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, Michael D. Geschwind 8,9

1. Prion Alliance, Cambridge, Massachusetts 02139
2. Broad Institute, Cambridge, Massachusetts 02142
3. Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114
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.
8. Memory and Aging Center, University of California San Francisco, San Francisco, California 94158
9. These authors contributed equally to the manuscript.

*Correspondence may be addressed to:

Eric Vallabh Minikel
eric.minikel@prionalliance.org

ABSTRACT

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 dataset, 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 = .002), but including individuals ascertained retrospectively through family history reduces this figure to 7 years (p = .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 = .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 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 may 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, 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 naïve paired t tests will detect a false signal of anticipation. We explore methods for detecting and controlling for

3

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

this ascertainment bias. We then apply our analytical framework to our Glu200Lys dataset and successfully reproduce the anticipation reported by other groups, but demonstrate that this anticipation is a false positive due to ascertainment bias.

MATERIALS 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 (ANCJDR): Details of 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 U.K. 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 (UCSF): 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. This 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 either (1) had a genotyping test indicating the presence of the mutation, or were related to someone with a positive test and either (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 (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 wild-type *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 wild-type individual. By Bayes' rule, the high penetrance of the mutation and rarity of non-genetic 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 due to uncertain diagnosis, uncertain relatedness to other individuals or uncertain age at death; none of the results reported here differed when the analysis was re-run excluding these individuals. Missing *PRNP* codon 129 information was imputed from affected family members when possible. For the purpose of assessing rates of predictive genetic testing, "at-risk" individuals were defined as those who were not symptomatic at last follow-up but who did have a parent deemed to have the mutation, per the above criteria. Recursively, individuals were also considered at-risk if they had a parent who qualified as at-risk according to the above definition.

Simulation. We hypothesized (see Introduction) that the selective ascertainment of individuals with disease onset within a specified window (for instance, 1989-2013) might explain reported anticipation in Glu200Lys prion disease. To assess whether such an "ascertainment window" was sufficient to create a false signal of anticipation, we created a simulation model in which parent and child ages of onset are drawn from the same specified distribution but ascertainment of parent-child pairs is selective based on year of onset. For each round of simulation we generated n =100,000 parent-child pairs. We only simulated parents and children with the causative mutation. We wished to model a situation in which individuals have been falling ill of prion disease continuously throughout history, and will continue to do so into the future, whereas our data are limited only by an artificial ascertainment window. We therefore generated parents with a year of birth uniformly distributed from 1700 to 2000, in order to ensure that (1) the density of disease onsets over our largest ascertainment window (1880 to 2013) would be uniform and (2) that observed years of onset would effectively be left-truncated only according to our ascertainment criteria and not according to the underlying simulated distribution. We set the child's birth to occur 28 ± 6 years after the parent's birth, per the actual distribution from our data. To simulate an age of onset distribution while accounting for censoring due to deaths due to other causes ("competing risks"), we first created a normally distributed 64 ± 10 year age of onset to approximate the distribution in our real data (see Results) and then randomly censored individuals according to the U.S actuarial life table for 2009 (see Web Resources). We did not model changes in the life expectancy over the time periods considered here. Further details of the specific simulations are explained in Results and in the legends of Tables 1 and 2.

Statistical analyses. All simulations and analyses of Glu200Lys pedigree data were conducted in R 3.0.2. In order to demonstrate the effects of ascertainment bias on the evidence for anticipation

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.
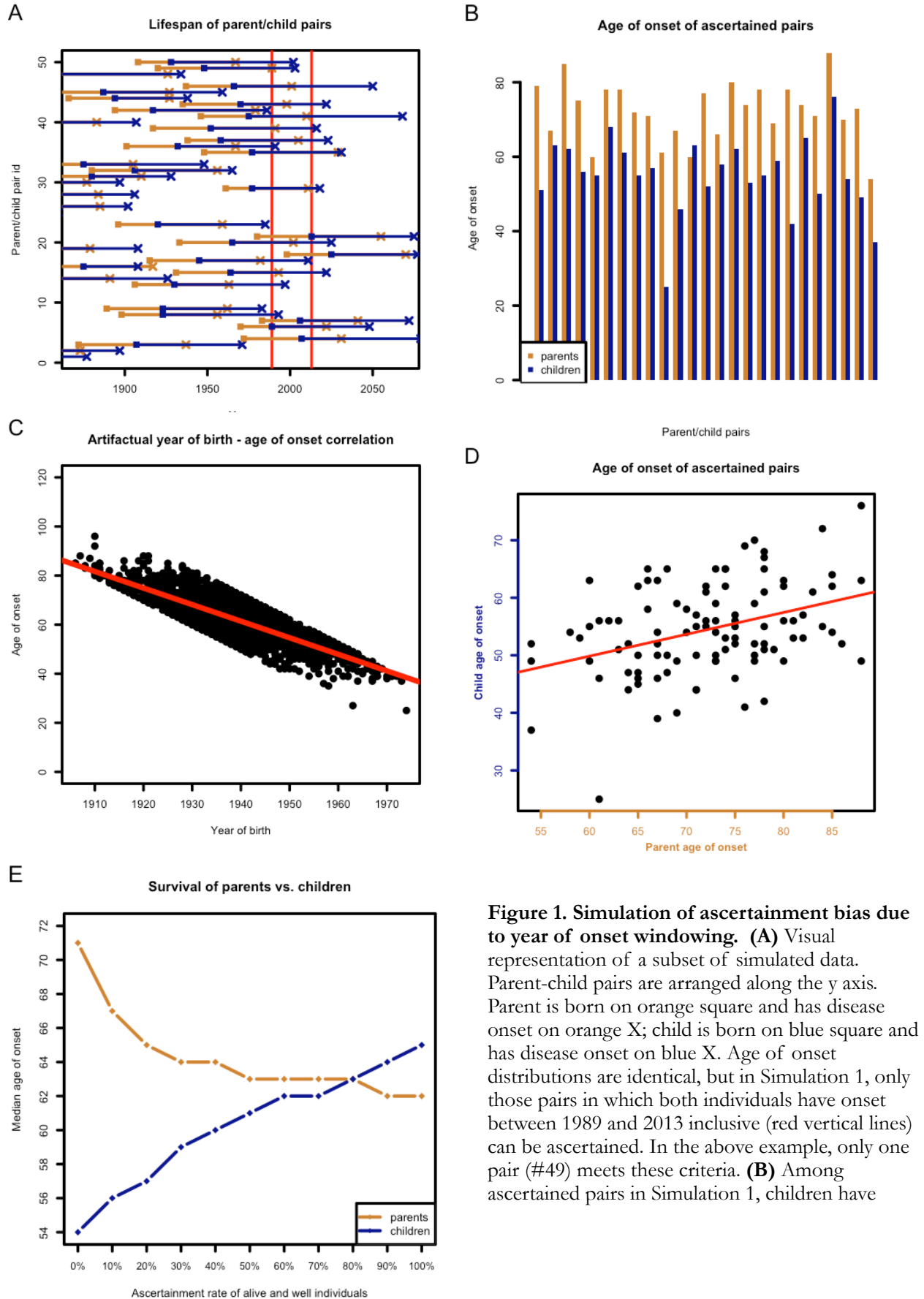
reported by previous studies of genetic prion disease, we adopted methods from those studies when possible. For paired t tests, all possible parent-child pairs were generated from Glu200Lys pedigree structures using a SQL *join* operation. This results in multiple-counting parents of multiple affected children, and counting both as parents and as children those individuals in the middle of pedigrees with three or more affected generations. Multiple counting means that pairs are not independent; this mirrors the methods from at least one prior study of prion disease anticipation.[13] Based on these paired lists of parents and children, differences in parent and child age of onset were assessed using 2-tailed paired t tests (the R *t.test* function) for naïve comparisons of observed age of onset distributions. Correlation between parent and child age of onset was assessed using linear regression (the R *lm* function). Survival analysis utilized the R *survival* package. For survival analysis, to avoid multiple-counting and ensure independence of all pairs, we randomly selected one parent-child pair from each pedigree and compared the survival curve for all parents to that of all children using a log-rank test (the R *survdiff* function). Because we randomized which pairs were included, we repeated this analysis over 1000 iterations and have reported aggregate statistics.. Source code and output are available online (see Web Resources).

