EUROPEAN COMMISSION
DIRECTORATE-GENERAL JRC
Institute for Reference Materials and Measurements (IRMM)
RM Unit

REPORT

THE EVALUATION OF FIVE RAPID TESTS FOR THE DIAGNOSIS OF TRANSMISSIBLE SPONGIFORM ENCEPHALOPATHY IN BOVINES (2ND STUDY)

27 March 2002

REPORT

THE EVALUATION OF FIVE RAPID TESTS FOR THE DIAGNOSIS OF TRANSMISSIBLE SPONGIFORM ENCEPHALOPATHY IN BOVINES (2ND STUDY)

Heinz Schimmel, Paolo Catalani, Lionel Le Guern, Jozeph Prokisch,
Wolfgang Philipp, Stefanie Trapmann, Reinhard Zeleny
Reference Materials Unit
Institute for Reference Materials and Measurements
Joint Research Centre
Geel
Belgium

Schimmel@irmm.jrc.be

Jim Moynagh
European Commission
Rue de la Loi 200
1049 Bruxelles
Belgium

James.moynagh@cec.eu.int

Vers 1.1

THE EVALUATION OF FOUR RAPID TESTS FOR THE DIAGNOSIS OF TSE IN BOVINES (2ND STUDY)

1. Background

1.1 Objective

In 1999 the European Commission evaluated 4 rapid diagnostic tests for BSE [1]. Three tests showed an excellent performance demonstrating 100 % sensitivity and 100 % specificity on the samples examined. All three tests have since been approved for programmes of BSE monitoring in cattle. Approximately 7 million tests per year are now being performed within the European Union and more worldwide.

Several new tests have been developed since 1999 possibly showing equal or better performance than those already evaluated. A second evaluation was carried out to examine these tests. Some tests may offer additional features like the ability to distinguish between various forms of TSE, more explicitly in sheep.

As in 1999 the objective of the evaluation exercise is to evaluate diagnostic tests for TSE in bovines. As a first step also for those tests which may distinguish between TSEs an acceptable performance level with bovine brain samples has to be demonstrated before further evaluation with specific sheep samples being infected with various TSE strains or non infected can take place. This evaluation is designed to obtain concrete information on the performance of the tests which are either currently available or at an advanced stage of development. The availability of this information will allow decisions to be taken on possible future uses of such tests.

This test evaluation has been designed and managed jointly by the Directorate General of the European Commission, Health and Consumer Protection in collaboration with an expert group and the Directorate General Joint Research Centre, Institute for Reference Materials and Measurements (IRMM) located at Geel, Belgium. All practical aspects of the sample preparation were carried out at IRMM.

1.2 Parameters Examined

Ideally a diagnostic test should perform with 100% accuracy, in respect of its ability to correctly detect infected and uninfected animals. In reality, however, this is seldom if ever the case. Before a test is used it is essential, particularly where its use impinges on decisions relating to public health, to know how it performs.

All diagnostic tests have a level of false positives and false negatives. The best tests minimise these factors. Because BSE is rare, at least in most animal populations, even a low level of false positives can greatly exceed the true positives detected by the test. Unless a good confirmatory test is available, this situation may simply cause confusion and be economically wasteful. On the other hand, low sensitivity, resulting in false negative reactions will result in diseased animals reaching the market with possible public health consequences. The sensitivity and the specificity of diagnostic tests can often be adjusted by varying the cut-off point, the point taken as the divider between the positive and negative readings. Usually increasing the sensitivity will reduce the specificity and vice versa.

For the initial assessment of the tests the cut-off point used was that proposed by the test developers. These values were sometimes arrived at on the basis of limited small scale trials. Following this assessment consideration can be given to selecting the optimum cut-off value or indeed to whether the inclusion of an inconclusive or doubtful category is useful.

The following parameters were examined in the evaluation:

- Diagnostic specificity: the ability to correctly identify non-infected animals;
- Diagnostic sensitivity: the ability to correctly identify infected animals with clinical symptoms;
- The detection limit or the ability of the tests to detect small quantities of PrP, the substance implicated in the causation and progression of TSE diseases. All methods, also qualitative tests, have a detection limit. The detection limit is an important parameter to be determined because it gives an indication on the risk of getting false negative results for samples containing low PrP levels. The detection limit could also give an indication of the capability of the tests to detect the presence of pre-clinical BSE.

Experience gathered after the introduction of large scale BSE screening in the European Union in 2000 shows that also the properties of the samples analysed may have an impact on the definition of the cut off value. For example our experience shows that unfrozen samples, as usually collected in slaughterhouses and submitted for analysis to testing laboratories, give consistently higher results than frozen samples used for the test evaluation. Such phenomena are well known in the field of clinical chemistry and the term commutability is used to describe them.
For the second evaluation it was decided to reduce the total number of sample slices provided to each test developer from 1400 in 1999 to 200. This number allows to assess the general performance of a test and the separation between data populations for negative and positive samples. It is highly recommended to fix the cut off value finally to be applied by field laboratories after analysis of a larger number of fresh field samples mostly being negative because of the low incidence of the disease. In case of a weak separation of the negative and positive data populations (an overlap would mean that there are false positive of negative identifications) the collection of data to fix the cut off should be done on a population with a rather high incidence of BSE. This is because an adaptation of the cut off to higher values bears the risk that weak positive samples may fall under the cut off since they may not increase in signal proportionally. If the cut off could be adapted to a lower level weak positive samples may drop over-proportionally consequently also leading to an overlap of the data populations. In both cases (i.e. good or weak separation of negative and positive data populations) all samples close to the preliminary cut off, i.e. either slightly below or above cut off, should be confirmed by other tests because the properties of the negative as well as positive samples may change with the sample history as the above mentioned example shows.
The evaluation was based on samples having been stored as such or prepared with material having been stored since 1999 under appropriate conditions in a freezer at – 70 ° C.

1.3 Additional parameters examined

The following aspects were examined with the contribution of several test developers:

- Comparison of the detection limits with homogenate samples used in 1999
- Assessment of the impact of homogenisation of brain tissue

Not from all rapid BSE tests, already evaluated in 1999 or part of this evaluation, data could be obtained due to limitations in samples available and the need to apply quantitative tests to quantify the effects. The selection has been also made on basis of different test principles. In addition data on the stability of test samples have been collected where available.

The results are presented in this report because of their relevance for the understanding of the outcome of this evaluation, the comparison of the evaluations carried out in 1999 and 2001, the set up of protocols for future test evaluations, for general scientific interest and their possible consequences for research projects. In particular the investigation on the impact of homogenisation on the performance of rapid BSE tests is important for the demonstration of the suitability of the use of homogenates for the assessment of the detection limits of test and their relative sensitivities.

1.4 Heterogeneity of PrP concentration and its implications

In the context of detection limit and sensitivity is the possible heterogeneity of the PrP distribution in bovine brains an important aspect. During the 1999 evaluation [1] it has been observed that occasionally samples gave considerable low readings. These samples originated predominantly from locations more distant from the obex and more particular form places posterior to the obex. However in a few cases low readings have also been observed in tissue originating from locations anterior and more often posterior to the obex directly neighbouring the obex region. Out of 300 positive brains and spinal cord samples 21 individual samples gave weak signals in one or more tests. In only five individual cows the signals were consistently low or at medium level for all three successful tests. In all other cases one slice has given high readings whereas other slices of the brain or spinal cord of the same cow gave very low values. This indicates a degree of heterogeneity of the prion concentration in the CNS tissue of positive cases at the clinical stage. A comparison of the dilution series used to assess the detection limits in 1999 with samples having given low signals indicates the following:
9 brain stem samples out of 334 analysed with the BioRad Platelia test corresponded to a dilution level of 1:10 to 1:20, 1 sample each to a dilution of 1:50, 1:160 and less than 1:300 respectively.
7 spinal cord samples out of 334 analysed with the Enfer test corresponded to a dilution of more than 1:10, the two lowest signals almost to a 1:20 dilution.
The application of the Prionics Check Western Blot test on the dilution series resulted in problems related to incomplete digestion of the samples. Data obtained within the frame of this evaluation indicate that 1 out of the 334 samples corresponded to a

dilution of approximately 1:80 (directly neighbouring the obex region anterior to the obex). There have been 2 other samples which were classified as weak signals which correspond to lower dilutions but in addition 88 samples which gave bands of medium intensity only.
Therefore the lower the detection limit for PrP the lower the probability that the test will produce a false negative result with sub-samples containing low PrP levels which obviously occur even in infected animals with clinical symptoms. In the same context attention has to be given to appropriate sampling i.e. location where the sample is taken and sample amount as well as number of independent replicates.

1.5 Planning

The organisation of the evaluation was divided into a number of phases.

1. Design of experimental protocol
2. Call for expression of interests to identify the most advanced rapid diagnostic tests
3. Selection of the most advanced tests for evaluation.
4. Composition of sample sets and where necessary preparation of tissue as blind coded samples in the P3 facilities at IRMM Belgium.
5. Testing of the blind samples by the participants.
6. Analysis of the results.

2. Selection of participants

2.1 Call for expression of interest

Expressions of interest were invited from those who had tests in advanced stages of development or available for use. Emphasis was laid on the ability of the tests to distinguish TSEs or on their sensitivity. The call was published in the Official Journal of the European Communities No. C42 of 15 February 2000. Applications were received from seven organisations which covered eight tests.

In order to ensure that useful tests would be widely available, applicants were also required to give assurances that they were prepared to make their tests available on a non discriminatory basis following the evaluation.

2.2 Assessment of applications

The applications were assessed by a scientific expert panel comprising four external scientists based on criteria covering the scientific basis for the test, available experimental evidence, practicality of the sampling and testing procedures and stage

of development of the test. Based on this assessment, five tests were selected for inclusion in the test evaluation exercise. These tests are listed in section 3 below.

3. Tests Selected for Evaluation

Tests from five organisations were selected for participation in the evaluation exercise on the basis of the information submitted by them in response to the call for expression of interest. So far only data relationg to the following four tests are available and can be included in this preliminary report.

Test A **ID Lelystad, The Netherlands**

Test B **PerkinElmer Life Sciences, United Kingdom**

Test C **Prionics A.G., Switzerland**

Test D **University of California in San Francisco, USA**

Test E **MRC Prion Unit, Imperial College, United Kingdom**

4. Evaluation Protocol

The evaluation exercise was designed primarily to evaluate the sensitivity, specificity and detection limits associated with each of the selected tests. All samples were presented for testing in a coded 'blind' format or coded on site by European Commission staff. The protocol used is attached in Annex 2.

4.1 Sensitivity

The sensitivity of a test is the proportion of infected reference animals that test positive in the assay. 48 samples from the same number of individual animals were used to assess this element. On purpose three of the positive cow brains having given consistently low results were included in the set of samples.

4.2 Specificity

The specificity of a test is the proportion of uninfected reference animals that test negative in the assay. A total of 152 samples from the same number of individual animals were included in the exercise for this purpose.

4.3 Detection limits and Titre

The material used to prepare the dilutions was nervous tissue that had been titrated in RIII mice, yielding a titre of $10^{3.1}$ mouse i.c./i.p LD50/g of tissue.

To assess the test detection limits, various dilutions of the positive brain homogenate were used. The sample was presented in various dilutions down to 10^{-3}. The objective of this was to determine the test detection limits and also to gain a perspective on the behaviour of the test in highly heterogeneous samples and in pre clinical animals. In order to achieve acceptable viscosity, 20% aqueous solution containing 5% sucrose was added to the homogenate.

The number of samples examined by each test is set out below.

Dilution	**Number of samples**	**Dilution**	**Number of samples**
Undiluted	1	$10^{-2.5}$	4
$10^{-1.0}$	4	$10^{-3.0}$	4
$10^{-1.5}$	4	Negative	4-5
$10^{-2.0}$	4		

4.3.1 Examination of the commutability issue

Because commutability problems were encountered with one test another experimental set to assess the detection limit was proposed to all test developers. The following conditions were applied:
The same material used to prepare the dilutions mentioned above was diluted serving as a common, defined and homogeneous starting point. This starting material did not contain any buffer. The tissue concentration at each dilution level had to be the same i.e. when diluting the positive sample negative material had to be used to maintain the same tissue concentration level. The matrix (tissue) background plays an important role in the assessment of the detection limit. If kept constant a much more realistic determination of the detection limit is possible compared to dilution curves where besides the analyte also the analytically challenging background is diluted and consequently reduced in its impact. Five brain stems from New Zealand cows were provided as blank material for dilution and as negative control. Suitable and equal amounts of negative material from these brains had to be homogenised according to a procedure deemed by the test developers to be most appropriate. Also the positive starting material was treated in the same way. The obtained positive homogenate had to be diluted with the negative homogenate in three fold dilutions (i.e. 1:3, 1:9, 1:27, 1:81, 1:243, 1:729 etc.). The same pipette had to be used to guarantee that the dilution factor was as expected even if the pipette volume was significantly deviating from the nominal value. At least four aliquots per concentration level had to be produced by the test developer and were coded on site by Commission staff. The identity of the samples was unknown to the test developers.
Three out of four test developers have performed this experiment whereas UCSF wanted to understand first all parameters having potentially impact on the assessment

of the detection limit. The proposal however remains still valid and the experiment could be carried out as soon as UCSF would be prepared.

4.4 Additional features

4.4.1 Comparison with the evaluation carried out in 1999

In order to allow comparison of data obtained in this evaluation with data gathered in the previous evaluation two test developers of the 1999 exercise where asked to analyse the newly prepared diluted homogenates provided in 2001 together with a limited number of remaining homogenates produced and used in 1999. BioRad agreed to carry out the analysis. Enfer was also asked to perform the analyses but they could not be carried out yet.

4.4.2 Assessment of the impact of homogenisation of brain tissue

So far for the evaluation of the detection limit of the tests homogenised samples have been used, at least as a starting material. The validity of this approach has been investigated by a targeted experimental set up.
There are indications that mechanical stress during homogenisation is reducing the measurable PrP concentration. If this reduction of the test signal is proportionally the same for every test their relative sensitivities can still be compared. The questions whether homogenisation of brain tissue is in principle leading to a reduction of test signal by biochemical reactions or whether the signal reduction for a test could be more than for others some tests based on different principles have been investigated.
A selection of tests based on different test principles was necessary because of the limited availability of well shaped and high quality positive samples required for such investigations. Only quantitative tests were considered. The test from UCSF, Prionics (ELISA), Enfer and BioRad Platelia have been selected. From Enfer no data are available yet. The information obtained is of large importance for future assessments of detection limits but also for the production of reference materials for quality control and calibration because they have to be homogenous but also representative. The set up of the experiment is outlined in chapter 6. It is based on the observation made during the 1999 evaluation [1] that the test signals were symmetrically distributed between both brain stem or spinal cord.

