## Cause-of-death statistics in public health and epidemiology

*Exploring new applications*

Mitratza, M.

**Publication date**
2021

**Document Version**
Final published version

**License**
Other

# Cause-of-Death statistics in public health and epidemiology:

## Exploring new applications

Marianna Mitratza

# Cause-of-Death statistics in public health and epidemiology: Exploring new applications

Marianna Mitratza


Cause-of-Death statistics in public health and epidemiology: Exploring new applications

**ISBN: 978-94-6421-195-5**


Cover: Marianna Mitratza
Printing: Ipskamp Printing, Enschede

# Cause-of-Death statistics in public health and epidemiology: Exploring new applications

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel

op dinsdag 23 februari 2021, te 15.00 uur

door

**Marianna Mitratza**

geboren te Amarousio

# Promotiecommissie

| | | |
|---|---|---|
| **Promotores:** | prof. dr. A.E. Kunst | AMC-UvA |
| | prof. dr. J.W.P.F. Kardaun | AMC-UvA |
| **Copromotores:** | dr. B. Klijs | Centraal Bureau voor de Statistiek |
| **Overige leden:** | prof. dr. L.J.A. Stalpers | AMC-UvA |
| | prof. dr. F.G. Schellevis | Vrije Universiteit Amstedam |
| | prof. dr. F. Janssen | Rijksuniversiteit Groningen |
| | dr. S.J. Pinto-Sietsma | AMC-UvA |
| | dr. F. Bakhshi-Raiez | AMC-UvA |

Faculteit der Geneeskunde

*"Medicine is a science of uncertainty and an art of probability."*

*William Osler*

*To my parents, to my twin-sister,*
*to Alex*

# Contents

# Chapter 1

General Introduction

Epidemiology, public health and medical knowledge use mortality statistics as an anchor outcome. Cause-of-death (CoD) databases are a rich data source of demographic and medical information with a well-recognized value. Already at the onset of the public health field in the 19th century, information collected from death certificates has been the major source to establish policies for prevention and control of diseases. In parallel with the demands of each era, the historical use of mortality data has shifted gradually from surveillance of communicable diseases towards monitoring of chronic noncommunicable diseases with high societal impact.

CoD registries offer many advantages for research and monitoring purposes. For instance, death is a clear endpoint, and data collection covers wide populations in space, such as countries, and over time, compared to other data. On top of that, a strong international coordination and strict regulations from national authorities are there to safeguard the consistency and quality of the data. Current applications involve global comparisons of mortality trends for main types of diseases, thanks to the vast amount of available data sources [1]. Several common uses of CoD data include exploring or testing new hypotheses with respect to the contribution of a disease to death. As detailed data for previously underrepresented populations become more easily accessible, inequalities in mortality are studied more [2, 3]. In addition, possibilities arise with developments of data linkages and data interoperability [4, 5].

However, there is room for improvement in the use of CoD data in order to capitalize on their potential value. It is not yet well-studied how many CoDs show a mortality time-trend when an agnostic search among the thousands of common and rare conditions would be performed. Moreover, new groups of CoDs could be formed according to updated medical knowledge, beyond the established categorizations, in order to measure their collective mortality burden in the population. Regarding the quality of CoD data use, only combinations with external morbidity information at the patient level can test their validity and broaden their utility, particularly for the study of lifelong chronic diseases.

In order to understand these opportunities in the use of CoD statistics, their main structure and properties should be briefly described first. The fundamental element of CoD statistics is the underlying cause-of-death. This is officially designated as "the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury", as defined by the World Health Organization (WHO) in 1948 [6]. Any other CoD mentioned on the death certificate is termed as the non-underlying cause-of-death. Collectively, any CoD reported on the death certificate, expressing any direct or indirect contribution to death, is usually termed

as multiple cause-of-death. Selection of the underlying CoD from the death certificate follows guidelines according to rules formulated by the WHO. This may have an impact on the produced CoD data, as several coding rules regulate the coding procedures with the purpose to prioritize more concretely defined conditions over ambiguous or deficient statements. Nevertheless, a known limitation of the CoD system is the room for subjective decision up to a certain extent by the certifying physician.

The development of standardized mortality statistics from death certificates requires the efficient classification of CoDs into meaningful categories. For the general purpose of the taxonomy of diseases in morbidity and mortality the International Classification of Diseases (ICD) has been developed and internationally adopted. The ICD was maintained from 1900 to 1948 by the International Statistical Institute, and since 1948 by WHO. Currently, the 10th version of the ICD is used for mortality statistics worldwide (ICD-10) by most countries [7]. Its basic structure is based on 3-digit codes organized into 21 Chapters, 19 of which are used for CoDs. The first five chapters cover more broad and general categories of disease. The following nine chapters focus on a system or organ, and the remaining chapters are more specific, for instance, describing external causes or congenital malformations. Occasionally, 3-digit codes may correspond to "other" and "unspecified" diseases. External causes get much attention, because of their legal and judicial consequences. More opportunities for the refinement of disease classification exist by aligning updated medical knowledge with the needs of the users.

In view of the rapid increase of the ageing population, and consequently of the rising age at death, it is important to assess the potentials and shortcomings of using existing CoD data in public health and epidemiology. Single underlying CoDs have been historically the main source of mortality data. However, currently the more complex multiple CoD data are available in many countries, allowing us to further explore interactions among diseases at the end of life. As life expectancy has increased significantly and non-fatal chronic diseases have become more important, more people suffer from multimorbidity, i.e. from more than one disorder, before they die. A natural upcoming challenge would be to study the different diseases that people die with, and not only the single disease from which a person would die from.

At the same time, new opportunities arise in the era of big data. Vast amounts of healthcare data can be easily linked at the individual level while respecting data privacy regulations. Therefore, we need to re-evaluate the place of CoD data in these new settings. For instance, diseases at the latest phase of life would be the most relevant to study with a design that incorporates a CoD database. The informative value of a CoD database along with

multiple other datasets is also a field that needs further exploration, in order to decide upon the value of each register for the particular disease.

In this thesis, new methods to use CoD statistics in the face of these challenges are explored in three parts. We develop innovative approaches in order to illustrate novel uses of cause-of-death data for monitoring time-trends (**Part I**), assessing the co-morbidities of diseases at the time of death (**Part II**), as well as estimating the lifetime risk for a disease (**Part III**).

## Part I

### Using underlying Cause-of-Death data to monitor trends over time

Monitoring mortality time-trends provides important epidemiological knowledge for research and planning of public health policies. Statistics on mortality trends are often presented in a public platform by national statistical offices; these are based on aggregate level data and include only the underlying CoDs for the majority of countries. Typically, the focus lies on few specific causes of interest or systematic, abbreviated lists of leading CoDs. Currently, applications include the identification of inequalities in mortality trends among demographic groups of age and sex and associations with risk factors, such as alcohol and smoking [8], or the discovery of clusters in mortality patterns [9]. Comparing mortality trends over time informs the prioritization of diseases for preventive actions, particularly regarding the most affected populations. When certain types of CoDs are monitored, for instance, as part of infectious diseases' surveillance, the implications for public health practice and medical research are obvious. Users can evaluate directly outcomes of the developed implementations, such as a vaccination program for varicella [10].

The use of selected CoDs, however, involves significant choices for some diseases at the expense of others. Diseases may be chosen based on their high prevalence, their known association with an exposure, or their high cost in terms of healthcare provisions. However, this approach may unintentionally restrict the range of CoDs for monitoring purposes, including potentially relevant ones, and thus could result in the limited use of available mortality data. When analysing mortality time-trends, it may be more pragmatic to complement established lists of CoDs with a selection of CoDs based on their corresponding annual number of deaths.

The ICD-10 classification system already includes more than 1,700 codes at the three-position level, and the forthcoming ICD-11 release will have a large scale up of codes' dimensions [11]. Mapping diseases in ICD using broad or narrow codes may have an impact on the analysis of mortality trends over time, as it has been shown, for instance, in idiopathic pulmonary fibrosis [12]. In this context, it remains unknown what level of detail in ICD coding researchers should use with existing mortality data to monitor trends. A similar challenge is posed for national statistical offices that have to decide on which cause-specific mortality trends to publish along with common pre-specified causes-of-death lists. These issues need further exploration, so that these decisions are shaped with a fair trade-off between producing valuable signals for health policy, and producing uninformative signals, such as large random fluctuations from more rare CoDs. An

additional point for consideration is how such decisions could be adapted in international settings, as smaller countries cannot monitor the same extended list of CoDs as the big ones, due to small numbers of observed deaths.

In Part I of this thesis, to investigate the above aspects of underlying CoD use, the relation of trends to CoD size is studied. We pay particular attention to the mean annual number of deaths, which expresses the rarity of a disease or condition that is selected as underlying CoD in a population. The CoD size can vary by a factor of 1,000 or more. We first use the Dutch CoD database to find long- and short-term mortality trends and assess the impact of CoD size to the detection of trends (**Chapter 2**). Importantly, population numbers, and hence, the number of deaths, varies by a factor of 100 between the countries of the European Union. As a second study, we extend our approach to European countries, and consider further the role of CoD type and country. Our aim is to generate a rule of thumb for identifying eligible CoDs in terms of a minimum CoD size that is expected to allow for the detection of a long-term trend (**Chapter 3**).

## Part II

### Using multiple Cause-of-Death data to assess comorbidities at the end-of-life stage

Multiple CoD data broaden the potentials of CoD data for research and public health practice. By incorporating non-underlying CoDs, mortality data offer a source with information on sometimes many diseases for a deceased person. Eight out of ten death certificates have more than one cause of death listed, and on average 2.4–3.1 diseases are mentioned for each natural death, depending on the country [13-15]. Non-underlying CoDs can play different roles on a death certificate, ranging from a complication of the underlying CoD to a condition which contributed to death via a different pathway compared to the causal chain initiated by the underlying CoD. With these data, more complex study methods are needed. The most common use of multiple CoD data is estimating the overall prevalence of a disease or condition by the time of death, counting diseases not only as the underlying CoD [16, 17].

In principle, a common single disease is studied, such as atrial fibrillation [18] or a more infrequent CoD, such as an infectious disease [19, 20] or a rare disease, such as neurofibromatosis type 1 [21]. In such cases, using multiple CoD data, the overall direct and indirect burden of a disease in mortality is measured. Thus, one can find important

information on the public health burden of deaths related to the disease under investigation. When a condition, for instance a surgical procedure or a complication, is not commonly recorded as the underlying CoD, but instead as a non-underlying CoD, the demographic, temporal, and spatial associations of this condition can be quantified in a more robust way using multiple CoD data. A more specialized application of multiple CoD data analysis is investigating the aetiology of a rare unspecified disease, such as unspecified encephalitis [22].

An increasingly acknowledged application of multiple CoD data in public health and epidemiology is for studying associations of comorbidities at the time of death. Often, multiple CoD studies employ methods to compare the occurrence of a non-underlying CoD in people dying from a specific underlying CoD compared to the general deceased population [23, 24]. In this manner, useful insights regarding the relations between diseases can be extracted.

The value of multiple CoD data can be realised especially for certain population groups, such as the elderly, who often suffer from multimorbidity [25, 26]. One reason is that a common complication of several diseases, such as heart failure [27], is discouraged by coding rules to be selected as underlying CoD when other diseases are present. A second reason is that many chronic diseases are rarely fatal, but serious enough to be listed on the death certificate, such as chronic obstructive pulmonary disease [28], psychiatric diseases [29], diabetes [30], or dementia [31].

Modern medicine advocates the wider movement from single diseases to pathophysiological mechanisms in order to classify them and study their epidemiology [32]. Multiple CoD data may be particularly important for several chronic rare diseases that represent a common underlying pathophysiologic mechanism [33, 34]. Thus, these CoDs may pose a significant mortality burden altogether that would go unnoticed when studied separately, such as the mortality burden of viral hepatitis [35] or infectious diseases [36]. This type of studies focusing on novel groupings of CoDs may have added value and reveal potential for common prevention and treatment strategies.

In Part II, **Chapter 4**, we move our attention towards the case study of systemic autoimmune diseases (SAIDs). This group of diseases serves as an illustration of performing a multiple CoD study for an understudied pathophysiological mechanism for various reasons. On the one hand, the research field of autoimmune diseases is continuously expanding, resulting in new discoveries of rare CoDs. On the other hand, the ICD-10 classification needs adjustment in order to reorganise the currently scattered SAID codes across tract

and organ-based chapters into a uniform class. Moreover, SAIDs are severe rare chronic diseases, which are often not selected as underlying CoDs in the presence of other very prevalent CoDs. This rarity would also make difficult and expensive to develop cohorts to study their association with mortality, making the option of multiple CoD studies even more attractive. Last, SAIDs, by definition, can cause multi-organ damage. Therefore, the study of comorbidities as reported on death certificates can fill knowledge gaps regarding the extent of known complications.

## Part III

### Enrichment of Cause-of-Death data with healthcare registries to assess disease importance

There are occasions where CoD data alone may not be adequate to assess the importance of a disease for mortality. For instance, certain diseases are often underreported on the death certificate, as they may be considered more as a risk factor, such as obesity [37] or not fitting the definition of underlying CoD, such as dementia [38] and hypertension [39]. Different types of external data can prove useful for complementing CoD data. More specifically, national healthcare registrations contain data based on healthcare interactions which are collected routinely for administrative purposes. These can offer new opportunities for medical research and public health studies. Primary care or long-term care data, hospital discharge registrations, claims, as well as medication registrations are examples of data sources that be used for enrichment of CoDs when linked to the CoD database.

So far, combined CoD and morbidity data in public health and epidemiology are used both for methodological and substantive studies. Main methodological applications involve validation purposes, such as the comparison of CoDs with cancer registers which can provide information on the coverage and the reliability of each register [40, 41]. Similarly, the completeness of the CoD database can be investigated by estimating the agreement between hospital and CoD data [42]. Substantive studies with mortality as the endpoint of interest are often enabled by linkage to patient level data. In that way, the mortality risks of patients suffering from a disease, such as affective psychotic disorder [43] can be studied.

Linkage of the CoD database with healthcare data sources enables the study of the concept of 'dying with a chronic disease' in a systematic manner [44-46]. This is complementary to

the concept of 'dying from a disease', which is implicit in underlying CoDs. Individuals are worried about how likely they will carry a disease until the end stages of their life, with its potential corresponding needs for care from family members or the healthcare system. From a more practical point of view, knowing how many people die with a disease in the population can be very useful for healthcare planning of human resources and facilities regarding the end-of-life phase [47]. Moreover, the concept of 'dying with a disease' offers public health practitioners a straightforward measure to communicate about the importance of chronic disease to individual people.

In Part III, we illustrate two chronic diseases for which enrichment of CoD with healthcare data can be valuable. The case study of diabetes mellitus, in **Chapter 5**, investigates the end-of-life occurrence of the disease in a sample of the Dutch population. This disease was selected because it may start at a younger age, with a large impact on the quality of life over lifetime, and it may require complex care at the end-of-life phase. Except for primary care and CoD data, medication and hospital discharge registrations are used. The case study of dementia, in **Chapter 6**, estimates the lifetime risk of the disease in the national Dutch population. Dementia affects populations from the mid-phase of life in the earliest cases up to the end of life, and requires complex type of care too. As there is no current treatment available, but only drugs aiming at delaying the course of the disease, intensive preventive efforts are deemed as the most responsible response to mitigate the problem as much as possible [48].

## General aim

Our general aim is to illustrate new potential applications of already available mortality data for public health monitoring and epidemiological purposes of researchers and statistical offices. We design explorative case studies, so that new applications are illustrated by example, rather than by theory.

## Specific objectives

Our objectives are to perform studies and develop methodologies in order to enhance (1) the use of underlying CoD data to monitor trends over time in cause-specific mortality, (2) the use of multiple CoD data to study associations between diseases at the end-of-life stage and (3) the use of CoD data linked with healthcare registries to assess the prevalence of diseases at the end of life.

**Table1.** Overview of studies presented in this thesis

| Part | Chapter | Data sources | Statistical methods | Disease |
|---|---|---|---|---|
| I | 2 | aggregated Underlying CoD (Netherlands) | polynomial regression models, analysis of outliers in time-trends | 625 Causes of Death |
| | 3 | aggregated Underlying CoD (21 European countries) | CoD size threshold with receiver operating characteristic curve diagnostics | 202–791 Causes of Death per country |
| II | 4 | Multiple CoD (Netherlands) | multiple-cause-of-death analysis; association measures with 20 comorbidities | systemic autoimmune diseases and subgroups |
| III | 5 | Multiple CoD, primary care, hospital discharge, dispensed medications (Netherlands) | overlaps of data sources, sequential addition | diabetes mellitus |
| | 6 | Multiple CoD, primary care, hospital discharge, dispensed medications, long-term care, specialized mental care, claims (Netherlands) | overlaps of data sources, sequential addition | dementia |

CoD = Cause of Death

# Outline of this thesis

This thesis consists of three main parts, as presented in **Table 1**. **Part I** consists of two chapters aimed to understand how the size of underlying CoD affects the opportunity to detect mortality time trends. **Chapter 2** illustrates our methodology to empirically assess mortality time-trends in the short and long term in the Netherlands, using polynomial regression models and outlier analysis. **Chapter 3** elaborates further on this topic by extending our method to all European countries and by deriving a threshold for detecting trends with the use of receiver operating characteristic curve diagnostics.

**Part II** consists of **Chapter 4**, which focuses on the use of multiple CoD data after re-classifying systemic autoimmune diseases as a single group of causes of death. We perform a multiple-cause-of-death analysis in Dutch CoD data. **Part III** consists of two chapters aimed to assess the contribution of a chronic disease to death, based on CoD data enriched with nationwide healthcare registry data. **Chapter 5** investigates different combinations of data sources to obtain estimates of the occurrence of diabetes mellitus at the end of life, by employing a sample of Dutch primary care registrations as a starting point. **Chapter 6**, follows a similar approach to estimate the lifetime risk of dementia, using several nationwide databases in the Netherlands.

In **Chapter 7**, the final chapter of this thesis, an overview of the main findings is provided, and the methodological strengths and limitations are discussed. Moreover, reflections on the key findings in relation to the objectives of this thesis are given. Finally, potential implications for regular users of mortality data are discussed and topics for future research are recommended.

# References


1. GBD 2017 Causes of Death collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1736-88.
2. Lewer D, Jayatunga W, Aldridge RW, et al. Premature mortality attributable to socioeconomic inequality in England between 2003 and 2018: an observational study. Lancet Public Health. 2020;5(1):e33-e41.
3. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends-an up-date. Cancer Epidemiol Biomarkers Prev. 2016;25(1):16-27.
4. Gini R, Francesconi P, Mazzaglia G, et al. Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey. BMC Public Health. 2013;13:15.
5. Pastorino R, De Vito C, Migliara G, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health. 2019;29(Suppl):23-7.
6. Manual of the international statistical classification of diseases, injuries, and causes of death. Sixth revision. Geneva, World Health Organization, 1949.
7. World Health Organization-WHO. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; World Health Organization: Geneva, Switzerland, 1992; Volume I-III.
8. Janssen F and van Poppel F. The Adoption of Smoking and Its Effect on the Mortality Gender Gap in Netherlands: A Historical Perspective. Biomed Res Int. 2015;2015:370274.
9. Frova L, Salvatore MA, Pappagallo M, Egidi V. The multiple cause of death approach to analyse mortality patterns. Genus. 2009;65(1):1-21.
10. Leung J, Bialek SR, Marin M. Trends in varicella mortality in the United States: Data from vital statistics and the national surveillance system. Hum Vaccin Immunother. 2015;11(3):662-8.
11. World Health Organization-WHO. Available online: https://icd.who.int/browse11/l-m/en.
12. Hutchinson JP, McKeever TM, Fogarty AW, et al. Increasing global mortality from idiopathic pulmonary fibrosis in the twenty-first century. Ann Am Thorac Soc. 2014;11(8):1176-85.
13. Australian Institute of Health and Welfare 2012. Multiple causes of death. Bulletin no. 105. Cat. no. AUS 159. Canberra: AIHW.
14. Koch TM. Multiple Underlying Causes of Death using Montana Death Certificates, 1999-2014. Montana: Office of Epidemiology Scientific Support; 2015
15. Désesquelles A, Gamboni A, Demuru E, Grundy E. We only die once… but from how many causes? Population and Societies 2016;534:1-4.
16. Redelings MD, Sorvillo F, Simon P. A comparison of underlying cause and multiple causes of death: US vital statistics, 2000-2001. Epidemiology (Cambridge, Mass.). 2006;17(1):100-3.
17. Moreno-Betancur M, Sadaoui H, Piffaretti C, Rey G. Survival analysis with multiple causes of death: Extending the competing risks model. Epidemiology. 2017;28(1):12-9.

18. Fedeli U, Avossa F, Ferroni E, et al. Contemporary Burden of Atrial Fibrillation and Associated Mortality in Northeastern Italy. Am J Cardiol. 2016;118(5):720-4.
19. Arif N, Yousfi S, Vinnard C. Deaths from Necrotizing Fasciitis in the United States, 2003-2013. Epidemiol Infect. 2016;144(6):1338-44.
20. Khuu D, Shafir S, Bristow B, Sorvillo F. Blastomycosis Mortality Rates, United States, 1990-2010. Emerg Infect Dis. 2014;20(11): 789-94.
21. Masocco M, Kodra Y, Vichi M, et al. Mortality associated with neurofibromatosis type 1: A study based on Italian death certificates (1995-2006). Orphanet J Rare Dis. 2011;6:11.
22. Tack DM, Holman RC, Folkema AM, et al. Trends in encephalitis-associated deaths in the United States, 1999-2008. Neuroepidemiology. 2014;43(1):1-8.
23. Avouac J, Amrouche F, Meune C, et al. Mortality profile of patients with rheumatoid arthritis in France and its change in 10 years. Semin Arthritis Rheum. 2017;46(5):537-43.
24. Cummings PL, Kuo T, Javanbakht M, Sorvillo F. Trends, productivity losses, and associated medical conditions among toxoplasmosis deaths in the United States, 2000-2010. Am J Trop Med Hyg. 2014;91(5):959-64.
25. Gorina Y and Lentzner H. Multiple causes of death in old age. Aging Trends. 2008;(9):1-9.
26. Goldberger N, Applbaum Y, Meron J, Haklai Z. High Israeli mortality rates from diabetes and renal failure – Can international comparison of multiple causes of death reflect differences in choice of underlying cause? Isr J Health Policy Res. 2015;4:31.
27. Rahimi K, Duncan M, Pitcher A, et al. Mortality from heart failure, acute myocardial infarction and other ischaemic heart disease in England and Oxford: a trend study of multiple-cause-coded death certification. J Epidemiol Community Health. 2015;69(10):1000-5.
28. Marcon A, Saugo M, Fedeli U. COPD-related mortality and co-morbidities in Northeastern Italy, 2008-2012: a multiple causes of death analysis. 2016;13(1):35-41.
29. Polednak AP. Trend in rates for deaths with mention of schizophrenia on death certificates of US residents, 1999-2010. Soc Psychiatry Psychiatr Epidemiol. 2014;49(7):1083-91.
30. Romon I, Jougla E, Balkau B, Fagot-Campagna A. The burden of diabetes-related mortality in France in 2002: an analysis using both underlying and multiple causes of death. Eur J Epidemiol. 2008;23(5):327-34.
31. Désesquelles A, Demuru E, Salvatore MA, et al. Mortality from Alzheimer's disease, Parkinson's disease, and dementias in France and Italy: a comparison using the multiple cause-of-death approach. J Aging Health. 2014;26(2):283-315.
32. Dozmorov MG. Reforming disease classification system-are we there yet? Ann Transl Med. 2018;6(Suppl 1):S30.
33. Arakelyan A, Nersisyan L, Poghosyan D, et al. Autoimmunity and autoinflammation: a systems view on signaling pathway dysregulation profiles. PLoS One. 2017;12:e0187572.
34. McGonagle D and McDermott MF. A proposed classification of the immunological diseases. PLoS Med. 2006;3:e297.
35. Ly KN, Xing J, Klevens RM, et al. Causes of Death and Characteristics of Decedents With Viral Hepatitis, United States, 2010. Clin Infect Dis. 2014;58(1):40-9.

36. Désesquelles A, Demuru E, Pappagallo M, et al. After the epidemiologic transition: a reassessment of mortality from infectious diseases among over-65s in France and Italy. Int J Public Health. 2015;60(8):961-7.

37. Barbieri M, Désesquelles A, Egidi V, et al. Obesity-related mortality in France, Italy, and the United States: a comparison using multiple cause-of-death analysis. Int J Public Health. 2017;62(6):623-9.

38. Romero JP, Benito-León J, Louis ED, Bermejo-Pareja F. Under reporting of dementia deaths on death certificates: a systematic review of population-based cohort studies. J Alzheimers Dis. 2014;41(1):213-21.

39. Kung HC and Xu J. Hypertension-related Mortality in the United States, 2000-2013. NCHS Data Brief. 2015;(193):1-8.

40. Fink AK, German RR, Heron M, et al. Impact of using multiple causes of death codes to compute site-specific, death certificate-based cancer mortality statistics in the United States. Cancer Epidemiol. 2012;36(1):22-8.

41. De Vries S, Schaapveld M, Kardaun JWPF, et al. Comparing causes of death of Hodgkin lymphoma and breast cancer patients between medical records and cause-of-death statistics. Clin Epidemiol. 2018;10: 1523-31.

42. Klijs B, Nusselder WJ, Mackenbach JP. Nationwide individual record linkage study showed poor agreement of causes of death and hospital diagnoses at individual level but reasonable agreement at population level. J Clin Epidemiol. 2014;67(2):160-8.

43. Termorshuizen F, Wierdsma AI, Visser E, et al. Psychosis and suicide risk by ethnic origin and history of migration in the Netherlands. Schizophr Res. 2012;138(2-3):268-73.

44. Murtaugh CM, Spillman BC, Wang XD. Lifetime risk and duration of chronic disease and disability. J Aging Health. 2011;23(3):554-77.

45. Haneuse S. Multiple causes of death: The importance of substantive knowledge in the Big Data era. Epidemiology. 2017;28(1):28-9.

46. Licher S, Heshmatollah A, van der Willik KD, et al. Lifetime risk and multimorbidity of non-communicable diseases and disease-free life expectancy in the general population: A population-based cohort study. PLoS Med. 20194;16(2):e1002741.

47. Bone AE, Gomes B, Etkind SN, et al. What is the impact of population ageing on the future provision of end-of-life care? Population-based projections of place of death. Palliat Med. 2018;32(2):329-36.

48. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020;396(10248):413-46.

# Part I

## Using underlying Cause-of-Death data to monitor trends over time

# Chapter 2

## Detecting mortality trends in the Netherlands across 625 Causes of Death

**Marianna Mitratza**, Anton E. Kunst, Jan W.P.F. Kardaun

*Published in*
*International Journal of Environmental Research and Public Health 2019;16(21):4150.*

## Abstract

**Background:** Cause of death (CoD) data are essential to public health monitoring and policy. This study aims to determine the proportion of CoDs, at ICD-10 three-position level, for which a long-term or short-term trend can be identified, and to examine how much the likelihood of identifying trends varies with CoD size.

**Methods:** We calculated annual age-standardized counts of deaths from Statistics Netherlands for the period 1996–2015 for 625 CoDs. We applied linear regression models to estimate long-term trends, and outlier analysis to detect short-term changes.

**Results:** The association of the likelihood of a long-term trend with CoD size was analyzed with multinomial logistic regression. No long-term trend could be demonstrated for 216 CoDs (34.5%). For the remaining 409 causes, a trend could be detected, following a linear (211, 33.8%), quadratic (126, 20.2%) or cubic model (72, 11.5%). The probability of detecting a long-term trend increased from about 50% at six mean annual deaths, to 65% at 22 deaths and 75% at 60 deaths. An exceptionally high or low number of deaths in a single year was found for 16 CoDs.

**Conclusions:** When monitoring long-term mortality trends, one could consider a much broader range of causes of death, including ones with a relatively low number of annual deaths, than commonly used in condensed lists.

# Introduction

Mortality statistics are a key source of information in public health, epidemiology and medicine. As part of the vital statistics registries, they cover entire national populations, extend over long periods of time and are easily accessible. Among others, these statistics are used to monitor cause-specific mortality, in order to identify trends that may inform disease prevention, screening and surveillance.



Both short- and long-term mortality trends can be monitored to identify changes that may prompt health policy actions. Trends in the short term, defined as a maximum period of a year, are particularly important to identify sudden population-level changes and events, such as outbreaks of communicable diseases. Long-term mortality trends, defined as secular changes, as observed across several years, may reflect gradual changes in the incidence or case-fatality of specific diseases or injuries.

Causes of death (CoDs) are – in most countries – classified according to the 10th revision of the International Classification of Diseases (ICD-10) [1]. The ICD-10 contains 1761 codes at a three-position level, which refers to a single condition, a group of diseases, "other" or "unspecified" conditions, and more than 14,000 codes at the four- and five-position level. Studies monitoring trends in mortality by cause of death are often restricted to a few specific causes of interest, or to a systematic, but abbreviated, list of leading CoDs [2, 3].

The large number of ICD-10 codes raises the question of how many of these codes should be distinguished when tabulating, analysing and publishing mortality trends. This applies especially to national statistical offices, both within Europe and in other continents, as they have to decide on the level of detail to include when publishing annual cause-specific mortality data. Monitoring all available ICD-10 codes, such as the 1761 codes at the three-position level, may produce valuable signals for health policy, that would go unnoticed when using abbreviated lists of leading CoDs. On the other hand, this approach may also produce uninformative signals, such as large random fluctuations in the annual number of deaths from smaller causes.

The size of a CoD – in terms of mean annual number of deaths – may offer a criterion to decide which ICD-10 codes should be distinguished. All else being equal, the likelihood of detecting a trend increases with an increase in CoD size. A detailed description of this relationship may, therefore, help develop guidelines for selecting CoDs. Such guidelines may be particularly relevant regarding the forthcoming release of the ICD-11 version, which may increase the number of possible CoD codes, and consequently decrease the

average number of deaths per CoD code. The 2018 version of ICD-11 for Mortality and Morbidity Statistics has more than double the number of disease entity codes, compared to ICD-10 [4].

The general aim of this study was to determine the level of detail, in terms of CoD size, at which the ICD-10 classification can be best used for presenting mortality trends in national populations. The specific objective of this study was to estimate the proportion of causes of death for which we can observe a long-term trend or a short-term change, and how the probability of finding a long-term trend relates to CoD size. To do this, we assessed trends in CoD in the Netherlands over a period of 20 years.

## Materials and Methods

We analyzed annual mortality data from Statistics Netherlands [5] for the period 1996–2015, regarding each underlying CoD at the ICD-10 three-position level, as reported on the death certificate and coded at Statistics Netherlands. This period was chosen as the ICD-10 was introduced in the Netherlands in 1996, and 2015 was the last year with available data at the time of initiation of this study.

We excluded 1136 CoDs with less than three annual deaths on average, because most of these CoDs had predominantly zero or only zero annual deaths. No death (only zero values) was observed for 415 CoDs throughout the 20-year period, while the rest of the excluded CoDs accounted together for 476 deaths per year on average, which corresponded to 0.4% of the total number of deaths.

For the remaining 625 CoDs, we calculated age-standardized counts of deaths using age-standardized mortality rates, calculated with the direct method, using the Dutch population in 2005 as a reference population (i.e., mid-period). This method intended to control for annual changes in the age distribution of the population. The population size was 15.5 million persons in 1996, with the age distribution (<20, 20–40, 40–65, 65–80, 80 years) being 24%, 32%, 31%, 10% and 3%, respectively. The corresponding numbers for 2015 were 16.9 million persons, with the age distribution being 23%, 25%, 35%, 13% and 4%, respectively. In the Netherlands, 137,561 deaths occurred in 1996, and 147,134 deaths in 2015. The age-standardized counts – further termed number of deaths – are available in an additional file (**Supplementary Table S1**).

We assessed time-trends in CoDs by applying two complementary methods: identification of gradual long-term trends by means of polynomial models, and identification of sudden year-by-year changes by means of outlier detection.

In more detail, we assessed how many CoDs (at the ICD-10 three-position level) would demonstrate a statistically significant change in mortality over a 20-year period. Possible trends were assessed by applying four different linear regression models, with polynomial terms of year added as independent covariates. We used orthogonal polynomials that account for multicollinearity of the polynomial components [6]. We applied the constant, the linear, the quadratic and the cubic model and evaluated them with the following hierarchical approach. Firstly, all four models were fitted, and we used the lowest corrected Akaike Information Criterion (AICc) to select the best model. We used the AICc as it performs better than the AIC for small sample sizes [7]. Secondly, the best model for each COD was compared with the constant model using the F-test, at a significance level of $\alpha = 0.05$. If the best model performed better than the constant model with statistical significance, it remained the final best model. Otherwise, the constant model became the final best model for this COD.

In the development of the four polynomial models, a bivariate step parameter (break point) for deaths before or during/after 2013 was introduced, to allow for the possible effect of the introduction of automated coding of mortality data, which replaced the manual method of coding in the Netherlands in 2013 [8], as well as the implementation of the cumulative WHO updates to ICD-10, for the period 1996–2013 [9].

The association between CoD size and likelihood of fitting a long-term trend (linear, quadratic, cubic) was analyzed with multinomial logistic regression, taking the logarithm of the mean number of deaths for the CoD size, since its distribution was highly skewed.

Next, we assessed the number of CoDs with significant sudden changes across the monitoring period. For each CoD, we aimed to detect years with extreme observations in the number of deaths, which were defined as residuals of the final best model, at the significance level (alpha) of 0.01 and 0.001. These cases were identified using the outlierTest function from the "car" package in R software, with a cut-off of 0.01 and 0.001, respectively. The Bonferroni p-values were obtained assuming a t-distribution with degrees of freedom (df) equal to the residual df for the model minus one [10].

In order to illustrate how the use of condensed lists of ICD-10 codes can lead to other selections when monitoring causes of death, we repeated the analysis of long-term trends

with ICD-10 three position codes, aggregated as "chapters" or "blocks" of conditions [1]. The ICD-10 consists of 22 chapters, 19 of which are used for coding the underlying cause of death, and each chapter consists of one or more blocks of three position codes. For example, Chapter IX (Diseases of the circulatory system) includes a block like I20-I25 (Ischaemic Heart disease), which includes diseases like I21 (Acute Myocardial Infarction).

In the analysis of long-term trends, we used linear regression models, assuming normal error distribution, as we were primarily concerned with changes in absolute terms. Yet, statistical theory says that numbers of deaths are Poisson distributed, and that regression models with a log-linear link function would be more appropriate. However, the approximation of a Poisson by a normal error distribution is said to be adequate if the mean number of observations is about five or more. Moreover, in sensitivity analysis, we found that the main findings would not substantially change if we were to use Poisson regression instead of normal regression analysis (results not shown).

All analyses were conducted with R software (3.3.1 version) [11].

No ethics approval or consent to participate was necessary, as we used publicly available population data.

# Results

For a substantial amount of CoDs (216, 34.5%), we could not detect a long-term (linear, quadratic, cubic) trend (**Table 1**). For the remaining 409 causes, an annual trend was detected, following a linear (211), quadratic (126) or cubic (72) model.

The descriptive statistics of the association between the probability of detecting an annual trend in mortality and the size of cause of death are given in **Table 1**. As expected, polynomial models of higher order were more often selected when the median number of deaths increased. The median number of deaths in the group of causes that were best described by the constant model, was 9.5 (20.4 for linear model, 38.1 for quadratic model and 56.2 for cubic model). Among all 625 CODs, a trend could be detected in 65.5% of cases (33.8% for linear model, 20.2% for quadratic model and 11.5% for cubic model).

**Table 1.** Distribution of the Causes of Death (ICD-10 three-position), according to the final best model, stratified by their size.

| | Final Best Model | | | | |
|---|---|---|---|---|---|
| | Constant (no detectable trend) | Linear | Quadratic | Cubic | Total |
| Cause of Death Size [a] | Number of Causes of Death (% of all Causes of Death with that Size) | | | | |
| [3,5) | 55 (59.2) | 27 (29.0) | 8 (8.6) | 3 (3.2) | 93 (100.0) |
| [5,15) | 78 (45.1) | 64 (37.0) | 24 (13.9) | 7 (4.0) | 173 (100.0) |
| [15,30) | 32 (29.4) | 38 (34.9) | 23 (21.1) | 16 (14.7) | 109 (100.0) |
| [30,100) | 35 (27.8) | 44 (34.9) | 31 (24.6) | 16 (12.7) | 126 (100.0) |
| [100,9750) | 16 (13.0) | 38 (30.6) | 40 (32.5) | 30 (24.4) | 124 (100.0) |
| Total | 216 (34.5) | 211 (33.8) | 126 (20.2) | 72 (11.5) | 625 (100.0) |
| | Median number of deaths [IQR] [b] | | | | |
| | 9.5 [4.9–27.4] | 20.4 [7.7–56.7] | 38.1 [14.8–162.4] | 56.2 [22.0–479.2] | 20.7 [7.6–71.8] |

a) [*i*,*j*] is an interval notation for all values between *i* (included) up to *j* (not included).
b) IQR: interquartile range.

