</td>
<td>54 (49,59)</td>
<td>34 (30,38)+</td>
<td>158 (146,171)</td>
<td>73 (68,79)+</td>
<td>95 (89,103)**</td>
</tr>
<tr>
<td>84-86</td>
<td>54 (49,60)</td>
<td>38 (34,43)</td>
<td>158 (143,173)</td>
<td>84 (77,91)**</td>
<td>107 (98,116)**</td>
</tr>
<tr>
<td>87-89</td>
<td>49 (43,56)</td>
<td>41 (36,47)</td>
<td>181 (162,202)+</td>
<td>95 (86,105)</td>
<td>149 (137,162)</td>
</tr>
<tr>
<td>90-92</td>
<td>55 (47,64)</td>
<td>39 (32,46)+</td>
<td>208 (180,239)+</td>
<td>108 (96,122)</td>
<td>169 (151,187)</td>
</tr>
<tr>
<td>93-95</td>
<td>53 (42,66)</td>
<td>31 (23,41)</td>
<td>224 (182,273)+</td>
<td>105 (88,124)</td>
<td>204 (176,234)+</td>
</tr>
<tr>
<td>96-98</td>
<td>57 (41,79)</td>
<td>35 (24,51)</td>
<td>285 (220,360)+</td>
<td>123 (98,155)</td>
<td>181 (143,226)+</td>
</tr>
<tr>
<td>99-101</td>
<td>53 (31,91)</td>
<td>26 (12,54)+</td>
<td>206 (123,324)</td>
<td>137 (94,194)+</td>
<td>199 (139,276)+</td>
</tr>
<tr>
<td>102-104</td>
<td>53 (19,140)+</td>
<td>44 (15,126)</td>
<td>243 (98,488)</td>
<td>108 (47,226)</td>
<td>177 (183,338)</td>
</tr>
<tr>
<td>105-107</td>
<td>10 (0,183)</td>
<td>51 (7,273)</td>
<td>256 (48,700)</td>
<td>95 (16,405)</td>
<td>160 (35,499)</td>
</tr>
</tbody>
</table>

Upper (lower) subscript shows level of systematic uncertainties. See text for details.

TRENDS Working Paper 06-06
Uncertainty in onset calculation. Sources for uncertainties related to overestimation and shift of dates of onsets are the enrollment of new beneficiaries and alteration of coverage by the Medicare program of certain beneficiaries within the observation period for legal (eligibility) or administrative (enrollment under another health insurance) reasons. Actually, enrollment of new beneficiaries does not lead to overestimates in our analysis because a) we use a 6-month cut to determine disease onset and b) Medicare data cover a longer time domain than that of individual observations from 1992 to 2001. To project estimates of incidence rates to the U.S. population we use the design weights, which become nonzero after the first survey in which the individual participates. Thus, participants of the fourth wave (1994) contribute to the incidence pattern only beginning from 1994 (when their weights become nonzero), but their diagnoses are analyzed from 1991. This provides sufficient time to avoid such bias.

Medicare coverage and censoring uncertainties. Here we consider two types of uncertainties. The first is related to partial coverage. However, we do not expect that it is important because of the relatively small fraction of individuals who are not under part A or B coverage. In January 2001, 3.46% of sample persons had Part A only, 0.69% had Part B only, and 95.85% had both Part A and B coverage. Second, sources are related with effects of effective censoring when an individual is “under observation;” however his/her records cannot appear in Medicare files. A well-known example of how information can be missed is the relocation of elderly abroad. Details of how such effects can influence mortality are discussed by Kestenbaum (1992) and Kestenbaum & Ferguson (2002).

Effect of such uncertainties can be investigated by applying different censoring schemes with different definitions of censoring dates. In the basic calculation, the final date of observation is the earliest date among dates of disease onset, death, and the last date of cohort observation; i.e., December 31, 2001. In the first alternative censoring scheme, the last day of observation is the last day of part B coverage. Since part B coverage requires monthly payments, this is a good indicator that an individual is alive and being observed. Comparison of these results with our original calculations shows that the ratio of incidences rates deviates noticeably from 1. only for ages ~100 and does not normally exceed 1.03. A measure of these uncertainties estimated in units of CI is presented in Table 1. If the deviation from the basic result exceeds 5% (10%, 15%) of CI of corresponding sign, it is marked by one (two, three) ‘+’ or ‘−’ as a subscript in a corresponding cell of Table 1. To evaluate a maximal level of censoring uncertainties, we used a second censoring scheme in which the last observation day is defined by the last record in either the NLTCS or Medicare
files. Uncertainties in this case are about 4-7% (i.e., the ratio is 1.04-1.07) increasing to 15% for younger ages. Such age dependence of the uncertainty allows us to speculate that this effect arises due to a decreasing contribution of person-years that is important for younger ages, rather than due to effective loss of individuals—which is more likely at advanced ages.

