Alzheimer's & Dementia 14 (2018) 130-139

Featured Article

Dementia prevalence and incidence in a federation of European Electronic Health Record databases: The European Medical Informatics Framework resource

Gayan Perera[a], Lars Pedersen[b], David Ansel[c], Myriam Alexander[d], H. Michael Arrighi[e], Paul Avillach[f,g], Nadia Foskett[h], Rosa Gini[i], Mark F. Gordon[j], Usha Gungabissoon[a,k], Miguel-Angel Mayer[l], Gerald Novak[m], Peter Rijnbeek[f], Gianluca Trifirò[f,n], Johan van der Lei[f], Pieter J. Visser[o,p], Robert Stewart[a,q,*]

[a]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
[b]Department of Clinical Medicine, Aarhus University, Aarhus N, Denmark
[c]THIN Contacts, THIN, 1 Canal Side Studios, London, United Kingdom
[d]Real World Data and Health Analytics Department, GSK, Uxbridge, Middlesex, United Kingdom
[e]Janssen Pharmaceuticals Research & Development, Mill Valley, South San Francisco, CA, USA
[f]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands
[g]Department of Biomedical Informatics, Harvard Medical School & Children's Hospital Informatics Program, Boston Children's Hospital, Boston, MA, USA
[h]Roche Products Ltd, Welwyn Garden City, United Kingdom
[i]Agenzia Regionale di Sanità della Toscana, Florence, Italy
[j]Clinical Development and Medical Affairs, Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT, USA
[k]Real World Evidence (Epidemiology), GSK R&D, Uxbridge, Middlesex, United Kingdom
[l]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Catalonia, Spain
[m]Janssen Pharmaceutical Research and Development, Titusville NJ, USA
[n]Dipartimento di Scienze Biomediche, Odontoiatriche e Immagini Morfologiche e Funzionali, Università degli Studi di Messina, Messina, Italy
[o]Alzheimer Centre, School for Mental Health and Neuroscience (MHeNS), University Medical Centre Maastricht, Maastricht University, Maastricht, The Netherlands
[p]Department of Neurology, Alzheimer Center, Neuroscience Amsterdam, VU University Medical Center, Amsterdam, The Netherlands
[q]South London and Maudsley NHS Foundation Trust, London, United Kingdom

Abstract

Introduction: The European Medical Information Framework consortium has assembled electronic health record (EHR) databases for dementia research. We calculated dementia prevalence and incidence in 25 million persons from 2004 to 2012.

Methods: Six EHR databases (three primary care and three secondary care) from five countries were interrogated. Dementia was ascertained by consensus harmonization of clinical/diagnostic codes. Annual period prevalences and incidences by age and gender were calculated and meta-analyzed.

Results: The six databases contained 138,625 dementia cases. Age-specific prevalences were around 30% of published estimates from community samples and incidences were around 50%. Pooled prevalences had increased from 2004 to 2012 in all age groups but pooled incidences only after age 75 years. Associations with age and gender were stable over time.

Discussion: The European Medical Information Framework initiative supports EHR data on unprecedented number of people with dementia. Age-specific prevalences and incidences mirror estimates from community samples in pattern at levels that are lower but increasing over time.

*Corresponding author. Tel.: +44-20-7848-0778; Fax: +44-20-7848-5408.

E-mail address: robert.stewart@kcl.ac.uk

http://dx.doi.org/10.1016/j.jalz.2017.06.2270

1552-5260/ © 2017 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

© 2017 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Dementia; Incidence; Prevalence; Electronic Health Records; European Medical Informatics Framework

1. Introduction

Electronic health records (EHRs) are increasingly replacing paper records across health care sectors, creating large volumes of digitized data on real-world clinical interventions and outcomes. Applications of such data extend beyond clinical care to planning and costing health services, and health surveillance. For research, EHR-derived databases offer large sample sizes and are particularly suited to investigations of intervention outcomes in routine care, such as predictors of response, safety, comparative effectiveness, and health economic evaluations, as well as etiologic investigations of rare exposures/outcomes.

EHRs have been underused in dementia research, although examples include identification in primary care [1], resource use in Alzheimer's disease [2], comorbidities [3], case capture efficiency [4], dementia incidence [5], dementia-free life expectancy [6], risks associated with medication exposures [7] and other disorders [8], atypical antipsychotics and mortality in vascular dementia [9], and cognitive trajectories before and after acetylcholinesterase inhibitor initiation [10].

EHR data on dementia, as with most clinical research data, are limited to people who have received a diagnosis. These are a subset of people living with dementia in the community, many of whom will not receive a diagnosis and can only be ascertained in surveys applying case-ascertainments. However, epidemiologic studies of dementia tend to focus on incidence and risk factors (i.e., investigating up to the point of onset), and most research on dementia outcome, including service costs, involves cohorts who have received a clinical diagnosis rather than screened community samples. The relationships between community incidence/prevalence and incidence/prevalence of diagnosed dementia are therefore important to understand because of these different samples. Such relationships may vary both geographically and temporally: settings with active dementia diagnosis centers are likely to see a higher proportion of community cases appear on health care databases, and these are likely to increase over time with greater public awareness and a higher salience in primary care.