The association between CoD size and selected polynomial degree was significant at $\alpha = 0.001$. For every 10% increase in the mean number of deaths, there was a 3% increase in the odds of having the linear, as opposed to the constant, model (OR: 1.03, 95%CI: 1.02 ; 1.05). Similarly, the odds of having the quadratic or the cubic model compared to the constant model increased by 6% (OR: 1.06, 95%CI: 1.04 ; 1.07) and 8% (OR: 1.08, 95%CI: 1.06 ; 1.09), respectively. When the mean number of deaths was doubled, the correspondent changes in the odds were 27%, 49% and 71%.

**Figure 1** gives the estimated probability of a CoD having any long-term trend, defined as either a linear, a quadratic or a cubic final best model. The probability of having any sort of long-term trend (represented by the continuous line at the top of the graph) increased from 50% at a mean number of six deaths, to about 65% at 22 deaths and 75% at 60 deaths. The probability of having a linear trend (represented by the dashed line) increased as the mean number of deaths rose, reaching a peak at about 50 deaths. The subsequent decline is due to the increased probability of observing a quadratic trend (which reaches

a peak at about 500 deaths) and a cubic trend (which steadily increased). Simultaneously, this reflects a shift towards higher order trends with increasing CoD size.

When we move from the individual three-position ICD-10 codes towards aggregated lists of ICD-10 chapters, or blocks, for the study of long-term trends of causes of death, the number of CoDs with a detectable trend decreased considerably, from 409 three-position codes to 121 blocks and 12 chapters, respectively (**Table 2**).

**Table 2.** Distribution of the final best model of ICD-10 Causes of Death at different levels of aggregation.

| Final Best Model | Three-Position Codes | Blocks | Chapters |
|---|---|---|---|
| Constant (no detectable trend) | 216 | 41 | 5 |
| Any detectable trend | 409 | 121 | 12 |
| linear | 211 | 56 | 6 |
| quadratic | 126 | 42 | 4 |
| cubic | 72 | 23 | 2 |
| Total | 625 | 162 | 17 |

Trend: the composite of linear, quadratic and cubic final best models.

**Supplementary Table S2** gives detailed information for each individual CoD with a detectable trend. A great variety of trends was observed. The 211 CoDs with a linear trend showed either a monotonous increase or a monotonous decrease. Of the 72 CoDs with a cubic model, some showed an initial decrease, interrupted by a stagnation or increase, and followed by a second decrease, while other CoDs followed the opposite pattern (initial increase, interrupted by stagnation or decrease, and subsequently a new increase). Only a few CoDs showed no long-term trend, but instead had a few peak years with high mortality.

Regarding the investigation of sudden changes in causes of death, in total, 43 out of the 625 causes of death had an outlier observation year in the 20-year period at the 0.01 alpha level (not reported here). For the outlier observations at the 0.001 alpha level (**Table 3**), there were 16 CoDs, with 14 having extremely high values, and two having an extremely low value, in one single year. Of these 16 CoDs, nine were best described by the constant model, whereas a long-term trend could be identified for seven CoDs.

**Table 3.** Causes of death (ICD-10 three-position) with a detectable mortality fluctuation at the alpha 0.001 level.

| ICD-10 Code | ICD-10 Code Label | Observation Year | Deaths in Observation Year (O) | Mean Deaths in Other Years (E) | Ratio O/E | Final Best Model |
|---|---|---|---|---|---|---|
| B24 | Unspecified human immunodeficiency virus [HIV] disease | 1996 | 97.9 | 18.7 | 5.24 | quadratic |
| B94 | Sequelae of other and unspecified infectious and parasitic diseases | 2013 | 2.4 | 3.1 | 0.77 | constant |
| F32 | Depressive episode | 2001 | 54.3 | 28.3 | 1.92 | constant |
| G31 | Other degenerative diseases of nervous system, not elsewhere classified | 2015 | 408.2 | 90.1 | 4.53 | constant |
| I08 | Multiple valve diseases | 1996 | 58.6 | 19.5 | 3.01 | cubic |
| I99 | Other and unspecified disorders of circulatory system | 2002 | 75.1 | 25.7 | 2.92 | quadratic |
| J09 | Influenza due to certain identified influenza virus | 2009 | 31.7 | 1.6 | 19.81 | linear |
| J10 | Influenza due to identified influenza virus | 2014 | 9.3[a] | 9.0 | 1.03 | constant |
| K62 | Other diseases of anus and rectum | 2001 | 40.8 | 11.5 | 3.55 | constant |
| K66 | Other disorders of peritoneum | 2002 | 14.8 | 3.2 | 4.63 | constant |
| N19 | Unspecified renal failure | 1996 | 696.3 | 361.6 | 1.93 | quadratic |
| Q27 | Other congenital malformations of peripheral vascular system | 2005 | 16 | 2.6 | 6.15 | constant |
| R17 | Unspecified jaundice | 1999 | 28.9 | 12.5 | 2.31 | constant |
| V45 | Car occupant injured in collision with railway train or railway vehicle | 1999 | 29.2 | 7.4 | 3.95 | linear |
| Y36 | Operations of war | 2014 | 183.8 | 1.0 | 183.8 | constant |
| P22 | Respiratory distress of newborn | 1996 | 52.4 | 7.9 | 6.63 | cubic |

a) Cause of death J10 has substantially decreased in number of deaths in 2014, compared to 2013 and 2015 (i.e., years after the change to automatic coding), but has a similar number of deaths compared to years in the period 1996–2012.

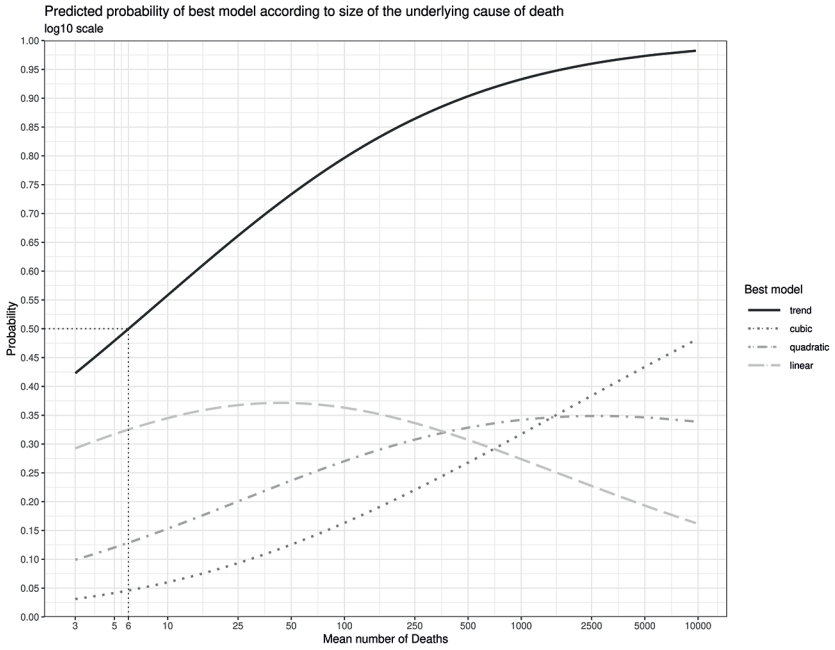


**Figure 1.** Predicted probability of detecting a long-term trend, according to size of the underlying cause of death.

# Discussion

The objective of this paper was to determine how many of the CoDs, at the ICD-10 three-position level, reveal an annual trend or a short-term fluctuation in the Netherlands, over a period of 20 years. The study outcomes could offer a criterion for deciding which ICD-10 codes should be distinguished when describing trends in a wide range of causes of death. This study could be particularly relevant in view of the forthcoming release of the ICD-11 version, which may increase the number of possible CoD codes.

A long-term trend could be identified for about two thirds of the CoDs with at least three annual deaths on average, and no long-term trend for the remaining one third. The probability of detecting a time trend increased from 50%, at a mean annual number of six deaths, to about 65% at 22 deaths, and 75% at 60 deaths. An exceptionally high or low number of deaths in one year could be demonstrated for only few CoDs.

## Evaluation of Data and Methods

The coding of causes of death at Statistics Netherlands may have affected observed trends for three different reasons: the delayed consequences of the introduction of the ICD-10, the change to automated coding, and other incidental changes.

Firstly, the introduction of the new ICD-10 classification version, in 1996, in the Netherlands, which replaced the ICD9, may have been followed by temporal re-adjustments in the coding of CoD during the first few years after 1996, such as the HIV codes (B20, B24), which were introduced for the first time in the ICD-10. Other examples are unspecified renal failure (N19), and respiratory distress of newborn (P22). Our finding that multiple valve diseases (I08) had an extremely high number of deaths in 1996 was already noted in a study aimed at detecting the effects of changes in data production during the period 1970–2006 [12].

Secondly, in our study we accounted, to a large extent, for the introduction of the automated CoD coding and related changes in 2013. Switching from manual to automated coding can result in significant changes in cause-specific mortality rates [13-15]. In further analyses, we found that our results would be substantially affected if we omitted the 2013 step parameter and, thus, ignored the switch to automated coding. This would have decreased the proportion of CODs without a detectable long-term trend (from 34.5% to 26.6%), or with a linear model (from 33.8% to 27.8%), and increased the proportion of CoDs with a quadratic model (from 20.2% to 23.5%), and cubic model (from 11.5% to 22.1%). While the inclusion of the step parameter was intended to prevent from identifying spurious time trends, it implements a conservative approach, that may come at the price of masking some real trends.

Thirdly, several incidental changes in the coding of CoD may underlie some of the large short-term fluctuations that we observed. Other degenerative diseases of the nervous system, not elsewhere classified (G31), showed a 6.5-fold increase in 2015 as compared to 2013, most likely because of the implementation of a WHO ICD-10 update [9]. Changes in the coding process in Statistics Netherlands occurred in the period 1999–2002, when less resources for quality control were available. This may have contributed to changes in the coding of conditions, such as depressive episode (F32), as well as codes described as “other” or “unspecified”, such as other diseases of anus and rectum (K62), other and unspecified disorders of circulatory system (I99) and unspecified jaundice (R17).

There are some limitations in the methodology used in our study. Firstly, we focused on codes at the ICD-10 three-position level and did not examine the more detailed codes at the ICD-10 four-position level. Although this could increase the number of the CoDs with a detectable long-term trend, the sporadic occurrence of most four-position codes would add the potential for redundant analyses. Secondly, we did not investigate changes within individual years. For some CoDs, such as contagious diseases with epidemic outbreaks, it may be more important to monitor day-to-day or week-to-week changes, rather than year-to-year trends. Thirdly, we included CoDs with three or four deaths annually, even though it is uncertain whether deaths were normally distributed, as we assumed by applying normal regression. In a sensitivity analysis, we restricted the regression analysis to CoDs with five or more deaths annually. We found that the probability of detecting a long-term trend was 50% at five mean annual deaths, 65% at 20 deaths and 75% at 60 deaths – results that are very close to our main findings.

## Interpretation of Trends

While we described long-term trends in mortality, we did not aim to explain the trends that were observed. In general, understanding long-term changes in mortality requires consideration of the incidence and case-fatality of a disease, the latter being largely determined by changes in medical practices and technologies, and in the provision of and access to health care [16, 17]. Generally, relevant changes occur slowly and with delayed effects. Without any additional external information, such as personal data obtained from record linkages, it is difficult to attribute these long-term changes to specific factors, such as changes in treatments or risk factor prevalence [18]. One factor, the ageing of the Dutch population, was controlled by using the age-standardized death counts.

Some of the observed short-term fluctuations may reflect real changes in the occurrence of specific diseases or injuries in the Netherlands. One example is influenza due to an identified influenza virus (J10) that had an extremely low observation in 2014, which was not attributable to an exchange with J11 (influenza, virus not identified). Real changes in external causes of death are exemplified by operations of war (Y36), which increased in 2014, due to the MH17 flight crash.

Further research might show whether, and how, our results are cross-nationally generalizable. Some results are not directly generalizable to other countries, as the Netherlands has a medium-sized population (mean population during the study period was 16.3 million persons), and the proportion of CoDs with a detectable trend is likely to

be higher (lower) in countries with a substantially larger (smaller) population. However, the fundamental relationship between CoD size and the likelihood of detecting a long-term trend might be applicable to other countries, and could be the focus of international comparative studies.

## Conclusions



This study demonstrates, for the Dutch population, that a long-term mortality trend can be identified for at least 409 CoDs, with a curved instead of linear trend in many cases. The size of a CoD is an important predictor of the probability of detecting a long-term trend. Year-to-year mortality fluctuations could be demonstrated in fewer CoDs and may result from problems with the coding of CoDs.

In the context of the ICD11 release, with a large increase in the number of possible CoD codes, and a decrease in the average number of deaths per CoD, this study provides new evidence to guide the level of CoD coding and reporting that should reasonably be recommended. Such recommendations are particularly important to national statistical offices, as they have to decide what level of detail to include when they publish annual cause-specific mortality data. Both producers of these statistics and users of the data benefit from realizing that, despite the large number of ICD codes that could be distinguished, the number of CoDs for which a trend can be detected is limited. At the same time, it is important to realize that there is a reasonable likelihood of detecting a long-term trend, even for CoDs with only three annual deaths. Our results therefore suggest the selection of a broad range of causes of death, preferably defined in terms of a minimum annual number of deaths, rather than condensed lists of CoDs, in order to not overlook CoDs with a significant long-term trend.

# References


1. World Health Organization-WHO. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; World Health Organization: Geneva, Switzerland, 1992; Volume I–III.
2. GBD 2015. Mortality and Causes of Death Collaborators Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: A systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388(10053):1459-544.
3. Meslé F and Vallin J. Chapter 10. General Trends in Mortality by Cause. In Mortality and Causes of Death in 20th-Century Ukraine. Demographic Research Monographs; Springer Science+Business Media B.V: Dordrecht, The Netherlands, 2012; pp. 153-72.
4. World Health Organization-WHO. Available online: https://icd.who.int/browse11/l-m/en
5. Statistics Netherlands, Netherlands. Available online: www.cbs.nl/en-gb/figures
6. Shacham M and Brauner N. Minimizing the Effects of Collinearity in Polynomial Regression. Ind Eng Chem Res. 1997;36:4405–12
7. Wong CS and Li WK. A note on the corrected Akaike information criterion for threshold autoregressive models. J Time Ser Anal. 1998;19:113-24.
8. Johansson LA and Pavillon G. IRIS: A language-independent coding system based on the NCHS system MMDS. In Proceedings of the WHO-FIC Network Meeting, Tokyo, Japan, 16-22 October 2005.
9. World Health Organization-WHO. Cumulative Official Updates to ICD-10, 1996-2015; WHO Collaborating Centres for the Family of International Classifications: Geneva, Switzerland, 2016; Volume 1. Available online: http://www.who.int/classifications/icd/icd10updates/en/
10. Fox J and Weisberg S. An {R} Companion to Applied Regression, 2nd ed.; Sage: Thousand Oaks, CA, USA, 2011. Available online: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion
11. R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2019. Available online: https://www.R-project.org
12. Rey G, Aouba A, Pavillon G, et al. Cause-specific mortality time series analysis: A general method to detect and correct for abrupt data production changes. Popul Health Metr. 2011;9:52.
13. World Health Organization-WHO. Strengthening Civil Registration and Vital Statistics for Births, Deaths and Causes of Death. Resource Kit; World Health Organization: Geneva, Switzerland, 2012.
14. Denissov G. Sharp decrease in observed cerebrovascular mortality may be due to certification and coding. Scand J Public Health. 2016;44:335-7.
15. Floristán YF, Delfrade JO, Carrillo JP, et al. Coding Causes of Death with IRIS Software. Impact in Navarre Mortality Statistic. Rev Esp Salud Publica. 2016;2:e1-9.
16. Deeg DJH, van Vliet MJG, Kardaun JW, Huisman M. Understanding the mortality decline at older ages: improved life course or improved present period? Annu Rev Gerontol Geriatr. 2013;33(1):259-91.

17. Peters F, Nusselder WJ, Mackenbach JP. A closer look at the role of healthcare in the recent mortality decline in the Netherlands: results of a record linkage study. J Epidemiol Community Health. 2015;69:536-42.

18. Koopman C, Vaartjes I, van Dis I, et al. Explaining the Decline in Coronary Heart Disease Mortality in the Netherlands between 1997 and 2007. PLoS One. 2016;11:e0166139.

# Supplementary Material

**Supplementary Table S1.** Age-standardized count of deaths for the three position ICD-10 codes reported in the Netherlands, for the period 1996–2015.

Available only online due to limited space: https://www.mdpi.com/1660-4601/16/21/4150

**Supplementary Table S2.** Detailed long-term trend patterns for the 625 three position ICD-10 codes with at least three average annual deaths as individual causes, or aggregated per block and chapter.

Available only online due to limited space: https://www.mdpi.com/1660-4601/16/21/4150

# Chapter 3

## Deriving a cut-off point of the size of Cause of Death for mortality trend analysis in 21 European countries

**Marianna Mitratza**, Jan W.P.F. Kardaun, Anton E. Kunst

*Published in*
*BMJ Open 2020;10:e031702.*

# Abstract

**Background:** The International Classification of Diseases (ICD-10) distinguishes a large number of causes of death (CoDs) that could each be studied individually when monitoring time-trends. We aimed to develop recommendations for using the size of CoDs as a criterion for their inclusion in long-term trend analysis.

**Methods:** We performed a retrospective trend analysis in 21 European countries of the WHO Mortality Database. Deaths from causes of death (3-position ICD-10 codes) with ≥5 average annual deaths in a 15-year period between 2000 and 2016 were used. Fitting polynomial regression models, we examined for each CoD in each country whether or not changes over time were statistically significant (with $\alpha = 0.05$) and we assessed correlates of this outcome. Applying receiver operating characteristic (ROC) curve diagnostics, we derived CoD size thresholds for selecting CoDs for trends analysis.

**Results:** Across all countries, 64.0% of CoDs had significant long-term trends. The odds of having a significant trend increased by 18% for every 10% increase of CoD size. The independent effect of country was negligible. As compared to circulatory system diseases, the probability of a significant trend was lower for neoplasms and digestive system diseases, and higher for infectious diseases, mental diseases and signs-and-symptoms. We derived a general threshold of around 30 (range: 28–33) annual deaths for inclusion of a CoD in trend analysis. The relevant threshold for neoplasms was around 65 (range: 61–70) and for infectious diseases was 20 (range: 19–20).

**Conclusions:** The likelihood that long-term trends are detected with statistical significance is strongly related to CoD size and varies between ICD-10 chapters, but has no independent relation to country. We recommend a general size criterion of 30 annual deaths to select CoDs for long-term mortality-trends analysis in European countries.

# Introduction

Mortality data are essential for the monitoring of population-wide trends in a large number of diseases and injuries, as well as for the evaluation of health policies. A common source for these data is the statistics maintained by national statistical offices [1, 2]. National statistics of causes of death (CoDs) include many codes of the10th revision of the International Classification of Diseases (ICD-10 codes) [3]. Given the detail of this classification – there are 1,752 3-position ICD-10 codes – a part of it may not be instrumental for monitoring long-term time-trends due to the small number of deaths for specific codes.

When using these statistics to monitor long-term trends in mortality, a main question is which of the many possible CoDs to include. At the very least, the selection should include only CoDs that are large enough to have a reasonable probability of detecting a long-term mortality trend. This probability may be influenced by several factors. One main factor is the CoD size, defined as the mean annual number of deaths, which expresses the rarity of a disease or condition that is selected as underlying cause of death in a population. Incidence changes or effects of interventions are common factors discussed in mortality trends analyses [4, 5]. In addition, this probability might depend on other factors, such as the type of CoD, or the country of interest. Certain types of CoDs may be more likely to present a long-term trend. For example, neoplasms have been shown to be more gradual in their annual changes [6], whereas infectious diseases [7] may have high year-to-year variation. As regards to different populations, the likelihood to detect a long-term trend for a CoD may vary between countries because of differences in population size, CoD coding practices that may also influence observed mortality trends [8], trends in prevalence of risk factors [9, 10, 11, 12], implementation of new prevention strategies [13, 14], treatment protocols [5] or healthcare reforms [15].

Due to the fact that the likelihood to detect a long-term trend of a CoD may depend on various factors, there is a need for an empirical assessment of such likelihood. Such analysis may provide an empirical basis for the identification of CoDs for which long-trends are likely to be detectable. More specifically, it may be used to define a criterion, or rule of thumb, that identifies eligible CoDs in terms of a minimum CoD size. When such a criterion allows for variation by CoD type and country, it may be used in national and international trend analysis across a broad range of CoDs.

The general objective of this study was to determine a CoD size criterion for the study of long-term mortality trends in European countries. The specific objectives were: (1) to assess the association between the size and the type of a CoD and the probability of detecting a

long-term trend in European countries, (2) to assess how this association varies according to country, and (3) to identify a minimum annual number of deaths recommended to monitor trends in cause-specific mortality.

# Methods

## Data

We used annual mortality data for 21 European countries of the WHO Mortality Database (1 October 2017 update) [16]. We included the 21 countries of the European Union (28 countries) or the European Free Trade Association (4 countries) that had been using ICD-10 (3- or 4-position) coding for at least 15 consecutive years. Iceland, Luxemburg and Malta were excluded because of their small population [17]. The most recent 15-year period was selected, which was 2001–2015 for all countries with few exceptions (Belgium, France and Switzerland: 2000–2014; Austria: 2002–2016). If the time series of a CoD in a country was interrupted by a year without any data on that CoD, we assumed that zero cases occurred.

## Statistical analysis

For each year and CoD in a country, we calculated an age-standardized count of deaths using the direct method. As reference population, we used the age-distribution of the European Standard Population 2013, scaled to the mid-period population of each country. This method intended to compensate for annual changes in the age-distribution of the population, while keeping the age-standardized count close to the observed absolute numbers.

For further analysis, we analyzed CoDs that had at least 5 average age-standardized annual deaths, because most of the smaller CoDs had predominantly zero or only zero annual deaths.

Long-term time-trends of the age-standardized count of deaths of each CoD in each country were analyzed using ordinary least squares regression models. Trends were fitted by applying linear regression models with polynomial terms of year as continuous, independent covariates [18]. We used orthogonal polynomials in order to account for multicollinearity of the polynomial components [19]. We fitted four models: the constant, the linear, the quadratic and the cubic model (with zero, first, second and third- degree

polynomials, respectively). The four models were applied for all CoDs in each country. We used the lowest corrected Akaike Information Criterion (AICc) to select the best model for each CoD in each country [20]. In a next step, the best model was compared with the constant model using the F-test, at the significance level of $\alpha = 0.05$. If the best model performed better than the constant model with statistical significance, it was kept as the final best model. Otherwise, the constant model was selected as the best model for this CoD. In the rest of the paper, the constant model is referred to as the absence of a demonstrable trend.

Next, using a multilevel logistic regression model, we determined how the categorization of a CoD as having a statistically significant trend (i.e. best model being the linear, quadratic or cubic model) was related to CoD size and CoD type. These variables were included in the model as fixed effects. The CoD size was defined as the mean annual number of deaths and the CoD type was defined as the ICD-10 chapter in which it is classified. The chapter of circulatory diseases was the reference category, as it had the largest number of deaths. As the distribution of the number of deaths across CoDs was highly skewed, we used its natural logarithm as a measure of CoD size. The model also included the level of countries as random effect, in order to investigate the variation of European countries in the likelihood of detecting a long-term trend. We calculated the Intra-class Correlation Coefficient (ICC), which expresses the proportion of the variance in the outcome that is attributable to variations between the countries [21]. The ICC was calculated both with and without controlling for the fixed effects of the size and type of the CoD.

Finally, we used receiver operating characteristic (ROC) curve diagnostics [22, 23, 24] to derive CoD size thresholds for detecting a long-term time-trend. We calculated the Area Under the Curve (AUC) of the logistic model with CoD size as the predictor and the binary categorization of a CoD as having a significant long-term time-trend as the outcome. We derived the CoD size thresholds using three indices. Firstly, we used the maximum Youden index [25, 26, 27], which represents the point of the ROC curve with the maximum sum of sensitivity (se) and specificity (sp). Secondly, we used the index measuring the minimum difference between sensitivity and specificity [23]. Thirdly, we estimated the index that represents the point closest to the top-left part of the ROC curve [22, 26].

All analyses were conducted using R statistical software (3.5.1 version) [28].

# Results

The number of CoDs with at least 5 annual deaths on average varied between 202 (Estonia) and 791 (Germany) (**Table 1**). Of these CoDs, 32.6%, 20.2% and 11.2% had a significant trend following a linear, quadratic or cubic model respectively. The percentage of CoDs with no significant trend (i.e. constant model) varied from 27.5% to 43.9%, and was highest in the Nordic countries, Switzerland and Slovenia. More detailed information on the best model for each CoD in each European country can be found in an additional file (**Supplementary Table S1**).

Both CoD size and CoD type were significantly associated with the likelihood of having a significant long-term trend (p-value <0.001) (**Table 2**). For every 10% increase in the CoD size, we observed a 18% increase (1.1^1.73 = 0.18) in the odds of having a significant trend (OR = 1.73, 95%CI = 1.67 ; 1.79). Regarding the CoD type, neoplasms and digestive system diseases had lower probability for detecting a trend in comparison to the circulatory system diseases. On the other hand, this probability was higher for infectious diseases, mental diseases and signs-and-symptoms. **Figure 1** shows for each CoD chapter in each country the estimated probability of having a significant long-term trend in relation to CoD size. The variation between CoD chapters was substantial, irrespective of CoD size. Neoplasms (chapter C00.D48) as a group of CoDs showed the lowest probability of having a detectable trend.

We found only small variation of countries in the likelihood of detecting a long-term trend, as the ICC for the country-level random effect was only 0.013 (without fixed effects for chapter and size) and 0.003 (with fixed effects) (**Table 2**). **Figure 2** illustrates the small differences between countries in the estimated probability of having a long-term trend.

**Table 1.** Frequencies of causes of deaths according to estimates of their long-term time-trend in 21 European countries.

| Country (sorted by mean population) | CoDs* analyzed (n) | Type of long-term time-trend (%) | | | | Mean population (thousands) | Mean crude annual deaths from all CoDs together | Mean crude annual deaths from CoDs analyzed* |
|---|---|---|---|---|---|---|---|---|
| | | No significant trend | Linear | Quadratic | Cubic | | | |
| Estonia | 202 | 34.7 | 35.6 | 15.3 | 14.4 | 1,343 | 16,642 | 16,006 |
| Slovenia | 228 | 43.9 | 29.4 | 17.1 | 9.6 | 2,029 | 18,830 | 18,163 |
| Latvia | 259 | 38.2 | 28.6 | 17.0 | 16.2 | 2,157 | 30,840 | 30,091 |
| Lithuania | 308 | 37.7 | 32.5 | 19.5 | 10.4 | 3,187 | 42,101 | 41,243 |
| Croatia | 315 | 34.6 | 37.1 | 15.9 | 12.4 | 4,381 | 51,444 | 50,844 |
| Norway | 356 | 42.4 | 26.4 | 23.6 | 7.6 | 4,804 | 41,764 | 40,886 |
| Finland | 355 | 42.8 | 35.2 | 16.6 | 5.4 | 5,302 | 49,869 | 48,933 |
| Denmark | 420 | 40.0 | 37.4 | 13.6 | 9.0 | 5,497 | 54,446 | 53,319 |
| Switzerland | 451 | 43.9 | 31.0 | 16.4 | 8.6 | 7,558 | 62,183 | 61,384 |
| Austria | 374 | 35.8 | 25.4 | 23.5 | 15.2 | 8,359 | 77,256 | 75,765 |
| Sweden | 457 | 43.3 | 30.6 | 19.5 | 6.6 | 9,259 | 91,504 | 90,516 |
| Hungary | 482 | 36.1 | 27.6 | 29.3 | 7.1 | 10,024 | 130,830 | 129,695 |
| Czech Republic | 443 | 31.8 | 35.9 | 23.0 | 9.3 | 10,383 | 107,448 | 106,636 |
| Belgium | 517 | 37.3 | 33.7 | 20.9 | 8.1 | 10,678 | 104,344 | 103,257 |
| Netherlands | 554 | 33.9 | 28.0 | 25.1 | 13.0 | 16,489 | 138,373 | 137,645 |
| Romania | 444 | 31.8 | 36.0 | 19.4 | 12.8 | 21,594 | 258,000 | 257,265 |

| Country (sorted by mean population) | CoDs* analyzed (n) | Type of long-term time-trend (%) | | | | Mean population (thousands) | Mean crude annual deaths from all CoDs together | Mean crude annual deaths from CoDs analyzed* |
|---|---|---|---|---|---|---|---|---|
| | | No significant trend | Linear | Quadratic | Cubic | | | |
| **Poland** | 581 | 27.7 | 33.2 | 22.9 | 16.2 | 38,114 | 375,231 | 374,295 |
| **Spain** | 672 | 33.0 | 33.9 | 18.2 | 14.9 | 44,594 | 385,512 | 383,204 |
| **United Kingdom** | 710 | 30.8 | 33.8 | 21.1 | 14.2 | 61,690 | 580,733 | 578,572 |
| **France** | 738 | 27.5 | 38.8 | 22.6 | 11.1 | 61,709 | 535,320 | 532,293 |
| **Germany** | 791 | 27.8 | 35.4 | 24.7 | 12.1 | 81,980 | 852,566 | 849,400 |

* coded at ICD-10, 3-position level; including all causes of death with at least 5 mean number of deaths in the 15-year-period.

CoD, cause of death; ICD, International Classification of Diseases.

**Table 2.** Relationship between the likelihood for a cause of death (CoD) to have a significant long-term trend with its size, corresponding ICD-10 chapter, and country.

| COD characteristic | | Number of CODs | Total number of deaths | Odds Ratio*** (95% Confidence Interval) |
|---|---|---|---|---|
| **Size** | | | | |
| log(mean deaths*) | | | — | **1.73** (1.67 ; 1.79) |
| **ICD10 Chapter** | | | | |
| I00.I99 | Diseases of the circulatory system | 988 | 23,610,116 | reference |
| A00.B99 | Certain infectious and parasitic diseases | 454 | 823,839 | **1.63** (1.25 ; 2.14) |
| C00.D48 | Neoplasms | 1871 | 15,632,543 | **0.57** (0.47 ; 0.69) |
| D50.D89 | Diseases of blood and blood-forming organs and certain disorders involving the immune mechanisms | 217 | 143,156 | 0.82 (0.59 ; 1.14) |
| E00.E90 | Endocrine, nutritional and metabolic diseases | 365 | 1,547,431 | 1.01 (0.76 ; 1.34) |
| F00.F99 | Mental, behavioural disorders | 230 | 1,725,881 | **1.62** (1.13 ; 2.30) |
| G00.G99 | Diseases of the nervous system | 536 | 1,888,629 | 0.90 (0.70 ; 1.15) |
| J00.J99 | Diseases of the respiratory system | 553 | 4,740,481 | 1.22 (0.94 ; 1.58) |
| K00.K93 | Diseases of the digestive system | 767 | 2,816,168 | **0.75** (0.59 ; 0.94) |
| M00.M99 | Diseases of the musculoskeletal system and connective tissue | 367 | 283,108 | 1.14 (0.86 ; 1.50) |
| N00.N99 | Diseases of the genitourinary system | 385 | 1,029,118 | 0.98 (0.74 ; 1.29) |
| Q00.Q99 | Congenital malformations, deformations and chromosomal abnormalities | 325 | 141,848 | 0.89 (0.67 ; 1.19) |
| R00.R99 | Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (signs-and-symptoms) | 258 | 2,062,587 | **1.68** (1.18 ; 2.38) |
| V01.Y98 | External causes of morbidity and mortality | 1992 | 3,005,539 | 1.20 (0.99 ; 1.46) |
| Other** | | 349 | 240,748 | 1.29 (0.97 ; 1.71) |
| **Intra-class correlation for the country level** | | | | |
| model with fixed effects for size and ICD10 chapter | | 0.003 | | |
| model with no fixed effects | | 0.013 | | |

* mean deaths: mean of the annual number of deaths for a cause of death monitored in the 15-year-period, measured per country. Only including CoDs with 5 or more deaths.
** "Other" consists of the causes of death classified in the ICD-10 chapters H00.H59: Diseases of the eye and adnexa, H60.H95: Diseases of the ear and mastoid process, L00.L99: Diseases of the skin and subcutaneous tissue, O00.O99: Pregnancy, childbirth and the puerperium and P00.P96: Certain conditions originating in the perinatal period.
*** odds ratios in bold were statistically significant with p-value <0.05.

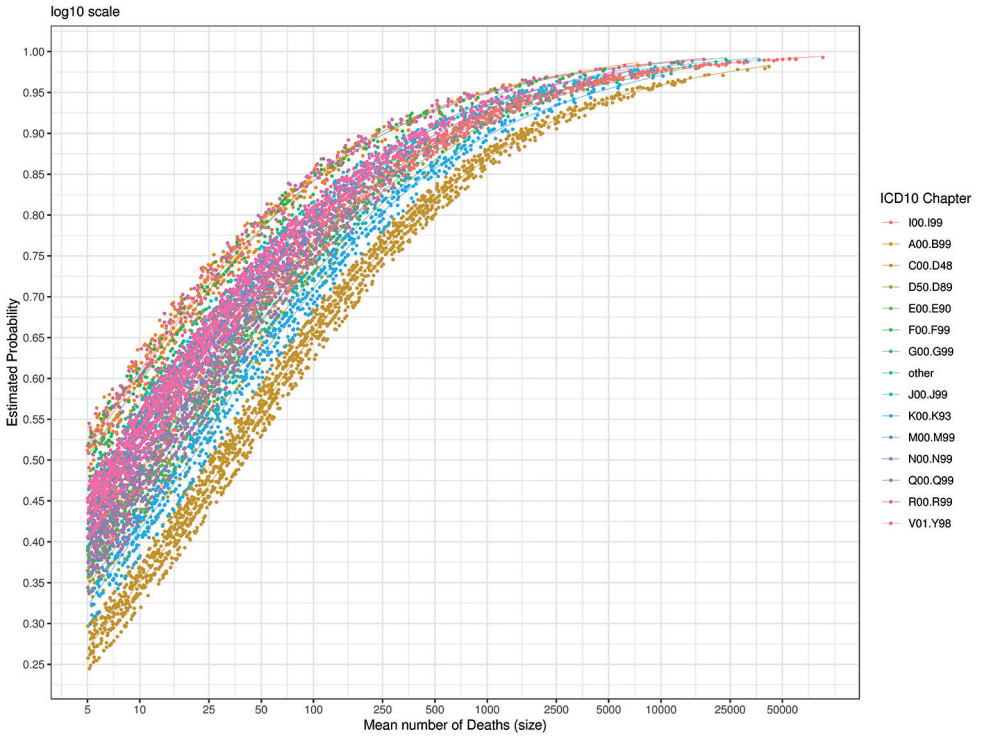


**Figure 1.** Estimated probability for an underlying cause of death to have a significant long-term trend according to its size, by ICD-10 chapter. See **Table 2** for the definition of the chapters. ICD, International Classification of Diseases.

**Figure 3a** describes the sensitivity (se) and specificity (sp) for detecting a significant long-term trend using different levels of thresholds in terms of any CoD size. The AUC corresponding to these se and sp values was 0.706, with 95%CI: 0.695 ; 0.716. The maximized sum index (Youden Index) was 32.7 annual deaths, with sensitivity (se) 61.4% and specificity (sp) 70.3%. The minimum difference index was 27.5 annual deaths (se = 65.5%, sp = 65.5%). The closest top-left index was 29.4 annual deaths (se = 64.0%, sp = 67.5%) (**Figure 3a**).

The corresponding analysis for the neoplasms yielded a similar ROC curve (AUC = 0.703, 95%CI: 0.680 ; 0.727) (**Figure 3b**). The Youden Index was 70.4, with sensitivity 61.5% and specificity 69.6%, and the minimum difference index was 60.5 annual deaths (se = 64.8%, sp = 64.8%). The closest top-left index was also 60.5 annual deaths (se = 64.1%, sp = 66.9%).

