






RESEARCH

Open Access



Validation of diagnostic codes and epidemiologic trends of Huntington disease: a population-based study in Navarre, Spain

Esther Vicente^{1,2*}, Ainara Ruiz de Sabando^{2,3,4†}, Fermín García³, Itziar Gastón⁵, Eva Ardanaz^{1,6}
and María A. Ramos-Arroyo^{2,3}

Abstract

Background: There is great heterogeneity on geographic and temporary Huntington disease (HD) epidemiological estimates. Most research studies of rare diseases, including HD, use health information systems (HIS) as data sources. This study investigates the validity and accuracy of national and international diagnostic codes for HD in multiple HIS and analyses the epidemiologic trends of HD in the Autonomous Community of Navarre (Spain).

Methods: HD cases were ascertained by the Rare Diseases Registry and the reference Medical Genetics Centre of Navarre. Positive predictive values (PPV) and sensitivity with 95% confidence intervals (95% CI) were estimated. Overall and 9-year periods (1991–2017) HD prevalence, incidence and mortality rates were calculated, and trends were assessed by Joinpoint regression.

Results: Overall PPV and sensitivity of combined HIS were 71.8% (95% CI: 59.7, 81.6) and 82.2% (95% CI: 70.1, 90.4), respectively. Primary care data was a more valuable resource for HD ascertainment than hospital discharge records, with 66% versus 50% sensitivity, respectively. It also had the highest number of “unique to source” cases. Thirty-five per cent of HD patients were identified by a single database and only 4% by all explored sources. Point prevalence was 4.94 (95% CI: 3.23, 6.65) per 100,000 in December 2017, and showed an annual 6.1% increase from 1991 to 1999. Incidence and mortality trends remained stable since 1995–96, with mean annual rates per 100,000 of 0.36 (95% CI: 0.27, 0.47) and 0.23 (95% CI: 0.16, 0.32), respectively. Late-onset HD patients (23.1%), mean age at onset (49.6 years), age at death (66.6 years) and duration of disease (16.7 years) were slightly higher than previously reported.

Conclusion: HD did not experience true temporary variations in prevalence, incidence or mortality over 23 years of post-molecular testing in our population. Ascertainment bias may largely explain the worldwide heterogeneity in results of HD epidemiological estimates. Population-based rare diseases registries are valuable instruments for epidemiological studies on low prevalence genetic diseases, like HD, as long as they include validated data from multiple HIS and genetic/family information.

*Correspondence: evicentc@navarra.es; esther.vicente@unavarra.es

†Esther Vicente and Ainara Ruiz de Sabando contributed equally to this paper and both are co-first authors

¹ Community Health Observatory Section, Instituto de Salud Pública y Laboral de Navarra, IdiSNA (Navarre Institute for Health Research), Pamplona, Spain

Full list of author information is available at the end of the article



Keywords: Huntington disease, Incidence, Mortality, Prevalence, Trends, Diagnostic codes, Positive predictive value, Sensitivity, Rare Diseases Registry

Background

Huntington disease (HD) is a rare, autosomal dominant neurodegenerative disorder caused by the abnormal expansion of a CAG repeat sequence in the Huntingtin (*HTT*) gene. An expansion of 36 or more CAGs can lead to the disease, with earlier onsets associated with longer CAG repeats. HD is characterised by progressive motor, cognitive and/or psychiatric dysfunction, with onset typically occurring in the fourth decade of life [1].

The discovery of the mutation causing the disease in 1993 enabled unambiguous genetic testing, having a profound effect in the ascertainment of HD cases. Since then, multiple studies and methodologies have aimed to estimate its prevalence across different populations, displaying a highly variable HD distribution. Although HD is universal, it presents notable geographical differences, with the highest prevalence rates in western European origin populations and the lowest in Asian and African populations [2–5]. More recently, some studies among Caucasians have reported a substantial increase in prevalence, incidence and/or mortality rates, which might indicate a time variation in HD epidemiology [6–10]. However, whether it is, in fact, a true trend or secondary to an improved ascertainment process in post molecular years has not been fully investigated.

Another factor that may contribute to the variation in results among HD studies is the validity of different sources of ascertainment. Because of the rarity of the disease, most epidemiologic studies use administrative databases or health information systems (HIS) to identify HD cases. Nevertheless, classification and coding systems in current HIS are frequently nonspecific, which may result in a lack of completeness and accuracy of HD diagnosis. In parallel, population-based registry/surveillance registries are considered key instruments to estimate incidence and prevalence rates, temporal trends and geographical distribution of low prevalent diseases [11]. Given that data sources are potential windows for ascertainment bias, selection of datasets and diagnostic validation seem critical to maximize the quality of registries and their potential success as valuable resources for epidemiological research studies.

The Population-based Rare Diseases Registry of Navarre (RERNA) is an on-going registry, created in 2013 [12], with a specific registration protocol that includes: (a) extraction of “potential cases” from all available HIS; (b) comparison of cases through health identification codes and elimination of duplicates, (c) validation

of diagnosis based on the criteria for each disease, (d) codification of validated diagnosis, (e) registration of socio-demographic variables of “confirmed cases”, and f) review of vital status and place of residence.

In addition to RERNA, Navarre counts with a clinical/genetic HD reference centre that provides services for HD patients and their families and collaborates in multicentre, multinational HD research studies. Our group has previously estimated the incidence and prevalence of HD in Navarre [13, 14]. In the present study, we aim to analyse the epidemiological trends of HD over a 27-year period in our community, and to examine the validity of different ascertainment sources, alone and in combination, used in population-based rare diseases registries (RDR).

Methods

Setting and study population

This study focuses on the population of Navarre, one of the 17 Autonomous Communities (AC) in northern Spain, with 647,554 inhabitants (50.51% women) in January of 2018, comprising 1.39% of the Spanish population [15].

The Spanish National Health System (S-NHS) is based on the principles of universality, free access, equity and fairness of financing, and is mainly funded by taxes [16]. Over 98% (637,683 individuals) of Navarre’s citizens have an individual health card with a unique 8-digit personal identification code (called CIPNA), which allows them to have access to the public health system. It contains information on birth date, sex and other socio-demographic conditions, and enables unique identification and matching of data among databases [17].

Systematic digital diagnostic coding in Navarre has not been evenly implemented for all HIS. Therefore, for the purpose of diagnostic code validation, we analysed data from the period 2000–2017 to ensure maximum ascertainment in all available data sources. Data for the epidemiology study included cases ascertained during a 27-year period, from January 1991 to December 2017. The study was approved by the Navarre Ethical Committee of Clinical Research.

Sources of ascertainment

A. Minimum Basic Data Set at Hospital Discharge (MBDS)

The MBDS is a mandatory registry for all hospitals in Spain (both public and private) which links administrative data with clinical diagnoses. Medical diagnoses are encoded using the International Classification of Diseases (ICD): the Clinical Modification of its ninth revision (ICD-9-CM) until 2015, and the Spanish Clinical Modification of its version 10 (ICD-10-ES) thereafter [18, 19]. For this study, all episodes containing 333.4 (from 2000–2015) and G10 (from 2016–2017), as primary or supplementary diagnostic codes, in the Navarre's MBDS were identified as potential HD cases.