4.4.3 Stability of test samples

From the data collected during the evaluation exercise some information on the influence of the storage temperature on the behaviour of the samples could be derived by comparing the results obtained for samples stored at higher temperature with those obtained for samples stored at lower temperature. If degradation of samples occurs it should be more pronounced at higher temperature according to the Arrhenius law. In principle equivalence of results for samples stored at different temperatures is indicating stability of samples.

5. Selection of Positive and Negative Tissue

As stated above sections of both CNS tissue and tissue homogenates were used in the evaluation of each test.

1. The whole central nervous tissue was of the type which was used routinely in each test and on which most of the test development work had been carried out. All four participants received brain stem tissue.

2. All participants received identical samples of the homogenised central nervous tissue.
3. All particpants received an identical sample of 0.5 g titrated brain homogenate to produce the dilution series according to their protocol.

The known positive and known negative material was obtained as described below.

5.1 Tissue from Infected Reference Animals

5.1.1 Criteria for selection

Brainstem and spinal cord samples were selected from bovines showing clinical signs of BSE and were included in the project following confirmatory tests as detailed below. These samples were supplied by the Veterinary Laboratory Agency, Weybridge, United Kingdom.

5.1.2 Collection methods

Following euthanasia, the brainstem was removed from each suspect animal by extraction via the foramen magnum.

5.1.2 Confirmatory tests

A slice of tissue containing the obex with a thickness of about 1cm was removed for diagnosis. This tissue was fixed in 10% formal saline, followed by standard histopathological processing and embedding in paraffin wax, and staining of 5µm sections with a modified HE stain. The diagnosis was based on light microscopic examination of the major BSE target sites at the obex i.e. the nucleus of solitary tract and nucleus of the spinal tract of the trigeminal nerve.

Samples were declared positive upon observation of characteristic vacuolation (spongiform change) affecting grey matter neuropil with a systematic and usually bilaterally symmetrical distribution in either the nucleus of the solitary tract or the nucleus of the spinal tract of the trigeminal nerve. At least three neurophil vacuoles were noted in a lesion target site.

5.2 Tissue from Uninfected Reference Animals

5.2.1 Criteria for selection

Samples were collected in New Zealand from healthy adult bovines (at least 4 years old) of mixed breeds. New Zealand was selected as the source for negative material because of its widely recognised high status as regards freedom from TSE diseases. No cases of BSE have ever been reported in New Zealand. In addition its cautious importation policy has minimised the risk of importing animals or feed that might carry the disease.

5.2.2 Collection methods

Following slaughter, the skull was opened to allow the removal of the brain.

5.2.3 Confirmatory tests

Animals were healthy and showed no clinical signs suggestive of nervous disease. As an additional check, the obex from each sample was removed and subjected to histological examination as described in 5.1.2 above. No samples were removed from the study as a results of this examination.

6 Sample Preparation and Testing Procedure

Preparation of sample slices

The positive and negative material was delivered end 1998, beginning 1999 to the EU IRMM at Geel Belgium. The samples have since that time been stored at –70 ºC. The sample preparation took place in 1999 for the negative samples and in 2001 for the positive samples.

Prior to the arrival of the samples the sample tissue containers were prepared and pre-labelled according to a randomised formula. The containers were then separated into two separate batches, those destined for positive and negative material.

Before being processed each sample was inspected visually and the code given by the suppliers was checked. Inappropriate samples or samples with inconsistencies concerning the code were not processed. The samples were then cut according to one of three schemes depending on the anatomical presentation of each sample. The presentation varied due to the natural fluctuation of the sample size and the difficulties in achieving fully reproducible sampling conditions in the abattoirs (NZ) and the veterinary centres (UK). Fig. 1 below shows the most frequently applied cutting scheme and all three schemes used are shown in Annex 3. The following criteria were applied for the selection of the sampling scheme:

	Minimum length of the brain stem posterior to the obex (L2)	Minimum length of the depression between the cerebellar penduncles (Ventricle) on the brain stem anterior to the obex (L1)
Scheme 1	2.1 to 3 cm	1.5 cm
Scheme 2	1.4 to 2 cm	2.0 cm
Scheme 3	0.7 to 1 cm	2.5 cm

Samples which did not fulfil these criteria were rejected. Of the 1064 negative samples 8 have been processed according to scheme 1, 899 to scheme 2 and 157 to scheme 3. One set of 1064 negative samples has been split in five sets of samples with consecutive numbers of 152 samples per set i.e. the majority of negative samples in each set were not identical to those in another set.

48 brain stem halves of positive samples (remainders of samples of which only side 1 was used in 1999) were all cut according to scheme 2. The 48 positive samples in each set originated from the same individual cows. The 48 samples chosen comprised on purpose 3 samples which tended to give medium or low intensity signals in the 1999 evaluation [1].

Five sets consisting of brain stem slices of 200 samples each have been prepared. As an average 16 % each of the negative brain stem samples were taken from positions 1.1, 1.2, 1.3, 1.4, 1.5 and 1.6. The slice 1.1 of the positive samples has been always reserved for the set kept by IRMM for confirmation purposes. This means that 20 % each of the positive brain stem samples were taken from positions 2.2, 2.3, 2.4, 2.5 and 2.6. The sampling scheme ensured that the brain stem samples supplied to the companies were taken at maximum 2 to 2.5 cm away from the obex.

The schemes allowed the recording of the anatomical location of each individual sample to allow analysis of any relationship between the location and the test response. The corresponding duplicate samples were always taken from the opposite side of the brain stem or spinal cord . Approximately 1g of tissue was used for each sample.

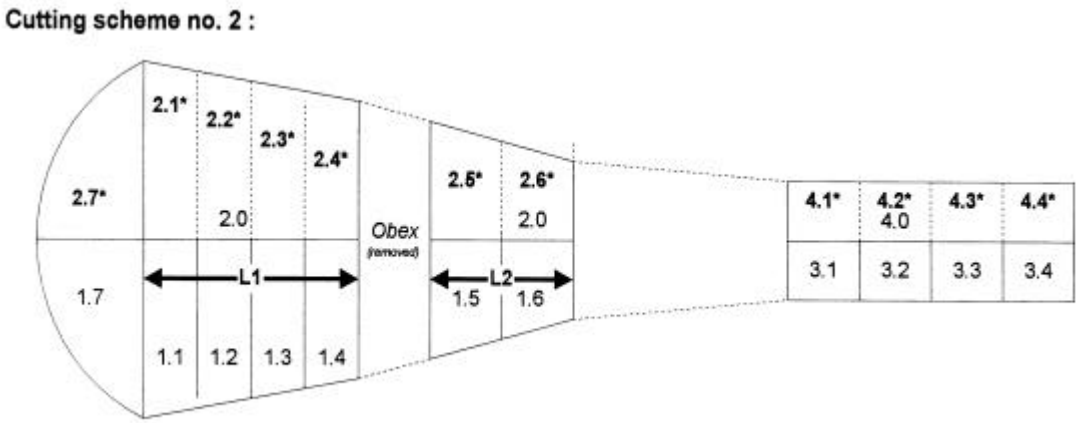


Figure 1: Typical cutting scheme used to prepare the samples

The preparation of the positive and negative samples was separated both in location and time to avoid the possibility of transfer of infectivity or mix up of samples.

Preparation of homogenates:

Series (a)
After addition of 20 % w/w of an aqueous solution of 5 % sucrose in total 172 negative brain stem samples were homogenised with an Ultraturrax mixer in smaller pools. The homogenates were pooled. The pool was filtered and used for the production of the dilution series.
The positive titrated homogenate (see chapter 4.4) was also diluted with 20 % w/w of aqueous 5 % sucrose solution and homogenised with an Ultraturrax. This step was carried out in 1999. The remaining material not having been used at that time has been stored in the meantime at –70 °C. In 2001 it was used for the production of a new series of diluted homogenates. The 10^{-1} and $10^{-1.5}$ dilution was prepared by gravimetrical mixing of appropriate amounts of the pooled negative material and the titrated homogenate, both containing 20 % w/w of aqueous 5 % sucrose solution. A

propeller stirrer was used as mixing device. The lower dilutions were each prepared by 1 in 10 dilution of the corresponding higher concentrated homogenate.

Series (b)
In addition another sub-batch of the same titrated pool as used in 1999 and 2001 has been obtained in 2001 from VLA, UK. It served as the starting material for the dilution series prepared according to a protocol of the test developers and for evaluation of the stability of the homogenates used in this study (see chapters 4.3 and 7.8).

Preparation of samples to study impact of homogenisation:

For studying the impact of homogenisation on the test signals sample slices have been cut from 5 well shaped brain stems. Care was taken to ensure that the brain stems were cut along the middle line and rectangular to this line. The weight per slice was approximately 1g meaning that each slice was about 3 mm thick. Alternating between the two brain stem halves slices have been kept frozen at –70 °C and homogenised with a turrax type mixer (Omni with disposable probes) for 45 seconds after addition of 20 % aqueous 5 % sucrose buffer. The samples were still cool when starting turraxing and immediately frozen after preparation. Sets of 10 pairs of corresponding slices and homogenates composed one sample set. Care was taken that the samples per set originated from at least three brains and within one brain were neighbouring each other. The samples were cut according to the scheme outlined in annex 4.

Testing procedure

P3 facilities were used for the positive material. Samples were then airfreighted in dry ice to each participant.

The testing of these samples took place in the participants' premises and was supervised by Commission officials. The following additional features were put in place to ensure confidence in the exercise. All samples were sealed for delivery with special numbered seals. Participants were required to test the samples in a particular order and to inform the commission on the opening of each container as well as forwarding each evening the results of that days testing. Further details of these requirements are contained in the test protocol in Annex 2.

In principle the test developers were allowed four weeks for completion of the testing however all four test developers were able to complete the testing of 200 slices within 2-4 days. The original intention was to evaluate most tests in a two stage approach i.e. first with the dilution series of homogenates and if performing well with the full set of 200 sliced samples. However after having observed in June 2001 major commutability problems of one test (third test to be evaluated) with homogenates it was decided that the test developers whose test still had to be evaluated and the previously in the first step evaluated test developers should get the full sample set. Due to delayed release from customs one set of samples thawed during transport, was destroyed under appropriate conditions and replaced by a new set. The first evaluation (first stage) took place in December 2000 and the data of the fourth test evaluated became available beginning of September 2001.

7. Results of the Evaluation

The sensitivity and specificity estimates are based on the predetermined cut off points supplied by the participants in advance of the evaluation. The selection of cut off points is considered in section 7.7 (see also remarks in chapter 1.2).

Sensitivity and specificity were calculated on the basis of 48 and 152 samples respectively. On occasion it proved necessary to repeat tests during the evaluation exercise. This was necessary either because the test protocol involved an inconclusive category which required a retest to clarify the status of the sample.
If samples had to be re-tested because it fell into the inconclusive category the result of the repetition is taken into consideration. If measurements were repeated on request of IRMM for scientific interest and not because they were classified by the test developer to fall in the inconclusive category based on the predefined cut off criteria the original result is considered.
The objective of the evaluation was to evaluate test performance not laboratory performance. Consequently efforts were made to resolve a small number of errors arising from factors such as misreporting, sample mix up and contamination. Results deemed to be due to laboratory mishaps, rather than a problem with the test or the sample, and which were later resolved by a repeat test where necessary or by tracking in the documentation have not been taken into account. If obvious technical problems occurred during the evaluation (pipetting errors due to apparatus failure or human error or obvious problems with reagents) the measurements were repeated and the results of the repetition have been taken into consideration.

All such cases are detailed below in the results section.

7.1 Test A ID - Lelystad

The test is based on the comparison of two spots of brainstem homogenate which are treated differently, one is treated with chaotropic agent, the other is non-treated.
Proteolytic treatment is applied to homogenates and detetction is by an enzyme linked monoclonal antibody directed against unfolded PrP^{Sc}. Proteinase K digested brain extracts are diluted and spotted (dot blots) on two membranes of which one is further exposed to chaotrophic solution (***T*** membrane) and the other processed without further treatment (***N*** membrane). The two separate membranes contain equal aliquots of the digested homogenates. Samples with proteinase K resistant PrP fractions can then be detected after an enzymatic reaction as darker spots in a photographic film on the ***T*** membrane, compared to the corresponding spot situated on the non-treated membrane ***N***. Each sample was carried out in duplicate (i.e. 2 spots on each membrane)
The test is performed in a 96 well format, results can be obtained after 6 hours.
Examination is carried out visually by two independent persons and alternatively automatically by a CCD-camera system.

Decision criteria (cut off definition)

The visual inspection of the blot images does not give a quantitative reading. It gives a qualitative response and is based on comparison of the dot intensity on the *T* and *N* membrane. In order to avoid distraction of the reader only the dots under evaluation are visible through appropriately positioned holes in a mask. If compared to the *N* membrane the spot on the *T* membrane is darker it is classified as positive. If only one of the duplicates is judged darker, the sample is classified as 'dubious'. The blots are independently read by two persons. The final result is positive if both persons classify the sample as being positive or if one person classes it as positive and the other as 'dubious'. If both classify the sample as 'dubious' or negative or if one person classifies it as positive and the other as negative the sample is considered negative. Thus, dubious final results are not occurring by visual inspection.
With the CCD-camera system the theoretical cut-off is defined as 0.015. This cut-off value was obtained from a preceding series of BSE negative samples. The value obtained for two corresponding spots on the *T* and the *N* membrane are subtracted following the equitation *t-n*. Values > 0.015 receive the score positive, equal or lower results score negative. If one of the two replicates scores positive (i.e. $t\text{-}n > 0.015$) and the other one negative ($t\text{-}n < 0.015$) the sample is classified 'dubious' and will be subjected to further visual inspection.

