

**Health Protection Analytical Team
Department of Health
April 2008**

vCJD Prevalence: Resolving the discrepancy in current estimates.

**What can we learn from testing up to 30,000 additional
appendix and tonsil samples?**

Advice sought from the Committee

SEAC is asked to consider the Paper and a suggestion to convene an expert advisory group to consider these issues in greater detail.

Introduction

As of early 2004, estimates for the number of vCJD cases expected from the primary vCJD outbreak (generally supposed to be due to BSE infectivity in the human food chain), ranged from around 250 to something of the order of 600¹. The range corresponded to differing views of the susceptibility of various genotypes to clinical disease. All currently identified clinical cases have been of the MM genotype and the lower figure corresponded to a scenario in which only MM genotypes would ever develop clinical disease. The larger number was based on the assumption that clinical disease would eventually also appear in MV and VV genotypes, assuming a similar a priori probability of infection. On the basis of these expected case numbers, a pessimistic estimate of vCJD prevalence would be about 1/100,000 (600 in a population of 60 million) – in fact rather less, given there had already been approximately 150 deaths.

In 2004, publication of the full results of the retrospective study of stored appendix and tonsil samples (the so called Hilton study) suggested this estimate was in need of major modification. Three positive samples were found in around 11,000 appendix samples. (One positive had already been reported in 2002.) Two of these samples were in VV genotypes. The simplest explanation of this data was a prevalence of around 1/4000, at least in the age cohort tested, with the majority of those infected of all genotypes having sub-clinical infection and possibly acting as infective carriers. This not only suggested that the possibility of secondary infection might be considerably higher than had been previously estimated, it also suggested transmission scenarios in which there was a much greater possibility of a self sustaining secondary epidemic.

The range of credible prevalence scenarios suggested by the Hilton study range from approximately 1/1500 to 1/15000². This again applies to the age range tested, and

¹ Original estimates had been much higher but by 2004 the number of new clinical cases had fallen to around 5 per year, in turn leading to relatively low estimates for future cases.

² Or 1/20,000 depending on the exact statistical definition of 'credible' used.

may overstate prevalence in other age groups. But it also assumes that all those infected with vCJD would be carrying abnormal prion protein at levels detectable by the test: if not, the results would understate true prevalence. It was hoped to decrease uncertainty by obtaining a new estimate based on tonsil samples (the prospective survey, NATA). The frequency of operations to remove tonsils allowed collection of a large number of samples relatively quickly, though not all would come from the most relevant age cohort.

So far around 45,000 tonsil samples have been tested. No positive samples have been found. The maximum likelihood estimate of prevalence from the study is thus 0 and the credible interval up to about 1/20,000.

At first sight, the two estimates seem to give very different results with only a small, if any, overlap of credible prevalence values.

Methods of analysis taking account of the age profile of the samples in the two studies can give, in principle, different results and may be able to resolve the apparent discrepancy. However, such methods also reduce the statistical power of the analysis and therefore reduce the probability of finding discrepant results between the two studies if they exist. Further, such analyses are model dependent³ and this in itself makes an assessment of the results more difficult. The combination of these two factors, suggests that it is premature to dismiss the discrepancy on the basis of the slightly different age ranges/birth cohorts sampled in the two studies.

Further information, which could resolve the apparent discrepancy, may become available from a proposed study of brain and/or spleen tissue taken at general autopsies. However, practical considerations may make such a study unworkable and even if the study is eventually found to be practical, it may take some time to fully resolve the outstanding issues. Alternatively, the NATA study may eventually throw up a number of positives more compatible with the Hilton prevalence estimates, however this is improbable given the lack of positives so far. In the sequel, we will consider other approaches to resolving the apparent discrepancy.

What can we conclude from the apparently discrepant estimates?

One could simply ignore the apparent discrepancy and calculate a best estimate combining the two results. Standard maximum likelihood techniques give a best estimate prevalence of 1/20,000 with a credible interval of around 1/10,000 to 1/100,000. The usefulness of this estimate is however, open to question. One should not combine obviously discrepant results unless one really believes they are, in fact, both genuine estimates of the same quantity. The discrepancy can be considered formally by calculating the probability of obtaining the individual results if the combined estimate were the true prevalence value. The probability for the NATA result is ~10% which is a reasonable probability: the result on the other hand for the Hilton study (including possibility of having obtained even more positive results) is less than 2%. This does not show that the appendix study is necessarily wrong - but it

³ For example, does asymptomatic infection peak in the same age groups/birth cohorts as clinical cases?

does suggest further investigation is required unless we have overriding non-statistical reasons to believe the Hilton study.