B. Electronic Clinical Records in Primary Care (ECRPC)

The ECRPC is implanted in all Spanish regions and currently provides an on-going population-wide data source, as Primary Health Care is the first and most frequent point of contact between the population and the S-NHS [20]. In Navarre, primary care episodes are coded as per the International Classification for Primary Care (ICPC) issued by the WONCA [21]. These codes are not specific for rare diseases, but Navarre's ECRPC includes additional literal descriptors for each ICPC code, some of which are specific for a rare disease or for a group of rare diseases. For the purpose of this study, the specific descriptor "Huntington disease, chorea" was used to identify potential HD cases during the study period.

C. Temporary Work Disability Registry (TWDR)

Workers who require a sick leave are given a temporary work disability initiation form, which entitles them to receive compensation payments from the Ministry of Work. Every temporary work disability episode has assigned an ICD-9-CM diagnostic code according to the cause reported by the primary care physician [22]. For this study, all temporary work disability episodes containing 333.4 code were selected from the Navarre's TWDR during the period 2000–2017.

D. Mortality Statistics (MS)

Regional Health Ministries are in charge of the process of coding and registering the health variables of the deaths, including, the underlying cause of death (UCD) and, since 2014, the contributing cause of death (CCD) that have occurred in their territory [23]. The tenth revision of ICD coding system was adopted by the World Health

Organization in 1989, and implemented in the Spanish MS as of 1999 [24]. For this study, mortality records from Navarre, containing G10 code (both UCD, and CCD since 2014) during 2000–2017 were identified.

E. Medical Genetics Centre (MGC)

Navarre has a public reference MGC, located at the tertiary-level public hospital of the AC. Since 1991, patients with clinical signs compatible with HD and their relatives are referred to the MGC for assessment, counselling and molecular testing, when appropriate, following the HD guidelines for genetic testing [25, 26] and the pertinent signed informed consent. CAG repeat lengths are determined using PCR amplification assays with fluorescently labelled primers flanking the CAG repeat sequence [27]. The fragment size is determined by capillary electrophoresis with 3500 Genetic Analyzer (Applied Biosystems) and GeneMapper Software 5.

Demographic, clinical, family history and genetic data are collected and recorded in a disaggregated format, assigning an independent genetic family number that links to the CIPNA. Information on age at onset, age at diagnosis, parental origin of the disease, origin of family ancestors and, at least, three-generation family history is regularly obtained and revised at follow-up visits. The MGC is a site research centre for collaborative HD studies (Registry and Enroll-HD), with yearly follow-up evaluation of participants.

Validation and diagnostic criteria

Case validation was performed using information from medical records and the clinical assessment of a neurologist and a clinical geneticist, both experts in HD. A detailed chart review was carried out and pertinent information was extracted from each chart. Pedigrees were also analysed to ascertain secondary cases, defined as symptomatic relatives who were not seen in the clinic, but were reported by family members as having signs compatible with HD.

Patients were diagnosed of HD if they fulfilled one of the following inclusion criteria: (1) Individuals with neurocognitive signs compatible with HD and a genetic test result of > 35 CAG repeats in the *HTT* gene; (2) Individuals showing neurocognitive signs compatible with HD, without a genetic test result available and with a genetically confirmed HD maternal or paternal family history.

Date at diagnosis was defined as that in which symptomatic patients were clinically diagnosed with HD or when a positive genetic test result (> 35 CAG repeats) was obtained. Patients who underwent presymptomatic testing and became symptomatic within the study period

were also included, setting the diagnosis date as that of disease initiation.

Epidemiology and demographic estimates

Point prevalence of HD was calculated annually using the number of HD symptomatic individuals per 100,000 inhabitants, resident in Navarre, on the 31st of December. Age- and sex-specific prevalence was estimated for the 31st of December, 2017. Incidence and mortality rates were defined as the number of newly diagnosed symptomatic HD cases and of HD-registered deaths, respectively, per 100,000 inhabitants per year. Mean annual incidence and mortality rates per 100,000 were analysed for three periods: 1991–1999, 2000–2008 and 2009–2017. For annual age-adjusted mortality rates, we used the 2013 European Standard Population as reference [28]. Overall trends were analysed for the three epidemiologic indicators.

Statistics

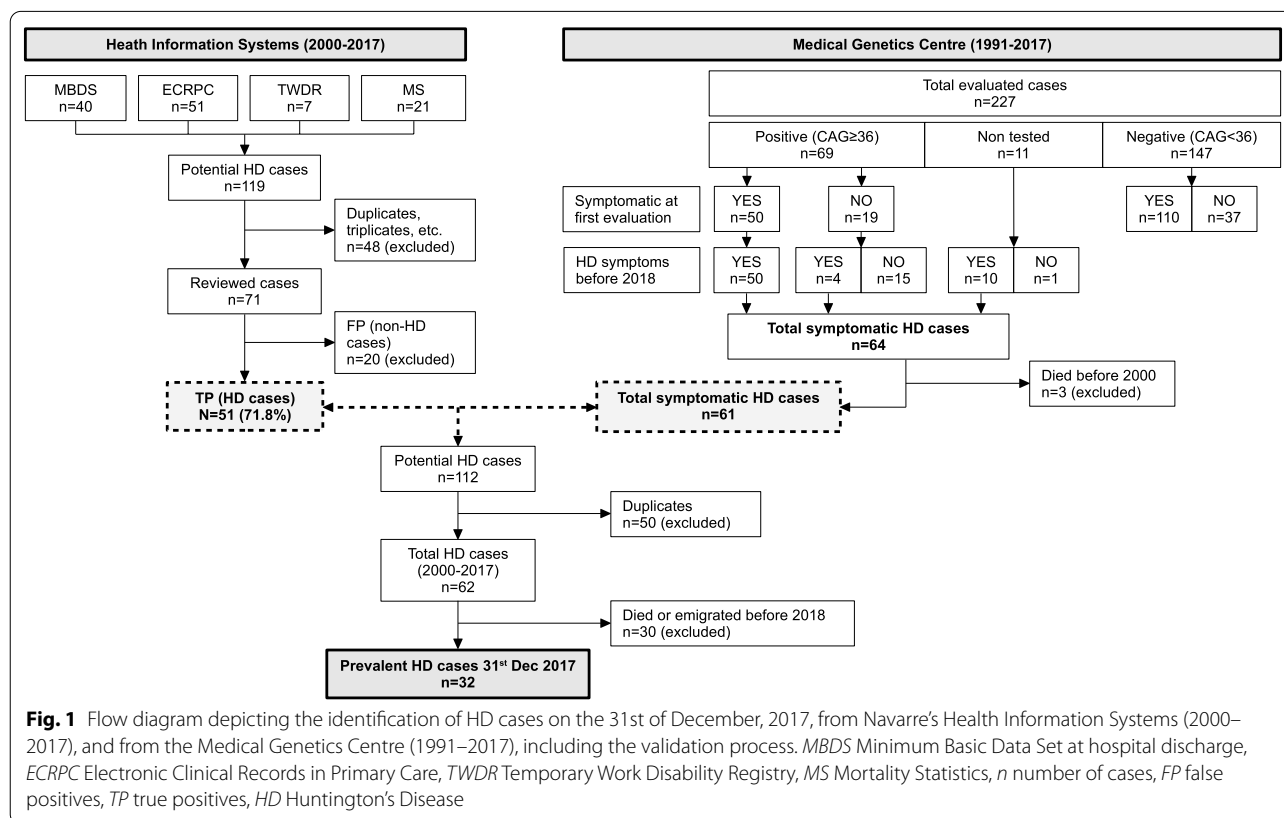
Results were summarised using descriptive statistics, such as mean and standard deviation, frequencies and proportions. Positive predictive values (PPV) and 95% confidence intervals (95% CI) were estimated for each source of ascertainment as the fraction of HD cases

