This is a repository copy of *The global prevalence of Huntington's disease: a systematic review and discussion*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/103389/

Version: Accepted Version

Article:
Baig, S.S., Strong, M. and Quarrell, O.W.J. (2016) The global prevalence of Huntington's disease: a systematic review and discussion. Neurodegenerative Disease Management, 6 (4). ISSN 1758-2024

https://doi.org/10.2217/nmt-2016-0008

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

The global prevalence of Huntington's disease: a systematic review and discussion

Sheharyar Sajjad Baig (1) , Mark Strong (2) & Oliver WJ Quarrell* (3)

1. Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

2. School of Health & Related Research, University of Sheffield, Sheffield, UK

3. Department of Clinical Genetics, Sheffield Children's Hospital, Northern General Hospital, Herries Road, Sheffield, S5 7AU, UK

*Author for correspondence: oliver.quarrell@sch.nhs.uk

Accepted for publication in Neurodegenerative Disease Management: 2 June 2016

Published online: 20 July 2016

DOI: 10.2217/nmt-2016-0008

Weblink: http://www.futuremedicine.com/doi/10.2217/nmt-2016-0008

PRACTICE POINTS

- Calculating something as simple as the prevalence of Huntington's disease (HD) is problematic and contentious.
- Multiple sources are available to ascertain HD cases in a given population.
- Populations that employed diagnostic testing of HD have increased their ascertained prevalence measures over the last two decades.
- The estimated prevalence of HD in North America, North Western Europe and Australia ranges from 5.96 to 13.7 cases per 100 000 population.
- The ascertained prevalence of HD in Asia is much lower than Western populations.
- Using multiple sources for ascertainment of HD cases, although time-consuming, is more likely to determine the true prevalence of the disease in a given population.

ABSTRACT

The ascertained prevalence of Huntington's disease (HD) increased significantly following the provision of diagnostic testing. A systematic review was conducted to estimate the prevalence of HD in the post-diagnostic testing era. 22 studies with original data pertaining to the prevalence of HD (1993-2015) were included and analysed. A global meta-analysis was not performed due to heterogeneity in study methods and geographical variation. The prevalence of HD is significantly lower in Asian populations compared to Western Europe, North America and Australia. The global variation in HD prevalence is partly explained by the average CAG repeat lengths and frequency of different *HTT* gene haplotypes in the general population. Understanding the prevalence of HD has significant implications for healthcare resource planning.

Key Words

Huntington's disease – Epidemiology – Diagnosis

INTRODUCTION

Huntington's disease (HD) is a slowly progressive autosomal dominant neurodegenerative disorder characterised by motor abnormalities, cognitive impairment and psychiatric disturbances [1]. The disease is caused by an expanded CAG triplet repeat in the *HTT* gene which encodes an abnormal polyglutamine expansion in the huntingtin protein [2].

HD was classically a clinical diagnosis made in the context of a positive family history of the condition. After the identification of the underlying genetic mutation in 1993 [2], diagnostic testing became widely available. This enabled clinicians to make a confident diagnosis of individuals with typical neurological features but without a known family history of the condition; this group may represent up to 10% of new HD cases [3]. As a consequence, the ascertainment of HD in populations has increased and the measured prevalence of HD in several populations is substantially higher in the post-diagnostic testing era [4–6]. Studies performed prior to 1993 may therefore underestimate the true prevalence of HD.

The management of HD requires the co-ordination of professionals from multiple domains including

neurologists, psychiatrists, psychologists, specialist nurses, physiotherapists, occupational therapists, social services and carer services. In order to allocate the optimal and appropriate amount of scarce resources, an accurate calculation of the scope of disease burden on the population is imperative. If previous estimates of prevalence underestimate the true prevalence, the current provision of health and social care services allocated to individuals with HD may be underequipped.

The second issue that arises from uncertain prevalence measures is that healthcare services are unable to identify the number of individuals at-risk of developing HD. The ratio of symptomatic individuals (prevalence) to individuals at 50% risk of developing HD has been described, on theoretical grounds, as being 1:5 [7] and approximately 1:4.2 in empirical studies [8,9]. At present, identifying these individuals is important to be able to offer predictive testing, genetic counselling, emotional support and recruitment for clinical research. In the future, characterising and quantifying this population is significant as future disease-modifying therapies may be targeted at gene positive individuals in the pre-symptomatic period of HD.

AIMS

The present study will attempt to:

1. Identify the published measurements of HD prevalence made in the era of diagnostic testing.
2. Reconcile the geographical variation in HD prevalence explaining the factors that determine variation in the true and ascertained (measured) prevalence of HD.

METHODS

Search Strategy

A systematic literature search was conducted using a predetermined protocol. Two computer-stored databases, MEDLINE (1993-2015) and EMBASE (Excerpta Medical Database; 1993-2015), were searched for studies investigating the prevalence of Huntington's disease in a defined population. The search strategy was developed after consultation with a research librarian and is detailed in Appendix 1.

Further studies were identified from the following sources

1. Searching within references of relevant articles.

2. Searching for articles that cited the studies identified using the search strategy.
3. Information from articles on the uptake of predictive testing
4. Specialist textbooks on Huntington's disease
5. Web searches.
6. Online databases [10,11]

Selection of Studies

All studies identified by the search strategy were screened by one reviewer (S.S.B.) who excluded those that were irrelevant. The abstracts of the remaining studies were screened by one reviewer (S.S.B.) who excluded studies which were not observational or did not investigate the epidemiology of Huntington's disease. Full texts of all the remaining studies and assessed by two independent reviewers (S.S.B. and O.W.J.Q).

Inclusion/Exclusion Criteria

Articles were included based on the criteria established in Table 1. Studies performed prior to 1993 were excluded for two major reasons. Firstly, as diagnostic genetic testing became available in 1993, studies before this relied solely on a clinical diagnosis of HD and, as such, had the possibility of incomplete ascertainment of HD cases. Secondly, as discussed in greater detail in the discussion, there is a suggestion that the true prevalence of HD may be increasing as the life expectancy in the general population rises [4], the most current studies were felt to be of most relevance. In several cases, HD prevalence measures on populations made before 1993 had been repeated and updated; it is these recent studies with higher ascertainment that were included in the present analysis.

The measurement of the prevalence of HD in a population is typically performed through a cross-sectional, observational study. In some cases, where a registry for HD was established, the prevalence is established by means of a cohort study. Our qualification of observational studies is important as there are several studies published in the literature which estimate the prevalence of HD in different populations by using computational models based on the mean CAG repeat length in the general population and the common *HTT*

gene haplotypes rather than on observed data on the number of individuals with a diagnosis of HD.

Data Extraction

For each study, data extracted included the region studied, population size, prevalence date, sources of case ascertainment, diagnostic criteria, number of cases of Huntington's disease, prevalence per 100 000 population and methodological limitations of the study. 95% confidence intervals were calculated for each prevalence estimate using the Agresti Coull method [12].

Data Analysis

Due to the heterogeneity between studies with respect to their methods of identifying, diagnosing and recording Huntington's disease cases, it was felt to be inappropriate to combine all the studies and perform a meta-analysis to provide pooling statistics. Where pooled estimates were reported, a DerSimonian and Laird random effects model for the logit transformed prevalences was assumed. [13]. All calculation was performed using the meta package in R 3.2.3.