Advances in medical research require an increasing quantity and detail of health data to answer today's complex questions. At the same time, huge volumes of health data are being collected and electronically stored, either in routine EHR databases or through research-driven cohort studies associated with biobanks and other efforts. To help improve access to patient-level data, the European Medical Information Framework (EMIF) was launched in 2013. As part of the broader European Innovative Medicines Initiative, EMIF aims to create an environment that allows efficient reuse of health data in two therapeutic areas: Alzheimer's disease and metabolic disorders. In the first year of operation, EMIF explored several EHR resources to render data available for analysis with robust data security and governance. In this study, we provide initial data on number of cases of dementia detected on six EHR sources—three primary care and three secondary care. Our objective was partly to describe this data resource and partly to provide prevalence and incidence estimates for diagnosed dementia in European populations, comparing these across the different databases and years of data collection.

2. Methods

2.1. Databases

Using the collaborations and platform for analysis set up by EMIF, the following six EHR databases were interrogated: (1) Agenzia regionale di sanità della Toscana (ARS), (2) Aarhus University Hospital (AUH), (3) the Health Search Database (GENOMEDICS), (4) the Information System of Parc de Salut Mar (IMIM-UPF), (5) Integrated Primary Care Information (IPCI), and (6) The Health Improvement Network (THIN). Characteristics of databases are summarized in Table 1. Number of active patients on each database on January 1, 2013 varied from around 513,000 in IMIM-UPF to around 3.8 million in THIN.

2.1.1. Catchment databases

The ARS database contains secondary health care data from the Tuscany region of Italy, including pharmacy, outpatient and hospital data along with linked data on death, and birth and malformation registries. The AUH database contains data from secondary health care registries covering the northern and central region of Jutland, Denmark, including information on all inpatient, emergency room, and outpatient visits, inpatient treatments, laboratory data, and prescriptions. The IMIM-UPF database contains secondary health care data from the Barcelona area, including acute, long term, and mental health services with information on inpatient, emergency room, and outpatient visits.

2.1.2. Primary care databases

IPCI, GENOMEDICS, and THIN contain primary care data from selected general practices in the Netherlands, Italy, and United Kingdom, respectively.

Considering broader health care settings, Italy has a government-funded universal health care system in which free health care is provided to all citizens, all persons in Italy

Table 1
Characteristics of the electronic health record databases participating in the study

Characteristics	ARS	AUH	GENOMEDICS	IMIM-UPF	IPCI	THIN
Geographic origin of data	Tuscany, Italy	Northern and central region of Jutland, Denmark	Italy	Barcelona, Spain	Netherlands	UK (England, Northern Ireland, Scotland, and Wales)
Source of data	Administrative/billing data on general pharmacy; inhabitant registry; outpatient care; hospital admission	Emergency ward; general pharmacy; hospital admission; inhabitant registry	Primary care records	Emergency ward; hospital admission; outpatient care	Primary care records	Primary care records
Year of initial data collection	1996	1996	1998	1990	1995	2002
Reasons for entry/exit	Patient moves in or out a specific geographical area	Patient moves in or out a specific geographical area	Patient enters or leaves the practice of a GP	Hospital admission/ discharge	Patient enters or leaves the practice of a GP	Patient enters or leaves the practice of a GP
Approximate total (cumulative) number of patients	5 millions	2.3 million	2.3 million	More than 1 million	2.8 million	12 million
Approximate number of active patients as at January 2013	3.6 million	1.8 million	900,000	513,000	1.8 million	3.8 million
Average follow-up period per patient in years	9 y	13 y	10 y	4.75 y	3 y, but with a wide range	Median follow-up of active patients is 9 y
Primary sources of clinical diagnoses	Death registry; inpatient care; exemptions from copayment	Outpatient care; inpatient care; emergency care; death registry; procedure registry	Primary care	Death registry; procedure registry; inpatient care; emergency care; outpatient care	Primary care and communications to/ from secondary care	Primary care
Terminology system(s) used to store diagnoses	ICD-9	ICD-10	ICD-9	ICD-9	Free text, codes from Dutch College of General Practitioners, International Classification of Primary Care	READ codes, free text
Death recorded	Yes	Yes	Yes	Yes	Yes	Yes

Abbreviation: GP, general practitioner.

NOTE. Databases: (1) Agenzia regionale di sanità della Toscana (ARS), (2) Aarhus University Hospital (AUH), (3) the Health Search Database (GENOMEDICS), (4) the Information System of Parc de Salut Mar (IMIM-UPF), (5) Integrated Primary Care Information (IPCI), (6) The Health Improvement Network (THIN).

Table 2
Definition of dementia incidence and prevalence used in each study database

Type of measure	ARS	AUH	GENOMEDICS	IMIM-UPF	IPCI	THIN
Dementia case (numerator)	Dementia recorded in hospital (ICD-9). Includes inpatient only	Dementia recorded in hospital (ICD-10). Included inpatient and outpatients	Dementia recorded in primary care records (ICD-9)	Dementia recorded in hospital (ICD-9). Include inpatient records only	Dementia recorded in primary care records (ICPC + free text)	Dementia recorded in primary care records (READ codes used—not free text)
Reference population	Patients living in Tuscany, a geographical area of Italy	Patients living in a geographical area/ region of Denmark	Primary care patients (cover around 1000 GPs in Firenze, Italy	All hospitalized patients in Parc Salut Mar, Barcelona area	Primary care patients in Holland. Restricted calculation to GP practices with at least 12 mo of reliable records	Primary care patients in the UK

