Ascertainment of Dementia Cases in the Spanish European Prospective Investigation into Cancer and Nutrition-Murcia Cohort

María Encarnación Andreu-Reinón[a, b]　Diana Gavrila[b, c]　María Dolores Chirlaque[b–d]
Sandra M. Colorado-Yohar[b, c, e]　Pilar Amiano[c, f]　Eva Ardanaz[c, g]
Fernando Navarro-Mateu[c, h]　Carmen Navarro[b–d]　José María Huerta[b, c]

[a]Murcia Regional Health Council, Murcia, Spain; [b]Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain; [c]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain; [d]Department of Health and Social Sciences, University of Murcia, Murcia, Spain; [e]Research Group on Demography and Health, National Faculty of Public Health, University of Antioquia, Medellín, Colombia; [f]Public Health Division of Gipuzkoa, Instituto BIO-Donostia, Basque Government, Donostia/San Sebastián, Spain; [g]Public Health Institute of Navarra, IDISNA, Pamplona, Spain; [h]Unidad de Docencia, Investigación y Formación en Salud Mental (UDIF-SM), Servicio Murciano de Salud, IMIB-Arrixaca, Murcia, Spain

Abstract

Background: Cohort studies generally focus on a particular disease, although they offer the possibility of evaluating different outcomes with minimal additional investment. The objective of this study was to describe the methodology used to assess dementia in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Murcia study. **Methods:** The EPIC-Murcia cohort consists of 8,515 healthy participants (68% women, aged 30–70 years), recruited between 1992 and 1996 and followed up for over 20 years. Incident cases were ascertained by a 2-step protocol: a record linkage with health databases to identify potential events and a review of medical records of potential cases to validate incident cases. **Results:** Overall, 1,202 potential cases were identified, and 275 dementia cases were validated. Medical reports were the source of information in 243 cases, with complete neurological information in 227, and a high degree of certainty of the diagnosis in 229 cases. P70 (dementia code) and/or antidementia drugs and/or ICD codes identified 259 cases (sensitivity: 94.2%, 95% CI 90.7–96.6; specificity: 98.1%, 95% CI 97.8–98.4). **Conclusion:** Ascertainment of incident dementia in the EPIC-Murcia cohort study was feasible using information from medical records. This systematic 2-step validation protocol is proposed as a feasible way to ascertain dementia in cohort studies originally designed for other endpoints.

© 2018 S. Karger AG, Basel

Background

Cohort studies provide a convenient framework for assessing the incidence rate of major chronic diseases and the best-quality observational data. However, setting up

María Encarnación Andreu-Reinón
Servicio de Epidemiología
Ronda de Levante, 11, 2ª planta
ES–30008 Murcia (Spain)
E-Mail m.encarna.andreu@gmail.com

a cohort study can be very expensive and time consuming, since it may require following up with a large number of participants for a long time [1]. In view of these considerations, cohorts that are initially conceived to study a certain disease can be later conveniently used to investigate different outcomes.

As such, the European Prospective Investigation into Cancer and Nutrition (EPIC) study was initially designed to investigate the relationships between nutritional and lifestyle factors with the incidence of cancer [2, 3]. Nevertheless, the project was later extended to evaluate additional endpoints, such as stroke and acute myocardial infarction (as part of the EPIC-Heart and EPIC-CVD projects) [4], diabetes (EPIC-Interact) [5] or Parkinson Disease (NeuroEPIC) [6], taking advantage of the vast amount of data and research infrastructure of the EPIC project in an efficient way. Similarly, other studies, such as the Framingham Cohort Study [7–9] or the Cardiovascular Health study [10, 11], initially focused on a primary outcome (cardiovascular disease), were later successfully used to assess other outcomes such as dementia.

Most cognitive cohort studies (or cognitive sub-studies) have ascertained dementia by conducting neuropsychological and neurological examinations of the cohort participants. By contrast, ascertainment of dementia is more difficult in those studies without cognitive screenings at baseline. In some cases, such as the above-mentioned examples [7, 8, 10], a random sample of the overall cohort was obtained, and the neuropsychological assessment was performed on this subcohort. However, this is still expensive and not always possible.

The diagnosis of dementia is based on clinical criteria, with the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) criteria being the most recognized and widely used until 2013, when a new version was published [12, 13]. Based on previous experiences within the EPIC study, we carried the identification and validation of incident cases of dementia through record linkage with existing health databases (primary care registries, hospital discharge databases, and mortality registries) and an exhaustive review of clinical records. The objective of this study was, thus, to describe the procedures used in the ascertainment of dementia and its subtypes in the EPIC-Murcia cohort.

Methods

Population and Study Design
EPIC is a large multicentre prospective study coordinated by the International Agency for Research on Cancer, which includes over half a million participants (70% women) from 10 European countries (France, Italy, Spain, the United Kingdom, the Netherlands, Greece, Germany, Sweden, Denmark and Norway) [14].

The Spanish branch of EPIC recruited 41,438 participants from 5 regions (Asturias, Gipuzkoa, Navarra, Murcia and Granada) between 1993 and 1996. Baseline dietary and lifestyle data was collected in face-to-face interviews using validated questionnaires [15]. Participants also underwent a physical examination to gather anthropometric information and provided details on their medical and reproductive history, together with a blood sample. A full description of the study methods has been published elsewhere [15].

EPIC-Spain was initially designed to identify new cases of cancer through periodic record linkage with population-based cancer registries covering the population of the 5 Spanish areas where the study was set. Vital status of the cohort and cause of death were assessed through record linkage with the regional mortality registry and the Spanish National Statistics Institute (www.ine.es).

The current study was conducted within the EPIC-Murcia cohort, which consisted of 8,515 participants (n = 5,831 women), recruited between October 1992 and February 1996, with ages ranging from 30 to 70 years at the time of enrolment (98% below 65 years).

