Ascertainment Bias Causes False Signal of Anticipation in Genetic Prion Disease

Eric Vallabh Minikel,[1,2,3,*] Inga Zerr,[4,5] Steven J. Collins,[6] Claudia Ponto,[4] Alison Boyd,[6] Genevieve Klug,[6] André Karch,[4] Joanna Kenny,[7] John Collinge,[7] Leonel T. Takada,[8] Sven Forner,[8] Jamie C. Fong,[8] Simon Mead,[7,9] and Michael D. Geschwind[8,9]

Anticipation is the phenomenon whereby age of onset in genetic disease decreases in successive generations. Three independent reports have claimed anticipation in Creutzfeldt-Jakob disease (CJD) caused by the c.598G>A mutation in *PRNP* encoding a p.Glu200Lys (E200K) substitution in the prion protein. If confirmed, this finding would carry clear implications for genetic counseling. We analyzed pedigrees with this mutation from four prion centers worldwide (n = 217 individuals with the mutation) to analyze age of onset and death in affected and censored individuals. We show through simulation that selective ascertainment of individuals whose onset falls within the historical window since the mutation's 1989 discovery is sufficient to create robust false signals both of anticipation and of heritability of age of onset. In our data set, the number of years of anticipation observed depends upon how strictly the data are limited by the ascertainment window. Among individuals whose disease was directly observed at a study center, a 28-year difference between parent and child age of onset is observed ($p = 0.002$), but including individuals ascertained retrospectively through family history reduces this figure to 7 years ($p = 0.005$). Applying survival analysis to the most thoroughly ascertained subset of data eliminates the signal of anticipation. Moreover, even non-CJD deaths exhibit 16 years anticipation ($p = 0.002$), indicating that ascertainment bias can entirely explain observed anticipation. We suggest that reports of anticipation in genetic prion disease are driven entirely by ascertainment bias. Guidelines for future studies claiming statistical evidence for anticipation are suggested.

Introduction

Prion diseases are uniformly fatal, progressive neurodegenerative disorders caused by the conversion of the cellular prion protein, PrP^{C}, to a misfolded conformation known as the prion, or PrP^{Sc}, in which Sc stands for scrapie, the prion disease of sheep and goats.[1] In humans, prion diseases have an incidence of approximately 1 death per 1 million individuals per year,[2] and usually occur as simplex cases in individuals with two wild-type (WT) copies of the prion protein gene (*PRNP* [MIM 176640]), commonly referred to as sporadic cases. A minority of cases are genetic and, very rarely, prion disease can be environmentally acquired.[1] Creutzfeldt-Jakob disease (MIM 123400) caused by the c.598G>A (dbSNP id rs28933385) mutation, which encodes a p.Glu200Lys (E200K) substitution in PrP, is the most common genetic form of prion disease worldwide.[3] This point mutation was first identified in 1989[4] and was established as a dominant Mendelian cause of disease by 1991.[5–7] Disease penetrance in mutation heterozygotes appears to reach 80%–100% by age 80.[8,9] Reported estimates of the mean age of onset in individuals with this mutation range from 53[7] to 63,[10] and the mean survival after disease onset is 7 months.[11]

Three reports[12,13] (see also Web Resources) have claimed statistical evidence that this genetic prion disease exhibits *anticipation*, a phenomenon in which successive generations exhibit progressively earlier disease onset or more severe presentation.[14] These studies reported a 7 to 14 year younger age of onset or death among children in affected parent-child pairs, suggesting implications for genetic counseling.

The only genetic mechanisms known to cause anticipation are the germline expansion of unstable repeats in disorders such as Huntington's disease and type 1 myotonic dystrophy,[15,16] and telomere shortening in disorders such as dyskeratosis congenita and breast cancer.[17,18] Anticipation in a genetic prion disease might raise the question of whether disease in children is accelerated by exposure to infectious material during their parents' illness, however, the only known routes of human-to-human prion transmission are cannibalism[19] and iatrogenic exposure.[20,21]

Because a variety of sources of ascertainment bias are known to contribute to false statistical signals of anticipation,[14,22–25] we set out to determine whether the anticipation reported for Glu200Lys genetic prion disease could be a statistical artifact. An individual's observed age of onset cannot be greater than the age at interview or ascertainment, and previous studies have modeled the effects of this right truncation of age of onset[22] and provided methods of correction.[23–26] These methods of correction,

[1]Prion Alliance, Cambridge, MA 02139, USA; [2]Broad Institute, Cambridge, MA 02142, USA; [3]Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; [4]National Reference Center for TSE, Georg-August University, 37073 Goettingen, Germany; [5]German Center for Neurodegenerative Disease (DZNE), 37073 Goettingen, Germany; [6]Australian National Creutzfeldt-Jakob Disease Registry, The University of Melbourne, Parkville, Australia 3010; [7]MRC Prion Unit, Department of Neurodegenerative Disease, University College London (UCL) Institute of Neurology, and NHS National Prion Clinic, National Hospital for Neurology and Neurosurgery, UCL Hospitals NHS Trust, Queen Square, WC1N 3BG London, UK; [8]Department of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, CA 94158, USA
[9]These authors contributed equally to this work
*Correspondence: eric.minikel@prionalliance.org
http://dx.doi.org/10.1016/j.ajhg.2014.09.003.

however, require either a consistent, known set of ascertainment criteria[23] or the use of only a subset of available data.[26] In rare diseases, data may be too sparse for subsetting, and may not represent a unified ascertainment effort, but rather consist of a mix of data points ascertained retrospectively (through family histories of varying depth and quality), directly (symptomatic individuals seen clinically), and prospectively (asymptomatic individuals with a mutation, followed for varying amounts of time). We therefore sought to model ascertainment bias due to left- and right-truncation not of the age of onset per se, but of the year of onset. Because the Glu200Lys substitution was discovered only 25 years ago and most prion surveillance programs and clinical centers have been established even more recently, we hypothesized that the selective ascertainment of parents and children whose deaths both occurred within this 25-year window could explain the reported differences in parent and child age of death.