Infectious and parasitic diseases (AUC = 0.706, 95%CI: 0.695 ; 0.716) yielded a Youden Index of 19.7 annual deaths, with sensitivity 67.1% and specificity 69.8%. The closest top-left index was identical, while the minimum difference threshold was 19.3 annual deaths (se = 67.4%, sp = 67.4%) (**Figure 3c**).

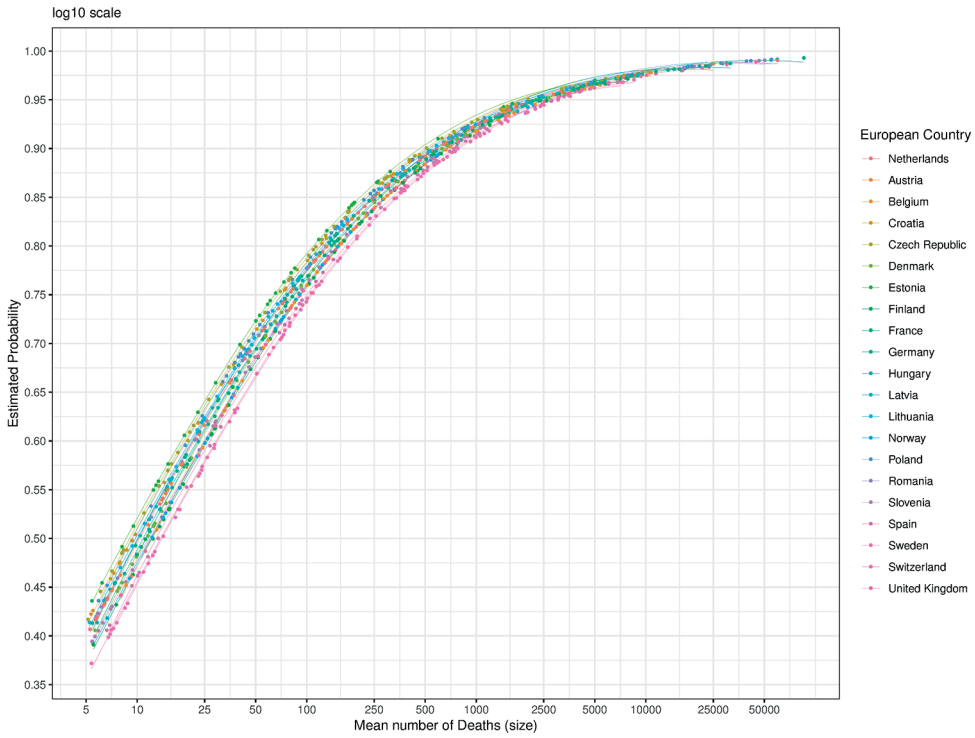


**Figure 2.** Estimated probability for a disease of the circulatory system to have a significant long-term mortality trend according to its size, by European country.

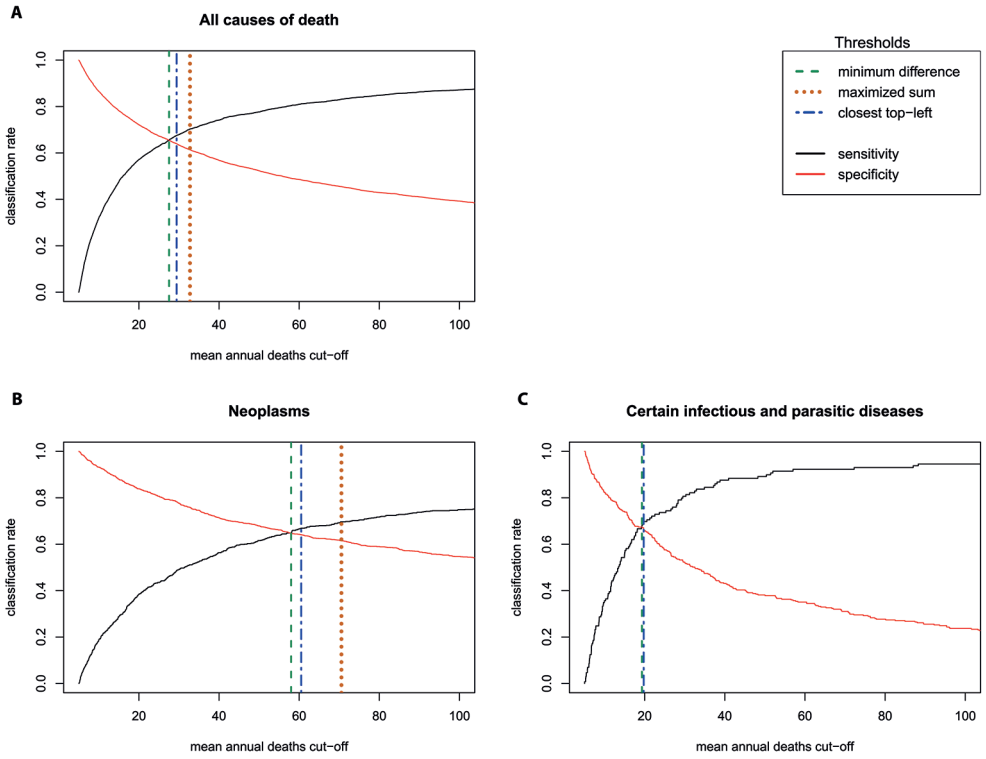


**Figure 3.** Sensitivity and specificity of the cause of death size for the detection of significant long-term time-trends, with thresholds for the optimal cause of death size for trend analysis.

# Discussion

CoD data are used in widely varying settings, ranging from detailed mortality profiles to macro estimates. Applications include studies in localized areas [29], single countries [30] or worldwide [2, 31]; for a single-disease [32, 33] or disease group [9]; monitored for days or a long-term period [7]; for specific age groups [33, 34] or specific situations (e.g. maternal mortality [5], external causes [35, 36, 37]). These settings all impose different requirements on the collected data. Here we focused on one particular application: national estimates of mortality time trends for a reasonably long period (15 years), for a considerable number of countries (21) that have quite comparable CoDs data collection and registration systems [38], covering as many CoDs as possible. Our aim was to investigate the effect of the size of a CoD on the probability to detect a significant trend, and how this is related to country and type of CoD (ICD-10 chapter).

Our results indicate that both the size and the type of a cause of death were associated with the probability of detecting a significant trend, while variations among European countries were negligible. Some types of CoDs, particularly neoplasms and digestive system diseases, had a lower probability for detecting a significant trend in comparison to the circulatory system diseases, whereas infectious and mental diseases had a higher probability. The results suggest a general size criterion of 30 annual deaths for selecting causes of death to include in long-term mortality trends analysis, and a more specific criterion of 65 deaths for neoplasms and 20 for infectious diseases.

We should outline the limitations of our study. Firstly, due to the exclusion of causes of death with less than 5 annual deaths on average, smaller countries were represented in our analysis with fewer causes of death. However, this is unlikely to have a strong influence on the results, as the suggested CoD size threshold of about 30 deaths is much higher than the lower limit of 5 mean annual deaths. Secondly, although we proposed the CoD size as a criterion to select CoDs for long-term trend analysis, we acknowledge that other criteria could be used, such as greater preference to CoDs that involve high healthcare costs or that are potentially modifiable by preventive or curative actions. Thirdly, the likelihood to demonstrate a time trend with statistical significance depends on the statistical method that is used to describe these trends. Our results are dependent on the balance between avoiding Type I error and Type II errors. As for Type I errors, we chose a significance level of $\alpha = 0.05$. A more restrictive significance level would have the consequence to increase Type II errors, i.e. to reduce the proportion of CoDs for which a trend would be detected based on our method.

Moreover, our results should be seen as conditional on our use of ordinary least squares regression (OLS) models with polynomial terms. The OLS approach may not be appropriate for small counts. However, the approximation of a Poisson by a normal error distribution is generally assumed to be adequate if the mean number of observations is about five or more. For larger counts, OLS has the benefit that a variance can be estimated, rather than postulated.

In addition, an alternative to the classic polynomial regression approach would have been to use Generalized Additive Models (GAMs). These models have the advantage of being able to pick up trends that are not polynomial. In a sensitivity analysis, we applied GAMs with Gaussian process smoothing function to our data. We found that a long-term trend could be detected in 71.7 percent of the CoDs, as compared to 64.0 percent in our original analysis. There were virtually no CoDs for which a trend could be detected when

using polynomial models but not when using GAMs. This would imply that our results are approximately robust to the method used, although somewhat conservative.

Finally, including spatial correlation in our model may have altered the chance of detecting a significant trend for CoDs with marked geographical patterns. We calculated Moran's I test for spatial correlation among countries regarding the proportion of CoDs in each country with a detected long-term trend. The Moran's I test was found to be not statistically significant for all CoDs collectively (p-value = 0.988). At the level of CoD chapters, we found significant spatial correlation for the chapters C-D (p-value = 0.002), E (p-value = 0.025), and V-Y (p-value = 0.001), but not for other chapters.

We found that mortality from neoplasms was less likely to have a significant trend, for a given size of CoD. This may relate to the fact that the neoplasm mortality levels tend to change gradually over time, without short-term trend changes [6]. Additionally, cancers are usually coded reliably and consistently over time [39, 40, 41], so that coding artefacts can rarely induce artificial changes. Conversely, the dynamic nature of infectious diseases may be responsible for their higher likelihood to change over time, and to have significant trends even with relatively small numbers of deaths. Similarly, the chapter of signs-and-symptoms is sensitive to changes in the coding rules and practices, thus creating significant changes even with small number of deaths.

Our study showed that European countries did not vary substantially in the probability of detecting a significant long-term trend in CoDs of the same size and type. This finding is surprising given the heterogeneity of the countries in terms of demographic characteristics, disease epidemiology, healthcare systems and coding practices. We found that differences between countries in the proportions of CoDs with a significant trend (shown in **Table 1**) can be related to differences in CoD size, which is strongly related to the differences in population size. Consequently, our analysis provides support for establishing one common CoD size threshold, applicable for all European countries and for use in international trend analyses.

Currently there is no gold standard for the selection of CoDs to analyze for long-term trends. In this study we attempted to set such a standard, based on the criterion of the CoD size, which is easy to measure for each single CoD. We calculated thresholds with three common methods, which came close enough (e.g. in the range of 28 to 33 deaths) to support one general recommendation for practical use. Of course, different thresholds may be preferred, depending on the user's preference to avoid either false positives (by selecting a higher threshold) or false negatives (lower threshold).

In our data, the number of CoDs that surpassed our recommended threshold of 30 annual deaths on average was around 500 for the biggest countries, 200–250 for the middle-sized countries and around 100 for the smaller European countries (results not shown). In total, 52 CoDs had over 30 annual deaths on average in each country included in our analysis. This implies that at least 52 CoDs could be included in the international comparison of long-term trends, but up to 100 if one is to accept a greater risk of false positives in smaller countries.

From the public health practitioner's perspective, the findings of our study can be used in order to set realistic expectations about the number of CoDs that are likely to have a significant long-term trend in populations. We recommend a size criterion of 30 annual deaths to be considered when planning for national or international monitoring and comparisons of cause-specific mortality.

# References


1. Mathers CD, Fat DM, Inoue M, et al. Counting the dead and what they died from: an assessment of the global status of cause of death data. Bull World Health Organ. 2005;83(3):171-7.
2. Naghavi M, Abajobir AA, Abbafati C, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1151-210.
3. World Health Organization. International statistical classification of diseases and related health problems. Tenth revision, vol. 2. Geneva, Switzerland: WHO; 1993.
4. Aregawi M, Malm KL, Wahjib M, et al. Effect of anti-malarial interventions on trends of malaria cases, hospital admissions and deaths, 2005-2015, Ghana. Malar J. 2017;16(1):177.
5. Mnyani CN, Buchmann EJ, Chersich MF, et al. Trends in maternal deaths in HIV-infected women, on a background of changing HIV management guidelines in South Africa: 1997 to 2015. J Int AIDS Soc. 2017;20(3):e25022.
6. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends-an update. Cancer Epidemiol Biomarkers Prev. 2016;25(1):16-27.
7. Armstrong GL, Conn, LA, Pinner RW. Trends in infectious disease mortality in the United States during the 20th century. JAMA. 1999;281(1):61-6.
8. Griffiths C, Brock A, Rooney C. The impact of introducing ICD-10 on trends in mortality from circulatory diseases in England and Wales. Health statistics quarterly / Office for National Statistics. 2004(22):14-20.
9. Choe YJ, Choe SA, Cho SI. Trends in infectious disease mortality, South Korea, 1983–2015. Emerg Infect Dis. 2018;24(2):320-7.
10. Van Dyke M, Greer S, Odom E, et al. Heart disease death rates among blacks and whites Aged ≥ 35 Years – United States, 1968-2015. MMWR Surveill Summ. 2018;67(5):1-11.
11. Hertel-Fernandez AW, Giusti AE, Sotelo JM. The Chilean infant mortality decline: improvement for whom? Socioeconomic and geographic inequalities in infant mortality, 1990–2005. Bull World Health Organ. 2007;85(10):798-804.
12. Mesalles-Naranjo O, Grant I, Wyper GMA, et al. Trends and inequalities in the burden of mortality in Scotland 2000-2015. PLoS One. 2018;13(8):e0196906.
13. Medina-Gomez OS and Medina-Reyes IS. Mortality from type 2 diabetes and implementation of the PREVENIMSS program: a time series study in Mexico, 1998–2015. Cad Saude Publica. 2018;34(5):e00103117.
14. Allanson ER and Pattinson RC. Quality-of-care audits and perinatal mortality in South Africa. Bull World Health Organ. 2015;93(6):424-8.
15. Notrica DM, Sayrs LW, Krishna N. The effect of verified pediatric trauma centers, state laws, and crash characteristics on time trends in adolescent motor vehicle fatalities, 1999–2015. J Trauma Acute Care Surg. 2018;85(5):944-52.
16. WHO Mortality Database. Geneva: World Health Organization; 2017. Available from: http://www.who.int/healthinfo/mortality_data/en/.

17. Eurostat – European Union. Chapter 7: Demographic challenges-population projections. In: People in the EU: Who are we and how do we live? Eurostat Statistical Books. Luxemburg: Publications Office of the European Union; 2015. p. 159.

18. Mitratza M, Kunst AE, Kardaun JWPF. Detecting Mortality Trends in the Netherlands Across 625 Causes of Death. Int J Environ Res Public Health. 2019;16(21):4150. doi: 10.3390/ijerph16214150.

19. Shacham M and Brauner N. Minimizing the effects of collinearity in polynomial regression. Ind Eng Chem Res. 1997;36:4405-12.

20. Wong CS and Li WK. A note on the corrected Akaike information criterion for threshold autoregressive models. J Time Ser Anal. 1998;19(1):113-24.

21. Diez Roux AV. A glossary for multilevel analysis. J Epidemiol Community Health. 2002;56:588-94.

22. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8(4):283-98.

23. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000;45(1-2):23–41.

24. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 2007;115(5):654–7.

25. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–5.

26. Perkins NJ and Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol. 2006;163(7):670–5.

27. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J. 2005;47(4):458–72.

28. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: https://www.R-project.org.

29. Bradshaw D, Groenewald P, Bourne DE, et al. Making COD statistics useful for public health at local level in the city of Cape Town. Bull World Health Organ. 2006;84(3):211-7.

30. Franca EB, Passos VMA, Malta DC, et al. Cause-specific mortality for 249 causes in Brazil and states during 1990–2015: a systematic analysis for the global burden of disease study 2015. Popul Health Metr. 2017;15(1):39.

31. Ebmeier S, Thayabaran D, Braithwaite I, et al. Trends in international asthma mortality: analysis of data from the WHO Mortality Database from 46 countries (1993–2012). Lancet. 2017;390(10098):935-45.

32. Burkill S, Montgomery S, Hajiebrahimi M, et al. Mortality trends for multiple sclerosis patients in Sweden from 1968 to 2012. Neurology. 2017;89(6):555-62.

33. Reynales-Shigematsu LM, Guerrero-Lopez CM, Hernandez Avila M, et al. Divergence and convergence in cause-specific premature adult mortality in Mexico and US Mexican Hispanics from 1995 to 2015: analyses of 4.9 million individual deaths. Int J Epidemiol. 2018;47(1):97-106.

34. Oza S, Lawn JE, Hogan DR, et al. Neonatal cause-of-death estimates for the early and late neonatal periods for 194 countries: 2000–2013. Bull World Health Organ. 2015;93(1):19-28.

35. Mack K, Clapperton A, Macpherson A, et al. Trends in the leading causes of injury mortality, Australia, Canada, and the United States, 2000–2014. Can J Public Health. 2017;108(2):e185-e91.

36. Padron-Monedero A, Damian J, Pilar Martin M, Fernandez-Cuenca R. Mortality trends for accidental falls in older people in Spain, 2000–2015. BMC Geriatr. 2017;17(1):276.

37. Ajdacic-Gross V, Weiss MG, Ring M, et al. Methods of suicide: international suicide patterns derived from the WHO mortality database. Bull World Health Organ. 2008;86(9):726-32.

38. Commission Regulation (EU) No 328/2011 of 5 April 2011 implementing Regulation (EC) No 1338/2008 of the European Parliament and of the Council on Community statistics on public health and health.and safety at work, as regards statistics on causes of death Text with EEA relevance. Official Journal of the European Union.

39. Harteloh P, de Bruin K, Kardaun J. The reliability of cause-of-death coding in The Netherlands. Eur J Epidemiol. 2010;25(8):531-8.

40. Lu TH, Lee MC, Chou MC. Accuracy of cause-of-death coding in Taiwan: types of miscoding and effects on mortality statistics. Int J Epidemiol. 2000;29(2):336-43.

41. Mieno MN, Tanaka N, Arai T, et al. Accuracy of death certificates and assessment of factors for misclassification of underlying cause of death. J Epidemiol. 2016;26(4):191-8.

# Supplementary Material

**Supplementary Table S1.** Best model for each cause of death in each European country.

Available only online due to limited space:
https://bmjopen.bmj.com/content/bmjopen/10/1/e031702/DC1/embed/inline-supplementary-material-1.pdf?download=true

# Part II

## Using multiple Cause-of-Death data to assess comorbidities at the end-of-life stage

# Chapter 4

## Systemic autoimmune disease as a Cause of Death: mortality burden and comorbidities

**Marianna Mitratza**, Bart Klijs, A. Elisabeth Hak, Jan W.P.F. Kardaun, Anton E. Kunst

*Published in*
*Rheumatology (Oxford) 2020;keaa537.*

# Abstract

**Background:** Systemic autoimmune diseases (SAIDs) have chronic trajectories and share characteristics of self-directed inflammation, as well as aspects of clinical expression. Nonetheless, burden-of-disease studies rarely investigate them as a distinct category. This study aims to assess the mortality rate of SAIDs as a group and to evaluate co-occurring causes of death.

**Methods:** We used death certificate data in the Netherlands, 2013–2017 (N = 711,247), and constructed a SAIDs list at the fourth-position ICD-10 level. The mortality rate of SAIDs as underlying cause of death (CoD), non-underlying CoD, and any-mention CoD was calculated. We estimated age-sex-standardized observed/expected (O/E) ratios to assess comorbidities in deaths with SAID relative to the general deceased population.

**Results:** We observed 3,335 deaths with SAID on their death certificate (0.47% of all deaths). The mortality rate of SAID was 14.6 per million population as underlying CoD, 28.0 as non-underlying CoD, and 39.7 as any-mention CoD. The mortality rate was higher for females and increased exponentially with age. SAID-related deaths were positively associated with all comorbidities except for solid neoplasms and mental conditions. Particularly strong was the association with diseases of the musculoskeletal system (O/E = 3.38; 95%CI: 2.98 ; 3.82), other diseases of the genitourinary system (O/E = 2.73; 95%CI: 2.18 ; 3.38), influenza (O/E = 2.71; 95%CI: 1.74 ; 4.03), blood diseases (O/E = 2.02; 95%CI: 1.70 ; 2.39), skin and subcutaneous tissue diseases (O/E = 1.95; 95%CI: 1.54 ; 2.45), and infectious diseases (O/E=1.85; 95%CI: 1.70 ; 2.01).

**Conclusions:** Systemic autoimmune diseases constitute a rare group of causes of death, but contribute to mortality through multiple comorbidities. Classification systems could be adapted to better encompass these diseases as a category.

# Introduction

Systemic autoimmune diseases (SAIDs), affect currently 400,000 people worldwide [1]. They constitute a subgroup of autoimmune diseases, a family of complex chronic diseases characterized by dysregulation of the adaptive and the innate immune system [2, 3]. SAIDs, such as SLE or sarcoidosis, share the potential for affecting multiple organs and tissues [2], and may cause similar clinical symptoms, including skin and joint manifestations, and are often treated with immunosuppressive drugs. Furthermore, several SAIDs have a shared genetic background and their development involves to some extent several overlapping molecular pathways that are activated by environmental triggers [4-7].

In spite of similarities in their pathophysiology and clinical expression, SAIDs are not commonly treated as one uniform group. This may be partially because there is no clear consensus on the list of disorders that belong to SAIDs. Moreover, disease classification systems such as the International Classification of Diseases (ICD) [8] are usually organized on the basis of the affected primary organ system, resulting in SAIDs being scattered across different disease chapters. For example, systemic lupus erythematosus (SLE) is classified in the chapter 'Diseases of the musculoskeletal system', whereas sarcoidosis is classified under 'Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism'. Because burden-of-disease studies often rely on standard disease categorizations, such as the ICD-10, SAIDs may be underrated in terms of their public health importance.



So far, only a limited number of studies have assessed the mortality burden of SAIDs together in one analysis. A recent study indicated an excess mortality risk for various SAIDs, but considered each SAID condition separately, reporting the highest burden for systemic sclerosis and systemic vasculitides [9]. Two other studies focussed on the mortality from autoimmune diseases as a group and found that 0.1–0.2% of all deaths had one of the included SAIDs reported as the underlying or any-mention cause of death (CoD) [10, 11]. Both studies, however, were restricted to women and did not cover the whole spectrum of SAIDs. Other previous studies report a prevalence of SAIDs ranging from 0.02% in children [12] to 0.05–0.4% in a nationwide population. However, these studies likely have underestimated the burden of SAIDs as a group, because of potential selection bias and the limited number of included SAIDs [13-16].

Because SAIDs can alter the risk of other diseases substantially [17], the causes of death (CoDs) reported among deaths with SAIDs may differ from those reported in the general population. In recent years, various population studies have assessed the cooccurrence of

single SAIDs with other diseases at the time of death, using death certificate data. SLE has been associated with a higher reporting of coagulation and haemorrhagic disorders, and renal failure [18]. Other multiple-cause-of-death studies reported associations of giant cell arteritis with cardiovascular diseases and infections [19], of systemic sclerosis with cardiopulmonary disorders [20], and of sarcoidosis with tuberculosis, chronic respiratory diseases and infections [21]. Furthermore, SAIDs, as members of the autoimmune diseases' family, have been found to be associated with a higher risk of a second autoimmune disease [14], vascular dementia [22], and mood disorders [23], as well as increased risk of death from cancer [24, 25].

The aim of this study was to investigate SAIDs as a group, joining CoDs from different chapters of the ICD-10 classification. We used this group to assess the mortality rate of SAID as underlying and non-underlying CoD. Furthermore, we assessed rates of co-occurring CoDs in deceased persons with SAID compared with the general population. For the purpose of this study, we presented a novel approach of using all information as reported on the death certificates in order to broaden our knowledge on the associations of SAIDs with serious comorbidities at the time of death.

# Methods

## Data collection

We obtained anonymous data from the Cause of Death database of Statistics Netherlands. In the Netherlands, for each deceased person, a death certificate is completed by the attending or the forensic physician, following the World Health Organization (WHO) model. Three lines are available to describe the causal chain of morbid events leading to death (Part I), and one line for diseases and conditions contributing to death (Part II). The ICD-coded output includes both the underlying CoD – i.e. the disease or injury initiating the chain of morbid events leading directly to death [8] – and the non-underlying – i.e. intermediate or contributory – CoD. A CoD is commonly termed as 'any-mention CoD' when reported either as the underlying or a non-underlying cause of death on the death certificate. Diseases are coded according to the 10$^{th}$ revision of the International Classification of Diseases (ICD-10) classification of the WHO [8]. Date of birth, date of death, sex and mention of a systemic autoimmune disease as underlying, non-underlying, and any-mention CoD were extracted from this data source for the years 2013–2017 for all deaths. Years before 2013 were omitted, because automatic coding and selection of the

underlying cause of death was introduced in 2013, with the adoption of Iris 4.4.1 software [26].

SAIDs were identified based on a list of ICD-10 fourth-position codes that we constructed based on relevant studies [1, 5, 7, 12, 15, 27-29] and expert knowledge of an internist-clinical immunologist who is co-authoring this paper. We made a distinction between vasculitides and non-vasculitides (**Supplementary Table S1**), because of differences in their clinical presentation and complications (mainly ischaemic in vasculitides). Comorbidities of systemic autoimmune diseases as mentioned on the death certificate were classified using ICD-10 third-position or fourth-position ICD-10 codes, grouped according to the European Shortlist for Causes of Death 2012 [30]. We modified this list to include diseases of the kidney and urinary tract infections as distinct subcategories of the Diseases of the genitourinary system and influenza and pneumonia as distinct subcategories of Diseases of the respiratory system, because of their clinical significance (**Supplementary Table S2**).

Our study population consisted of decedents in the Netherlands (2013–2017). General population data for this period were provided by Statistics Netherlands. Death certificates for persons ≤1 year old were excluded from the study population as they follow a different process of coding. According to the Dutch law on medical research, no ethical approval was required for this study, as no living subjects were involved and all data were anonymized.



## Statistical analysis

We calculated the number of deaths that listed each of the main SAID subclasses on the death certificate as the underlying CoD, a non-underlying CoD, and any-mention CoD. We estimated the ratio of underlying CoD to any-mention CoD by SAID subclass, to determine which diseases are relatively more important as underlying rather than contributing causes of death [31]. We calculated the absolute number of deaths and annual age-specific mortality rates (per 1,000,000 population) for the total study population, and by sex. These results are presented stratified by mention of SAID as the underlying CoD, a non-underlying CoD, and any-mention CoD on the death certificate for the total SAID group, as well as the subgroups of vasculitides and non-vasculitides.

To assess the association of SAIDs with other causes of death, we performed a multiple-cause-of-death analysis. We analysed the CoDs in death certificates of deceased persons with SAID as any-mention, non-underlying or underlying CoD, respectively, and present these figures as absolute numbers and percentages. A death certificate may have more

than one non-underlying CoD listed, therefore the non-underlying CoDs examined were non-exclusive. In addition, we calculated the ratio of the observed (O) number of deaths to the expected (E) number of deaths, termed as the O/E ratio [32], with the expected cases adjusted for sex and age using as reference population the general Dutch deceased population between 2013 and 2017 [711,247 deaths, mean age 78 years (S.D. = 13.6), 369,255 females]. An O/E ratio over 1 means that the combination of SAID and the CoD under examination was more frequent than would be expected if these diseases were independent. The CIs of the O/E ratios were estimated using Byar's method [33].

We used R software for the statistical analyses (The R Foundation for Statistical Computing, version 3.2.3).

# Results

In the 5-year period of study (2013–2017), a SAID was reported on 3,335 death certificates as underlying or non-underlying CoD. This corresponds to 0.47% of all 711,247 deaths in the Netherlands.

**Table 1** presents the number of deaths with SAID reported as underlying, non-underlying or any-mention CoD, according to the main subclasses of the disease group. Overall, among non-vasculitides and vasculitides, systemic sclerosis (n = 268) was the main SAID subclass reported as underlying CoD, while PMR (n = 844) was the main subclass as non-underlying CoD. Systemic vasculitis was the main subclass among vasculitides, both as underlying CoD (n = 224), and non-underlying CoD (n = 178).

Among non-vasculitides, the ratio of underlying CoD to any-mention CoD was the highest for other specified systemic involvement of connective tissue (0.89) and systemic sclerosis (0.67). Thus, these diseases were more likely to be considered as underlying CoD, when reported. Polymyalgia rheumatica (PMR) had the lowest ratio (0.09), excluding subclasses with very few cases. Among vasculitides, the ratio of underlying CoD to any-mention CoD was the highest for Goodpasture syndrome and Behcet's disease (0.71), and the lowest for giant cell arteritis (0.33).

**Table 1.** Deaths with SAID as cause of death, by subclass, the Netherlands 2013–2017.

| Systemic Autoimmune Disease | Underlying CoD, number | Non-underlying CoD, number | Any-mention CoD, number | Underlying CoD / Any-mention CoD ratio |
|---|---|---|---|---|
| **Non-Vasculitides** [a] | | | | |
| Systemic lupus erythematosus | 100 | 222 | 251 | 0.40 |
| Dermatopolymyositis | 100 | 112 | 212 | 0.47 |
| Sjögren syndrome | 48 | 156 | 204 | 0.24 |
| Systemic sclerosis | 268 | 265 | 403 | 0.67 |
| Mixed connective tissue disease | 17 | 17 | 34 | 0.50 |
| Polymyalgia rheumatica | 81 | 844 | 925 | 0.09 |
| Antiphospholipid syndrome | 24 | 21 | 45 | 0.53 |
| Sarcoidosis | 201 | 304 | 500 | 0.40 |
| Amyloidosis | 0 | 2 | 2 | 0.00 |
| Still disease | 2 | 4 | 6 | 0.33 |
| IgG4-related disease | 18 | 83 | 101 | 0.18 |
| Cogan's disease | 0 | 3 | 3 | 0.00 |
| Other specified systemic involvement of connective tissue | 17 | 2 | 19 | 0.89 |
| Systemic involvement of connective tissue, unspecified | 66 | 75 | 141 | 0.47 |
| **Vasculitides** | | | | |
| ANCA- associated vasculitis | 205 | 144 | 349 | 0.59 |
| Giant cell arteritis | 52 | 105 | 157 | 0.33 |
| Other systemic vasculitis | 30 | 41 | 71 | 0.42 |

a) There are no deaths with mention of relapsing polychondritis.
SAID, systemic autoimmune disease; CoD, cause of death; ANCA, antineutrophil cytoplasmic antibody

**Table 2** shows the age-specific mortality from SAID, according to the type of report on death certificates, by sex. About 60.2% (n = 2,007) of the any-mention SAID deaths occurred in females. The corresponding percentages for the deaths with SAID as the underlying CoD and SAID as a non-underlying CoD were 57.2% (n = 703) and 62.2% (n = 1,461), respectively. At all ages, and in each of the three CoD reporting types, females

had a higher mortality rate than males. For both genders, and in each of the three CoD reporting types, there was an exponential increase of the annual mortality rate with age. Females over 80 years had the highest mortality burden, with a mortality rate of 453.9 deaths with any mention of SAID per 1,000,000 population. The median age of death of decedents with SAID as the underlying CoD was 74 years (IQR = 18), 6 years lower than the general population, with 17.3% dying younger than 60 years old. **Supplementary Table S3** shows the mortality rates of vasculitides and non-vasculitides separately.

**Table 2.** Deaths with SAID as cause of death, by age and sex.

| | Underlying CoD | | Non-underlying CoD | | Any-mention CoD | |
|---|---|---|---|---|---|---|
| | number | annual rate per 1,000,000 population | number | annual rate per 1,000,000 population | number | annual rate per 1,000,000 population |
| **Females** | | | | | | |
| <40 | 17 | 0.9 | 25 | 1.3 | 35 | 1.8 |
| 40–59 | 99 | 8.2 | 110 | 9.1 | 179 | 14.7 |
| 60–79 | 317 | 36.9 | 493 | 57.4 | 726 | 84.5 |
| ≥80 | 270 | 114.9 | 833 | 354.4 | 1,067 | 453.9 |
| Total | 703 | 16.6 | 1,461 | 34.5 | 2,007 | 47.4 |
| **Males** | | | | | | |
| <40 | 16 | 0.8 | 18 | 0.9 | 32 | 1.6 |
| 40–59 | 80 | 6.6 | 97 | 8.0 | 160 | 13.1 |
| 60–79 | 278 | 34.0 | 427 | 52.2 | 651 | 79.5 |
| ≥80 | 152 | 112.2 | 345 | 254.6 | 485 | 357.9 |
| Total | 526 | 12.7 | 887 | 21.3 | 1,328 | 31.9 |
| **Females and Males** | | | | | | |
| <40 | 33 | 0.8 | 43 | 1.1 | 67 | 1.7 |
| 40–59 | 179 | 7.4 | 207 | 8.5 | 339 | 13.9 |
| 60–79 | 595 | 35.5 | 920 | 54.8 | 1,377 | 82.1 |
| ≥80 | 422 | 113.9 | 1,178 | 317.9 | 1,552 | 418.8 |
| Total | 1,229 | 14.6 | 2,348 | 28.0 | 3,335 | 39.7 |

SAID, systemic autoimmune disease; CoD, cause of death.

The report of different types of disorders as any-mention CoD on death certificates with any mention of a SAID is provided in **Table 3**. Leading any-mention CoDs were diseases of the circulatory system (55.5%), followed by diseases of the respiratory system (35.9%), endocrine and metabolic disorders (21.6%), and neoplasms (20.9%). Deaths with SAID as a cause of death had a higher report of all types of conditions as cooccurring CoDs than deaths in the general population, with the exception of solid neoplasms and mental disorders. Particularly high was the O/E ratio for diseases of the musculoskeletal system (3.38), other diseases of the genitourinary system (2.73), influenza (2.71), and diseases of the blood (2.02). Diseases of the skin and subcutaneous tissue and infectious diseases also presented high O/E ratios (1.95 and 1.85, respectively). A low O/E ratio was observed for solid neoplasms (0.48) and mental disorders (0.71).

An overview of the comorbidities reported on the death certificate when SAID was the underlying CoD or a non-underlying CoD, respectively, is given in **Table 4**. For most comorbidities, the associations with SAID had the same direction – positive or negative – regardless of the type of reporting on the death certificate. For some comorbidities, there were opposite directions when SAID was the underlying CoD compared with when SAID was a non-underlying CoD. For example, a positive association with pneumonia was observed only when SAID was the underlying CoD (O/E = 2.77 vs 0.81). For endocrine, nutritional and metabolic disease, a positive association was found only when SAID was a non-underlying CoD (O/E = 0.71 vs 1.68). **Supplementary Table S4** shows the corresponding results in detail for different diseases of the circulatory system as a CoD.

Of all SAID cases, 2.4% (79 deceased) had a second type of SAID mentioned on their death certificates. Report of a SAID as non-underlying CoD was much more frequent when another SAID was the underlying or non-underlying CoD, and vice versa (O/E =10.45 vs 10.81). The most common SAID pairs in our study were giant cell arteritis with PMR (9 deaths); SLE with antiphospholipid syndrome (6); systemic vasculitis with IgG4-related disease (4); SLE with Sjögren syndrome (4); systemic sclerosis with mixed connective tissue disease (5)/Sjögren syndrome (5)/SLE (3).