Sensitivity and Specificity (visual inspection)

	True positive	**True negative**	**Totals**
Test Positive	47	0	47
Test negative	1*	152	153
Totals	48	152	200

	%	95% Confidence Interval (one sided Poisson)
Sensitivity	97.9*	90.1 %
Specificity	100	98.0 %

* one false negative sample which was tested positive when repeated upon request and which may be considered as one of the lower concentrated sub-samples taking the aspects summarised in chapters 1.4, 7.8 and the detection limit as summarised below into consideration

Sensitivity and Specificity (CCD-camera evaluation)

Recording blot images with a cooled CCD-camera (12 bit resolution, program: Gel-Pro Analyser, Media Cybernetics) revealed certain difficulties, i.e. light spots could sometimes not be transformed by the software when surrounded by a high number of very positive (dark) spots. This concerned 25 spots on two photographs that contained a particularly high number of positive samples.
One sample (IRMM 3474) was judged 'dubious' (single *t-n* values 0.0195 and –0.0133, mean 0.0031, cut-off = 0.015) which was correctly scored negative in the visual inspection.

Comparison of results from visual inspection and CCD-camera system
(classification: BSE-positive; NEG-negative; DUB-'dubious')

	Visual inspection		**CCD-camera**			
Result	**BSE**	**NEG**	**BSE**	**NEG**	**DUB**	**Not read**
Blot 1	0	45	0	45	0	0
Blot 2	6	39	6	39	0	0
Blot 3	23	22	23	4	0	18
Blot 4	18	27	18	19	**1**	7
Blot 5	0	20	0	20	0	0
Total	**47**	**153**	**47**	**127**	**1**	**25**

Samples with false classification

2 samples were initially reported as false negative, however from the original blots and evaluation sheets made available a transcription error was identified by which one positive sample was erroneously reported as a negative. Consequently only one false negative is mentioned in the table above (sample 2874, position 2.6, see cutting

scheme in annex 2). The false negative was read both visually and with the camera as negative (see Fig.3).
For scientific interest the test developer was asked by IRMM to repeat the measurement of sample 2874 having been tested false negative. The repetition was made blindly in a series of 9 samples in the sequence of the initial run and including sample 2874.
The measurements were carried out on basis of remaining spotting dilution, remaining digest, remaining homogenate and remaining tissue. The remaining spotting dilution was again classified negative whereas the remaining digest, the remaining homogenate and the freshly processed remaining tissue were read as positive, but the signal was weak. The fact that sample 2874 has been tested first negative and when repeated on basis of the same homogenate positive reflects the variation in reproducibility of the overall measurement procedure becoming obvious with such weak positive samples close to the detection limit. Additional confirmation of sample 2874 was received with the standard Prionics Check Western blot system where the same sample showed a single (weak) PrP^{Sc} typical band. It should be noted that this sample slice originated from one of the three positive samples tending to give low readings in the 1999 evaluation [1].
An interesting phenomenon appeared during the repetition when a clearly positive sample in the first run gave a negative reading (sample number 2394). However, in this case due to shortage of remaining tissue only 0.16 g of brainstem was subjected to homogenisation. This demonstrates the importance of taking sufficiently large samples reducing the risk of missing areas with accumulated PrP^{Sc} and of being very critical in taking sub-samples.

Dilution Series

Two different dilution series were tested to determine the detection limit of the test.
(a) Dilutions of homogenates in the range of 10^{-1} to 10^{-3} prepared at IRMM in 2001.
(b) Dilutions of homogenates prepared freshly on site with the same mouse titrated positive brain homogenate used to produce series (a) (another sub-batch without addition of any buffer) and a pool of 5 negative brain stem material originating from areas posterior to the obex. Equal amounts (0.47 to 0.58 g) of the five brain stems were collected, homogenised and mixed with the positive material in serial 1:3 dilution steps.
Series (a) was tested in a blind manner and detected all 1:10 dilutions (positive score 4/4, see table). The 1:30 dilution was not detected as positive (positive score 0/4) but gives a somewhat darker spot on the T membrane compared to higher dilutions (see Fig. 4). The CCD-camera failed to recognise the 1:10 dilution (score 1 positive, 1 ‘dubious’, 2 negative, mean 0.011) and had a positive sore of 0/4 for the 1:30 dilution.
After visual inspection by two independent, trained inspectors dilution of 1:81 (positive score 4/4) was detectable in series (b). However the CCD-camera scored positive results up to 1:27 (4/4) but failed to clearly recognise higher dilutions (1:81 was read 3 times ‘dubious’ and 1 time negative by the camera). It may be more difficult for an unexperienced inspector to recognise the 1:81 dilution as a positive especially if compared to much higher dilutions looking not much different (see Fig. 5).

***t-n* values obtained by CCD-camera for dilution series (a), cut off = 0.015**

dilution factor	visual scoring	CCD-camera		
	result*	mean *t-n* value (n = 4)	SD	result*
1	1/1	0.3329	-	1/1
10	4/4	0.0110	0.0088	1(1)/4
30	0/4	-0.0046	0.0043	0/4
100	0/4	-0.0175	0.0050	0/4
300	0/4	-0.0146	0.0036	0/4
1000	0/4	-0.0220	0.0055	0/4
negative	0/4	-0.0277	0.0090	0/4

*Columns with "result" present number of samples scored as BSE positive per total number of samples within a dilution category. Between parentheses are indicated the number of samples that scored 'dubious' by the CCD-camera system.

***t-n* values obtained by CCD-camera for dilution series (b), cut off = 0.015.**

dilution factor	Visual scoring	CCD-camera		
	result*	*mean t-n value* (n = 4)	SD	Result*
1	4/4	0.3145	0.0213	4/4
3	4/4	0.1829	0.0103	4/4
9	4/4	0.0729	0.0052	4/4
27	4/4	0.0306	0.0035	4/4
81	4/4	0.0107	0.0039	0(3)/4
243	0/4	0.0051	0.0022	0(1)/4
729	0/4	0.0053	0.0055	0(1)/4
2187	0/4	0.0021	0.0035	0(1)/4

*Columns with "result" present number of samples scored as BSE positive per total number of samples within a dilution category. Between parentheses are indicated the number of samples that scored 'dubious' by the CCD-camera system.

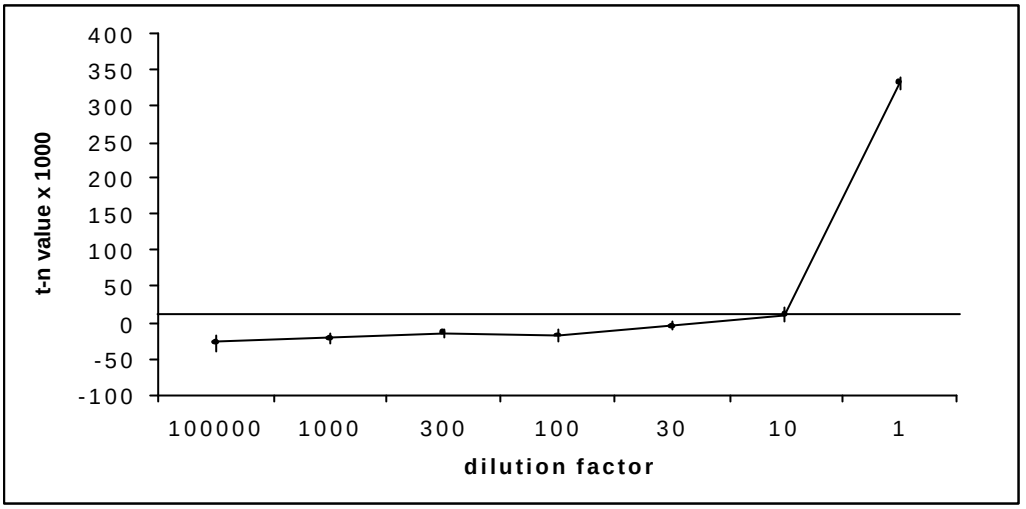


Fig.1. Results for dilution series (a) by CCD-camera measurement. The cut-off is set at 15 (0.015 x 1000).

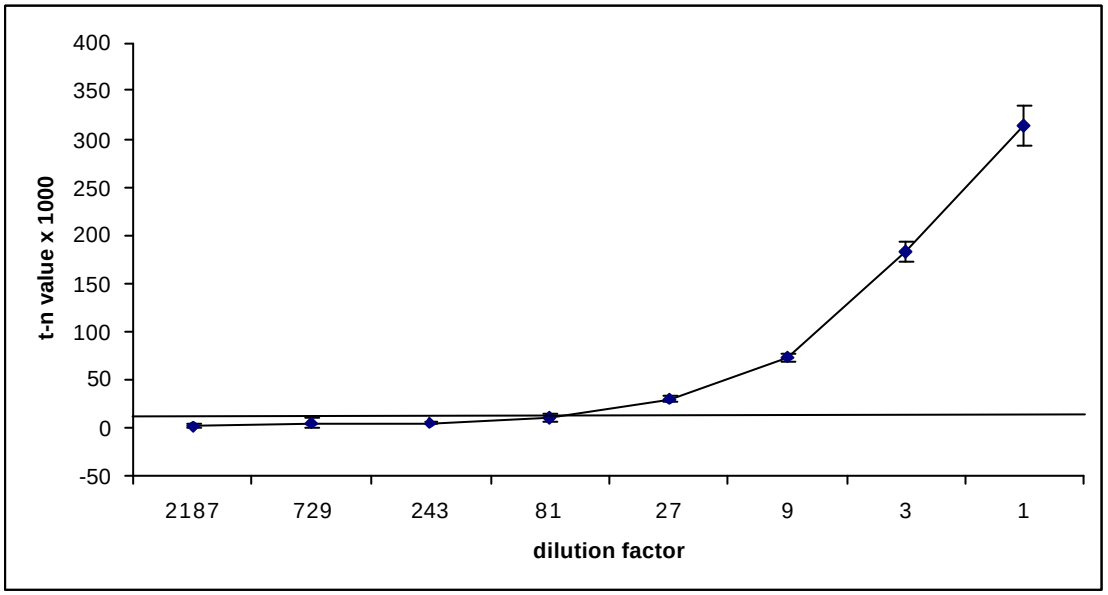


Fig.2. Results for dilution series (b) by CCD-camera. The cut off is set at 15 (0.015 x 1000).

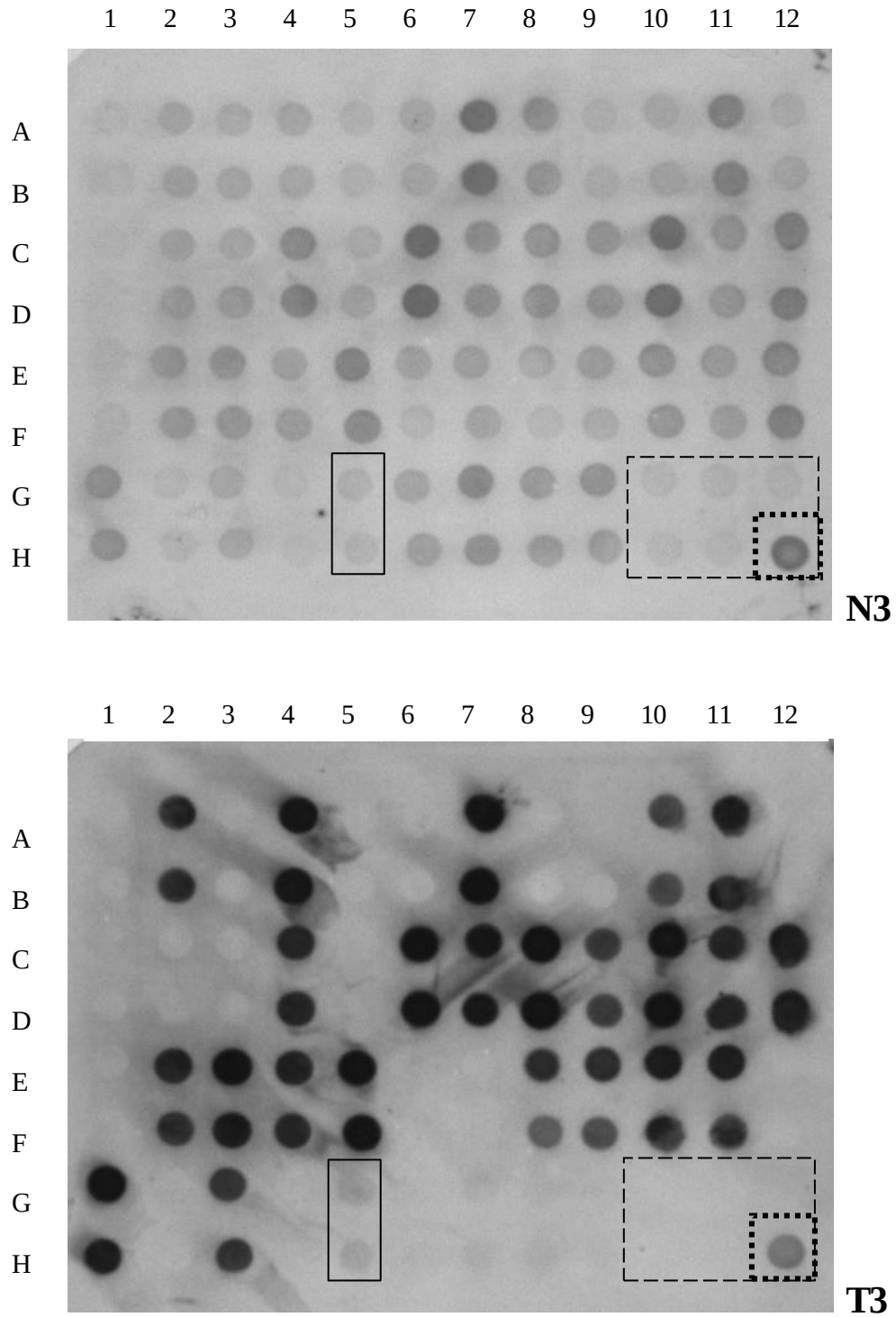
Another aspect to be considered if comparing the evaluation of the two dilution series is the overall darkness of the N membranes (background areas surrounding the dots and the dots themselves). The N membrane for the series (a) (Fig.4) is overall darker making the visual distinction of lighter spots on the T membrane compared to the N

membrane more difficult than in the blot of series (b) showing less intensive dots and background (Fig.5). In addition the blot for series (b) has been made separately whereas the blot for series (a) was run together in a series of five blots.

There are indications that there is a difference in detection limits as determined by series (a) or (b) of up to a factor of 8. However, the factor is more likely closer to 3 because the 1:10 dilutions in series (a) gave a visual reaction closer in appearance to the 1:27 rather than the 1:81 dilutions in series (b).

