



Published in final edited form as:

J Am Geriatr Soc. 2012 February ; 60(2): 323–327. doi:10.1111/j.1532-5415.2011.03786.x.

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

Igor Akushevich, PhD^a, Julia Kravchenko, MD, PhD^b, Svetlana Ukraintseva, PhD^{a,b}, Konstantin Arbeev, PhD^a, and Anatoliy I. Yashin, PhD, Sc.D^{a,b}

^aCenter for Population Health and Aging, Duke University, Durham, NC 27708

^bDuke Cancer Institute, Duke University, Durham, NC 27705

Abstract

Background/Objectives—To utilize the Medicare Files of Service Use (MFSU) to evaluate patterns in the incidence of aging-related diseases in the U.S. elderly population.

Design—Age-specific incidence rates of nineteen aging-related diseases were evaluated with the National Long Term Care Survey (NLTC) and the Surveillance, Epidemiology, and End Results (SEER) Registry data both linked to MSUF (NLTC-M and SEER-M, respectively), using a developed algorithm for individual date at onset evaluation.

Setting—A random sample from the entire U.S. elderly population (Medicare beneficiaries) was used in NLTC, and 26% of U.S. population is covered by the SEER Registry data.

Participants—34,077 individuals from NLTC-M and 2,154,598 from SEER-M.

Measurements—Individual medical histories were reconstructed using information on diagnoses coded in MFSU, dates of medical services/procedures, and Medicare enrollment/disenrollment.

Results—The majority of diseases (e.g., prostate cancer, asthma, diabetes) had a monotonic decline (or decline following short period of increase) in incidence with age. A monotonic increase of 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. The results obtained from the NLTC-M and SEER-M were in agreement (excluding an excess for circulatory diseases in the NLTC-M). A sensitivity analysis proved the stability of the evaluated incidence rates.

Conclusion—The developed computational approaches applied to the nationally representative Medicare-based datasets allows reconstruction of age patterns of disease incidence in the U.S. elderly population at the national level with unprecedented statistical accuracy and stability with respect to systematic biases.

Corresponding Author: Igor Akushevich, PhD, Center for Population Health and Aging, Duke University, 002 Trent Dr., Durham, NC 27708, Phone: (919) 668-2715, Fax: (919) 684-3861, igor.akushevich@duke.edu, ia6@duke.edu. Alternate Corresponding Author: Konstantin Arbeev, PhD, Center for Population Health and Aging, Duke University, 006 Trent Dr., Durham, NC 27708, Phone: (919) 668-2707, Fax: (919) 684-3861, ka29@duke.edu.

Conflict of Interest: The editor in chief has reviewed the conflict of interest checklist provided by the authors and has determined that the authors have no financial or any other kind of personal conflicts with this paper.

Author Contributions:

IA: study design, data analysis and interpretation, manuscript preparation; JK: data analysis and interpretation, manuscript preparation; SU: data analysis and interpretation; KA: data analysis and interpretation; AI: study design, data analysis and interpretation, manuscript preparation.

Keywords

Medicare; chronic disease onset; comorbidity

INTRODUCTION

Understanding morbidity and mortality trends in the U.S. population with growing proportions of elderly is a major public health concern and an important issue for policymakers and governmental institutions, as well as for health care agencies such as Medicare and Medicaid. Studying age patterns of aging-related diseases requires large population-based datasets that are costly to collect and maintain; therefore, studies on age patterns of diseases in the U.S. elderly are not common in the U.S., especially in incidences of aging-associated chronic diseases. Incidence rates can be estimated using data from the national registries or from surveys representing the U.S. population; in the latter case, special procedures are required to generalize survey respondent results to the national level. One way to do that is to use a weight function (possibly, time-dependent) assigned to each individual such that “weighted” sums (or means) over individuals in the sample give estimates at the national level. A national survey with such a design is the 1982–2005 NLTCs which focuses on the U.S. elderly (65+) population¹. Another dataset to be used in the analysis is the SEER Registry data linked to Medicare Service Use Files (SEER-M). Using these extensive sources of information allows us to validate specific algorithms of incidence rate identification from administrative data and to prove their capability to evaluate the incidence rates of major aging-related diseases in U.S. elderly.

In this study, for the first time age patterns of incidence of common geriatric diseases were evaluated at the national U.S. level.