RESULTS

Simulated effects of ascertainment window. We generated 100,000 parent-child pairs with all individuals harboring a dominant genetic mutation and ages of onset for all individuals independent and identically distributed, corresponding to a scenario in which no anticipation or heritability of age of onset are present (see Materials and Methods and Figure 1A). The mean age of onset without accounting for censoring was 62, whereas the median age of onset in survival analysis accounting for censoring due to intercurrent deaths was 64. When considering all simulated individuals, there was no difference in age of onset (p = .91, 2-tailed paired t-test) no correlation between parent and child age of onset (p = .53, linear regression) and no correlation between the year of birth and age of onset for all individuals (p = .59, linear regression).

Next we considered the effects of selectively ascertaining individuals whose disease onset occurs within an "ascertainment window," which we define as a range of years in which disease onsets can be observed. It is impossible to ascertain ages of onset after 2013, the last year in which our data were collected. Right-truncating the year of onset at 2013 and considering all affected pairs introduced only a very small difference between parent and child age of onset (.37 years, p < .0001, 2-tailed paired t test). Because we had generated parent-child pairs with years of birth as early as 1700, only a minority of pairs was affected by this right truncation. Spreading the simulated pairs out over an even greater range of years of parent birth (year 0 – 2000) made this anticipation cease to be significant (p = .31, 2-tailed paired t test), indicating that right truncation of year of onset is not sufficient cause a false signal of anticipation. In practice, any real dataset will be both left-truncated (when researchers began to study the disease) and right-truncated (at the present year or when the study stopped, whichever is earlier). For Glu200Lys CJD, we reasoned that ascertainment rates would be higher after the discovery of the disease's causal mutation in 1989 [4], so we next considered the effects of ascertaining only those individuals whose year of onset occurs between 1989 and 2013 inclusive.

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

A Lifespan of parent/child pairs

B Age of onset of ascertained pairs

C Artifactual year of birth - age of onset correlation

D Age of onset of ascertained pairs

E Survival of parents vs. children

Figure 1. Simulation of ascertainment bias due to year of onset windowing. (A) Visual representation of a subset of simulated data. Parent-child pairs are arranged along the y axis. Parent is born on orange square and has disease onset on orange X; child is born on blue square and has disease onset on blue X. Age of onset distributions are identical, but in Simulation 1, only those pairs in which both individuals have onset between 1989 and 2013 inclusive (red vertical lines) can be ascertained. In the above example, only one pair (#49) meets these criteria. **(B)** Among ascertained pairs in Simulation 1, children have

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

(Figure 1 continued) almost categorically younger onset than their parents, leading to 17 years of observed anticipation (p < .0001, 2-tailed paired t-test). For visibility, a subset of simulated points are shown. **(C)** Ascertainment windowing introduces an artifactual correlation between year of birth and age of onset, slope = -.67 (p < .0001, linear regression). **(D)** Ascertainment windowing also leads to a correlation between parent and child age of onset and thus a false signal of heritability (slope = .48, p < .0001, linear regression). **(E)** Ascertained pairs are supplemented with those pairs where one individual's onset occurs in 1989-2013 and the other is alive and well as of 2013. The plot shows that the median survival of parents and children in Kaplan-Meier curves depends upon what proportion of alive and well individuals are included. If 0% are included, then 17 years anticipation are observed, just as in (B). As ascertainment increases, the anticipation is reduced. As inclusion of alive and well individuals approaches 100%, children have longer survival than parents. This is because more of the children are censored, so their age of onset distribution better reflects the hypothetical distribution of age of onset without the influence of competing risks.

When only those simulated pairs in which parent and child disease onset both occurred within this window were considered (Figure 1A), we observed 17 years of anticipation (p < .0001, 2-tailed paired t-test, Figure 1B). Because parents were born, on average, 28 years earlier than their children, yet the window of observation was only 25 years long, the ascertained pairs were vastly enriched for those in which the child dies at least 3 years younger than the parent. Under these conditions, we also observed an artifactual correlation between year of birth and age of onset, with slope of -.67, (p < .0001, linear regression, Figure 1C). This relationship is intuitive because an individual born in 1970, for instance, can only be included in the dataset if onset occurs by age 43 (year 2013), and an individual born in 1940 can only be included if onset occurs at age 49 or later (year 1989). Year of birth – age of onset correlation is therefore a consequence of ascertaining only individuals with the year of onset within a particular window. This problem has previously been noted in Huntington's disease pedigrees[38]. Among the ascertained pairs, parent and child age of onset were also correlated with a slope of 0.48 (p < .0001, linear regression, Figure 1D). In single parent-offspring regression, the slope can be doubled to obtain an estimate of phenotypic variance explained by additive genetic heritability plus environmental effects[39]. Applying this formula to the ascertained data would suggest that age of onset is up to 96% heritable. Because the true distributions of parent and child ages of onset in our simulation were neither different nor correlated, this shows that the year of onset "windowing" simulated here is sufficient to create false signals of both anticipation and heritability. We confirmed this finding through 1000 iterations with the same set of simulation conditions and n = 100 ascertained pairs to simulate a realistic sample size for a rare disease (Table 1, Simulation 1)

Because we knew the anticipation and heritability identified in our above simulation to be false signals, we asked whether improved ascertainment or simple analytical methods could disprove them. We considered a two-step ascertainment model; in the first step, we ascertained all individuals with disease onset between 1989 and 2013 (simulating clinical visits), and in the second step we ascertained any and all individuals still alive and well as of 2013 who were parents or children of the individuals ascertained in step 1 (simulating genetic testing and prospective follow-up of individuals with the mutation). When we compared parent and child survival curves under this two-step ascertainment model in Kaplan-Meier survival analysis, the median age of onset of children was greater than that of parents (64 vs. 62 years, p < .0001, log-rank test). This counter-intuitive finding of later child onset can be explained as follows. Because parents are born earlier than children, pairs in this two-step ascertainment model are predominantly ones in which the parent dies of genetic prion disease between 1989 and 2013 while the child is alive and well in 2013. In such pairs, the parent's age of onset distribution is truncated by competing risks (parents who *would* have had older disease onsets were never ascertained because they died of other causes first, thus making disease

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

onset appear younger). The child's age of onset distribution, however, is less affected by competing risks (children who would have older disease onsets are ascertained through genetic testing even if they will never develop the disease in their lifetimes).