A second approach is to accept the discrepancy as indicating that there is a more fundamental problem with one or the other of the two studies than simple random variation. One hypothesis is that the technique used to test the tonsils is less effective than that used on the appendix samples, either because infectivity arises later in tonsil tissue compared to appendix tissue or because the EIA assays used for initial screening in the NATA study were (relatively⁴) insensitive to infectivity.

Further testing

The latter possibility discussed above, is to be investigated by testing 10,000 tonsil samples by the techniques used to test the appendix samples (IHC). If the Hilton prevalence estimate is in fact correct and the infectivity is present in the tonsil samples, there is a 95% probability of finding at least one positive sample amongst the 10,000 tonsil samples (even at 1/10,000 prevalence there is ~2/3 chance of finding at least one positive result).

Finding even a single positive result could then be good evidence that the Hilton study gives a credible estimate of the prevalence whereas the EIA part of the NATA study does not. However, an alternative hypothesis is that the IHC assays are, in fact, throwing up false positives. The actual status of IHC positives can (at least in principle) be confirmed for the tonsil samples by carrying out a variety of other tests. Failure to find a single confirmed positive would suggest *either* that the Hilton estimate is a significant over-estimate of prevalence or that infectivity is not present (detectably and consistently through the incubation period) in tonsils.

These considerations also indicate the need to independently verify the Hilton results and the case for doing so will be strengthened if there are no confirmed positive results from the 10,000 tonsil samples to be tested by IHC. If the purpose is simply to verify the Hilton finding, relatively small numbers of appendix samples would be required compared with the hundred thousand or so required for an accurate prevalence estimate. A study of say 20,000 archived appendices analysed by the methods of the original Hilton study would have a less than 1% probability of finding no results if the 1/4000 estimate is accurate. Testing 30,000 appendices would give a 95% probability of finding at least one positive at a prevalence of 1/10,000⁵.

A positive result from such a study would indicate either that the Hilton study gives a credible estimate of the prevalence, whereas the NATA study does not, or that the IHC method is throwing up false positives. On the other hand, failure to find any positive appendix samples in this repeat experiment would strongly suggest that the original Hilton study is not a reliable indicator of prevalence.

The nature of the fixed archived samples precludes any extensive further analysis to see if these are false or true positives. The hypothesis can, however, be investigated

⁴ In the case of NATA eventually producing say one EIA positive result in 100,000 tonsil samples.

⁵ The collection of 30,000 samples would also allow a significant portion (say 1/3) to be collected outside the main age range of the Hilton study, which will be important for both the consideration of age based models and certain specific policy questions.

by examining further NATA EIA negative tonsil samples. At the apparent prevalence (that is including any IHC false positives) of 1/10,000 there is a 95% probability of finding at least one sample in 30,000 tonsil samples. If such a sample is found it could then be investigated to determine its status. If no such sample is found it would be strong evidence that infectivity is not developing in tonsil tissue at an early enough stage of infection to be of use in prevalence studies.

An efficient testing algorithm

At first sight the considerations above suggest that we would need to test of the order of 30,000 archived appendix samples and 30,000 NATA tonsil samples by IHC to resolve the prevalence discrepancy. Currently available capacity in the UK for IHC testing could probably support up to 15,000 tests per year but with a wastage rate of the order of one third. This suggests that 10,000 successful tests might be possible each year. Testing 60,000 samples would then take six years (a little less because some NATA EIA negative samples have already been tested by IHC). It is however possible to construct a more efficient testing algorithm which would typically allow reasonably firm conclusions to be drawn after the testing of 30,000 to 40,000 samples (possibly many fewer) and reasonable provisional conclusions to be drawn earlier, in most cases by the end of 2009.

The algorithm is discussed in annex B. Annex A discusses the timescale on which both provisional and relatively definitive results should become available.

Conclusion

A repetition of the Hilton study with some 30,000 appendix samples would illuminate the apparently discrepant results of the NATA and Hilton studies. A single positive result would indicate that prevalence is indeed better represented by the estimates of the Hilton study than those of NATA, or alternatively that the IHC assay is producing a significant number of false positives. The possibility of a significant false positive rate from IHC testing can be investigated if any positive results are found in the 10,000 NATA EIA negative samples currently being tested by IHC. Should no positives be found, testing another 20,000 such samples should find one such positive if the Hilton estimates for positive IHC results are correct and infectivity is indeed present in tonsil tissue. Whatever the outcome such investigations should provide a substantially firmer basis for future policy decisions with major resource implications.