RESULTS

Figure 1 shows the selection of cases for the systematic review. 3397 studies were identified through MEDLINE, EMBASE, web searches, citation searches, searches within references from previous review articles and selected studies, textbooks and from prior knowledge. Titles were screened for 2030 non-duplicate studies and 217 abstracts were screened. 175 abstracts were excluded as they were either duplicates or did not meet the inclusion criteria. 41 full text articles and 8 conference abstracts were assessed in detail for eligibility with 19 excluded. In addition, twelve review articles on HD epidemiology were identified but searches through the references did not yield any additional studies [14–25].

Of the excluded studies: five estimated the HD prevalence before 1993 [26–30], four studied specific subgroups that were not representative of the whole population [31–34] , four were not population-based observational studies and estimated the prevalence indirectly [35–38], two studied small geographical clusters of high prevalence [39,40], two had insufficient information regarding case ascertainment [41,42] ,one did not differentiate between symptomatic individuals and asymptomatic mutation positive individuals [43] and one

was not written in English [44].

Twenty-two studies examining HD prevalence (eighteen original articles and four conference abstracts) were included in the qualitative analysis. Fifteen studies were conducted in European populations, one in North America, two in Australia and four in Asia. The hypothetical global mean prevalence based on pooling all the data from the studies included in the present systematic review in a meta-analysis would be 5.5 per 100,000; However, the interpretation and application this figure as an average global prevalence of HD would be inappropriate due to the heterogeneity between the included studies.

Table 2 details the results of the systematic review. It contains the ascertained prevalence of HD in different populations from four continents. Figure 2 shows a funnel plot of prevalence (per 100,000 population) against population size. The hypothetical global mean prevalence is shown as a dashed vertical line, as are 95% control limits. Significant overdispersion is evident, suggesting that variation in prevalence estimates is due to causes other than simple sampling variability. There is no evidence of a relationship between prevalence and population size, though regional differences are clearly seen.

Figure 3 shows Forest Plots representing studies of HD prevalence from four continents. Figure 4 illustrates the ascertained prevalence of HD in different studies geographically.

DISCUSSION

The present study is the most comprehensive systematic review of Huntington's disease (HD) epidemiology conducted in the post-diagnostic testing era. It identifies prevalence estimates from populations in four continents and indicates marked variation in the prevalence of HD. It indicates that the ascertained prevalence of HD has increased significantly following the advent of diagnostic testing and details the higher prevalence of HD in European, North American and Australian populations relative to Asian populations.

The recorded prevalence of HD in several individual populations has increased after the introduction of genetic testing [4–6,9,45,46]. The study performed in Finland showed a four-fold increase in the prevalence of HD following the introduction of genetic testing [6]. This may partly be explained by the ability to diagnose

individuals with a negative family history (new mutations, historical misdiagnosis in family members, non-penetrance, non-paternity) through genetic testing [1]. Additionally, as the life expectancy in the general population increases, individuals may present with HD in later life; this may be particularly relevant individuals with reduced penetrance alleles who develop symptoms in later life [4,47]. Other factors that may contribute towards the increase in recorded prevalence of HD over time include the use of diagnostic testing earlier in the course of the illness e.g. with early cognitive or behavioural symptoms with subtle motor symptoms in the context of a positive family history. In populations where the prevalence of HD has previously been low, increased clinician familiarity with the disease entity may contribute to the increase in recorded prevalence.

In the UK, two recent studies used primary care research databases to determine the current prevalence of HD which resulted in two strikingly different estimates of 5.96 [48] and 12.3 [5] per 100,000 of the population. The larger estimate, however, describes the prevalence in the over 20 population where HD is far more common. When the findings of Evans *et al* were combined with an additional publication by their group describing the prevalence of HD in the under-21 population [49] , the HD prevalence in the UK in 2010 was estimated to be 9.28 per 100 000 population. The residual difference between the two primary care research databases remains unaccounted for.

There is significant global variation in the prevalence of HD. A substantial proportion of the measured differences in HD prevalence is secondary to variation in the true prevalence of HD i.e. geographical differences that would persist even if there was complete ascertainment of every case of HD. Nevertheless, this variation may, in part, be explained by factors that affect the complete ascertainment of individuals with HD. The possible reasons for differences in true and ascertained prevalence of HD are summarised in Table 3.

A major biological determinant of differences in the true prevalence of HD between populations is the mean CAG repeat length in the general population. Populations with a higher prevalence of HD e.g. European populations have been shown to have a higher mean CAG repeat length in the *HTT* gene in the non-affected l population when compared to populations with a lower prevalence of HD e.g. Japan and China [50,51]. There is thought to be a causal relationship between the two factors as populations with a greater proportion of individuals with CAG repeat lengths in the high-normal range serve as a pool of potential new mutations with

expansion of the CAG repeat length in subsequent generations, first into the intermediate allele range (27-35 repeats) and then into the affected range (≥ 36 repeats) [47]. Another significant biological determinant of variation in the true prevalence of HD is the haplotype of the *HTT* gene. Warby *et al* (2009) determined that, in a European population, CAG expansion in the *HTT* gene occurs with significantly increased frequency on two haplotypes, A1 and A2, compared to haplogroups B and C [52]. In East Asian individuals, however CAG expansions are associated most with haplotype C [53]. Warby *et al* (2011) further demonstrated that these high risk haplotypes, A1 and A2, are present in 20% of the individuals from the general European population (with < 27 CAG repeats) but were absent in a sample of the general population of East Asia [53]. The proposed explanation of these findings is that the mutation rate of the CAG expansion in the *HTT* gene is more likely to occur on haplotypes A1 and A2 because other *cis* elements make these CAG repeat length on these chromosomes more unstable. As these haplotypes are more common in European populations compared to East Asian populations, this may explain the markedly higher prevalence of HD in the former. Thirdly, in geographically isolated populations such as Iceland and Malta, the founder effect may explain some of the variation seen. [54,55].

As mentioned, variation in HD prevalence may be explained by factors that affect the ascertainment of individual cases of HD when healthcare researchers attempt to determine prevalence measures. There are several data sources utilised by healthcare workers in order to identify individuals with HD; each of these these has its own advantages, disadvantages, sensitivity, specificity and error rate. For instance, a study which takes data from a centralised testing centre which runs a regional HD service led by a small number of clinicians who are intimately involved in the local HD community and who actively characterise HD pedigrees in order to determine accurately the prevalence [4,56] is more likely to have a higher prevalence figure than a data source which relies on coding such as hospital discharge summaries.

Errors in the measured prevalence of HD prevalence can arise through multiple routes. For instance, if individual cases are not cross-referenced with death notifications, deceased individuals may incorrectly be included in point prevalence measures; in essence, the reported prevalence may in fact be the cumulative incidence over the study period. In addition, the onset of HD is insidious; therefore, ideally, a prevalence date needs to be a little earlier than the study date to allow for the fact that some individuals in the study

population will be symptomatic but undiagnosed at the time of the study but were affected at the time of the earlier prevalence date. Individuals who have been identified as having an abnormal CAG expansion through a predictive testing but who are currently presymptomatic should not be included within prevalence measures of HD. However, in studies where data on individuals with HD is extracted from the relevant administrative code on a large databases e.g. primary care records and national insurance databases, there is a possibility that some presymptomatic individuals may have been incorrectly coded as having a diagnosis of HD. This can be overcome by healthcare researchers accessing the clinical records of all cases of HD identified in large datasets to confirm the diagnosis, however, this requires additional ethical approval and a greater number of resources. There are a number of conditions which may be incorrectly diagnosed as HD but are not caused by an abnormal CAG expansion in the *HTT* gene. These conditions, termed 'HD phenocopy syndromes' can clearly be ruled out by the use of diagnostic genetic testing, however, in individuals with a purely clinical diagnosis of HD, upto 1% of cases actually represent HD phenocopy syndromes [57]. Further, poor response rates and incomplete information from clinician surveys, family surveys and family pedigrees can lead to an underascertainment of cases.