Recall that in the two-step ascertainment model above, an affected individual is ascertained in step 1, and if that individual has a parent or child alive and well in 2013, then that parent or child is *always* ascertained. In practice, however, it is difficult to ascertain 100% of asymptomatic relatives of an affected individual. We therefore also modeled *incomplete* ascertainment of asymptomatic individuals. When we did so we saw that the false signal of anticipation from naïve t tests was not fully corrected (Figure 1E). The difference in median survival according to the log-rank test was 17 years when no censored individuals were included, and this difference shrunk to 7 years when 20% of individuals alive and well in 2013 were included, and 2 years when 50% of such individuals were included. This indicates that accounting for censored observations is not sufficient to remove false signals of anticipation unless ascertainment of censored individuals is exhaustive. In long-term studies of predictive genetic testing for neurodegenerative diseases, it has been reported that only 3% to 24% of individuals with risk of having inherited a mutation pursue predictive genetic testing.[40–45] Survival analysis including asymptomatic individuals with a mutation is therefore unlikely to mitigate false signals of anticipation and heritability in this disease setting.

We also asked whether including year of birth as a covariate could eliminate the false signal of heritability we observed in our simulation. When both parent age of onset and child year of birth were used in a linear model to explain child age of onset, only the child year of birth proved correlated (slope = -.72, p < .0001, linear regression), whereas parent age of onset was no longer significant (p = .73, linear regression). When we simulated an age of onset that truly was heritable, this method reduced, but did not eliminate, the estimate of heritability (Table S1).

Simulated effects of different ascertainment criteria. In Simulation 1 (Table 1) we assumed that it was only possible to include individuals directly ascertained within the window from 1989 to 2013. In reality, some Glu200Lys families have been followed since as early as 1963[7] and it is also possible to obtain information on deceased individuals retrospectively through interview with family members, though it can be more difficult for individuals whose year of death is long past[46]. We therefore asked whether false anticipation and heritability would still be observed with larger or more flexible ascertainment windows. In this analysis we limited the number of ascertained pairs to 100 in order to simulate a realistic sample size for a rare disease. Expanding the size of the direct ascertainment window back to 1950 or 1880 (but still not adding indirect ascertainment) decreased the strength of the false signals quantitatively, but did not reliably eliminate them, even when the window was longer than a human lifespan (Table 1, Simulations 1-3).

Next we considered including some indirectly ascertained individuals. To simulate retrospective ascertainment, we first directly ascertained individuals with onset within the window, then ascertained any parents or children thereof whose onset had already occurred prior to the beginning of the window. We hypothesized that it might be more difficult to ascertain individuals who died long ago, so for indirectly ascertained individuals we applied a linear reduction in ascertainment probability according to how many years prior to the beginning of the ascertainment window the individual had disease onset. In Simulation 4 (Table 1), for instance, direct ascertainment from 1989-2013 is exhaustive, but the probability of indirect ascertainment declines by 5% for each year prior to 1989 that an individual has onset so that, for instance, an individual with onset in 1988 can be retrospectively ascertained with 95% probability, in 1987 with 90% probability, and so on. Including some indirectly ascertained individuals in this manner further reduced the magnitude of

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

anticipation but did not completely eliminate it in most scenarios tested (Simulations 4-9). The false signal of heritability was weaker than that of anticipation and could be reliably reduced to statistical insignificance in the most extreme scenarios tested (Simulations 6-9). Only when indirect ascertainment was 100% exhaustive and independent of the year of onset (Simulation 10) did a signal of anticipation cease to be detected, though a year of birth – age of onset correlation still existed.

Sim-ul-ation	Direct ascertain-ment window	Indirect ascertain-ment	Mean anticip-ation (years)	Anticip-ation significant [b]	Mean additive heritab-ility	Heritab-ility significant [b]	Mean year of birth – age of onset slope	Slope signifi-cant
1	1989-2013	None	16.4	100%	93%	100%	-0.67	100%
2	1950-2013	None	4.9	98%	19%	16%	-0.22	100%
3	1880-2013	None	1.6	23%	3%	5%	-0.06	100%
4	1989-2013	Declining 5%/year	10.7	100%	59%	83%	-0.55	100%
5	1950-2013	Declining 5%/year	3.4	79%	14%	11%	-0.18	100%
6	1880-2013	Declining 5%/year	1.4	18%	3%	6%	-0.05	98%
7	1989-2013	Declining 1%/year	2.0	33%	6%	5%	-0.26	100%
8	1950-2013	Declining 1%/year	0.9	12%	4%	5%	-0.12	100%
9	1880-2013	Declining 1%/year	0.4	6%	1%	4%	-0.04	85%
10	1989-2013	Exhaustive	0.2	5%	-1%	5%	-0.20	99%

Table 1. Simulations of ascertainment bias. Ascertaining only those individuals with disease onset within a historical window creates false signals of anticipation and heritability (Simulation 1). These false signals are reduced but not eliminated by expanding the ascertainment window (Simulations 2-3). Probabilistic retrospective ascertainment (see text) reduces these signals further (Simulations 4-9). Only when retrospective ascertainment is 100% exhaustive does anticipation become reliably non-significant (Simulation 10). [a]Averages from 1000 simulations. [b]Percentage of 1000 simulations in which this figure was statistically significant at p < .05. Figures which were significant no more often than expected by chance are shown in gray.

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

Together our simulations demonstrate that ascertaining individuals over a larger time window or including some indirect ascertainment (retrospective phenotyping) can quantitatively reduce the false signals of anticipation and heritability, but that these false signals are still likely to be observed as statistically significant under anything other than almost complete ascertainment. Our results indicate that whenever retrospective ascertainment is less than exhaustive, there is a risk of observing false signals of anticipation and heritability of age of onset.

Stratification of simulated data. One recent report of anticipation in Glu200Lys genetic prion disease[13] argued that the observed anticipation must be real because anticipation was observed, using paired t tests, within every subset of 26 parent-child pairs when these were stratified by age of death, year of birth or year of death. To determine whether such an analysis could indeed rule out a false positive due to ascertainment bias, we stratified 26 simulated parent-child pairs with onset between 1989 and 2013 in the same way described in the previous report (Table 2). In our simulation, anticipation is not real, yet a strong signal of anticipation is nevertheless observed in all eight strata. Therefore, these methods of stratification do noteliminate spurious signals of anticipation due to ascertainment bias.

Cohort name	Inclusion criteria	Anticipation	p
Child early birth year	Child born < 1939	16 years	.009
Child late birth year	Child born >= 1939	18 years	< .0001
Child early death year	Child year of onset < 2000	26 years	.0004
Child late death year	Child year of onset >= 2000	15 years	< .0001
Child early death age	Child age onset < 61	21 years	< .0001
Child late death age	Child age onset >= 61	11 years	.003
Parent early death age	Parent age onset < 70	12 years	.01
Parent late death age	Parent age onset >= 70	20 years	< .0001

Table 2. Stratification of simulated data. In n = 26 simulated parent-child pairs with independent and identically distributed ages of onset but ascertaining only those pairs whose onsets both fall within 1989-2013, all subsets of data stratified according to the variables defined in Table 3 of Pocchiari et al.[13] still show significant anticipation.

Characteristics of Glu200Lys pedigree data. We combined data on Glu200Lys families from four independent study centers worldwide (Table 3). Each individual dataset reflected a different method of ascertainment (see Materials and Methods) and accordingly, the proportion of individuals ascertained indirectly or asymptomatic as of last follow-up varied considerably (Table 3).

Of individuals whose disease onset or death was directly observed by one of the study centers, 65% had no reported family history of prion disease. This figure may reflect some combination of (1) incomplete reporting of family history, (2) under-diagnosis of affected individuals in earlier generations, (3) censoring of asymptomatic individuals with the mutation upon death due to other causes, and (4) *de novo* mutations. Similarly, we knew the genotypes of only 22% of at-risk individuals (see Methods) in our datasets. Together, these figures suggest that neither retrospective ascertainment nor prospective following of individuals with the mutation are exhaustive in our dataset.