Abbreviation: ICPC, International Classification of Primary Care; GP, general practitioner.
NOTE. Databases: (1) Agenzia regionale di sanità della Toscana (ARS), (2) Aarhus University Hospital (AUH), (3) the Health Search Database (GENOMEDICS), (4) the Information System of Parc de Salut Mar (IMIM-UPF), (5) Integrated Primary Care Information (IPCI), (6) The Health Improvement Network (THIN).

are registered with a general practitioner (GP) who acts as a gatekeeper for all primary care health care services, the system is mostly region-based although citizens may access (free of charge) any hospital in Italy. The Danish National Health Service provides universal tax-supported health care, guaranteeing unfettered access to GPs and hospitals, and partial reimbursement for prescribed medications, and accurate linkage of all registries at the individual level is possible in Denmark using the unique Central Personal Register number assigned to each Danish citizen at birth and to residents on immigration. The Spanish health care system consists of both private and public health care (although more than 90% of the population makes use of the public system for their medical care), the financing of the health care system is the responsibility of the National Health System and the autonomous communities and is based on a universal coverage model, although information linkage between health care levels and autonomous communities' EHR systems is not complete. In the Netherlands, all citizens are registered with one GP, who forms the first point of care for all medical complaints, and electronic medical records contain all journal entries written by the GPs, and coded and anonymous data on patient demographics, diagnoses using the International Classification for Primary Care, referrals, laboratory findings, and drug prescriptions. The UK has a government-funded National Health Service delivering health care free at the point of delivery, although with standard charges for some services (e.g., dentistry) and prescription medication, and with GP referral required for all nonemergency secondary care.

2.2. Dementia definitions

Codes used to identify dementia are detailed in Appendix A. The aim for the databases was to capture as broad as possible a range of indicators, restricting to those that clearly indicated a dementia diagnosis, rather than those that were suggestive. Therefore we opted not, at this stage, to define dementia subgroups, on the assumption that more specific diagnoses would be more likely to have higher heterogeneity. The process of code assembly was a reciprocal one between academics and database holders, based on knowledge of dementia diagnoses and of local clinical practice. The Unified Medical Language System was used to identify medical concepts and corresponding codes in each terminology, and a common database model was used to share and pool data and verify the semantic basis of the event extraction queries [11]. Beginning with long lists of terms from constituent databases and diagnostic systems, feedback from the database holders was obtained at several stages to refine the extraction queries, and definitions were finalized by consensus.

2.3. Prevalence and incidence calculations

Definitions used for case prevalence and incidence estimations are summarized in Table 2. We computed database- and year-specific annual incidences and period

Table 3
Distribution of data availability before and after the index date (first instance of recorded dementia) for cases in EMIF-AD EHR databases

Database	Number with dementia 2004–2012	Years of data before diagnosis for individuals with dementia				Years of data after diagnosis for individuals with dementia			
		Mean	SD	Range	Total person-years	Mean	SD	Range	Total person-years
ARS	52,296	4.66	2.65	0–9	244,828	2.09	2.18	0–9	109,649
AUH	20,765	4.88	2.82	0–10	111,976	2.67	2.22	0–10	61,300
GENOMEDICS	13,826	5.52	3.40	0–12	98,427	3.35	2.70	0–12	59,700
IMIM-UPF	1566	6.40	4.84	0–22	17,936	0.85	1.62	0–14	2390
IPCI	6128	2.29	1.81	0–11	19,814	1.46	1.26	0–9	12,583
THIN	44,044	8.13	5.71	0–26	556,964	2.28	2.25	0–22	156,202
Combined data	138,625	5.80*	3.71*	0–26	1,049,945	2.32*	2.21*	0–14	401,824

Abbreviations: EHR, electronic health record; EMIF, European Medical Information Framework; AD, Alzheimer's disease; SD, standard deviation.
NOTE. Databases: (1) Agenzia regionale di sanità della Toscana (ARS), (2) Aarhus University Hospital (AUH), (3) the Health Search Database (GENOMEDICS), (4) the Information System of Parc de Salut Mar (IMIM-UPF), (5) Integrated Primary Care Information (IPCI), (6) The Health Improvement Network (THIN).
*Arithmetically weighted by case numbers.

prevalences by age and gender for each database and for each calendar year, restricting the analyses to the calendar years 2004 to 2012 when data were available from all sources. Denominator populations for the three catchment databases (ARS, AUH, and IMIM-UPF) are defined from resident lists (inhabitant registries; see Table 1), regardless of health care contact and were ascertained for the calendar years described in this analysis based on routinely recorded and updated entry/exit dates (from the catchment) and national mortality records. Denominator populations for the primary care databases consist of total populations registered, again updated for each year of analysis by entry and exit dates (including mortality). The date of the first recorded dementia coding for a given person within each database (across all years of data since initial collection—see Table 1) was therefore ascertained and these provided the numerator population for incidence estimates (i.e., for the respective calendar years of the first dementia coding), with the denominator population defined for that year in the catchment or primary care database, having taken into account entry and exit dates. Each case of dementia then contributed to prevalence estimates (applying the same year-specific denominators) from the age at first diagnosis to the age at exit. Denominator populations were further subclassified by age ranges at the mid-point for each year and both numerator and denominator samples were restricted to people aged 50 years or older at respective time points.