The use of multiple sources to identify individuals with HD has been instrumental in improving the ascertainment of HD prevalence. In British Columbia, the use of several sources for identifying individuals with HD yielded the highest prevalence estimate of HD in a Western population [4]. The issues that arise with multiple source ascertainment include its time-consuming and costly nature, the possibility of including the same individual twice or more in prevalence measures ('double-counting') and the practical difficulties in carrying this out in a large population.

A key limitation of the current study is the absence of studies that were not conducted in the English language. The authors are aware of one such study in the San-in area of Japan [44]; however, the estimated prevalence in the abstract of this study does not appear dissimilar to quoted figures from Japan in 1996 [58] and 2015 [59].

Conclusions

The present study demonstrates an increase in the ascertained prevalence of Huntington's disease (HD) in several populations and indicates marked global geographical variation in the prevalence of the disease which is likely explained by the mean CAG repeat length in the unaffected population, *HTT haplotypes* and the variable use of multiple sources of ascertainment to determine the prevalence of HD. Optimising the ascertainment of HD cases in a given population requires the recording of cases from multiple sources with safeguards to prevent double-counting of individuals in the reported estimates.

FUTURE PERSPECTIVE

Five studies on HD prevalence were published in 2015 suggesting there is continued interest in the epidemiology of HD [6,39,46,59,60]. Accurately characterising the prevalence of the condition is necessary to allocate the optimal amount of resources for health and social care resource provision, research funding and psychological counselling.

The aim of the future treatment for HD is to alter the natural history of the disease. Ideally, treatment should start in the pre-symptomatic phase. The ratio of 50% at-risk individuals to symptomatic individuals is either 4.2:1 or 5:1 [7,8]. There are currently several active clinical trials for drug therapy in HD; if even a single study shows a neuroprotective effect, it is likely that the demand for predictive testing services will markedly increase. Therefore, accurately determining the prevalence of HD, and thereby the at-risk population size, may become increasingly important in the future.

REFERENCES

1. Bates GP, Dorsey R, Gusella JF, *et al.* Huntington disease. *Nat. Rev. Dis. Prim.* [Internet]. (April), 15005 (2015). Available from: http://www.nature.com/articles/nrdp20155.
   **This is an important up to date and authoratitive review of Huntington's disease.
2. Macdonald M. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* [Internet]. 72(6), 971–983 (1993). Available from: http://www.sciencedirect.com/science/article/pii/009286749390585E.
   **This paper describes the identification as an unstable expansion of the CAG repeat length as the cause of Huntington's disease.
3. McCusker EA, Casse RF, Graham SJ, Williams DB, Lazarus R. Prevalence of Huntington disease in New South Wales in 1996. *Med. J. Aust.* [Internet]. 173(4), 187–90 (2000). Available from: http://www.ncbi.nlm.nih.gov/pubmed/11008591.
4. Fisher ER, Hayden MR. Multisource ascertainment of Huntington disease in Canada: Prevalence and population at risk. *Mov. Disord.* [Internet]. 29(1), 105–114 (2014). Available from:

http://doi.wiley.com/10.1002/mds.25717.
**This paper is the most recent estimate of the prevalence of Huntington's disease using multiple methods of ascertainment.

5. Evans SJ, 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* [Internet]. 84(10), 1156–1160 (2013). Available from: http://jnnp.bmj.com/cgi/doi/10.1136/jnnp-2012-304636.
*This paper describes the latest estimate of the prevalence of Huntington's disease in the UK . It gives the prevalence for the population ≥21 years and needs to be read in conjunction with the paper from Douglas *et al* 2012 (Ref 49) if a comparison is to be made with other prevalence studies.

6. Sipilä JOT, Hietala M, Siitonen A, Päivärinta M, Majamaa K. Epidemiology of Huntington's disease in Finland. *Parkinsonism Relat. Disord.* [Internet]. 21(1), 46–49 (2015). Available from: http://linkinghub.elsevier.com/retrieve/pii/S1353802014004106.

7. Conneally PM. Huntington disease: genetics and epidemiology. *Am. J. Hum. Genet.* 36(3), 506–526 (1984).

8. Tassicker RJ, Teltscher B, Trembath MK, *et al.* Problems assessing uptake of Huntington disease predictive testing and a proposed solution. *Eur. J. Hum. Genet.* [Internet]. 17(1), 66–70 (2009). Available from: http://www.nature.com/doifinder/10.1038/ejhg.2008.142.

9. Morrison P, Harding-Lester S, Bradley A. Uptake of Huntington disease predictive testing in a complete population. *Clin. Genet.* [Internet]. 80(3), 281–286 (2011). Available from: http://doi.wiley.com/10.1111/j.1399-0004.2010.01538.x.

10. Frequency of Inherited Disorders Database [Internet]. Inst. Med. Genet. Cardiff Univ. Available from: http://medic.cardiff.ac.uk/fidd/search.aspx.

11. Published reports on worldwide prevalence of Huntington's disease [Internet]. Cent. Mol. Med. Ther. Available from: http://www.cmmt.ubc.ca/research/diseases/huntingtons/HD_Prevalence.

12. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *Am. Stat.* [Internet]. 52(2), 119–126 (1998). Available from: http://www.jstor.org/stable/2685469?origin=crossref&seq=1#page_scan_tab_contents.

13. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control. Clin. Trials*. 7(3), 177–188 (1986).

14. Xu M, Wu Z-Y. Huntington Disease in Asia. *Chin. Med. J. (Engl).* [Internet]. 128(13), 1815–1819 (2015). Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84933556800&partnerID=tZOtx3y1.

15. Vieira RT, Caixeta L, Machado S, *et al.* Epidemiology of early-onset dementia: a review of the literature. *Clin. Pract. Epidemiol. Ment. Health* [Internet]. 9, 88–95 (2013). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3715758&tool=pmcentrez&rendertype=abstract.

16. 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.* [Internet]. 27(9), 1083–1091 (2012). Available from: http://doi.wiley.com/10.1002/mds.25075.

17. Hoppitt T, Pall H, Calvert M, *et al.* A systematic review of the incidence and prevalence of long-term neurological conditions in the UK. *Neuroepidemiology* [Internet]. 36(1), 19–28 (2011). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2997442&tool=pmcentrez&rendertype=abstract.

18. Quarrell O, O'Donovan KL, Bandmann O, Strong M. The Prevalence of Juvenile Huntington's Disease: A Review of the Literature and Meta-Analysis. *PLoS Curr.* [Internet]. 4, e4f8606b742ef3 (2012). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3426104&tool=pmcentrez&rendertype=a

bstract.

19. Jader L. An overview of neurological disorders in Wales. *Neuroepidemiology* [Internet]. 28(2), 65–78 (2007). Available from: http://www.ncbi.nlm.nih.gov/pubmed/17230026.

20. Bargiela D, Yu-wai-man P, Keogh M, Horvath R. Prevalence of neurogenetic disorders in the North of England. 0, 1195–1201 (2015).