that fulfilled the HD diagnostic criteria, or true positives (TP), with respect to all potential HD cases: TP and false positives (FP). Sensitivity and 95% CI were estimated as the fraction of confirmed HD cases identified by each source, with respect to the total number of HD individuals ascertained in the study (for MBDS, ECRPC, TWDR or MGC), or to the total deceased HD patients (for MS). Change-points, slopes and average annual per cent changes (AAPC) were assessed by Joinpoint regression, annually for prevalence and biannually for incidence and mortality.

Results

HD case ascertainment (HIS and MGC)

HIS captured a total number of 119 potential HD cases between 2000 and 2017: 40 from MBDS, 51 from ECRPC, 7 from TWDR, and 21 from MS. Forty-eight of them (40.3%) were identified in more than one source, and duplicates were excluded from the analysis. The remaining 71 potential HD cases were reviewed to verify the diagnosis. Fifty-one (71.8%) were confirmed as TP HD cases and 20 (28.2%) were ruled out and classified as FP. Of these, eight were incorrectly coded (50% with unspecified chorea), 10 had negative genetic test results, and two had a positive family history but the presence



of clinical signs could not be confirmed. A flow diagram of the identification and validation process of potential cases with HD is illustrated in Fig. 1.

From 1991 to 2017, a total of 227 individuals with clinical signs compatible with HD and/or family history of this disease, were evaluated at the MGC. Of them, 147 tested negative (110 symptomatic and 37 asymptomatic), 69 tested positive (50 symptomatic and 19 asymptomatic) and 11 did not get or want to be tested (10 symptomatic and one asymptomatic). Before 2018, four of the 19 asymptomatic positive cases showed neurocognitive signs, giving a total number of 64 HD manifest cases during the period (53.1% women). Three of them died before 2000. These findings are illustrated in Fig. 1.

PPVs and sensitivity

For the period 2000–2017, 62 HD cases were ascertained combining HD cases notified by MGC and/or captured by the explored HIS. Of them, 29 were registered deceased and one emigrated.

The overall PPV for the combined HIS was 71.8% (95% CI: 59.7, 81.6). All FP cases were captured by only one source: nine by MBDS, 10 by ECRPC, none by TWDR and one by MS, yielding a PPV (95% CI) of 75.5% (61.2, 88.6), 80.4% (66.5, 89.7), 100% (56.1, 100) and 95.2% (74.1, 99.8), respectively (Table 1). ECRPC had the highest number of “unique to source” cases, in contrast to TWDR, with none. Eighteen cases (35%) were identified in a unique source and only two by all explored sources (4%). A Venn diagram illustrating HD confirmed cases and their overlap in the different HIS is presented in Fig. 2.

Sensitivity (95% CI), estimated for all combined HIS, was 82.2% (70.1, 90.4), with great variation among them: 50% (37.2, 62.8) for MBDS, 66.1% (52.9, 77.4) for ECRPC, 11.3% (5, 22.5) for TWDR, and 69% (49.1, 84) for MS (Table 1).

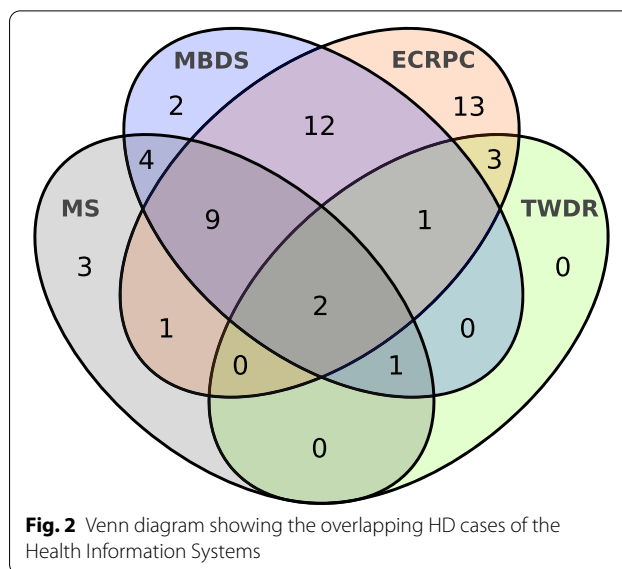


Fig. 2 Venn diagram showing the overlapping HD cases of the Health Information Systems

Prevalence, incidence and mortality rates

A total of 80 HD cases were identified between 1991 and 2017, corresponding to 42 families. Of them, 69 cases were genetically confirmed, while 11 patients were diagnosed based on a genetically positive HD first-degree relative and the manifestation of neurological signs. On average, three new HD positive cases were identified per year throughout the study period (1991–2017), 2.3 symptomatic and 0.7 asymptomatic (Table 2), showing an increasing trend of pre-manifest testing in detriment of the symptomatic.