While it may be necessary to test 30,000 archive appendix and 30,000 NATA EIA negative samples by IHC, reasonable conclusions may be drawn after testing substantially fewer samples using the algorithm described in annex B. Provisional conclusions may be reasonably drawn even earlier, generally before the end of 2009 assuming the testing schedule of annex A.

Annex A: Timescales

The following timescales for reaching the various endpoints of the decision tree described in annex B are based on three assumptions.

1. That some 15,000 tonsil tissues samples (either archived appendix samples or NATA tonsil samples) can be tested by IHC each year. This is higher than current throughput rates for the initial 10,000 NATA EIA negative samples being tested by IHC but can reasonably be expected after more experience has been gained.
2. That some two thirds of these tests produce acceptable results.
3. That the IHC testing of the initial 10,000 NATA samples will be completed by September 2008.

Working from the top of the diagram in annex B:

The first three endpoints indicate decisions which could be made **without further testing** if current biological knowledge supported the appropriate inferences i.e.

1. Accept the Hilton estimate, if the NATA assays are considered insensitive to infectivity or that detectable infectivity was thought to arise only close to the development of clinical symptoms.
2. Accept the NATA estimate, if it were considered that the Hilton study was simply unlucky in picking up so many positives or that IHC has a significant false positive rate.
3. Accept a combined estimate on the basis that there was no reason to doubt either study and that the statistical discrepancies were simply due to the occurrence of improbable events.

The fourth endpoint represents a situation where an analysis of the first 10,000 NATA EIA negative samples has produced some IHC positives, which have been further confirmed by other tests and observations. In this case the Hilton estimate is to be preferred. Such a result is possible by **September 2008**.

The next endpoint represents a case where an analysis of the first 10,000 NATA EIA negative samples has produced some IHC positives which have not been confirmed by other tests and observations. A *provisional* decision in favour of the NATA estimate could be made by **September 2008** but a definitive result might not be available until **September 2010**. (A similar situation exists if the IHC result cannot be confirmed or refuted by other tests, except that it may be preferable to make a preliminary decision in favour of the Hilton study and if no further evidence is obtained by testing up to 30,000 tonsils it will be necessary to begin appendix testing).

Moving back to the left hand edge of the diagram, if no IHC positives are found in the 10,000 NATA tonsil samples then appendix testing takes place until either an IHC positive is found or 30,000 appendix samples have been tested. Assuming no positives are found a preliminary decision in favour of the NATA estimate is possible in **September 2009** with definitive results in September 2011.

If an IHC positive appendix sample is found then a preliminary decision in favour of the Hilton estimate could be made sometime between **September 2008 and September 2011** with up to two years additional testing of tonsil samples to confirm or refute the preliminary result.

*In summary at least a preliminary determination in favour of one estimate or another should be possible **by September 2009**. Most routes through the decision tree produce definitive results by **September 2011** but some may take until 2013.*

Annex B: Test Decision Tree

The following diagram indicates a decision tree designed to minimise the number of tissue samples which would need to be tested by IHC to arrive at reasonably firm estimates of vCJD prevalence to inform policy decisions. The tree is based on the assumption that NATA estimates of prevalence remain significantly smaller than those from the Hilton study.

Rectangular boxes indicate states of knowledge. When attached to ‘end-point’ circles they indicate the conclusion which can reasonably drawn by reaching that end point. At other points they indicate reasonable hypotheses. These hypotheses may be based on a ‘balance of probability’ argument rather than definitive states of knowledge.

Diamonds indicate some form of test. Typically these will be of a certain number of tissue samples. A plus sign indicates a number additional samples above an initial 10,000 that will have been tested by that point. There are a few routes through the diagram where these additional tests will have already been undertaken and the decision can be made on existing results

Circles indicate points where the algorithm halts because one may draw a reasonably definitive conclusion for policy purposes. These conclusions are indicated in the attached ‘state-of-knowledge’ rectangle.

The decision route starts in the top left hand corner with a box indicating the current results of the NATA and Hilton studies.

Test Decision Tree - vCJD Prevalence