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

Study center	In operation since	n c.598G>A individuals	n c.598G>A parent-child pairs	Proportion of c. 598G>A individuals ascertained indirectly	Proportion of c.598G>A individuals who are asymptomatic
ANCJDR	1993	24	7	42%	8%
German CJD Surveillance	1993	32	3	38%	3%
MRC Prion Unit	1991	57	29	25%	23%
UCSF	2001	104	58	73%	37%
All		**217**	**97**	**52%**	**25%**

Table 3. Characteristics of Glu200Lys datasets. Data from the four study centers varied in terms of the length of time for which data had been collected, the degree of indirect ascertainment, and the ascertainment of asymptomatic individuals with the mutation. See Methods for details.

Disease duration and genotypic influence. Disease duration was defined as time from first symptom (see Data Collection in Materials and Methods) to death. Disease duration followed a non-normal distribution (p < .0001, Shapiro-Wilk normality test) with a median of 162 days (n = 61, interquartile range 205 days), similar to the figure reported elsewhere[11]. Disease duration did not differ by study center (p = .35, Kruskal-Wallis test, n = 61) or between directly and indirectly ascertained individuals (p = .08, Kruskal-Wallis test, n = 38 and 23 respectively). *PRNP* codon 129 information was available for a subset of individuals with known disease duration. Among individuals with a haplotype encoding Glu200Lys cis 129Met, disease duration was shorter for individuals with a trans allele encoding 129Met (median of 137 days, n = 25) than a trans allele encoding 129Val (median of 426 days, n = 7) and this difference was significant (p = .02, Kolmogorov-Smirnov test), consistent with previously reported data[11]. Disease duration appeared to differ between haplotypes encoding Glu200Lys cis 129Met (n = 42) and Glu200Lys cis 129Val (n = 6) proteins, with a median duration of 137 vs. 331 days respectively (p = .04, Kolmogorov-Smirnov test), although trans codon 129 information was not available for all of these individuals, so it is possible that the longer duration among cis 129Val individuals might be due to higher rates of codon 129 heterozygosity. Because disease duration was generally a year or less and our data included age of onset for some individuals but age of death for others, we decided to consider age of onset and death interchangeably in the anticipation analysis, preferring age of death when both were available.

Age of onset or death and genotypic influence. Overall, the mean age of onset or death in affected individuals was 62 ± 10 years (± sd, n = 158) and was approximately normally distributed

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

(p = .36, Shapiro-Wilk normality test; Figure 2A). The median observed age of onset or death was 63 years (n = 158). The median rose to 64 years when we included asymptomatic individuals with the mutation, censored at date of last follow-up or death due to other causes, and applied a survival analysis (n = 207). This survival analysis indicated 93% disease penetrance by age 80, within the range of previously reported estimates,[8,9] but we note that this figure is biased upward by the ascertainment of affected families and incomplete genotyping of unaffected individuals . 64 years is, to our knowledge, older than any other measure of central tendency of age of onset for this mutation yet reported, probably because our use of survival analysis accounts for censoring due to competing risks, whereas other published estimates have been based simply on observed ages of onset[3,7,10,47]. Even our estimate is likely biased downward due to incomplete genotyping of unaffected individuals.

Age of onset or death did not differ by study center (p = .18, Model I ANOVA, n = 158), sex (p = .74, unpaired t test, n = 132), or mode of ascertainment (direct vs. indirect, p = .66, unpaired t test, n = 158). Among individuals with a haplotype encoding Glu200Lys cis 129Met, age of onset or death did not differ between trans 129Met (n = 46) and trans 129Val (n = 10) genotypes (p = .71, 2-tailed unpaired t test; p = .39, log-rank test, n = 64 Met-Met vs. 17 Met-Val including censored individuals), consistent with a previous report[48]. It was unclear whether age of onset or death might differ between individuals with haplotypes encoding Glu200Lys cis 129Met versus Glu200Lys cis 129Val, as only a weak trend could be detected whether censored individuals were included (censored median 64 vs. 57 years, p = .05, log-rank test, n = 142 Met vs. 13 Val) or excluded (mean 63 vs. 56 years, p = .11, 2-tailed unpaired t-test, n = 105 Met vs. 10 Val). Because trans codon 129 did not influence age of onset or death and because a child's cis codon 129 is inherited from the affected parent, we deemed it unnecessary to consider codon 129 genotype in our anticipation analyses.

Ascertainment bias in Glu200Lys pedigree data. Excluding asymptomatic individuals, the year of disease onset or death was known for 94% of directly ascertained individuals (68 of 72) and 55% of indirectly ascertained individuals (50 of 91). Year of disease onset or death occurred between 1989 and 2013 in 92% of all individuals for whom this variable was known (Figure 2B). This 25-year window was shorter than the typical difference between parent and child year of birth, which was 28 ± 6 years (mean ± sd, n = 150 including unaffected individuals), thus partially limiting ascertainment to parent-child pairs in which the child died at least 3 years younger than the parent. For individuals ascertained directly, year of birth and age of onset were even more strongly correlated than in our simulation (slope = -.79, p < .0001, linear regression, Figure 2C, black line). The correlation weakened only slightly when indirectly ascertained individuals were included as well (slope = -.48, p < .0001, linear regression, Figure 2C, dashed line).

Our simulation (see above) indicated that ascertainment bias could introduce false signals of heritability. We did not, however, observe any correlation between parent and child age of onset (p = .99, linear regression, n = 39 pairs) nor in sibling pair ages of onset (p = .63, linear regression, n = 26 pairs). Thus we do not find any signal, whether real or artifactual, that age of onset in this disease is heritable. This is not surprising because our simulation indicated that false signals of heritability are weaker than false signals of anticipation (Table 1), and the small size of our dataset may also leave us underpowered to detect any true heritability of age of onset which may exist.