**Table 3.** Deaths with SAID as any-mention CoD, and with other condition as any-mention CoD.

| Condition | SAID as any-mention cause of death | | |
|---|---|---|---|
| | number | % [a] | Age-sex-standardized observed/expected (O/E) ratios [95%CI] |
| Infectious and parasitic diseases | 552 | 16.6 | 1.85 [1.70 ; 2.01] |
| Neoplasms | 696 | 20.9 | 0.55 [0.51 ; 0.60] |
| Malignant neoplasms | 643 | 19.3 | 0.53 [0.49 ; 0.58] |
| Solid malignant neoplasms | 534 | 16.0 | 0.48 [0.44 ; 0.52] |
| Malignant neoplasms of lymphoid, haematopoietic tissue | 112 | 3.4 | 1.09 [0.90 ; 1.32] |
| Non-malignant neoplasms | 76 | 2.3 | 1.16 [0.92 ; 1.46] |
| Diseases of the blood and blood-forming organs | 137 | 4.1 | 2.02 [1.70 ; 2.39] |
| Endocrine, nutritional and metabolic diseases | 721 | 21.6 | 1.27 [1.18 ; 1.36] |
| Mental and behavioural disorders | 379 | 11.4 | 0.71 [0.64 ; 0.79] |
| Diseases of the nervous system and the sense organs | 432 | 13.0 | 1.18 [1.07 ; 1.30] |
| Diseases of the circulatory system | 1,852 | 55.5 | 1.20 [1.15 ; 1.26] |
| Ischaemic heart diseases | 336 | 10.1 | 1.11 [0.99 ; 1.23] |
| Acute myocardial infarction | 130 | 3.9 | 0.69 [0.57 ; 0.82] |
| Other ischaemic heart diseases | 227 | 6.8 | 1.08 [0.94 ; 1.23] |
| Other heart diseases | 1,189 | 35.7 | 1.20 [1.13 ; 1.27] |
| Cerebrovascular diseases | 321 | 9.6 | 0.90 [0.80 ; 1.00] |
| Other diseases of the circulatory system | 758 | 22.7 | 1.86 [1.73 ; 1.99] |
| Diseases of the respiratory system | 1,197 | 35.9 | 1.44 [1.36 ; 1.52] |
| Influenza | 24 | 0.7 | 2.71 [1.74 ; 4.03] |
| Pneumonia | 553 | 16.6 | 1.54 [1.41 ; 1.67] |
| Chronic obstructive pulmonary disease | 309 | 9.3 | 0.87 [0.78 ; 0.98] |
| Other diseases of the respiratory system | 600 | 18.0 | 1.67 [1.54 ; 1.81] |
| Diseases of the digestive system | 326 | 9.8 | 1.28 [1.15 ; 1.43] |

| Condition | SAID as any-mention cause of death | | |
|---|---|---|---|
| | number | % [a] | Age-sex-standardized observed/expected (O/E) ratios [95%CI] |
| Diseases of the skin and subcutaneous tissue | 75 | 2.2 | 1.95 [1.54 ; 2.45] |
| Diseases of the musculoskeletal system | 255 | 7.6 | 3.38 [2.98 ; 3.82] |
| Diseases of the genitourinary system | 578 | 17.3 | 1.65 [1.52 ; 1.79] |
| Diseases of the kidney | 444 | 13.3 | 1.73 [1.58 ; 1.90] |
| Urinary tract infections | 106 | 3.2 | 1.22 [1.00 ; 1.48] |
| Other diseases of the genitourinary system | 85 | 2.5 | 2.73 [2.18 ; 3.38] |
| Symptoms, signs, ill-defined causes | 873 | 26.2 | 1.12 [1.05 ; 1.19] |
| External causes of morbidity and mortality | 195 | 5.8 | 0.82 [0.71 ; 0.95] |

a) % represents the proportion of subjects with a specific any-mention cause of death among all subjects with SAID as any-mention cause of death.

SAID, systemic autoimmune disease; CoD, cause of death.

**Table 4:** Number of deaths with SAID as the underlying CoD (or a non-underlying CoD) and with other conditions as a non-underlying CoD (or the underlying CoD), by condition, the Netherlands 2013–2017.

| Condition | SAID as underlying cause of death | | | SAID as non-underlying cause of death | | |
|---|---|---|---|---|---|---|
| | number | % [a] | Age-sex-standardized observed/expected (O/E) ratios [95%CI] | number | % [b] | Age-sex-standardized observed/expected (O/E) ratios [95%CI] |
| Infectious and parasitic diseases | 215 | 17.5 | 2.69 [2.34 ; 3.08] | 102 | 4.3 | 2.12 [1.72 ; 2.57] |
| Neoplasms | 69 | 5.6 | 0.26 [0.20 ; 0.33] | 504 | 21.5 | 0.72 [0.66 ; 0.79] |
| Malignant neoplasms | 56 | 4.6 | 0.23 [0.17 ; 0.29] | 478 | 20.4 | 0.71 [0.65 ; 0.78] |
| Solid malignant neoplasms | 42 | 3.4 | 0.17 [0.12 ; 0.23] | 395 | 16.8 | 0.64 [0.58 ; 0.70] |
| Malignant neoplasms of lymphoid, haematopoietic tissue | 14 | 1.1 | 1.91 [1.04 ; 3.21] | 83 | 3.5 | 1.58 [1.26 ; 1.96] |
| Non-malignant neoplasms | 16 | 1.3 | 1.81 [1.03 ; 2.93] | 26 | 1.1 | 0.99 [0.65 ; 1.46] |
| Diseases of the blood and blood-forming organs | 42 | 3.4 | 2.24 [1.61 ; 3.02] | 10 | 0.4 | 1.51 [0.72 ; 2.78] |
| Endocrine, nutritional and metabolic diseases | 114 | 9.3 | 0.71 [0.59 ; 0.86] | 93 | 4.0 | 1.68 [1.35 ; 2.05] |
| Mental and behavioural disorders | 48 | 3.9 | 0.49 [0.36 ; 0.66] | 150 | 6.4 | 0.87 [0.73 ; 1.02] |
| Diseases of the nervous system and the sense organs | 99 | 8.1 | 1.45 [1.18 ; 1.76] | 127 | 5.4 | 1.09 [0.91 ; 1.30] |
| Diseases of the circulatory system | 558 | 45.4 | 1.52 [1.39 ; 1.65] | 627 | 26.7 | 1.09 [1.00 ; 1.18] |
| Ischaemic heart diseases | 56 | 4.6 | 1.25 [0.95 ; 1.62] | 142 | 6.0 | 1.06 [0.89 ; 1.24] |
| Acute myocardial infarction | 20 | 1.6 | 1.69 [1.03 ; 2.60] | 80 | 3.4 | 1.02 [0.81 ; 1.27] |
| Other ischaemic heart diseases | 36 | 2.9 | 1.06 [0.74 ; 1.47] | 62 | 2.6 | 1.23 [0.94 ; 1.58] |
| Other heart diseases | 372 | 30.3 | 1.50 [1.35 ; 1.66] | 217 | 9.2 | 0.96 [0.84 ; 1.10] |
| Cerebrovascular diseases | 68 | 5.5 | 1.30 [1.01 ; 1.65] | 125 | 5.3 | 0.86 [0.72 ; 1.02] |
| Other diseases of the circulatory system | 215 | 17.5 | 1.96 [1.71 ; 2.24] | 143 | 6.1 | 1.86 [1.57 ; 2.20] |

| Condition | SAID as underlying cause of death | | | SAID as non-underlying cause of death | | |
|---|---|---|---|---|---|---|
| | number | % [a] | Age-sex-standardized observed/expected (O/E) ratios [95%CI] | number | % [b] | Age-sex-standardized observed/expected (O/E) ratios [95%CI] |
| Diseases of the respiratory system | 506 | 41.2 | 2.22 [2.03 ; 2.43] | 208 | 8.9 | 1.16 [1.01 ; 1.33] |
| Influenza | 3 | 0.2 | 3.80 [0.76 ; 11.12] | 19 | 0.8 | 4.08 [2.45 ; 6.36] |
| Pneumonia | 264 | 21.5 | 2.77 [2.44 ; 3.12] | 38 | 1.6 | 0.81 [0.57 ; 1.11] |
| Chronic obstructive pulmonary disease | 70 | 5.7 | 1.13 [0.88 ; 1.43] | 90 | 3.8 | 0.91 [0.73 ; 1.12] |
| Other diseases of the respiratory system | 291 | 23.7 | 2.57 [2.29 ; 2.89] | 61 | 2.6 | 2.14 [1.64 ; 2.75] |
| Diseases of the digestive system | 77 | 6.3 | 1.18 [0.93 ; 1.47] | 101 | 4.3 | 1.47 [1.20 ; 1.79] |
| Diseases of the skin and subcutaneous tissue | 21 | 1.7 | 2.16 [1.34 ; 3.30] | 7 | 0.3 | 1.61 [0.64 ; 3.31] |
| Diseases of the musculoskeletal system | 55 | 4.5 | 3.20 [2.41 ; 4.17] | 41 | 1.7 | 2.86 [2.06 ; 3.89] |
| Diseases of the genitourinary system | 210 | 17.1 | 2.32 [2.02 ; 2.66] | 70 | 3.0 | 1.49 [1.16 ; 1.88] |
| Diseases of the kidney | 183 | 14.9 | 2.69 [2.31 ; 3.10] | 29 | 1.2 | 1.07 [0.72 ; 1.54] |
| Urinary tract infections | 22 | 1.8 | 1.16 [0.73 ; 1.76] | 27 | 1.1 | 1.73 [1.14 ; 2.52] |
| Other diseases of the genitourinary system | 22 | 1.8 | 2.58 [1.61 ; 3.90] | 14 | 0.6 | 3.19 [1.74 ; 5.36] |
| Symptoms, signs, ill-defined causes | 275 | 22.4 | 1.21 [1.07 ; 1.36] | 12 | 0.5 | 0.23 [0.12 ; 0.41] |
| External causes of morbidity and mortality | 54 | 4.4 | 0.63 [0.48 ; 0.83] | 48 | 2.0 | 0.47 [0.35 ; 0.63] |
| **SAID** | 43 | 3.5 | 10.45 [7.56 ; 14.07] | 43 | 1.8 | 10.81 [7.82 ; 14.56] |

a) % represents the proportion of subjects with a specific non-underlying cause of death among the subjects with SAID as the underlying cause of death.

b) % represents the proportion of subjects with the specific underlying cause of death among the subjects with SAID as a non-underlying cause of death.

SAID, systemic autoimmune disease; CoD, cause of death.

# Discussion

In this study, we analysed >3,000 Dutch death certificates with mention of a systemic autoimmune disease. Our findings suggest that SAID is a rare group of causes of death, with a mortality rate of about 15 and about 40 per million population measured as an underlying CoD and any-mention CoD, respectively. The mortality rate is higher for female decedents and is rising exponentially with age. We identified a pattern of comorbidities predominant in decedents dying with SAID that involved musculoskeletal, genitourinary and blood disorders, as well as infections – including influenza and pneumonia – and skin disorders.

This is one of the first studies to construct an exhaustive list of SAIDs and investigate SAIDs as a group of causes of death. We used multiple-cause-of-death data, covering the total Dutch population. The use of national data, which are universal and cover all sociodemographic backgrounds, avoids the potential selection bias of cohort studies. We conducted a multiple-cause-of-death analysis that allows quantification of the association between each comorbidity at the time of death and SAID, taking into consideration the role of diseases in the dying causal chain.

The limitations of this study include properties of the data used. First, multiple-cause-of-death data come with the known risk of occasional arbitrary classification of a condition as non-underlying instead of underlying CoD or generally its omission from the death certificates. Diseases may be omitted due to restricted relevance to death, either because of low severity, or being in remission. The way that certificates are filled in depends on the subjective judgement of the physician, who is supposed to decide between reporting every possible condition or only the conditions that evidently contributed to death. In the absence of clinical records to verify each individual diagnosis and to capture all comorbidities, our findings may be conservative and underestimate the contribution of SAID to mortality in the Netherlands. A recent study on SLE found substantial underreporting on death certificates of cases, especially for older people [34]. However, the extent of the potential underreporting for SAID as a group has not been studied explicitly in the literature.

Second, we cannot make causal inferences about the relationship of SAID with another disease at the time of death, as we examine cross-sectional associations. However, when a disease tends to occur with SAID as underlying CoD, but not with SAID as non-underlying CoD, this would indicate that this disease acts more likely as a complication of SAID rather than a concurrent independent condition contributing to death. Although, observer-

certifier variation in the reporting of associations affects the accuracy of information for individual patients or groups of patients, for large-scale population-based estimates a cancelling-out of inaccuracies might be expected. Third, unfortunately, there were not enough cases to perform a separate analysis of comorbidities for patients dying from either vasculitides or non-vasculitides, nor patients by sex and age class.

Our finding of ~15 persons per million population dying annually from a SAID as the underlying CoD (0.2% of all deaths) is in relative agreement with previous evidence of 10–17 persons per million in western countries, based on numbers for selected SAIDs analysed among the much larger group of autoimmune diseases in nationwide data [10, 11] or as part of registries [9]. When investigating SAID as any-mention CoD, we found that ~40 persons per million had mention of a SAID on their death certificates (0.5% of all deaths). This estimate was much higher than a study reporting 17 persons per million population, but missed common SAIDs such as PMR, sarcoidosis and giant cell arteritis [10].

The predominance of systemic sclerosis and systemic vasculitis, particularly ANCA-associated vasculitis, as leading underlying CoD among SAIDs in our population is consistent with a recent Norwegian study [9].



The considerably higher mortality rate of SAID for females in our study (50% higher in any-mention CoD analysis) may reflect the worldwide prevalence estimates of three times more women suffering from a SAID [1]. Contrary to our expectations, a recent registry study reported higher mortality for males [9]. Given the fact that we have no information on the clinical course of SAIDs in those who died from or with the disease, we cannot make claims on the ways in which sex is associated with mortality of SAID patients or what caused this excess female mortality. For some diseases, like SLE, studies are inconclusive about the role of sex in mortality, reporting higher incidence in females, but inconsistent evidence on the clinical course and mortality [35].

Our finding that the mortality risk from SAID in the population is increasing with age may be explained by lifetime accumulation of damage involving several vital systems and more rarely by occurrence of elderly-onset cases [36].

Deaths with report of one type of SAID were proportionally more frequent in cases with another SAID as CoD compared with the general population. This finding implies that patients with a SAID are more likely to suffer from a concomitant type of SAID at the time of death. Other studies have found a higher risk for a second autoimmune disease

for people who suffer from one autoimmune disease [14]. Shared determinants, such as genes, environment and lifestyle have been proposed as underlying mechanisms [4-7].

The more frequent report of cardiovascular disease on death certificates of subjects with SAID as the underlying CoD compared with the general population is in agreement with previous evidence regarding the occurrence of complications, such as stroke and heart disease [17, 37, 38]. Mechanisms involved in the development of cardiovascular disease include the formation of atherosclerotic plaques via inflammation, most studied for SLE and systemic sclerosis, as well as procoagulant activity predominant in antiphospholipid syndrome. The role of treatment-related cardiac toxicity in SAID has been also described [38].

Our results corroborate previous findings regarding the importance of infections as a cause of death in patients with SAID [39, 40]. About 17% of the deceased with SAID in our study had an infection that contributed directly or indirectly to death, in line with a recent SAID mortality study [9]. General infections, pneumonia, genitourinary infections and, most of all, influenza were reported more frequently on death certificates with a mention of SAID. However, it is not known whether an infection can be attributed to disease activity or the effect of immunosuppressive drugs, therefore is a topic for further study. Our finding that pneumonia had more frequent report on death certificates with SAID only as an underlying CoD may imply that it is recognized as a particular lethal pathway for these patients.

Although 16% of our study population had a solid malignant neoplasm at the time of death, the rate of deaths with any mention of a solid neoplasm was substantially lower compared with the general population. This finding may seem surprising, as SAID cannot be considered as a protective factor, but is consistent with prior publications [19-21, 39]. A potential explanation may be surveillance bias, with SAID patients having more frequent visits to specialists, therefore getting an earlier detection and treatment of cancer. At the same time, cancer has high prominence in the death certification process, often resulting in fewer mentions of other diseases, because it might be assumed to have contributed to death on its own. On the contrary, malignant neoplasms of lymphoid, and haematopoietic tissue were more frequently reported as a cause of death in deceased with SAID, although still rare. A meta-analysis found that SLE and Sjögren syndrome are associated with occurrence of non-Hodgkin lymphoma [41]. Immunomodulatory therapies have been associated with development of lymphomas as a rare adverse effect.

In our study, several other conditions were associated with mention of SAID on the death certificate. The high number of mentions of musculoskeletal disease may be partially explained by the fact that musculoskeletal disease is often an inherent part of SAID. Disorders of the blood and blood-forming organs, as well as diseases of the genitourinary, the respiratory and the digestive system were associated with SAID, especially when SAID was reported as underlying CoD. This reflects to a large extent known complications, such as haemolytic anaemia in SLE [42], renal failure in ANCA-associated vasculitides and SLE [40, 42], lung failure in systemic sclerosis and dermatopolymyositis [27], and gastrointestinal ulcers or perforation in vasculitides and dermatopolymyositis [28]. Of note, skin and subcutaneous tissue disease, a not so well-studied complication or comorbidity, was found to have a strong association with SAID reporting on the death certificate. Antiphospholipid syndrome, vasculitides and Sjögren syndrome have been documented as SAIDs with potentially severe skin involvement [29]. Similarly, neurological disorders may occur more often in SAID patients in the form of peripheral neuropathy [43]. Endocrine, nutritional and metabolic diseases had a high occurrence in deceased with SAID, which was driven only via their reporting as an underlying CoD. This evidence reinforces the hypothesis that diseases such as hypothyroidism [44] and diabetes mellitus [45], may not act as a complication of SAID, but rather a comorbidity.



Although survival of patients with SAID has improved during recent years, our findings extend our knowledge on ways to further advance it. Clinically relevant comorbidities may be targets for secondary prevention, with complete vaccination coverage for influenza and pneumonia infections being the most prominent. In addition, it is recommended that physicians monitor closely the respiratory, renal and cardiovascular functions of the patients and treat early known complications that may become life-threatening. At the same time, more rare but severe conditions that may contribute to death directly or indirectly, such as musculoskeletal, skin and blood disorders, as well as lymphomas, should be taken into consideration when assessing the overall physical state and quality of life of the patients. Further research is needed to fully understand the mechanisms driving mortality in SAID patients and to define subgroups vulnerable for certain complications.

Finally, we presented a novel approach of using all death certificate information in order to assess how SAID relates to other diseases in affecting mortality in the general population. Whereas previous studies applied multiple-cause-of-death analysis to specific SAID diseases, we were the first to apply this approach to SAID as one class. Countries following the WHO model of death certificate and coding systematically all mentioned causes of death could adopt our method. Such results may be used also to monitor mortality related with SAIDs and inform policies to address them. This is relevant particularly in view

of the forthcoming ICD-11 introduction [46], in which the problem of SAID classification is mitigated, but there is room for improvement, especially when new autoinflammatory diseases are being discovered. Mortality classification systems need more flexibility to allow for such new views on the grouping of diseases to be incorporated.

To our knowledge, this is the first study to investigate systemic autoimmune diseases as a group using the multiple-cause-of-death approach. Systemic autoimmune diseases compose a rare group of causes of death, but can contribute to mortality through various comorbidities. Reclassification of readily available data provides useful estimates of the mortality burden of systemic autoimmune diseases in the general population.

# References


1. Ramos-Casals M, Brito-Zeron P, Kostov B, et al. Googledriven search for big data in autoimmune geoepidemiology: analysis of 394,827 patients with systemic autoimmune diseases (review). Autoimmun Rev. 2015;14:670-9.
2. Arakelyan A, Nersisyan L, Poghosyan D, et al. Autoimmunity and autoinflammation: a systems view on signaling pathway dysregulation profiles. PLoS One. 2017;12:e0187572.
3. McGonagle D and McDermott MF. A proposed classification of the immunological diseases. PLoS Med. 2006;3:e297.
4. Zhernakova A, Withoff S, Wijmenga C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases (review). Nat Rev Endocrinol. 2013;9:646-59.
5. Teruel M, Chamberlain C, Alarcon-Riquelme ME. Omics studies: their use in diagnosis and reclassification of SLE and other systemic autoimmune diseases (review). Rheumatology (Oxford). 2017;56:i78-i87.
6. Hudson M, Bernatsky S, Colmegna I, et al. Novel insights into systemic autoimmune rheumatic diseases using shared molecular signatures and an integrative analysis. Epigenetics. 2017;12:433-40.
7. Wahren-Herlenius M and Dörner T. Immunopathogenic mechanisms of systemic autoimmune disease (review). Lancet. 2013;382:819-31.
8. World Health Organization (WHO). International statistical classification of diseases and related health problems, 10th revision: Volume II. Geneva: World Health Organization, 1993.
9. Garen T, Lerang K, Hoffmann-Vold AM, et al. Mortality and causes of death across the systemic connective tissue diseases and the primary systemic vasculitides. Rheumatology (Oxford). 2019;58:313-20.
10. Thomas SL, Griffiths C, Smeeth L, et al. Burden of mortality associated with autoimmune diseases among females in the United Kingdom. Am J Public Health. 2010;100:2279-87.
11. Walsh SJ and Rau LM. Autoimmune diseases: a leading cause of death among young and middle-aged women in the United States. Am J Public Health. 2000;90:1463-6.
12. Shiff NJ, Lix LM, Joseph L, et al. The prevalence of systemic autoimmune rheumatic diseases in Canadian pediatric populations: administrative database estimates. Rheumatol Int. 2015;35:569-73.
13. Jacobson DL, Gange SJ, Rose NR, Graham NMH. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. Clin Immunol Immunopathol. 1997;84:223-43.
14. Sardu C, Cocco E, Mereu A, et al. Population based study of 12 autoimmune diseases in Sardinia, Italy: prevalence and comorbidity. PLoS One. 2012;7:e32487.
15. Bernatsky S, Lix L, Hanly JG, et al. Surveillance of systemic autoimmune rheumatic diseases using administrative data. Rheumatol Int. 2011;31:549-54.
16. Eaton WW, Pedersen MG, Atladottir HO, et al. The prevalence of 30 ICD-10 autoimmune diseases in Denmark. Immunol Res. 2010;47:228-31.

17. Hak AE, Karlson EW, Feskanich D, et al. Systemic lupus erythematosus and the risk of cardiovascular disease: results from the Nurses' health study. Arthritis Rheum. 2009;61:1396-402.

18. Falasinnu T, Chaichian Y, Simard JF. Impact of sex on systemic lupus erythematosus-related causes of premature mortality in the United States. J Womens Health (Larchmt). 2017;26:1214-21.

19. Aouba A, Gonzalez Chiappe S, Eb M, et al. Mortality causes and trends associated with giant cell arteritis: analysis of the French national death certificate database (1980–2011). Rheumatology (Oxford). 2018;57:1047-55.

20. Elhai M, Meune C, Boubaya M, et al. Mapping and predicting mortality from systemic sclerosis. Ann Rheum Dis. 2017;76:1897-905.

21. Jamilloux Y, Maucort-Boulch D, Kerever S, et al. Sarcoidosis-related mortality in France: a multiple-cause-of-death analysis. Eur Respir J. 2016;48:1700-9.

22. Wotton CJ, Goldacre MJ. Associations between specific autoimmune diseases and subsequent dementia: retrospective record-linkage cohort study, UK. J Epidemiol Community Health. 2017;71:576-83.

23. Benros ME, Waltoft BL, Nordentoft M et al. Autoimmune diseases and severe infections as risk factors for mood disorders: a nationwide study. JAMA Psychiatry. 2013;70:812-20.

24. Hemminki K, Liu X, Ji J, et al. Effect of autoimmune diseases on mortality and survival in subsequent digestive tract cancers. Ann Oncol. 2012;23:2179-84.

25. Yu KH, Kuo CF, Huang LH, et al. Cancer risk in patients with inflammatory systemic autoimmune rheumatic diseases: a nationwide population-based dynamic cohort study in Taiwan. Medicine (Baltimore). 2016;95:e3540.

26. Harteloh PPM. The implementation of an automated coding system for cause-of-death statistics. Inform Health Soc Care. 2020;45:1-14.

27. Rayner CF and Grubnic S. Pulmonary manifestations of systemic autoimmune disease. Best Pract Res Clin Rheumatol. 2004;18:381-410.

28. Cojocaru M, Cojocaru IM, Silosi I, Vrabie CD. Gastrointestinal manifestations in systemic autoimmune diseases. Maedica (Buchar). 2011;30:1-51.

29. Rashtak S, Pittelkow MR. Skin involvement in systemic autoimmune diseases (review). Curr Dir Autoimmun. 2008;10:344-58.

30. European shortlist for causes of death, 2012 [Internet]. Luxembourg: Eurostat, 2016. http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=COD_2012&StrLanguageCode=EN&IntPcKey=&StrLayoutCode=HIERARCHIC

31. Redelings MD, Sorvillo F, Simon P. A comparison of underlying cause and multiple causes of death: US vital statistics, 2000–2001. Epidemiology. 2006;17:100-3.

32. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause-of-death data. Am J Epidemiol. 1986;124:161-79.

33. Rothman KJ, Boice JD Jr. Epidemiologic analysis with a programmable calculator. NIH Pub No. 79-1649. Bethesda, MD: National Institutes of Health, 1979: 31-2.

34. Falasinnu T, Rossides M, Chaichian Y, Simard JF. Do death certificates underestimate the burden of rare diseases? The example of systemic lupus erythematosus mortality, Sweden, 2001–2013. Public Health Reports. 2018;133:481-8.

35. Murphy G and Isenberg D. Effect of gender on clinical presentation in systemic lupus erythematosus. Rheumatology (Oxford). 2013;52:2108-15.

36. Diep JT and Gorevic PD. Geriatric autoimmune diseases: systemic lupus erythematosus, Sjogren's syndrome, and myositis. Geriatrics. 2005;60:32-8.

37. Prasad M, Hermann J, Gabriel SE, et al. Cardiorheumatology: cardiac involvement in systemic rheumatic disease (review). Nat Rev Cardiol. 2015;12:168–76.

38. Lee KS, Kronbichler A, Eisenhut M, et al. Cardiovascular involvement in systemic rheumatic diseases: an integrated view for the treating physicians (review). Autoimmun Rev. 2018;17:201-14.

39. Thomas G, Mancini J, Jourde-Chiche N, et al. Mortality associated with systemic lupus erythematosus in France assessed by multiple-cause-of-death analysis. Arthritis Rheumatol. 2014;66:2503-11.

40. Flossmann O, Berden A, de Groot K, et al. Long-term patient survival in ANCA-associated vasculitis. Ann Rheum Dis. 2011;70:488-94.

41. Zintzaras E, Voulgarelis M, Moutsopoulos HM. The risk of lymphoma development in autoimmune diseases: a meta-analysis. Arch Intern Med 2005;165:2337-44.

42. Fors Nieves CE, Izmirly PM. Mortality in systemic lupus erythematosus: an updated review (review). Curr Rheumatol Rep. 2016;18:21.

43. Voss EV and Stangel M. Nervous system involvement of connective tissue disease: mechanisms and diagnostic approach (review). Curr Opin Neurol. 2012;25:306-15.

44. Partington R, Helliwell T, Muller S, et al. Comorbidities in polymyalgia rheumatica: a systematic review (review). Arthritis Res Ther. 2018;20:258.

45. Bao YK, Weide LG, Ganesan VC, et al. High prevalence of comorbid autoimmune diseases in adults with type 1 diabetes from the HealthFacts database. J Diabetes. 2019;11:273-9.

46. World Health Organization (WHO). International classification of diseases for mortality and morbidity statistics, 11th Revision. Geneva: World Health Organization, 2018. https://icd.who.int/browse11/l-m/en.

# Supplementary Material

**Supplementary Table S1.** Deaths with a systemic autoimmune disease, by ICD-10 4-position code, the Netherlands 2013–2017.

| Systemic Autoimmune Disease | ICD-10 code | ICD-10 label | Underlying CoD, number | Non-underlying CoD, number | Any-mention CoD, number | Underlying CoD / Any-mention CoD ratio |
|---|---|---|---|---|---|---|
| **Non-Vasculitides** | | | | | | |
| Systemic lupus erythematosus | M32.1 | SLE with organ/system involvement | 77 | 2 | 79 | 0.97 |
| | M32.8 | other forms of SLE | 1 | 1 | 2 | 0.50 |
| | M32.9 | SLE, unspecified | 22 | 219 | 241 | 0.09 |
| Dermatopolymyositis | M33.0 | juvenile dermatomyositis | 0 | 0 | 0 | — |
| | M33.1 | other dermatomyositis | 20 | 17 | 37 | 0.54 |
| | M33.2 | polymyositis | 27 | 28 | 55 | 0.49 |
| | M33.9 | dermatopolymyositis, unspecified | 0 | 0 | 0 | — |
| | M36.0 | dermatopolymyositis in neoplastic disease | 0 | 3 | 3 | 0.00 |
| | M60.8 | other myositis | 42 | 42 | 84 | 0.50 |
| | M60.9 | myositis, unspecified | 11 | 21 | 32 | 0.34 |
| | G72.4 | inflammatory myopathy, not elsewhere classified | 0 | 2 | 2 | 0.00 |
| Sjögren syndrome | M35.0 | sicca syndrome [Sjögren] | 48 | 156 | 204 | 0.24 |

| Systemic Autoimmune Disease | ICD-10 code | ICD-10 label | Underlying CoD, number | Non-underlying CoD, number | Any-mention CoD, number | Underlying CoD / Any-mention CoD ratio |
|---|---|---|---|---|---|---|
| Systemic sclerosis | M34.0 | Progressive SSc | 0 | 0 | 0 | — |
| | M34.1 | CREST syndrome | 41 | 41 | 82 | 0.50 |
| | M34.2 | SSc induced by drugs and chemicals | 0 | 0 | 0 | — |
| | M34.8 | other forms of SSc | 131 | 0 | 131 | 1.00 |
| | M34.9 | SSc, unspecified | 96 | 228 | 324 | 0.30 |
| Mixed connective tissue disease | M35.1 | other overlap syndromes | 17 | 17 | 34 | 0.50 |
| Polymyalgia rheumatica | M35.3 | polymyalgia rheumatica | 81 | 844 | 925 | 0.09 |
| Antiphospholipid syndrome | D68.6 | other thrombophilia, includes: | 24 | 21 | 45 | 0.53 |
| | | – anticardiolipin syndrome | | | | |
| | | – antiphospholipid syndrome | | | | |
| | | – presence of the lupus anticoagulant | | | | |
| Sarcoidosis | D86.0 | sarcoidosis of lung | 16 | 6 | 22 | 0.73 |
| | D86.1 | sarcoidosis of lymph nodes | 0 | 0 | 0 | — |
| | D86.2 | sarcoidosis of lung with sarcoidosis of lymph nodes | 2 | 0 | 2 | 1.00 |
| | D86.3 | sarcoidosis of skin | 1 | 1 | 2 | 0.50 |

| Systemic Autoimmune Disease | ICD-10 code | ICD-10 label | Underlying CoD, number | Non-underlying CoD, number | Any-mention CoD, number | Underlying CoD / Any-mention CoD ratio |
|---|---|---|---|---|---|---|
| | D86.8 | sarcoidosis of other and combined sites, includes: | 17 | 10 | 27 | 0.63 |
| | | – iridocyclitis in sarcoidosis | | | | |
| | | – multiple cranial nerve palsies in sarcoidosis | | | | |
| | | – sarcoid: | | | | |
| | | – arthropathy | | | | |
| | | – myocarditis | | | | |
| | | – myositis | | | | |
| | | – uveoparotid fever [Heerfordt] | | | | |
| | D86.9 | sarcoidosis, unspecified | 165 | 289 | 454 | 0.36 |
| Amyloidosis | E85.0 | non-neuropathic heredofamilial amyloidosis, includes: | 0 | 2 | 2 | 0.00 |
| | | – familial Mediterranean fever | | | | |
| | | – hereditary amyloid nephropathy | | | | |
| Still disease | M06.1 | adult-onset Still disease | 2 | 2 | 4 | 0.50 |
| | M08.2 | juvenile arthritis with systemic onset | 0 | 2 | 2 | 0.00 |
| Relapsing polychondritis | M94.1 | relapsing polychondritis | 0 | 0 | 0 | — |
| IgG4-related disease | D89.9 | disorder involving the immune mechanism, unspecified | 18 | 83 | 101 | 0.18 |

| Systemic Autoimmune Disease | ICD-10 code | ICD-10 label | Underlying CoD, number | Non-underlying CoD, number | Any-mention CoD, number | Underlying CoD / Any-mention CoD ratio |
|---|---|---|---|---|---|---|
| Cogan's disease | H16.3 | interstitial and deep keratitis | 0 | 0 | 0 | — |
| | H51.8 | other specified disorders of binocular movement | 0 | 3 | 3 | 0.00 |
| Other specified systemic involvement of connective tissue | M35.8 | | 17 | 2 | 19 | 0.89 |
| Systemic involvement of connective tissue, unspecified | M35.9 | | 66 | 75 | 141 | 0.47 |
| **Vasculitides** | | | | | | |
| Antineutrophil cytoplasmic antibody (ANCA)- associated vasculitis | M30.1 | polyarteritis with lung involvement [Churg-Strauss], includes: allergic granulomatous angiitis | 16 | 14 | 30 | 0.5 |
| | M31.3 | Wegener granulomatosis | 188 | 129 | 317 | 0.6 |
| | M31.7 | microscopic polyangiitis | 1 | 1 | 2 | 0.5 |
| Giant cell arteritis | M31.5 | giant cell arteritis with polymyalgia rheumatica | 0 | 0 | 0 | — |
| | M31.6 | other giant cell arteritis | 52 | 105 | 157 | 0.3 |
| Other systemic vasculitis | M30.0 | polyarteritis nodosa | 9 | 13 | 22 | 0.4 |
| | M30.3 | mucocutaneous lymph node syndrome [Kawasaki] | 0 | 1 | 1 | 0.0 |
| | M30.8 | other conditions related to polyarteritis nodosa | 3 | 1 | 4 | 0.8 |

| Systemic Autoimmune Disease | ICD-10 code | ICD-10 label | Underlying CoD, number | Non-underlying CoD, number | Any-mention CoD, number | Underlying CoD / Any-mention CoD ratio |
|---|---|---|---|---|---|---|
| | M31.4 | aortic arch syndrome [Takayasu] | 2 | 5 | 7 | 0.3 |
| | D69.0 | allergic purpura, includes: | 5 | 16 | 21 | 0.2 |
| | | – porpura | | | | |
| | | – anaphylactoid | | | | |
| | | – Henoch-Schönlein | | | | |
| | | – nonthrombocytopenic (haemorrhagic/ idiopathic) | | | | |
| | | – vascular | | | | |
| | | – vasculitis, allergic | | | | |
| | M31.0 | hypersensitivity angiitis [Goodpasture syndrome] | 6 | 3 | 9 | 0.7 |
| | M35.2 | Behcet's disease | 5 | 2 | 7 | 0.7 |

**Supplementary Table S2:** List of comorbidities used in the study (SAIDs are excluded).

| Level | Code | Description | ICD-10 code |
|---|---|---|---|
| 1 | 1. | Infectious and parasitic diseases | A00-B99 |
| 1 | 2. | Neoplasms | C00-D48 |
| 2 | 2.1 | Malignant neoplasms | C00-C97 |
| 3 | 2.1.1 | Solid malignant neoplasms | C00-C80, C97 |
| 3 | 2.1.2 | Malignant neoplasms of lymphoid, haematopoietic and related tissue | C81-C96 |
| 2 | 2.2 | Non-malignant neoplasms (in situ, benign and uncertain) | D00-D48 |
| 1 | 3. | Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism | D50-D89 |
| 1 | 4. | Endocrine, nutritional and metabolic diseases | E00-E89 |
| 1 | 5. | Mental and behavioural disorders | F01-F99 |
| 1 | 6. | Diseases of the nervous system and the sense organs | G00-H95 |
| 1 | 7. | Diseases of the circulatory system | I00-I99 |
| 2 | 7.1 | Ischaemic heart diseases | I20-I25 |
| 3 | 7.1.1 | Acute myocardial infarction | I21-I22 |
| 3 | 7.1.2 | Other ischaemic heart diseases | I20, I23-I25 |
| 2 | 7.2 | Other heart diseases | I30-I51 |
| 2 | 7.3 | Cerebrovascular diseases | I60-I69 |
| 2 | 7.4 | Other diseases of the circulatory system | I00-I15, I26-I28, I70-I99 |
| 1 | 8. | Diseases of the respiratory system | J00-J99 |
| 2 | 8.1 | Influenza | J09-J11 |
| 2 | 8.2 | Pneumonia | J12-J18 |
| 2 | 8.3 | Chronic obstructive pulmonary disease | J40-J44, J47 |
| 2 | 8.4 | Other diseases of the respiratory system | J00-J06, J20-J39, J45-J46, J60-J99 |
| 1 | 9. | Diseases of the digestive system | K00-K92 |
| 1 | 10. | Diseases of the skin and subcutaneous tissue | L00-L99 |
| 1 | 11. | Diseases of the musculoskeletal system | M00-M99 |
| 1 | 12. | Diseases of the genitourinary system | N00-N99 |
| 2 | 12.1 | Diseases of the kidney | N00-N19 |
| 2 | 12.2 | Urinary tract infections | N30.0, N30.9, N34.1, N34.2, N39.0 |
| 2 | 12.3 | Other diseases of the genitourinary system | N20-N29, N30.1-N30.8, N31-N33, N34.0, N34.3, N35-N37, N39.1-N39.9, N40-N99 |
| 1 | 13 | Symptoms, signs, ill-defined causes | R00-R99 |
| 1 | 14 | External causes of morbidity and mortality | V01-Y89 |
| 1 | 15 | Other diseases * | O00-O99, P00-P96, Q00-Q99 |