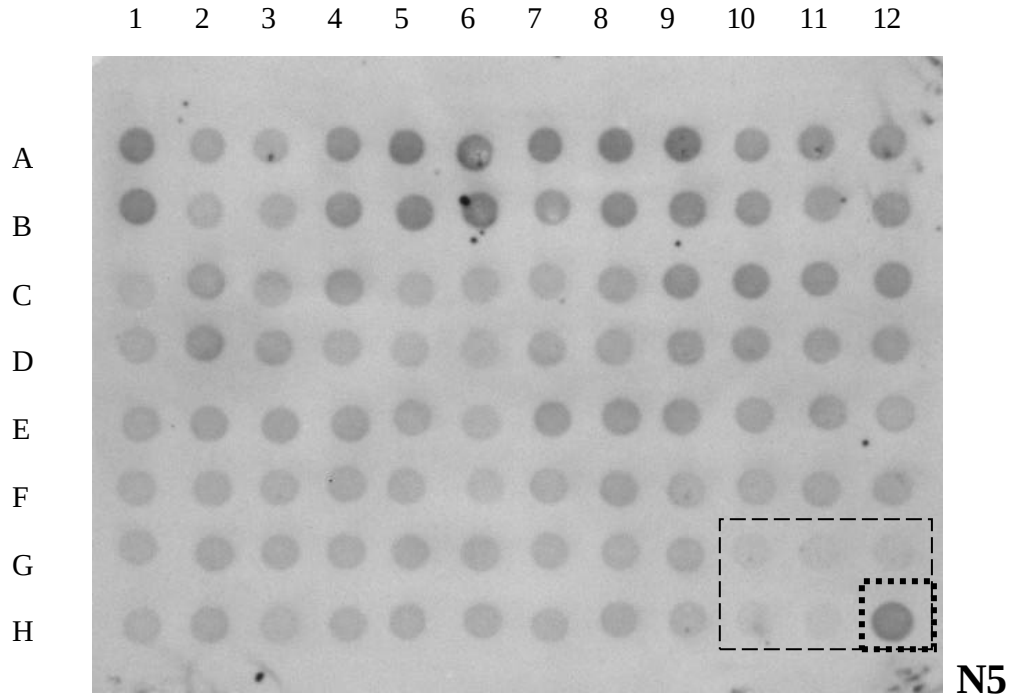
Fig. 3: Blot 3 - false negative sample

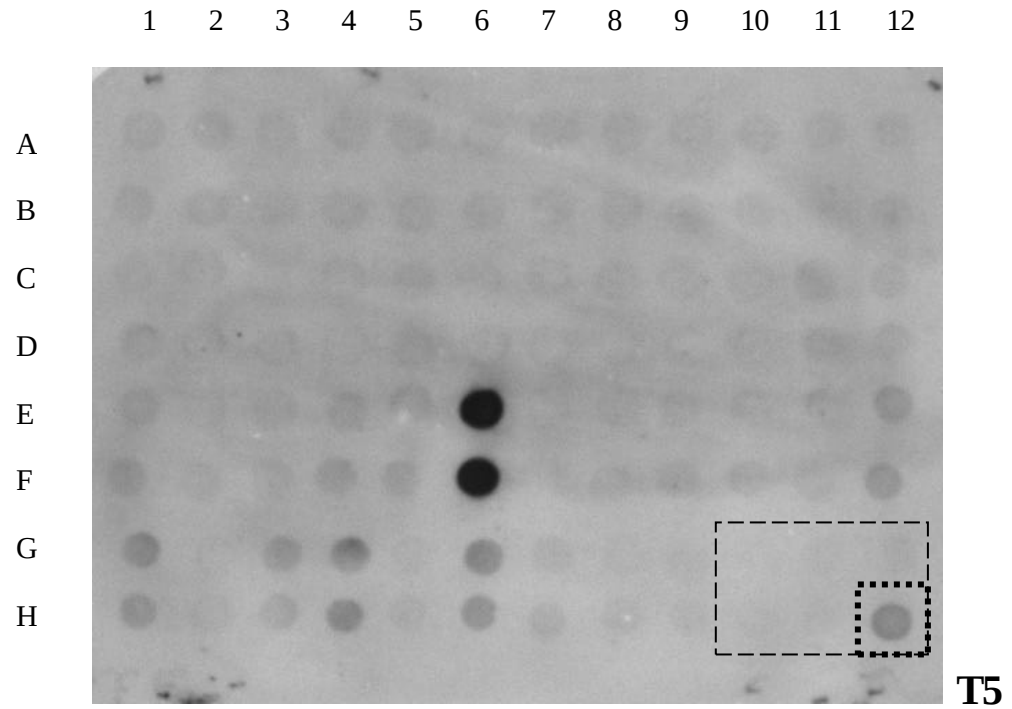
(#2874 in position G5/H5, boxed)



G10, G11, G12, H10, H11 represent digested negative controls.
H12: non-digested negative control.

Fig. 4: Dilution series (a) of homogenates prepared by IRMM



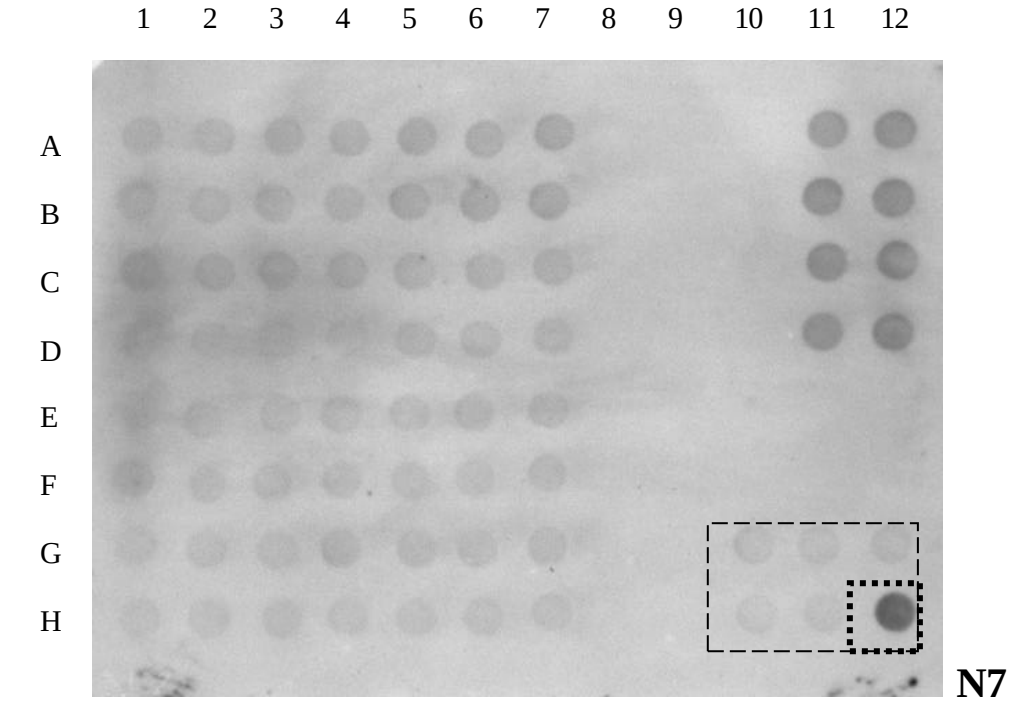


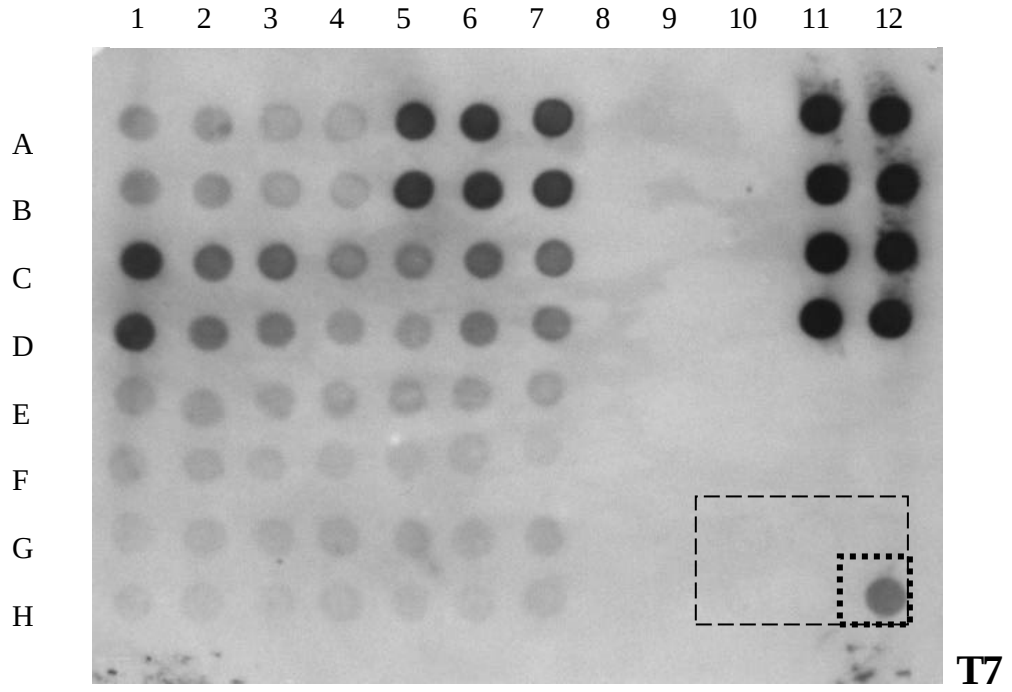
G10, G11, G12, H10, H11 represent digested negative controls.
H12: non-digested negative control.

	1	2	3	4	5	6	7	8	9	10	11	12
A & B												
C & D									neg.	neg.	10^{-3}	10^{-3}
E & F	$10^{-1.5}$	10^{-3}	$10^{-2.5}$	$10^{-1.5}$	$10^{-1.5}$	10^{0}	neg.	10^{-2}	10^{-2}	10^{-2}	10^{-2}	10^{-1}
G & H	10^{-1}	10^{-3}	$10^{-1.5}$	10^{-1}	$10^{-2.5}$	10^{-1}	$10^{-2.5}$	neg.	$10^{-2.5}$			

Dilution for N5 and T5

Fig. 5: Homogenate dilution series (b) prepared freshly





G10, G11, G12, H10, H11 represent digested negative controls.
H12: non-digested negative control.

	1	2	3	4	5	6	7	8	9	10	11	12
A & B	27	27	81	81	3	3	3				1	1
C & D	3	9	9	27	27	9	9				1	1
E & F	81	81	243	243	729	729	2187					
G & H	2187	729	729	243	243	2187	2187					

Dilution-factor for N7 and T7

7.2 Test B PerkinElmer Life Sciences

This test is based on the isolation of both the soluble (supposed to be PrP^{C}) and the insoluble prion protein fraction (presumed to be PrP^{Sc}); subsequently, measurement of the two fractions is accomplished by a non-competitive immunoassay using a DELFIA detection technique. The result is calculated as the percentage of insoluble PrP of the total PrP.
The samples (brain stem, cervical spinal cord) are turraxed in a buffer. A chaotrophic agent is added to an aliquot of the homogenate. After centrifugation the soluble fraction is collected. Again chaotrophic agent is added to the remaining pellet at a higher concentration than previously to solubilise the insoluble PrP fraction. Both fractions are applied to a microtitre plate pre-coated with a monoclonal capture antibody and incubated overnight. After a washing step, the europium-labelled monoclonal detection antibody is applied. After incubation for 2 hours at room temperature, the plate is extensively washed. Detection is accomplished by adding DELFIA enhancement solution, incubation at room temperature for 5 minutes and measurement by time-resolved fluorescence. Three in-house controls were run on each plate, one positive, prepared from a pool of positive samples (received from IRMM in 1999), one negative from a pool of negative samples (received from IRMM in 1999), and a sample prepared by titrating the positive pool in the negative such that a result of about 8% was obtained (value of the cut-off). Human platelet-enriched plasma was used to establish a calibration curve. Since the calibration solutions do not contain the same amount of reagents, especially with regard to the chaotrophic agent, the units are arbitrary and are defined in units per ml (U/ml).
Each sample was analysed in duplicate; the mean value of the 2 measurements was taken into account for calculating the final results (% insoluble PrP of total PrP).
Carried out in the way mentioned above the test takes approximately 17 hours including an over night incubation (12 hours).

Decision criteria (cut off definition)

The cut-off was established analysing about 10000 samples collected within the OTMS-2 scheme and fixed at 8% (% insoluble PrP of total PrP). However it was mentioned that the quality of the material (storage temperature, age) and the choice of the internal reference samples have a substantial influence on the cut-off as degradation will most likely occur in a discriminative manner. No inconclusive category was defined for repeating measurements. However duplicates exhibiting more than 20 % standard deviation are re-analysed.

Sensitivity and Specificity

	True positive	**True negative**	**Totals**
Test Positive	48	1*	49
Test negative	0	151	151
Totals	48	152	200

	%	95% Confidence Interval (one sided Poisson)
Sensitivity	100	93.8 %
Specificity	99.3*	96.9 %

* one false positive. Separation of the positive and negative data population can be achieved by moving the cut off (see below)

Samples with false classification

Sample 7004 is a false positive showing a value of 8,1%. The false positive showed a mean of 8,1% with 0,5% STDEV.