There were 65 HD manifest cases (54% women) during 1991–2017, which gave an estimated prevalence per 100,000 (95% CI) of 5.7 (3.69, 7.71) for 31/12/1999, 5.07 (3.31, 6.83) for 31/12/2008 and 4.94 (3.23, 6.65) for 31/12/2017 (Table 2). Of them, 63 were incident throughout the study period, 1991–2017. By 9-year intervals, 1991–1999, 2000–2008 and 2009–2017, mean annual incidence rates per 100,000 (95% CI) were 0.46 (0.30, 0.69), 0.39 (0.25, 0.59) and 0.35 (0.22, 0.53), respectively (Table 2). A total of 32 deaths were recorded throughout

Table 1 Potential HD cases, true positives (TP), false positives (FP), positive predictive value (PPV) and sensitivity of all sources of ascertainment

Source of ascertainment	Potential HD cases	TP	FP	PPV (95% CI)	Sensitivity (95% CI)
MBDS	40	31	9	75.5% (61.2, 88.6)	50.0% (37.2, 62.8)
ECRPC	51	41	10	80.4% (66.5, 89.7)	66.1% (52.9, 77.4)
TWDR	7	7	0	100% (56.1, 100)	11.3% (5.0, 22.5)
MS	21	20	1	95.2% (74.1, 99.8)	69.0% (49.1, 84.0)
MGC	61	61	0	100% (92.6, 100)	98.4% (90.2, 100)

MBDS minimum basic data set at hospital discharge, ECRPC electronic clinical records in primary care, TWDR Temporary Work Disability Registry, MS mortality statistics, MGC Medical Genetics Centre

Table 2 Sex-, age- and period-specific estimates of prevalence and incidence of Huntington disease (HD) in Navarre, for the period 1991–2017

	Number of HD cases			Prevalence		Incidence	
	Total	Symptomatic	Asymptomatic (n) ^a	Total prevalent cases	Point prevalence rate per 100,000 (95% CI)	Total incident cases	Annual incidence rate per 100,000 (95% CI)
Total	80	61	19	65		63	0.40 (0.31, 0.50)
Sex							
Men	38 (47%)	26	12 (4)	15 (47%) ^b	4.68 (2.31, 7.05) ^b	29 (46%)	0.37 (0.25, 0.52)
Women	42 (53%)	35	7	17 (53%) ^b	5.20 (2.73, 7.67) ^b	34 (54%)	0.43 (0.30, 0.59)
Age (years)							
≤ 20	4	1	3 (1)	0 ^b	0 ^b	1	0.03 (0, 0.16)
21–40	22	8	14 (2)	4 ^b	2.64 (0.05, 5.22) ^b	10	0.19 (0.10, 0.35)
41–60	28	26	2 (1)	10 ^b	5.11 (1.94, 8.27) ^b	26	0.63 (0.42, 0.91)
> 60	26	26	0	18 ^b	11.53 (6.20, 16.85) ^b	26	0.78 (0.52, 1.12)
Period							
1991–1999	26	23	3 (2)	31 ^c	5.70 (3.69, 7.71) ^c	22	0.46 (0.30, 0.69)
2000–2008	28	21	7 (1)	32 ^d	5.07 (3.31, 6.83) ^d	21	0.39 (0.25, 0.59)
2009–2017	24	15	9 (1)	32 ^b	4.94 (3.23, 6.65) ^b	20	0.35 (0.22, 0.53)

^a Number of HD presymptomatic cases that became symptomatic throughout the study period

Point prevalence dates: ^b31/12/2017, ^c31/12/1999, ^d31/12/2008

the study, with mean annual rates per 100,000 (95% CI) for each 9-year period of 0.06 (0.02, 0.17), 0.3 (0.18, 0.49) and 0.23 (0.13, 0.38), respectively. Figure 3 shows the annual rates of prevalence, incidence and mortality during 1991–2017.

Regarding evolution, Joinpoint analysis for prevalence showed a change-point in 1999 (95% CI: 1996–2002), with an increasing slope until 1999 at an AAPC of 6.1 (95% CI: 3.8, 8.5) and a slightly decreasing slope from 1999 onwards with AAPC of -0.7 (95% CI: -1.4, -0.2) (Fig. 4). Incidence and mortality experienced unstable annual rates during the first period (1991–1999), showing a distinctive incidence peak in 1993–1994 and a very low number of deaths, respectively. However, from the biennium 1995–1996 onwards, both annual rates remained stable, with no relevant changes. Joinpoint results for incidence and mortality during this 23-year period did not show trend changes or significant slopes, with AAPC (95% CI) of 0.4 (-3.4, 4.3) and 2.1 (-4.1, 8.7), respectively. Overall annual incidence and mortality rates per 100,000 (95% CI) during 1995–2017 were 0.36 (0.27, 0.47) and 0.23 (0.16, 0.32), respectively. Age-adjusted mortality rate was 0.24 per 100,000 (95% CI: 0.09, 0.76).

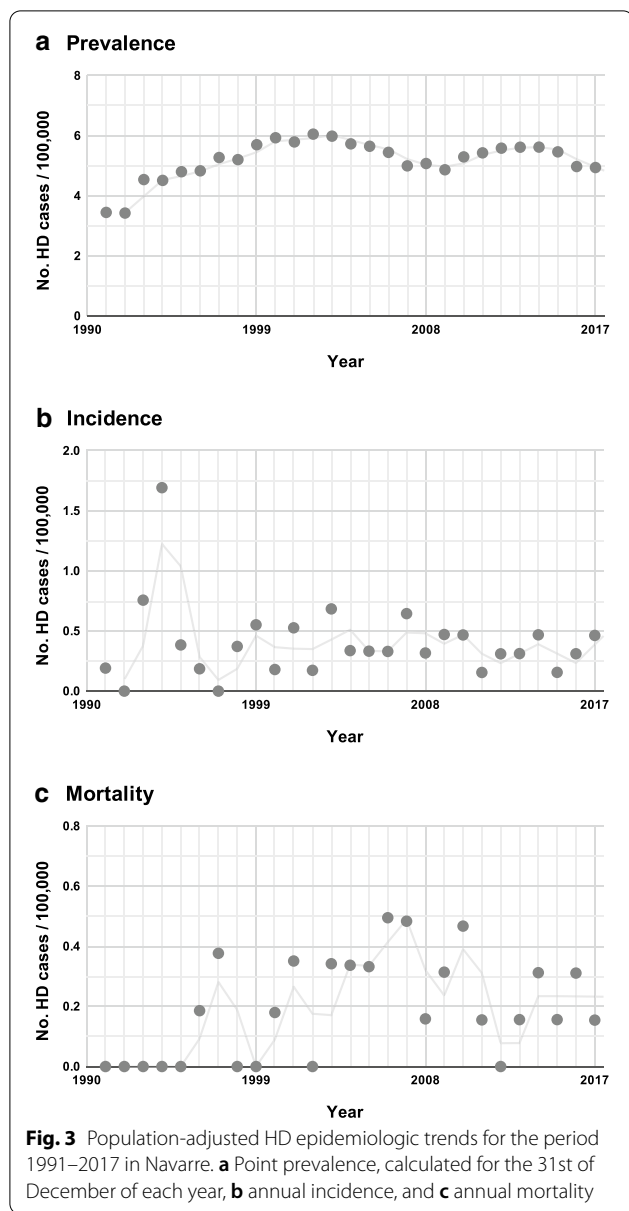
The mean number of CAG repeats in the *HTT* gene was 43 ± 3.8 (range 37–57 CAGs, $n=69$). HD cases had a mean age at onset of 49.8 ± 18.8 years (range 14–82, $n=63$) (Fig. 5) and a mean age at incidence of 55.7 ± 16.5 years (range 15–85, $n=63$). Juvenile forms of HD represented 6.2% of cases, and 23.1% showed

symptoms after 60 years of age (late-onset HD). Mean CAG length for late-onset cases was 39.5 ± 0.91 (range 38–41, $n=12$). Across the studied period, 32 patients were deceased at an average age of 66.6 ± 17.1 years (range 27–90) and, overall, the disease duration was 16.7 ± 8.1 years (range 4–39, $n=31$). Suicide was the cause of death in 6.3% of cases.