2.4. Analysis

The following output was generated and displayed graphically: (1) case prevalence and incidence by age group and year for each database; (2) case prevalence and incidence by age group and database for each year of calculation; and (3) female to male ratios of prevalence and incidence by year for each database. As an initial summary, a table was generated, which contained annual period prevalence, annual incidence, and female to male ratio statistics of these two estimates, by age group and database, averaged across the 9 years of observation. Prevalences and incidences were then meta-analyzed across databases using random-effects models (because of observed high levels of heterogeneity) and pooled estimates graphically displayed by age group and year. This was followed by a meta-analysis of individual linear regression coefficients of logged prevalence or incidence outcomes against age group increments, approximate estimates of associations with age. I^2 statistics were calculated in all instances as estimates of heterogeneity. Final analyses estimated case duration (approximating survival after diagnosis) by dividing prevalence and annual incidence (percentages), displaying these for age groups by database and year.

3. Results

Numbers of cases detected with dementia in each database from 2004 to 2012 are displayed in Table 3. A total of 138,625 cases were represented, varying from 1566 in the IMIM-UPF database to 52,296 in ARS. Illustrative data on periods of time represented before and after the first dementia diagnosis are likewise displayed. Overall, the total number of person-years was estimated to be more than 1 million before the first diagnosis and more than 400,000

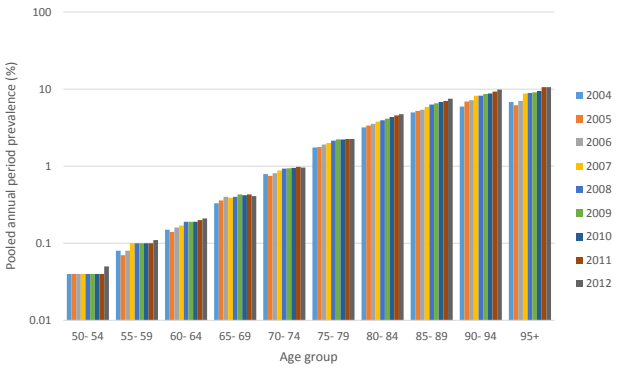

Fig. 1. Pooled dementia prevalence by age group and year. A logarithmic y-axis scale is used to enable relative time trends to be compared between age groups.

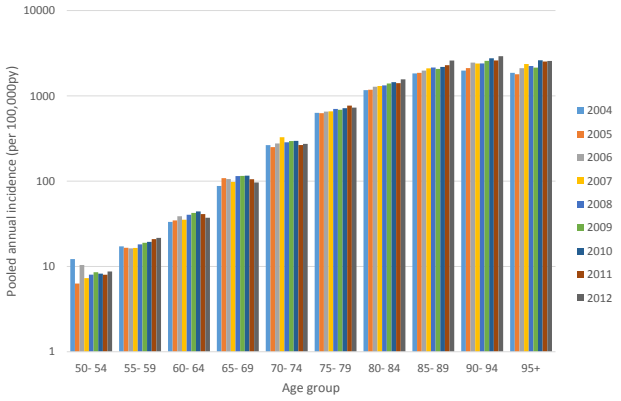

Fig. 2. Pooled dementia incidence by age group and year. A logarithmic y-axis scale is used to enable relative time trends to be compared between age groups.

after this date; mean follow-up periods for the six databases averaged 5.3 and 2.1 years, respectively.

Pooled prevalences and incidences of dementia by age group and year, meta-analyzed across databases are displayed in Figs. 1 and 2 and detailed in Table B1 of Appendix B. Because we wanted to visualize relative changes in estimates over the sampled years, logarithmic scales were used for the y-axes to illustrate such changes in different age groups. High levels of heterogeneity were found for all cells: I^2 statistics were >95% for all prevalence estimates and >90% for all incidence estimates apart from those in the 50- to 54-year-old age group, which ranged from 50% to 87%. For dementia prevalence estimates, reasonably consistent increases across the years analyzed were observed in most age groups; however, these period effects were only evident for dementia incidence estimates in older age groups. Pooled associations with age are displayed in Table B2 of Appendix B. From I^2 estimates, considering the final column, coefficients showed a high level of homogeneity between years for all databases apart from IMIM-UPF (which also had the smallest slope coefficients). On the other hand, considering pooled estimates by year, between-database heterogeneity was frequently marked, tended to be lowest in earlier years and highest in more recent years. Considering the coefficients themselves, there were no noticeable trends across the years analyzed.

Estimated annual period prevalence and incidence of dementia averaged across the observed years and stratified by database are given in Table 4; they are further displayed by age and year for each database in Figs. C1 and C2 of Appendix C and by age and database for each year in Figs. C3 and C4 of Appendix C. All databases showed increases in prevalence and incidence by age and generally patterns were similar, particularly across the larger databases. Most databases also appear to show a leveling off of prevalence and incidence trajectories in the oldest age group—in that estimates for the 95+ age group are lower than would be expected from curve trajectories in preceding age groups. Prevalences varied around twofold across databases: for example for 85- to 89-year olds, mean 1-year period prevalences ranged from 4.9% to 9.4% and mean annual incidences from 1793 to 3650 per 100,000 person-years (Table 4). Female to male prevalence ratios are displayed in Fig. C5 of Appendix C; these showed a modest increase across age groups from equality in younger age groups to a female predominance in older age groups, but little or no change over the years sampled, particularly in the larger data sets. Female to male incidence ratios (Fig. C6 of Appendix C) were closer to unity, did not change substantially over time, and only increased slightly in strength across the age range.