Data distribution

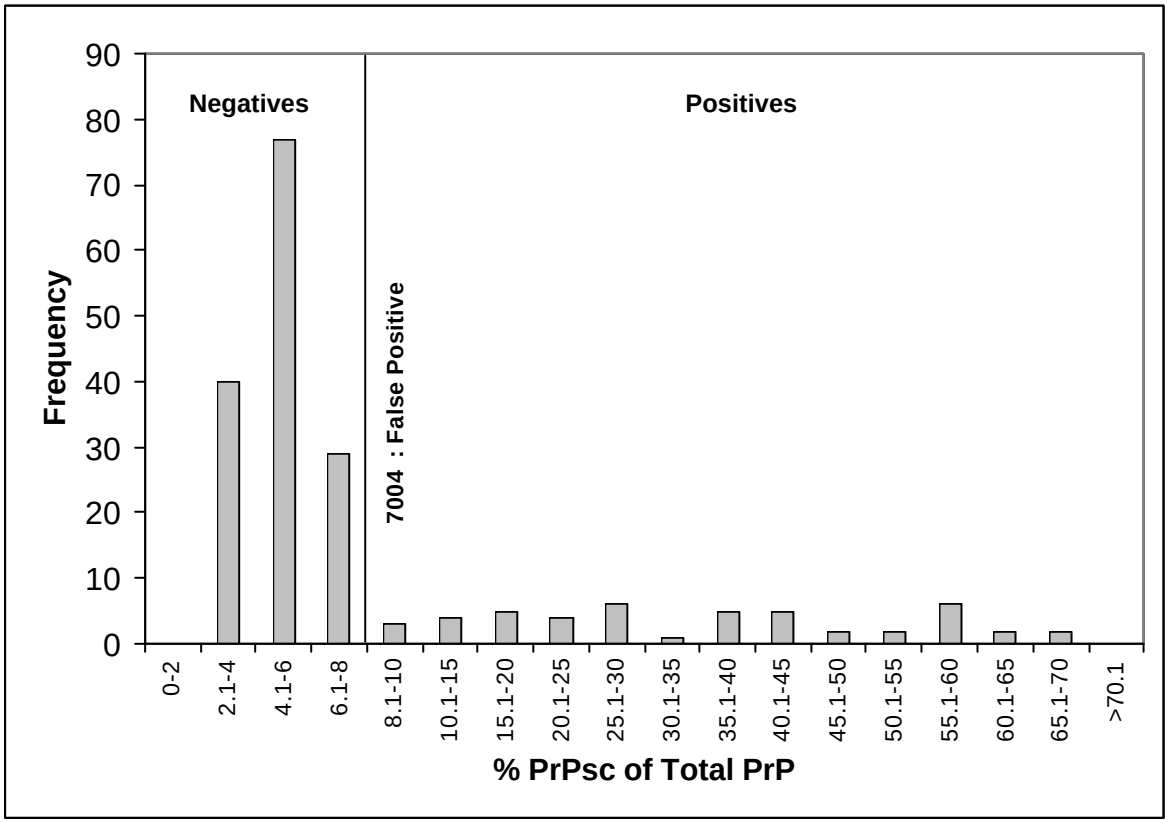


Fig 6: PerkinElmer Life Sciences test - results for positive and negative samples

The data populations for positive and negative samples are close together and differ only by 1.4 %. If three times the standard deviation is added to the mean for the

negative sample an upper value of 8.7 is obtained meaning that 99 % of the negative samples are expected to fall below that value. However if 5 times the standard deviation is added to the mean of the negative samples (frequently done for the definition of the cut off value) a value of 11.2 is obtained falling already into the range of positive samples. The eight highest negative samples gave a reading of 8.1, 7.9, 7.8, 7.8, 7.7, 7.5, 7.2 and 7.0 %. The 7 lowest positive samples gave a reading of 9.5 (slice 2.6, see cutting scheme in annex 2), 9.9 (slice 2.3), 10.4 (slice 2.4), 11.4 (slice 2.6), 11.9 (slice 2.4), 12.3 (slice 2.6) and 15.1 % (slice 2.2). In other words 6 of the 152 negatives samples lay within 0.9 times the cut-off value (7,2%), 4 of these within 0.95 times the cut-off value (7,6%). Two of the low positive samples (samples giving 9.9 and 10.4 %) belonged to the group of three samples which tended to give lower results in the 1999 evaluation [1]. The third sample gave 19.3 %. On the other side other randomly selected samples gave comparably low results (such as 9.5, 11.4 etc.).
In fact by a redefinition of the cut off e.g. at 8.7 % (mean of negative samples + 3 SD) the positive and negative population can be separated. However since it is expected (see above) that degradation or ageing processes of samples may have an influence on the cut off additional data would be desirable for the redefinition of the cut off. Those data should be obtained with samples being in its properties closer to what is to be expected in the field (i.e. samples not having been stored over prolonged periods at low temperature).

Average of negative and positive data populations, insoluble and soluble PrP fractions

	Average in % insoluble PrP / total PrP	SD	Average Soluble fraction U/ml	SD	Average Insoluble Fraction U/ml	SD
Positive	36.8	19.4	96.5	51.3	79.3	84.2
Negative	4.86	1.26	86.8	28.1	4.49	1.89

Dilution Series

Two different dilution series were tested to determine the detection limit of the test.
(a) Dilutions of homogenates in the range of 10^{-1} to 10^{-3} prepared at IRMM in 2001.
(b) Dilutions of homogenates prepared freshly on site with the same mouse titrated positive brain homogenate used to produce series (a) (another sub-batch not containing any buffer) and a pool of 5 negative brain stem material originating from areas posterior to the obex. Equal amounts of the five brain stems were collected, homogenised and mixed with the positive material in serial 1:3 dilution steps. The samples were coded and measured as described above.

Series (a) prepared by IRMM:

Dilution	Mean	Standard deviation
Undiluted	95.4	-
10-1	7.6	0.5
10-1.5	5.4	0.2
10-2	5.0	0.5
10-2.5	4.8	0.3
10-3	4.5	0.4
Negative controls	5.2	0.2

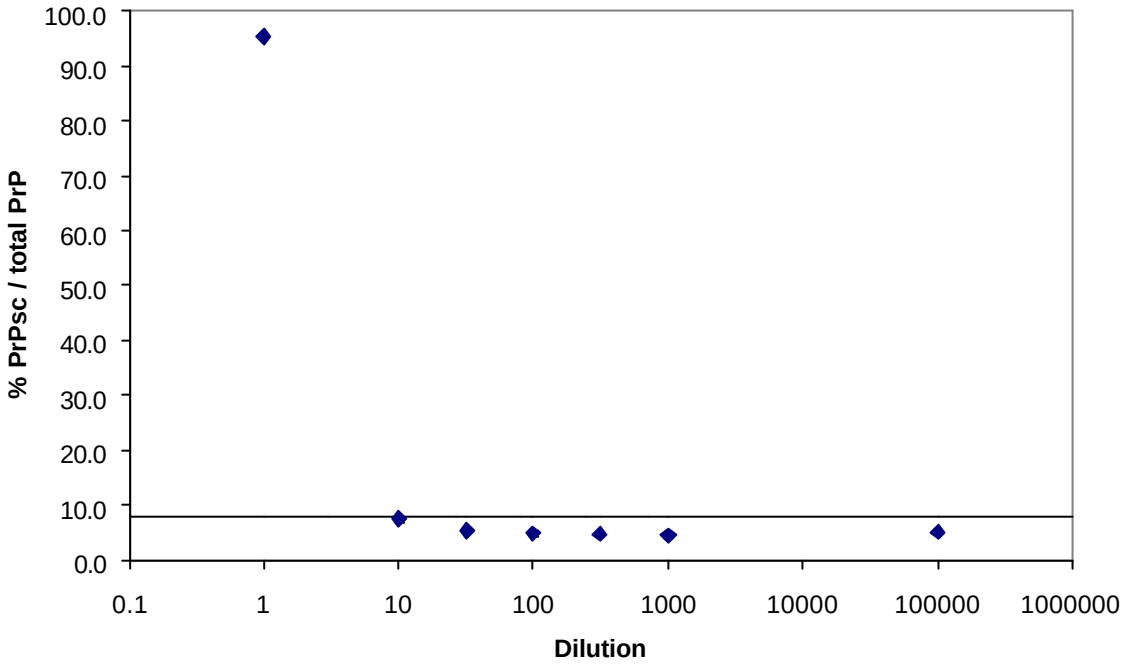


Fig. 7: Dilution series (a) prepared by IRMM (negative controls are shown as a 1:100000 dilution)

The 1:10 dilution was around cut off. One replicate at that dilution level was just above cut off (8.07 %) whereas the three others were below cut off (7.83, 7.57 and 6.83 %)

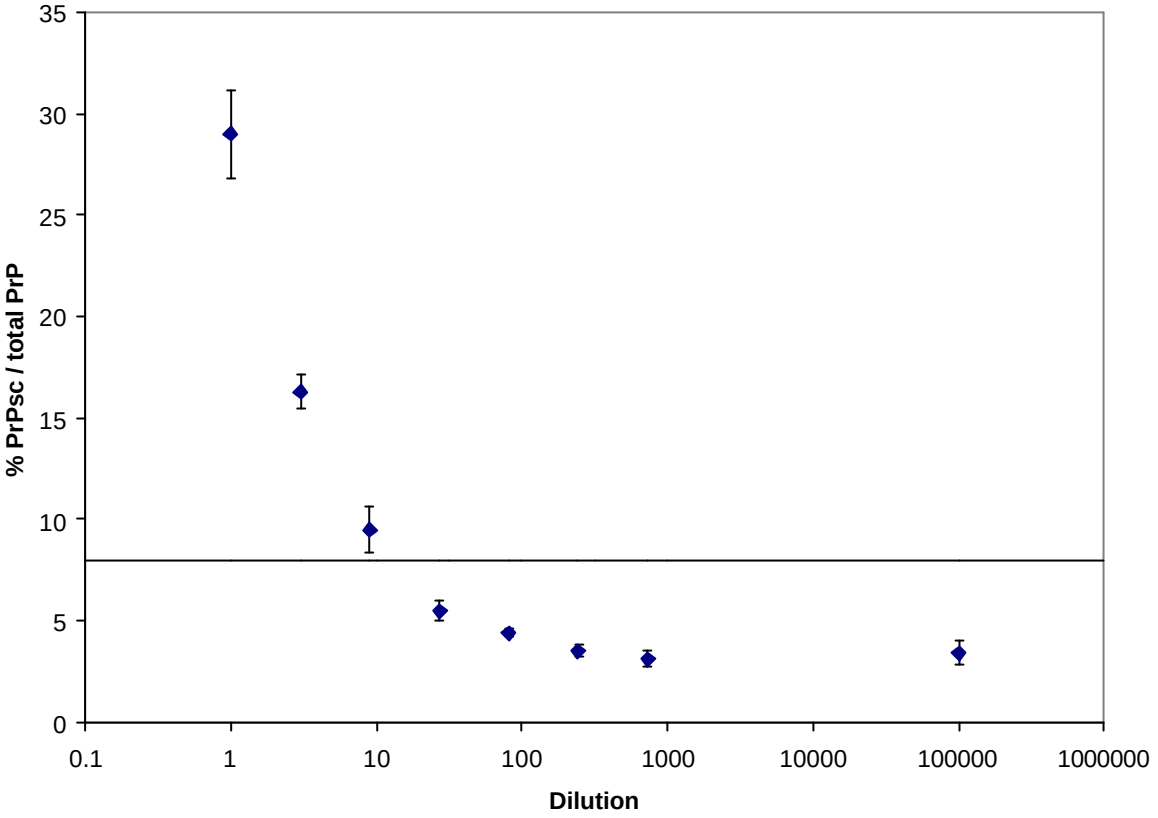


Fig. 8: Dilution series (b) prepared according to test developer's procedure (negative controls are shown as a 1:100000 dilution)

Series (b) prepared according to test developers procedure:

Dilution	Mean	Standard deviation
Undiluted	29	2.2
1:3	16.3	0.8
1:9	9.5	1.1
1:27	5.5	0.5
1:81	4.4	0.2
1:243	3.5	0.3
1:729	3.1	0.4
Negative controls	3.4	0.6

The 1:9 dilutions were clearly recognised as positives, the 1:27 dilutions were significantly higher than the negatives (5,5 ± 0,5% vs. 3,4 ± 0,6%), but the value itself suggested a clear negative (cut-off of 8%).

There is a good agreement between both dilution series. Series (b) is indicating a slightly lower detection limit (approximately 1:15 dilution compared to 1:10 with series (a)). However the starting point (undiluted positive) behaved differently

although the undiluted positives originated from the same pool. Whether the phenomenon is an analytical artefact or a true sample effect could not be determined. The effect is also addressed in chapter 7.7.

Although the % insoluble PrP / total PrP level seemed to be higher in the undiluted positive material used for series (a) prepared by IRMM the high ratio is caused by a low level of insoluble and even lower level of soluble PrP compared to the undiluted material used for series (b). Only about 1 unit/ml soluble PrP was measured in the material used for series (a) compared to 285 units/ml in the material used for series (b). Also a lower level of insoluble PrP namely only about 20 units/ml insoluble PrP compared to 118 units/ml in the material used for series (b) was measured.
Since the test result is expressed as a ratio between insoluble PrP and total PrP the detection limit is influenced by the level of soluble PrP in the negative pool used for the dilution and by the fraction of insoluble PrP in the starting positive material. The higher the level of soluble PrP using the same positive starting material the lower the dilution factor of the samples laying above cut off. The soluble PrP fraction measured in the negative samples of series (a) was at 126 units/ml whereas the mean of the negative slices was 86.8 units/ml with 39.8 units/ml being the lowest reading and 171.8 units/ml being the highest. The negative pool used for the dilutions is therefore in average leading to a slight overestimation of in average a factor of 1.5. The mean of the negative samples in series (b) was 52.6 units/ml soluble PrP (mean of 4 samples, insoluble PrP^{sc} fraction at 1.9 units/ml, mean insoluble PrP 3.4 %) i.e being somewhat lower than an average negative sample and giving less conservative estimates. The data for the 1 in 10 respectively 1 in 9 dilutions of series (a) and (b) correspond to the expected values on the basis of the negative and positive pools demonstrating consistency of data. The discrepancy between the absolute readings obtained at different occasions for the positive pool used for the preparation of series (b) indicate difficulties in the reproducibility of the calibration, however if expressed in relative % insoluble PrP / total PrP the values are the same.

	% insoluble PrP / total PrP	Soluble PrP	Insoluble PrP	No samples	No replicates per sample
Positive slices	36.8	96.5	79.3	48	1
Negative slices	4.9	86.8	4.5	152	1
Positive pool series (a)	95.4	1.1	22.8	1	3
Negative pool series (a)	5.2	126	6.9	5	3
1:10 dilution series (a)	7.6	116	9.6	4	3
Positive pool series (b) [1]	29.2	285	118	1	3
Positive pool series (b) [2]	29	121	49.8	4	1
Negative pool series (b)	3.4	52.7	1.9	4	1
1:9 dilution series (b)	9.5	76.6	8.0	4	1

1: Sample measured together with samples of series (a)
2: Sample measured together with samples of series (b)

The results of the dilution series are also in line with the closeness of the data populations of negative and positive samples (see above).