* Other diseases: Complications of pregnancy, childbirth and puerperium, Certain conditions originating in the perinatal period, Congenital malformations and chromosomal abnormalities

**Supplementary Table S3.** Deaths with non-vasculitides and vasculitides as cause of death, by age and sex.

| | Non-vasculitides | | | | | | Vasculitides | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Underlying CoD | | Non-underlying CoD | | Any-mention CoD | | Underlying CoD | | Non-underlying CoD | | Any-mention CoD | |
| | number | annual rate per 1,000,000 population | number | annual rate per 1,000,000 population | number | annual rate per 1,000,000 population | number | annual rate per 1,000,000 population | number | annual rate per 1,000,000 population | number | annual rate per 1,000,000 population |
| **Females** | | | | | | | | | | | | |
| <40 | 17 | 0.9 | 24 | 1.2 | 34 | 1.8 | 0 | 0.0 | 1 | 0.1 | 1 | 0.1 |
| 40–59 | 94 | 7.7 | 106 | 8.7 | 170 | 14.0 | 5 | 0.4 | 4 | 0.3 | 9 | 0.7 |
| 60–79 | 274 | 31.9 | 445 | 51.8 | 636 | 74.1 | 43 | 5.0 | 50 | 5.8 | 92 | 10.7 |
| ≥80 | 200 | 85.1 | 744 | 316.5 | 909 | 386.7 | 70 | 29.8 | 94 | 40.0 | 164 | 69.8 |
| Total | 585 | 13.8 | 1,319 | 31.2 | 1,749 | 41.3 | 118 | 2.8 | 149 | 3.5 | 266 | 6.3 |
| **Males** | | | | | | | | | | | | |
| <40 | 16 | 0.8 | 17 | 0.9 | 31 | 1.6 | 0 | 0.0 | 1 | 0.1 | 1 | 0.1 |
| 40–59 | 68 | 5.6 | 88 | 7.2 | 139 | 11.4 | 12 | 1.0 | 9 | 0.7 | 21 | 1.7 |
| 60–79 | 201 | 24.6 | 355 | 43.4 | 507 | 61.9 | 77 | 9.4 | 76 | 9.3 | 150 | 18.3 |
| ≥80 | 72 | 53.1 | 292 | 215.5 | 354 | 261.2 | 80 | 59.0 | 53 | 39.1 | 132 | 97.4 |
| Total | 357 | 8.6 | 752 | 18.1 | 1,031 | 24.8 | 169 | 4.1 | 139 | 3.3 | 304 | 7.3 |
| **Females and Males** | | | | | | | | | | | | |
| <40 | 33 | 0.8 | 41 | 1.0 | 65 | 1.7 | 0 | 0.0 | 2 | 0.1 | 2 | 0.1 |
| 40–59 | 162 | 6.7 | 194 | 8.0 | 309 | 12.7 | 17 | 0.7 | 13 | 0.5 | 30 | 1.2 |
| 60–79 | 475 | 28.3 | 800 | 47.7 | 1,143 | 68.1 | 120 | 7.2 | 126 | 7.5 | 242 | 14.4 |
| ≥80 | 272 | 73.4 | 1,036 | 279.5 | 1,263 | 340.8 | 150 | 40.5 | 147 | 39.7 | 296 | 79.9 |
| Total | 942 | 11.2 | 2,071 | 24.7 | 2,780 | 33.1 | 287 | 3.4 | 288 | 3.4 | 570 | 6.8 |

CoD: Cause of Death

**Supplementary Table S4.** Deaths with SAID as non-underlying (underlying) CoD and circulatory disease as underlying (non-underlying) CoD.

| SAID main subclasses | SAID as Underlying CoD – Circulatory disease as Non-underlying CoD | | | SAID as Non-underlying CoD – Circulatory disease as Underlying CoD | | |
|---|---|---|---|---|---|---|
| | number | % of the SAID subclass | Age-sex-standardized observed/expected (O/E) ratios [95%CI] | number | % of the SAID subclass | Age-sex-standardized observed/expected (O/E) ratios [95%CI] |
| Systemic lupus erythematosus | 53 | 53.0 | 2.04 [1.53 ; 2.67] | 39 | 17.6 | 0.86 [0.61 ; 1.17] |
| Dermatopolymyositis | 30 | 30.0 | 1.11 [0.75 ; 1.58] | 35 | 31.3 | 1.21 [0.84 ; 1.68] |
| Sjögren syndrome | 19 | 39.6 | 1.36 [0.82 ; 2.12] | 41 | 26.3 | 0.97 [0.70 ; 1.32] |
| Systemic sclerosis | 156 | 58.2 | 2.09 [1.77 ; 2.44] | 44 | 16.6 | 0.75 [0.54 ; 1.00] |
| Polymyalgia rheumatica | 35 | 43.2 | 1.31 [0.91 ; 1.82] | 275 | 32.6 | 1.06 [0.94 ; 1.19] |
| Sarcoidosis | 90 | 44.8 | 1.54 [1.24 ; 1.90] | 86 | 28.3 | 1.17 [0.93 ; 1.44] |
| ANCA-associated vasculitis | 75 | 36.6 | 1.06 [0.83 ; 1.33] | 37 | 25.7 | 1.01 [0.71 ; 1.39] |
| Giant cell arteritis | 20 | 38.5 | 1.09 [0.67 ; 1.69] | 38 | 36.2 | 1.20 [0.85 ; 1.65] |
| Other SAID | 112 | 64.4 | 1.98 [1.61 ; 2.40] | 42 | 16.9 | 0.76 [0.55 ; 1.03] |

SAID: systemic autoimmune disease; CoD: Cause of Death

# Part III

## Enrichment of Cause-of-Death data with healthcare registries to assess disease importance

# Chapter 5

## Prevalence of diabetes mellitus at the end of life: an investigation using individually linked Cause-of-Death and healthcare register data

**Marianna Mitratza**, Anton E. Kunst, Peter P.M. Harteloh, Markus M.J. Nielen, Bart Klijs

*Published in:*
*Diabetes Research and Clinical Practice 2020;160(108003).*

# Abstract

**Background:** Although diabetes mellitus at the end of life is associated with complex care, its end-of-life prevalence is uncertain. Our aim is to estimate diabetes prevalence in the end-of-life population, to evaluate which medical register has the largest added value to cause-of-death data in detecting diabetes cases, and to assess the extent to which reporting of diabetes as a cause of death is associated with disease severity.

**Methods:** Our study population consisted of deaths in the Netherlands (2015–2016) included in Nivel Primary Care Database (Nivel-PCD; N = 18,162). The proportion of deaths with diabetes (Type 1 or 2) within the last two years of life was calculated using individually linked cause-of-death, general practice, medication, and hospital discharge data. Severity status of diabetes was defined with dispensed medicines.

**Results:** According to all data sources combined, 28.7% of the study population had diabetes at the end of life. The estimated end-of-life prevalence of diabetes was 7.7% using multiple cause-of-death data only. Addition of general practice data increased this estimate the most (19.7%-points). Of the cases added by primary care data, 76.3% had a severe or intermediate status.

**Conclusions:** More than one fourth of the Dutch end-of-life population has diabetes. Cause-of-death data are insufficient to monitor this prevalence, even of severe cases of diabetes, but could be enriched particularly with general practice data.

# Introduction

The end phase of life is characterized by accumulation of health problems, high health care costs, and a need for tailored medical treatment [1-3]. In the coming decades, the number of deaths, and thus the number of persons in the last phase of their life is expected to steadily increase [4, 5]. The implications of this increase for the provision of health care and community services are not well understood, particularly because it is more common to study the prevalence of diseases in relation to age than in relation to the end of life [4]. For diabetes, it is unknown what proportion of the end-of-life population has the disease and how many patients have a mild, intermediate or severe form of the disease. Estimates of the proportion with diabetes at the end of life are essentially synonymous with the lifetime prevalence of diabetes, because diabetes is a chronic disease with a practically irreversible course [6].

Studies investigating causes of death give a first impression of the prevalence of diabetes at the end of life. These studies show that 5–10% of all deaths in the community have diabetes recorded on their death certificate [7, 8]. However, these percentages are likely an underestimation of the prevalence of diabetes at the end of life, because the cause of death concept is developed to describe the frequency of diseases that are died from rather than with. Studies investigating causes of death in cohorts of patients with diabetes show that only 35–58% of all persons who die with diabetes have it recorded as a cause of death [7, 9-14].

To achieve a more complete picture of all persons who have diabetes at the end of life, the cause-of-death registration can be linked with other medical registrations. A review of the literature has concluded that administrative databases may be adequately sensitive and highly specific to identify diabetes [14]. Linkage studies have shown that 75–91% of all persons who have diabetes are recorded as such in the primary care registration [15, 16]. Another study found that using combined information from national lists of drug dispensing, outpatient attendance, laboratory tests (HbA1c) and hospital diagnoses performs well to detect persons with diabetes [17]. In the Netherlands, various medical registrations are available containing information on the presence of diabetes in individuals, including registrations covering general practice, hospital discharges and administered medication.

It is expected that the cause-of-death registration is more likely to have patients with more severe levels of diabetes reported, as its purpose is focused on the causal role of the disease to death. Previous studies have suggested that insulin-treated patients with

diabetes are more likely to have diabetes mentioned on their death certificate [7, 9, 10], while those treated with lifestyle interventions may have this information missing [18]. It becomes, therefore, relevant to study whether enrichment of mortality records with administrative data allows to capture the less severe cases of diabetes.

The aim of our study was to estimate the prevalence of diabetes mellitus at the end of life using cause-of-death data individually linked with data from the Dutch general practice, medication and hospital registrations. We assessed which of these medical registrations has the largest added value to using cause- of-death data in detecting people dying with diabetes. Furthermore, we investigated the distribution according to the diabetes severity status of cases that could not be identified using cause-of-death data.

# Materials and Methods

## Study population

Our study population (N = 18,162) consisted of all deaths in the Netherlands between 1 January 2015 and 31 December 2016 among persons included in the general practitioner registration of the Nivel Primary Care Database (Nivel-PCD) 2015–2016. Nivel-PCD contains routinely recorded data from a sample of 435 general practices in the Netherlands and covers approximately 10% of the Dutch population. The age and sex distribution of the registered patients is representative of the general Dutch population [19]. Nivel-PCD does not include patients living in nursing homes, as they receive general health care by the institutional physician.

This study has been approved according to the governance code of the Nivel Primary Care Database, under number NZR-00318.033. Dutch law allows the use of electronic health records without informed consent of patients and without approval by a medical ethics committee for observational studies, like ours, that do not use directly identifiable data (Dutch Civil Law, Article 7:458).

## Prevalence of diabetes according to different data sources

For each person in our study population the presence of Type 1 and 2 diabetes mellitus at the end of life was assessed according to the cause-of-death registration, and according to information from three medical registrations, which were the Nivel-PCD (general practice),

the hospital discharge register (main and secondary diagnoses) and the dispensed medication database. To identify the presence of diabetes at the end of life according to each medical registration, a time frame of exactly two years prior to death was chosen, based on the date of death. This timeframe was motivated by the intention to enable a relatively large period to increase the probability of detecting persons with diabetes, but at the same time remain close to the time of death. The information from the different data sources was linked at the level of individuals, using a unique person identifier.

## Cause of death

The Cause-of-Death registration covers the total population of the Netherlands. Within the Dutch system, physicians are obliged to report the cause of death to the civil register of the municipality where the person died. Causes of death are coded at Statistics Netherlands using the 10th revision of the International Classification of Diseases (ICD-10) classification of the WHO and include both the Underlying Cause of Death (CoD) – i.e. the disease or injury initiating the chain of morbid events leading directly to death [20] – and the non-underlying – i.e. intermediate or contributory – CoD. A CoD is commonly termed as a "Multiple CoD" when it is reported either as the underlying or a non-underlying cause of death on the death certificate. In the Netherlands Iris 4.4.1 software is adopted for the automated coding in accordance with the ICD-10 rules for coding and selecting an underlying cause of death and with the WHO ICD-10 updates for the years 2015 and 2016 [21]. Date of birth, date of death, sex and mention of diabetes mellitus (ICD-10 codes E10-E14) as the Underlying CoD and as a Multiple CoD were extracted from this data source for the years 2015 and 2016.

## General practice

In general practice (GP), all consultations and procedures for a single health problem are clustered into an episode of care that also incorporates diagnoses of chronic diseases in previous years. An algorithm is used to construct episodes of illness that include an estimated date of diagnosis and, for non-chronic diseases, a date of recovery [22]. Diagnoses in Nivel-PCD are coded according to the International Classification of Primary Care version 1 (ICPC-1). We used data for 2015 and 2016 and the presence of diabetes was assessed using the ICPC-1 code T90, which includes both Type 1 and 2 diabetes mellitus.

## Hospital discharges

The Dutch Hospital Discharge Register (owned by Dutch Hospital Data) includes hospital discharges of all general and academic hospitals and two short-stay hospitals in the Netherlands. Overnight stays and day cases are included. For each stay, one primary and one or more secondary diagnoses (optionally added) of discharge are available, coded according to the ICD-10 classification. We used data for 2013–2016 to extract all primary and secondary diagnoses of diabetes mellitus (ICD-10 codes E10-E14) for admissions within two years prior to date of death.

## Dispensed medicines

The Dispensed Medicines database (owned by the National Health Care Institute) contains information of all dispensed medication reimbursed under the statutory basic medical insurance. Medication dispensed in residential homes for the elderly are included, but medicines provided in hospitals or nursing homes are not. The presence of diabetes within two years prior to death was established on the basis of prescriptions of Insulin and analogues (ATC-4: A10A) and Blood glucose lowering drugs, excluding insulins (ATC-4: A10B) using data for 2013–2016. Dispensed medicines for diabetes were also used as an indicator of the severity of diabetes at the end of life. In accordance with guidelines of the International Diabetes Federation (IDF) [23] the following categories were distinguished for the severity of diabetes [24, 25]: mild: managed only with lifestyle changes (no antidiabetic medication dispensed); intermediate: single or multiple oral antidiabetic agents dispensed only; and severe: insulin prescribed (with or without oral antidiabetic agents).

## Statistical analysis

We first present the number of persons and the number of deaths in Nivel-PCD, as well as the number of hospitalizations prior to death. The absolute number and the proportion of persons with diabetes at the end of life was calculated using each data source in isolation and in combination. We calculated the increase in the number and proportion of deaths with diabetes when information from Nivel-PCD, the Hospital Discharge Register and the Dispensed Medicines Database was used in addition to information from the CoD registration. The overlap of diabetes diagnoses in the different data sources was visualized in a Venn-diagram. The number and proportion of persons with diabetes according to

severity levels was calculated for those with diabetes as a cause of death, those with diabetes as a cause of death or GP diagnosis, and those with diabetes as a GP diagnosis only. All results are presented as a total and by age and sex.

In order to evaluate the representativeness of our study sample, from which we had to exclude the institutionalised, we compared the prevalence of diabetes in our study population with the prevalence in the total Dutch (deceased) population. For this comparison we used all data sources with coverage of the total Dutch population (causes of death, and hospital- and medication register, but not Nivel-PCD). The results for the total Dutch population are also presented according to living situation (living in an institution or not) which was measured using data from the Household Database of Statistics Netherlands. Based on the length of stay in an institution during the last two years of life, we categorized persons as institutionalized exclusively, partially or not at all.

All analyses were performed using the R software (The R Foundation for Statistical Computing, version 3.2.3).

According to the Dutch law on medical research, no ethical approval was required for this study, as no living subjects were involved and all data were anonymized.

## Results

**Table 1** presents the characteristics of the study population, which constituted of 18,162 patients registered in Nivel-PCD 2015–2016 who died in 2015-2016. This group of deceased patients was 1.4% of the total number of registered patients. Of all deaths in our study population, 79.0% had one or more hospitalizations in the last two years of their life.

The end-of-life prevalence of diabetes according to each data source in isolation and for all registrations combined is presented in **Table 2**. Of all deaths, 2.2% had diabetes recorded as the Underlying CoD and 7.7% as a Multiple CoD. Nivel-PCD could identify most of the cases identified with diabetes at the end of life (27.1%), followed by the medication registration (22.4%) and the hospital-any diagnosis registration (17.1%). The lowest number was achieved by the hospital-main diagnosis registration (1.1%). According to all data sources combined, 28.7% of all deaths had diabetes at the end of life. Diabetes is mentioned on the death certificates in 28% of these cases (7.7/28). In 8% of these cases, diabetes is the underlying cause of death (2.2/28).

**Table 1.** Description of the study population by age and sex.

| | Source of the study population: Nivel-Primary Care Database 2015–2016 n | Study population: Deaths in Nivel-Primary Care Database 2015–2016 n (%) | One or more hospitalizations in the study population* n (%) |
|---|---|---|---|
| **Age, years** | | | |
| <50 | 817,139 | 660 (0.1) | 473 (71.7) |
| 50–60 | 191,387 | 1,205 (0.6) | 984 (81.7) |
| 60–70 | 162,429 | 2,692 (1.7) | 2,314 (86.0) |
| 70–80 | 99,193 | 4,254 (4.3) | 3,688 (86.7) |
| 80–90 | 49,773 | 6,230 (12.5) | 4,958 (79.6) |
| >90 | 9,295 | 3,131 (33.7) | 1,924 (61.5) |
| **Sex** | | | |
| female | 672,233 | 8,809 (1.3) | 6,710 (76.2) |
| male | 656,983 | 9,353 (1.4) | 7,631 (81.6) |
| **Total** | 1,329,317 | 18,162 (1.4) | 14,341 (79.0) |

* hospitalizations refer to a hospital admission for any cause during the last 2 years of life

There was an increasing trend with age, with the highest prevalence estimates for the 70–80 years old, in the proportion of deceased who had diabetes according to the Nivel-PCD, the medication and the hospital-any diagnosis registrations. After 80 years old, these percentages declined. The cause of death registration showed a peak of deaths with diabetes among those who died at 80–90 years. For the younger group (<50 years), the medication registration had the highest percentage. The prevalence of diabetes was similar for both sexes according to each data source.

**Table 2:** Prevalence of diabetes in the last two years of life according to different data sources among persons who were part of Nivel-Primary Care Database and died in 2015–2016 (n = 18,162).

| | Number and proportion (%) of deaths with diabetes | | | | | | |
|---|---|---|---|---|---|---|---|
| | Cause of Death registration | | Nivel-Primary Care Database | Hospital Discharge registration | | Dispensed Medicines registration | All registrations combined |
| | Underlying CoD | Multiple CoD | | Main diagnosis | Any diagnosis | | |
| **Age, years** | | | | | | | |
| <50 | 13 (2.0) | 24 (3.6) | 53 (8.0) | 8 (1.2) | 33 (5.0) | 55 (8.3) | 62 (9.4) |
| 50–60 | 18 (1.5) | 54 (4.5) | 184 (15.3) | 18 (1.5) | 124 (10.3) | 165 (13.7) | 209 (17.3) |
| 60–70 | 39 (1.4) | 150 (5.6) | 626 (23.3) | 33 (1.2) | 441 (16.4) | 570 (21.2) | 695 (25.8) |
| 70–80 | 93 (2.2) | 358 (8.4) | 1,373 (32.3) | 66 (1.6) | 945 (22.2) | 1,178 (27.7) | 1,434 (33.7) |
| 80–90 | 170 (2.7) | 585 (9.4) | 1,951 (31.3) | 58 (0.9) | 1,232 (19.8) | 1,556 (25.0) | 2,052 (32.9) |
| >90 | 68 (2.2) | 230 (7.3) | 730 (23.3) | 16 (0.5) | 334 (10.7) | 537 (17.2) | 762 (24.3) |
| **Sex** | | | | | | | |
| female | 188 (2.1) | 700 (7.9) | 2,368 (26.9) | 82 (0.9) | 1,435 (16.3) | 1,901 (21.6) | 2,511 (28.5) |
| male | 213 (2.3) | 701 (7.5) | 2,549 (27.3) | 117 (1.3) | 1,674 (17.9) | 2,160 (23.1) | 2,703 (28.9) |
| **Total** | 401 (2.2) | 1,401 (7.7) | 4,917 (27.1) | 199 (1.1) | 3,109 (17.1) | 4,061 (22.4) | 5,214 (28.7) |

CoD: Cause of Death

In **Table 3**, we show how many new cases of diabetes are identified when information of each medical registration is used in addition to the cause of death registration. Adding information of Nivel-PCD to the cause of death registration resulted in the highest increase in the observed diabetes prevalence (19.7% points). The medication registration followed (15.7%), while fewer new cases could be identified with hospital discharge data (12.6% for the any-diagnosis registration and 0.6% for the main diagnosis registration).

**Table 3.** Increase in the number and proportion (%-points) of deaths with diabetes in the last two years of life when combining multiple causes of death with medical registration data compared to using multiple causes of death in isolation.

| | Increase in the number and proportion (%-points) of deaths with diabetes by data source when combined with Multiple CoD | | | |
|---|---|---|---|---|
| | Nivel-Primary Care Database | Hospital Discharge registration | | Dispensed Medicines registration |
| | | Main diagnosis | Any diagnosis | |
| **Age, years** | | | | |
| <50 | 31 (4.7) | 4 (0.6) | 16 (2.4) | 35 (5.3) |
| 50–60 | 134 (11.1) | 11 (0.9) | 90 (7.5) | 117 (9.7) |
| 60–70 | 485 (18.0) | 25 (0.9) | 349 (13.0) | 439 (16.3) |
| 70–80 | 1,026 (24.1) | 41 (1.0) | 709 (16.7) | 852 (20.0) |
| 80–90 | 1,396 (22.4) | 30 (0.5) | 893 (14.3) | 1058 (17.0) |
| >90 | 505 (16.1) | 7 (0.2) | 224 (7.2) | 352 (11.2) |
| **Sex** | | | | |
| female | 1,698 (19.3) | 49 (0.6) | 1036 (11.8) | 1305 (14.8) |
| male | 1,879 (20.1) | 69 (0.7) | 1245 (13.3) | 1548 (16.6) |
| **Total** | 3,577 (19.7) | 118 (0.6) | 2281 (12.6) | 2853 (15.7) |

CoD: Cause of Death

The overlap between the different registrations in detecting people who died with diabetes is displayed in **Figure 1**. This Venn diagram shows the absolute number of patients with diabetes identified in multiple registers (in overlapping areas), and the number of those identified in only unique cases (in peripheral areas). Nivel-PCD identifies the most unique cases (610).

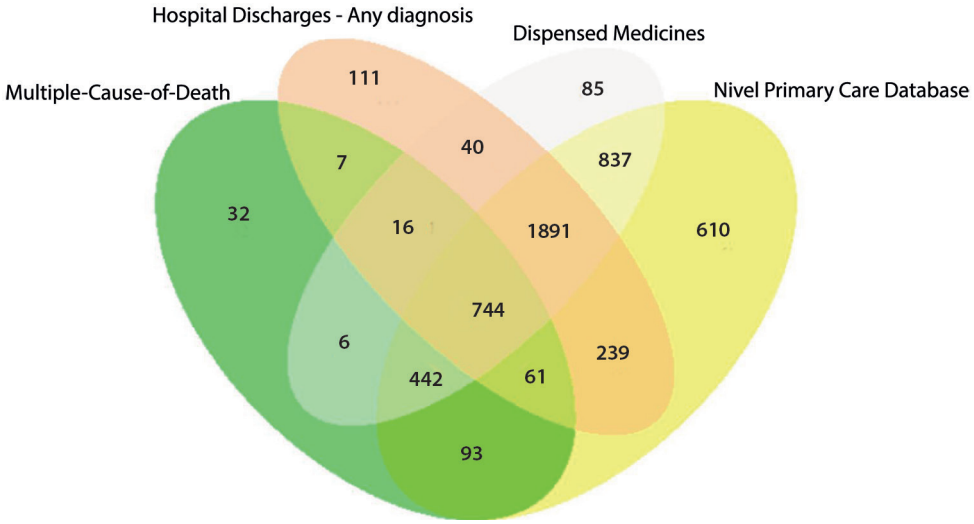


**Figure 1.** Prevalence of the end-of-life diabetes in absolute numbers using combinations of registers.

The distribution of diabetes at the end of life by severity level, as measured according to dispensed medication data, is provided in **Table 4**. According to the Multiple CoD registration enriched with Nivel-PCD records, 41% of the deceased had intermediate status (only oral antidiabetic medication) and 38% had severe diabetes (insulin prescribed), compared to only 21% being mild cases (no medication). Severe cases were more likely to be reported as the Underlying CoD and the less severe as non-underlying CoD. One third of the new cases added from Nivel-PCD (33.5%) were severe, while 42.8% were of intermediate and 23.7% were of mild severity. The distribution according to diabetes severity was similar for both sexes. The proportion of severe cases declined with increasing age (**Figure 2**). The highest decline was observed for the new Nivel-PCD cases.

**Table 4.** Prevalence of diabetes in the last two years of life according to different data sources among persons who were part of the Nivel-Primary Care Database and died in 2015–2016, stratified by severity using dispensed medicines data.

| | Number and proportion (%) of deaths with diabetes stratified by severity | | |
|---|---|---|---|
| | **Mild** | **Intermediate** | **Severe** |
| **Multiple CoD** | | | |
| **Sex** | | | |
| female | 104 (14.8) | 249 (35.6) | 347 (49.6) |
| male | 89 (12.7) | 263 (37.5) | 349 (49.8) |
| **Total** | 193 (13.8) | 512 (36.5) | 696 (49.7) |
| **Cases added by Nivel-Primary Care Database*** | | | |
| **Sex** | | | |
| female | 449 (26.4) | 708 (41.7) | 541 (31.9) |
| male | 400 (21.3) | 823 (43.8) | 656 (34.9) |
| **Total** | 849 (23.7) | 1,531 (42.8) | 1,197 (33.5) |
| **Multiple COD enriched with Nivel-Primary Care Database** | | | |
| **Sex** | | | |
| female | 553 (23.1) | 957 (39.9) | 888 (37.0) |
| male | 489 (19.0) | 1,086 (42.1) | 1,005 (38.9) |
| **Total** | 1,042 (20.9) | 2,043 (41.0) | 1,893 (38.0) |

* Cases of diabetes recorded in the Nivel-Primary Care Database, but not in the multiple CoD registration

The representativeness of our sample of the noninstitutionalised population for the general (deceased) Dutch population is evaluated in **Supplementary Table S1**. The proportion of deaths with diabetes in our sample (25.3%) was similar to that in the general Dutch population (23.5%), if diabetes is identified using the same set of sources – i.e. Cause of death, Hospital Discharges and Medication registration combined.

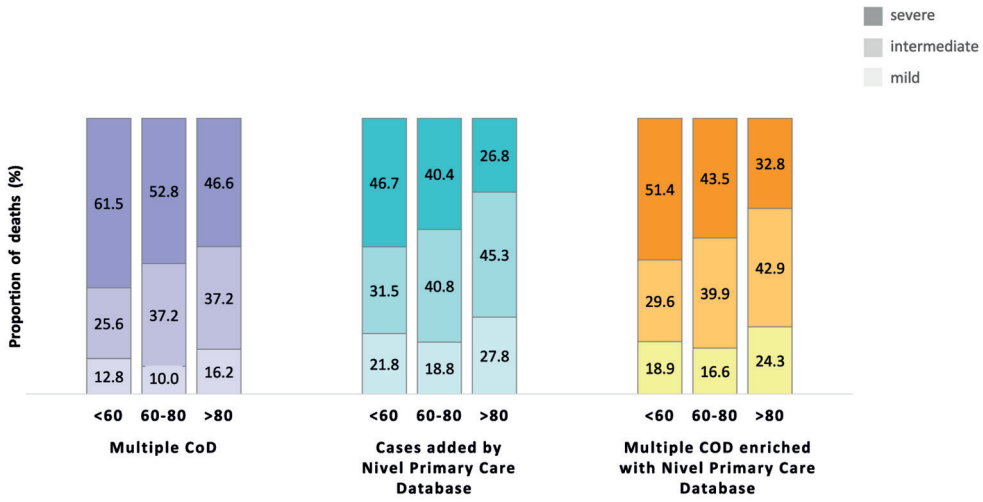


**Figure 2.** Proportion of deaths with diabetes at the end of life, stratified by severity.

# Discussion

## Summary of findings

This study estimated the prevalence of diabetes mellitus at the end of life in the Netherlands using multiple administrative data sources. Using all data sources combined, 28.7% of the study population was found to have diabetes at the end of life. Our findings suggest that only a minority of those who died with diabetes have this condition mentioned on their death certificate. By using primary care data to enrich cause of death data, we identified more additional persons with diabetes as compared to using medication or hospital discharge data. These added cases are distributed over all severity levels.

## Comparison with existing literature

The prevalence of diabetes mellitus in the elderly as observed in this study is higher than in previous Dutch studies, which found a prevalence of only 17–21% for the age-group 75–84 [26, 27]. However, we examined the end-of-life prevalence of diabetes, which can be interpreted as the lifetime prevalence, because diabetes is largely irreversible. This lifetime prevalence is expected to be higher than the point prevalence of diabetes in the population. Indeed, the prevalence of diabetes at the time of death was as high as 38%

according to a Canadian study investigating the accumulation of chronic conditions at the end of life based on real administrative data from the general population [1] or an Australian study estimating the lifetime prevalence based on multi-state models [28].

We found similar end-of-life prevalence rates for women and men, with a slightly higher percentage in men (28.9% vs 28.5%). This is in line with some previous studies [26, 29, 30]. However, other studies found a higher burden of diabetes for men in terms of prevalence and incidence [28, 31], but a higher burden of diabetes for women in terms of mortality [28]. This sex difference may be due to differential survival, as females with diabetes have a greater risk of mortality than their male counterparts, partially due to a greater impact of cardiovascular disease [32].

Our observation of a declining end-of-life diabetes prevalence after the age of 80 years is in line with previous evidence [27, 31]. This may be explained by the reduced survival of patients with diabetes, because of uncontrolled diabetes, multiple complications and disability [33], which result in fewer patients surviving until 80 years of age [34]. Of note, with increasing age, we found a gradual shift towards lower levels of severity. This may be due in part to the predominance of Type 1 diabetes among younger patients [35], which is treated with insulin, and therefore always classified as a severe form in our study.

The fact that we detected more diabetes cases using general practice data in comparison with other administrative data is consistent with previous studies [15, 36, 37]. Many cases were found only in general practice registration probably because of the gate-keeping system in the Netherlands, with general practitioners being the first point of contact of any patient within healthcare. Moreover, the Dutch general practice registrations are cumulative, covering diagnoses made across all previous years [22], which is not the case with cause-of-death data, hospital discharge data or dispensed medicines data.

The few cases of diabetes that were found only in the cause-of-death registration, were institutionalized in 22 out of 32 cases. The lack of detection by other sources may reflect poorer coverage of these sources for the institutionalized population.

About 21% of the persons with diabetes that we could identify in the primary care or cause-of-death registration did not use any medication. This proportion is in agreement with estimates from a Dutch care utilization study (20.5%) [38] and a survey (25%) [35]. Contrary to expectations, many of the severe cases were not identified by the cause-of-death registration, even though diabetes may have contributed to the death of the patient. This finding may reflect the subjective judgement of the certifying physician as

to whether diabetes played a role in the death [39]. Diabetes may be ignored especially when a common complication such as cardiovascular disease is absent [9,10].

## Methodological considerations

In our study, we have a good representation of the general Dutch population, thus avoiding selection bias which might be inherent to population surveys and epidemiologic follow-up studies subject to selective non-response and attrition. Furthermore, we used individually linked administrative data from various health care registries, which enhanced the completeness of our estimate of the end-of-life prevalence of diabetes. Our method is potentially generalizable to other countries.

The present study has some limitations. Given that the study population does not include institutionalized people, who have in principle a higher prevalence of diabetes [40, 41], it is expected that we missed some cases with diabetes. However, the fact that the general practice data are cumulative may compensate to an extent this missing information. Moreover, in sensitivity analysis we found a similar prevalence rate for our non-institutionalized study population and the general population of Dutch residents in their last years of life.

Another potential methodological limitation of the study regards the use of dispensed medicines as an indicator of diabetes severity among those approaching death. When life expectancy is limited, the medication may be less intensified in order to balance out the benefit-harm outcomes [42]. Therefore, our approach may have underestimated the proportion of the study population with severe diabetes. One or more additional indicators of disease severity, such as diabetic complications or duration of diabetes, may be more accurate in these cases. Unfortunately, these indicators were not available in the data sources. Finally, the general practice data did not allow for the distinction between Type 1 and 2 diabetes. Of note, 84% of the cases with diabetes have been found to be of Type 2 diabetes in the Netherlands [35].

## Implications of findings

This study has highlighted the high burden of diabetes mellitus at the end of life. The end-of-life prevalence of diabetes can be interpreted as the lifetime prevalence, which may be a more intuitive risk estimate to communicate to the general public for awareness. This

is important, considering that a substantial proportion of persons living with diabetes remains undiagnosed [43]. If diabetes is prevented, or its progression in individuals is slowed down, the end-of-life burden will be reduced.

It is important for health care provision and planning to know how many people have diabetes at the end of life. These people have higher needs for intensive and complex care, hospitalizations, and polypharmacy at the end of life [44, 45], which may be prolonged [46]. For health care policies to respond to these needs, it is important to forecast the number of people dying with diabetes in the next decades [47].

We showed that combining cause-of-death data with general practice data can provide a database to accurately monitor the end-of-life prevalence of diabetes. According to our findings, cause-of-death data are insufficient to detect the end-of-life population with even severe diabetes. Our approach of combining different sources of administrative data for estimating the end-of-life prevalence of a disease could be followed for other common chronic conditions known to be associated with high disability and complex care at the end of life, such as chronic obstructive pulmonary disease [48], dementia [1], and multimorbidity [49].

# Conclusion

This study using administrative data in the Netherlands showed that more than one fourth of the Dutch population has diabetes at the end of life. Only a minority of even the severe cases with diabetes at the end of life are reported on death certificates. Enrichment of cause-of-death data with general practice data enables the monitoring of the end-of-life prevalence of diabetes in order to raise public awareness and prepare the health care system for diabetes care in an ageing population.

# References


1. Rosella L, Kornas K, Huang A, et al. Accumulation of chronic conditions at the time of death increased in Ontario from 1994 To 2013. Health Affairs (Millwood). 2018;37(3):464-72.
2. Payne S, Hawker S, Kerr C, et al. Experiences of end-of-life care in community hospitals. Health Soc Care Community. 2007;15(5):494-501.
3. Munshi MN, Florez H, Huang ES, et al. Management of diabetes in long-term care and skilled nursing facilities: a position statement of the American Diabetes Association. Diabetes Care. 2016;39(2):308–18.
4. Bone AE, Gomes B, Etkind SN, et al. What is the impact of population ageing on the future provision of end-of-life care? Population-based projections of place of death. Palliat Med. 2018; 32(2):329-36.
5. Statistics Netherlands. Statline: Key figures of various population forecasts and observation. Netherlands: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70737NED/table?ts=1564395 103872
6. Ang GY. Reversibility of diabetes mellitus: narrative review of the evidence. World J Diabetes. 2018;9:127-31.
7. Cheng WS, Wingard DL, Kritz-Silverstein D, Barrett-Connor E. Sensitivity and specificity of death certificates for diabetes: as good as it gets? Diabetes Care. 2008;31(2):279-84.
8. Rodriguez F, Blum MR, Falasinnu T, et al. Diabetes-attributable mortality in the United States from 2003 to 2016 using a multiple-cause-of-death approach. Diabetes Res Clin Pract. 2019;148:169-78.
9. McEwen LN, Kim C, Haan M, et al. Diabetes reporting as a cause of death: results from the Translating Research Into Action for Diabetes (TRIAD) study. Diabetes Care. 2006;29(2):247-53.
10. Evans JM, Barnett KN, McMurdo ME, Morris AD. Reporting of diabetes on death certificates of 1872 people with type 2 diabetes in Tayside, Scotland. Eur J Public Health 2008;18(2):201–3.
11. Andersson DK and Svärdsudd K. The value of death certification statistics in measuring mortality in persons with diabetes. Scand J Prim Health Care. 1994;12(2):114-20.
12. Morgan CL, Currie CJ, Peters JR. Relationship between diabetes and mortality: a population study using record linkage. Diabetes Care. 2000;23(8):1103-7.
13. Raymond NT, Langley JD, Goyder E, et al. Insulin treated diabetes mellitus: causes of death determined from record linkage of population based registers in Leicestershire. UK J Epidemiol Community Health. 1995;49(6):570-4.
14. Saydah SH, Geiss LS, Tierney E, et al. Review of the performance of methods to identify diabetes cases among vital statistics, administrative, and survey data. Ann Epidemiol. 2004;14(7):507-16.
15. Harris SB, Glazier RH, Tompkins JW, et al. Investigating concordance in diabetes diagnosis between primary care charts (electronic medical records) and health administrative data: a retrospective cohort study. BMC Health Serv Res. 2010;10:347.
16. Morris AD, Boyle DI, MacAlpine R, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. BMJ. 1997;315(7107):524-8.