21. Pringsheim T, Fiest K, Jette N. The international incidence and prevalence of neurologic conditions: How common are they? . *Neurol.* [Internet]. 83 (18), 1661–1664 (2014). Available from: http://www.neurology.org/content/83/18/1661.short.

22. Al-Jader LN, Harper PS, Krawczak M, Palmer SR. The frequency of inherited disorders database: prevalence of Huntington disease. *Community Genet.* [Internet]. 4(3), 148–57 (2001). Available from: http://www.ncbi.nlm.nih.gov/pubmed/14960907.

23. Borlongan C V, Burns J, Tajiri N, *et al.* Epidemiological Survey-Based Formulae to Approximate Incidence and Prevalence of Neurological Disorders in the United States: a Meta-Analysis. *PLoS One* [Internet]. 8(10), e78490 (2013). Available from: http://dx.doi.org/10.1371%2Fjournal.pone.0078490.

24. de Pedro-Cuesta J, Rábano A, Martínez-Martín P, *et al.* Comparative Incidence of Conformational, Neurodegenerative Disorders. *PLoS One* [Internet]. 10(9), e0137342 (2015). Available from: http://dx.doi.org/10.1371%2Fjournal.pone.0137342.

25. Lekoubou A, Echouffo-Tcheugui JB, Kengne AP. Epidemiology of neurodegenerative diseases in sub-Saharan Africa: a systematic review. *BMC Public Health* [Internet]. 14(1), 653 (2014). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4094534&tool=pmcentrez&rendertype=abstract.

26. Morrison PJ, Nevin NC. Huntington disease in County Donegal: epidemiological trends over four decades. *Ulster Med. J.* [Internet]. 62(2), 141–4 (1993). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2449037&tool=pmcentrez&rendertype=abstract.

27. Morrison PJ, Johnston WP, Nevin NC. The epidemiology of Huntington's disease in Northern Ireland. *J. Med. Genet.* [Internet]. 32(7), 524–30 (1995). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1050544&tool=pmcentrez&rendertype=abstract.

28. Kokmen E, Özekmekçi F, Beard C, PC O, LT K. INcidence and prevalence of huntington's disease in olmsted county, minnesota (1950 through 1989). *Arch. Neurol.* [Internet]. 51(7), 696–698 (1994). Available from: http://dx.doi.org/10.1001/archneur.1994.00540190076018.

29. El Tallawy HN a, Farghaly WM a, Rageh T a., *et al.* Door-to-door survey of major neurological disorders (project) in Al Quseir City, Red Sea Governorate, Egypt. *Neuropsychiatr. Dis. Treat.* 9, 767–771 (2013).

30. Shiwach RS. Prevalence of Huntington's disease in the Oxford region. *Br. J. Psychiatry* [Internet]. 165(3), 414–415 (1994). Available from: http://bjp.rcpsych.org/content/165/3/414.2.abstract.

31. Kovalchuk AU, Bykanova MA, Pizova NV, Kuznetcova IP. An epidemiology of movement disorders in Yaroslavl district by materials of our register. *Eur. J. Neurol.* 17(Suppl 3), 389 (2010).

32. Danila O, Hirdes JP, Maxwell CJ, *et al.* Prevalence of neurological conditions across the continuum of care based on interRAI assessments. *BMC Health Serv. Res.* [Internet]. 14(1), 29 (2014). Available from: http://www.biomedcentral.com/1472-6963/14/29.

33. Withall A, Draper B, Seeher K, Brodaty H. The prevalence and causes of younger onset dementia in Eastern Sydney, Australia. *Int. Psychogeriatrics*. 26(Special Issue 12), 1955–1965 (2014).

34. Alonso ME, Ochoa A, Boll M-C, *et al.* Clinical and genetic characteristics of Mexican Huntington's disease patients. *Mov. Disord.* [Internet]. 24(13), 2012–5 (2009). Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-70450159037&partnerID=tZOtx3y1.

35. Paradisi I, Hernández A, Arias S. Huntington disease mutation in Venezuela: age of onset, haplotype

analyses and geographic aggregation. *J. Hum. Genet.* [Internet]. 53(2), 127–135 (2008). Available from: http://www.nature.com/doifinder/10.1007/s10038-007-0227-1.

36. Gatto E, Parisi VL, Sanguinetti A, Persi G, Etcheverry JL. Estimate Huntington Disease Prevalence in Latin America (P4.062). *Neurol.* [Internet]. 82 (10 Supplement) (2014). Available from: http://www.neurology.org/content/82/10_Supplement/P4.062.abstract.

37. Gatto E, Parisi V, Persi G, *et al.* Clinical and genetic characteristics in patients with Huntington's Disease from Argentina. *Parkinsonism Relat. Disord.* [Internet]. 18(2), 166–9 (2012). Available from: http://www.sciencedirect.com/science/article/pii/S1353802011003087.

38. Hećimović S, Klepac N, Vlasić J, *et al.* Genetic background of Huntington disease in Croatia: Molecular analysis of CAG, CCG, and Delta2642 (E2642del) polymorphisms. *Hum. Mutat.* [Internet]. 20(3), 233 (2002). Available from: http://www.ncbi.nlm.nih.gov/pubmed/12204002.

39. Agostinho LA, da Silva IS, Maia LA, *et al.* A Study of a Geographical Cluster of Huntington's Disease in a Brazilian Town of Zona da Mata, Minas Gerais State. *Eur. Neurol.* [Internet]. 74(1-2), 62–68 (2015). Available from: http://www.karger.com/DOI/10.1159/000434630.

40. Alencar MA, Lopez AM, Figueiredo E, Porciúncula CG, Monlleó I. E05 Prevalence of Huntington's disease in Feira Grande, a small city in Northeastern Brazil. *J. Neurol. Neurosurg. Psychiatry* [Internet]. 81 (Suppl 1), A22–A22 (2010). Available from: http://jnnp.bmj.com/content/81/Suppl_1/A22.3.abstract.

41. Bernhardt C, Schwan A-M, Kraus P, Epplen JT, Kunstmann E. Decreasing uptake of predictive testing for Huntington's disease in a German centre: 12 years' experience (1993–2004). *Eur. J. Hum. Genet.* [Internet]. 17(3), 295–300 (2009). Available from: http://www.nature.com/doifinder/10.1038/ejhg.2008.164.

42. Kirilenko NB, Fedotov VP, Baryshnikova N V, Dadali EL, Poliakov a V, Zinchenko R a. [Genetic and epidemiologic analysis of hereditary diseases of the nervous system in the cities of Volgograd and Volzhskii]. *Genetika* [Internet]. 40(9), 1256–1261 (2004). Available from: http://www.ncbi.nlm.nih.gov/pubmed/15559154.

43. Dastgiri S, Bonyadi M, Mizani T. Epidemiology of neuro-genetic disorders in Northwestern Iran. *Neurosciences*. 17(2), 171–172 (2012).

44. Adachi Y, Nakashima K. Population genetic study of Huntington's disease--prevalence and founder's effect in the San-in area, western Japan. [Japanese]. *Nippon Rinsho - Japanese J. Clin. Med.* [Internet]. Japanese j(4), 900–904 (1999). Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed4&AN=10222787\nhttp://openurl.ac.uk/athens:_edu//lfp/LinkFinderPlus/Display?sid=OVID:Embase&id=pmid:10222787&id=&issn=0047-1852&isbn=&volume=57&issue=4&spage=900&pages=900-904&date=1.