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.
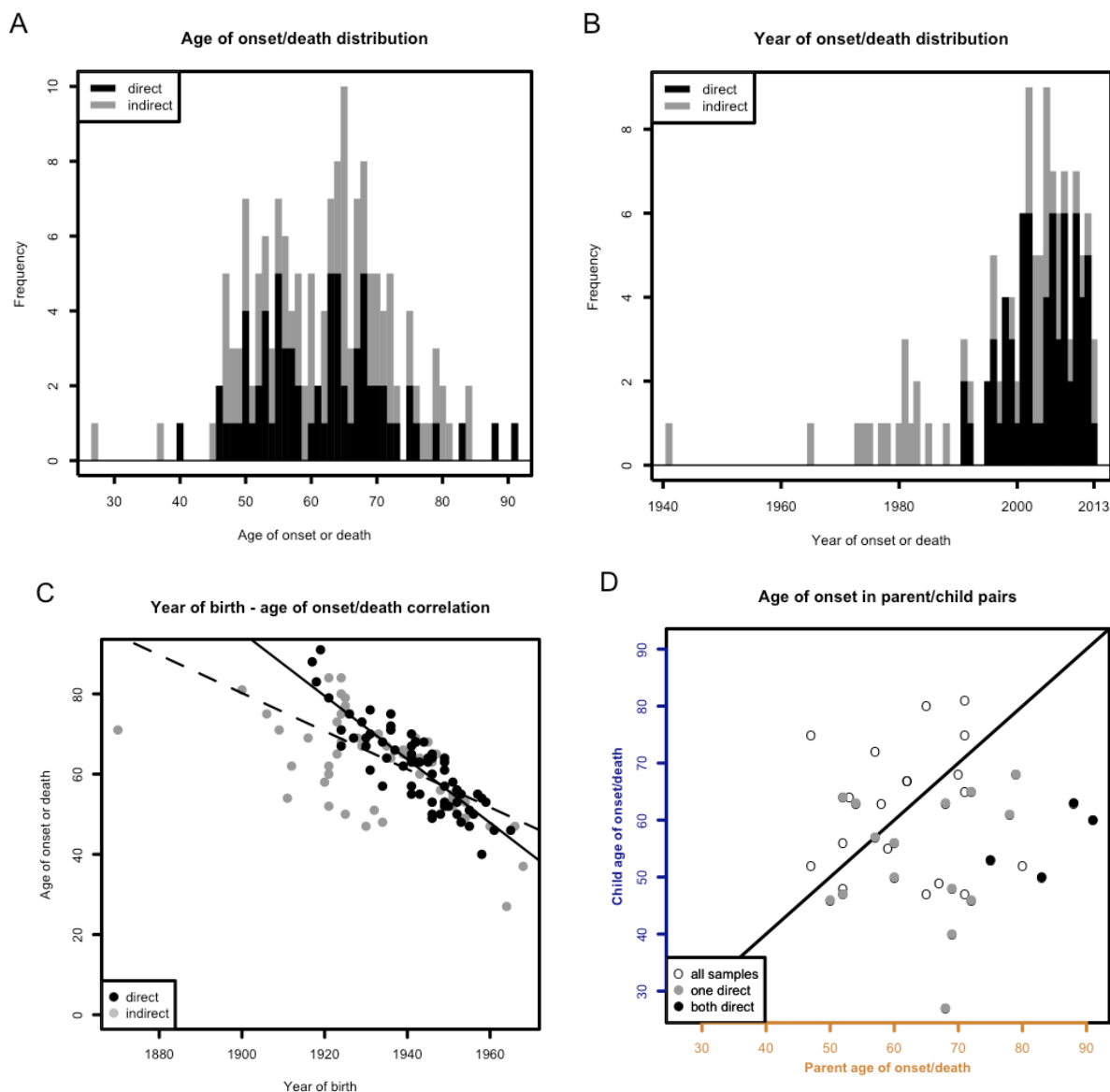

Figure 2. Ascertainment bias and anticipation in Glu200Lys pedigree data. (A) Distribution of age of onset or death by ascertainment mode. (B) Years of onset or death, when known, occur overwhelmingly in the 25-year window since the mutation's discovery in 1989. (C) Year of birth and age of onset are artifactually correlated. This correlation is strongest in directly ascertained individuals (black line) but remains highly significant when indirectly ascertained individuals (gray points) are included in addition (dashed black line). (D) Parent versus child age of onset or death. Points below the diagonal line indicate pairs in which the child dies younger than the parent. Directly ascertained pairs (black points) fall categorically below the line, pairs with one directly ascertained member are largely below the line (gray points), and pairs in which both individuals are indirectly ascertained fall on both sides (white points).

14

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

Strength of anticipation signal depends upon degree of ascertainment bias. Because signatures of ascertainment bias were strongly evident in our data, we expected to see a difference between parent and child age of onset in observed pairs. Based on our simulation results (Table 1), we hypothesized that the strength of the anticipation signal from naïve paired t tests would vary depending on how flexible we were able to make our ascertainment window, within the constraints of our available data. When we considered only pairs in which both parent and child were ascertained directly, we observed 28 years of anticipation (p = .002, 2-sided paired t-test, n = 4 pairs). When we additionally included pairs with one indirectly ascertained individual, this figure dropped to 15 years (p = .0002, 2-sided paired t-test, n = 20 pairs). When we also included pairs in which both individuals were ascertained indirectly, this dropped further to 7 years (p = .005, 2-sided paired t-test, n = 39 pairs). Thus, the number of years of anticipation observed depends upon how narrow the ascertainment window is and the mechanism of case reporting.

The strength of the anticipation signal also differed by study center according to the extent of ascertainment bias in each individual dataset. The UCSF dataset, which was the largest and had both the highest percentage of indirectly ascertained individuals (73%) and the least correlation between year of birth and age of onset or death (slope = -.41), showed the least evidence of anticipation (1 year, p = .68, 2-sided paired t-test). The MRC Prion Unit showed a marginal paired difference (7 years, p = .10, 2-sided paired t test), whereas the Australian and German cohorts showed larger differences (18 years, p = .05 and 31 years, p = .03 respectively, 2-sided paired t test).

Application of survival analysis to Glu200Lys anticipation. The foregoing analysis indicated that our data exhibited a nominal signal of anticipation when analyzed with naïve paired t tests. Because this signal depended upon the degree of ascertainment bias, and because our data exhibited signatures of ascertainment bias similar to what we had observed in our simulation, we suspected that the anticipation we observed was an artifact. We next set out to determine whether a survival analysis including asymptomatic individuals with the c.598G>A mutation would correct this artifact. Based on our simulation results (Figure 1E) and the fact that our ascertainment of asymptomatic individuals with the mutation was not exhaustive, we expected that survival analysis would not fully eliminate the signal of anticipation in our Glu200Lys data. For the survival analysis, we considered all parent-child pairs in which both individuals harbored the mutation, regardless of censored status. To ensure independence of tested pairs and avoid multiple-counting, we randomly selected one such pair from each pedigree and compared parent and child survival curves. Across 1000 iterations, parent and child survival curve medians differed by an average of 7 years (log-rank test p < .05 in 82% of iterations). Because our simulation indicated that more thorough ascertainment of censored individuals reduces artifactual anticipation signals (Figure 1E), we repeated the same analysis on only the UCSF and MRC Prion Unit cohorts, which had the highest proportion of asymptomatic individuals with the mutation (Table 3). Within these cohorts, the median difference in parent and child survival curves was only 1 year and was usually not significant (p < .05 in only 8% of 1000 iterations). This is consistent with the hypothesis that observed parent-child age of onset differences in Glu200Lys prion disease are due to incomplete ascertainment and not due to true anticipation.

Anticipation in age at death due to other, non-prion, causes. The foregoing analyses suggested that the anticipation we observed in Glu200Lys families could be attributed entirely to ascertainment bias. We reasoned that if this was true, then anticipation in age of death might be observed in deaths attributable to other causes as well, with these unrelated deaths providing a kind of negative control. When we compared age at death for all parent-child pairs in our dataset in which neither individual had CJD listed as cause of death, we observed 14 years of anticipation (p =

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

.002, 2-sided paired t-test, n = 37 pairs). In many cases the cause of death was not listed or was ambiguous, so we could not rule out the possibility that some of those individuals may actually have died of undiagnosed or unreported prion disease. When we therefore considered only those pairs in which each individual was known not to harbor the mutation (see Methods) and/or died of a known cause clearly distinct from prion disease (cancer, heart attack, accident, etc.), the anticipation grew to 35 years (p = .001, 2-sided paired t-test, n = 9 pairs). This set included some individuals who had died much younger than the typical age of onset for Glu200Lys prion disease (Figure 2A), introducing a potential new source of ascertainment bias as individuals who die young may be counted as children but are unlikely to ever become parents[14]. When we further filtered our data to include only individuals with age at death greater than 40, 16 years of anticipation were still observed (p = .002, 2-sided paired t-test, n = 5 pairs).