17. Thornley S, Wright C, Marshall R, et al. Can the prevalence of diagnosed diabetes be estimated from linked national health records? The validity of a method applied in New Zealand. J Prim Health Care. 2011;3(4):262-8.

18. Andresen EM, Lee JA, Pecoraro RE, et al. Underreporting of diabetes on death certificates, King County, Washington. Am J Public Health. 1993;83(7):1021-4.

19. de Hoon S, Verberne L, Davids R, et al. Zorg doorde huisarts: Jaarcijfers 2015 entrendcijfers 2011–2015. Netherlands institute for health services research 2016.

20. World Health Organization – WHO. International statistical classification of diseases and related health problems, tenth revision. Volume II. Geneva: World Health Organization; 1993.

21. Harteloh PPM. The implementation of an automated coding system for cause-of-death statistics. Inform Health Soc Care. 2018;1-14.

22. Nielen MMJ, Spronk I, Davids R, et al. Estimating Morbidity Rates Based on Routine Electronic Health Records in Primary Care: Observational Study. JMIR 2019;7(3).

23. Aschner P. New IDF clinical practice recommendations for managing type 2 diabetes in primary care. Diabet Res Clin Pract. 2017;132:169-70.

24. Overbeek JA, Heintjes EM, Prieto-Alhambra D, et al. Type 2 diabetes mellitus treatment patterns across Europe: a population-based multi-database study. Clin Ther. 2017;39:759-70.

25. Bouma M, Rutten GE, de Grauw WJ, et al. Nederlands Huisartsen Genootschap. [Summary of the practice guideline 'Diabetes mellitus type 20 (second revision) from the Dutch college of general practitioners]. Ned Tijdschr Geneeskd. 2006;150(41):2251-6.

26. Overbeek JA, van der Heijden AW, Herings RM, Nijpels G. More than doubling the prevalence of diabetes. Ned Tijdschr Medic. 2017;161:D673.

27. Vektis. Chapter 1. Ouderenzorg anno 2018. In: Zorgthermometer ouderenzorg – Inzicht in de ouderenzorg; 2018: p.15.

28. Magliano DJ, Shaw JE, Shortreed SM, et al. Lifetime risk and projected population prevalence of diabetes. Diabetologia. 2008;51(12):2179-86.

29. National Center for Chronic Disease Prevention and Health Promotion; Division of Diabetes Translation. National Diabetes Statistics Report; 2017. https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetesstatistics-report.pdf.

30. Licher S, Heshmatollah A, van der Willik KD, et al. Lifetime risk and multimorbidity of non-communicable diseases and disease-free life expectancy in the general population: a population-based cohort study. PLoS Med. 2019;16(2).

31. Jacob L, Breuer J, Kostev K. Prevalence of chronic diseases among older patients in German general practices. Ger Med Sci. 2016;14.Doc03.

32. Roche MM and Wang PP. Sex Differences in All-Cause and Cardiovascular Mortality, Hospitalization for Individuals With and Without Diabetes, and Patients With Diabetes Diagnosed Early and Late. Diabetes Care. 2013;36(9):2582-90.

33. Bourdel-Marchasson I, Helmer C, Fagot-Campagna A, et al. Disability and quality of life in elderly people with diabetes. Diabetes Metab. 2007;33(Suppl 1):S66-74.

34. Rockwood K, Tan MH, Phillips S, Mcdowell I. Prevalence of diabetes mellitus in elderly people in Canada: report from the Canadian Study of Health and Aging. Age Ageing. 1998;27:573-7.

35. CBS-Centraal Bureau of Statistics. Chapter 2. Diabetes in beeld vanuit verschillende bronnen. In: Gezondheid en zorg in cijfers 2014. CBS. Den Haag/Herleen; 2014:27-46.

36. Gini R, Francesconi P, Mazzaglia G, et al. Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey. BMC Public Health. 2013;13:15.

37. De Bruin A, Voorips L, Ploemacher J, et al. Netherlands Pilot Project on Morbidity Statistics. The Haag/Bilthoven; 2011.

38. InEen. Transparante Ketenzorg Diabetes Mellitus, COPD en VRM Rapportage Zorggroepen Over; 2013. 2014:85.

39. Lu TH, Walker S, Huang CN. It is not appropriate to record diabetes on death certificates for every diabetic patient. Diabetes Res Clin Pract. 2004;65:293-5.

40. Newton CA, Adeel S, Sadeghi-Yarandi S, et al. Prevalence, quality of care, and complications in long term care residents with diabetes: a multicenter observational study. J Am Med Dir Assoc. 2013;14(11):842-6.

41. Dybicz SB, Thompson S, Molotsky S, Stuart B. Prevalence of diabetes and the burden of comorbid conditions among elderly nursing home residents. Am J Geriatr Pharmacother. 2011;9(4):212-23.

42. LeRoith D, Biessels GJ, Braithwaite SS, et al. Treatment of diabetes in older adults: an endocrine society clinical practice guideline. J Clin Endocrinol Metab. 2019;104(5):1520-74.

43. Cho NH, Shaw JE, Karuranga S, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract. 2018;138:271-81.

44. Blecker S, Herbert R, Brancati FL. Comorbid diabetes and end-of-life expenditures among Medicare beneficiaries with heart failure. J Card Fail. 2012;18(1):41-6.

45. Hamada S and Gulliford MC. Drug prescribing during the last year of life in very old people with diabetes. Age Ageing. 2017;46(1):147-51.

46. Dunning T, Duggan N, Savage S. Caring for People with Diabetes at the End of Life. Curr Diab Rep. 2016;16(11):103.

47. National Institute for Public Health and the Environment. Public Health Foresight Study 2018 – A healthy prospect Synthesis. RIVM. Bilthoven; 2018.

48. Klijs B, Mackenbach JP, Kunst AE. Disability occurrence and proximity to death. Disabil Rehabil. 2010;32(21):1733-41.

49. Pefoyo AJK, Bronskill SE, Gruneir A, et al. The increasing burden and complexity of multimorbidity. BMC Public Health. 2015;15:415.

# Supplementary Material

**Supplementary Table S1.** Prevalence of diabetes in the last two years of life (according to Cause of death, Hospital Discharges and Dispensed Medicines registration combined) among persons who were part of the Nivel-Primary Care Database and died in 2015–2016 and among deaths in the Dutch population 2015–2016 stratified by institutionalization status.

| | Proportion of deaths with diabetes n/N* (%) | | | | |
|---|---|---|---|---|---|
| | | Dutch population stratified by institutionalization status | | | |
| | Nivel-Primary Care Database 2015–2016 $N_1$* | Institutionalized exclusively $N_2$* | Institutionalized partially $N_3$* | Not institutionalized $N_4$* | Total $N_5$* |
| **Age, years** | | | | | |
| <50 | 8.9 | 8.6 | 7.0 | 7.9 | 7.9 |
| 50–60 | 15.7 | 10.9 | 18.4 | 16.1 | 16.0 |
| 60–70 | 23.3 | 16.6 | 25.4 | 23.2 | 23.0 |
| 70–80 | 30.5 | 22.6 | 29.4 | 29.4 | 29.0 |
| 80–90 | 28.8 | 22.3 | 27.4 | 26.9 | 26.4 |
| >90 | 20.5 | 15.7 | 18.5 | 18.6 | 17.9 |
| **Sex** | | | | | |
| female | 24.8 | 18.8 | 24.0 | 22.9 | 22.4 |
| male | 25.9 | 19.0 | 24.8 | 25.3 | 24.8 |
| **Total** | 25.3 | 18.9 | 24.3 | 24.2 | 23.5 |

* N = $N_1$, $N_2$, $N_3$, $N_4$ or $N_5$

# Chapter 6

## Estimating the lifetime risk of dementia using nationwide individually linked Cause-of-Death and healthcare register data

Bart Klijs, **Marianna Mitratza**, Peter P.M. Harteloh, Eric P. Moll van Charante, Edo Richard, Markus M.J. Nielen, Anton E. Kunst

*Published in*
*International Journal of Epidemiology 2020;dyaa219.*

# Abstract

**Background:** Previous estimates of the lifetime risk of dementia are restricted to older age groups and may suffer from selection bias. In this study, we estimate the lifetime risk of dementia starting at birth using nationwide integral linked health register data.

**Methods:** We studied all deaths in the Netherlands in 2017 (N = 147,866). Dementia was assessed using the cause-of-death registration, individually linked with registers covering long-term care, specialized mental care, dispensed medicines, hospital discharges and claims, and primary care. The proportion of deaths with dementia was calculated for the total population and according to age at death and sex.

**Results:** According to all data sources combined, 24.0% of the population dies in presence of dementia. This proportion is higher for females (29.4%) than for males (18.3%). Using multiple causes of death only, the proportion with dementia is 17.9%. Sequential addition of long-term care and hospital discharge data increased the estimate with 4.0 and 1.5%-points. Further addition of dispensed medicines, hospital claims and specialized mental care data added another 0.6%-points. Among persons who die at age 65-70 or younger, the proportion with dementia is 6.2% or lower. After age 70, the proportion rises sharply, with a peak of 43.9% for females and 33.1% for males at age 90–95.

**Conclusions:** Around one fourth of the Dutch population is diagnosed with dementia at some point in life and dies in presence of dementia. It is a major challenge to arrange optimal care for this group.

# Introduction

Dementia is one of the world's major public health problems. Worldwide, around 44 million people have the disease and it is the fifth leading cause of death [1]. Between 1990 and 2006, the global burden of dementia more than doubled [1]. It is estimated that in 2050 more than 106 million people will have the disease [2]. Dementia is characterized by a progressive loss of cognitive functioning and increasing dependency on all levels of daily activities [3, 4]. Currently there is no treatment to cure dementia or to alter its progressive course. Surveys among adults in the western world show that 30–60% are at least a little worried about developing dementia [5]. Besides physical and mental suffering, people fear to lose their independence and dignity, or to become a burden to others [6].

The lifetime risk of dementia is the probability that a person will develop dementia before death. The lifetime risk of dementia complements traditional population health measures such as point prevalence, incidence and mortality. The measure is relevant to individuals who have questions about the probability to end their life with dementia. Estimates of the lifetime risk of dementia are also relevant for end-life care planning because they quantify how many people per year die with dementia. Previous studies have provided estimates starting at a relatively old age, between age 55 and 70 years, while none of the studies investigated the risk starting at birth [7-13]. Furthermore, all previous studies used (longitudinal) sample data in combination with epidemiological models [7-13]. A disadvantage of using sample data is that it can lead to selection bias, and the choice of the epidemiological model used likely affects the study result. Partly, these factors may explain the large variation of estimates across studies. At age 65–70, for example, the estimated lifetime risk varies from 7.1% in males and 16.6% in females to 42.4% for males and females together [7, 8].

One way to avoid selection bias and the dependence on epidemiological models is using national or regional health register data instead of sample data. Cause-of-death data, for example, are nationwide and contain information on the occurrence of dementia at the end of life. However, by default, causes of death describe the occurrence of diseases that are died from rather than with. In a recent study in England and Wales, only 45% of all deaths with dementia had it reported as a cause of death [14]. More complete information on the end of life occurrence of dementia can be obtained by linking cause-of-death data with data from other medical registrations [15-17]. In the Netherlands, various registers with information on the occurrence of dementia in individuals are available. These registers cover hospital care, long-term care, specialized mental care, dispensed medicines, and

primary care. All registers (except primary care) are nationwide and contain data that can be linked at the individual level.

The aim of this study is to estimate the lifetime risk of dementia starting at birth, using integral data for a whole population. We investigate how the proportion of deaths with dementia varies according to age of death and sex. We use nationwide and individually linked data from various Dutch health registrations. We investigate which data sources contain most information on the occurrence of dementia at the end of life and to what extent data sources contain overlapping or unique information.

# Methods

## Study population

Our study population consisted of all deaths in the Netherlands in 2017, recorded in the Dutch cause-of-death registration (N = 147,866). The average age of death is 78.6 years (standard deviation is 14.3 years), and 51.8% of the study population is female. Of all deaths, 86.0% had been admitted to the hospital within 2 years prior to death, 39.0% used long-term care and 5.4% used specialized mental healthcare. According to the Dutch law, patients' informed consent or approval by a medical ethics committee are not required for pseudonymised health record linkage studies like ours (Dutch Civil Law, Article 7:458).

## Occurrence of dementia according to different data sources

For each person in our study population, the presence of dementia at the end of life was assessed using causes of death and information from health registries covering long-term care, hospital discharges, hospital claims, specialized mental care, dispensed medicines and primary care. The information from the different data sources was linked at the level of individuals using a unique person identifier. To identify dementia at the end of life, a period of exactly two years prior to death was chosen. This period allowed assessing the presence of dementia in a relatively long period prior to death, without using an overly excessive amount of data. The codes used to define cases of dementia in the cause-of-death registration and in the various medical registrations are presented in **Supplementary Table S1**.

The cause-of-death registration covers the total population of the Netherlands. In the Netherlands, causes of death are coded according to the 10th revision of the International Classification of Diseases (ICD-10) using Iris 4.4.1 software [18]. The cause-of-death registration includes the single Underlying cause of death, i.e. the disease or injury initiating the chain of morbid events leading directly to death, and one or more non-underlying (intermediate or contributory) causes of death. Mentions of dementia as the single underlying cause of death or as one of the non-underlying causes of death were extracted from this data source for the year 2017 [19]. The term 'multiple causes of death' is used when mentions of dementia, either as the single underlying cause or as one of the non-underlying causes, are counted.

The long-term care register contains information about indications for use of long-term continuous (24 hours per day) care that is financed through the Dutch Long-term Care Act. According to the Long-term Care Act, each potential candidate for long-term care is submitted to a needs' assessment. When this needs assessment confirms that institutionalized long-term care is required, the patient receives one out of nine available 'indications' for long-term care. One of these indications was 'Nursing and care in an institution with intensive dementia care', which we used to define dementia (**Supplementary Table S1**).

The hospital discharge register is owned by Dutch Hospital Data and contains information on hospital discharges of all general and academic hospitals in the Netherlands. Diagnoses in the hospital discharge register are coded according to the ICD-10. Dementia was defined using the primary and (one or more) secondary diagnoses of discharge available (**Supplementary Table S1**).

Dutch medical specialist care is financed through Diagnosis-Treatment-Combinations (DTC's), which are similar to Diagnosis-Related-Groups (DRG's). The hospital declarations register contains information about all DTC financed treatments by medical specialists. The DTC-system uses a nonstandard classification of diagnoses, with separate categorizations per medical specialism.

Similar to medical specialist care, specialized mental care is financed through DTC's. The specialized mental care register contains information on mental care treatments in mental care institutions or independent practices. The specialized mental care register contains diagnostic information following the classification used in the Diagnostic and Statistical Manual of Mental Disorders (DSMVI).

The dispensed medicines register is owned by the National Health Care Institute and contains information of all dispensed medication reimbursed under the statutory basic medical insurance. Medication dispensed in residential homes for the elderly are included, but medication provided to hospitals or nursing homes are not. Within the dispensed medicines data that was available to us, ATC5 coding was used.

The Nivel primary care database contains information on disease episodes of patients registered in a subset of Dutch general practices covering approximately 10% of the Dutch population. Disease episodes are constructed using information from patient records and are coded according to the ICPC-1 classification. Once established, diagnoses of chronic diseases are carried onward into the data of more recent years [20].

Information on age, sex, and institutionalization were available through linkage with the Dutch population register and data on household type.

## Statistical analysis

The absolute number and the proportion of deaths with dementia at the end of life according to each single data source was calculated. Furthermore, the increase in the number and proportion of deaths with dementia, when consecutively adding information from the different medical registrations to the information from the cause-of-death registration, was calculated. In this analysis, the medical registrations were added in descending order of the number of dementia cases in each of the medical registrations. The number and proportion of deaths with dementia were also calculated according to institutionalization (yes/no). The overlap of dementia diagnoses in the different data sources was visualized in a Venn-diagram. The number and proportion of deaths with dementia, and the percentage institutionalized among persons with dementia were plotted in relation to the age of death.

The primary care register was not included in our main analysis, because it covers only 10% of the Dutch population. However, within the (non-institutionalized) primary care subpopulation, it was evaluated whether the primary care register contained cases of dementia that were not contained in any of the other medical registrations.

To select the codes to define the presence of dementia according to each data source, a conservative approach was taken. A sensitivity analysis was performed in which the proportion of deaths with dementia was calculated, taking a less restrictive approach

to the define of dementia. The codes used to define dementia in both approaches are presented in **Supplementary Table S1**.

All analyses were performed using the R software (version 3.2.3).

## Results

Using all data sources in combination resulted in an estimated proportion of deaths with dementia of 24.0% (**Table 1**). 10.8% of all deaths had dementia registered as the underlying cause of death, 7.1% as one of the non-underlying causes of death and 17.9% either as the underlying or as a non-underlying cause of death. Sequential addition of the long-term care data and hospital discharge data to the cause-of-death data increased the estimated proportion with dementia with 4.0 and 1.5%-points, respectively. Further addition of the dispensed medicines, hospital claims and specialized mental care data also increased the proportion, but only with 0.4, 0.1 and 0.1%-points. Of all persons in our study population, 26.3% had lived in an institution in the last two years of their life (**Table 2**). Among persons who had dementia at the end of their life, this was 62.0%. The proportion of deaths with dementia was higher among persons who had lived in an institution at the end of their life (56.7%) than among persons who had not been institutionalized (12.4%).

**Figure 1** presents the overlap and complements of the presence of dementia according to the cause-of-death registration, hospital discharge and claims data, and long-term care data. In these data sources, 34,781 cases of dementia were identified. Of these cases, 4,717 (13.6%) were found in all data sources, and 23,343 (67.1%) were found in at least two of the data sources. Although the data sources have substantial overlap, each source also provides unique cases of dementia. That is, the cause-of-death registration, the long-term care registration and the hospital discharge registration provide 4,932, 4,074, and 2,432 unique cases, respectively.

The proportion of deaths with dementia is higher among females (29.4%) than among males (18.3%; **Figure 2**, upper panel). The proportion also varies according to age. Among men and women who die at age 65–70 or younger, the proportion with dementia is 6.2% or lower. After age 70, the proportion increases sharply, with a peak of 43.9% for females and 33.1% for males at age 90–95.

**Table 1.** Number and percentage of deaths with dementia according to the national cause-of-death and medical registrations for the years 2017.

| | Number (and percentage) of deaths with dementia | | |
|---|---|---|---|
| | **Single data source** | **Cumulative** | **Added to previous** |
| **Total Dutch population (N = 147,866)** | | | |
| 1. Underlying causes of death | 15,951 (10.8) | 15,951 (10.8) | — |
| 2. Non-underlying causes of death | 10,492 (7.1) | 26,443 (17.9) | 10,492 (7.1) |
| 3. Long-term care | 26,014 (17.6) | 32,349 (21.9) | 5,906 (4.0) |
| 4. Hospital discharges | 9,797 (6.6) | 34,603 (23.4) | 2,254 (1.5) |
| 5. Dispensed medicines | 4,540 (3.1) | 35,257 (23.8) | 654 (0.4) |
| 6. Hospital claims | 1,009 (0.7) | 35,415 (24.0) | 158 (0.1) |
| 7. Specialized mental care | 1,036 (0.7) | 35,513 (24.0) | 98 (0.1) |
| **Primary care subpopulation (N = 10,601)** | | | |
| Registrations 1–7 combined | | 1,512 (14.3) | — |
| 8. Primary care | 1,107 (10.4) | 1,711 (16.1) | 199 (1.9) |

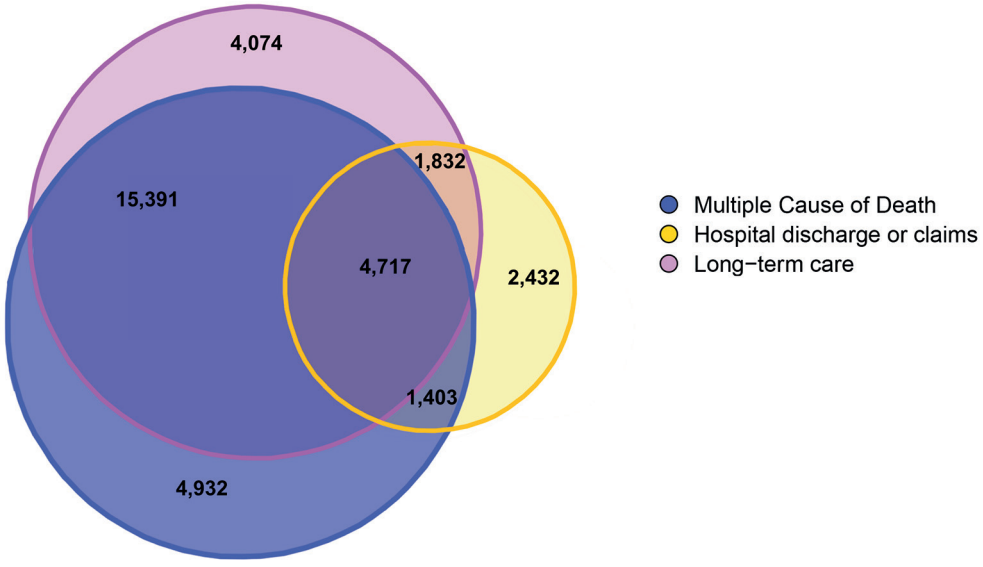


**Figure 1.** Overlap and complements of deaths with dementia according to different data sources.

**Table 2.** Number (and percentage) of deaths with dementia according to data source and Institutionalization.

| | Number (and percentage) of deaths with dementia | | Percentage institutionalization among deaths with dementia |
|---|---|---|---|
| | Institutionalized (N = 38,834) | Non-institutionalized (N = 108,844) | |
| All data sources combined | 22,005 (56.7) | 13,508 (12.4) | 62.0 |
| 1. Underlying causes of death | 10,701 (27.6) | 5,250 (4.8) | 67.1 |
| 2. Non-underlying causes of death | 6,509 (16.8) | 3,983 (3.7) | 62.0 |
| 1+2. Multiple causes of death | 17,210 (44.3) | 9,233 (8.5) | 65.1 |
| 3. Long-term care | 18,031 (46.4) | 7,983 (7.3) | 69.3 |
| 4. Hospital discharges | 4,815 (12.4) | 4,982 (4.6) | 49.1 |
| 5. Dispensed medicines | 2,322 (6.0) | 2,218 (2.0) | 51.1 |
| 6. Hospital claims | 360 (0.9) | 649 (0.6) | 35.7 |
| 7. Specialized mental care | 564 (1.5) | 472 (0.4) | 54.4 |

188 deaths for which no information on institutionalization prior to death was available are excluded from this table.

Of all deaths with dementia, 63.3% were female (**Figure 2**, middle panel). Of all deaths with dementia, 7.3% died before the age of 75, 29.8% died between age 75–85, 52.2% between age 85–95, and 9.6% at age 95 or older.

The percentage of institutionalization in relation with age has a v-shaped pattern (**Figure 2**, bottom panel). At the age of 50–55 years, 72.7% of all males and females who died with dementia lived in an institution. At the age of 70–75, this is 44.3% among males and 48.7% among females. From age 70–75 and onward, there is again a positive relation of institutionalization with age. At the age of 100 or older, institutionalization is highest, namely 79.5% among males and 81.1% among females. In general, the level of institutionalization is higher among females than among males.

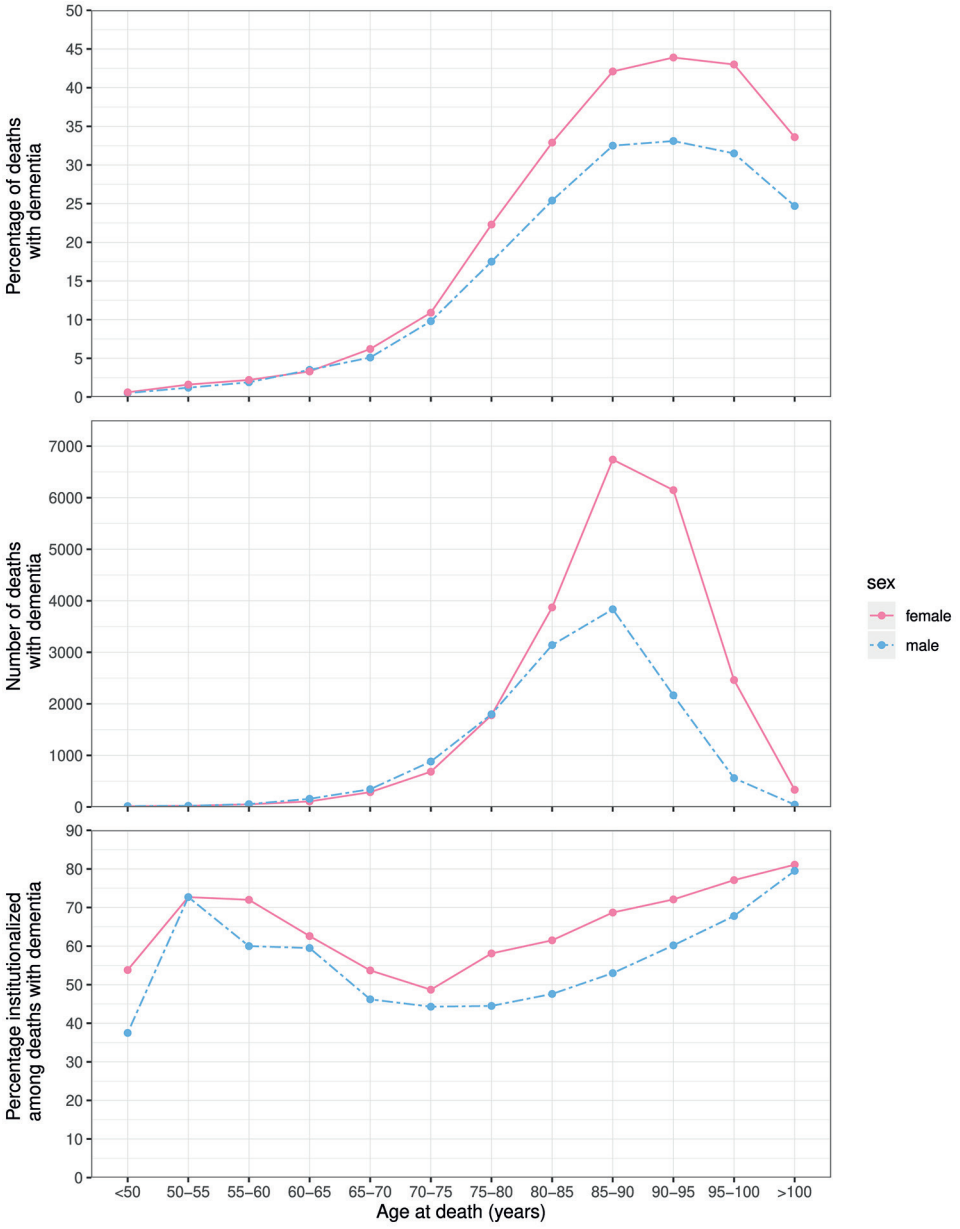


**Figure 2.** Number and percentage of deaths with dementia (N = 35,513), and percentage institutionalized, according to age and sex in 2017.

Within the primary care sub-population, 169 cases of dementia were identified in the primary care registration that were not found in the other data sources. Using the primary care data in addition to all other data sources resulted in a 1.9%-point higher estimate of proportion of deaths with dementia (16.1%; **Table 1**).

Using a less restrictive approach to select the codes to define cases of dementia leads to a slightly higher estimate of the proportion of deaths with dementia, i.e. 25.5% instead of 24.0% (**Supplementary Table S2**).

# Discussion

## Summary of findings

Linking the cause-of-death registration with other healthcare registrations, particularly with the long-term are and hospital discharge registration, provides a much better estimate of the lifetime risk of dementia. Almost one out of four persons die in presence of dementia. This is one out of five among males and almost one out of three among females. The proportion of deaths with dementia is highest for those who die at the age of 90–95 years. Within this group, one third of all males and almost half of all females die with dementia.

## Methodological considerations, comparison with previous studies and interpretation of findings

Our study is the first to estimate the proportion of deaths with dementia, using integral data for a whole population [7-13]. For various reasons, the estimated 24.0% of all deaths with dementia, is likely conservative. Firstly, a restrictive approach was taken in the selection of the diagnosis codes to define dementia. Secondly, as was demonstrated within the primary care sub-population, a small proportion of all persons who have dementia in the population are not identified in any of the registrations used in our analysis (**Table 1**). Thirdly, our analysis does not include undiagnosed dementia [24].

Previous studies investigating the lifetime risk of dementia all took a longitudinal perspective, using longitudinal cohort data [7-13]. Basically, all studies used a cohort life table approach in which the age-specific incidence of dementia was combined with mortality rates. Our study is different in that we used cross-sectional data and a

period life table approach to combine dementia prevalence rates with mortality rates. In this period life table approach, the lifetime risk of dementia is estimated for 'average individuals', under the assumption that these individuals, throughout their life, would be exposed to the age-specific prevalence and mortality rates as observed in our study. This assumption is equivalent to the assumption underlying the well-known period (healthy) life expectancies.

The lifetime risk of dementia estimated in previous studies showed great heterogeneity, which may be due to differences in the study samples and epidemiological models used [7-13]. One of the previous studies was conducted in the Netherlands [10]. In the Rotterdam study, 7,046 non-demented subjects were followed up for, on average, 2.1 years. During follow-up, 162 persons developed dementia and the estimated lifetime risk of dementia at age 55 was 15.9% among males and 32.6% among females. These estimates are reasonably in line with the estimates in our study (18.3% among males and 29.4% among females).

We found a higher lifetime risk of dementia among women than among men, which is consistent with previous studies [7, 9, 10]. A traditional explanation for the higher lifetime risk among women is that the risk of dementia increases with age and that women have a higher life expectancy than men [12]. However, this cannot explain that the age-specific proportion of deaths with dementia is higher among women than among men. Other factors may play a role here. Various studies show that the incidence of dementia is higher among women than among men, while case-fatality of dementia is higher among men [10, 21-23]. This means that the 'inflow' in the pool of prevalent dementia is higher among females, while the 'outflow' is lower. Another part of the explanation may be that the rate of undetected dementia is higher among men than among women [24].

According to our results, the percentage of deaths with dementia decreases with increasing age after the age of 90–94, which appears to be counterintuitive. According to the literature, there is no indication that the prevalence of dementia decreases after the age of 90 [25]. Moreover, the mortality rate among persons with dementia increases consistently towards older ages [23]. An alternative explanation may be that the registration of dementia in health register data is relatively incomplete for persons older than 90. A high number of co-occurring diseases among the oldest old, for example, may be a cause of under-registration of certain diseases, including dementia.

Almost 14% of all deaths with dementia have it recorded as a cause of death but are not found in the hospital- or long-term care registrations (**Figure 2**). In part, these deaths can

be individuals who were not institutionalized or hospitalized. Moreover, in the long-term care data, dementia was defined present when the care required was characterized as 'Nursing and care in an institution with intensive dementia care', which may exclude dementia cases requiring less intensive dementia care.

## Implications

Around one fourth of the population is diagnosed with dementia at some point in life and dies in presence of dementia. The coming decades, the proportion of the population that dies in presence of dementia will increase due to ageing of the population. Between now and 2050, the proportion of the population that dies at age 85 or older will increase from 39% to 67% [26]. Using the expected number of deaths according to age and sex for the year 2050, combined with the age-and sex-specific lifetime risk of dementia as observed in our study, we project that between now and 2050, the lifetime risk of dementia would increase from 24.0% to 32.3% [26]. The absolute number of persons who die in presence of dementia will almost double (from 35,500 to 67,000).

The challenge ahead is to provide optimal care for a rapidly growing group of persons with dementia, also in the last phase of life. According to the World Health Organization, dementia care should aim at optimizing physical health, cognition and well-being, and treat accompanying illness [27]. Behavioural and psychological symptoms should be treated and support must be given to long-term support to carers [27].

# References


1. GBD 2016 dementia collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurology. 2019;18:88-06.
2. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement. 2007;3(3):186-91.
3. World Health Organization. Dementia: a public health priority (Infographic; https://www.who.int/mental_health/neurology/dementia/infographic_dementia/en/). World Health Organization 2007.
4. Mitchell SL, Teno JM, Kiely DK, et al. The clinical course of advanced dementia. N Engl J Med. 2009;361(16):1529-38.
5. Bowen CE, Kessler EM, Segler J. Dementia worry in middle-aged and older adults in Germany: sociodemographic, health-related and psychological correlates. Eur J Ageing. 2018;16(1):39-52.
6. Volicer L. Fear of Dementia. J Am Med Dir Assoc. 2016;17(10):875-8.
7. Beiser A, D'Agostino RB Sr, Seshadri S, et al. Computing estimates of incidence, including lifetime risk: Alzheimer's disease in the Framingham Study. The Practical Incidence Estimators (PIE) macro. Stat Med. 2000;19(11-12):1495-522.
8. Carone M, Asgharian M, Jewell NP. Estimating the lifetime risk of dementia in the Canadian elderly population using cross-sectional cohort survival data. J Am Stat Assoc. 2014;109(505):24-35.
9. Fishman E. Risk of Developing Dementia at Older Ages in the United States. Demography. 2017;54(5):1897-919.
10. Ott A, Breteler MM, van Harskamp F, et al. Incidence and risk of dementia. The Rotterdam Study. Am J Epidemiol. 1998;147(6):574-80.
11. Brookmeyer R and Abdalla N. Estimation of lifetime risks of Alzheimer's disease dementia using biomarkers for preclinical disease. Alzheimer's & Dementia. 2018;14(8):981-88.
12. Seshadari S and Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. Lancet Neurol. 2007;6:1106-14.
13. Seshadari S, Wolf PA, Beiser A, et al. Lifetime risk of dementia and Alzheimer's disease. Neurology. 1997;49(6):1498-504.
14. Gao L, Calloway R, Zhao E, et al. Accuracy of death certification of dementia in population-based samples of older people: analysis over time. Age and Ageing. 2018;47(4):589-94.
15. Zilkens RR, Spilsbury K, Bruce DG, Semmens JB. Linkage of hospital and death records increased identification of dementia cases and death rate estimates. Neuroepidemiology. 2009;32(1):61-9.
16. Pujades-Rodriguez M, Assi V, Gonzalez-Izquierdo A, et al. The diagnosis, burden and prognosis of dementia: A record-linkage cohort study in England. PLoS One;13(6):e0199026.
17. Mitratza M, Kunst AE, Harteloh PPM, et al. Prevalence of diabetes mellitus at the end of life: An investigation using individually linked cause-of-death and medical register data. Diabetes Res Clin Pract. 2020;160(108003).

18. Harteloh PPM. The implementation of an automated coding system for cause-of-death statistics. Inform Health Soc Care. 2018;1-14.
19. World Health Organization-WHO. International statistical classification of diseases and related health problems, tenth revision. Volume II. Geneva: World Health Organization, 1993.
20. Nielen MMJ, Spronk I, Davids R, et al. Estimating morbidity rates based on routine electronic health records in primary care: observational study. JMIR Med Inform. 2019;26;7(3):e11929.
21. Fratiglioni L, Viitanen M, von Strauss E. et al. Very old women at highest risk of dementia and Alzheimer's disease: incidence data from the Kungsholmen Project, Stockholm. Neurology. 1997;48(1):132-8.
22. Letenneur L, Gilleron V, Commenges D, et al. Are sex and educational level independent predictors of dementia and Alzheimer's disease? Incidence data from the PAQUID project. J Neurol Neurosurg Psychiatry. 1999;66(2):177-83.
23. Todd S, Barr S, Roberts M, Passmore AP. Survival in dementia and predictors of mortality: a review. Int J Geriatr Psychiatry. 2013;28:1109-24.
24. Lang L, Clifford A, Wei L, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open. 2017;7:e011146.
25. Cao Q, Tan CC, Xu W, et al. The prevalence of dementia: A systematic review and meta-analysis. J Alzheimer Dis. 2020;73(3):1157-66.
26. CBS Statline, 28-1-2020. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82690NED/table?dl=30937.
27. World Health Organization. Towards a dementia plan: a WHO guide. Geneva: World Health Organization; 2018.

