**Supplementary methods**

**Literature review of published Huntington’s disease epidemiology data**

A pragmatic review of published epidemiology data in Huntington’s disease (HD) was conducted. As a first step, general web searches, using Google Scholar and the Google search engine, were performed to identify relevant research articles on the epidemiology of HD published up to May 2019. Secondly, a pragmatic literature search of PubMed was conducted in April and May 2019 using terms specific to HD as well as incidence, prevalence, and epidemiology to identify articles on the epidemiology of HD published between January 2016 and May 2019 (shown in Supplementary Table 6). A manual screen of the search results was conducted and papers with titles suggesting inclusion of comorbid conditions, disease other than HD, or studies from regions outside the countries/regions of interest were excluded. The abstracts of the remaining studies were reviewed to understand their objectives and the results, and to identify potential data sets for the analysis process. The full articles of the selected abstracts were reviewed and shortlisted for inclusion if they contained data within the scope of the research. One article by Ohlmeier et al. [1] published after May 2019 was included in the analysis during refinement. Data extraction from the identified articles captured geographical region, time period studied, sample size, methodology including limitations, epidemiology data (incidence and prevalence), case definition, median age at diagnosis, definitions of staging criteria, stage distributions among diagnosed patients, and survival.

**Extrapolation of survival curves**

Parametric survival analysis methods were used to derive complete time-to-event curves. Six possible statistical distributions and fits of the models were evaluated using the flexsurv R package [2] The best-fitting models were identified based on comparison of Akaike information criterion values and the Bayesian information criterion values, as well as plausibility of the extrapolated values. Time-varying annual probabilities of death or progression estimated from the chosen progression-free survival (PFS) model were applied to each cohort to estimate the probability of remaining in that stage for the subsequent year. Because PFS was defined as not moving out of the current state (composite of death or stage progression), the proportion of patients moving from one stage to the next stage each year was estimated as the difference between the values of the extrapolated overall survival (OS) and PFS curves at each time point.

**One-way sensitivity analysis**

A one-way sensitivity analysis was performed using high-, medium-, and low-incidence inputs (shown in Supplementary Table 1), based on published ranges in the studies identified in the literature review and with respect to geographical specificity where possible.

**Supplementary tables**

**Supplementary Table 1.** Base and peak diagnosed incidences used to estimate diagnosed incidence of manifest Huntington’s disease

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Incidence input | Base value (1950), per 100,000 | Peak value (2006a), per 100,000 |
| Brazil | High | 0.50b | 0.72c |
| Medium | 0.21d | 0.29e |
| Low | 0.13d | 0.16f |
| Canada | High | 0.50b | 0.90g |
| Medium | 0.69h |
| Low | 0.61i |
| France, Germany, and UK | High | 0.50b | 0.90g |
| Medium | 0.72c |
| Low | 0.61i |
| Italy | High | 0.50j | 0.90g |
| Medium | 0.50j | 0.72c |
| Low | 0.30j | 0.30j |
| Spain | High | 0.50k | 0.90g |
| Medium | 0.50k | 0.72c |
| Low | 0.47k | 0.47k |
| USA | High | 0.50b | 0.90g |
| Medium | 0.69h |
| Low | 0.50b |

Linear extrapolation was used from 1950 to 2006.

a2006 is the midpoint year of the recent incidence inputs identified in the literature review.