Discussion

The present study is the first to validate the accuracy and sensitivity of the main HD diagnostic codes in different routinely collected health-care datasets, alone and in combination, using medical records as the gold standard. We also provide unbiased HD epidemiological estimates and trends from 1991 to 2017, in a well-defined geographic region in northern Spain, Navarre, using supplementary clinical, genetic and family data from the genetic reference centre of the region.

Approximately 2/3 of HD cases identified across all four HIS were confirmed by review of clinical/genetic records, with individual dataset PPVs ranging from 76% for hospital discharges data to 100% for temporary working leave information. PPVs for primary care and mortality data codes were 80% and 95%, respectively. These figures are within the range of those reported for other neurodegenerative diseases, such as Parkinson Disease [29], Charcot Marie Tooth [30], dementia [31], Guillain-Barré syndrome [32] or Duchenne/Becker muscular dystrophy [33].



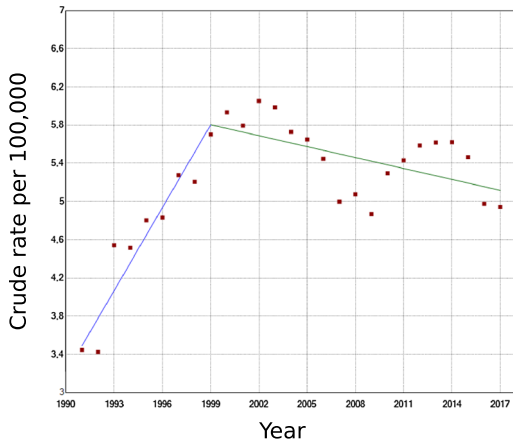
Sensitivity, however, was more diverse among datasets (from 11% in TWDR to 69% in MS). Primary care data was a more valuable resource for HD case ascertainment than hospital discharge electronic records. Combination of both sources identified 77% of all HD cases with a PPV of 72%. TWDR and MS databases presented the highest PPV of explored HIS (100% and 95%, respectively); on the contrary, they had lower capacity to identify HD cases. Similar variation in sensitivity of routinely collected health-care data has also been observed for Parkinson disease [29]. We are not aware, however, of analogous studies for HD or other rare diseases with genetic

diagnosis; therefore, further comparative analysis of our results is not yet feasible.

Epidemiologic findings show that, on 31st December 2017, the prevalence of manifest HD in Navarre was 4.94 per 100,000, with an average annual incidence rate of 0.36 per 100,000 inhabitants (1995–2017). These estimates are within the contemporary European range [3, 34] and in concordance with our general population CAG repeat length distribution (data not shown). Prevalence, however, is lower than that reported in the United Kingdom [7] or Ireland [35], higher than in some northern European countries like Finland [36] or Iceland [37], and in line with other southern European populations [38–41]. Other post molecular studies carried out in Spain have also reported comparable HD prevalence figures (4.6 and 4.0, per 100,000 in Asturias and Murcia, respectively [42, 43]; in contrast, lower prevalence was observed for Balearic Islands (2 per 100,000) [43]. It is likely, however, that this low prevalence rate reflects incomplete ascertainment, given the limitations in data sources and length of the study period (four years). Similarly, we observed a higher mean annual adjusted mortality rate (0.24 per 100,000 during 1991–2017) than that previously reported in Spain for an overlapping period (0.08 and 0.15 per 100,000 in 1991 and 2013, respectively) [9]. Worldwide, however, HD mortality remains understudied, with a few, mainly pre-molecular reports, showing comparable rates to ours in the United States (0.23/100,000) [44] and lower in Austria (0.13/100,000) [45].

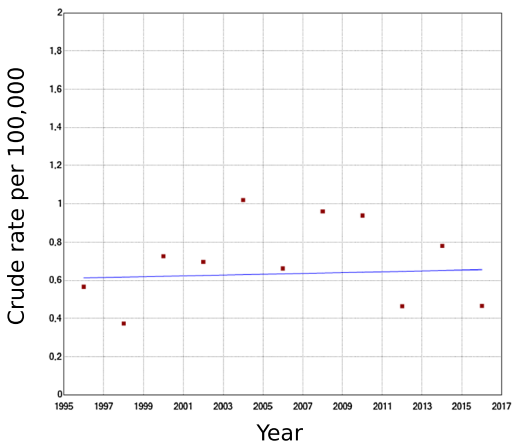
It is generally accepted that availability of direct HD testing increased ascertainment of cases, by the identification of patients with unknown HD family history, which occurs in approximately 10–16% of cases [7, 13]. Consequently, HD overall prevalence and incidence estimates are higher than in pre-testing decades [7, 8, 34, 36]. Nevertheless, there is still wide geographic variation among studies, which cannot be fully explained by the population genetic background, including the pool of intermediate CAG repeat alleles and *HTT* haplotypes. Moreover, there is some evidence of a potential trend of increasing HD rates in some populations [7, 10]. However, interpretation of results is controversial, as most studies differ in demographic characteristics, case-sources and case-ascertainment methods. Our study demonstrates that prevalence, incidence and mortality rates of HD in our population did not experience a true increase over time, showing stable estimated trends over the last 23-years of post-molecular HD testing. Interestingly, prevalence rates showed an increasing trend during 1991–1999, while incidence was slightly higher than in the following years (2000–2017). As shown in Fig. 3, annual incidence experienced a distinctive peak in 1993–1994, suggesting that the excess of incident cases in the

a Prevalence



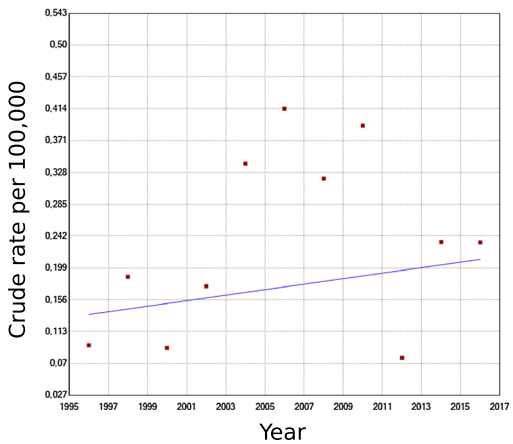
Joinpoints (95% CI)	Period	AAPC (95% CI)	Slopes		Slope 1 – slope2	
			b	p-value	b	p-value
1999 (1996, 2002)	1991-1999 (slope 1)	6.1 (3.8, 8.5)	0.06	<0.001	-0.067	<0.001
	2000-2017 (slope 2)	-0.7 (-1.3, -0.2)	-0.007	0.012		