# Supplementary Material

**Supplementary Table S1.** Codes used to define dementia in the Dutch cause-of-death and other medical registrations.

| | Years | Coding system | Codes for dementia |
|---|---|---|---|
| **Causes of death** | 2017 | ICD10 | F00 – Dementia in Alzheimer disease |
| | | | F01 – Vascular dementia |
| | | | F02 – Dementia in other diseases classified elsewhere |
| | | | F03 – Unspecified dementia<br>F05.1 – Delirium superimposed on dementia |
| | | | G30 – Alzheimer's disease |
| | | | G31.0 – Circumscribed brain atrophy (Frontotemporal dementia, Pick disease and Progressive isolated aphasia) |
| | | | G31.8 – Other specified degenerative diseases of nervous system (Grey matter degeneration, Lewy bodies dementia, Subacute necrotizing encephalopathy) |
| | | | G31.1– Senile degeneration of the brain, not elsewhere classified[a] |
| | | | R41.3 – Other amnesia[a] |
| **Long-term care** | 2015–2017 | ZZPcode | 754 – Nursing and care in an institution with intensive dementia care |
| | | Dominant ground | 2 – Psychogeriatric condition[a] |
| **Hospital discharges** | 2015–2017 | ICD10 | See cause-of-death codes |

| | Years | Coding system | Codes for dementia |
|---|---|---|---|
| **Hospital claims** | 2015–2017 | DBC-MSZ coding | 0330-0401 – Neurology – Dementia syndrome |
| | | | 0313-0091 – Internal Medicine – Memory problems and dementiaa |
| | | | 0335-0242 – Geriatrics – Memory problems and dementiaa |
| **Dispensed medicines** | 2015–2017 | ATC5 | N06D (excl. N06DX02) – Anti dementia drugs |
| **Specialized mental care** | 2015–2016 | ICD10 | See cause-of-death codes |
| **Primary care** | 2017 (retrospect) | ICPC1 | P70 – Senile dementia/Alzheimer |

a) Only used in sensitivity analysis (Results in **Supplementary Table S2**)

**Supplementary Table S2.** Number and percentage of deaths with dementia; sensitivity analysis using a less conservative selection of codes to define dementia.

| | Number (and percentage) of deaths with dementia | | | |
|---|---|---|---|---|
| | **Single data source** | **Cumulative** | **Added to previous** | **Cumulative in conservative approach*** |
| **Total Dutch population (N = 147,866)** | | | | |
| 1. Underlying causes of death | 15,952 (10.8) | 15,952 (10.8) | — | 15,951 (10.8) |
| 2. Non-underlying causes of death | 10,500 (7.1) | 26,452 (17.9) | 10,500 (7.1) | 26,443 (17.9) |
| 3. Long-term care | 29,924 (20.2) | 33,754 (22.8) | 7,302 (4.9) | 32,349 (21.9) |
| 4. Hospital discharges | 10,036 (6.8) | 35,865 (24.3) | 2,111 (1.4) | 34,603 (23.4) |
| 5. Dispensed medicines | 4,540 (3.1) | 36,466 (24.7) | 601 (0.4) | 35,257 (23.8) |
| 6. Hospital claims | 4,716 (3.2) | 37,566 (25.4) | 1,100 (0.7) | 35,415 (24.0) |
| 7. Specialized mental care | 1,067 (0.7) | 37,642 (25.5) | 76 (0.0) | 35,513 (24.0) |
| **Primary care subpopulation (N = 10,601)** | | | | |
| Registrations 1–7 combined | | 1,654 (15.6) | — | 1,512 (14.3) |
| 8. Primary care registration | 1,107 (10.4) | 1,823 (17.2) | 169 (1.6) | 1,711 (16.1) |

* Additional dementia codes:
Underlying and Non-underlying causes of death, Hospital discharges, and specialized mental care:
G31.1 – Senile degeneration of the brain, not elsewhere classified; R41.3 – Other amnesia
Long-term care: 2 – Psychogeriatric condition,
Hospital claims: 0313-0091 – Internal Medicine – Memory problems and dementia; 0335-0242 – Geriatrics – Memory problems and dementia

# Chapter 7

General Discussion

# General discussion outline

This chapter is organized on the basis of the three different ways of applying of Cause-of-Death (CoD) data in public health research and epidemiology that are investigated throughout this thesis. More specifically, we applied novel methods to use CoD data to monitor time-trends (**Part I**), to assess co-morbidities at the end-of-life stage (**Part II**), and to estimate the lifetime risk of diseases (**Part III**). Various data sources were utilised for this purpose, with CoD data as the main core, allowing for an in-depth exploration of the existing opportunities of using CoD data combined with other sources.

Below, for each part, we briefly present the topic of illustrative application, and then give a summary of the key findings of our corresponding studies. Subsequently, we reflect on the quality of data sources and methods used by discussing both strengths and limitations. Finally, potential implications of the findings and future challenges for public health and research are discussed.

# Part I
## Using underlying Cause-of-Death data to monitor trends over time

### I.a. Illustrative application of CoD data

In this part, the objective was to assess the level of detail of cause-of-death (CoD) data – in terms of annual number of deaths and granularity – that is possible for monitoring mortality trends. Currently, either excessively long lists of all available International Classification of Diseases (ICD) codes or abbreviated lists of leading high-prevalent CoDs can be used in the analysis of trends [1, 2]. Both approaches may offer biased views of the importance of CoDs for health policy, because less common conditions may be lost among thousands of ICD codes or, in the other approach, be ignored. Ultimately, we aimed to develop a rule of thumb regarding the minimum number of annual deaths of a CoD for inclusion in monitoring and research on long-term trends in mortality in European countries. This rule had to be based on a fair trade-off between producing valuable and invaluable signals.

### I.b. Key findings

In **Chapter 2**, as a first step, we utilized the Dutch CoD database in order to assess the role of CoD size in the detection of short- and long-term mortality trends. For each of the 625 analyzed CoDs, the best fitted model was selected among polynomial regression models of degree zero up to three. According to our findings, about two thirds of the CoDs (i.e. 409 CoDs) had a demonstrable long-term trend. A curved instead of linear trend was identified in many cases across a 20-year period, sometimes even for CoDs with few annual deaths on average. CoD size was an important predictor of the probability of detecting a long-term trend. With outlier detection analysis, an exceptionally high or low number of deaths in one year was identified for only few CoDs. Our main conclusion was that a much broader range of CoDs, as compared to abbreviated lists, could be included in the monitoring of long-term mortality trends.

In **Chapter 3**, the second study of this part, we replicated our findings from the Netherlands to European countries with data available for a 15-year period covered by ICD-10. Our purpose was to study the effect of CoD size in larger as well as smaller countries, in addition to a middle-sized country like the Netherlands. We verified the strong relationship between CoD size and the possibility to detect a long-term trend in mortality. Every 10% increase of CoD size was associated with an 18% increase in the probability of detecting

a trend, with a negligible effect of country. Neoplasms and digestive disorders were less likely to have a demonstrable trend. In contrast, infectious and mental diseases were more likely to have a demonstrable trend. By applying receiver operating characteristic (ROC) curve diagnostics, we derived a general threshold of 30 annual deaths for inclusion of a CoD in long-term mortality trends analysis, and a more specific criterion of 65 deaths for neoplasms and 20 for infectious diseases. Overall, this study showed that it may be feasible to select CoDs for long-term mortality-trends analysis across European countries on the basis of a common general CoD size threshold.

This finding led to the identification of 52 CoDs that surpassed the cut-off point of 30 annual deaths on average, and, thus, could be used for comparison of long-term mortality trends in European countries. **Table 1** shows that the majority of these CoDs are malignant neoplasms (22), diseases of the circulatory system (13), and diseases of the digestive system (6). Mental and behavioural disorders due to alcohol use and Parkinson's disease or Alzheimer's disease are part of the list of CoDs, which could be useful for monitoring long-term trends. When the specific threshold of 65 annual deaths on average is applied for each country, 13 malignant neoplasms are included (**Table 1**, ICD-10 codes highlighted in bold). When the specific threshold of 20 annual deaths on average is applied for each country, only one infectious disease, unspecified pneumonia, can be monitored.

**Table 1.** List of ICD-10 codes that can be monitored in Europe.

| ICD-10 Code | ICD-10 Code Label | ICD-10 Code | ICD-10 Code Label |
|---|---|---|---|
| C15 | Malignant neoplasm of oesophagus | I11 | Hypertensive heart disease |
| **C16** | Malignant neoplasm of stomach | I21 | Acute myocardial infarction |
| **C18** | Malignant neoplasm of colon | I25 | Chronic ischaemic heart disease |
| **C20** | Malignant neoplasm of rectum | I26 | Pulmonary embolism |
| **C22** | Malignant neoplasm of liver and intrahepatic bile ducts | I35 | Nonrheumatic aortic valve disorders |
| **C25** | Malignant neoplasm of pancreas | I42 | Cardiomyopathy |
| C32 | Malignant neoplasm of larynx | I60 | Subarachnoid haemorrhage |
| **C34** | Malignant neoplasm of bronchus and lung | I61 | Intracerebral haemorrhage |
| C43 | Malignant melanoma of skin | I63 | Cerebral infarction |
| **C50** | Malignant neoplasm of breast | I64 | Stroke, not specified as haemorrhage or infarction |
| C53 | Malignant neoplasm of cervix uteri | I70 | Atherosclerosis |
| C54 | Malignant neoplasm of corpus uteri | I71 | Aortic aneurysm and dissection |
| **C56** | Malignant neoplasm of ovary | I80 | Phlebitis and thrombophlebitis |
| **C61** | Malignant neoplasm of prostate | **J18** | Pneumonia, organism unspecified |
| **C64** | Malignant neoplasm of kidney, except renal pelvis | J44 | Other chronic obstructive pulmonary disease |
| **C67** | Malignant neoplasm of bladder | J45 | Asthma |
| **C71** | Malignant neoplasm of brain | K25 | Gastric ulcer |
| **C80** | Malignant neoplasm, without specification of site | K26 | Duodenal ulcer |
| C85 | Other and unspecified types of non-Hodgkin lymphoma | K55 | Vascular disorders of intestine |
| C90 | Multiple myeloma and malignant plasma cell neoplasms | K70 | Alcoholic liver disease |
| C91 | Lymphoid leukaemia | K74 | Fibrosis and cirrhosis of liver |
| C92 | Myeloid leukaemia | K85 | Acute pancreatitis |
| E10 | Type 1 diabetes mellitus | R99 | Other ill-defined and unspecified causes of mortality |
| E11 | Type 2 diabetes mellitus | W19 | Unspecified fall |
| F10 | Mental and behavioural disorders due to use of alcohol | X70 | Intentional self-harm by hanging, strangulation and suffocation |
| G20 | Parkinson disease | | |
| G30 | Alzheimer disease | | |

## I.c. Reflections on the data source of interest and methods

### I.c.1. Strengths

The rich data source used in Part I, in particular the WHO Mortality Database, is a publicly available and reliable compilation of annual mortality data produced by national statistical offices [2], and the Dutch CoD database is part of it. This extensive coverage of underlying CoD data for so many countries allowed the analysis for most European countries, and enabled inferences about different populations. Providing data for a large sample of CoDs and disease groups, ranging from very prevalent to more rare diseases was a prominent advantage in our studies. Thus, we could consider also conditions that might have been overlooked otherwise, because of their exclusion from standardized lists that focus on CoDs with high burden in terms of prevalence or mortality. At the same time, investigating long periods of consecutive years, and the general stability of the ICD-10 classification system over time, made the detection of long-term trends feasible. Given that the CoD coding system is relatively stable among European countries and is subject to common regulation, the data collection procedure is considered robust to a large extent [3, 4].

A novel feature in our methodology was the use of ROC curve analysis. To our knowledge, our study was one of the first to connect disease time-trends with development of a cut-off point using ROC diagnostics [5]. For the purpose of developing a rule-of-thumb for a CoD size threshold that allows for the detection of a long-term trend, we simultaneously considered sensitivity and specificity as criteria. This provided us with an objective method to optimize the trade-off between aiming to capture most signs of mortality changes while avoiding too much noise.

### I.c.2. Limitations

An overarching limitation of the type of data used in this part of the thesis concerns the reporting and coding of underlying CoDs. At the individual level, the reporting of CoDs, and – up to a certain extent – the selection of the underlying CoD in a death certificate, may be affected by the subjective judgement of the certifying physician who is called to give a balanced judgement [6]. Some additional characteristics of the CoD registration, such as the occasionally inconsistent mapping of medical information to ICD-10 codes may entail limitations in the quality and usability of the data. This type of bias is inherent in CoD data. Therefore, completeness or validation studies of death certificates could help to estimate the extent of the bias and yield suggestions to further optimize use of CoDs

in public health practice and research [7]. At the population level, issues of comparability over time and between countries still remain, despite the general international guidelines for underlying CoD registration [8].

Another limitation of our data source is the fact that CoD data alone present limited possibilities for the interpretation of trends. In our study, we found that infectious diseases and mental disorders are more likely to have demonstrable long-term mortality trends, while for neoplasms this is less likely. This might reflect the more variable or more gradual nature of these different types of CoDs, respectively. However, we could only describe the trends, but we cannot draw conclusions about the origin of disease-specific patterns without additional data, such as risk factors' prevalence trends and medical progress relevant to specific diseases [9].

Some of the methodological choices in our studies may have influenced the results of our analyses. Firstly, even though the standard Ordinary Least Squares regression models (OLS) with polynomials of low degrees are not very flexible to model trends in detail, they still allow for a curve and are less likely to overfit trends, as compared to other common methods like Generalized Additive models [10]. For our approach aimed at identifying past mortality time-trends and their relation to the CoD size, and without the intention to make projections for future years, the OLS models can be considered appropriate.

A second limitation in the methodological choices is our focus on the relationship between the number of annual deaths and the detection of long-term trends. Our study intended to identify CoDs with significant changes over time, rather than CoDs with a high impact at any moment of time. Additional criteria for the selection of CoDs to be included in long-term trend analysis were not considered, such as the level of healthcare costs involved in the prevention or management of CoDs, the preventability of certain CoDs or the burden of mortality at younger ages and their corresponding years-of-life-lost. Considering such additional criteria was beyond the general aims of our studies, but could be a next step in further research.

## I.d. Implications and future challenges

The findings presented in Part I reveal new potential applications of underlying CoD data. We found a reasonable likelihood of detecting a long-term mortality trend, even for infrequent CoDs, as defined by their annual number of deaths. Thus, we are able to develop a list of 52 eligible CoDs to be monitored for all European countries with a few

exceptions (i.e. the very small countries that were excluded from our study). This range of CoDs is more limited compared to the Eurostat shortlist which contains 86 codes [1]. However, if more false positive trends for small countries were to be allowed, international comparisons for up to 100 CoDs could be made, as shown in **Table 2**.

**Table 2.** Number of CoDs with at least 30 average annual deaths per country.

| Country | number of CoDs with ≥ 30 deaths | Country | number of CoDs with ≥ 30 deaths |
|---|---|---|---|
| Austria | 111 | Lithuania | 91 |
| Belgium | 112 | Netherlands | 116 |
| Croatia | 102 | Norway | 92 |
| Czech Republic | 118 | Poland | 124 |
| Denmark | 100 | Romania | 119 |
| Estonia | 63 | Slovenia | 76 |
| Finland | 99 | Spain | 123 |
| France | 124 | Sweden | 110 |
| Germany | 125 | Switzerland | 107 |
| Hungary | 123 | United Kingdom | 125 |
| Latvia | 78 | | |

An important insight of our study is that it is not only common to detect linear trends, which could be simply extrapolated, but also numerous curved trends. For policy making, a CoD with a curved behaviour might mean that developments, in general, could be anticipated in a less predictable, depending on the exact point of curve at the year under examination. Public health practitioners may become more vigilant to monitor the most recent trends in mortality of those CoDs.

A remaining challenge is to create a more detailed, but at the same time relevant list of underlying CoDs for regular monitoring of mortality trends. This need will become more compelling when the new ICD-11 classification, with thousands of additional codes, is implemented. Our cut-off criterion to detect long-term trends may help to define a set of

diseases that need particular attention given the likelihood to demonstrably change over time. Our cut-off criterion may be adjusted in practice, rather than used as a fixed, and static value.

Our findings do not indicate significant variation between the countries in the relationship between the CoD size and the detection of a long-term mortality trend. As the trends could be identified with a similar behaviour across countries, a European-wide policy for presenting annual mortality trends could be proposed. For instance, such European-wide list already exists to inform European policies regarding cancer [11].

The methodology developed for our studies has useful implications for future research. A similar approach of developing a practical rule of thumb may have added value for the detection of trends in other types of data used for monitoring, such as annual morbidity data. A noteworthy challenge lies in the difference between morbidity and mortality data, as morbidity data usually are less stable in terms of registration practices over the years and across regions [12].

## Part II
## Using multiple Cause-of-Death data to assess comorbidities at the end-of-life stage

### II.a. Illustrative application of CoD data

In this part, we aimed to illustrate the concept of measuring the mortality burden of several related CoDs when studied altogether, instead of studied separately. As proof-of-concept, we studied Systemic Autoimmune Diseases, a group of serious chronic conditions which share several characteristics in their pathophysiology and their clinical manifestation. However, they are not treated as a uniform disease group in classification systems, for instance in ICD-10, where they are scattered across various chapters [13]. At the same time, SAIDs share the potential for multiple comorbidities, thus, they offer the opportunity of exploring the value of information on non-underlying CoDs at death certificates.

## II.b. Key findings

In **Chapter 4**, we constructed a single SAID class by combining ICD-10 codes that are categorized in distinct ICD-10 chapters. We found that 0.47% of all deaths were related to SAID. The mortality rate was about 15 per million population with SAID measured as an underlying CoD, and about 40 per million population when measured as any-mention CoD. The mortality rate for SAID was higher for females and showed an exponential increase with age. Using a multiple-cause-of-death analysis, we demonstrated that SAID-related deaths were positively associated with different types of comorbidities, and especially with musculoskeletal, genitourinary and blood disorders, as well as infections and skin disorders. When SAID is a non-underlying CoD, the more common underlying CoDs, as compared to the general deceased population, were influenza and other infectious diseases, musculoskeletal diseases, as well as other diseases of the respiratory and the genitourinary system.

## II.c. Reflections on the data source of interest and methods

### II.c.1. Strengths

In this part, we used information on all underlying and non-underlying CoDs reported on death certificates in order to make population estimates on the mortality burden of SAIDs. An about three-fold prevalence of SAID among deceased was found, compared to underlying CoD alone, thus producing an estimate of the overall importance of these chronic diseases in terms of mortality.

Another strength of analysing a nationwide multiple CoD database is the potential to study all types of comorbidities of SAID at the time of death. Estimating the occurrence of less prevalent comorbidities is of particular value when derived from a data source with national coverage and representativeness. The multiple-cause-of-death analysis allowed us to assess the direction and the strength of the relationship between SAID and comorbidities from each system that may be affected in patients suffering from SAID. This is a known advantage of this type of analysis of multiple CoD data, but our study showed that it can also be applied to a large group of closely related diseases.

We were among the first to compile an exhaustive list of SAID diseases that are scattered across ICD-10 categories and to study them as a single class of CoD. In that way, CoD data

were used in order to better estimate the mortality burden and comorbidities of diseases with a common underlying pathophysiological mechanism [14, 15].

### II.c.2. Limitations

The use of individual death certificates as data source in this part entailed some limitations. Firstly, the registration of the non-underlying CoDs depends on the decision of the certifying physician whether or not to report these. It is not clear whether patients suffering from any SAID would have non-underlying CoDs reported on their death certificates with a greater or lower probability in comparison to other deceased people. A linkage with electronic health records might have partially resolved this limitation by providing more insights on the medical history of the patients.

Second, the cross-sectional nature of the death certificate data did not allow to make claims about causality between conditions. Nevertheless, we should recognize that the concept of causal chain is inherent to the structure of the reporting of underlying CoD. The underlying CoD should have started the process that lead to death via other reported intermediate conditions. By closer examination of both underlying and non-underlying CoDs we can get an indication for the direction of associations. Thus, if a disease occurs with SAID as underlying CoD, but not with SAID as non-underlying CoD, it may be implied that this disease lies within the lethal causal chain of SAID and is not likely to act as an independent concurrent disease at the time of death. When one is aware of this limitation, the corresponding findings can be interpreted accordingly, and set the basis for further targeted studies to test for causality, using, for instance, longitudinal cohorts.

## II.d. Implications and future challenges

Future studies may elaborate on the findings of our study, as presented in Chapter 4, or address some of the gaps identified specifically regarding SAID. Firstly, as SAIDs encompass an evolving research field [16], our developed list can be extended when new rare autoinflammatory diseases are identified, and be of general use for researchers in the field of rheumatology and immunology. SAIDs can be assessed in many other data sources except of the mortality database [17]. Secondly, the impact of diseases that do not usually get much attention, such as skin disorders, can be investigated more thoroughly, and taken into account for further prevention and management of the patients suffering from a SAID. Thirdly, an interesting topic would be the examination of multiple CoDs

separately for vasculitides and non-vasculitides, something our study was not able to address due to low numbers. ANCA-associated vasculitis has been associated with poor prognosis [18], so areas for timely management of the patients could be emphasized even more. Fourthly, other studies could aim to ascertain whether an infection at the time of death could be attributed to disease activity or the effect of immunosuppressive drugs in SAID patients, and thus inform the development of drug regulatory policies accordingly. Finally, our finding of higher SAID mortality rate in females could be further explored in future research, given that the role of sex in SAID mortality has not been studied explicitly. One approach would be to disentangle the sex differences in incidence from differences in the clinical course of SAID. Instead of using multiple CoD data alone, the design of longitudinal cohorts could be valuable for this purpose.

Taking into consideration the time when this text is being written, one can think also of ways in which our findings can be of value in the battle against SARS-CoV-2 which is responsible for the new COVID-19 disease [19]. In our study, a strong link between SAID and influenza in mortality was demonstrated. SARS-CoV-2, as a virus that targets primarily the respiratory system, has been often described as resembling influenza. Further studies are required to examine the role of COVID-19 infection in the mortality of patients with SAID, especially in the elderly who have an impaired immune system [20]. This may prove valuable for the prioritization of this vulnerable population in view of the administration of vaccines, which are currently under development.

# Part III
# Enrichment of Cause-of-Death data with healthcare registries to assess disease importance

## III.a. Illustrative application of CoD data

In this part, our aim was to study the potential of linking CoD data with nationwide healthcare registries in order to quantify the prevalence and impact of diseases at the time of death. The lifetime burden of several severe chronic diseases may be underestimated or not measured accurately when using CoD data only. Main reasons for this bias are the fact that chronic diseases, such as diabetes mellitus or dementia, are not always reported on the death certificates [21], or the exclusion of oldest age-groups in epidemiological cohort studies [22].

## III.b. Key findings

In **Chapter 5**, we used a representative sample of Dutch primary care registrations, individually linked with CoD data, as well as medication and hospital discharge registrations, to estimate the prevalence of diabetes mellitus at the end of life. We found that diabetes was present among 30% of all deaths, according to all data sources combined. Using multiple CoD data only, the estimated end-of-life prevalence of diabetes was about 8%, whereas using underlying CoD data only the corresponding percentage was about 2%. We identified general practice data as the medical register with the largest added value to CoD data in detecting cases of diabetes, as it increased this estimate with almost 20%. Almost 80% of all deceased with diabetes used antidiabetic drugs. 76% of the cases with diabetes added by general practice data had a severe or intermediate status of diabetes control, as determined by the use of antidiabetic drugs. We demonstrated that CoD data enriched with general practice data may be sufficient to monitor the end-of-life prevalence of diabetes.

In **Chapter 6**, we estimated the lifetime risk of dementia by combining nationwide registrations for the total Dutch population. CoD data were individually linked with registers covering long-term care, specialized mental care, dispensed medicines, hospital discharges and claims, and primary care. The current lifetime risk of dementia was found to be almost one fourth in the Dutch population, according to all data sources combined. This proportion was higher for women and (for obvious reasons) it was higher for people who had lived longer. According to underlying CoD data only, 11% of persons who died had dementia, while with multiple CoD data, the proportion was 18%. The sequential addition of long-term care and hospital discharge data increased this estimate with 4- and 1.5%-points, respectively. An interesting finding was that among persons who die at age 65–70 or younger, the proportion with dementia was up to 6%. After age 70, the proportion increased sharply, with a peak of 44% for women and 33% for men at age 90–95.

## III.c. Reflections on the data source of interest and methods

### III.c.1. Strengths

The combined use of several data sources is one of the main strengths of our studies. More specifically, in Chapter 5 individual patient-level data from several healthcare data sources were linked to a small, but validated and representative sample of primary care registries – the Nivel Primary Care Database (Nivel-PCD). In Chapter 6 only national data sources were used. In this manner, the representativeness of our study populations was ensured, and robust inferences about the presence of a disease at the time of death for the general Dutch population were possible. In this way, we avoided the potential selection bias that are inherent to most epidemiological cohort studies [23].

Another strength of the studies derives from the generally acknowledged advantages of record linkage [24, 25]. The collection of large volumes of data representing the patient's journey across the healthcare system is of high value. For example, using individually linked patient data, we had the opportunity to derive more robust estimates of dementia prevalence in younger populations.
Our approach for the estimation of lifetime risk of a disease avoids the underlying assumptions of epidemiological models that try to estimate such risks by modelling incidence and survival estimates from limited study samples [26, 27]. In our studies we only used observational data, starting from the deceased population and investigating retrospectively their diagnoses in the immediate past two years. Thus, it can be assumed that our findings may reflect more accurately the real-world situation for a given moment in time.

### III.c.2. Limitations

A limitation is the potential difficulties in the generalization of our methods and findings to other countries. For instance, a primary care registration covering the entire local or national population, which is established only in few countries, may be a prerequisite for estimating the prevalence of diabetes at the end of life. Another limitation to our data is that some patients may remain undiagnosed and therefore not included in the registries [28, 29].

There are some methodological concerns regarding several definitions used in our studies. In Chapter 5, diabetes at the late stages of life may be treated less aggressively, so medication

might not be a representative indicator of severity, resulting in some underestimation in the prevalence of severe diabetes. External data from special laboratory measurements, such as HbA1c, or information on the occurrence of known diabetes complications may have provided a more robust classification of diabetes severity. In Chapter 6, dementia was not defined and registered in a consistent way across data sources. However, this may not necessarily have had biased our findings. To mitigate the inconsistency, we conducted a sensitivity analysis with more broad definition of dementia and we reached similar conclusions.

## III.d. Implications and future challenges

Our findings presented in Part III may be used to inform public health policies. As part of these policies, public awareness could be raised by communicating the high lifetime risk of severe diseases such as diabetes and dementia. Another target of public health policy would be to arrange complex care and healthcare capacities for older people suffering from these diseases. This would be of great value as the burden of diabetes and dementia is expected to rise steeply according to recent projections [30, 31]. For both types of policies, the lifetime risk approach we adopted may be more informative than traditional prevalence estimates outcomes These lifetime estimates emphasize the accumulation of risks over the life course, thus make the concept of disease burden more intuitive. We recommend monitoring lifetime risk of dementia regularly with a multiple-data-source approach, for instance every 5 years, in order to inform stakeholders about foreseen increase of dementia and associated healthcare needs.

Along with the linkage of healthcare data sources per se, we also investigated the impact of the sequence of addition of each data source to our estimates, starting from the CoD database. If resource constraints exist or the time for planning and execution of a record linkage from multiple sources is limited, certain choices need to be made. For the choice of the next data source, it was demonstrated that one should be practical and flexible, and take different approaches for different diseases. This data source should represent the registries most likely to capture the target population, i.e. people with disease in the end-of-life stage in the real world. For example, regarding diabetes, it is reasonable to add primary care data [32]. For dementia, using CoD as the basic source of information was not very far away from the cumulative estimate lifetime risk from all available national sources. Yet, registries from special care facilities, such as mental care and long-term care, could be considered as additional data sources to use in a record linkage.

# Future research

Real-world data from routine healthcare registrations have the potential to offer valuable insights for both public health policy, and medical research and practice, particularly in the era of an ageing population which is characterized also as the era of big data. Cause-of-death data, a source of data derived from the death certificate, may contribute to the study of diseases in such times and populations by several means.

Firstly, as a source with publicly available data of relatively high quality, at least for many countries, cause-of-death databases can serve as an open area where new statistical methods and approaches can be freely tested by researchers on aggregated data. The level of granularity can be selected by the researcher according to the specific question of the study, ranging from national or international populations to specific population subgroups [33]. Currently, an unprecedented challenge for the scientific community arises as new studies are needed to provide estimates of short-term trends of infectious diseases like COVID-19 [34]. Infectious diseases' surveillance is nowadays mostly based on laboratory data rather than CoD data. CoD data are, in general, too slow in providing much information for policy making during the course of an epidemic. However, methods developed for CoD monitoring, such as those presented in this thesis, may be adapted to assist the surveillance of COVID-19. Monitoring of trends at different time-scales, including week-to-week or day-to-day trends, could be useful for monitoring the spread of the epidemic in the future. Other factors available to CoD registries, such as population and regional characteristics, have an important role to play in the monitoring of trends and, therefore, to assist the more accurate prediction of trends.

Secondly, cause-of-death data are typically organized based on standardized classification systems of diseases. We already showed that the systems could be adapted to enable analyses of disease groups based on their underlying pathophysiology and allow all comorbidities to be considered in research and care [35]. From a broader perspective, one could envision a more general transformation of the disease classification system from a static structure based - to a large extent - on distinct anatomical systems towards a dynamic framework based on different views that would otherwise overlap. These views may be built, for instance, on groups of diseases affecting different organs (pancreas, skin, etc.), on different mode of exposure (any organ and systemic infectious diseases together), on different treatment classes available (antimicrobial drugs, biological therapies, etc.), or on different underlying pathways. Regarding the latter point, we showed how developing a class of SAID, we were able to quantify the burden and to identify comorbidities that may

relate to the shared underlying mechanisms of SAIDs. From the public health research perspective, the underlying pathway view may be more actionable, for example, by shifting the attention of health policy priorities from a single organ-related disease towards a pathway that is amenable to action. In this context, future research using multiple CoD data can re-classify frequent classes of CoDs, such as diabetes and cardiovascular diseases, or focus on groups of rare CoDs with no common ICD-10 class. This more holistic view of a disease may align better with the modern concept of systems medicine and integrated care [36].

Third, analyses of CoD data can benefit from the linkage with multiple healthcare and medical registrations [37]. One way is the design of common epidemiological studies with death as the endpoint under investigation, such as in studies on the end-of-life prevalence of disease. Another way is the evaluation of the quality of the CoD database, by estimating its completeness as compared to information known from other registries. A key challenge would be to develop a national platform of individually linked administrative and health data from all relevant data sources – including CoD data – in order to capitalize on the value of each single source [38, 39]. Such an integral platform could offer a tool to researchers to obtain more robust results regarding the life-time burden of specific diseases [40] and to investigate population inequalities in disease burden.

Last, but not least, it would be interesting for future epidemiological research to study lifetime risk as a potential useful health-related indicator for the ageing population. Moving away from traditional point prevalence towards lifetime risk of a serious chronic disease, the time dimension of disease trajectories leading to the end-of-life stage may be recognized more effectively. Therefore, complementary to the indicators of incidence and disease progress, lifetime risk could quantify the cumulative burden [41] and the severity of a disease in a population. Based on lessons learnt from studying diabetes mellitus and dementia, we recommend that future research may focus on the study of the lifetime risk of other serious chronic diseases, such as chronic obstructive pulmonary disease, or even for specific chronic infections, such as HIV. Attention should also be given to lifetime prevalence of multi-morbidity, as older people rarely suffer from only one chronic disease. Future epidemiological research could analyse trends of lifetime risk, study differences between populations, and make comparisons of lifetime risks between diseases.

# Overall Conclusion

In conclusion, our studies based on CoD data highlighted their value for research, monitoring and policies in the public health domain and in the medical field. This added value can become more evident when the current potentials of big data in health are being considered, especially through linkage of several registries in healthcare.

Our findings demonstrated that CoD data can be used to study mortality as a single endpoint, and as a data source to be combined with other data sources for the study of lifetime risk of disease. We may fuel the current growth of public health and epidemiological evidence by further studies that combine routine medical and healthcare data with innovative approaches.

# References


1. European shortlist for causes of death, 2012 [Internet]. Luxembourg: Eurostat; 2016. Available from: http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=COD_2012&StrLanguageCode=EN&IntPcKey=&StrLayoutCode=HIERARCHIChttp://.
2. WHO Application of ICD-10 for low-resource settings initial cause of death collection: The Startup Mortality List (ICD-10-SMoL), V2.1. Geneva: World Health Organization; 2018.
3. Causes of death (hlth_cdeath). Reference Metadata in Euro SDMX Metadata Structure (ESMS). [Internet]. Luxembourg: Eurostat; 2020. Available from: https://ec.europa.eu/eurostat/cache/metadata/en/hlth_cdeath_esms.htm.
4. European Statistics on Causes of Death – COD Methodological information. Implementation of the recommendations of the ad-hoc Task Force on Quality Control. Luxembourg: Eurostat; 2007.
5. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation. 2007;115(5):654-7.
6. Percy C, Stanek E, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. Am J Public Health. 1981;71(3):242-50.
7. Bonneux LG, Huisman CC, de Beer JA. Mortality in 272 European regions, 2002-2004. An update.
8. Janssen F and Kunst AE. ICD coding changes and discontinuities in trends in cause-specific mortality in six European countries, 1950–99. Bull World Health Organ. 2004;82(12):904-13.
9. Gianino MM, Lenzi J, Fantini MP, et al. Declining amenable mortality: a reflection of health care systems? BMC Health Serv Res. 2017;17(1):735.
10. Machine Learning with R, tidyverse, and mlr. Hefin Rhys. Chapter 10
11. D'Onofrio A, Mazzetta C, Robertson C, et al. Maps and atlases of cancer mortality: a review of a useful tool to trigger new questions. Ecancermedicalscience. 2016;10:670.
12. Weiskopf NG and Weng G. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.
13. World Health Organization-WHO. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; World Health Organization: Geneva, Switzerland, 1992; Volume I–III.
14. Arakelyan A, Nersisyan L, Poghosyan D, et al. Autoimmunity and autoinflammation: a systems view on signaling pathway dysregulation profiles. PLoS One. 2017;12:e0187572.
15. Ralli M, D'Aguanno V, Di Stadio A, et al. Audiovestibular Symptoms in Systemic Autoimmune Diseases. J Immunol Res. 2018:5798103.
16. Psarras A, Emery P, Vital EM. Type I interferon-mediated autoimmune diseases: pathogenesis, diagnosis and targeted therapy. Rheumatology (Oxford). 2017;56(10):1662-75.
17. Gossec L, Kedra J, Servy H, et al. EULAR points to consider for the use of big data in rheumatic and musculoskeletal diseases. Ann Rheum Dis. 2020;79(1):69-76.
18. Westman K, Flossmann O, Gregorini G. The long-term outcomes of systemic vasculitis. Nephrol Dial Transplant. 2015;30(Suppl 1):i60-6.

19. European Centre for Disease Prevention and Control. Increased transmission of COVID-19 in the EU/EEA and the UK – thirteenth update. 23 October 2020. Stockholm: ECDC; 2020.

20. Ramos-Casals M, Brito-Zerón P, López-Soto A, Font J. Systemic autoimmune diseases in elderly patients: Atypical presentation and association with neoplasia. Autoimmun Rev. 2004;3(5):376-82.

21. Cheng WS, Wingard DL, Kritz-Silverstein D, Barrett-Connor E. Sensitivity and specificity of death certificates for diabetes: as good as it gets? Diabetes Care. 2008;31(2):279-84.

22. Ott A, Breteler MM, van Harskamp F, et al. Incidence and risk of dementia. The Rotterdam Study. Am J Epidemiol. 1998;147(6):574-80.

23. Nohr EA and Liew Z. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand. 2018;97(4):407-16.

24. Curbelo RJ, Loza E, de Yébenes MJC, Carmona L. Databases and registers: useful tools for research, no studies. Rheumatol Int. 2014;34(4):447-52.

25. Bohensky MA, Jolley D, Sundararajan V, et al. Data Linkage: A powerful research tool with potential problems. 2010;10(1):346.

26. Miller BG, Hurley JF. Life table methods for quantitative impact assessments in chronic mortality. J Epidemiol Community Health. 2003;57(3):200-6.