7.3 Test C (Prionics-Check LIA) Prionics A.G.

This sandwich ELISA system (Prionics – Check LIA) uses two different monoclonal antibodies to detect Proteinase K resistant PrP fractions. Homogenates of brain tissues are prepared by vigorous mixing in buffer and then subjected to proteolytic treatment. The immunoassay is directly applied to the digested homogenates without further clean up or centrifugation reducing the measurement time considerably. Following an incubation step with the first enzyme linked antibody the protein – antibody complexes are transferred into a new microtitre plate coated with a second monoclonal antibody directed against PrP. The complexes are bound to this capture antibody and become accessible for detection by a chemiluminescence producing reaction. The system is designed in a 96 well format and can be fully automated from homogenisation to detection and analysis.
The test is performed in a convenient 96 well format, results can be obtained after 4 hours. The ELISA technique enables full automation. Raw data obtained from chemiluminescence measurements are imported into a program that calculates the cut-off value and assigns the samples to be negative, positive, weak positive or suspect. In the latter cases a confirmation with an alternative test system is recommended.

Decision criteria (cut off definition)

The cut-off value is calculated as 5 times the mean of 4 negative controls measured in duplicate on each 96 well plate. The negative controls consist of a pool of negative samples having been homogenised and digested according to the test protocol. The digested pool is kept as negative control. A 'grey zone' around the cut-off of +/- 25% is defined and used to identify suspect and weak positive samples that need to be subjected to further testing (normally Prionics-Check Western Blot or concentration Western Blot including a centrifugation step for the enrichment of PrP^{res}).
If both replicates are above the grey zone the sample is classified as positive. If one replicate is above the grey zone and one in the grey zone the sample is rated as weak positive. Samples giving either two values in the grey zone (either both above cut off or one above and the other replicate below cut off) or one value in the grey zone above cut off and the second replicate below the grey zone are classified as suspect and are repeated. If the first replicate is laying in the grey zone below cut off and the second replicate below the grey zone or if both replicates are below the grey zone the sample is classified as negative and not repeated.
Even though the test had been designed as a single well test, the evaluation was carried out in duplicate.

Sensitivity and Specificity

	True positive	**True negative**	**Totals**
Test Positive	47	0	47
Test negative	1*	152	153
Totals	48	152	200

Results of the third measurement series (for other series see text below)

	%	95% Confidence Interval
Sensitivity	97.9*	90.1 %
Specificity	100	98.0 %

* false negative sample which was tested positive when repeated upon request (if either the residual homogenate or tissue was used) and which may be considered as one of the exceptionally low sub-sample taking the aspects summarised in chapters 1.4, 7.8 and the detection limit as summarised below into consideration

Repeated samples

In total three measurement series were performed with all 200 sliced samples. The results of the first series for negative samples including the cut off were lower than expected (at background level at about 60 chemiluminescence units compared to normally 100 – 200, cut off 296 to 436 units, normally between around 1000) causing two false negative results and one suspect (wrongly classified through software as negative) indicating technical difficulties. The technical problems were realised by the test developer completely on its own without disclosure of the sample status. Assuming that the results were caused by low detection antibody concentration the concentration level was doubled (instead of 1 in 1000 a dilution of 1 in 500 was applied). The results were somewhat higher and closer to the expected levels but highly fluctuating (cut off values between 1165 and 3092 units). During the repetition of the measurements the originally used batch of chemiluminescence substrate ran out and using the new batch results were obtained more corresponding to the normal range. Therefore the whole series of measurements has been repeated a second time using the originally foreseen antibody concentration and the new substrate batch (the test developer confirmed later on that the substrate batch used at the beginning of the experiments had been defective. The test developer expects that such a situation can be prevented if the substrate will be included in the final diagnostic kit ensuring respective quality control). The signal level corresponded consequently to the expected levels above (cut off 616 to 2261, negative samples around 150 units) and in addition an antibody concentration has been applied as specified in the test protocol submitted. Consequently the outcome of the third measurement series is reported the data being based on measurements performed with a new batch of substrate for the chemiluminescence reaction. Extension of the delay before measurements form 2-10 minutes to 45 minutes, other than stated in the protocol, led values for negative samples generally to drop considerably thus resulting in a better separation of negative and positive populations. This incident stresses the importance of quality control measures during the production and before release of test kit batches.

Samples with false classification

Sample number 5634 has been wrongly classified as a negative sample in the third series of measurements. One of the duplicate values reached the 25% grey zone below the cut-off (the second replicate was clearly below cut off). The sample was scored negative on bases of the predefined decision criteria.

If the decision criteria would be modified such that also samples which show one replicate in the grey zone below cut off and the second replicate below the grey zone would be classified as suspect and would be further investigated the sample would have become in the repeat positive. In fact for scientific interest a repetition of sample 5634 and surrounding samples (to keep the test blind) has been requested by IRMM using the residual homogenate and tissue. Both repeats turned out weak positive (signal relative to cut off; homogenate: 1.43 and 0.78; tissue: 1.62 and 0.80) thus leading to the recommendation of a further inspection of the sample. However two other negative sample showed one replicate in the grey zone below cut off (sample numbers 4964 (0.76) and 4994 (1.0)) and it cannot be predicted how the results of a repetition would look like.
It is also worth mentioning what how the 3 weakest samples (none identical with those which tended to give lower results in the 1999 evaluation [1]) behaved under the different conditions (see table below).

Sample	Conditions	Replicate 1 Relative to cut off	Replicate 2 Relative to cut off	Classification
4044	AB 1:1000, Old substrate	4.57	0.57	Suspect
	AB 1:500 New substrate[1]	122	186	Positive
	AB 1:1000 New substrate	40.1	77.1	Positive
4254	AB 1:1000, Old substrate	0.88	0.62	False negative
	AB 1:500 Old substrate[1]	112	8.94	Positive
	AB 1:1000 New substrate	15.6	8.66	Positive
5634	AB 1:1000, Old substrate	0.79	0.30	False negative
	AB 1:500 Old substrate[1]	1.71	0.51	Suspect
	AB 1:1000 New substrate	0.85	0.18	False negative

1: Samples 5634 and 4254 have been measured with the older batch of substrate whereas sample 4044 was measured with the new batch.

The data show that sample 4044 being a borderline case in the first measurement series and sample number 4254 appearing to be negative in the first series become clearly positive in the following series. Sample 5634 appears negative in the first series, becomes a suspect sample (erroneously classified by the computer program as negative) if the two fold antibody concentration is applied and is classified as a false negative when the foreseen antibody concentration was applied and the new substrate.

Data distribution (3rd series)

The five highest negative sample gave a mean of two replicates relative to cut off of 0.596, 0.608, 0.625, 0.628 and 0.694. The positive sample 5634 gave a mean signal relative to cut off of 0.517 (slice 2.4, cutting scheme see annex 2). The five other lowest positives gave a mean relative to cut off of 4.85 (slice 2.5), 10.8 (slice 2.3), 12.14 (slice 2.5), 30.36 (slice 2.5) and 49.3 (slice 2.6). Therefore with the exception of sample 5634 the system differentiates clearly between negative and positive cases.

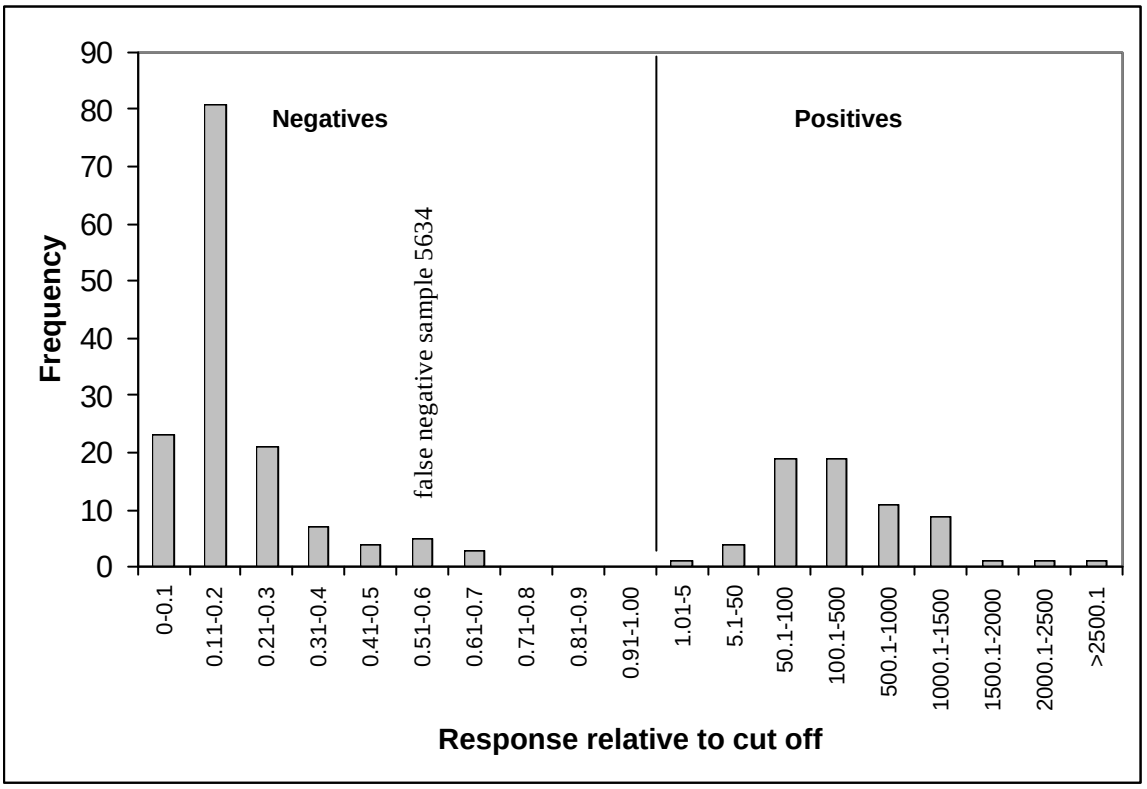


Fig 9: Prionics Check LIA test - results for positive and negative samples

As already mentioned above only one of the 3 positive samples which tended to give low results in the 1999 evaluation [1] is amongst the samples giving the lowest signals (10.8 relative to cut off). Regarding also the results obtained with the diluted homogenates sample 5634 may therefore be considered as one of the exceptionally low sub-samples similar to those which appeared in the 1999 evaluation [1] (see chapter 1.2) and which are explained by heterogeneity of PrP^{sc}.

Average of negative and positive data populations (3rd series)

	Average relative to cut off	SD
Positive	609	649
Negative	0.20	0.12

Dilution Series

Two different dilution series were tested to determine the detection limit of the test.
(a) Dilutions of homogenates in the range of 10^{-1} to 10^{-3} prepared at IRMM in 2001.
(b) Dilutions of homogenates prepared freshly on site with the same mouse titrated positive brain homogenate used to produce series (a) (another sub-batch not containing any buffer) and a pool of 5 negative brain stem material originating from areas posterior to the obex. Equal amounts of the five brain stems (about 0.5 g) were collected, homogenised and mixed with the positive material in serial 1:3 dilution steps. The samples were coded and measured according to the standard protocol.

Series (b) has been measured together with the slices. For the measurement series summarised below the standard testing protocol has been applied equivalent to the third measurement series of the slices.

Series (b) prepared according to test developer's procedure:

Dilution	Mean	Standard deviation
Undiluted	Not measured	-
1:3	348.29	33.42
1:9	62.98	7.34
1:27	14.97	2.56
1:81	3.16	0.44
1:243	1.57	0.38
1:729	1.46	0.35
1:2187	1.03	0.16
Negative controls	0.88	0.08

The average of the negative controls is just below cut off. However all 8 replicates of the four negative control samples were laying in the grey zone (± 25 % around cut off). One out of four negative samples shows one replicate in the grey zone above cut off whereas the other replicate is below cut off but still in the grey zone. According to the decision criteria the sample should be classified as suspect. The reason for the high signals of the negative samples was not found but since the same pool was used to prepare all dilutions all samples will show signals which are higher than expected. The negative controls can be expected to show readings close to the mean of the negative slices i.e. 0.2 relative to the cut off. This means that the mean for the negative controls is about 0.6 units relative to the cut off too high. In order to estimate

a realistic detection limit of the test this contribution to the elevated baseline could be subtracted or the cut off could be elevated form 1 to 1.6. Consequently the test has with a high probability the potential to detect approximately a 1:240 dilution (the 1 in 243 is just below the “elevated cut off” of 1.6).
However the example with the dilution series (a) as summarized below shows that the detection limit is significantly influenced by the sample behaviour and modifications of the procedure and therefore the afore mentioned can only be considered as an estimate and may either higher or lower under conditions giving readings in the expected range for negative samples.