DISCUSSION

The prediction of age of onset in those carrying mutations causal of neurodegenerative diseases is important for genetic counseling, clinical trial design,[49,50] and understanding of fundamental disease mechanisms. We chose to investigate bias in a specific genetic prion disease as it has been reported[12,13] that Glu200Lys genetic prion disease exhibits anticipation, with children succumbing to disease 7 to 14 years younger than their parents on average. We approached this claim with skepticism; the causal mutation is not a repeat expansion and human prion strains are not transmitted in childbirth or through casual contact between people[51–53], leaving no obvious mechanism for anticipation. Such a large decrease in age of onset in germline transmission ought to lead to juvenile onset cases within a few generations, as observed in repeat expansion disorders[15], yet no such cases are observed in Glu200Lys families (Figure 2A). In the present study, we considered the possibility that reported evidence of anticipation in Glu200Lys prion disease arises instead from ascertainment bias.

We tested for anticipation in our data using naïve paired t tests and we did observe a statistically significant difference between parent and child age of onset, but four separate lines of evidence argue that this difference is artifactual rather than biological. First, 92% of known years of onset or death for CJD-affected individuals in our dataset fell within the 25-year window from the mutation's discovery in 1989 to the time of our study in 2013, whereas parents were born on average 28 years earlier than their children. This enriches for pairs in which children die at least 3 years younger than their affected parent. Pairs in which children die older than their parents are much less likely to be observed during this time window. Second, we observed an artifactual correlation between year of birth and age of onset reflective of this "windowing" effect because individuals born later in time are only captured in this dataset if they have earlier onset. The distribution of ages of onset is therefore shifted older for a parent's year of birth (for instance, 1920) than for their child's year of birth (for instance, 1950). Third, the strength of the anticipation signal we observed depended upon how strictly the data were limited by the ascertainment window. When we required that both parent and child had to be directly observed as patients by the study centers contributing data, we observed 28 years of anticipation. When all indirectly ascertained individuals were also included this figure dropped to 7 years, and when we performed survival analysis on only the data from two centers most active in performing predictive genetic testing, we saw no difference in age of onset at all. Fourth, when using naïve paired t tests we observed anticipation even among individuals in our pedigrees who did not harbor the mutation and/or died of causes unrelated to prion disease.

None of the three previous studies reporting Glu200Lys anticipation[12,13] (see also Web Resources) presented data on the distribution of year of onset, tested for correlation between year of birth and

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

age of onset, or carried out a negative control by assessing anticipation in unrelated deaths as we have done here. One of these studies[12] found a stronger anticipation (14 years rather than 7) when only directly genotyped individuals were included, similar to our finding that the strength of anticipation depends upon the duration of the ascertainment window. We showed that the stratification analyses, as performed in one study,[13] do not remove the effects of ascertainment bias. Because anticipation was observed in certain analyses of our data as well, we consider it likely that the differences between our study and the previous studies arise from differences in methodology rather than differences in whether anticipation is truly present in the study population.

In assessing the evidence for anticipation in Glu200Lys prion disease, we have created a simulation model with relevance to other diseases as well. In contrast to previous studies of anticipation, we consider ascertainment to be limited by year of onset, not by age of onset per se. In general we find that ascertaining only those individuals with onset within a restricted time period is sufficient to create robust false signals of anticipation. This is of general concern because the genetic causes of many dominant Mendelian disorders were identified within the past 25 years, leaving a relatively narrow window for ascertainment. Our simulation shows that even high rates of retrospective ascertainment are unlikely to completely remove the resulting false signal of anticipation. Although certain statistical methods can help to correct for this bias when ascertainment criteria are consistent or datasets large enough to allow subsetting, we believe our own data may be typical of some rare diseases in representing a mixture of ascertainment modes, with both retrospective and prospective phenotyping present but non-exhaustive. For such datasets, we agree with the view[54] that statistical tests alone may be inadequate to discriminate between situations with and without anticipation. We believe that one route towards preventing spurious claims of anticipation such as seen in Glu200Lys genetic prion disease may lie not in the requirement of more rigorous statistical tests but in an expectation of thorough and transparent assessment of a dataset's degree of bias. Toward that end, we propose that future studies reporting factors that determine age of onset in adult-onset dominant conditions should be expected to provide (1) a histogram of year of onset or death, (2) a test for correlation between year of birth and age of onset, (3) descriptive statistics on the extent of retrospective ascertainment and predictive genetic testing or prospective follow-up, and (4) a test for anticipation in deaths of other causes. Our results also have implications for study design in adult onset genetic diseases; year of birth information should be collected, asymptomatic individuals with a mutation should be tracked, and retrospective ascertainment should be as thorough as possible. These measures will make it easier to quantify and reduce ascertainment bias.

Though heritability of age of onset was not observed in our Glu200Lys data, we noticed in our simulation that windowed ascertainment can also create a false signal that age of onset is heritable. This false signal could easily be disproven by including child year of birth as a covariate in parent-offspring regression. Our result suggests that estimates of the heritability of age of onset in genetic disease made based on parent-offspring regression ought to be scrutinized carefully and accepted only if the parent-child age of onset correlation remains after controlling for year of birth.

In summary, the phenomenon of anticipation previously reported in Glu200Lys prion disease is likely an artifact. Ascertainment bias is a pervasive problem in age of onset in genetic disease but can be reduced through appropriate data collection methods and recognized with simple analytical tools.

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

WEB RESOURCES

Online Mendelian Inheritance in Man (OMIM): http://www.omim.org

Mitrova, E. (2012). The role of anticipation in the incidence of CJD with the mutation E200K. Presented at the CJD Family Foundation Conference, Washington, D.C., July 14, 2012. Archived on August 5, 2014: http://web.archive.org/web/20140805095200/http://www.cjdfoundation.org/shared/cjdfoundation/files/webfm/admin/2012-conference/Eva-Mitrova.pdf

U.S. Social Security Administration 2009 Life Tables. Archived on June 2, 2014: http://web.archive.org/web/20140602090427/http://www.ssa.gov/oact/STATS/table4c6.html

Source code and output for this study: https://github.com/ericminikel/e200k-anticipation

REFERENCES

1. Prusiner, S.B. (1998). Prions. Proc. Natl. Acad. Sci. U. S. A. *95*, 13363–13383.

2. Holman, R.C., Belay, E.D., Christensen, K.Y., Maddox, R.A., Minino, A.M., Folkema, A.M., Haberling, D.L., Hammett, T.A., Kochanek, K.D., Sejvar, J.J., et al. (2010). Human prion diseases in the United States. PloS One *5*, e8521.

3. Kong, Q., Surewicz, W.K., Petersen, R.B., Chen, S.G., Gambetti, P., Parchi, P., Capellari, S., Goldfarb, L., Montagna, P., Lugaresi, E., et al. (2004). Inherited Prion Diseases. In Prion Biology and Diseases, (Cold Spring Harbor Laboratory Press),.

4. Goldgaber, D., Goldfarb, L.G., Brown, P., Asher, D.M., Brown, W.T., Lin, S., Teener, J.W., Feinstone, S.M., Rubenstein, R., and Kascsak, R.J. (1989). Mutations in familial Creutzfeldt-Jakob disease and Gerstmann-Sträussler-Scheinker's syndrome. Exp. Neurol. *106*, 204–206.

5. Goldfarb, L.G., Korczyn, A.D., Brown, P., Chapman, J., and Gajdusek, D.C. (1990). Mutation in codon 200 of scrapie amyloid precursor gene linked to Creutzfeldt-Jakob disease in Sephardic Jews of Libyan and non-Libyan origin. Lancet *336*, 637–638.