**Age reporting uncertainties.** An important source of uncertainty, especially at advanced ages, might be represented by errors in the reported date of birth. Estimation of the percentage of individuals in sex- and race-specific populations reporting the wrong date of birth and probability distribution over years added/subtracted from original date of birth was performed by Rosenwaike and Stone (2003) and Preston et al. (1996). They focused on individuals born around 1875. Using their results, we performed a simulation study to estimate the effect on the age-specific incidence rates. Rosenwaike and Stone (2003) report that since the early 1990s (we assume 1990), individuals who have applied for Medicare have had to meet “rather strict evidentiary requirements” regarding proof of age. Therefore, the situation with errors in date of birth (i.e., the effect of “age misreporting”) can be considered improved for those born in 1925 and later. Therefore we have to provide estimates for those born before 1925. Two alternative strategies are used for that. First, assuming a constant percentage of people misreporting age from 1875 to 1925, we assess the upper limit of uncertainty. A more realistic situation is that the percentage of people misreporting age is not constant and declines—being maximal for those born in 1875 and reaching a minimum for those born in 1925. Therefore a second strategy is to linearly interpolate this decline. The initial distribution of age-misreporting people conditional on sex and race is taken from Table 5 of Rosenwaike and Stone (2003). In both strategies, we keep this distribution fixed. For each individual, the offset in date of birth is simulated according to this distribution in calculating age-specific incidence rates. For the first strategy, we also keep fixed percentages for people misreporting age. In the second case, this percentage linearly decreases. We simulated 500 samples, which is sufficient to have statistically stable results. Averaging over these samples, we calculated final incidence rates with adjusted dates of birth.

The results for evaluation of the systematic uncertainty due to the effect of age misreporting using the linear model are shown in Table 1. If deviations from the basic result exceed 5% (10\%, 15\%) of CI of corresponding sign, it is marked by one (two, three) ‘+’ or ‘−’ as a superscript in the corresponding cell of Table 1. Estimation of the upper level of the uncertainty using the model with a constant percentage of people misreporting age gives modest results: the incidence rate may change 2-5\% only for the last age interval (102-104
DISCUSSION

To compare our results with those in the epidemiologic literature, we focus on each group of diseases, emphasizing 1) the shape of the age pattern of incidence rates, 2) the absolute incidence level, and 3) sex differences.

The most detailed U.S. data on cancer incidence come from the SEER Registry for invasive cancers (Ries et al., 2004). Although we did not distinguish between total malignant cancer incidence and incidence of only its invasive forms, we hypothesize that both forms are adequately addressed by considering onset for inpatients only. Therefore, for comparison we calculated age patterns of incidence rates using i) all sources for onsets and ii) an inpatient-only subset. Results are presented in Figure 1. There is excellent agreement for cancer based on inpatient onset. Incidence rates from all onsets are higher than what can be attributed to the contribution of non-invasive cancer forms.

Estimates of age-specific CVD and CBVD incidence were recently published by Arnold et al. (2005). They used CHS data to estimate incidence rates of major categories of CVD in older Americans. Our estimates are compared with those for myocardial infarction (410) and stroke (436). Good agreement is seen for the incidence of myocardial infarction for males based on all sources of onsets. For females, results are in better agreement with inpatient rates. Incidence rates for stroke are in poor agreement with those provided by Arnold et al (2005). This can be attributed to the choice of the cohort in the CHS and uncertainties in both studies related to inadequate quality of self-report data and mismatch in ICD codes.

Incidence rates of diabetes were the focus of many studies. However, there are only a few results showing age distribution of diabetes above age 65. Recent results are provided by McBean et al. (2004) who examined diabetes prevalence, incidence, and mortality from 1993 to 2001 among fee-for-service Medicare beneficiaries aged 67+ using a 5% random sample of enrollees. The shape and absolute level of incidence rates (Table 1) are in agreement with those provided by McBean et al. (2004). Note, however, that another scheme for diabetes onset identification (Hebert et al., 1999) is often used. In this scheme, it is required that a second record with diabetes ICD must be observed if the first one was registered as an ambulatory claim (i.e., a physician/supplier or hospital outpatient claim). This is a reason for the relative excess in the incidence rate in Table 1. Incidence rates calculated using this scheme decrease up to 10%. Figure 2 demonstrates excellent
agreement between the diabetes age pattern calculated using the second strategy and the results of McBean et al. (2004).

The incidence rate of NDD is large and is second to CVD. The most common disorders in the NDD group are Alzheimer’s disease and dementia. They have been the focus of several studies and meta-analyses that estimated incidence rates and their age patterns in elderly populations. We compare the results of our calculation to the meta-analysis of Gao et al. (1998) and to the analysis of CHS by Fitzpatrick et al. (2004). Results of both of those studies are in good agreement with our study (Fig.3). Since CHS includes only community-dwelling individuals, the minor excess in our results for Alzheimer’s disease for individuals aged 80+ can be explained by the contribution of institutionalized individuals. Our results are also in good agreement with dementia and Alzheimer’s disease incidence observed in the Cache County study (Miech et al., 2002) with exception of the case of females aged 85+, for which our results are 1.5-2.0 times lower.