b Incidence



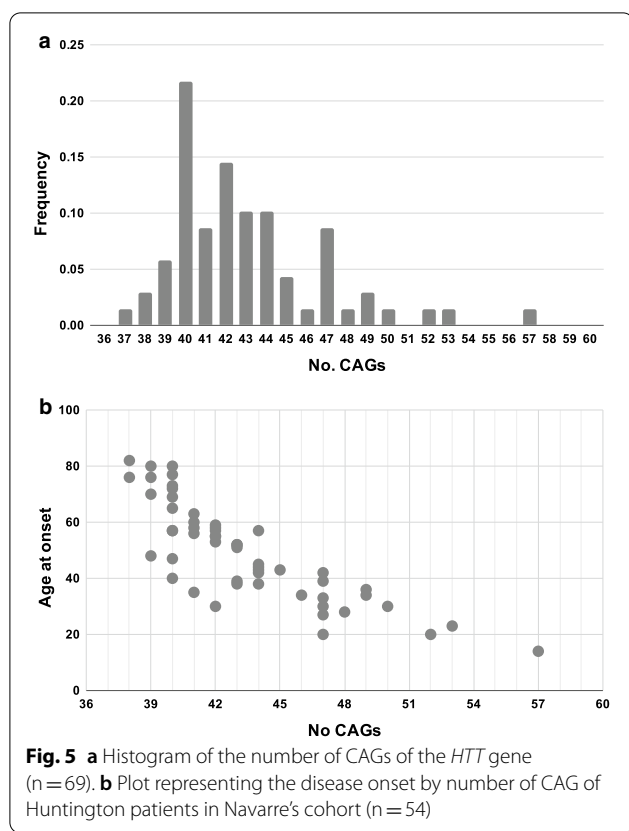
Joinpoints (95% CI)	Period	AAPC (95% CI)	Slopes	
			b	p-value
none	biannually 1995-2016	0.4 (-3.4, 4.3)	0.004	0.830

c Mortality



Joinpoints (95% CI)	Period	AAPC (95% CI)	Slopes	
			b	p-value
none	biannually 1995-2016	2.1 (-4.1, 8.7)	0.021	0.472

Fig. 4 Huntington disease prevalence **a**, incidence **b** and mortality **c** trends, change points, average annual per cent changes (AAPC) and slopes using a Joinpoint regression model



first study period is, most likely, a consequence of the availability of direct HD testing which allowed the identification of previously suspected, but undiagnosed, HD cases. It resulted in a 6.1% average increase in annual prevalence until 1999, followed by a very slight decrease (0.7%) thereafter. The number of prevalent cases did not vary significantly over this period, but the total population experienced a 20% increase since 1991 (data not shown). It is, therefore, conceivable that demographic changes in the population might have contributed to slightly decrease the prevalence trends. With respect to mortality, no deaths from HD were recorded before 1996, but improvement in HD ascertainment resulted in stable annual rates thereafter. Very low mortality rates have been also observed in Spain in the late 1980s with increasing trends until 2013 [9].

Most epidemiologic HD studies use health-care databases as the main source of ascertainment. Our study strongly suggests that ascertainment bias may be an important factor that could explain, at least in part, geographic and temporary differences in reported HD prevalence and incidence rates. According to our results, individual hospital discharges and primary health-care datasets might miss 30–50% of prevalent HD cases and include over 20% of non-HD patients.

Misclassification of cases mainly involved: a) under-ascertainment of late-onset patients showing neurocognitive signs commonly seen in other relatively frequently diseases, like Alzheimer disease, obsessive-compulsive disorder and other dementias and psychiatric illnesses, and, b) inclusion of asymptomatic mutation carriers and negative HD family members. In the present study, 47% of FP in hospital and primary care datasets were asymptomatic members of HD families, either with unknown or negative genetic testing. In this regard, it is worth mentioning that, although uptake of HD predictive testing is overall low [46], the expectation for better medical interventions, including the availability of preconceptual diagnosis and potentially promising new treatments, may result in a temporary increase of genetic testing in asymptomatic individuals. We, in fact, observed that over one-third of all HD positive cases identified during the most recent study period (2000–2017), were asymptomatic, a proportion three times higher than that during 1991–1999. It would be interesting to investigate this issue in other and larger populations and its possible effect as a potential bias in HD ascertainment.

To overcome the above-mentioned limitations of individual HIS ascertainment, several studies used multiple sources of information, yielding higher true prevalence HD rates [8, 10]. As a counterpart, however, these studies are more likely to double/triple-count a relatively large number of individuals. In our analysis, 43% of potential HD cases were included in both primary care and hospital databases, and 64.5% would be double-counted when combined with mortality dataset. Consequently, minimizing overestimation of true HD prevalence/incidence in multiple sources ascertainment studies requires further highly time-consuming investigations, which may not be feasible when dealing with large populations. Finally, we also proved that the inclusion of genetic and family data is a relevant source that adds high validity to case ascertainment. The MGC ascertained 14 ‘unique to source’ cases, corresponding to 23% of HD cases, and identified 65% of FP. We, therefore, conclude that population-based RDR are potentially a highly valuable instrument to conduct epidemiological studies on low prevalence, like HD, providing the inclusion of multiple health and administrative validated sources of information, in conjunction with genetic and family data.

As expected, demographic characteristics of HD patients in Navarre were similar to those reported in most Caucasian populations. We observed, however, some interesting differences. HD natural history seems to present with a wider range in the timing of initiation of signs, higher proportion of late-onset HD cases (23.1%), and longer overall survival than previously estimated [8,

38, 39, 41, 47–49]. This is most likely due to improvements in clinical and molecular HD ascertainment, which, in our population, resulted in the identification of a high proportion of cases with low-penetrant alleles (10%). Additional circumstances, such as better health-care interventions may have also contributed to increase quality of life and extended life expectancy.

The main limitation of the present study is the small population coverage and sample size of HD cases, given the low prevalence of the disease. In addition, variability in access to health-care systems and diagnostic coding specificity in different populations could limit the possibility of extrapolating our diagnostic code validation results to a national or international scale. We must mention, to this respect, that the annual adjusted mortality rate in our regional study was 40% higher (0.24 per 100,000) than the overall rate previously reported in Spain (0.15 per 100,000) using MS for an overlapping period (1991–2013) [9]. This difference is in concordance with our results on sensitivity of the national mortality dataset, supporting the value of this validation study. Finally, the strength of the present work lies in the study design, an HD population-based analysis of nearly three decades, with complete case ascertainment, using clinical and genetic data as reference standards.

Conclusions

We present the first HD diagnostic code validation analysis for different HIS, and demonstrate that epidemiological estimates for this rare disease in Navarre do not show true temporary variations during the last decades of post molecular testing. Improved HD ascertainment may decrease heterogeneity among worldwide HD epidemiological studies and result in a higher identification of low-penetrant allele carriers that will widen the knowledge of the natural history of the disease.