45. 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* [Internet]. 76 (3), 337–342 (2005). Available from: http://jnnp.bmj.com/content/76/3/337.abstract.

46. 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.* , 1–4 (2015).

47. Kay C, Fisher ER, Hayden MR. Huntington's Disease (Oxford Monographs on Medical Genetics). 4th ed. Oxford University Press.

48. Sackley C, Hoppitt TJ, Calvert M, *et al.* Huntington's Disease: Current Epidemiology and Pharmacological Management in UK Primary Care. *Neuroepidemiology* [Internet]. 37(3-4), 216–221 (2011). Available from: http://www.karger.com/doi/10.1159/000331912.

49. Douglas I, Evans S, Rawlins MD, Smeeth L, Tabrizi SJ, Wexler NS. Juvenile Huntington's disease: a population-based study using the General Practice Research Database. *BMJ Open* [Internet]. 3(4), e002085–e002085 (2013). Available from: http://bmjopen.bmj.com/cgi/doi/10.1136/bmjopen-2012-002085.
* This paper needs to be read in conjunction with Evans *et al* 2013 (Ref 5). It focuses on the UK

prevalence of patients with with Huntington's disease who are ≤20 years.

50. Squitieri F, Andrew SE, Goldberg YP, *et al.* DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. *Hum. Mol. Genet.* [Internet]. 3 (12), 2103–2114 (1994). Available from: http://hmg.oxfordjournals.org/content/3/12/2103.abstract.
* This paper and the the one below are important for understanding the difference in CAG repeat lengths in different populations

51. Rubinsztein DC, Amos W, Leggo J, *et al.* Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. *Nat Genet* [Internet]. 7(4), 525–530 (1994). Available from: http://dx.doi.org/10.1038/ng0894-525.
*This paper and the one above are important for understanding the difference in CAG repeat lengths in different populations.

52. Warby SC, Montpetit A, Hayden AR, *et al.* CAG Expansion in the Huntington Disease Gene Is Associated with a Specific and Targetable Predisposing Haplogroup. *Am. J. Hum. Genet.* [Internet]. 84(3), 351–366 (2009). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2668007/.
** This paper and the one below describes HD mutations occurring on different haplotype backgrounds and provides evidence for true differneces in the prevalence of HD in different populations.

53. Warby SC, Visscher H, Collins JA, *et al.* HTT haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia. *Eur. J. Hum. Genet.* [Internet]. 19(5), 561–566 (2011). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3083615/.
** This paper and the one above describes HD mutations occurring on different haplotype backgrounds and provides evidence for true differneces in the prevalence of HD in different populations.

54. Sveinsson Ó, Halldórsson S, Olafsson E. An Unusually Low Prevalence of Huntington’s Disease in Iceland. *Eur. Neurol.* [Internet]. 68(1), 48–51 (2012). Available from: http://www.karger.com/doi/10.1159/000337680.

55. Gassivaro Gallo P, Buhagiar M, Cuschieri A, Viviani F. Huntington's chorea (HD) in Malta: Epidemiology and origins. *Int. J. Anthropol.* [Internet]. 14(2-3), 115–125 (1999). Available from: http://dx.doi.org/10.1007/BF02443891.

56. Morrison PJ. Accurate prevalence and uptake of testing for Huntington's disease. *Lancet Neurol.* [Internet]. 9(12), 1147 (2010). Available from: http://linkinghub.elsevier.com/retrieve/pii/S1474442210702878.

57. Wild EJ, Tabrizi SJ. Huntington's disease phenocopy syndromes. *Curr. Opin. Neurol.* [Internet]. 20(6) (2007). Available from: http://journals.lww.com/co-neurology/Fulltext/2007/12000/Huntington_s_disease_phenocopy_syndromes.14.aspx.

58. Nakashima K, Watanabe Y, Kusumi M, *et al.* Epidemiological and Genetic Studies of Huntington's Disease in the San-in Area of Japan. *Neuroepidemiology* [Internet]. 15(3), 126–131 (1996). Available from: http://www.karger.com/DOI/10.1159/000109899.

59. Hasegawa K. Epidemiology of Huntington's disease in Japan [Abstract]. *J. Neurol. Sci.* 357, e269 (2015).

60. Kim HS, Lyoo CH, Lee PH, *et al.* Current Status of Huntington's Disease in Korea: A Nationwide Survey and National Registry Analysis. *J. Mov. Disord.* [Internet]. 8(1), 14–20 (2015). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4298714&tool=pmcentrez&rendertype=abstract.

61. James CM, Houlihan GD, Snell RG, Cheadle JP, Harper PS. Late-onset Huntington's Disease: A Clinical and Molecular Study. *Age Ageing* [Internet]. 23 (6), 445–448 (1994). Available from: http://ageing.oxfordjournals.org/content/23/6/445.abstract.

62. Vlis MV Der, Zoeteweij M, Losekoot M, Haeringen a Van. Paradox of a better test for Huntington ' s disease. , 579–583 (2000).

63. Reverberi L, Contardi S, Fioravanti V, Cavallieri F, Codeluppi L, Valzania F. Epidemiological study on Huntington's disease in the districts of Modena and Reggio Emilia, Italy [abstract]. *Mov. Disord.* 29, Suppl 1: 582 (2014).

64. Panas M, Karadima G, Vassos E, *et al.* Huntington's disease in Greece: the experience of 14 years. *Clin. Genet.* [Internet]. 80(6), 586–590 (2011). Available from: http://doi.wiley.com/10.1111/j.1399-0004.2010.01603.x.

65. Vicente E, Garcia-Amigot F, Gaston M, *et al.* J37 Prevalence Of Huntington Disease In Navarra (spain). Sensitivity And Positive Predictive Value Of Different Sources Of Asecertaiment. *J. Neurol. Neurosurg. Psychiatry* [Internet]. 85(Suppl 1), A77–A77 (2014). Available from: http://jnnp.bmj.com/cgi/doi/10.1136/jnnp-2014-309032.220.

66. Peterlin B, Kobal J, Teran N, Flisar D, Lovrečić L. Epidemiology of Huntington's disease in Slovenia. *Acta Neurol. Scand.* [Internet]. 119(6), 371–375 (2009). Available from: http://doi.wiley.com/10.1111/j.1600-0404.2008.01110.x.

67. Magzhanov R, Saifullina E, Kutuev I, Khidiyatova I, Khusnutdinova E. M06 Epidemiology of Huntington's disease in the Republic of Bashkortostan. *J. Neurol. Neurosurg. Psychiatry* [Internet]. 83 (Suppl 1), A48–A48 (2012). Available from: http://jnnp.bmj.com/content/83/Suppl_1/A48.2.abstract.

68. Chen Y-Y, Lai C-H. Nationwide Population-Based Epidemiologic Study of Huntington's Disease in Taiwan. *Neuroepidemiology* [Internet]. 35(4), 250–254 (2010). Available from: http://www.karger.com/doi/10.1159/000319462.

FINANCIAL AND COMPETING INTERESTS DISCLOSURE

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

APPENDIX 1

Search strategy of electronic databases (EMBASE and MEDLINE).