bBased on data from Kokmen et al. 1994 [3] cBased on data from Wexler et al. 2016 [4] dBased on proportional adjustment of data from Kokmen et al. 1994 [3] eBack-calculated from prevalence based on unpublished data from the Brazilian Huntington’s Disease Patients’ Association. fBack-calculated from prevalence based on Castilhos et al. 2019 [5] gBased on data from Ohlmeier et al. 2019 [1] hBased on data from Almqvist et al. 2001 [6] iBased on data from Sackley et al. 2011 [7] jBased on data from Carrassi et al. 2017 [8]. kBased on data from Ramos-Arroyo et al.2005 [9]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Supplementary Table 2.** Overview of included studies from the literature search | | | | | |
| Publication | Country/region | Objectives | Population | Study design | Key outcomes |
| Kokmen et al.1994 [3] | USA | To calculate average annual incidences (1950–1989) and estimate prevalences (1960–1990) | 205 medical records were reviewed; 10 definite and 7 probable cases of HD were recorded.a Individuals whose symptoms began while they were residents of Olmsted County between January 1, 1950 and December 31, 1989 were included. | Retrospective review of healthcare records from all sources, including hospitals, outpatient clinics, and long-term care institutions. Incidences were estimated using the definite and probable cases as the numerator and the population estimates based on the decennial census, with linear interpolation for intercensal years as the denominator. | The average annual incidence of definite HD per 100,000 person-years for 1950 through 1989 was 0.3 (95% CI, 0.1–0.5) per 100,000; average prevalence was 6.3 (95% CI, 1.7–16) per 100,000 for the total population |
| Almqvist et al. 2001 [6] | Canada | To determine the uptake of the genetic mutation test for HD in the province of BC, Canada and assess the incidence and rate of identification of new mutations for HD | 205 individuals were referred for HD testing.b Of these, 141 individuals (68.8%) had a CAG tract length ≥36 repeats. | Retrospective review of data for symptomatic patients residing in BC referred to the University of BC Department of Medical Genetics for the diagnostic genetic test for HD between September 1, 1993 and June 30, 2000 | The mean minimum incidence in BC was 6.9 per million per year.c The mean annual incidence estimate for the Greater Vancouver Region was 6.85 per million per year and for the Greater Victoria Region was 10.42 per million per year. |
| Ramos-Arroyo et al. 2005 [9] | Spain | To evaluate the uptake of HD genetic analysis in Spain, and to provide additional information on the epidemiology of this disease from the experience of 9 years of direct genetic testing | 317 individuals with symptoms compatible with HD were tested. HD diagnosis (CAG tract length ≥36 repeats) was confirmed in 166 of these individuals. | Retrospective review of data for individuals referred to the Medical Genetics Laboratory of the Hospital Virgen del Camino, Pamplona, Spain, for diagnostic testing for HD between October 1993 and December 2002 | The incidence for the autonomous communities of Navarra and the Basque country, based on the number of newly diagnosed cases by genetic testing, was 4.7 per million per year |
| Sackley et al. 2011 [7] | UK | To provide epidemiological data and describe the pharmacological management of HD (2004–2008) | *N* = 324d | Retrospective review of primary care database, including data from >8 million patients from 400 general practices across the UK. Incidences between January 1 and December 31 of each year from 2004 to 2008 were identified and calculated as the number of new cases per 100,000 person-years. | The reported incidence of HD ranged from 0.44 to 0.78 per 100,000 person-years, and prevalence ranged from 5.96 to 6.54 per 100,000 for those with definite diagnoses |
| Wexler et al.2016 [4] | UK | To determine the prevalence of HD in the UK between 1990 and 2010 | The source population was all patients aged ≥21 years who were registered with general practices contributing to the CPRD between 1990 and 2010.e A total of 393 incident cases of HD were identified from the CPRD database between 1990 and 2010 from a total population of 9,282,126 persons (corresponding to 54,907,468 person-years). | Retrospective review of incident patients with a new diagnosis of HD who were identified from the primary care records of the CPRD. Incidence was calculated from the ratio of number of persons with a new recorded diagnosis of HD for each year from 1990 to 2010, divided by the total number of persons in the database for that year who had also had ≥1 year in the database and were ≥21 years old. | The average incidence of HD, during the entire period, was 7.2 per million patient-years (95% CI, 6.5–7.9) |
| Carrassi et al. 2017 [8] | Italy | To reappraise HD epidemiology in a northern Italian population, in relation to the introduction of genetic testing | 22 patients in the province of Ferrara who showed early signs and symptoms of HD (1990–2009) | HD cases were identified from administrative health data and medical records from the Units of Neurology and Genetics, Ferrara University Hospital, and from other provincial neurological structures | HD mean annual incidence in 1990–2009 was 0.3 per 100,000 (95% CI, 0.2–0.5). On December 31, 2014, HD prevalence was 4.2 per 100,000 (95% CI, 2.4–7.0). |
| Ohlmeier et al. 2019 [1] | Germany | To estimate the incidence and prevalence of HD and analyze the current routine care based on German claims data | 308 patients with HDf | Retrospective evaluation of data. Source of data was a sample of the Institute for Applied Health Research Berlin (InGef) Research Database, comprising data on approximately 4 million insured persons from approximately 70 German statutory health insurances. | The overall 2‑yearg incidence of HD was 1.8 per 100,000 (95% CI, 1.4–2.4). The overall 2-year prevalence of HD was 9.3 per 100,000 (95% CI, 8.3–10.4). |
| Castilhos et al. 2019 [5] | Brazil | To estimate the minimal prevalence of HD in Brazil (2013–2106) | In total, 179 Brazilian families were identified of whom 99 were from Rio Grande do Sul state. Prevalence was estimated for this state, as data obtained for this state were broader than those obtained in other parts of the country. Families from Rio Grande do Sul included 209 symptomatic individuals with HD, 690 individuals at 50% risk, and 515 individuals at 25% risk alive by December 2016. | Symptomatic individuals with HD with molecular diagnosis in one institution, as well as their care partners, were contacted from September 2013 to December 2016 | The minimal prevalence of symptomatic patients with HD was 1.85 per 100,000 |