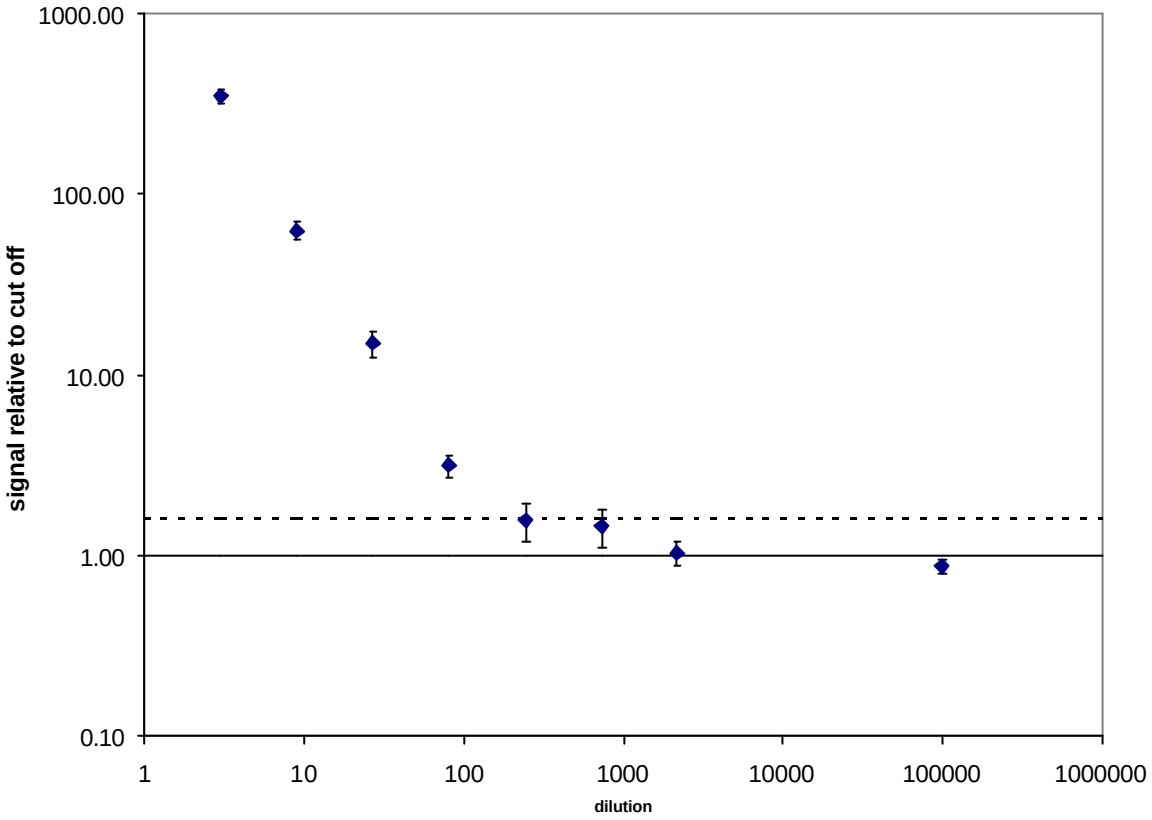


Fig 10: Dilution series (b) prepared according to test developer’s procedure (negative controls are shown as a 1:100000 dilution)

The series (a) has been measured at an earlier occasion and therefore the technical problems with the substrate did not occur. However it appeared that if the homogenate samples (80 % tissue) as prepared by IRMM are analysed according to the standard protocol that all samples including the negative ones are a factor 11 to 18 above cut off i.e. according to the decision criteria positive. If the same test set up was applied to fresh negative samples only signals in the expected range were observed. An explanation for this phenomenon could not be found but the effects observed are clearly related to the way the material was produced rather than the performance of

the test. In the field of clinical testing effects where test samples do not behave like fresh field samples are summarised under the term commutability. Commutability problems can appear in cases with artificial or highly processed samples (including freeze drying) but also occasionally with samples which have been only frozen.
The digestion of the homogenates was not complete which could be even visually seen because of increased turbidity of the digested solutions. Readings up to 40.000 chemiluminescence units indicate that as much as 1 % of the PrP^c remains undigested since the signal of an undigested sample is in the range of 4.000.000 chemiluminescence units. The standard deviation on the results is also very high indicating that the incomplete digestion is not reproducible. A Western Blot was carried out with these samples showing the typical bands for undigested PrP^c (two bands above and below the typical more intense band in the middle).

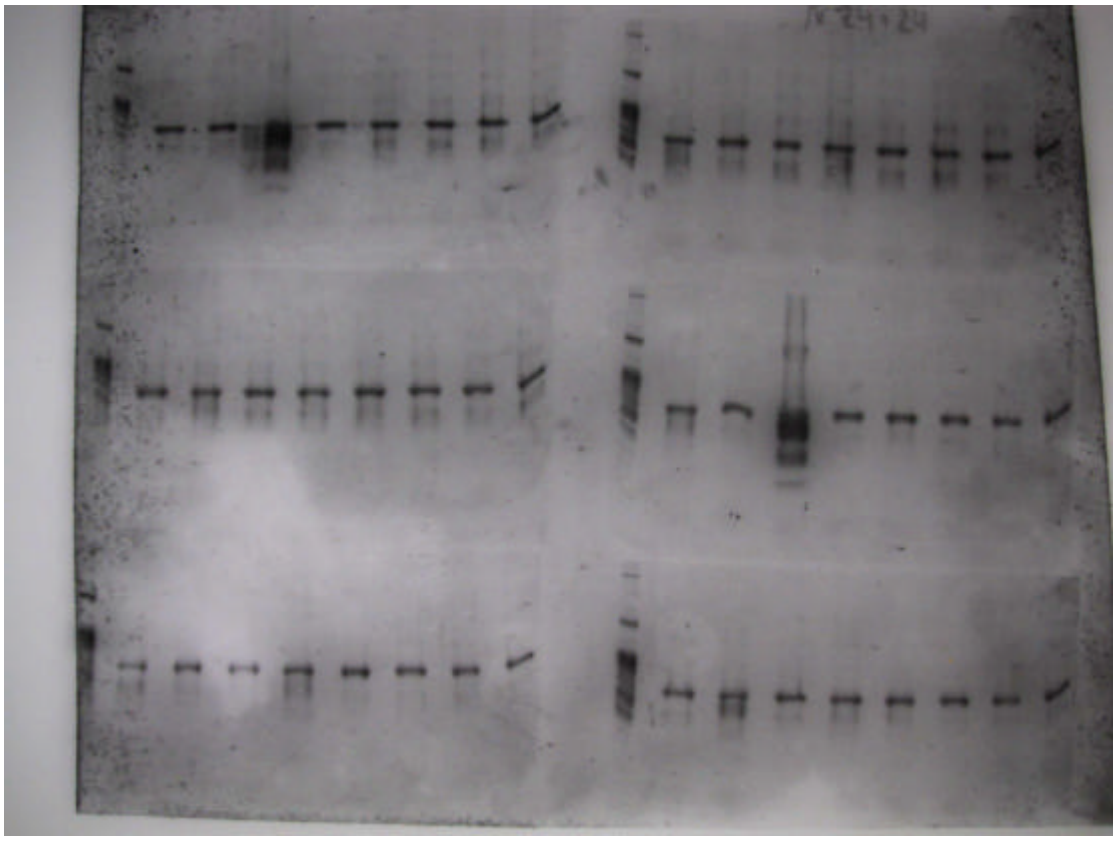


Fig. 11: Western Blot of homogenate series (a) produced by IRMM

These bands have also been observed in the 1999 evaluation [1] of the Prionics Check Western Blot.

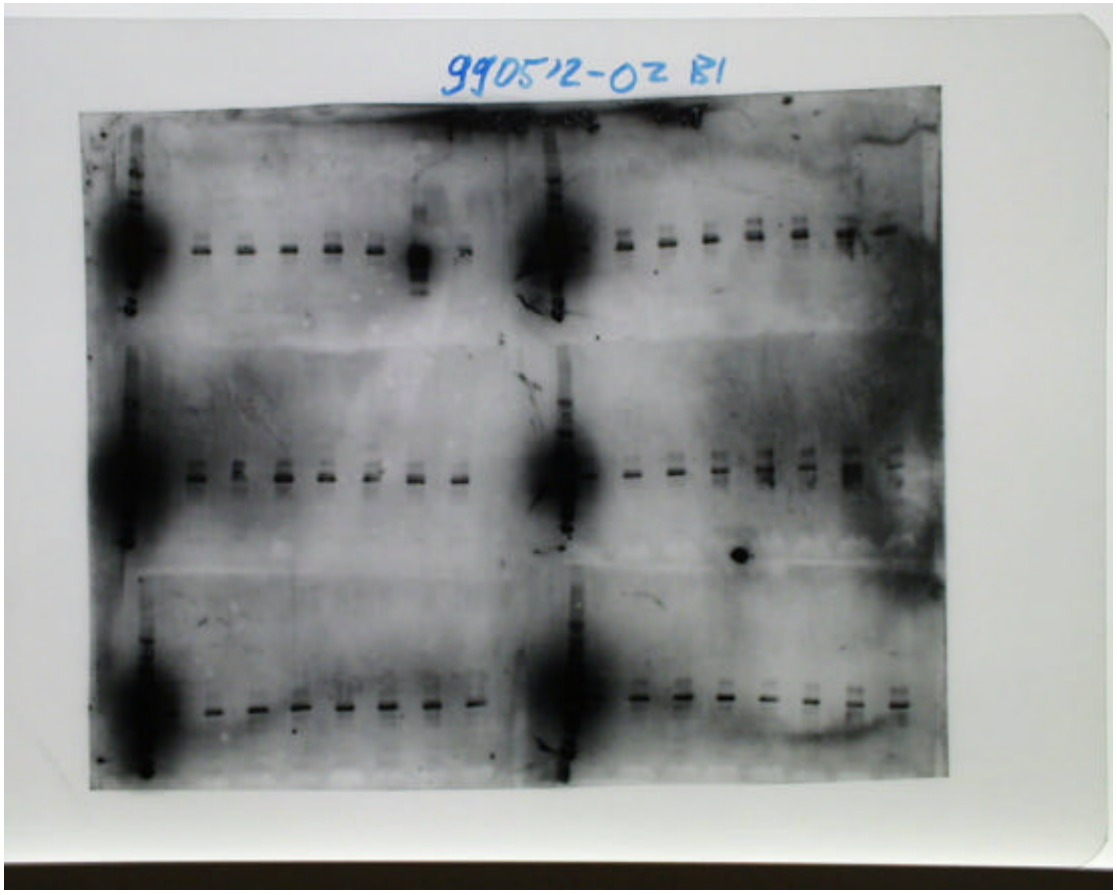


Fig. 12: Western Blot of homogenate series produced by IRMM and analysed in 1999

Hence one can conclude that the commutability problems are related to the homogenisation procedure rather than to the prolonged storage of the homogenates or the raw materials used for their production.

All results for series (a) samples including those for negative samples are above cut off using the standard conditions and therefore cannot be used for the assessment of the detection limit.
Because of the observed commutability problems the test developer was allowed to modify the procedure such that complete digestion is possible. Based on past experience it was suggested to dilute the homogenates by a factor of two (5 % tissue homogenates instead of 10 % homogenates). A dilution of 1:10 could be detected under these conditions. Besides one out of four samples the 1:32 dilutions gave signals slightly below cut off. The signals for the higher dilutions were clearly below cut off. The level of the negative controls corresponds to the mean of the negative sample slices (signal 0.20 relative to cut off).
However the test developer was concerned that the digestion conditions are too stringent leading to partial digestion of PrP^{sc} and therefore to an increase of the detection limit. Consequently the series (a) was once again measured with the originally foreseen tissue concentration in the homogenates (10%) and 2 hours incubation time for the digestion instead of one hour in the standard protocol. 4 out of 4 samples at 10^{-2} dilution level were detected positive. All 4 samples at $10^{-2.5}$ dilution

level were classified negative. The negative controls are close to the mean value of the negative slices (0.2 relative to cut off).

Series (a) prepared by IRMM:

Incubation time for digestion in hours	1		1		2	
% tissue in homogenate	10		5		10	
Dilution	Mean relative to cut off	SD	Mean relative to cut off	SD	Mean relative to cut off	SD
Undiluted	809	439	842	12	1499	0.4
10-1	56	23	8.0	2.3	88	19
10-1.5	17	11	0.80	0.19	11.9	2.2
10-2	11	8	0.36	0.06	1.92	0.33
10-2.5	14	10	0.26	0.04	0.59	0.20
10-3	11	9	0.24	0.05	0.26	0.03
Negative controls	18	13	0.35	0.08	0.27	0.07

In the case of non commutable samples the variation of the test protocol has a strong influence on the detection limit. Whereas with homogenates containing 5 % tissue the detection limit is at about a 1 in 30 dilution it is approximately 1 in 200 for the homogenates containing 10 % tissue but digested two hours instead of one in order to reduce the background of protease K resistant PrP^c.
The detection limit of 1:200 with homogenates containing 10 % tissue and 2 hours incubation time for digestion is also in good agreement with the detection limit of 1:240 obtained with series (b).

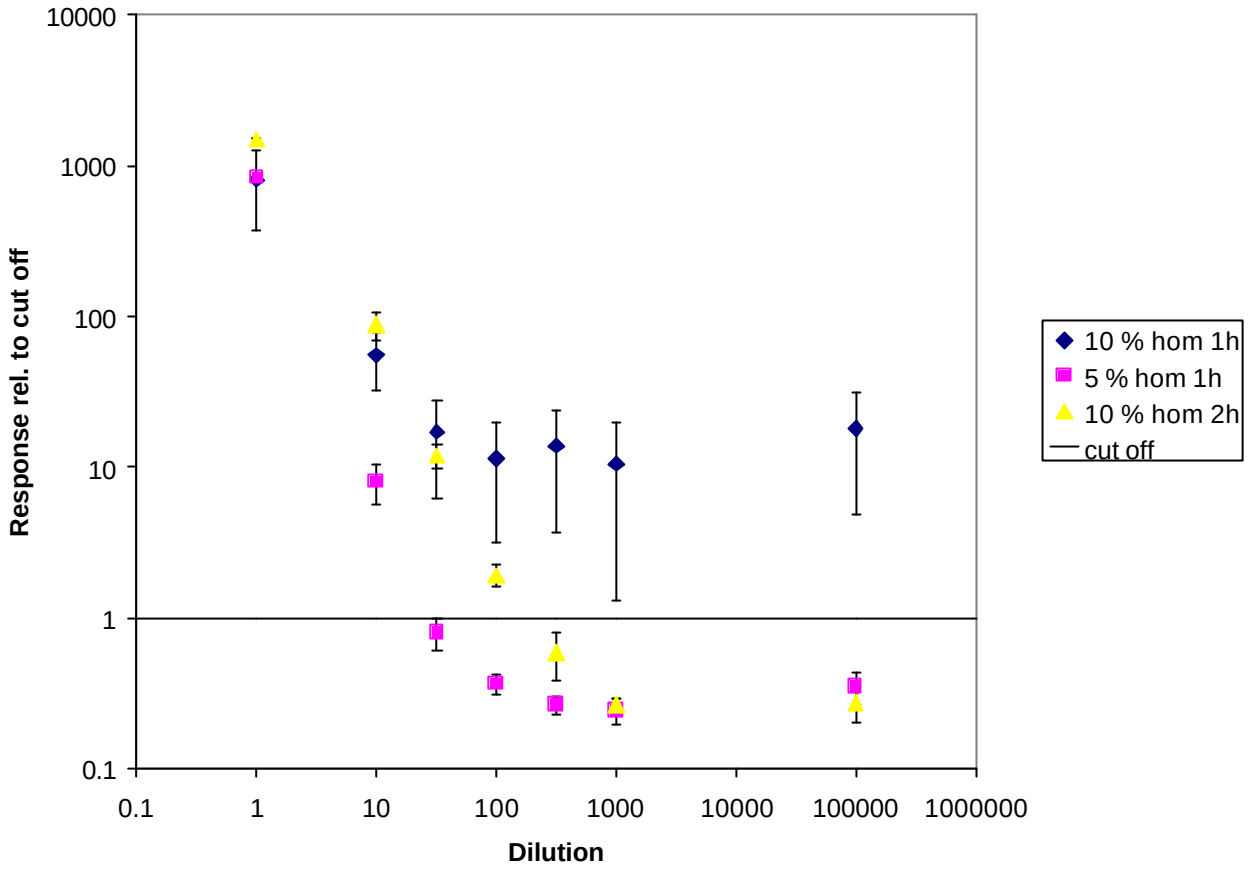


Fig 13: Dilution series (a) prepared by IRMM (negative controls are shown as a 1:100000 dilution)

Prionics Check Western Blot

Dilution series (b) has also been analysed with the Prionics Check Western Blot already evaluated in 1999. Due to the incomplete digestion of PrP^c giving an interfering band only 1 in 10 dilutions could be unambiguously identified as positive. Since after the digestion of samples from series (b) a lower level of PrP^c remains undigested the samples give a better estimate of the detection limit compared to those used in 1999. The Western Blot is shown in Annex 4.