DATA AND METHODS

Both Medicare-linked datasets (SEER-M and NLTCs-M) contain information from MSUF since 1991. These data sets allow for reconstruction of individual histories of medical service use and, therefore, for modeling of individual follow-up from age 65 to death or onset of a disease of interest. All individuals in the SEER-M and NLTCs-M are longitudinally tracked for Medicare Part A and Part B service use. The records are available for each institutional (inpatient (INP), outpatient (OTP), skilled nursing facility (SNF), hospice (HSP), or home health agency (HHA)) and non-institutional (Carrier-Physician-Supplier (CAR) and durable medical equipment (DME) providers) claim types.

Two of the six NLTCs waves, namely cohorts of 1994 and 1999, were used for analysis: they were chosen primarily because the high quality Medicare follow-up data are available only since 1991 and the complete 5-year follow-up after the NLTCs interview is accessible only for these two waves after 1991. The NLTCs uses a sample of individuals drawn from the national Medicare enrollment files. In total, 34,077 individuals were followed-up for 5 years. So-called “screeener weights” released with the NLTCs were used in this study to produce the national population estimates.

The collection of SEER data began in 1973 and currently covers about 26% of the U.S. population. The SEER-M dataset includes Medicare records for individuals with diagnosed breast (n=353,285), colon (n=222,659), lung (n=342,961), prostate (n=448,410) cancers and skin melanoma (n=101,123), as well as Medicare records for 5% control. In total, Medicare records for 2,154,598 individuals are available in SEER-M.

Date of onset definitions

Nineteen diseases with high prevalence in elderly were selected for analyses (no diseases were initially selected but later excluded from analysis). The ages at onset of all studied diseases were reconstructed from the MFSU using the following scheme. First, the individual medical histories of the applicable disease were reconstructed from the Medicare files combining all records with their respective ICD-9 codes (listed in Table 1). Then a special procedure was applied for individuals with the history of the considered disease to separate the incident and prevalent cases. In the procedure used for identification of the date at onset, a date of a Medicare record (referred to as “*this record*” below in this subsection) was identified with the date of onset of applicable disease if both of the conditions mentioned below were met:

- i. *This record* was the earliest record with respective ICD code as a primary diagnosis in one of the four Medicare sources (inpatient care, outpatient care, physician services, and skilled nursing facilities). This choice is in accordance with the general practice of reconstruction of the date at onset from Medicare data²⁻⁴.
- ii. In addition to *this record*, there was another record with its respective ICD code as the primary diagnosis from one of the four Medicare sources listed in (i), which appeared with a date different from the date of *this record* and not later than 0.3 years after *this record*. Death occurred during this period (i.e., 0.3 years after *this record*) was also considered as the second record.

The first condition allows for identifying the first occurrence of disease code, and the second condition is required for confirmation of disease presence. This algorithm was used in studies of recovery after stroke⁵, medical cost trajectories before and after disease onset⁶, and the role of behavior factors in cancer risk⁷. The algorithm was implemented using SAS (SAS Institute, Inc., Cary, NC). Required running time for the computational server (2 quad core Xeon X5570 processors running at 2.93 Ghz with 32 GB of ram.) is several hours per disease.

RESULTS

For most of studied diseases, age-adjusted incidence rates obtained using the NLTCs-M and SEER-M were in an excellent agreement (see Figure 1). As an additional control for reasonability of the obtained incidence rates, the total mortality rate was also estimated. Among studied diseases, incidence rates of Alzheimer’s disease, stroke, and heart failure increased with age, while the rates of cancers of lung and breast, angina pectoris, diabetes, asthma, emphysema, arthritis, and goiter became lower at advanced ages. The rates obtained from the NLTCs-M were a little higher for incidence of most of non-cancer diseases (such as myocardial infarction, stroke, heart failure, diabetes, and ulcer) than the rates calculated using the SEER-M; the same difference was observed for the total mortality rates at ages 85+. For possible explanations of these findings, both methodological and substantive hypotheses are discussed in *Results* (those which can be tested using the NLTCs-M and SEER-M data) and *Discussion* (those requiring usage of external datasets) sections. For the majority of diseases, a decline of incidence rates with age was detected. Due to the use of weights in this analysis, the obtained age-specific rates and standard errors could provide estimates which are valid for general U.S. elderly population. Several types of age-patterns of disease incidence were observed in this study. The first one was a monotonic increase until age 85–95, with a subsequent slowing down, leveling off, and decline at age 100: for myocardial infarction, stroke, heart failure, ulcer, and Alzheimer’s disease. The second type had an earlier maximum and a more symmetric shape (an inverted U-shape): for cancers of the lung and colon, Parkinson’s disease, and renal failure. The majority of diseases (e.g., prostate cancer, asthma, and diabetes mellitus) demonstrated the third shape: a monotonic