Abbreviations

95% CI: 95% Confidence intervals; AAPC: Average annual per cent changes; AC: Autonomous Communities; CCD: Contributing cause of death; ECRPC: Electronic Clinical Records in Primary Care; FP: False positives; HD: Huntington Disease; HIS: Health information systems; ICD: International Classification of Diseases; ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification; ICD-10-ES: International Classification of Diseases, Tenth Revision, Clinical Modification (Spanish adaptation); ICPC: International Classification for Primary Care; MBDS: Minimum Basic Data Set at Hospital Discharge; MGC: Medical Genetics Centre; MS: Mortality Statistics; PPV: Positive predictive values; RDR: Rare diseases registries; RERNA: Population-based Rare Diseases Registry of Navarre; S-NHS: Spanish National Health System; TP: True positives; TWDR: Temporary Work Disability Registry; UCD: Underlying cause of death.

Acknowledgements

We very much appreciate the collaboration of HD families for their contribution in obtaining reliable personal and family information. We would also like to thank Dr. Berta Ibañez from the Methodology Unit of Navarrabiomed for her assistance in the biostatistical analysis.

Authors' contributions

EV conception and design, database management, writing—original draft, review and editing. ARS clinical database management, data analysis, writing—original draft, review and editing. FG genetic analysis. IG collection and revision of clinical data. EA database management, writing—review and editing. MARA conception and design, collection and revision of clinical data, witting—original draft and critical revision, study supervision. All authors read and approved the final manuscript.

Funding

This study was funded by the Instituto de Salud Carlos III (Grant No. FIS PI15/02227), and the CHDI Foundation (Enroll-HD study). ARS was supported by a Fellowship of Departamento de Salud, Gobierno de Navarra (ref. 0011-4809-2018-000001). Funders had no role in the design, conduct, analysis or reporting of this study.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate

This research study was conducted retrospectively from data obtained for clinical purposes. The study was approved by the "Comité de Ética de la Investigación con medicamentos (CEIm)" of Gobierno de Navarra (PI_2015/78. Res. 662, 18th October 2019).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Community Health Observatory Section, Instituto de Salud Pública y Laboral de Navarra, IdiSNA (Navarre Institute for Health Research), Pamplona, Spain. ² Department of Health Sciences, Universidad Pública de Navarra, IdiSNA, Pamplona, Spain. ³ Department of Medical Genetics, Complejo Hospitalario de Navarra, IdiSNA, Pamplona, Spain. ⁴ Fundación Miguel Servet-Navarrabiomed, IdiSNA, Pamplona, Spain. ⁵ Department of Neurology, Complejo Hospitalario de Navarra, IdiSNA, Pamplona, Spain. ⁶ CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain.

Received: 13 August 2020 Accepted: 19 January 2021

Published online: 10 February 2021

References

- Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt BR, et al. Huntington disease. *Nat Rev Dis Prim*. 2015;1:15005.
- Chen Y-Y, Lai C-H. Nationwide population-based epidemiologic study of Huntington's disease in Taiwan. *Neuroepidemiology*. 2010;35:250–4.
- Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N. The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. *Mov Disord*. 2012;27:1083–91.
- Rawlins MD, Wexler NS, Wexler AR, Tabrizi SJ, Douglas I, Evans SJW, et al. The prevalence of Huntington's disease. *Neuroepidemiology*. 2016;46:144–53.
- Kay C, Hayden MR, Leavitt BR. Epidemiology of Huntington disease. In: Feigin AS, Anderson KE, editors. *Handbook of clinical neurology*, vol. 144 (3rd series): Huntington disease. Amsterdam: Elsevier B.V.; 2017. p. 31–46.
- Morrison PJ. Prevalence estimates of Huntington disease in Caucasian populations are gross underestimates. *Mov Disord*. 2012;27:1707–8.
- Evans SJW, Douglas I, Rawlins MD, Wexler NS, Tabrizi SJ, Smeeth L. Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records. *J Neurol Neurosurg Psychiatry*. 2013;84:1156–60.
- Fisher ER, Hayden MR. Multisource ascertainment of Huntington disease in Canada: prevalence and population at risk. *Mov Disord*. 2014;29:105–14.