The participants were healthy volunteers, mostly blood donors (67%) but also employees from private companies (3%), civil servants (5%), and the general population (23%) [14], all of whom were fully covered by the public health system. The exclusion criteria were pregnancy, lactation and not being physically or mentally capable of participating.

Cognitive status was not assessed at baseline. However, participants had to complete extensive baseline questionnaires on pathologies, medication, smoking, diet and lifestyles, during face-to-face interviews that were demanding enough to require normal cognition.

Ethics
The EPIC-Spain study was approved by the Ethics Committee of the Bellvitge Hospital (Barcelona, Spain), and all participants gave written informed consent. The consent form included the express authorization to access the participants' medical records to update clinical information in the future for research purposes within the EPIC study.

Ascertainment of Dementia Cases in the EPIC-Murcia Cohort
In the present study, incident cases of dementia occurring from the date of recruitment (1992–1996) through December 31st, 2016, were ascertained. Using the validation of Parkinson´s Disease in the NeuroEPIC4PD study [6] as reference, we conducted a validation process that consisted of 2 steps: the identification of potential incident cases by linkage of the EPIC database with medical datasets (step 1) and the subsequent validation of potential cases after careful revision of the medical records (step 2).

Step 1: Identification of Potential Incident Cases
For the identification of potential cases, a record linkage using unique identifiers was performed separately between the EPIC Murcia dataset and each available health database that could include dementia-related clinical information: primary care registries, hospital discharge databases (Minimum Dataset), and regional and national mortality registries, using selected International Classification of Diseases and Related Health Problems, 9th

Table 1. Health databases and description and classification of the selected codes for the identification and validation of potential cases of dementia in the EPIC-Murcia cohort

Health databases	Code	Classification	Description
Primary care registry	P20	ICPC-2[1]	Memory disturbance
	P70	ICPC-2	Dementia
	N29	ICPC-2	Neurological symptoms/complt. Other
	N99	ICPC-2	Neurological diseases, other
Primary care registry/MDS[5]	N06DA02[6]	ATC[2]	Donepezil
	N06DA03	ATC	Rivastigmine
	N06DA04	ATC	Galantamine
	N06DX01	ATC	Memantine
MOS/mortality registry	290	ICD-9[3]	Dementias
	331	ICD-9	Other cerebral degenerations
	F00	ICD-10[4]	Dementia in Alzheimer disease
	F01	ICD-10	Vascular dementia
	F02	ICD-10	Dementia in other diseases classified elsewhere
	F03	ICD-10	Unspecified dementia
	G30	ICD-10	Alzheimer disease

[1] ICPC-2, International Classification of Primary Care, Second edition.
[2] ATC, anatomical, therapeutic, chemical classification system.
[3] International Statistical Classificalion of Disease and Related Health Problems, 9th Revision.
[4] International Statistical Classification of Disease and Related Health Problems, 10th Revision.
[5] Minimum dataset (hospital discharge databases).
[6] N06D, antidementia drugs.

and 10th revision (ICD 9 and 10), International Classification of Primary Care, second edition (ICPC2), and Anatomical, Therapeutic, Chemical classification system (ATC) codes. The following codes were sought: 290, and 331 (ICD-9); F00, F01, F02, F03, and G30 (ICD-10); P20, P70, N29, and N99 (ICPC2); and N06DA02, N06DA03, N06DDA04, and N06DX01 (ATC; Table 1).

All participants with at least 1 record of any of the mentioned codes in at least one of the selected databases were considered *potential cases* of dementia.

Step 2: Validation of Potential Cases of Dementia

The validation procedure to ascertain incident dementia was performed by an expert neurologist who thoroughly reviewed all the available medical information of each potential case; the information was either in the form of electronic data or in the form of paper medical records. The validation of incident cases of dementia was based on different types (sources) of evidence: outpatient and hospital medical reports (admission, discharge and emergency reports), drug prescriptions, diagnostic tests and information of the previously selected codes recorded in health databases. The expert reviewer then decided whether or not the available data sufficed to establish the diagnosis of dementia. The decision-making process was guided by an ad hoc algorithm revised by the panel of neurologists participating in the study (Fig. 1).

A hierarchical procedure was established for data searching (Fig. 1) by reviewing electronic medical records first, and, if they did not contain enough information, then caution was taken to ensure the diagnosis of dementia through the revision of medical records on paper. In a minority of cases, when no medical reports were available, the case status was defined only by codes ($n = 32$, 11.6%), using the same codes selected for the identification of potential cases (Table 1).

A specific study database was designed where the data were collected; this included general information, clinical data (DSM IV criteria, clinical features and subtypes of dementia, incidence date and neuropsychological assessments), functional status, results of the diagnostic tests, use of antidementia drugs, source, quality and amount of information, and the clinical judgement of the reviewer (independent rater), who finally established the diagnosis of dementia, its subtype and the degree of certainty.

Diagnostic Criteria

Consequently, based on the available information, potential cases were validated as incident cases of dementia if a clinical diagnosis of dementia appeared in a medical report. If no medical reports were available, dementia was established if any of the following codes appeared in the medical databases: P70 and/or antidementia drug prescription, and/or ICD 9-10 codes (Table 2).

Quality Criteria

The quality of the data was defined as *high*, *medium*, or *low*, based on the amount of information obtained from the medical reports.

a) *High quality data:* when complete neurological information and neuropsychological information was available.

b) *Medium quality data:* when complete neurological information, but no neuropsychological information or information on the subtype, was available.