decline or a decline after a short period of rate of increase. Melanoma and emphysema can also be assigned to this pattern, yet their patterns could also be considered flat.

The date of chronic disease onset cannot be defined with the same precision as the date of the death. A specific record with the disease code does not contain information for concluding whether the disease was actually diagnosed on the date of the record or whether it was diagnosed earlier; in the latter case, the record relates to the visit to treat the already-diagnosed disease. Therefore, the date of onset can be identified using information collected in the MFSU with specific assumptions outlining a specific algorithm of calculation. Besides, the approaches to the registration of certain chronic disease onset could vary by medical service provider and by the specialization of medical staff. Thus, there is certain arbitrariness in defining the date of onset which can be used for constructing a unified definition of date of onset appropriate for population studies. The scheme used in this paper has resulted from the review of the approaches used in several published studies for different diseases²⁻⁴. The evaluated incidence rates are in close agreement with those obtained in other studies. For example, age-adjusted cancer rates (65+, per 100,000) for five considered cancers calculated using the NLTCs-M and standard SEER Registry data are 521/448 (as in the NLTCs/SEER, respectively) for female breast, 1042/991 for prostate, 346/345 for lung, and 243/232 for colon. The age-specific rates of diabetes incidence obtained in our study are in agreement with those from the Canadian Study of Health and Aging, the Zwolle Outpatient Diabetes project Integrating Available Care (ZODIAC-1, the Netherlands⁸), and the UK Pooled Diabetes Study⁹. For stroke, our results are in a good agreement with those obtained in the Cardiovascular Health Study (CHS) and in the Framingham Heart Study (FHS) cohort studies (reviewed in NIH/NHLBI¹⁰), as well as from the Health Cost and Utilization Project (HCUP) study¹¹, and the Greater Cincinnati-Northern Kentucky Stroke Study (GCNKSS)¹². The FHS and CHS results better correspond to the rates obtained for NLTCs-M rather than for SEER-M data. Estimates for mortality rates from the Human Mortality database also better correspond to the estimates of the NLTCs-M.

When analyzing large administrative datasets, there is often an issue of existence of factors which could produce systematic over- or underestimation of the number of diagnosed diseases or of the age at onset. The reasons for such uncertainties could be the incorrect date of the disease onset, latent disenrollment, and incorrect reporting of date of birth and date of death: while the first affects the age at onset, the latter tends to reduce or increase the number of person-years at risk. To evaluate the effect of these uncertainties, we performed the calculations with different definitions of disease onset, and used several alternative censoring schemes to define individual observation periods. The following calculations were performed: 1) only inpatient records were kept, 2) confirmation of the diagnosis was not required, 3) confirmation of inpatient cases was not required, 4) not only primary code contributed to diagnosis, and 5) different cut on the frequencies of the health maintenance organization (HMO) coverage (i.e., coverage by an alternative insurance). We found that qualitatively the picture is similar for all considered cases. It could be concluded, that the reason for the differences between the circulatory disease rates obtained from the NLTCs-M and SEER-M are unlikely methodological.

DISCUSSION AND CONCLUSION

Age-specific incidence rates representing the major groups of chronic diseases in elderly (and therefore, resulting in high medical expenditures) were calculated using the NLTCs-M and SEER-M data. These datasets cover the same time range (1991–2005) and both are designed to represent estimates at the national level. The strategy for identifying the dates of onset is based on the analysis of complete trajectories of individual records associated with the selected diseases. The most appropriate scheme for the onset identification requires