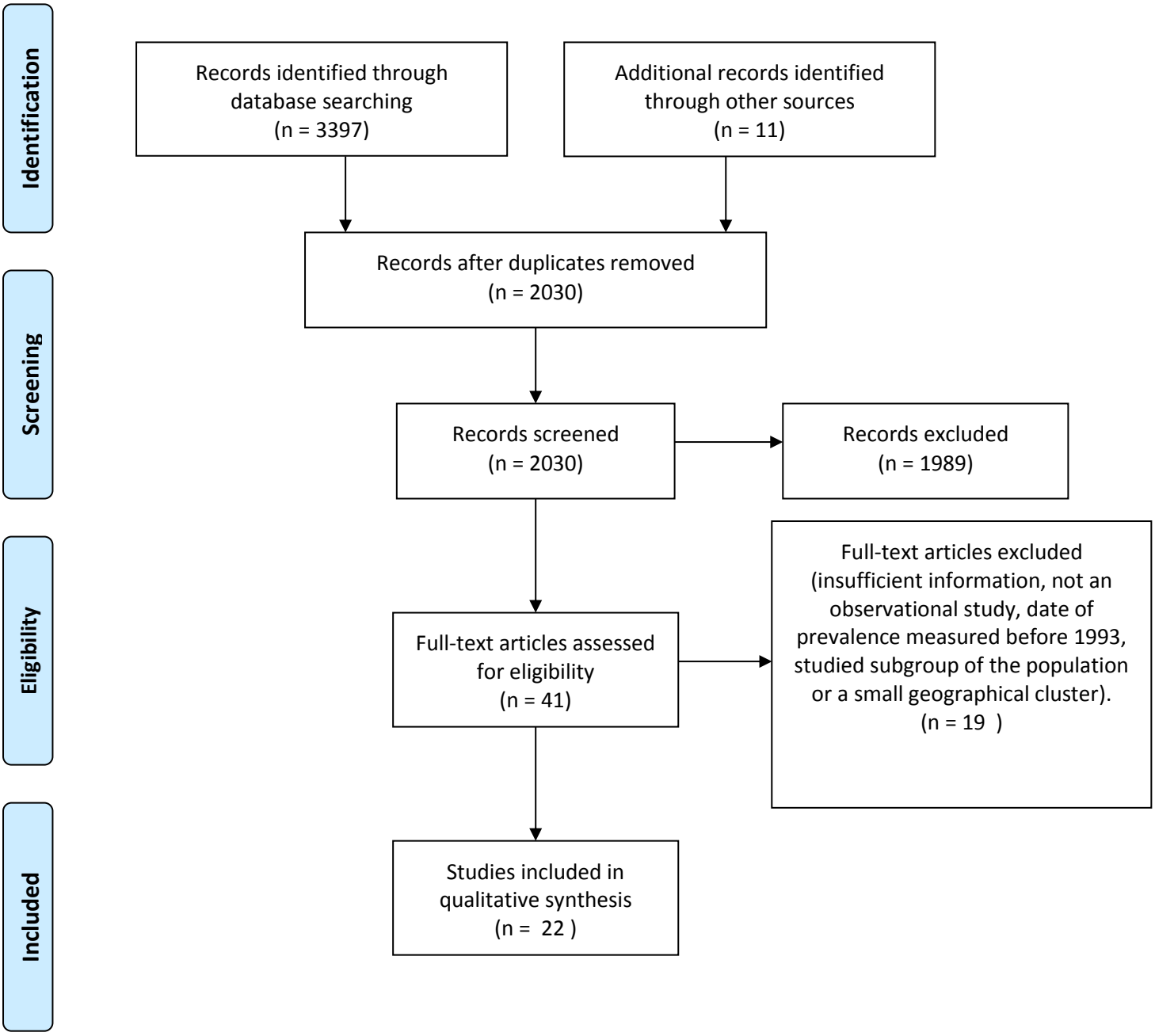
Search History

1. EMBASE; (Huntington* AND prevalence).ti,ab [Limit to: Publication Year 1993-2015]; 292 results.
2. EMBASE; (Huntington* AND population).ti,ab [Limit to: Publication Year 1993-2015]; 718 results.
3. EMBASE; (Huntington* AND incidence).ti,ab [Limit to: Publication Year 1993-2015]; 151 results.
4. EMBASE; (Huntington* AND epidemiology).ti,ab [Limit to: Publication Year 1993-2015]; 54 results.
5. EMBASE; 1 OR 2 OR 3 OR 4 [Limit to: Publication Year 1993-2015]; 1022 results.
6. Medline; exp HUNTINGTON DISEASE/ [Limit to: Publication Year 1993-2015]; 6845 results.
7. Medline; (prevalence OR population OR epidemiology OR incidence).ti,ab [Limit to: Publication Year 1993-2015]; 1435583 results.
8. Medline; 6 AND 7 [Limit to: Publication Year 1993-2015]; 409 results.
9. Medline; (Huntington* AND prevalence).ti,ab [Limit to: Publication Year 1993-2015]; 177 results.
10. Medline; (Huntington* AND population).ti,ab [Limit to: Publication Year 1993-2015]; 455 results.
11. Medline; (Huntington* AND epidemiology).ti,ab [Limit to: Publication Year 1993-2015]; 32 results.
12. Medline; (Huntington* AND incidence).ti,ab [Limit to: Publication Year 1993-2015]; 93 results.
13. Medline; 9 OR 10 OR 11 OR 12 [Limit to: Publication Year 1993-2015]; 772 results.
14. EMBASE; exp HUNTINGTON CHOREA/; 19324 results.
15. EMBASE; (prevalence OR population OR epidemiology OR incidence).ti,ab; 2278710 results.
16. EMBASE; 14 AND 15; 1194 results.
17. EMBASE; 5 OR 16 [Limit to: Publication Year 1993-2015]; 1199 results.
18. Medline; 8 OR 13 [Limit to: Publication Year 1993-2015]; 820 results.

Date of search: 19/10/2015.

FIGURES

Figure 1: Flow chart of systematic review procedure for identifying and selecting studies for reporting the prevalence of Huntington's disease in discrete populations.