Fig. 1. Flowchart of the process used for validation of the dementia cases in the EPIC-Murcia cohort. * Neurological or no neurological reports where the diagnosis of dementia appears. ** At least enough information to establish the diagnosis of dementia and the incidence date.
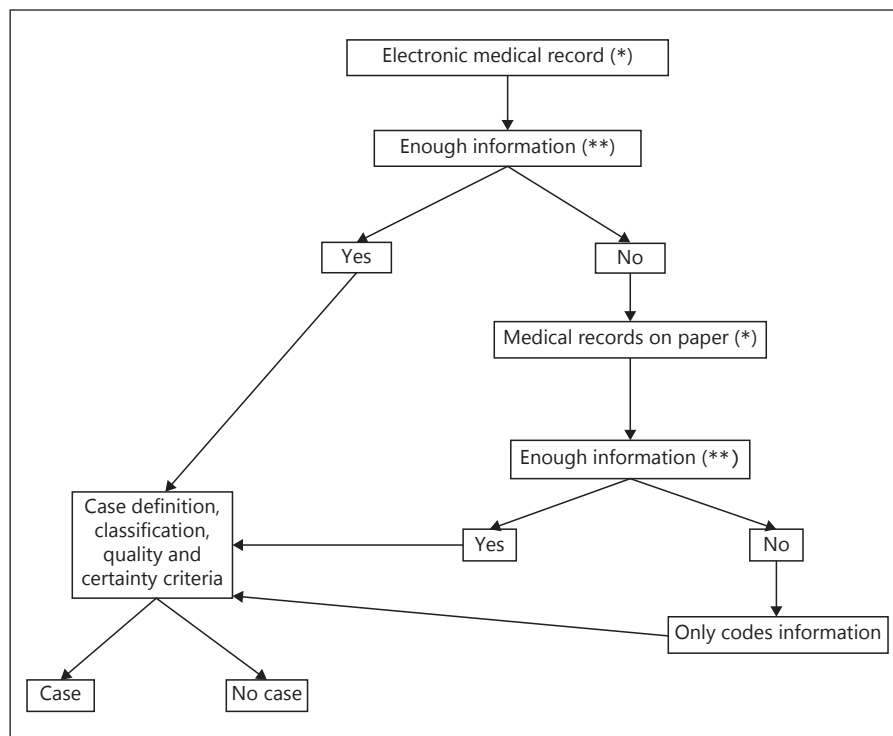
c) *Low quality data:* when clinical or neurological information available was not enough to establish the subtype, but it sufficed to ascertain the diagnosis of dementia.

Complete neurological information comprehended sufficient clinical data obtained from medical reports written by a specialist in dementia (neurologist, psychiatrist or geriatrist). This included clinical information (family history, first symptom, date of symptom onset, incidence date, symptom evolution, death), functional status at diagnosis, and dependence (measured by functional scales such as Global Deterioration Scale, Clinical Dementia Rating Scale, Blessed Dementia Scale, Barthel Index for Activities of Daily Living), diagnostic criteria (DSM IV), diagnosis of the subtype of dementia (Alzheimer's Disease, vascular dementia, Lewy Body disease, other subtypes of dementia and unspecified dementia), diagnostic tests (which tests were practiced and the results: neuroimaging, other tests – SPECT: single photon emission computed tomography, neurosonology, EEG), and pharmacological treatment (antidementia drugs).

The neuropsychological assessment included a complete cognitive evaluation of the patient and the identification of the clinical profile and severity carried out either by a neuropsychologist or a neurologist specialized in dementia. This included the date of the evaluation, test or battery of tests performed in each case, neuropsychological dysfunction pattern (referring to cortical, subcortical, cortico-subcortical), localization (frontal, temporal, parietal, occipital) and severity (mild, moderate, severe).

Certainty

Taking into account the quality and the amount of medical information, dementia was defined as *probable* or *possible*, according to the following criteria:

Probable Dementia: When enough information was obtained to ensure the validation of dementia (diagnosis of dementia in a medical report or a prescription code for anti-dementia drugs (either isolated or in combination with P70 or ICD codes).

Possible Dementia: When no medical reports were obtained, and P70 or ICD codes were not associated with antidementia drug prescriptions.

If the revision based on the available data did not allow the possibility to ensure the diagnosis of dementia, the potential case was discarded and recorded as a non-case.

Subtype of Dementia

Dementia subtype was specified when detailed clinical information about the subtype was available in medical reports. If no detailed information or if only codes were available, the subtype was considered non-specified.

Incidence Date

According to the available information, the date of the onset of dementia in the EPIC-Murcia cohort was determined as shown in Figure 2. The incidence date was the date of the first neurological report where the diagnosis of dementia appeared or if only codes were available, the earliest date of record of any code or the death date when no other information was available.

Statistical Analyses

The validity of the selected codes was assessed by calculating their sensitivity, specificity, and positive and negative predictive values with 95% CIs [16].

Participants of the EPIC-Murcia cohort were classified as cases (those with dementia) and non-cases (those without dementia).

Andreu-Reinón et al.

Table 2. Diagnostic criteria of suspected dementia cases according to the source, amount and quality of medical information