Estimated case duration is summarized in Appendix D. Modest increases were seen over time for most age groups, of the order of half a year improvement from 2004 to 2012 (Table D1 and Fig. D1 of Appendix D). Marked heterogeneity, however, was observed between databases with lowest duration in IMIM-UPF and highest in GENOMEDICS. Associations with age were not consistent between databases.

4. Discussion

The EMIF consortium has assembled what we believe may be the world's largest EHR resource for dementia research, and the primary objective of this initial study was to describe prevalences and incidences of cases across data sets and years of data collection. Random-effects meta-analyses indicated increasing prevalences across the 9 years analyzed but less marked increases in incidences apart from in older age groups, high between-database heterogeneity for prevalence/incidence estimates, and moderate heterogeneity for age-associated trajectories (highest in more recent years), but homogeneity of between-year age-associated trajectories. Gender ratios were relatively stable by database and time. Estimated case duration increased over time, did not vary consistently with age, and was heterogeneous between databases.

A recent systematic review of community surveys in Europe estimated dementia prevalences (%) to be 0.2, 1, 4, 8, 13, 21, 32, and 45 in 60 to 65, 65 to 70, 70 to 75, 75 to 80, 80 to 85, 85 to 90, 90 to 95, and 95+ age groups, respectively [12]. Taking the most recent 2012 data, our pooled dementia prevalence estimates from EHRs are 0.2, 0.4, 1, 2, 5, 8, 10, and 11 for these respective age groups—i.e., 100%, 40%, 25%, 25%, 38%, 38%, 31%, and 24% of estimates from community samples. These discrepancies are more marked than a 50% underestimation suggested by other direct comparisons of administrative and community data [13], but are similar to 2005 estimates from Danish hospital registers [14]. Our dementia incidence rates were also similar to these Danish data [14] and to ranges reported for people aged 65+ from US health care data [5]. Considering community studies of dementia incidence, estimates from EURODEM were 205, 489, 1623, 2975, and 5356 per 100,000 person-years in age groups 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85 to 89, respectively [15]. Our 2012 pooled estimates for these age groups were 96, 274, 728, 1564, and 2595 respectively—i.e., 47%, 56%, 45%, 53%, and 48% of the community estimates.

Table 4
Averaged annual period prevalence and incidence between 2004 and 2012 by age and EHR database

			Averaged annual estimates from 2004 to 2012					
Database	Age	Total number of dementia cases from 2004 to 2012	Annual period prevalence (%) (95% CI)		Female to male prevalence ratio	Annual incidence rate/ 100,000 person-years (95% CI)		Female to male incidence ratio
ARS	<70	2659	0.1	(0.1–0.2)	0.9	30	(29–32)	0.8
	70–74	3608	0.6	(0.6–0.7)	0.9	188	(182–194)	0.8
	75–79	7926	1.5	(1.4–1.5)	1.1	479	(469–490)	1.0
	80–84	13,917	3.0	(2.9–3.1)	1.3	1046	(1029–1063)	1.1
	85–89	14,212	5.3	(5.1–5.4)	1.3	1898	(1867–1929)	1.1
	90–94	7574	6.9	(6.6–7.1)	1.4	2464	(2409–2520)	1.1
	95+	2400	7.1	(6.7–7.5)	1.4	2004	(1925–2085)	1.1
AUH	<70	2183	0.3	(0.3–0.3)	0.7	58	(56–60)	0.8
	70–74	2188	1.4	(1.3–1.5)	1.0	383	(367–399)	1.0
	75–79	3965	3.2	(3.0–3.3)	1.2	915	(887–944)	1.1
	80–84	5509	6.1	(5.8–6.3)	1.3	1789	(1743–1837)	1.1
	85–89	4635	9.4	(9.0–9.8)	1.3	2637	(2562–2714)	1.1
	90–94	1844	10.8	(10.2–11.5)	1.4	2628	(2511–2751)	1.1
	95+	441	10.1	(8.9–11.4)	1.4	2384	(2171–2617)	1.1
GENOMEDICS	<70	1139	0.2	(0.2–0.2)	1.2	45	(43–48)	1.2
	70–74	1568	1.3	(1.2–1.4)	1.3	310	(295–326)	1.4
	75–79	2862	3.1	(3.0–3.3)	1.4	685	(660–710)	1.3
	80–84	3774	5.7	(5.5–6.0)	1.4	1222	(1183–1261)	1.3
	85–89	2854	8.7	(8.3–9.1)	1.4	1793	(1729–1860)	1.2
	90–94	1342	10.7	(10.0–11.4)	1.4	2215	(2100–2337)	1.1
	95+	287	12.1	(10.5–13.9)	1.6	2688	(2394–3017)	1.3
IMIM-UPF	<70	129	0.3	(0.1–0.7)	0.6	208	(175–247)	0.7
	70–74	118	0.7	(0.4–1.0)	0.7	454	(379–543)	0.9
	75–79	304	1.7	(1.3–2.2)	1.0	1049	(938–1174)	0.9
	80–84	429	3.0	(2.4–3.7)	1.2	1937	(1762–2129)	1.1
	85–89	400	4.9	(3.8–6.2)	1.6	3650	(3309–4026)	1.2
	90–94	147	7.7	(5.4–10.6)	2.0	4427	(3766–5204)	1.4
	95+	39	10.9	(5.7–18.7)	0.8	5513	(4028–7545)	0.6
IPCI	<70	659	0.2	(0.2–0.3)	1.0	69	(64–74)	1.0
	70–74	632	1.4	(1.2–1.6)	1.0	436	(404–472)	1.0
	75–79	1210	3.0	(2.7–3.3)	1.0	1076	(1017–1138)	1.1
	80–84	1632	5.7	(5.3–6.2)	1.1	2085	(1987–2189)	1.1
	85–89	1283	8.2	(7.4–8.9)	1.1	3012	(2852–3182)	1.1
	90–94	579	10.4	(9.1–11.9)	1.1	3968	(3658–4305)	1.1
	95+	133	11.0	(8.3–14.3)	1.7	4043	(3411–4792)	1.3
THIN	<70	3145	0.1	(0.1–0.2)	1.0	38	(37–40)	1.0
	70–74	4055	0.9	(0.9–1.0)	1.1	292	(283–301)	1.1
	75–79	8001	2.2	(2.1–2.3)	1.2	708	(692–723)	1.1
	80–84	11,477	4.6	(4.4–4.7)	1.4	1389	(1364–1414)	1.2
	85–89	10,739	7.9	(7.7–8.1)	1.6	2259	(2216–2302)	1.2
	90–94	5125	11.0	(10.6–11.4)	1.6	2711	(2638–2786)	1.3
	95+	1502	12.2	(11.5–13.0)	1.7	2638	(2508–2775)	1.3