6. Goldfarb, L.G., Mitrová, E., Brown, P., Toh, B.K., and Gajdusek, D.C. (1990). Mutation in codon 200 of scrapie amyloid protein gene in two clusters of Creutzfeldt-Jakob disease in Slovakia. Lancet *336*, 514–515.

7. Hsiao, K., Meiner, Z., Kahana, E., Cass, C., Kahana, I., Avrahami, D., Scarlato, G., Abramsky, O., Prusiner, S.B., and Gabizon, R. (1991). Mutation of the prion protein in Libyan Jews with Creutzfeldt-Jakob disease. N. Engl. J. Med. *324*, 1091–1097.

8. Chapman, J., Ben-Israel, J., Goldhammer, Y., and Korczyn, A.D. (1994). The risk of developing Creutzfeldt-Jakob disease in subjects with the PRNP gene codon 200 point mutation. Neurology *44*, 1683–1686.

9. Spudich, S., Mastrianni, J.A., Wrensch, M., Gabizon, R., Meiner, Z., Kahana, I., Rosenmann, H., Kahana, E., and Prusiner, S.B. (1995). Complete penetrance of Creutzfeldt-Jakob disease in Libyan Jews carrying the E200K mutation in the prion protein gene. Mol. Med. Camb. Mass *1*, 607–613.

10. Schelzke, G., Kretzschmar, H.A., and Zerr, I. (2012). Clinical aspects of common genetic Creutzfeldt-Jakob disease. Eur. J. Epidemiol. *27*, 147–149.

11. Pocchiari, M., Puopolo, M., Croes, E.A., Budka, H., Gelpi, E., Collins, S., Lewis, V., Sutcliffe, T., Guilivi, A., Delasnerie-Laupretre, N., et al. (2004). Predictors of survival in sporadic Creutzfeldt-Jakob disease and other human transmissible spongiform encephalopathies. Brain J. Neurol. *127*, 2348–2359.

12. Rosenmann, H., Kahana, E., Korczyn, A.D., Kahana, I., Chapman, J., and Gabizon, R. (1999). Preliminary evidence for anticipation in genetic E200K Creutzfeldt-Jakob disease. Neurology *53*, 1328–1329.

13. Pocchiari, M., Poleggi, A., Puopolo, M., D'Alessandro, M., Tiple, D., and Ladogana, A. (2013). Age at Death of Creutzfeldt-Jakob Disease in Subsequent Family Generation Carrying the E200K Mutation of the Prion Protein Gene. PloS One *8*, e60376.

14. Penrose, L.S. (1948). The problem of anticipation in pedigrees of dystrophia myotonica. Ann. Eugen. *14*, 125–132.

15. Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. Annu. Rev. Neurosci. *30*, 575–621.

16. La Spada, A.R., and Taylor, J.P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat. Rev. Genet. *11*, 247–258.

17. Vulliamy, T., Marrone, A., Szydlo, R., Walne, A., Mason, P.J., and Dokal, I. (2004). Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. Nat. Genet. *36*, 447–449.

18. Martinez-Delgado, B., Yanowsky, K., Inglada-Perez, L., Domingo, S., Urioste, M., Osorio, A., and Benitez, J. (2011). Genetic anticipation is associated with telomere shortening in hereditary breast cancer. PLoS Genet. *7*, e1002182.

19. Gajdusek, D.C. (1977). Unconventional viruses and the origin and disappearance of kuru. Science *197*, 943–960.

20. Brown, P., Preece, M.A., and Will, R.G. (1992). "Friendly fire" in medicine: hormones, homografts, and Creutzfeldt-Jakob disease. Lancet *340*, 24–27.

21. Hewitt, P.E., Llewelyn, C.A., Mackenzie, J., and Will, R.G. (2006). Creutzfeldt-Jakob disease and blood transfusion: results of the UK Transfusion Medicine Epidemiological Review study. Vox Sang. *91*, 221–230.

22. Heiman, G.A., Hodge, S.E., Wickramaratne, P., and Hsu, H. (1996). Age-at-interview bias in anticipation studies: computer simulations and an example with panic disorder. Psychiatr. Genet. *6*, 61–66.

23. Vieland, V.J., and Huang, J. (1998). Statistical evaluation of age-at-onset anticipation: a new test and evaluation of its behavior in realistic applications. Am. J. Hum. Genet. *62*, 1212–1227.

24. Tsai, W.Y., Heiman, G.A., and Hodge, S.E. (2005). New simple tests for age-at-onset anticipation: application to panic disorder. Genet. Epidemiol. *28*, 256–260.

25. Boonstra, P.S., Gruber, S.B., Raymond, V.M., Huang, S.-C., Timshel, S., Nilbert, M., and Mukherjee, B. (2010). A review of statistical methods for testing genetic anticipation: looking for an answer in Lynch syndrome. Genet. Epidemiol. *34*, 756–768.

26. Rabinowitz, D., and Yang, Q. (1999). Testing for age-at-onset anticipation with affected parent-child pairs. Biometrics *55*, 834–838.

27. Collins, S., Law, M.G., Fletcher, A., Boyd, A., Kaldor, J., and Masters, C.L. (1999). Surgical treatment and risk of sporadic Creutzfeldt-Jakob disease: a case-control study. Lancet *353*, 693–697.

28. Collins, S., Boyd, A., Lee, J.S., Lewis, V., Fletcher, A., McLean, C.A., Law, M., Kaldor, J., Smith, M.J., and Masters, C.L. (2002). Creutzfeldt-Jakob disease in Australia 1970-1999. Neurology *59*, 1365–1371.

29. Windl, O., Giese, A., Schulz-Schaeffer, W., Zerr, I., Skworc, K., Arendt, S., Oberdieck, C., Bodemer, M., Poser, S., and Kretzschmar, H.A. (1999). Molecular genetics of human prion diseases in Germany. Hum. Genet. *105*, 244–252.

30. Grasbon-Frodl, E., Lorenz, H., Mann, U., Nitsch, R.M., Windl, O., and Kretzschmar, H.A. (2004). Loss of glycosylation associated with the T183A mutation in human prion disease. Acta Neuropathol. (Berl.) *108*, 476–484.

31. Zerr, I., Kallenberg, K., Summers, D.M., Romero, C., Taratuto, A., Heinemann, U., Breithaupt, M., Varges, D., Meissner, B., Ladogana, A., et al. (2009). Updated clinical diagnostic criteria for sporadic Creutzfeldt-Jakob disease. Brain J. Neurol. *132*, 2659–2668.

32. Beck, J.A., Poulter, M., Campbell, T.A., Adamson, G., Uphill, J.B., Guerreiro, R., Jackson, G.S., Stevens, J.C., Manji, H., Collinge, J., et al. (2010). PRNP allelic series from 19 years of prion protein gene sequencing at the MRC Prion Unit. Hum. Mutat. *31*, E1551–1563.

33. Geschwind, M.D., Haman, A., and Miller, B.L. (2007). Rapidly progressive dementia. Neurol. Clin. *25*, 783–807, vii.

34. Geschwind, M.D., Shu, H., Haman, A., Sejvar, J.J., and Miller, B.L. (2008). Rapidly progressive dementia. Ann. Neurol. *64*, 97–108.

35. Paterson, R.W., Torres-Chae, C.C., Kuo, A.L., Ando, T., Nguyen, E.A., Wong, K., Dearmond, S.J., Haman, A., Garcia, P., Johnson, D.Y., et al. (2012). Differential diagnosis of jakob-creutzfeldt disease. Arch. Neurol. *69*, 1578–1582.