Source	Data available	Quality of the data[5]	Reviewer diagnosis of dementia	Reviewer diagnosis of subtype	Degree of certainty[6]
Medical reports					
Neurological report + neuropsychological assessment	Complete neurological and neuropsychological information to ensure the diagnosis of dementia, the incidence date and the subtype of dementia	High		Yes (always)	
Neurological report without neuropsychological assessment	Complete neurological information to ensure the diagnosis of dementia, but incomplete neuropsychological assessment or information on the subtype: (a) no neuropsychological assessment was done; (b) no functional assessment; (c) no subtype defined in the clinical information; (c) limited clinical information to verify the subtype of dementia in case the neurologist indicates this	Medium	Yes	(a) If the subtype of dementia is in the report this will be the subtype diagnosis. (b) If there is no subtype diagnosis in the report or if there are discrepant options (for example: Alzheimer versus Lewy Body disease): unspecified dementia	Probable
Other specialist/ emergency report					
Neurological report without neuropsychological assessment	Enough information to ensure the diagnosis of dementia, but uncertainty about the subtype of dementia or inconsistent information to ensure the diagnosis of dementia subtype: (a) mention to dementia only appears in medical history without any other information (quality of life, symptoms, treatment, etc.); (b) dementia diagnosis appearing only in some reports but not in others; (c) dementia diagnosis appears in some reports, but normal cognition is reflected in others	Low			
Other specialist/ emergency report					
Only codes[1]					
P20[2]	No diagnosis of dementia	Only codes	No	–	–
p70[2]	Yes	Only codes	Yes	No (unspecified dementia)	Possible*
ICD codes[3,4]	Yes	Only codes	Yes	No (unspecified dementia)	
P20 and p70[2]	Yes	Only codes	Yes	No (unspecified dementia)	
ICD codes[3,4] and p70[2]	Yes	Only codes	Yes	No (unspecified dementia)	
Antidementia drugs[2,3]	Yes	Only codes	Yes	No (unspecified dementia)	Probable*

[1] When no medical reports were available, the same codes used for the initial selection of potential cases were employed as clinical information to assess the diagnosis of dementia. [2] Primary care registries. [3] Hospital discharge databases (minimum dataset, MDS). [4] Regional mortality registry. [5] Quality of the data: 3 degrees of quality, only applied to medical reports: high, medium, and low. [6] Degree of certainty: Probable and Possible. * Only codes information. Degree of certainty based on the selected codes: antidementia drug, Probable; the remaining codes, Possible.
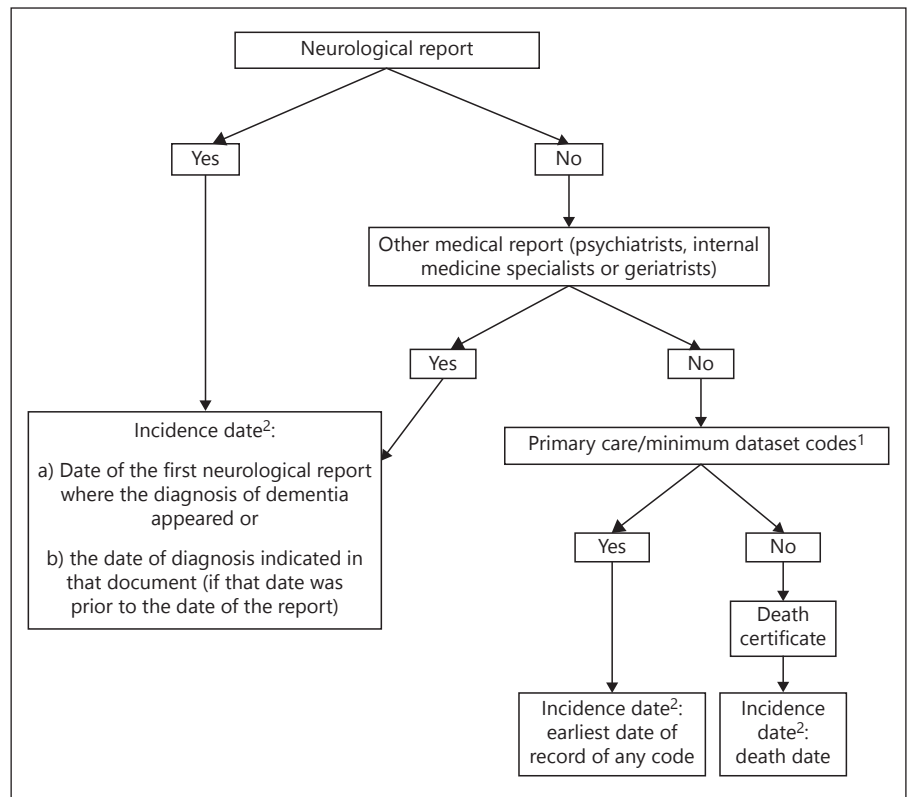
Fig. 2. Flowchart of the process used to determine the incidence date of dementia in the EPIC-Murcia cohort. [1] Minimum dataset (hospital discharge databases). [2] Date format: complete date available: dd/mm/yyyy; only month and year available: 01/mm/yyyy; only year available: 30/06/yyyy.

The statistical measures were estimated for each individual code and for the combinations of codes. Each code (or combination) was identified as "code present" or "code absent" to distinguish the participants with or without the selected codes. Table 3 shows the definition of the true and false positives and negatives used in the statistical analyses.

Statistical analyses were performed using STATA/SE version 14 (Stata Corp., College Station, TX, USA).

Table 3. Definition of true and false positives and negatives for the estimations of sensitivity, specificity, and PPV and NPVs

	Cases (*n* = 275)	Non-cases (*n* = 927)
Code present	True positives	False positives
Code absent	False negatives	True negatives

Potential cases (*n* = 1,202) were those initially identified by the selected ICD, ICPC and ATC codes. Cases: those "potential cases" established as incident cases after the process of validation of the reviewer (*n* = 275). *Non-cases:* the rest of the "potential cases" not identified as cases (*n* = 927).

PPV, positive predictive value; NPV, negative predictive value.

Results

Altogether, 1,202 potential cases were identified and reviewed, and 275 dementia cases (262 probable cases and 13 possible cases) were finally validated. The validation process confirmed that no participants had prevalent dementia at enrolment.

The subtype of dementia, established after thoroughly checking the medical records and codes, was "Alzheimer" in 160 cases (58%), "vascular" in 18 (7%), "Lewy bodies disease" in 9 (3%), "other type of dementia" (such as frontotemporal disease, supranuclear paralysis, or dementia associated with Parkinson's Disease, among others) in 33 cases (12%), and "unspecified dementia" in 55 cases (20%).