BC, British Columbia; CI, confidence interval; CPRD, Clinical Practice Research Datalink; HD, Huntington’s disease; ICD-10-GM, International Statistical Classification of Diseases and Related Health Problems, 10th revision, German modification; NHS, National Health Service.

aFor a definite diagnosis of HD, documented evidence of progressive choreiform movement disorder, evidence of inheritance with autosomal dominant patterns, and progression of cognitive, behavioral, or emotional dysfunction was required. Individuals who lacked one of the three criteria were identified as having probable HD.

bThe requirement for performing the HD test was a referral of an individual with neurological signs and symptoms that could be compatible with HD and a pedigree, including information about a possible family history of HD. A CAG tract length ≥36 repeats was considered potentially associated with the clinical presentation of HD.

cThe incidence of HD was calculated for the province of BC based on the number of symptomatic individuals with CAG tract lengths in the affected range (≥36 repeats) who were diagnosed between 1996 and 1999 (*n* = 110).

dIndividuals were identified as diagnosed with HD if they had a Read Code indicating the condition. (Read Codes were a clinical terminology system that was in widespread use in general practice in the UK until around 2018, when NHS England switched to using SNOMED CT.)

eEligible cases were defined as persons with one or more recorded diagnoses of HD or Huntington’s chorea in their medical records. The Read Codes used to identify cases of HD were F134.00 (Huntington’s chorea) and Eu02200 (dementia in HD).

fIndividuals were identified as prevalent HD cases if at least two diagnoses of HD (ICD-10-GM: G10) were recorded during the study period.

gA 24-month observation period was chosen instead of a 12-month observation period to enable identification of a sufficiently large number of HD cases and thus ensure a comparatively robust estimation of the HD burden.

**Supplementary Table 3.** Probability of progression-free survival at 3 years per Shoulson–Fahn total functional capacity-based stage (**a**) and Cox proportional hazards model overall (**b**), illustrating that time spent in stage does not vary substantially across stages

**a** Probability of progression-free survival at 3 years, per Shoulson–Fahn stage

|  |  |
| --- | --- |
| Stage | Probability (95% confidence interval) |
| 1 | 0.56 (0.44–0.72) |
| 2 | 0.68 (0.61–0.75) |
| 3 | 0.66 (0.59–0.74) |
| 4 | 0.45 (0.34–0.58) |
| 5 | 0.51 (0.35–0.74) |

**b** Cox model overall: there was no significant difference in progression-free survival between stages (reference group = stage 1) except for marginal significance for stage 4. There may be some differences, but these are not ordinal and are mostly not statistically significant, so a common survival time in all stages was assumed.