36. Parchi, P., Giese, A., Capellari, S., Brown, P., Schulz-Schaeffer, W., Windl, O., Zerr, I., Budka, H., Kopp, N., Piccardo, P., et al. (1999). Classification of sporadic Creutzfeldt-Jakob disease based on molecular and phenotypic analysis of 300 subjects. Ann. Neurol. *46*, 224–233.

37. Rabinovici, G.D., Wang, P.N., Levin, J., Cook, L., Pravdin, M., Davis, J., DeArmond, S.J., Barbaro, N.M., Martindale, J., Miller, B.L., et al. (2006). First symptom in sporadic Creutzfeldt-Jakob disease. Neurology *66*, 286–287.

38. Squitieri, F., Sabbadini, G., Mandich, P., Gellera, C., Di Maria, E., Bellone, E., Castellotti, B., Nargi, E., de Grazia, U., Frontali, M., et al. (2000). Family and molecular data for a fine analysis of age at onset in Huntington disease. Am. J. Med. Genet. *95*, 366–373.

39. Visscher, P.M., Hill, W.G., and Wray, N.R. (2008). Heritability in the genomics era--concepts and misconceptions. Nat. Rev. Genet. *9*, 255–266.

40. Owen, J., Beck, J., Campbell, T., Adamson, G., Gorham, M., Thompson, A., Smithson, S., Rosser, E., Rudge, P., Collinge, J., et al. (2014). Predictive testing for inherited prion disease: report of 22 years experience. Eur. J. Hum. Genet. EJHG.

41. Goizet, C., Lesca, G., Dürr, A., and French Group for Presymptomatic Testing in Neurogenetic Disorders (2002). Presymptomatic testing in Huntington's disease and autosomal dominant cerebellar ataxias. Neurology *59*, 1330–1336.

42. Laccone, F., Engel, U., Holinski-Feder, E., Weigell-Weber, M., Marczinek, K., Nolte, D., Morris-Rosendahl, D.J., Zühlke, C., Fuchs, K., Weirich-Schwaiger, H., et al. (1999). DNA analysis of Huntington's disease: five years of experience in Germany, Austria, and Switzerland. Neurology *53*, 801–806.

43. Harper, P.S., Lim, C., and Craufurd, D. (2000). Ten years of presymptomatic testing for Huntington's disease: the experience of the UK Huntington's Disease Prediction Consortium. J. Med. Genet. *37*, 567–571.

44. Tassicker, R.J., Marshall, P.K., Liebeck, T.A., Keville, M.A., Singaram, B.M., and Richards, F.H. (2006). Predictive and pre-natal testing for Huntington Disease in Australia: results and challenges encountered during a 10-year period (1994-2003). Clin. Genet. *70*, 480–489.

45. Morrison, P.J., Harding-Lester, S., and Bradley, A. (2011). Uptake of Huntington disease predictive testing in a complete population. Clin. Genet. *80*, 281–286.

46. Baker, H.E., Poulter, M., Crow, T.J., Frith, C.D., Lofthouse, R., and Ridley, R.M. (1991). Aminoacid polymorphism in human prion protein and age at death in inherited prion disease. Lancet *337*, 1286.

47. Kovács, G.G., Puopolo, M., Ladogana, A., Pocchiari, M., Budka, H., van Duijn, C., Collins, S.J., Boyd, A., Giulivi, A., Coulthart, M., et al. (2005). Genetic prion disease: the EUROCJD experience. Hum. Genet. *118*, 166–174.

48. Gabizon, R., Rosenmann, H., Meiner, Z., Kahana, I., Kahana, E., Shugart, Y., Ott, J., and Prusiner, S.B. (1993). Mutation and polymorphism of the prion protein gene in Libyan Jews with Creutzfeldt-Jakob disease (CJD). Am. J. Hum. Genet. *53*, 828–835.

49. Ryman, D.C., Acosta-Baena, N., Aisen, P.S., Bird, T., Danek, A., Fox, N.C., Goate, A., Frommelt, P., Ghetti, B., Langbaum, J.B.S., et al. (2014). Symptom onset in autosomal dominant Alzheimer disease: A systematic review and meta-analysis. Neurology.

50. Bateman, R.J., Xiong, C., Benzinger, T.L.S., Fagan, A.M., Goate, A., Fox, N.C., Marcus, D.S., Cairns, N.J., Xie, X., Blazey, T.M., et al. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N. Engl. J. Med. *367*, 795–804.

51. Belay, E.D. (1999). Transmissible spongiform encephalopathies in humans. Annu. Rev. Microbiol. *53*, 283–314.

52. Xiao, X., Miravalle, L., Yuan, J., McGeehan, J., Dong, Z., Wyza, R., MacLennan, G.T., Golichowski, A.M., Kneale, G., King, N., et al. (2009). Failure to detect the presence of prions in the uterine and gestational tissues from a Gravida with Creutzfeldt-Jakob disease. Am. J. Pathol. *174*, 1602–1608.

53. Murray, K., Peters, J., Stellitano, L., Winstone, A.M., Verity, C., and Will, R.G. (2011). Is there evidence of vertical transmission of variant Creutzfeldt-Jakob disease? J. Neurol. Neurosurg. Psychiatry *82*, 729–731.

54. Hodge, S.E., and Wickramaratne, P. (1995). Statistical pitfalls in detecting age-of-onset anticipation: the role of correlation in studying anticipation and detecting ascertainment bias. Psychiatr. Genet. *5*, 43–47.

Minikel et al. 2014. Ascertainment bias causes false signal of anticipation in genetic prion disease.

<u>Supplemental table</u>

Table S1. Correcting for ascertainment bias by including year of birth in parent-offspring regression reduces, but does not eliminate, true signals of heritability.

In order to determine whether our proposed method of including year of birth in parent-offspring regression models to eliminate false signals of heritability would result in false negatives when heritability is truly present, simulations 1-10 from Table 1 were re-run with the same ascertainment conditions but with a different formula for child age of onset. Here, child age of onset was computed as the unweighted mean of (1) a random variable drawn from the original age of onset distribution described in Methods, and (2) parent age of onset. This simulates a scenario in which age of onset is 100% additively heritable – the affected parent contributes half of the child's genome and half of the child's age of onset; the random portion of the child's age of onset represents the contribution from the unaffected parent's genome. Under these conditions, parent-offspring regression without year of birth correctly detects ~100% heritability under all conditions tested. Including year of birth in the model reduces this estimate, particularly in the simulations where ascertainment bias is most extreme (e.g. Simulations 1-2, 4-5), but the signal of heritability is almost always still significant. We conclude that adding year of birth to these regression models is an imperfect but still useful tool. In cases such as Simulation 1, where ascertainment bias has created a large amount of covariance between year of birth and age of onset, we suspect it may simply not be possible to determine how much of the correlation between parent and child age of onset is artifactual versus biological.

Simulation	Direct ascertainment window	Indirect ascertainment	Mean heritability estimate without year of birth	Significant	Mean heritability estimate with year of birth	Significant
1	1989-2013	None	> 100%	100%	41%	88%
2	1950-2013	None	100%	100%	86%	100%
3	1880-2013	None	98%	100%	96%	100%
4	1989-2013	Declining 5%/year	> 100%	100%	60%	100%
5	1950-2013	Declining 5%/year	99%	100%	90%	100%
6	1880-2013	Declining 5%/year	98%	100%	96%	100%
7	1989-2013	Declining 1%/year	98%	100%	65%	100%
8	1950-2013	Declining 1%/year	98%	100%	91%	100%
9	1880-2013	Declining 1%/year	98%	100%	96%	100%
10	1989-2013	Exhaustive	97%	100%	64%	97%