Medical reports were the source of information in 243 cases (88%): neurological reports were the source in 205 cases, and non-neurological reports were the source in 38 cases (14%), whereas codes were the only source of information in 32 cases (12%). The quality of the data obtained from the medical reports was "medium" or "high" in 227 cases (82.5%), and almost all cases were considered "probable cases" (*n* = 262; 95%).

Furthermore, the validation process confirmed that no participants had prevalent dementia at the time of enrolment.

In order to evaluate the feasibility and efficiency of the procedure, we also aimed to evaluate the usefulness of the

Andreu-Reinón et al.

search based on the codes defined in the identification step (step 1) for the prospective ascertainment of dementia cases. Thus, after the validation of the dementia cases was completed, sensitivity, specificity, and positive (PPV) and negative predictive values were estimated for each code or combination of codes that had led to identify the potential cases at step 1 [16] (Tables 4, 5, and online supplementary table S1; for all online suppl. material, see www.karger.com/doi/10.1159/000493209).

Regarding the isolated codes, "antidementia drugs" (N06DA02 Donepezil, N06DA03 Rivastigmine, N06DA04 Galantamine, and N06DX01 Memantine, ATC classification) showed the highest sensitivity: 74.2% for all dementia cases and 82.5% for Alzheimer, with high specificity, positive and negative predictive values. The second most useful code was "P70", identifying 175 cases of dementia (63.6% sensitivity and 99.2% specificity), and 116 cases of Alzheimer (72.5% sensitivity and 98.5% specificity). For the associations of the codes, the combination of "P70 and/or antidementia drugs" identified 241 cases of dementia (87.6%), and 146 cases of Alzheimer (91.3%), with 98.6% and 97.5% specificity respectively. The PPV reached 67.9 and 41.1% for dementia and Alzheimer cases respectively.

The code P20, when considered isolated from any other code, was the least useful code, identifying only 16 cases of dementia (5.8%), 11 cases of Alzheimer (6.9%), and no cases of dementia based only on code information, with a very low sensitivity (5.8% for dementia and 6.9% for Alzheimer).

All codes and their combinations had specificities and negative predictive values over 90%.

Discussion

In the EPIC-Murcia cohort, 275 incident cases were identified from a total of 1,202 potential cases (22.9%), with medium or high-quality data in 227 of them (82.5%), and a high degree of certainty in the diagnosis in 83.2%.

From the selected isolated codes, antidementia drugs showed the highest sensitivity, reaching values over 70% for dementia cases and over 80% for Alzheimer cases. The reason why the sensitivity of codes referring to antidementia drugs is higher for Alzheimer disease is because these medications (anticholinesterase drugs and memantine) are rather specific for this subtype of dementia. On the other hand, P20 was the least useful code, with sensitivity values below 10% for both dementia and Alzheimer cases.

Table 4. Estimated sensitivity, specificity, and PPV and NPV with 95% CIs of the selected codes used for the identification of potential cases of dementia

	Cases (true positives)[1]	Non cases (false positives)[2]	Total (code present)	Sensitivity (%, 95% CI)	Specificity (%, 95% CI)	PPV[3] (%, 95% CI)	NPV[4] (%, 95% CI)
Single codes							
P20[3]	80	755	835	29.1 (23.8–34.9)	90.8 (90.2–91.5)	9.6 (8.0–11.4)	97.5 (97.0–97.6)
P70[4]	175	65	240	63.6 (57.7–69.3)	99.2 (99.0–99.4)	72.9 (67.5–77.7)	98.8 (98.6–99.0)
Antidementia drugs[5]	204	71	275	74.2 (68.6–79.3)	99.1 (98.9–99.3)	74.2 (69.3–78.5)	99.1 (99.0–99.3)
ICD 9, 10 codes[6]	59	50	109	21.5 (16.8–26.8)	99.4 (99.2–99.6)	54.1 (45.2–62.8)	97.4 (97.3–97.6)
Other codes (N29, N99)	0	42	42	0.0 (0.0–1.3)	99.5 (99.3–99.6)	0	96.8 (96.8–96.8)
Combination of codes							
P20 without any other codes	16	728	744	5.8 (3.3–9.2)	91.2 (90.5–91.8)	2.2 (1.3–3.4)	96.9 (96.8–97.0)
P70 and/or antidementia drugs	241	114	355	87.6 (83.2–91.3)	98.6 (98.3–98.9)	67.9 (63.7–71.8)	99.6 (99.4–99.7)
P70 and/or antidementia drugs and/or ICD	259	157	416	94.2 (90.7–96.6)	98.1 (97.8–98.4)	62.3 (58.5–65.9)	99.8 (99.7–99.9)

[1] Total cases: 275.
[2] Total non-cases: 8240 (total EPIC Murcia cohort: 8515).
[3] P20, memory disturbance.
[4] P70 dementia.
[5] Antidementia drugs: N06DA, anticholinesterases N06DA02, Donepezil N06DA03, Rivastigmine and N06DA04, Galantamine. N06DX, Other antidementia drugs: N06DX01, Memantine.
[6] ICD-9 codes: 290 (dementias) and 331 (other cerebral degenerations). ICD-10 codes: F00-03 (dementias) and G30 (Alzheimer's disease).
ICD, International Classification of Diseases; NPV, negative predictive value; PPV, positive predictive value.