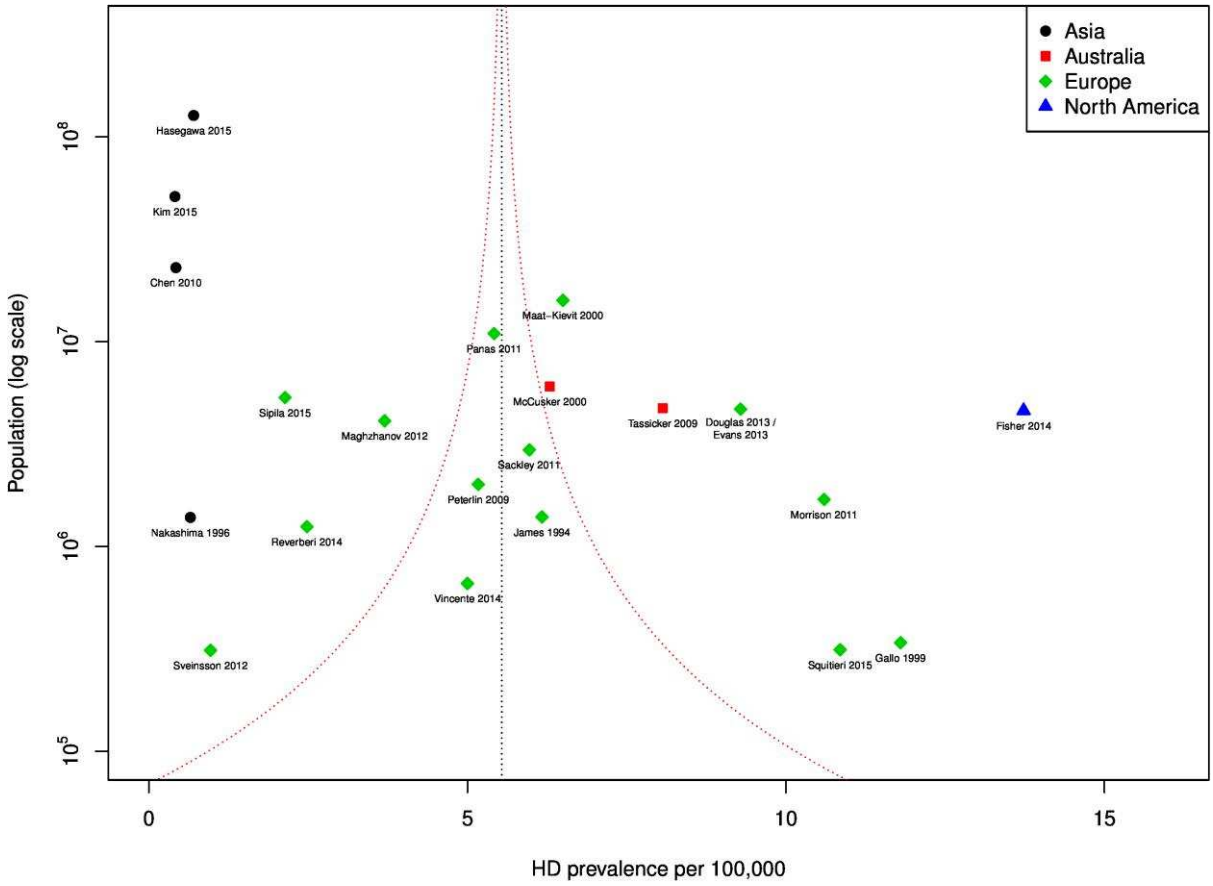
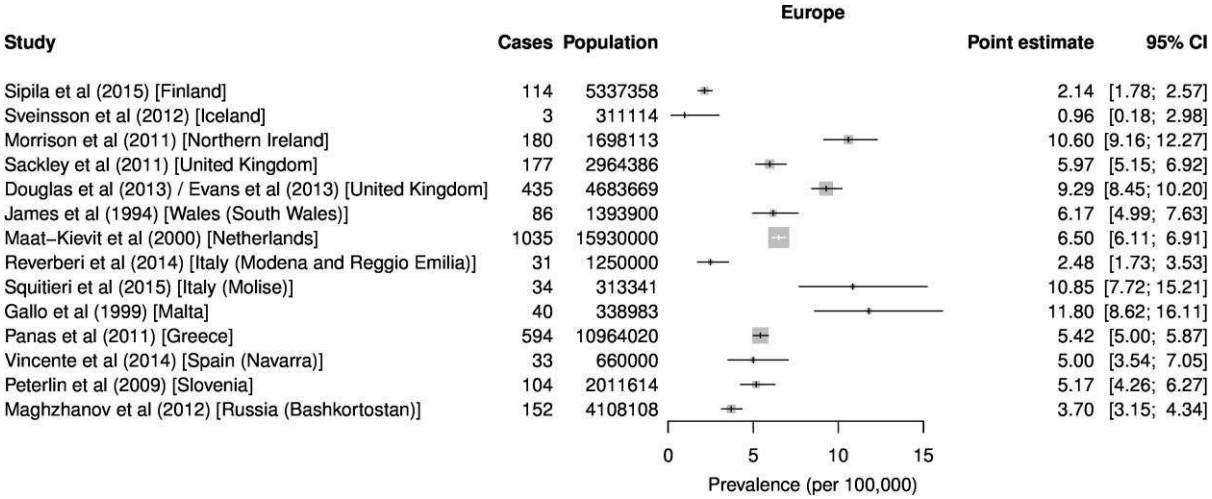
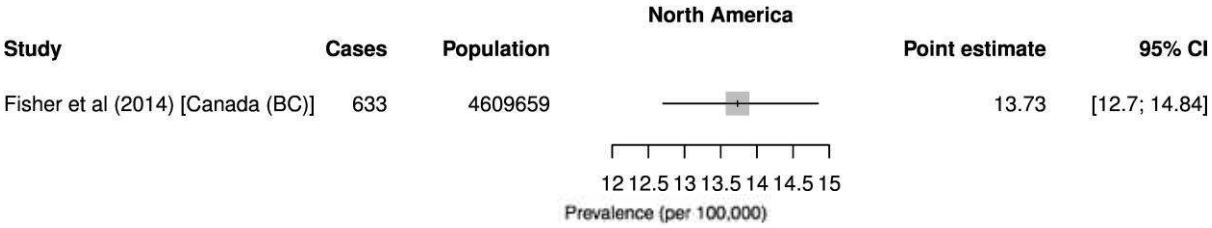


Figure 2 - Funnel plot of population size against HD prevalence using data from studies meeting the inclusion criteria.

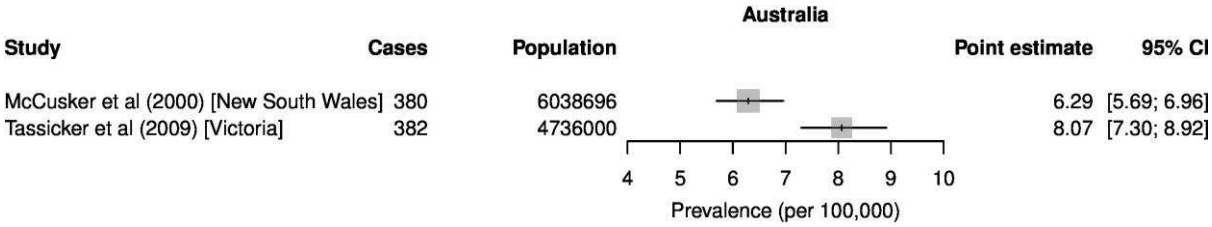
A



B



C



D

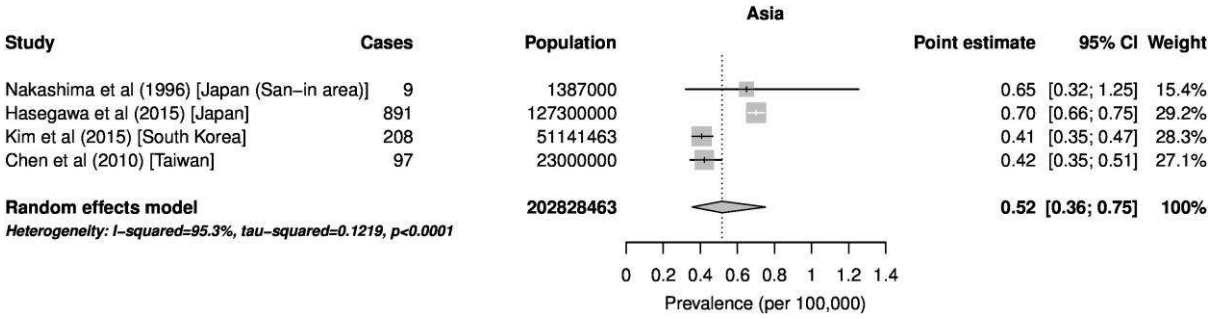


Figure 3 – Forest plots of studies of Huntington's disease prevalence by continent.
A – Europe, B – North America, C – Australia, D – Asia.