Since we considered broad disease categories covering a large number of specific diseases, we did not typically observe plateaus, or clear peaks in age patterns with the exception of CBVD for males (aged 94) and females (aged 100) [Table 1]. Nevertheless, such effects become more pronounced if we focus on specific diseases, as shown in Figures 1 and 3. Appearance of such effects can be attributed to the effect of selection (Vaupel et al., 1998), where frail individuals do not survive to advanced ages. Another explanation is related to possible under-registration of diagnoses at advanced ages – which, however, cannot be proved with available data. Other possible biologically and physiologically motivated explanations are discussed by Ukraintseva and Yashin (2001).

Uncertainties affecting the results include statistical error and systematic bias. Statistical uncertainties are characterized by SE or CI. We employed procedures for CI estimation (see Eq. (2)) that take into account study-design effects and that make the procedure applicable for small numbers of cases and/or person-years. One way to compare statistical uncertainties among age and disease groups is to define a ratio of left 95% CI to the respective mean value. The ratio i) increases with age for all diseases, ii) does not exceed 0.1 until age 85, and iii) exceeds 0.5 only for the last age groups for both sexes. Sensitivity analysis also demonstrated the possible influence of systematic biases on the observed age patterns. We addressed all significant sources of uncertainties discussed in the demographic literature that can affect age patterns. The effect of a given source of uncertainty is estimated by analysis of the difference between incidence rates for a standard (basic) calculation and for a calculation in which the effect of the given source of uncertainty is simulated or taken into account in an alternative way. We evaluated two computational schemes that provided i) the most realistic and ii) maximum estimates.
Specifically, we considered the effects of partial Medicare coverage, uncertainties in onset identification, the influence of different censoring schemes, and the effect of incorrect reporting of date of birth. We show that such uncertainties (both realistic and the maximum possible) do not seriously distort our findings on U.S. incidence age patterns, being a maximum of 10% of the main effect.

CONCLUSION

Age-specific disease incidences were calculated for CVD, NDD, cancer, diabetes mellitus and CBVD for the U.S. elderly population. National age patterns of incidence rates were first obtained for males aged 66-104 and for females aged 66-107. The demographic outcomes were studied using National Long Term Care Survey (NLTCS) longitudinal data collected from 1982 to 1999. Diagnoses and dates of onset were identified using ICD-9-CM codes from Medicare service use files linked to NLTCS. Possible sources of biases were analyzed and their contribution to age-specific incidence rates was estimated. Comparison of these age patterns with those available in the literature shows good agreement for cancer, diabetes mellitus, myocardial infarction (for males), and Alzheimer’s disease and related dementia. This suggests that the national age specific incidence patterns can be adequately evaluated from Medicare service use files. Usefulness of Medicare data is important because there are few data sources to study such incidence patterns to late ages in the national population. In addition, an advantage of the Medicare data is that they relate these age specific incidence patterns to Medicare costs and, using NLTCS files, to disability incidence. Such set of data is extremely important in projecting future Medicare costs.

We believe that these results are timely and important as they may inform current scientific and policy debates about the effects of biomedical research and therapeutic innovations on disease incidence at increasingly advanced ages – ages at which, in the recent past, therapeutic interventions were not even attempted. Future efforts will be directed toward i) investigation of time trends of major age-specific incidence rates, ii) detailed sensitivity analyses based on microsimulation, with special attention to data quality; and iii) mathematical modeling intended to use a broader class of models with substantive interpretation of parameters.
REFERENCES


Figure 1. Age specific incidence rates for selected diseases.

Closed dots (squares) correspond to rates obtained using all sources (inpatient only). Open dots denote SEER based incidence rates (Ries et al., 2004) for invasive cancers and the incidence of myocardial infarction and stroke from analysis by Arnold et al. (2005)
Figure 2. Age specific incidence rates for Diabetes mellitus.

Closed dots correspond to rates obtained using a strategy with the requirement of second record if first one was in outpatient sources. Squares denote results of analysis by McBean et al., (2004).

Figure 3. Age specific incidence rates for dementia and Alzheimer disease.

Closed dots correspond to the rates obtained using all sources. Squares, open dots, and triangles denote results of meta analysis by Gao et al. (1998), of CHS data analysis by Fitzpatrick et al. (2004), and of the Cache County study analysis by Miech et al. (2002), respectively.
Supported by the National Institute on Aging, TRENDS is a network of researchers working to accelerate scientific understanding of old-age disability trends. We do this in three broad ways.

- We conduct original research, looking for the causes and consequences of past, current, and future disability trends.
- We review findings from new studies, reconciling differences across studies and populations in the U.S. and abroad.
- We create an environment that is collegial, facilitating open discussion of salient issues and cutting-edge research.

In particular, we are focused on fostering international and cross-institutional collaboration to enhance our understanding of trends in old-age disability, and providing a mechanism for interactive discussion.

**TRENDS Working Papers** are prepublication papers that report on the current work of TRENDS-affiliated researchers. Written for timely dissemination, these papers may be submitted for publication in scholarly journals.

**Copyrights for all TRENDS Working Papers are held by the authors.** Readers may quote from this work only if they properly acknowledge the authors and the Series and do not alter the original work.