Table 5. Estimated sensitivity, specificity, and PPV and NPV with 95% CIs of selected codes used for the identification of potential cases of Alzheimer

	Cases EA (true positives)[1]	Total cases no EA + non-cases (false positives)	Total (code present)	Sensitivity (%, 95% CI)	Specificity (%, 95% CI)	PPV[3] (%, 95% CI)	NPV[4] (%, 95% CI)
Single codes							
P20[3]	54	781	835	33.8 (26.5–41.6)	90.7 (90.0–91.3)	6.5 (5.2–8.0)	98.6 (98.5–98.8)
P70[4]	116	124	240	72.5 (64.9–79.3)	98.5 (98.2–98.8)	48.3 (43.4–53.3)	99.5 (99.3–99.6)
Antidementia drugs[5]	132	143	275	82.5 (75.7–88.1)	98.3 (98.0–98.6)	48.0 (43.6–52.4)	99.7 (99.5–99.8)
ICD 9, 10 codes[6]	25	84	109	15.6 (10.4–22.2)	99.0 (98.8–99.2)	22.9 (16.4–31.1)	98.4 (98.3–98.5)
Other codes (N29, N99)	0	42	42	0.0 (0.0–2.3)	99.5 (99.3–99.6)	0	98.1 (98.1–98.1)
Combination of codes							
P20 without any other codes	11	733	744	6.9 (3.5–12.0)	91.2 (90.6–91.8)	1.5 (0.8–2.6)	98.1 (98.0–98.2)
P70 and/or antidementia drugs	146	209	355	91.3 (85.8–95.1)	97.5 (97.1–97.8)	41.1 (37.7–44.6)	99.8 (99.7–99.9)
P70 and/or antidementia drugs and/or ICD	149	267	416	93.1 (88.1–96.5)	96.8 (96.4–97.2)	35.8 (33.0–38.8)	99.9 (99.9–99.9)

[1] Total Alzheimer's cases: 160.
[2] Total no Alzheimer's cases (non-aises and other subtypes of dementia): 8355 (total EPIC Murcia cohort: 8515).
[3] P20, memory disturbance.
[4] P70, dementia.
[5] Antidementia drugs: N06DA, anticholinesterases; N06DA02, Donepezil; N06DA03, Rivastigmine; and N06DA04, Galantamine. N06DX, other antidementia drugs: N06DX01, Memantine.
[6] ICD-9 codes: 290 (dementias) and 331 (other cerebral degenerations). ICD-10 codes: F00-03 (dementias) and G30 (Alzheimer's disease).
ICD, International Classification of Diseases; NPV, negative predictive value; PPV, positive predictive value.

To date, there are only a few cohort studies that have published the methodology of their case validation. Most of them include neuropsychological screening and neurological or neuropsychological assessments of potential cases during the follow-up of the study [7, 8, 10, 17–31]. Some examples are the Personnes Agées QUID cohort in France [26], the Washington Hamilton Heights and Inwood neighbourhood in the United States [32] or the CFAS (The medical research council cognitive function and ageing study) [29] in the United Kingdom. This may be the most accurate method for identifying incident cases, but its cost is high in large cohort studies and it is not always possible to put into practice. Consequently, different pragmatic strategies have been used in other studies.

One previous study by Chien et al. [33] was designed to estimate the prevalence and incidence of dementia in Taiwan. It used a population-based registry and the diagnosis of dementia was established using only ICD-9-Clinical Modification codes (290: senile and presenile organic psychotic condition, or 331: Alzheimer disease) [33]. In the Rochester Epidemiology Project, the identification of potential cases was done using the medical record linkage system, and the validation was done by the revision of all medical records by a trained nurse and the subsequent confirmation by a neurologist [34]. The same method was used to identify other neurological diseases, such as migraine [35] or stroke [36]. However, so far there is no "gold standard" methodology to ascertain the incidence of dementia using existing data in cohort studies [37].

Some previous studies have evaluated the validity of ICD codes [38–42] and ATC codes (antidementia drugs) in the diagnosis of incident cases of dementia. Pippenger et al. [43] evaluated the usefulness of ICD-9 codes by comparing them to the diagnosis of a neurologist (considered the "gold standard"). In their study, potential cases were identified through ICD-9 "290" and "331" codes and the clinical records of such potential cases were checked afterwards in order to validate the diagnosis of incident cases. Our method basically followed the same approach (the use of codes to identify potential cases and a thorough subsequent expert revision of medical records).

Of note, P20 and P70, the most useful codes in our study, have not been previously evaluated in the identification of incident cases of dementia. However, the main objectives of both studies differed. Pippenger et al. [43] estimated validity indices for the case ascertainment approach in order to validate the diagnosis of incident dementia based on codes. On the contrary, we did not intend to validate the method in the absence of an independent gold standard but to describe the overall procedure

(search strategy, sources of data available, hierarchy of the validation protocol) and to give an indication (in terms of validity) of the performance of the codes for the identification of potential cases (not as diagnostic tools).

In a paper published by Liisa Jaakkimainen et al. [40], they reported a retrospective validation of different algorithms to diagnose dementia. It was found that the best algorithm to diagnose dementia in usual clinical practice, compared to the reference standard (family practice-based electronic medical records, a tool used in Canada) was "*one hospitalization code* [ICD-9 or 10] *OR claims codes at least 30 days apart in a 2-year period OR a prescription filled for Alzheimer's disease*". This algorithm had a high sensitivity and PPV (approximately 80%) and higher specificity and negative predictive value (approximately 99%).

Van de Vorst et al. [39] evaluated the validity of the diagnosis of dementia in a population-based registry with ICD-9 codes compared to medical records. They found that ICD-9 codes related to dementia (codes "290" and "331") showed a high PPV for the diagnosis of dementia and Alzheimer (93.2% for overall dementia), being higher in the group over 65 years old (95.5 vs. 67.9% among those <60 years). Finally, in other cohort studies, such as the Atherosclerosis Risk in Communities cohort [41], ICD-9 codes were used to define the incidence date.

We have established the diagnosis of dementia based on the individual revision of medical records in a process that implied at least 2 expert opinions: the clinical judgement of the medical doctor who carried out the in-person evaluation of the patients, and the clinical judgement of the neurologist that reviewed the medical records, acting as an independent rater. Only for a minority of the cases (<12%), the lack of medical reports led to a diagnosis based solely on codes. Nevertheless, since we have used those codes that have shown good diagnostic validity in different previous epidemiologic studies (plus additional information from ICPC-2 codes), the diagnosis of dementia could be reasonably established, albeit with lower certainty, in those cases.