To test this hypothesis, here we combine data from four national prion study centers to assemble the largest Glu200Lys cohort (n = 217 individuals) yet reported. We first create a simulation of the ascertainment of parent-child pairs with a mutation to identify conditions under which naive paired t tests will detect a false signal of anticipation. We explore methods for detecting and controlling for this ascertainment bias. We then apply our analytical framework to our Glu200Lys data set and successfully reproduce the anticipation reported by other groups but demonstrate that this anticipation is a false positive due to ascertainment bias.

Material and Methods

Data Collection

We combined data collected on Glu200Lys individuals and their families from four research centers with data collection practices as follows.

Australian National Creutzfeldt-Jakob Disease Registry

Details of Australian National Creutzfeldt-Jakob Disease Registry (ANCJDR) surveillance mechanisms, as well as data collection and analysis methods, have been reported previously.[27,28] In brief, prospective national surveillance of CJD has been undertaken since 1993 with CJD a Notifiable Disease throughout Australia since 2006. ANCJDR collects detailed medico-demographic information on suspect cases, including family histories, and provides diagnostic tests including *PRNP* genotyping. Year and age of death are primary variables with information on age at onset of first symptom collected if available. Informed, written consent was obtained from participants or legal next of kin. Ethical approval was obtained from the Office of Research Ethics and Integrity at The University of Melbourne.

German CJD Surveillance Unit

Details of German CJD surveillance have been reported previously.[29–31] In brief, the Surveillance Unit in Goettingen has collected data on all suspected prion disease cases in Germany since 1993. Diagnostic information is obtained from reporting hospitals and where possible, confirmation by autopsy is sought. The Surveillance Unit also accepts clinical referrals, provides diagnostic tests including *PRNP* genotyping, and, where possible, collects family history. Age of onset is defined from first symptom of a progressive neuropsychiatric disorder by interview with family members. Informed, written consent was obtained from participants or legal next of kin. Ethical approval was obtained from the Ethical Committee at the University Medical School, Georg-August University Goettingen.

MRC Prion Unit/NHS National Prion Clinic

The UK has had a centralized tertiary clinical referral service for CJD since 1991. Since 2004, all suspected CJD cases from the UK are referred to the NHS National Prion Clinic at the National Hospital for Neurology and Neurosurgery (NHNN) at University College London Hospitals NHS Trust. Age of onset was defined from first symptom of a progressive neuropsychiatric disorder and family history was obtained by interview with family members. Other details of data collection have been described previously.[32] Informed, written consent was obtained from participants or legal next of kin. Ethical approval was obtained from the NHNN/Institute of Neurology Joint Research Ethics Committee.

Memory and Aging Center, University of California San Francisco

The UCSF cohort comprises symptomatic and asymptomatic individuals from Glu200Lys families referred from the U.S. and abroad to the rapidly progressive dementia and Prion Disease research program since August 2001.[33–35] *PRNP* genotyping[36] was performed at the National Prion Disease Pathology Surveillance Center (Cleveland, OH), or by outside laboratories in some of the individuals who were tested prior to UCSF referral or lived abroad. Symptom onset was determined as previously reported.[37] A detailed, usually three generation, family pedigree was made by a neurologist and/or clinical genetic counselor for individuals participating in research. Further data were collected from medical records sent by referring physicians and/or from direct contact with family members (by email or telephone). Informed, written consent was obtained from research participants or legal next of kin. The UCSF data included in this study have been collected through UCSF Institutional Review Board-approved research protocols.

Data Annotation

Directly observed individuals were defined as those either seen clinically at one of the four centers or officially reported to one of the centers in its prion disease surveillance role. Indirectly observed individuals were those ascertained through interview with family members. Individuals were considered to have the c.598>A mutation if they (1) had either a genotyping test indicating the presence of the mutation or were related to someone with a positive test and (2) had a diagnosis of CJD or (3) were deemed to have died of CJD based on information obtained from interviewed family members. Individuals were considered to not have the mutation if they (1) had a genotyping test indicating the absence of the c.598G>A mutation or (2) were related to the family only by marriage and thus lacked a blood relationship to any affected individual. The Glu200Lys substitution causes CJD with nearly 100% penetrance,[8,9] but individuals with two WT *PRNP* alleles have a very low disease incidence of only about 1 in 1 million per year.[2] This incidence translates into a lifetime risk roughly on the order of 1 in 10,000 for a WT individual. By Bayes' rule, the high penetrance of the mutation and rarity of nongenetic CJD cases mean that any CJD case in a Glu200Lys pedigree is overwhelmingly likely to be genetic. Therefore, we assumed that all CJD cases in these pedigrees were due to the c.598G>A mutation. Data on four individuals with CJD were flagged as questionable