9. Sánchez-Díaz G, Arias-Merino G, Villaverde-Hueso A, Morales-Piga A, Abaitua-Borda I, Hens M, et al. Monitoring Huntington's disease mortality across a 30-year period: geographic and temporal patterns. *Neuroepidemiology*. 2016;47:155–63.
10. Squitieri F, Griguoli A, Capelli G, Porcellini A, D'Alessio B. Epidemiology of Huntington disease: first post-HTT gene analysis of prevalence in Italy. *Clin Genet*. 2016;89:367–70.
11. Kodra Y, Posada de la Paz M, Coi A, Santoro M, Bianchi F, Ahmed F, et al. Data quality in Rare Diseases Registries. In: Posada de la Paz M, Taruscio D, Groft SC, editors., et al., *Rare diseases epidemiology: update and overview*. 2nd ed. Cham: Springer; 2017. p. 149–64.
12. Vicente E, Guevara M, Lasanta MJ, Ramos-Arroyo MA, Ardanaz E. Implementing a population-based rare diseases registry in Spain: Navarre's experience. *Rev Esp Salud Publica*. 2018;92:e201811082.
13. Ramos-Arroyo MA, Moreno S, Valiente A. Incidence and mutation rates of Huntington's disease in Spain: experience of 9 years of direct genetic testing. *J Neurol Neurosurg Psychiatry*. 2005;76:337–42.
14. Vicente E, García-Amigot F, Gaston MI, Nuin-Villanueva MA, Hernandez B, Ardanaz E, et al. Prevalence of Huntington disease in Navarra (Spain). Sensitivity and positive predictive value of different sources of ascertainment. *J Neurol Neurosurg Psychiatry*. 2014;85:A77.
15. Instituto Nacional de Estadística. <http://www.ine.es>. Accessed 15 July 2020.
16. Bernal-Delgado E, García-Armesto S, Oliva J, Sánchez-Martínez FI, Repullo JR, Peña-Longobardo LM, et al. Spain: health system review. *Health Syst Transit*. 2018;20:1–179.
17. Servicio Navarro de Salud - Osasunbidea. Memoria 2018 del Servicio Navarro de Salud - Osasunbidea. http://www.navarra.es/home_es/Temas/Portal+de+la+Salud/Profesionales/Documentacion+y+publicaciones/Memorias+de+actividad/Ano+2018/Memoria+del+SNS-O+2018.htm. Accessed 15 July 2020.
18. Fernández-Navarro P, López-Abente G, Salido-Campos C, Sanz-Anquela JM. The Minimum Basic Data Set (MBDS) as a tool for cancer epidemiological surveillance. *Eur J Intern Med*. 2016;34:94–7.
19. Hernández Medrano I, Guillán M, Masjuan J, Alonso Cánovas A, Gogorcena MA. Reliability of the minimum basic dataset for diagnoses of cerebrovascular disease. *Neurología*. 2017;32:74–80.
20. Esteban-Vasallo MD, Domínguez-Berjón MF, Astray-Mochales J, Génova-Maleras R, Pérez-Sania A, Sánchez-Perruca L, et al. Epidemiological usefulness of population-based electronic clinical records in primary care: estimation of the prevalence of chronic diseases. *Fam Pract*. 2009;26:445–54.
21. WONCA. ICDPC-2-R: International Classification of Primary Care (Oxford Medical Publications). 2005. <http://www.ph3c.org/PH3C/docs/27/000496/0000908.pdf>. Accessed 15 July 2020.
22. Abásolo L, Lajas C, León L, Carmona L, Macarrón P, Candelas G, et al. Prognostic factors for long-term work disability due to musculoskeletal disorders. *Rheumatol Int*. 2012;32:3831–9.
23. Floristán-Floristán Y, Delfrade-Osinaga J, Carrillo-Prieto J, Aguirre-Pérez J, Moreno-Iribas C. Codificación de las causas de muerte con el software IRIS. Impacto en las estadísticas de mortalidad de la Comunidad Foral de Navarra. *Rev Esp Salud Publica*. 2016;90:e1–9.
24. Gómez-Redondo R, Faus-Bertomeu A, Poniakina S, Cámara-Izquierdo N. About Spain data on causes of death. 2017. https://www.causesofdeath.org/Data/ESP/20170126/ESP_bd.pdf. Accessed 15 July 2020.
25. Guidelines for the molecular genetics predictive test in Huntington's disease. International Huntington Association (IHA) and the World Federation of Neurology (WFN) Research Group on Huntington's Chorea. *Neurology*. 1994;44:1533–6.
26. Ramos Arroyo MA, Trujillo-Tiebas MJ, Milá M. Recomendaciones de buena práctica para el diagnóstico genético de la enfermedad de Huntington. *Med Clin (Barc)*. 2012;138:584–8.
27. Rubinsztein DC, Barton DE, Davison BCC, Ferguson-Smith MA. Analysis of the huntingtin gene reveals a trinucleotide-length polymorphism in the region of the gene that contains two CCG-rich stretches and a correlation between decreased age of onset of Huntington's disease and CAG repeat number. *Hum Mol Genet*. 1993;2:1713–5.
28. The 2013 European Standard Population. <https://www.causesofdeath.org/docs/standard.pdf>. Accessed 15 July 2020.
29. Harding Z, Wilkinson T, Stevenson A, Horrocks S, Ly A, Schnier C, et al. Identifying Parkinson's disease and parkinsonism cases using routinely collected healthcare data: a systematic review. *PLoS ONE*. 2019;14:e0198736.
30. Vaeth S, Jensen UB, Christensen R AH. Validation of diagnostic codes for Charcot – Marie – Tooth disease in the Danish National Patient Registry. *Clin Epidemiol*. 2016;783–7.
31. Wilkinson T, Schnier C, Bush K, Rannikmäe K, Henshall DE, Lerpiniere C, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur J Epidemiol*. 2019;34:557–65.
32. Levison LS, Thomsen RW, Christensen DH, Mellemkjær T, Sindrup SH, Andersen H. Guillain-Barré syndrome in Denmark: validation of diagnostic codes and a population-based nationwide study of the incidence in a 30-year period. *Clin Epidemiol*. 2019;11:275–83.
33. Soslow JH, Hall M, Burnette WB, Hor K, Chisolm J, Spurney C, et al. Creation of a novel algorithm to identify patients with Becker and Duchenne muscular dystrophy within an administrative database and application of the algorithm to assess cardiovascular morbidity. *Cardiol Young*. 2019;29:290–6.
34. Baig SS, Strong M, Quarrell OW. The global prevalence of Huntington's disease: a systematic review and discussion. *Neurodegener Dis Manag*. 2016;6:331–43.
35. Morrison P, Harding-Lester S, Bradley A. Uptake of Huntington disease predictive testing in a complete population. *Clin Genet*. 2011;80:281–6.
36. Sipilä JOT, Hietala M, Siitonen A, Päiväranta M, Majamaa K. Epidemiology of Huntington's disease in Finland. *Parkinsonism Relat Disord*. 2015;21:46–9.
37. Sveinsson Ó, Halldórsson S, Ólafsson E. An unusually low prevalence of Huntington's disease in Iceland. *Eur Neurol*. 2012;68:48–51.
38. Carrassi E, Pugliatti M, Govoni V, Sensi M, Casetta I, Granieri E. Epidemiological study of Huntington's disease in the province of Ferrara. *Italy Neuroepidemiology*. 2017;49:18–23.
39. Demetriou CA, Heraclides A, Salafoi C, Tanteles GA, Christodoulou K, Christou Y, et al. Epidemiology of Huntington disease in Cyprus: a 20-year retrospective study. *Clin Genet*. 2018;93:656–64.
40. Peterlin B, Kobal J, Teran N, Flisar D, Lovrečić L. Epidemiology of Huntington's disease in Slovenia. *Acta Neurol Scand*. 2009;119:371–5.
41. Panas M, Karadima G, Vassos E, Kalfakis N, Kladi A, Christodoulou K, et al. Huntington's disease in Greece: the experience of 14 years. *Clin Genet*. 2011;80:586–90.
42. Margolles P. Enfermedades raras en Asturias: enfermedad de Huntington. 2011. https://www.astursalud.es/documents/31867/36150/ER+IB+04+ENFERMEDAD+DE+HUNTINGTON_DL.pdf/53e6f6b8-2603-7c65-17fc-44a0d6836a11. Accessed 15 July 2020.
43. Cárffaro Rovira M, Salom Castell MM. Prevalencia y mortalidad de la enfermedad de Huntington a partir del Registro Poblacional de Enfermedades Raras de las Islas Baleares durante el período 2010–2013. *Rev Esp Salud Publica*. 2017;91:e201702020.
44. Lanska DJ, Lavine L, Lanska MJ, Schoenberg BS. Huntington's disease mortality in the United States. *Neurology*. 1988;38:769–72.
45. Eksterne E, Lebhart G. Long-term monitoring of the mortality trend of Huntington's disease in Austria. *Eur J Epidemiol*. 2005;20:169–72.
46. Baig SS, Strong M, Rosser E, Taverner NV, Glew R, Miedzybrodzka Z, et al. 22 Years of predictive testing for Huntington's disease: the experience of the UK Huntington's Prediction Consortium. *Eur J Hum Genet*. 2016;24:1396–402.
47. Chaganti SS, McCusker EA, Loy CT. What do we know about late onset Huntington's disease? *J Huntingtons Dis*. 2017;6:95–103.
48. Kim H-J, Shin C-W, Jeon B, Park H. Survival of Korean Huntington's disease patients. *J Mov Disord*. 2016;9:166–70.
49. Rodrigues FB, Abreu D, Damásio J, Goncalves N, Correia-Guedes L, Coelho M, et al. Survival, mortality, causes and places of death in a European Huntington's disease prospective cohort. *Mov Disord Clin Pract*. 2017;4:737–42.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.