Abbreviations: CI, confidence interval; EHR, electronic health record.

NOTE. Databases: (1) Agenzia regionale di sanità della Toscana (ARS), (2) Aarhus University Hospital (AUH), (3) the Health Search Database (GENOMEDICS), (4) the Information System of Parc de Salut Mar (IMIM-UPF), (5) Integrated Primary Care Information (IPCI), (6) The Health Improvement Network (THIN).

Differences in dementia prevalence and incidence can be expected between EHR databases and screened community samples. On the one hand, many people with dementia do not receive a diagnosis, and not all diagnoses are recorded in an accessible (i.e., structured) or sufficiently specific format within an EHR. On the other hand, community surveys often do not sample long-term care facilities, where dementia is highly prevalent [16] and whose residents are more likely to be represented in EHRs. Of interest, in the EMIF EHR samples there was more underestimation of community dementia prevalence than incidence. This may reflect different underlying reasons for discrepancies: prevalence underestimates being accounted for by underdiagnosis and incidence underestimates by differential survival (i.e., prediagnosis mortality). An underestimated prevalence would have resulted from an overestimated denominator population, whereas the opposite would be more likely if there were problems with enumeration. An overestimated denominator might have resulted from cases being seen in alternative facilities, although this would have given rise to a more visible difference between primary care and hospital databases (because diagnoses would be expected to be communicated to GPs regardless of their source).

Considering survival, postdiagnosis case duration was around 3 years in most age groups, which is shorter than the 5 to 9 years estimated from first symptom emergence [17,18] and which might be accounted for by delayed diagnosis, although might also reflect the capture of people with rapid-progression disorders missed by conventional cohort studies [19].

Between-database heterogeneity was observed for dementia prevalence and incidence estimates in all age groups. This is likely to arise from between-database differences in clinical ascertainment and/or recording of dementia rather than from underlying differences in risk. Considering data sources, the lowest prevalences in later years were consistently found in two secondary care databases (ARS and IMIM-UPF), which also tended to have the highest incidences, and later diagnosis and lower case survival was evident for IMIM-UPF at least; the other secondary care database (AUH) showed estimates much closer to those from primary care, possibly reflecting the additional use of outpatient records for case definition (Table 2). The consistently higher prevalence of dementia at the uppermost end of the age range in two primary care databases (GENOMEDICS and THIN) might reflect institutional facilities served by participating practices but might also reflect the high postdiagnosis case duration in these databases (Fig. D2 of Appendix D).

Danish register data [14] showed an upward trend in dementia incidence and prevalence over time, assumed to be due to improved access to diagnosis. Similarly, dementia diagnoses recorded on death certificates have been increasing [20,21]. Again, time trends are likely to be determined by health service access/responsiveness rather than changes in risk. In the EMIF databases, age-specific prevalences of dementia increased across the 9 years analyzed in all age groups; however, incidence rates only increased in older age groups. One possibility is that dementia ascertainment by clinical services in younger age groups has remained constant, but case survival has been improving, accounting for increasing prevalence but stable incidence; in older age groups increases in both ascertainment/recording and case survival would account for increases in both prevalence and incidence. Increases by year in case duration were observed across most age groups, consistent with this and with at least one previous report of improvement over time in dementia survival [22].