forthcoming occurrence of repeated claims containing chosen ICD codes as a prime diagnosis in basic Medicare sources. Comparison of age-patterns observed in our study in both datasets revealed a qualitative (i.e., the form of shapes) and quantitative (i.e., the magnitude of rates) agreement for almost all diseases. Specifically, an excellent agreement was found for all types of cancers, angina pectoris, asthma, arthritis, and goiter. For Parkinson's, Alzheimer's disease, emphysema, and renal disease the agreement was also good: slightly higher for NLTCs but not exceeding $2\sigma_E$ (corresponds to $p=0.05$). For diabetes, COPD, and ulcer several points with such increase were above the level of $2\sigma_E$. For myocardial infarction, stroke, and heart failure the systematic increase of rates observed in the NLTCs was detected. This increase correlated with an increase in total mortality rate. One hypothesis allowing us to explain the trend is different populations represented in both datasets: the entire U.S. population in NLTCs, and population of the SEER regions in SEER-Medicare dataset.

The observed patterns suggest declining incidence rates of aging-related diseases with age. This is in agreement with the results of several recent studies on declining cancer incidence among the oldest elderly (e.g., older than 85 years old)^{13–19}. This is also in agreement with the autopsy studies²⁰ and with the observed morbidity profiles in the elderly (e.g., among centenarians, 87% of males and 83% of females delayed onset or completely bypassed the most lethal diseases^{21, 22}).

An occurrence of shapes with a maximum and, especially, with monotonic decline contradict the hypothesis that risk of geriatric diseases correlates with accumulation of adverse health events such as genetic mutations, deterioration of vascular system, immunosenescence, etc. Three basic concepts could be appealing in explaining such shapes. First, it could be attributed to the effect of selection²³ when frail individuals do not survive to the advanced ages; this approach is popular in cancer modeling and was successfully applied to the SEER data^{19, 24, 25}. The second explanation is related to the issue of a possible under-diagnosis of chronic diseases at advanced ages (due both to less pronounced disease symptoms in the elderly and to infrequent doctor's office visits); however, it cannot be proven with the available data^{26, 27}. It has also been suggested²⁸ that three components contributing to mortality while aging (i.e., basal, ontogenetic and time-dependent) may interact leading to different patterns of morbidity and mortality in the elderly population.

Comparing the obtained age patterns with the results of other studies demonstrated their close similarities between the U.S. and other countries. The patterns of the majority of diseases could be well described by the developed algorithm, with its most important feature such as an occurrence of primary diagnosis in one of four Medicare sources (inpatient care, outpatient care, physician services, and skilled nursing facilities) and confirmed diagnoses by a subsequent record. The performed analyses suggested that the national age-specific incidence patterns can be adequately evaluated from the MSUF.

The usefulness of the Medicare data is important, because there are few data sources to study such incidence patterns at advanced ages in the U.S. population. For example, heart disease and stroke account for more than 40% of all deaths among persons aged 65–74 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, the nationally representative datasets linked to Medicare data could be very useful in estimating the incidence events of aging-related diseases and associated medical costs. An advantage of the approach is in using specific information in datasets to which the Medicare data are linked, e.g., the NLTCs can provide disability-specific incidence rates allowing for projecting the estimates for the whole U.S.

population (so, the rates are valid at the national level) and the SEER-M allows us to investigate effects of comorbidity and specific characteristics of cancer (such as histotype-specific cancer rates). Disease-specific age at onset evaluated from these data could be used for estimation of the screening strategies, where population groups at the highest risk will be evaluated at specific age or age interval. Besides, the Medicare data also allows for investigating the relation of these age specific incidence patterns to Medicare costs, including future Medicare cost projection. Thus, the results reported in this study 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 when the effective therapeutic interventions became actively introduced in recent decades.

Acknowledgments

The research reported in this paper was supported by the National Institute on Aging grants R01AG027019, R01AG032319 and R01AG028259. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

Sponsor's Role: none.