Our study has strengths and limitations. The case validation was done by 1 expert reviewer (a neurologist). Although this guaranteed a high internal validity of the method, which avoided any interobserver variability, it could also be seen as a limitation. Nevertheless, the case ascertainment protocol was revised by a panel of neurologists, and the reviewer established the diagnosis according to the data available in the medical records, adhering to this protocol. Furthermore, we used a sensitive approach to identify all potential cases of dementia by searching for all relevant dementia-related codes (plus others related to neurological conditions) in all available health databases (primary care records, hospital discharge databases, and mortality registries).

On the other hand, we consider the absence of a baseline cognitive assessment one of the main limitations of our study. However, after the case ascertainment was completed, we could verify that no cohort participants had prevalent dementia at baseline. The potential for reverse causation will be addressed in future etiological analyses by excluding cases diagnosed within the first years of follow-up as a sensitivity analysis.

As another limitation, dementia subtypes were diagnosed based on clinical information obtained from medical reports and diagnostic tests (when available). This led to a considerably high rate of unspecified subtypes of dementia (20%). A conclusive diagnosis of the dementia subtypes cannot be established in the absence of a complete neuropsychological evaluation and biomarker information. However, this information was not available for all cases. A further drawback is that this method may only be used in countries with electronic health data. Finally, lack of representativity of the EPIC-Murcia cohort would limit the external validity of the results.

In addition to the feasibility of this method, assessing the validity of the codes for the identification of dementia cases could be helpful to improve the efficiency of the identification step by maximizing its sensitivity. Our extensive search identified 1202 potential cases, whose medical histories were further individually revised to give the final number of 275 validated cases of dementia (23% of all revised cases). Taking into account our experience, the proportion of potential cases to validate could be drastically reduced using a combination of codes with high specificity and balanced sensitivity.

Conclusion

In conclusion, the systematization of our case ascertainment protocol could be taken as a reference to a pragmatic approach for the identification and validation of incident cases of dementia, when no routine cognitive testing or continuous follow-up screening is available.

Acknowledgements

References

1 Song JW, Chung KC: Observational studies: cohort and case-control studies. Plast Reconstr Surg 2010;126:2234–2242.

2 Riboli E: Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). Ann Oncol Off J Eur Soc Med Oncol 1992;3: 783–791.

3 Riboli E, Kaaks R: The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26(suppl 1):S6–S14.

4 Danesh J, Saracci R, Berglund G, Feskens E, Overvad K, Panico S, et al: EPIC-Heart: The cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. Eur J Epidemiol 2007;22:129–141.

5 InterAct Consortium TI, Langenberg C, Sharp S, Forouhi NG, Franks PW, Schulze MB, et al: Design and cohort description of the interact project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. Diabetologia 2011;54:2272–2282.

6 Gallo V, Brayne C, Forsgren L, Barker RA, Petersson J, Hansson O, et al: Parkinson's disease case ascertainment in the EPIC cohort: the neuroEPIC4PD study. Neurodegener Dis 2015;15:331–338.

7 Satizabal CL, Beiser AS, Chouraki V, Chêne G, Dufouil C, Seshadri S: Incidence of dementia over three decades in the Framingham heart study. N Engl J Med 2016;374:523–532.

8 Chene G, Beiser A, Au R, Preis S, Wolf P, Dufouil C, et al: Gender and incidence of dementia in the Framingham heart study from mid-adult life. Alzheimers Dement 2015;11:310–320.

9 Friedman G, in Hyg S, Kannel W, Dawber T, McNamara P: An evaluation of follow-up methods in the Framingham heart study. Am J Public Heal Nation's Heal 1967;57:1015–1024.

10 Fitzpatrick AL, Kuller LH, Ives DG, Lopez OL, Jagust W, Breitner JC, et al: Incidence and prevalence of dementia in the cardiovascular health study. J Am Geriatr Soc 2004;52:195–204.

11 Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al: The cardiovascular health study: design and rationale. Ann Epidemiol 1991;1:263–276.

12 American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (4th ed, text revision). Washington, American Psychiatric Association, 2000.

13 American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (ed 5). Arlington, American Psychiatric Association, 2013.

14 Riboli E, Hunt K, Slimani N, Ferrari P, Norat T, Fahey M, et al.: European prospective investigation into cancer and nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5:1113–1124.

15 González CA, Navarro C, Martínez C, Quirós JR, Dorronsoro M, Barricarte A, et al: El estudio prospectivo Europeo sobre cáncer y nutrición (EPIC). Rev Esp Salud Publica 2004; 78:167–176.

16 Rothman KJ, Greenland S, Lash TL: Chapter 9- validity in epidemiologic studies; in Wilkins LW (eds): Modern Epidemiology (ed 3). Philadelphia, Wolters Kluwer, 2008.

17 Hall CB, Verghese J, Sliwinski M, Chen Z, Katz M, Derby C, et al: Dementia incidence may increase more slowly after age 90: results from the bronx aging study. Neurology 2005; 65:882–886.

18 Paykel ES, Brayne C, Huppert FA, Gill C, Barkley C, Gehlhaar E, et al: Incidence of dementia in a population older than 75 years in the United Kingdom. Arch Gen Psychiatry 1994; 51:325–332.

19 Borenstein AR, Wu Y, Bowen JD, McCormick WC, Uomoto J, McCurry SM, et al: Incidence rates of dementia, Alzheimer disease, and vascular dementia in the Japanese American population in Seattle, WA: the Kame Project. Alzheimer Dis Assoc Disord 2014;28:23–29.