|  |  |  |
| --- | --- | --- |
| Stage | Hazard ratio  (95% confidence interval) | *p* value |
| 1 (reference) | 1 |  |
| 2 | 0.72 (0.49–1.06) | 0.10 |
| 3 | 0.84 (0.57–1.23) | 0.37 |
| 4 | 1.52 (1.02–2.27) | 0.04 |
| 5 | 1.38 (0.84–2.26) | 0.21 |

**Supplementary Table 4.** Comparison of parametric functions on the progression endpoint (**a**) and overall survival endpoint (**b**) (see next page)

**a** Progression endpoint. For progression, a common survival curve was assumed. The table shows the model fit statistics.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Function | Log likelihood | Number of parameters, K = | Number of contributing observations, *n* | AIC = –2log-likelihood + 2K | BIC = –2log-likelihood + 2ln(n) |
| Exponential | –2,954.544 | 1 | 1,678 | 5,911.088 | 5,923.939 |
| Weibull | –2,899.948 | 2 | 1,678 | 5,803.897 | 5,814.747 |
| Gamma | –2,900.006 | 2 | 1,678 | 5,804.012 | 5,814.863 |
| Log-normal | –2,916.420 | 2 | 1,678 | 5,836.839 | 5,847.691 |
| Log-logistic | –2,899.120 | 2 | 1,678 | 5,802.240 | 5,813.091 |
| Gompertz | –2,915.034 | 2 | 1,678 | 5,834.068 | 5,844.919 |

AIC, Akaike information criterion; BIC, Bayesian information criterion.

The Weibull model was used.

The fitted model was:

where S(*t*) is the probability of survival at time *t* (in days).

**b** Overall survival endpoint. For overall survival, it is known that survival varies by Shoulson–Fahn total functional capacity-based stage (giving four additional parameters: 5 stages – 1). The table shows the model fit statistics.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Function | Log likelihood | Number of parameters, K = | Number of contributing observations, *n* | AIC = –2log-likelihood + 2K | BIC = –2log-likelihood + 2ln(n) |
| Exponential | –1,132.553 | 5 | 1,678 | 2,275.106 | 2,279.957 |
| Weibull | –1,128.702 | 6 | 1,678 | 2,269.404 | 2,272.255 |
| Gamma | –1,128.933 | 6 | 1,678 | 2,269.866 | 2,272.717 |
| Log-normal | –1,134.623 | 6 | 1,678 | 2,281.246 | 2,284.097 |
| Log-logistic | –1,129.376 | 6 | 1,678 | 2,270.752 | 2,273.603 |
| Gompertz | –1,126.031 | 6 | 1,678 | 2,264.062 | 2,266.913 |

AIC, Akaike information criterion; BIC, Bayesian information criterion.

The Weibull model was used.

The fitted model was:

where S(*t*) is the probability of survival at time *t* (in days), and C represents the following values for the different stages:

|  |  |  |  |
| --- | --- | --- | --- |
| Shoulson–Fahn stage | C | Estimated median overall survival, years | Estimated 1-year probability of overall survival |
| 1 | 1 | NAa | 1.000a |
| 2 | –13.7 | 24.6 | 0.988 |
| 3 | –14.5 | 11.0 | 0.967 |
| 4 | –15.2 | 5.5 | 0.922 |
| 5 | –15.7 | 3.3 | 0.859 |

NA, not applicable.

aOwing to the limited number of patients who died in stage 1 in the database, the estimated probability of death in stage 1 is <1 but rounded to 1 at 3 decimal places. For the same reason, the model estimate of median survival is unrealistically high.