References

1. Manton KG, Gu XL. Changes in the prevalence of chronic disability in the United States black and nonblack population above age 65 from 1982 to 1999. *Proc Natl Acad Sci U S A*. 2001; 98:6354–6359. [PubMed: 11344275]
2. Nattinger AB, Laud PW, Bajorunaite R, et al. An algorithm for the use of Medicare claims data to identify women with incident breast cancer. *Health Serv Res*. 2004; 39:1733–1749. [PubMed: 15533184]
3. Nattinger AB, Laud PW, Bajorunaite R, et al. Clarification note to an algorithm for the use of medicare claims data to identify women with incident breast cancer (vol 39, pg 6, 2004). *Health Serv Res*. 2006; 41:302–302.
4. Sloan FA, Brown DS, Carlisle ES, et al. Estimates of incidence rates with longitudinal claims data. *Arch Ophthalmol*. 2003; 121:1462–1468. [PubMed: 14557184]
5. Yashin A, Akushevich I, Ukraintseva S, et al. Trends in Survival and Recovery From Stroke: Evidence From the National Long-Term Care Survey/Medicare Data. *Stroke*. 2010; 41:563–565. [PubMed: 20093634]
6. Akushevich I, Kravchenko J, Akushevich L, et al. Medical Cost Trajectories and Onsets of Cancer and NonCancer Diseases in US Elderly Population. *Comput Math Methods Med*. 2011; 2011:857–892.
7. Akushevich I, Kravchenko J, Akushevich L, et al. Cancer risk and behavioral factors, comorbidities, and functional status in the US elderly population. *ISRN Oncology*. 2011; 2011:9.
8. Ubink-Veltmaat L, Bilo H, Groenier K, et al. Prevalence, incidence and mortality of type 2 diabetes mellitus revisited: A prospective population-based study in The Netherlands (ZODIAC-1). *Eur J Epidemiol*. 2003; 18:793–800. [PubMed: 12974556]
9. Gatling W, Guzder RN, Turnbull JC, et al. The Poole Diabetes Study: How many cases of Type 2 diabetes are diagnosed each year during normal health care in a defined community? *Diabetes Res Clin Pract*. 2001; 53:107–112. [PubMed: 11403859]
10. NIH/NHLBI. Incidence and Prevalence: 2006 Chart Book on Cardiovascular and Lung Diseases. Vol. 2010. Bethesda, MD: National Institutes of Health, National Heart, Lung, and Blood Institute; 2006.
11. Williams G. Incidence and characteristics of total stroke in the United States. *BMC Neurol*. 2001; 1. [PubMed: 11208259]
12. Feigin V, Lawes C, Bennett D, et al. Stroke epidemiology: A review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol*. 2003; 2:43–53. [PubMed: 12849300]

13. Bonafe M, Valensin S, Gianni W, et al. The unexpected contribution of immunosenescence to the leveling off of cancer incidence and mortality in the oldest old. *Crit Rev Oncol Hematol*. 2001; 39:227–233. [PubMed: 11500264]
14. de Rijke JM, Schouten LJ, Hillen HFP, et al. Cancer in the very elderly Dutch population. *Cancer*. 2000; 89:1121–1133. [PubMed: 10964343]
15. DePinho RA. The age of cancer. *Nature*. 2000; 408:248–254. [PubMed: 11089982]
16. Piantanelli L. Cancer and aging - from the kinetics of biological parameters to the kinetics of cancer incidence and mortality. *Ann N Y Acad Sci*. 1988; 521:99–109. [PubMed: 3377369]
17. Miyaishi O, Ando F, Matsuzawa K, et al. Cancer incidence in old age. *Mech Ageing Dev*. 2000; 117:47–55. [PubMed: 10958922]
18. Stanta G, Campagner L, Cavalieri F, et al. Cancer of the oldest old - What we have learned from autopsy studies. *Clin Geriatr Med*. 1997; 13:55–68. [PubMed: 8995100]
19. Manton, K.; Akushevich, I.; Kravchenko, J. *Cancer Mortality and Morbidity Patterns in the US population: An Interdisciplinary Approach*. 1. New York, NY: Springer; 2009.
20. Bernstein AM, Willcox BJ, Tamaki H, et al. First autopsy study of an Okinawan centenarian: Absence of many age-related diseases. *J Gerontol A Biol Sci Med Sci*. 2004; 59:1195–1199. [PubMed: 15602075]
21. Evert J, Lawler E, Bogan H, et al. Morbidity profiles of centenarians: Survivors, delayers, and escapers. *J Gerontol A Biol Sci Med Sci*. 2003; 58:232–237. [PubMed: 12634289]
22. Hitt R, Young-Xu Y, Silver M, et al. Centenarians: The older you get, the healthier you have been. *Lancet*. 1999; 354:652–652. [PubMed: 10466675]
23. Vaupel J, Carey J, Christensen K, et al. Biodemographic trajectories of longevity. *Science*. 1998; 280:855–860. [PubMed: 9599158]
24. Trussell J, Richards T. Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. *Soc Method*. 1985; 15:242–276.
25. Yashin AI, Akushevich I, Arbeev K, et al. Studying health histories of cancer: A new model connecting cancer incidence and survival. *Math Biosci*. 2009; 218:88–97. [PubMed: 19167410]
26. Enright P, McClelland R, Newman A, et al. Underdiagnosis and undertreatment of asthma in the elderly. *Chest*. 1999; 116:603–613. [PubMed: 10492260]
27. Solomon PR, Murphy CA. Should we screen for Alzheimer's disease? A review of the evidence for and against screening for Alzheimer's disease in primary care practice. *Geriatr Gerontol Int*. 2005; 60:26–31.
28. Ukraintseva S, Yashin A. How individual age-associated changes may influence human morbidity and mortality patterns. *Mech Ageing Dev*. 2001; 122:1447–1460. [PubMed: 11470132]