Figure 4 - Ascertained Prevalence of Huntington's Disease in Different Populations (1993-2015). Bubble diameter proportional to prevalence per 100 000 population. (Figure created using http://cartodb.com)

TABLES

Table 1 - Study Design and Selection Criteria

Study Design	
Selection Criteria for Studies	Population-based observational studies
	Defined population
	Ascertainment of symptomatic cases of Huntington's disease
	Study conducted from 1993 onwards
Population	Individuals with a diagnosis of Huntington's disease.
Outcomes	Prevalence of Huntington's disease in defined geographical populations.

Table 2: Studies of the Prevalence of Huntington's Disease

Region	Prevalence Date	Sources of Case Ascertainment	Diagnostic Criteria	Population Size	Number of Cases on Prevalence Date	Prevalence per 100,000 population (95% CI)	Reference
EUROPE							
Finland	2010	HR, CR, Lab, Family Federation of Finland records, DC	Clinical phenotype plus either a family history of HD, a family history or motor symptoms suggesting HD or a positive DNA analysis (CAG repeat length ≥ 37)	5 337 358 (calculated)	114	2.14 (1.78 – 2.57)	Sipila *et al* (2015) [6]
Iceland	2007	HR, CR, FS, DC	Clinical phenotype plus either a family history of HD or a positive DNA analysis (CAG repeat length unstated)	311 114	3	0.96 (0.18 - 2.98)	Sveinsson *et al* (2012) [54]
Northern Ireland	2001	Prospective: CTC, HDR	Clinical phenotype with a positive DNA analysis (CAG repeat length ≥ 36)	1 698 113 (calculated)	180	10.6 (9.16 – 12.27)	Morrison *et al* (2011) [9]
United Kingdom	2008	PCD (THIN)	Administrative Read code	2 964 386	Unspecified	5.97 (5.15 – 6.92)	Sackley *et al* (2011) [48]
United Kingdom	2010	PCD (GPRD)	Administrative Read code	4 683 669	435	9.29 (8.45 – 10.20)	Douglas *et al* (2013) [49] and Evans *et al* (2013) [5]
Wales (South Wales)	1994	Prospective: HDR	Clinical phenotype or clinical phenotype with positive DNA analysis (CAG repeat length ≥ 36)	1 393 900	86	6.17 (4.99 – 7.63)	James *et al* (1994) [61]
Netherlands (Leiden)	2000	HDR (Leiden Roster)	Unspecified	15 930 000	Unspecified	6.50 (6.11 - 6.91)	Maat-Kievit *et al* (2000) [62]
Italy (Modena and Reggio Emilia)	2013	Unspecified (likely used death certificates to ascertain living cases)	Unspecified	1 250 000	31	2.48 (1.73 – 3.53)	Reverberi *et al* (2014) [63]
Italy (Molise)	2013	HR, CR, HDA RDR, FS	Clinical phenotype or clinical phenotype with positive DNA analysis (CAG repeat length unstated).	313 341	34	10.85 (7.72-15.21)	Squitieri *et al* (2015) [46]
Malta	1994	CTC	Unspecified	338 983 (calculated)	40	11.80 (8.62 – 16.11)	Gallo *et al* (1999) [55]
Greece	2008	Prospective: CTC	Clinical phenotype or clinical phenotype with a positive DNA analysis (CAG repeat length ≥ 36)	10 964 020	594	5.42 (5.00 – 5.87)	Panas *et al* (2011)[64]

Spain (Navarra)	2014	HR, CR, PCD	Clinical phenotype plus a family history of HD or a positive DNA analysis (CAG repeat length unstated)	660 000 (calculated)	33	5.00 (3.54 – 7.05)	Vincente *et al* (2014) [65]
Slovenia	2006	HR, CR, HDR, CTC, HDA	Clinical phenotype with a positive DNA analysis (CAG repeat length ≥ 37).	2 011 614	104	5.17 (4.26 - 6.27)	Peterlin *et al* (2009) [66]
Russia (Bashkortostan)	2012	National Genetic Register	Unspecified	1 250 000	152	3.70 (3.15 – 4.34)	Maghzhanov *et al* (2012) [67]
NORTH AMERICA							
Canada (British Columbia)	2012	HR, CR, Lab, CS, FS, NH, PCR, HDA	Clinical phenotype or clinical phenotype with a positive DNA analysis (CAG repeat length ≥ 36)	4 609 659	633	13.7 (12.7-14.8)	Fisher and Hayden (2014) [4]
AUSTRALIA							
Australia (New South Wales)	1996	HR, CS, HAD, FS	Clinical phenotype plus a family history of HD or a positive DNA analysis (CAG repeat length unstated)	6 038 696	380	6.29 (5.69-6.96)	McCusker *et al* (2000) [3]
Australia (Victoria)	1999	Lab, CTC	Unspecified	4 736 000	382	8.07 (7.30 – 8.92)	Tassicker *et al* (2009) [8]
ASIA							
Japan (San-in area)	1993	HR, CS	Clinical phenotype with a positive DNA analysis (CAG repeat length unstated) and atrophy of the caudate nucleus on CT/MRI	1 387 000	9	0.65 (0.32 – 1.25)	Nakashima *et al* (1996) [58]
Japan	Unspecified	Japan Intractable Disease Information Center, Department of the Specific Disease Control, Japanese Ministry of Health, Labor and Welfare.	Unspecified	127 300 000 (calculated)	891 (calculated)	0.70 (0.66 – 0.75)	Hasegawa *et al* (2015)[59]
South Korea	2013	HR, HDR, RDR	Administrative codes on NHI database or clinical phenotype with a positive DNA analysis	51 141 463	208	0.41 (0.35 – 0.47)	Kim *et al* (2015) [60]
Taiwan	2007	NHI	Administrative code (ICD-9 code 333.4)	23 000 000	97	0.42 (0.35 -0.51)	Chen *et al* (2010) [68]

Legend for Table 2

HR	Hospital Records and Hospital Discharge Registers
CR	Clinic Records
CS	Clinician Surveys
Lab	Genetic Testing Laboratories
CTC	Centralised Testing Centre
HDR	Huntington's Disease Registry
RDR	Rare Disease Registry
PCD	Primary Care Database
NHI	National Health Insurance Database
HDA	Huntington's Disease Association
FS	Family Surveys and Family Pedigrees
NH	Nursing Homes
VA	Veteran Affairs
SS	Social Services
DC	Death Certificates
THIN	The Health Improvement Network
GPRD	General Practice Research Database

Table 3 - Factors that may explain the geographical variation in HD prevalence.

Factors that may explain the geographical variation in HD prevalence	
Differences in the true prevalence	Average length of CAG repeat in the unaffected population which correlates to the new mutation rate [50,51]
	Frequency of A1 and A2 *HTT* haplotypes in the unaffected population. [53]
	The founder effect in small, geographically isolated populations. [54,55]
	Life expectancy in the general population. [47]
Differences in the ascertainment of HD cases.	Sensitivity and specificity of the data sources used for case ascertainment.
	The use of single or multiple sources for case ascertainment.
	Ease of accessing healthcare services in order to diagnose HD.
	Clinician familiarity with HD as a disease entity.
	The presence of large private or informal healthcare sector leads to an underascertainment of HD cases in national registers.
	Different incentives to hide a diagnosis of HD depending on local social stigma, real or perceived employment discrimination or insurance-based healthcare provision.