Despite lower prevalence/incidence, dementia associations with age and gender in the EMIF EHR databases were similar to community estimates, for example, increasing exponentially with age. Although underestimations of community incidence were relatively constant across age groups, these varied more for prevalence, being strongest in 70- to 80-year olds. Considering the most recent year of data, only two EHRs (THIN and GENOMEDICS, both primary care) showed continued increases in prevalence into the 95+ age group and only one (GENOMEDICS) showed this for incidence (Figs. C3 and C4 of Appendix C); however, it should be borne in mind that cell sizes in the 95+ age group are likely to have given unstable estimates in the smaller databases (particularly IMIM-UPF and IPCI), although ought to be robust in the larger ones (ARS and THIN). Age-specific dementia prevalence is generally found to be higher in women than men in community studies, although differences range from 1% to 2% [13,23–25] to a doubling or more [26–28]. Female to male prevalence ratios in the EHR data rose with age, whereas incidence ratios rose more modestly; this would be consistent with increased duration of diagnosed dementia in older women because of lower case mortality.

A major strength of these EHR resources is that they cover large populations, with data accruing over long periods. However, it is important to bear in mind the limitations of routine data. In particular, dementia diagnosis and recording may vary substantially between individuals, localities and internationally, influenced, for example, by clinician attitudes to the usefulness of the diagnosis, its requirement or not for treatment initiation, family pressures to assign (or not) a diagnosis, and the clinical or research criteria adopted. In addition, EHRs from nonspecialist care do not generally contain data (e.g., from cognitive/functional assessments), which indicate the dementia severity at diagnosis. Despite this, there are important potential research applications. Investigations of prognosis, impact, and intervention response are perhaps the most important, because alternative data sources (randomized trials and conventional clinical cohorts) are limited in size and generalizability. However, there are also possibilities for novel risk factor research, particularly where these tend to be well recorded in health records and missed in traditional cohort studies. Medication exposures and comorbidities are the most available data; however, data linkages offer novel additional possibilities in individual settings. For example, AUH can be linked to data on education, school performance, income, employment, and retirement, to an increasing extent over time, as well as potentially with comparable spouse and sibling data; THIN data can likewise be linked to those from the spouse, and thus, for example, there are important opportunities for investigating the shared environment or the influence of one person's dementia on another's.

In summary, we describe trends in dementia prevalence and incidence extracted from what collectively may form the world's largest database on this condition. EHR data contain a wealth of longitudinal information on very large samples, although limitations inherent in administrative information resources need to be borne in mind when studies are designed to use these.

Acknowledgments

Funding: The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under European Medical Informatics Framework (EMIF) grant agreement no. 115372, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in kind contribution. R.S. and G.P. are partly funded by the National Institute for Health Research

(NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London.
Authors' contributions: Analyses were carried out by L.P. and G.P. The manuscript was written by G.P. and finalized by R.S. and with substantial text contribution from all authors. All other authors contributed to the conceptualization, discussion, and design of this project, review and critical comments of draft manuscript, and review and approval of final manuscript.
Authors' information: M.A., H.M.A., N.F., M.F.G., U.G. and G.N. are or have been full-time salaried employees of their respective institutions. R.S. has received research funding from Pfizer, Lundbeck, Roche, GSK, and Janssen.
The authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE) and were fully responsible for all aspects of manuscript development. The authors received compensation (i.e., salary) as employees of their respective organizations. Any views expressed in this publication represent the personal opinions of the authors and not those of their respective employer. The authors' respective organizations were given the opportunity to review the manuscript for medical and scientific accuracy as well as intellectual property considerations.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2017.06.2270.

RESEARCH IN CONTEXT

1. Systematic review: Using PubMed for English language publications, we identified and reviewed all publications on incidence and prevalence of dementia in Europe, focusing on health records databases.
2. Interpretation: Analyses of 138,625 dementia cases demonstrated age-associated increases in prevalence and incidence followed by a flattening of trajectories in oldest age groups, modest increases in prevalence across the 9 years analyzed but relatively stable incidence and strengths of association with age and gender.
3. Future directions: The European Medical Information Framework consortium has assembled what we believe may be the world's largest electronic health record resource for dementia research. This presents potentially transformative opportunities in terms of sample size and naturalistic follow-up, although strengths and limitations of administrative data need to be carefully considered in designing future projects using this resource.