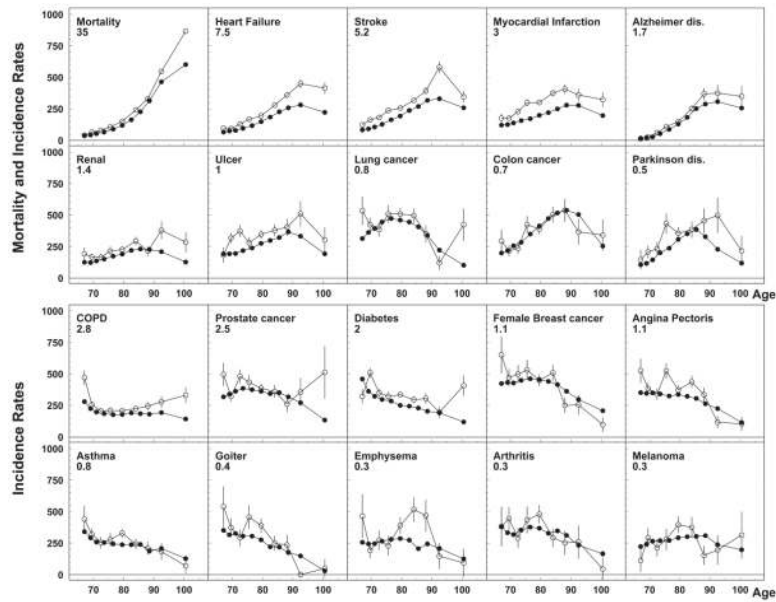


Figure 1.

Age-specific rates of total mortality and disease incidence calculated using NLTCs-Medicare (open dots) and SEER-Medicare (close dots). Rates for different diseases are rescaled to use the same scale on all plots to compare rates for different diseases: the original rate (per 100,000) can be calculated by multiplying the values obtained from plot by the rescaled factor. Upper panel represents diseases with age patterns of the first and second types (i.e., a monotonic increase of incidence with age with subsequent leveling off and decline) and lower panel represents the results of the age pattern of third type (a monotonic decline in incidence with age or decline following short period of increase)

Table 1

The ICD-9 codes used for the considered conditions

Group of diseases	Disease with ICD-9 codes
Cardio- and cerebrovascular	Myocardial infarction (410.xx), angina pectoris (413.xx), stroke (431.xx, 433.x1, 434.x1, 436.xx), heart failure (428.xx)
Malignancies	Lung cancer (162.xx), colon cancer (153.xx), breast cancer (females) (174.xx), prostate cancer (185.xx), skin melanoma (172.xx)
Neurogenerative	Parkinson's disease (332.xx), Alzheimer's disease (331.0)
Pulmonary	Chronic obstructive pulmonary disease (COPD) (490.xx, 491.xx, 492.xx, 493.xx, 494.xx, 495.xx, 496.xx), asthma (493.xx), emphysema (492.xx),
Endocrine and metabolic	Diabetes mellitus (250.xx), goiter (240.xx, 241.xx, 242.0x, 242.1x, 242.2x, 242.3x)
Miscellaneous	Chronic renal diseases with renal failure (403.xx, 404.xx, 585.xx, 250.4x, 249.4x), ulcer (531.xx, 532.xx, 533.xx, 534.xx), arthritis (714.0x, 714.1x, 714.2x, V82.1x)