20 Di Carlo A, Baldereschi M, Amaducci L, Lepore V, Bracco L, Maggi S, et al: Incidence of dementia, Alzheimer's disease, and vascular dementia in Italy. The ILSA study. J Am Geriatr Soc 2002;50:41–48.

21 Copeland JR, Dewey ME, Davidson IA, Saunders PA, Scott A: Geriatric Mental State-AGECAT: prevalence, incidence and long-term outcome of dementia and organic disorders in the liverpool study of continuing health in the community. Neuroepidemiology 1992;11:84–87.

22 Copeland JR, McCracken CF, Dewey ME, Wilson KC, Doran M, Gilmore C, et al: Undifferentiated dementia, Alzheimer's disease and vascular dementia: age- and gender-related incidence in Liverpool. The MRC-ALPHA Study. Br J Psychiatry 1999;175:433–438.

23 Matthews FE, Stephan BC, Robinson L, Jagger C, Barnes LE, Arthur A, et al: A two decade dementia incidence comparison from the cognitive function and ageing studies I and II. Nat Commun 2016;7:11398.

24 Corrada MM, Brookmeyer R, Paganini-Hill A, Berlau D, Kawas CH: Dementia incidence continues to increase with age in the oldest old: the 90+ study. Ann Neurol 2010;67:114–121.

25 Prince M, Acosta D, Ferri CP, Guerra M, Huang Y, Llibre Rodriguez JJ, et al: Dementia incidence and mortality in middle-income countries, and associations with indicators of cognitive reserve: a 10/66 dementia research group population-based cohort study. Lancet 2012;380:50–58.

26 Grasset L, Brayne C, Joly P, Jacqmin-Gadda H, Peres K, Foubert-Samier A, et al: Trends in dementia incidence: evolution over a 10-year period in France. Alzheimers Dement 2016; 12:272–280.

27 Aronson MK, Ooi WL, Geva DL, Masur D, Blau A, Frishman W: Dementia. Age-dependent incidence, prevalence, and mortality in the old old. Arch Intern Med 1991;151:989–992.

28 Brayne C, Gill C, Huppert FA, Barkley C, Gehlhaar E, Girling DM, et al: Incidence of clinically diagnosed subtypes of dementia in an elderly population. Cambridge project for later life. Br J Psychiatry 1995;167:255–262.

29 Brayne C: Incidence of dementia in England and Wales: the MRC cognitive function and ageing study. Alzheimer Dis Assoc Disord 2006;20:S47–S51.

30 Andersen K, Nielsen H, Lolk A, Andersen J, Becker I, Kragh-Sørensen P: Incidence of very mild to severe dementia and Alzheimer's disease in Denmark: the odense study. Neurology 1999;52:85–90.

31 Letenneur L, Commenges D, Dartigues JF, Barberger-Gateau P: Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. Int J Epidemiol 1994;23:1256–1261.

32 Scarmeas N, Stern Y, Tang M-X, Mayeux R, Luchsinger JA: Mediterranean diet and risk for Alzheimer's disease. Ann Neurol 2006;59: 912–921.

33 Chia Chien I, Lin YC, Chou YJ, Lin CH, Bih SH, Lee CH, et al: Treated prevalence and incidence of dementia among national health insurance Enrollees in Taiwan, 1996–2003. J Geriatr Psychiatry Neurol 2008;21:142–148.

34 Knopman DS, Petersen RC, Rocca WA, Larson EB, Ganguli M: Passive case-finding for Alzheimer's disease and dementia in two U.S. communities. Alzheimer's Dement 2011;7: 53–60.

35 Stang PE, Yanagihara T, Swanson JW, Beard CM, Melton LJ: A population-based study of migraine headaches in Olmsted County, Minnesota. Case ascertainment and classification. Neuroepidemiology 1991;10:297–307.

36 Whisnant JP, Melton LJ, Davis PH, O'Fallon WM, Nishimaru K, Schoenberg BS: Comparison of case ascertainment by medical record linkage and cohort follow-up to determine incidence rates for transient ischemic attacks and stroke. J Clin Epidemiol 1990;43:791–797.

37 Sibbett RA, Russ TC, Deary IJ, Starr JM: Dementia ascertainment using existing data in UK longitudinal and cohort studies: a systematic review of methodology. BMC Psychiatry 2017;17:239.

38 St Germaine-Smith C, Metcalfe A, Pringsheim T, Roberts JI, Beck CA, Hemmelgarn BR, et al: Recommendations for optimal ICD codes to study neurologic conditions: a systematic review. Neurology 2012;79:1049–1055.

39 van de Vorst IE, Vaartjes I, Sinnecker LF, Beks LJM, Bots ML, Koek HL: The validity of national hospital discharge register data on dementia: a comparative analysis using clinical data from a university medical centre. Neth J Med 2015;73:69–75.

40 Liisa Jaakkimainen R, Bronskill SE, Tierney MC, Herrmann N, Green D, Young J, et al: Identification of physician-diagnosed Alzheimer's disease and related dementias in population-based administrative data: a validation study using family Physicians' elec-tronic medical records. J Alzheimers Dis 2016;54:337–349.

41 Schneider AL, Gottesman RF, Mosley T, Alonso A, Knopman DS, Coresh J, et al: Cognition and incident dementia hospitalization: results from the atherosclerosis risk in communities (ARIC) study. Neuroepidemiology 2013;40:117–124.

42 Bharmal MF, Weiner M, Sands LP, Xu H, Craig BA, Thomas J: Impact of patient selection criteria on prevalence estimates and prevalence of diagnosed dementia in a medicaid population. Alzheimer Dis Assoc Disord 2007;21:92–100.

43 Pippenger M, Holloway RG, Vickrey BG: Neurologists' use of ICD-9CM codes for dementia. Neurology 2001;56:1206–1209.