References

[1] Williamson T, Michael EG, Birtwhistle R, Khan S, Garies S, Wong ST, et al. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. Ann Fam Med 2014;12:367–72.
[2] Chen L, Reed C, Happich M, Nyhuis A, Lenox-Smith A. Health care resource utilisation in primary care prior to and after a diagnosis of Alzheimer's disease: a retrospective, matched case-control study in the United Kingdom. BMC Geriatr 2014;14:76.
[3] Poblador-Plou B, Calderón-Larrañaga A, Marta-Moreno J, Hancco-Saavedra J, Sicras-Mainar A, Soljak M, et al. Comorbidity of dementia: a cross-sectional study of primary care older patients. BMC Psychiatry 2014;14:84.
[4] Francesconi P, Gini R, Roti L, Bartolacci S, Corsi A, Buiatti E. The Tuscany experimental registry for Alzheimer's disease and other dementias: how many demented people does it capture? Aging Clin Exp Res 2007;19:390–3.
[5] Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. Alzheimers Dement 2016;12:216–24.
[6] Sauvaget C, Tsuji I, Haan MN. Trends in dementia-free life expectancy among elderly members of a large health maintenance organization. Int J Epidemiol 1999;28:1110–8.
[7] Gray SL, Anderson ML, Dublin S, Hanlon JT, Hubbard R, Walker R, et al. Cumulative use of strong anticholinergic medications and incident dementia. JAMA Intern Med 2015;175:401–7.
[8] Exalto LG, Quesenberry CP, Barnes D, Kivipelto M, Biessels GJ, Whitmer RA. Midlife risk score for the prediction of dementia four decades later. Alzheimers Dement 2014;10:562–70.
[9] Sultana J, Chang CK, Hayes RD, Broadbent M, Stewart R, Corbett A, et al. Associations between risk of mortality and atypical antipsychotic use in vascular dementia: a clinical cohort study. Int J Geriatr Psychiatry 2014;29:1249–54.
[10] Perera G, Khondoker M, Broadbent M, Breen G, Stewart R. Factors associated with response to acetylcholinesterase inhibition in dementia: a cohort study from a secondary mental health care case register in London. PLoS One 2014;9:e109484.
[11] Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour JC, et al.EU-ADR consortium. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. J Am Med Inform Assoc 2013;20:184–92.
[12] Alexander M, Perera G, Ford L, Arrighi HM, Foskett N, Debove C, et al. Age-stratified prevalence of mild cognitive impairment and dementia in European populations: a systematic review. J Alzheimers Dis 2015;48:355–9.
[13] Eefsting JA, Boersma F, Van den Brink W, Van Tilburg W. Differences in prevalence of dementia based on community survey and general practitioner recognition. Psychol Med 1996;26:1223–30.
[14] Phung TK, Waltoft BL, Kessing LV, Mortensen PB, Waldemar G. Time trend in diagnosing dementia in secondary care. Dement Geriatr Cogn Disord 2010;29:146–53.
[15] Launer LJ, Andersen K, Dewey ME, Letenneur L, Ott A, Amaducci LA, et al. Rates and risk factors for dementia and Alzheimer's disease: results from EURODEM pooled analyses. EURODEM Incidence Research Group and Work Groups. European Studies of Dementia. Neurology 1999;52:78–84.
[16] Riedel-Heller SG, Schork A, Matschinger H, Angermeyer MC. Recruitment procedures and their impact on the prevalence of dementia. Results from the Leipzig Longitudinal Study of the Aged (LEILA75+). Neuroepidemiology 2000;19:130–40.
[17] Mölsa PK, Marttila RJ, Rinne UK. Survival and cause of death in Alzheimer's disease and multi-infarct dementia. Acta Neurol Scand 1986;74:103–7.

[18] Walsh JS, Welch HG, Larson EB. Survival of outpatients with Alzheimer-type dementia. Ann Intern Med 1990;113:429–34.
[19] Wolfson C, Wolfson DB, Asgharian M, M'Lan CE, Østbye T, Rockwood K, et al. A reevaluation of the duration of survival after the onset of dementia. N Engl J Med 2001;344:1111–6.
[20] Kochanek KD, Murphy SL, Xu J, Arias E. Mortality in the United States, 2013. NCHS Data Brief 2014;178:1–8.
[21] Perera G, Stewart R, Higginson IJ, Sleeman KE. Reporting of clinically diagnosed dementia on death certificates: retrospective cohort study. Age Ageing 2016;45:668–73.
[22] Beard CM, Kokmen E, O'Brien PC, Kurland LT. Are patients with Alzheimer's disease surviving longer in recent years? Neurology 1994;44:1869–71.
[23] Benedetti MD, Salviati A, Filipponi S, Manfredi M, De Togni L, Gomez Lira M, et al. Prevalence of dementia and apolipoprotein e genotype distribution in the elderly of Buttapietra, Verona province, Italy. Neuroepidemiology 2002;21:74–80.
[24] Prencipe M, Casini AR, Ferretti C, Lattanzio MT, Fiorelli M, Culasso F. Prevalence of dementia in an elderly rural population: effects of age, sex, and education. J Neurol Neurosurg Psychiatry 1996;60:628–33.
[25] Rodríguez-Sánchez E, Mora-Simón S, Patino-Alonso MC, García-García R, Escribano-Hernández A, García-Ortiz L, et al. Prevalence of cognitive impairment in individuals aged over 65 in an urban area: DERIVA study. BMC Neurol 2011;11:147.
[26] Ravaglia G, Forti P, Maioli F, Sacchetti L, Mariani E, Nativio V, et al. Education, occupation, and prevalence of dementia: findings from the Conselice study. Dement Geriatr Cogn Disord 2002;14:90–100.
[27] De Deyn PP, Goeman J, Vervaet A, Dourcy-Belle-Rose B, Van Dam D, Geerts E. Prevalence and incidence of dementia among 75-80-year-old community-dwelling elderly in different districts of Antwerp, Belgium: the Antwerp Cognition (ANCOG) Study. Clin Neurol Neurosurg 2011;113:736–45.
[28] Ott A, Breteler MM, van Harskamp F, Claus JJ, van der Cammen TJ, Grobbee DE, et al. Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study. BMJ 1995;310:970–3.

Did you know?

You can track the impact of your article with citation alerts that let you know when your article (or any article you'd like to track) has been cited by another ***Elsevier***-published journal.

Visit www.alzheimersanddementia.org today!