**Supplementary Table 5.** Current Shoulson–Fahn total functional capacity-based stage to next observed stage (Enroll-HD data suggest that it is unusual for individuals with Huntington’s disease to skip a Shoulson–Fahn stage when the disease progresses). In our model, owing to the assumption of a common PFS curve and the model extrapolation, at certain time points (which vary by stage), estimated cumulative progression (OS–PFS) begins to decrease. This is due to the higher rate of decline in OS compared with that in PFS. Therefore, from that time point, progression was capped in the model and remaining patients were assumed to leave the stage through death

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Shoulson–Fahn stage | | | | |
|  | 1 | 2 | 3 | 4 | 5 |
| *n* | 36 | 93 | 100 | 72 | 28 |
| Progressed stage, *n* (%) |  |  |  |  |  |
| 2 | 35 (97.2) | 0 | 0 | 0 | 0 |
| 3 | 1 (2.8) | 76 (81.7) | 0 | 0 | 0 |
| 4 | 0 | 2 (2.2) | 59 (59.0) | 0 | 0 |
| 5 | 0 | 1 (1.1) | 2 (2.0) | 35 (48.6) | 0 |
| Died, *n* (%) | 0 | 14 (15.1) | 39 (39.0) | 37 (51.4) | 28 (100) |

OS, overall survival; PFS, progression-free survival.

**Supplementary Table 6.** Search string used to determine the incidence and prevalence estimates reported in the literature through PubMed

|  |  |
| --- | --- |
| Search string: | (((((huntington's disease[MeSH Terms]) AND (Epidemiology OR Incidence OR Prevalence OR Diagnosed prevalence OR Diagnosis OR Incidence rate OR Prevalence rate OR Incidence trends OR Prevalence trends)) AND (Manifest HD OR Manifest huntington's disease OR Motor onset OR Motor manifestation OR Choreic movements OR Motor diagnosis OR Parkinsonian Huntington's Disease OR Parkinsonism in Huntington's Disease OR Extrapyramidal symptoms)) AND (Longitudinal study OR Cross-sectional study OR Population based study OR Real world data OR Retrospective study OR Observational study OR Long term follow-up study OR Follow-up study OR Registry OR Patient registry OR Patient tracking OR Cohort study OR Prospective study OR Long-term outcomes)) AND ("2016/01/01"[PDat] : "2019/05/01"[PDat] ) |

**Supplementary figures legends**

**Supplementary Fig. 1.** Structure and inputs to the model to estimate prevalence and survival of patients with HD. HD, Huntington’s disease.

**Supplementary Fig. 2.** Time to progression by Shoulson–Fahn total functional capacity-based stage, illustrating that time spent in stage does not vary substantially across stages.

# **References**

1. Ohlmeier C, Saum K-U, Galetzka W, Beier D, Gothe H. Epidemiology and health care utilization of patients suffering from Huntington’s disease in Germany: real world evidence based on German claims data. BMC Neurology. 2019 2019/12/10;19(1):318.

2. Jackson C. flexsurv: a platform for parametric survival modeling in R. J Stat Softw. 2016 2016-05-12;70(8):33.

3. Kokmen E, Ozekmekci FS, Beard CM, O'Brien PC, Kurland LT. Incidence and prevalence of Huntington's disease in Olmsted County, Minnesota (1950 through 1989). Arch Neurol. 1994 Jul;51(7):696–8.

4. Wexler NS, Collett L, Wexler AR, Rawlins MD, Tabrizi SJ, Douglas I, et al. Incidence of adult Huntington's disease in the UK: a UK-based primary care study and a systematic review. BMJ Open. 2016 Feb 23;6(2):e009070.

5. Castilhos RM, Santos JAD, Augustin MC, Pedroso JL, Barsottini O, Saba R, et al. Minimal prevalence of Huntington's disease in the South of Brazil and instability of the expanded CAG tract during intergenerational transmissions. Genet Mol Biol. 2019 Apr-Jun;42(2):329–36.

6. Almqvist E, Elterman D, MacLeod P, Hayden M. High incidence rate and absent family histories in one quarter of patients newly diagnosed with Huntington disease in British Columbia. Clin Genet. 2001;60(3):198–205.

7. Sackley C, Hoppitt TJ, Calvert M, Gill P, Eaton B, Yao G, et al. Huntington's disease: current epidemiology and pharmacological management in UK primary care. Neuroepidemiology. 2011;37(3–4):216–21.

8. 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(1–2):18–23.

9. 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 Mar;76(3):337–42.