27. Ligthart S. van Herpt TTW, Leening MJG, et al. Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: a prospective cohort study. Lancet Diabetes Endocrinol. 2016;4(1):44-51.

28. Cho NH, Shaw JE, Karuranga S, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract. 2018;138:271-81.

29. Lang L, Clifford A, Wei L, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open. 2017;7:e011146.

30. Lin J, Thompson TJ, Cheng YJ, et al. Projection of the future diabetes burden in the United States through 2060. Popul Health Metr. 2018;16(1):9.

31. Ahmadi-Abhari S, Guzman-Castillo M, Bandosz P, et al. Temporal trend in dementia incidence since 2002 and projections for prevalence in England and Wales to 2040: modelling study. BMJ. 2017;358:j2856.

32. Saydah SH, Geiss LS, Tierney E, et al. Review of the performance of methods to identify diabetes cases among vital statistics, administrative, and survey data. Ann Epidemiol. 2004;14(7):507-16.

33. Becker R, Silvi J, Fat DM, et al. A method for deriving leading causes of death. Bulletin of the World Health Organization. 2006;84(4):297-304.

34. Lewnard JA, Liu VX, Jackson ML, et al. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: prospective cohort study. BMJ. 2020;369:m1923.

35. Dozmorov MG. Reforming disease classification system-are we there yet? Ann Transl Med. 2018;6(Suppl 1):S30.

36. Bousquet J, Anto JM, Sterk PJ, et al. Systems medicine and integrated care to combat chronic noncommunicable diseases. Genome Med. 2011;3(7):43.

37. Wilkinson T, Ly A, Schnier C, et al. Identifying dementia cases with routinely collected health data: A systematic review. Alzheimers Dement. 2018;14(8):1038-51.

38. Pujades-Rodriguez M, Assi V, Gonzalez-Izquierdo A, et al. The diagnosis, burden and prognosis of dementia: A record-linkage cohort study in England. PLoS One. 2018;13(6):e0199026.

39. Delmestri A and Prieto-Alhambra D. CPRD GOLD and linked ONS mortality records: Reconciling guidelines. Int J Med Inform. 2020;136:104038.

40. Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J. 2018;39(16):1481-95.

41. Carone M, Asgharian M, Jewell NP. Estimating the lifetime risk of dementia in the Canadian elderly population using cross-sectional cohort survival data. J Am Stat Assoc. 2014;109(505):24-35.

# Summary

## Cause-of-Death statistics in public health and epidemiology: Exploring new applications

Mortality statistics are an essential part of epidemiology, public health and medicine. Death reports often are used as a clear endpoint in epidemiological studies. Mortality databases are commonly used in public health research and monitoring thanks to their large coverage in time and space. They are regulated under a strong national and international coordination.

Briefly, the underlying cause-of-death is the condition which initiated the causal chain of morbid events leading directly to death. Any other reported condition is a non-underlying cause-of-death, and every cause of death reported on the death certificate is a multiple cause-of-death. However, there is room for improvement in the applications of mortality data. It is not yet well-studied how many causes of death show a mortality time-trend when an agnostic search among the thousands of common and rare conditions would be performed. Moreover, new groups of causes of death could be formed according to updated medical knowledge on pathophysiological mechanisms, beyond the established categorizations, in order to measure their collective mortality burden in the population. Regarding the quality of cause-of-death data use, combinations with external morbidity information at the patient level can test their validity and broaden their utility to assess the importance of lifelong chronic diseases in an ageing population.

This thesis aims to illustrate new potential applications of already available mortality data for public health monitoring and epidemiological purposes of researchers and statistical offices. Our objectives are to perform explorative case studies and develop methodologies in order to better:

1. use underlying cause-of-death data to monitor trends over time in cause-specific mortality
2. use multiple cause-of-death data to study associations between diseases at the time of death
3. use cause-of-death data linked with healthcare registries to assess the prevalence of diseases at the time of death.

To address these objectives, this thesis is organized in three parts. **Part I** consists of two chapters focusing on how the average number of annual deaths as underlying causes-of-death influences the opportunity to detect mortality time-trends. **Chapter 2** illustrates our methodology to empirically assess mortality time-trends in the short and long term.

We analyzed the Dutch Cause-of-Death database using polynomial regression models and outlier analysis. This chapter shows that two thirds of the 625 causes of death had a demonstrable long-term trend. The average number of annual deaths was an important predictor of the probability of detecting a long-term trend. Yet, even for causes with relatively few annual deaths on average, long-term trends could often be detected. In contrast, an exceptionally high or low number of deaths in one year was identified for only few causes of death. We concluded that when monitoring long-term mortality trends, one could consider a much broader range of causes of death, including ones with a relatively low number of annual deaths, than commonly used in condensed lists.

**Chapter 3** reports on a replication of our method of detecting long-term mortality trends in European countries. This chapter reaffirms the strong relationship between the average number of annual deaths and the likelihood to detect a long-term trend as an outcome, independently of country. With the same annual number of deaths, neoplasms and digestive disorders were less likely to have a demonstrable trend, in contrast with infectious and mental diseases. By applying receiver operating characteristic (ROC) curve diagnostics, a general threshold of 30 annual deaths for detecting trends was derived, and a more specific criterion of 65 deaths for neoplasms and 20 for infectious diseases. We recommend a general criterion of 30 annual deaths to select causes of death for long-term mortality-trends analysis in European countries.

**Part II** consists of **Chapter 4**, which focuses on the use of multiple cause-of-death data after re-classifying systemic autoimmune diseases into a single group of causes of death. This chapter shows that the collective mortality burden of this new disease group was only modest in the Netherlands. Yet, there was an almost three-fold mortality rate when multiple cause-of-data were analyzed in comparison to underlying cause-of-death data. The multiple-cause-of-death analysis demonstrated that deaths related to systemic autoimmune diseases were positively associated with most comorbidities except for solid neoplasms and mental conditions. This positive association was especially strong with musculoskeletal, genitourinary, blood disorders, infections, and skin disorders. When systemic autoimmune disease was a non-underlying cause-of-death, underlying causes-of-death that were more common – compared to the general deceased population - were influenza and other infectious diseases, musculoskeletal diseases, and diseases of the respiratory and the genitourinary system. We concluded that systemic autoimmune diseases constitute a rare group of causes of death, but contribute to mortality through multiple comorbidities. Classification systems could be adapted to better encompass these diseases as a category.

**Part III** consists of two chapters aimed to assess the prevalence of a chronic disease at the time of death by enriching cause-of-death data with nationwide healthcare data. **Chapter 5** investigates different combinations of data sources in order to get estimates of the occurrence of diabetes mellitus at the end of life. This study was based on a sample of Dutch primary care registrations. These were individually linked with cause-of-death data, and medication and hospital discharge registrations. This chapter shows that almost one third of those who died in The Netherlands had diabetes at the end of life, according to all data sources combined. A four-fold end-of-life prevalence of diabetes was found when using multiple cause-of-death data compared to underlying cause-of-death data. The vast majority of people who died with diabetes used antidiabetic drugs, thus had severe or intermediate status. We concluded that primary care data could best be used to detect cases not identified in cause-of-death registries.

**Chapter 6** follows a similar approach in assessing the lifetime risk of dementia. Cause-of-death data were individually linked with registers covering long-term care, specialized mental care, dispensed medicines, hospital discharges and claims, and primary care. According to the combined data, the current lifetime risk of dementia is about one fourth for the national Dutch population. This estimate is higher among women (almost one third) than among men (one fifth), and higher among persons who lived longer, with a sharp increase after age 70. A less than two-fold lifetime risk of dementia was found when using multiple cause-of-death data compared to underlying cause-of-death data. The sequential addition of long-term care and hospital discharge data increased this estimate slightly. We concluded that around one fourth of the Dutch population is diagnosed with dementia at some point in life and dies in presence of dementia. It is a major challenge to arrange optimal care for this group.

Overall, we conclude that cause-of-death databases have much potential for public health and epidemiological research in an ageing population. Firstly, based on **Part I**, we conclude that it may be feasible to select a broad group of causes of death, perhaps much broader than recognized so far, for the analysis of long-term mortality time-trends in European countries. Our results call for the development of a more detailed, but at the same time relevant list of causes of death to be used for public statistics, monitoring and research. This challenge will become more compelling when the new ICD-11 classification, with thousands of additional codes, will be implemented.

Secondly, based on **Part II**, we conclude that the analysis of all data reported on the death certificate made possible to more accurately quantify the mortality burden of the group of systemic autoimmune diseases. Moreover, it made it possible to identify a characteristic

comorbidity pattern. This disease classification that we developed can be extended when new rare autoinflammatory diseases come to light, and be of general use for researchers in the field of rheumatology and immunology. Classification systems could be adapted to better encompass these disease groups, especially for the study of comorbidities that may not get much attention. The strong link between systemic autoimmune disease and influenza in mortality was demonstrated, and calls for urgent research in this patient population in the COVID-19 pandemic.

Thirdly, based on **Part III**, we conclude that combining cause-of-death with multiple healthcare databases provides much accurate estimates of the lifetime risk of severe chronic diseases. Different healthcare databases are indicated for different diseases. The end-of-life prevalence of diabetes can be reasonably monitored when cause-of-death data get enriched with primary care data, while for dementia cause-of-death data may be provide a rough estimate although not complete. Diabetes and dementia are examples of diseases with great personal, societal, and financial impact, and their burden is expected to rise steeply in the coming decades, to a great extent due to the ageing population. By communicating the importance of those diseases in terms of high lifetime risk, which is an intuitive indicator of disease burden, professional and public awareness can be enhanced.

# Samenvatting

## Doodsoorzakenstatistiek in volksgezondheid en epidemiologie: Onderzoek naar nieuwe toepassingen

Sterftecijfers zijn een essentieel onderdeel van epidemiologie, volksgezondheid en geneeskunde. Overlijdensrapporten worden vaak gebruikt als een duidelijk eindpunt in epidemiologische studies. Sterftedatabases worden veel gebruikt bij volksgezondheidsonderzoek en monitoring dankzij hun grote dekking in tijd en ruimte. Ze worden gereguleerd onder een sterke nationale en internationale coördinatie.

In het kort is de onderliggende doodsoorzaak de aandoening die de oorzakelijke keten van ziekelijke gebeurtenissen in gang heeft gezet en die direct tot de dood hebben geleid. Elke andere gerapporteerde aandoening is een niet-onderliggende doodsoorzaak en elke doodsoorzaak die op de overlijdensakte wordt vermeld, is een meervoudige doodsoorzaak. De toepassing van sterftegegevens is echter voor verbetering vatbaar. Er is nog onvoldoende onderzoek gedaan naar de hoeveelheid doodsoorzaken die een sterftetijdstrend vertonen wanneer een agnostische zoektocht onder de duizenden veel voorkomende en zeldzame aandoeningen zou worden uitgevoerd. Bovendien zouden er nieuwe groepen doodsoorzaken kunnen worden gevormd op basis van bijgewerkte medische kennis over pathofysiologische mechanismen, buiten de bestaande categorisaties, om hun collectieve sterftelast in de bevolking te meten. Wat betreft de kwaliteit van het gebruik van gegevens over doodsoorzaken, kunnen combinaties met externe morbiditeitsinformatie op patiëntniveau hun validiteit testen en hun bruikbaarheid verbreden om het belang van levenslange chronische ziekten bij een vergrijzende bevolking te beoordelen.

Dit proefschrift heeft tot doel nieuwe mogelijke toepassingen van reeds beschikbare sterftedata te illustreren voor volksgezondheidsbewaking en epidemiologische doeleinden van onderzoekers en statistiek bureaus. Onze doelstellingen zijn om exploratieve casestudy's uit te voeren en methodologieën te ontwikkelen om:

1. de onderliggende doodsoorzaakgegevens te benutten, zodat trends in de tijdsspecifieke mortaliteit gevolgd kunnen worden
2. de data van meerdere doodsoorzaakgegevens te gebruiken en zo associaties tussen ziekten op het moment van overlijden te kunnen bestuderen
3. de gegevens over doodsoorzaken, die zijn gekoppeld aan gezondheidsregistraties, te gebruiken zodat de prevalentie en impact van ziektes op het moment van overlijden kunnen worden beoordeeld.

Om deze doelstellingen te bereiken, is dit proefschrift onderverdeeld in drie delen. **Deel I** bestaat uit twee hoofdstukken die zich richten op hoe het gemiddelde aantal jaarlijkse sterfgevallen als onderliggende doodsoorzaken de mogelijkheid beïnvloedt om sterftetrends te detecteren. **Hoofdstuk 2** illustreert onze methodologie om de sterftetrends op korte en lange termijn empirisch te beoordelen. We analyseerden de Nederlandse doodsoorzaakdatabase met behulp van polynoomregressiemodellen en uitbijteranalyse. Dit hoofdstuk laat zien dat tweederde van de 625 doodsoorzaken een aantoonbare langetermijntrend vertoonde. Het gemiddeld aantal sterfgevallen per jaar was een belangrijke voorspeller van de kans op het ontdekken van een langetermijntrend. Maar zelfs voor oorzaken met gemiddeld relatief weinig sterfgevallen per jaar, konden vaak langetermijntrends worden vastgesteld. Daarentegen werd een uitzonderlijk hoog of laag aantal sterfgevallen in één jaar vastgesteld voor slechts enkele doodsoorzaken. We concludeerden dat bij het monitoren van langetermijnsterftetrends, men een veel breder scala van doodsoorzaken zou kunnen beschouwen, inclusief die met een relatief laag aantal sterfgevallen per jaar, dan vaak wordt gebruikt in verkorte lijsten.

**Hoofdstuk 3** rapporteert over een replicatie van onze methode voor het detecteren van langetermijn sterftetrends in Europese landen. Dit hoofdstuk bevestigt opnieuw de sterke relatie tussen het gemiddelde aantal jaarlijkse sterfgevallen en de waarschijnlijkheid dat een langetermijntrend als resultaat wordt gedetecteerd, onafhankelijk van het land. Met hetzelfde jaarlijkse aantal sterfgevallen hadden neoplasmata en spijsverteringsstoornissen minder kans op een aantoonbare trend, in tegenstelling tot infectie- en psychische aandoeningen. Door ROC-curve-diagnostiek toe te passen, werd een algemene drempel van 30 jaarlijkse sterfgevallen voor het detecteren van trends afgeleid, en een meer specifiek criterium van 65 sterfgevallen voor neoplasmata en 20 voor infectieziekten. We bevelen een algemeen criterium van 30 jaarlijkse sterfgevallen aan om doodsoorzaken te selecteren voor analyse van langetermijnsterfte-trends in Europese landen.

**Deel II** bestaat uit **Hoofdstuk 4**, dat zich richt op het gebruik van meerdere doodsoorzakengegevens na herclassificatie van systemische auto-immuunziekten in een enkele groep doodsoorzaken. Dit hoofdstuk laat zien dat de collectieve sterftelast van deze nieuwe ziektegroep in Nederland slechts bescheiden was. Toch was er een bijna drievoudig sterftecijfer wanneer meerdere doodsoorzaken werden geanalyseerd in vergelijking met onderliggende doodsoorzaken. De analyse van meerdere doodsoorzaken toonde aan dat sterfgevallen gerelateerd aan systemische auto-immuunziekten positief geassocieerd waren met de meeste comorbiditeiten, behalve solide neoplasmata en mentale aandoeningen. Deze positieve associatie was vooral sterk

wat betreft musculoskeletale, urogenitale aandoeningen, bloedaandoeningen, infecties en huidaandoeningen. Toen systemische auto-immuunziekte een niet-onderliggende doods-oor-zaak was, waren onderliggende doodsoorzaken die vaker voorkwamen – in vergelijking met de algemene overleden bevolking – influenza en andere infectieziekten, aandoeningen van het bewegingsapparaat en aandoeningen van de luchtwegen en het urogenitaal systeem. We concludeerden dat systemische auto-immuunziekten een zeldzame groep doodsoorzaken vormen, maar bijdragen aan mortaliteit door meerdere comorbiditeiten. Classificatiesystemen zouden kunnen worden aangepast om deze ziekten beter als categorie te kunnen omvatten.

**Deel III** bestaat uit twee hoofdstukken die bedoeld zijn om de prevalentie van een chronische ziekte op het moment van overlijden te beoordelen door gegevens over doodsoorzaken te verrijken met landelijke gegevens van de gezondheidszorg. **Hoofdstuk 5** onderzoekt verschillende combinaties van gegevensbronnen om schattingen te krijgen van het voorkomen van diabetes mellitus aan het einde van het leven. Dit onderzoek was gebaseerd op een steekproef van Nederlandse eerstelijnsregistraties. Deze werden individueel gekoppeld aan doodsoorzakengegevens en registraties van medicatie en ontslag uit het ziekenhuis. Dit hoofdstuk laat zien dat volgens alle databronnen bij elkaar bijna een derde van de in Nederland overleden mensen aan hun levenseinde diabetes had. Een viervoudige prevalentie van diabetes rond het levenseinde werd gevonden bij gebruik van meerdere doodsoorzakengegevens in vergelijking met onderliggende doodsoorzakengegevens. De overgrote meerderheid van de mensen die aan diabetes stierven, gebruikte antidiabetica en had dus een ernstige of middelmatige status. We concludeerden dat gegevens uit de eerstelijnszorg het beste kunnen worden gebruikt om casussen op te sporen die niet in doodsoorzakenregisters zijn geïdentificeerd.

**Hoofdstuk 6** volgt een vergelijkbare benadering bij het beoordelen van het levenslange risico van dementie. Gegevens over doodsoorzaken werden individueel gekoppeld aan registers met betrekking tot langdurige zorg, gespecialiseerde geestelijke zorg, verstrekte medicijnen, ontslagen en claims uit het ziekenhuis en de eerstelijnszorg. Volgens de gecombineerde gegevens is het huidige levenslange risico op dementie ongeveer een vierde voor de nationale Nederlandse bevolking. Deze schatting is hoger bij vrouwen (bijna een derde) dan bij mannen (een vijfde), en hoger bij mensen die langer leefden, met een sterke stijging na de leeftijd van 70 jaar. Een minder dan tweevoudig levenslange kans op dementie werd gevonden bij gebruik van multiple doodsoorzaakgegevens vergeleken met onderliggende doodsoorzakengegevens. De opeenvolgende toevoeging van gegevens over langdurige zorg en ontslag uit het ziekenhuis heeft deze schatting licht verhoogd. We concludeerden dat ongeveer een kwart van de Nederlandse bevolking

op enig moment in het leven de diagnose dementie krijgt en met dementie sterft . Het is een grote uitdaging om voor deze groep optimale zorg te regelen.

Over het algemeen concluderen we dat databases met doodsoorzaken veel potentie hebben voor volksgezondheid en epidemiologisch onderzoek bij een vergrijzende bevolking. Ten eerste concluderen we op basis van **Deel I** dat het mogelijk is om een brede groep van doodsoorzaken te selecteren, misschien veel breder dan tot dusverre werd erkend, voor de analyse van de langetermijnsterftetrends in Europese landen. Onze resultaten vragen om de ontwikkeling van een meer gedetailleerde, maar tegelijkertijd relevante lijst van doodsoorzaken die gebruikt kan worden voor openbare statistieken, monitoring en onderzoek. Deze uitdaging wordt nog dwingender wanneer de nieuwe ICD-11-classificatie, met duizenden aanvullende codes, wordt geïmplementeerd.

Ten tweede concluderen we op basis van **Deel II** dat de analyse van alle gegevens op de overlijdensakte het mogelijk maakte om de sterftelast van de groep van systemische auto-immuunziekten nauwkeuriger te kwantificeren. Bovendien werd het mogelijk een karakteristiek comorbiditeitspatroon te identificeren. Deze door ons ontwikkelde ziekteclassificatie kan worden uitgebreid wanneer nieuwe zeldzame auto-inflammatoire ziekten aan het licht komen, en kan van algemeen nut zijn voor onderzoekers op het gebied van reumatologie en immunologie. Classificatiesystemen zouden kunnen worden aangepast om deze ziektegroepen beter te om-vatten, vooral voor comorbiditeitstudies die misschien niet veel aandacht krijgen. Het sterke verband tussen systemische auto-immuunziekte en influenza bij sterfte werd aangetoond, en vereist dringend onderzoek bij deze patiëntenpopulatie in de COVID-19-pandemie.

Ten derde concluderen we op basis van **Deel III** dat het combineren van doodsoorzaken met meerdere databases in de gezondheidszorg veel nauwkeurige schattingen oplevert over het levenslange risico van ernstige chronische ziekten. Voor verschillende ziekten zijn verschillende zorgdatabases aangegeven. De prevalentie van diabetes rond het levenseinde kan redelijk worden gecontroleerd wanneer gegevens over doodsoorzaken worden verrijkt met gegevens uit de eerstelijnszorg, terwijl gegevens over doodsoorzaken voor dementie een ruwe schatting kunnen zijn, hoewel ze niet volledig zijn. Diabetes en dementie zijn voorbeelden van ziekten met grote persoonlijke, maatschappelijke en financiële gevolgen, en de verwachting is dat hun last de komende decennia sterk zal stijgen, grotendeels als gevolg van de vergrijzing. Door het belang van deze ziekten te benadrukken in termen van hoog levenslang risico, dat een intuïtieve indicator voor ziektelast is, kan het professionele en publieke bewustzijn worden vergroot.

# Acknowledgements

My PhD trajectory has been an exciting adventure filled with challenging moments and small victories, and now it has been successfully completed! This thesis could not have been written without the direct or indirect contribution of several people, to whom I am deeply grateful!

Prof. dr. Anton Kunst, my main supervisor, you have offered your excellent guidance throughout this journey, while also being the head of the Public Health department at the Academic Medical Center (AMC). We were able to design together a whole PhD project in a relatively short period of time, and to even complete this in four years, with all our articles published! Throughout this intense procedure, you taught me that good research means to maintain scientific rigor, while keeping the quest for innovation and going a step further. Valuable lessons for anyone who wants to grow as an independent researcher.

Prof. dr. Jan Kardaun, my second supervisor, I am thankful for your valuable advice during the course of my PhD trajectory. With your vision, this PhD project was initiated in collaboration with Statistics Netherlands (CBS) some years before my start. We started over again this project, collaborated in the early studies, and managed to bring the project to a fruitful end. I appreciate your comforting comments in difficult times and your enthusiastic comments in times of celebration!

Dr. Bart Klijs, my co-supervisor, I am glad you joined my supervising team after the first year and a half of my PhD trajectory. Our weekly meetings and your contribution in our research projects gave a serious boost to the progress of the studies.

I would like to sincerely thank also the Doctoral committee for devoting their time to read and comment on my manuscript, and acting as opponents in my PhD defence.

Moreover, I am thankful to my co-authors, who are experts in their field of research. Liesbeth, Peter, Edo, Erick, and Mark, I really enjoyed our collaboration and producing these nice publications. I hope we can work together again sometime in the future.

I am grateful for having met and spent time in and out of the office with my colleagues at the department of Public Health. Especially, I want to thank my AMC roomies, Paulien, Marloes, Lilian, Heike, and Esther for their support and relaxing coffee-moments, lunch breaks and dinners. I want to also thank my colleagues at the department, Karlijn, Mirte, Ben, Els, Joko, Kenneth, Reinie, Sigrid, and Gertrude for our chats and good times in courses, conferences, or just gatherings!

Irene, thank you for our discussions and being approachable to me, as an early-stage PhD researcher in the department. Henrike and Mirte, I find our discussions over statistics and methodology enjoyable and thought-provoking. Prof. dr. Karien Stronks and Prof. dr. Charles Agyemang, both your work and your personalities have been a great inspiration for me. Henriette, Armien, Nita, and Noor, a big thank you for your secretarial support and your quick responses.

Janneke, my esteemed colleague from AMC and cordial friend, I can truly say that this thesis could not have been written without your support. From the very beginning until the end of my PhD trajectory, I felt that you were always there to give me positive vibes and advice. It was incredibly important for me, particularly during isolation in the corona pandemic, to have regular discussions on my progress – among others – while enjoying our virtual coffee in English, Greek and een kleine beetje Dutch!

I would like to sincerely thank my colleagues from CBS. Especially, Kim, thank you for always being there as someone to tend to for all sorts of issues, ranging from kindly navigating me to the labyrinths of the building to helping me get access to the data and explaining the figures to me. Floor, Laura, Corine, Gerard, thank you for safeguarding and providing me with the raw data. Bas, managing the progress of our team at CBS was really helpful, and I appreciate your kind and honest attitude. Eva and Wendy, I would like to thank you for our nice conversations. Nelleke, Marjian, Jan, Marco, and Jan, thank you for all the chats and jokes, while eating so many cakes by the white table! Mark van der Loo, I enjoyed the UROS hackathon on the use of R in Official Statistics, and being member of the international team that you lead. It has been a wonderful experience to meet international colleagues, have fun, collaborate and learn at the same time!

Marie-Ange, I wish to thank you for your long-lasting encouragement and for helping me with the translation of the thesis summary in Dutch.

My two lovely paranymphen, whom I admire and feel lucky to call them friends, Eva and Tsveta. Eva, your positive spirit and your talented personality, seeing you work so hard, was, and continues to be refreshing and inspiring in so many ways. I can tell that our office room is very happy to host your family of plants! Tsveta, I am so glad to have met you at an AMC graduate course. Our dinners, parties, and walks make such nice memories in between our busy schedules during our PhD projects. Your gift, the “no stress” bracelet, will always decorate my desk as a gentle reminder of the needed work-life balance!

My dearest friends, your thoughts and support were a wonderful company throughout my PhD journey. Evangelia, my travel buddy and brilliant friend, thank you for your steady support. I cannot wait to arrange our next trip! Natasa and Georgia, our warm chats, many times from distance, were the best distraction from work. Iro and Kiara, my close friends since middle school, I am lucky to feel your positive vibes for so many years, and to continue sharing our lifechanging experiences. Athina, George, Foteini, and all the "Athenians", our daily chats and sharing our small and big news, made me start my day with a smile. Gaia and Dario, my Italian friends from the Dutch language courses, our dinners and gatherings in Amsterdam were a great pleasure for me.

Last, but not least, nothing can be more important than the incredible support of my family. My parents, Maria and Dimitris, I want to thank you for always being there for me, and encourage me through each and every little step of my PhD journey. Sharing your belief in me and talking about the rewards that come with all the hard work made my worries and anxieties fade away. Christina, my lovely twin-sister, I feel so lucky to share every stage of my life with you! You are my strong pillar and the person who can understand me the most in the world. I know your thoughts can travel to me despite our physical distance. Alex, my greatest supporter, I want to deeply thank you for your understanding for the long working hours, and for your motivating behaviour. You make me believe that everything is possible!

# About the author

Marianna Mitratza was born on the 3rd of March 1989, in Amarousio, Athens, Greece. She completed her secondary education at the Hellenic-American Education Foundation (HAEF) in 2006. Afterwards, Marianna proceeded to pursue her Medical Degree (MD) at the National and Kapodistrian University of Athens (UOA) in 2012, and obtained the licence to practice medicine in 2013. During a clinical research internship in the Pathology Department of Hygeia Hospital in Athens, she analysed patient data and wrote a scientific report on the EGFR mutations in lung cancer.

Marianna's interest in research and epidemiology grew, and in 2013, she was admitted to the two-year Master of Science (MSc) in Biostatistics at the Medical School of UOA in conjunction with the Department of Mathematics of UOA. She completed her thesis on the drug survival and effectiveness of biological agents in patients with psoriasis, while working as a researcher in Clinical Epidemiology and Biostatistics at the Department of Dermatology of Attikon University Hospital in Athens.

In 2015, Marianna moved to the Netherlands for an Erasmus+ traineeship in Pharmaco-epidemiology and Precision Medicine at the University of Utrecht, studying the pharmacogenetics of paediatric asthma exacerbations. In November 2016, she initiated her PhD project as a Doctor-researcher (Arts-onderzoeker) at the Public Health Department of the Academic Medical Center in Amsterdam, focusing on new applications of cause-of-death statistics in public health and epidemiology. As part of her PhD trajectory, Marianna was a co-worker at the Health and Care Department of Statistics Netherlands (CBS), where she worked with multiple Dutch healthcare administrative data.

Marianna is currently continuing her career in research, working as an Assistant Professor at the Department of Epidemiology, Julius Center for Health Sciences and Primary Care, at the University Medical Center Utrecht. She gets inspired following the advances in the field of medical data science and intends to contribute with her research to improving the lives of patients.

# List of publications

## Articles included in this thesis

**Mitratza M**, Kunst AE, Kardaun JWPF. Detecting Mortality Trends in the Netherlands Across 625 Causes of Death. Int J Environ Res Public Health. 2019;16(21):4150. doi: 10.3390/ijerph16214150.

**Mitratza M**, Kardaun JWPF, Kunst AE. How large should a cause of death be in order to be included in mortality trend analysis? deriving a cut-off point from retrospective trend analyses in 21 European countries. BMJ Open. 2020;10:e031702. doi: 10.1136/bmjopen-2019-031702.

**Mitratza M**, Klijs B, Hak AE, Kardaun JWPF, Kunst AE. Systemic autoimmune disease as a cause of death: mortality burden and comorbidities. Rheumatology. 2020; doi: 10.1093/rheumatology/keaa537.

**Mitratza M**, Kunst AE, Harteloh PPM, Nielen MMJ, Klijs B. Prevalence of diabetes mellitus at the end of life: an investigation using individually linked cause-of-death and medical register data. Diab Res Clin Pract. 2020;160(108003). doi: 10.1016/j.diabres.2020.108003.

Klijs B, **Mitratza M**, Harteloh PPM, Moll van Charante EP, Richard E, Nielen MMJ, Kunst AE. Estimating the lifetime risk of dementia using nationwide individually linked cause-of-death and health register data. Int J Epidemiol. 2020;dyaa219. doi: 10.1093/ije/dyaa219.

## Articles not included in this thesis

Dijk FN, Vijverberg SJ, Hernandez-Pacheco N, Repnik K, Karimi L, **Mitratza M**, Farzan N, Nawijn MC, Burchard EG, Engelkes M, Verhamme KM, Potočnik U, Pino-Yanes M, Postma DS, Maitland-van der Zee AH, Koppelman GH. IL1RL1 gene variations are associated with asthma exacerbations in children and adolescents using inhaled corticosteroids. Allergy. 2020;75(4):984-9. doi: 10.1111/all.14125. Epub 2019 Dec 17.

# PhD Portfolio

| | |
|---|---|
| Name PhD candidate: | **Marianna Mitratza** |
| PhD period: | November 2016 – October 2020 |
| Department: | Department of Public Health<br>Amsterdam UMC – University of Amsterdam |
| PhD supervisors: | Prof. dr. Anton E. Kunst,<br>Prof. dr. Jan W.P.F. Kardaun,<br>Dr. Bart Klijs |

| | Organisation | Year | ECTS* |
|---|---|---|---|
| **General courses** | | | **3.8** |
| Project Management | AMC | 2017 | 0.6 |
| Scientific Writing in English for Publication | AMC | 2018 | 1.5 |
| E-Science | AMC | 2018 | 0.5 |
| Research Data Management | AMC | 2018 | 0.5 |
| Medical Literature: EndNote | AMC | 2018 | 0.1 |
| Medical Literature: Embase/Medline via Ovid | AMC | 2018 | 0.1 |
| Unix | AMC | 2018 | 0.5 |
| **Specific courses** | | | **12.2** |
| Clinical Epidemiology: Observational Epidemiology | AMC | 2016 | 0.6 |
| Clinical Epidemiology: Systematic Reviews | AMC | 2017 | 0.7 |
| Clinical Epidemiology: Randomized Clinical Trials | AMC | 2017 | 0.6 |
| Causal inference | Erasmus MC | 2017 | 1.4 |
| Meta-research Short Course | WGH | 2017 | 0.9 |
| The Data Scientist's Toolbox (Coursera certificate) | Johns Hopkins | 2017 | 0.7 |
| R Programming (Coursera certificate) | Johns Hopkins | 2017 | 2.0 |
| Clinical Epidemiology: Evaluation of Medical Tests | AMC | 2018 | 0.9 |
| Utrecht Summer School – Big Data in Health Research | UU | 2018 | 1.4 |
| Getting and Cleaning Data (Coursera certificate) | Johns Hopkins | 2018 | 0.7 |

| | Organisation | Year | ECTS* |
|---|---|---|---|
| Exploratory Data Analysis (Coursera certificate) | Johns Hopkins | 2018 | 1.9 |
| Design and Interpretation of Clinical Trials (Coursera certificate) | Johns Hopkins | 2020 | 0.4 |
| | | | |
| **Seminars, workshops and master classes** | | | **5.3** |
| Anatomy lesson J. Ioannidis: Why most published research findings are false | AMC | 2016 | 0.1 |
| Ruysch Lecture B. Carleton: Precision Medicine – From population to patient-focused drug therapy | AMC | 2017 | 0.1 |
| What values are needed for Personalized Medicine | APH Personalized Medicine | 2017 | 0.1 |
| Annual Dutch Epidemiology congress | WEON | 2017 | 0.6 |
| WEON preconference – A glimpse into the future of epidemiology | WEON | 2017 | 0.1 |
| Masterclass M. Langendam: Using GRADE to rate the certainty in the evidence and strength of recommendations in clinical practice | APH Methodology | 2018 | 0.1 |
| Masterclass D. Ubbink: How to make a decision aid | APH Methodology | 2018 | 0.1 |
| Masterclass P.M.M. Bossuyt: Evaluating predictive markers for precision medicine | APH Methodology | 2018 | 0.1 |
| The Use of R in Official Statistics: uRos2018 | CBS | 2018 | 0.9 |
| Workshop Second-generation p-values: Introduction and applications | AMC | 2019 | 0.1 |
| Annual Dutch Epidemiology congress | WEON | 2019 | 0.6 |
| WEON course – Economic evaluation | WEON | 2019 | 0.2 |
| Workshop IEA Longitudinal cross-study research – opportunities, obstacles and the role of data harmonisation | IEA, SSM | 2019 | 0.1 |
| KEBB/BITE seminars | AMC | 2017-2019 | 2.0 |
| Seminar A. Abu-Hanna: AI and machine learning in medicine: should you care? | APH Methodology | 2020 | 0.1 |

| | Organisation | Year | ECTS* |
|---|---|---|---|
| **Presentations** | | | **2.0** |
| Public Health Department seminar | AMC | 2017 | 0.5 |
| Amsterdam Public Health meeting [poster presentation] | APH | 2017 | 0.5 |
| Amsterdam Public Health meeting [poster presentation] | APH | 2018 | 0.5 |
| The Use of R in Official Statistics: uRos2018 | CBS | 2018 | 0.5 |
| **(Inter)national conferences** | | | **7.5** |
| 10th European Public Health Conference – Stockholm [oral presentation] | EPHA | 2017 | 1.5 |
| 75th Annual Meeting of the American Academy of Dermatology – Orlando [poster presentation] | AAD | 2017 | 1.5 |
| 43rd Panhellenic Medical Conference – Athens [oral presentation] | AMS | 2017 | 1.5 |
| International Epidemiological Association and Society for Social Medicine & Population Health European Congress Annual Scientific Meeting – Cork [oral presentation] | IEA, SSM | 2019 | 1.5 |
| Annual Dutch Epidemiology congress – Groningen [oral presentation] | WEON | 2019 | 1.5 |
| **Other** | | | **1.7** |
| uRos2018 Hackathon project | | | 0.7 |
| Peer reviewing | | | |
| Journal of the American Academy of Dermatology (x3) | | 2017 | 0.3 |
| European Journal of Epidemiology (x2) | | 2019 | 0.2 |
| Journal of the American Academy of Dermatology (x4) | | 2020 | 0.4 |
| International Journal of Epidemiology | | 2020 | 0.1 |

| | Organisation | Year | ECTS* |
|---|---|---|---|
| **Awards and Prizes** | | | |
| IEA and SSM Joint Scientific Meeting,<br>Award for Best rapid fire Presentation – Chronic Disease | IEA, SSM | 2019 | |
| **Total ECTS** | | | **32.5** |

* ECTS = European Credit Transfer and Accumulation system; 1 ECTS = 28 hours
AMC = Academic Medical Center; Erasmus MC = Erasmus Medical Center; WGH = Western General Hospital; UU = Utrecht University; APH = Amsterdam Public Health; WEON = Dutch Epidemiological Conference; CBS = Statistics Netherlands; IEA = International Epidemiological Association; SSM = Society for Social Medicine and Population Health; EPHA = European Public Health Association; AAD = American Academy of Dermatology; AMS = Athens Medical Society