Number of homogenate samples prepared according to test developer's protocol scoring positive (above cut-off) at each dilution level.

	Test C
Titre $10^{3.1}$	-
1:3	4/4
1:9	4/4
1:27	4/4
1:81	$3(1^1)/4$
1:243	$(1^1)/4$
1:729	$(1^1)/4$
1:2181	0/4
Negative	0/4

1: classified as suspect

7.4 Test D (CDI) UCSF

This test is a sandwich immunoassay with time resolved fluorescence detection. It is based on the different affinity of the detection FAB to the conformations of normal PrP^{c} and abnormal PrP^{sc}. Two recombinant FABs are used. The epitope where the detection FAB binds to is accessible in normal PrPc but inaccessible for the FAB in undenatured abnormal PrPsc. The procedure consists of a homogenisation step in a test buffer, incomplete digestion with a low quantity of protease K, enrichment of PrP^{sc} and splitting of the sample in two aliquots of which one is treated with a chaotrophic agent (D sample) the other is not (N sample). The PrP is bound on microtitre plates which are coated with the recombinant capture FAB just before the test is performed. After incubation the labelled detection FAB is added. The pipetting steps within the procedure including removal of supernatant after centrifugation have been automated with a robot system. Each sample slice is cut in at least four sample aliquots (approximately 0.25 to 0.35 g each). The slices from locations posterior to the obex are cut rectangular to the length axis of the brain stem whereas the slices anterior to the obex are cut from the centre to the peripheral parts of the brain stem. The sample sequence is A, C, B, D. Samples A and B are taken for analysis whereas samples C and D are kept for verification purposes. From each sub-sample two replicates under native and denaturing conditions are performed i.e. per sample 8 wells on the microtitre plate are used. Hence per microtitre plate 11 samples can be analysed.
In its current state of development the test takes less than 8 hours to complete. The procedure is automated using a robot for all pipetting steps including removal of supernatant after centrifugation and allowing full sample traceability through the testing process.

Decision criteria (cut off definition)

The cut-off value is defined as 2400 cpm for the normalised difference between the D and the N sample (TRF_D-TRF_N). The normalised difference means that from the difference of the means of both replicates of the D and the N sample (per sub-sample) the reading for the negative control has been subtracted. An inconclusive range has also been defined for the normalised difference between 2400 and 4572 cpm. Samples falling in this range are classified as borderline high.
The following reasons for repetitions were defined:

- If (sub-sample) duplicates display discordant readings (one above and one below cut off)
- Samples where one or both replicates have a borderline high
- If quality control indicated machine or human error, either documented or suspected from discordant duplicates of N or D sample reading, or from presence of outlier

If both results of the repeated test (carried out with sub-sample C and D and not with the homogenate of sub-samples A and B) are negative, the sample is classified as BSE negative. If one or both replicates of the sample is positive and the sample tested previously tested positive, the sample is classified as BSE positive.

Sensitivity and Specificity

	True positive	True negative	Totals
Test Positive	48	0	48
Test negative	0	152	152
Totals	48	152	200

	%	95% confidence limit One sided Poisson
Sensitivity	100	93.8%
Specificity	100	98.0%

Repeated samples

14 out of 200 samples had to be repeated in accordance with the above criteria. All samples concerned were true negative. The results are summarised below:

	1st run					Repetition				Concl.
Sample	TRF_D-TRF_N		TRF_D-TRF_N		Plate	TRF_D-TRF_N		TRF_D-TRF_N		
8644	2938	Bl+	1446	-	5	1921	-	95	-	-
8694	5404	+	-222	-	6	840	-	241	-	-
9114	2619	Bl+	1205	-	10	282	-	442	-	-
9124	3231	Bl+	1833	-	10	192	-	-45	-	-
9164	2608	Bl+	1243	-	10	205	-	262	-	-
9184	6076	+	2800	Bl+	10	461	-	639	-	-
9244	2408	Bl+	2241	-	11	89	-	400	-	-
9284	3321	Bl+	1143	-	11	146	-	441	-	-
9314	4151	Bl+	4753	+	12	1034	-	1366	-	-
9334	2722	Bl+	1113	-	12	934	-	1047	-	-
9574	2412	Bl+	1005	-	14	328	-	36	-	-
9624	4208	Bl+	1787	-	14	141	-	311	-	-
9634	2590	Bl+	850	-	14	-267	-	201	-	-
9674	6108	+	-467	-	15	-1458	-	-1426	-	-

Bl+: borderline high positive

In total 10 samples had a borderline high positive and negative replicate. 2 had a borderline high positive and a positive replicate and 2 others had discordant positive and negative replicates. After repetition all 14 samples were correctly classified as negatives.

It is worth to mention that 11 out of 14 samples to be repeated were located on 4 plates out of 18. 7 samples to be repeated were located on two plates only. All four plates concerned were run on the second day. The total of 200 samples was analysed on three days, 77 samples on the first (two samples to be repeated), 75 samples on the second (11 samples to be repeated) and 48 samples on the third day (one sample to be repeated). The accumulation of samples to be repeated on one day and 4 plates only

may indicate a technical problem in the measurement procedure which might be overcome by improved quality assurance measures.

Samples with false classification

All samples were correctly classified.

Data distribution

For this evaluation the data of the results of the 14 repeated samples have been considered and the data of all other samples which did not have to be retested.

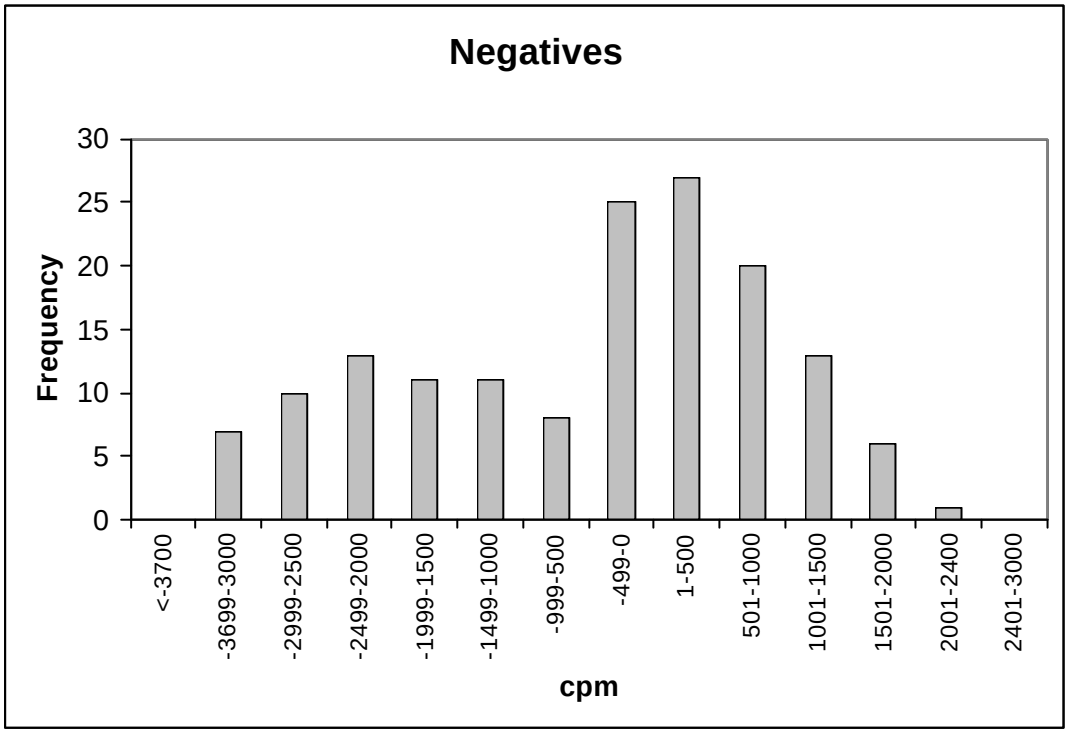


Fig 14: UCSF CDI test – negative samples (mean of two replicates)

There is a good separation between the positive and negative data populations. The five highest negative normalised differences TRF_D-TRF_N (mean of two replicates per sample) were 2098, 1956, 1952, 1800, 1714. The five lowest positive normalised differences TRF_D-TRF_N were 18592 (slice 2.4, cutting scheme see annex 2), 45535 (slice 2.5), 46217 (slice 2.6), 60163 (slice 2.6) and 60506 (slice2.5). Only one of the 3 positive samples which tended to give low results in the 1999 evaluation [1] is amongst the samples giving the lowest signals (60506). In principle it would be possible to increase the cut off, thus requiring less repeats. However this would have an impact on the detection limit and would increase the probability that weak positive samples would not be detected.

Since the results are based on two independently processed sub-samples also the distribution of the individual results is given below.

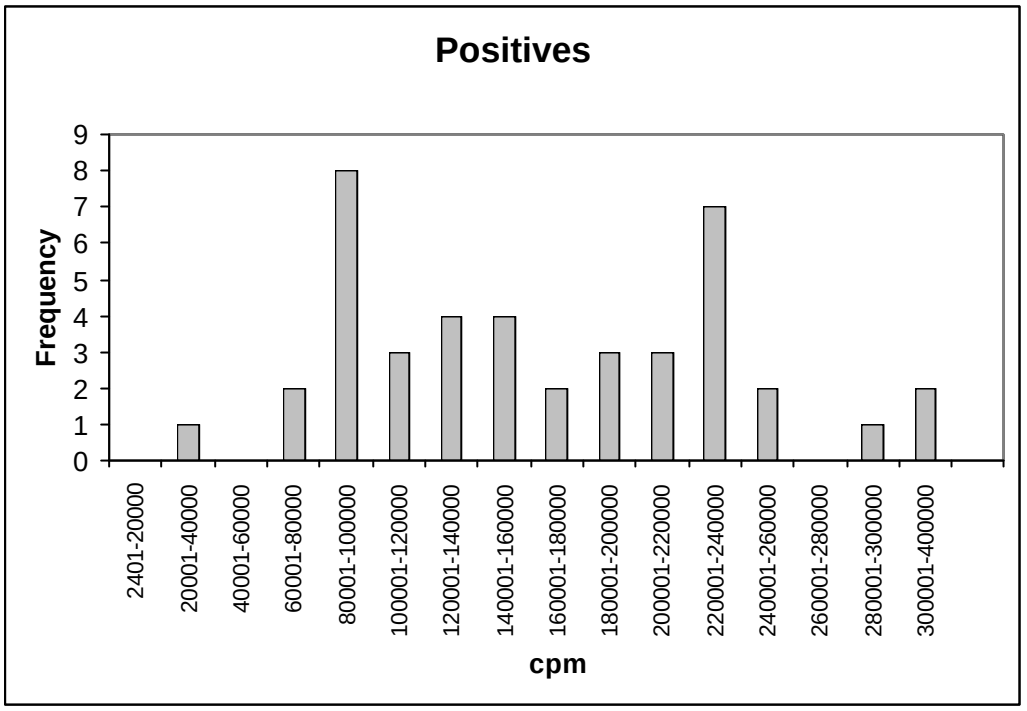


Fig 15: UCSF CDI test – positive samples (mean of two replicates)

Average of negative and positive data populations (mean of two replicates)

	Average cpm	SD
Positive	163868	87942
Negative	-499	1422

Dilution Series

Only dilutions of homogenates in the range of 10^{-1} to 10^{-3} prepared at IRMM in 2001 (series (a)) were used to assess the detection limit of the test.
The possibility to dilute the same mouse titrated positive brain homogenate used to produce the series (a) (i.e. another sub-batch not containing any buffer) with a pool of 5 negative brain stem material was offered. The test developer did not perform this experiment because they felt that possible effects of repeated freezing thawing cycles etc., required for making the dilution series, were not fully understood.
During the processing of the series (a) samples a pipetting error concerning four samples occurred (wrong tips leading to dosage of too large reagent volumes). The measurements were repeated with the remainders of the samples homogenised in UCSF buffer. Since the amount of sample per unit of series (a) only allowed one additional measurement per sample one result instead of two as foreseen in the testing protocol is mentioned in the table below.

Although measurements of other samples should have been repeated according to the decision criteria given above no further repetitions were carried out for the following reasons in addition to the limited amount of sample available:

- Samples close to the detection limit are expected to give signals around the cut off and repetitions should only confirm these readings (however outliers can only be identified and eliminated by performing repetitions).
- The protocol envisages repetition of measurements using tissue samples rather than using remainders of homogenised samples in homogenisation buffer.
- Because the influence of freezing and thawing on such homogenates was not known and believed to have a detrimental effect on the signal.

Consequently only the repetitions of four samples were taken into consideration where severe pipetting errors due to human error occurred (one 10^{-1} and 10^{-3} dilution respectively and two $10^{-2.5}$ dilutions).

Individual results obtained for series (a)

Dilution	Replicate	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Undiluted	1	130269				
	2	139704				
10-1	1	13508	12784	12998	8807	
	2	13030	13338	11762		
10-1.5	1	5287	4455	4503	7407	
	2	5079	5059	5633	7307	
10-2	1	1343	1587	2407	7257	
	2	2372	2818	3973	5910	
10-2.5	1	763	2302	120	184	
	2	862	3335			
10-3	1	464	2358	2081	49	
	2	266	2996	2198		
Negative	1	-197	-123	2298	1629	3813
	2	203	632	2148	2316	4308

Series (a) prepared by IRMM (first measurement series):

Dilution	Mean relative to cut off	Standard deviation
Undiluted	134987	6672
10-1	12318	1647
10-1.5	5591	1156
10-2	3458	2117
10-2.5	1261	1285
10-3	1487	1190
Negative controls	1703	1582

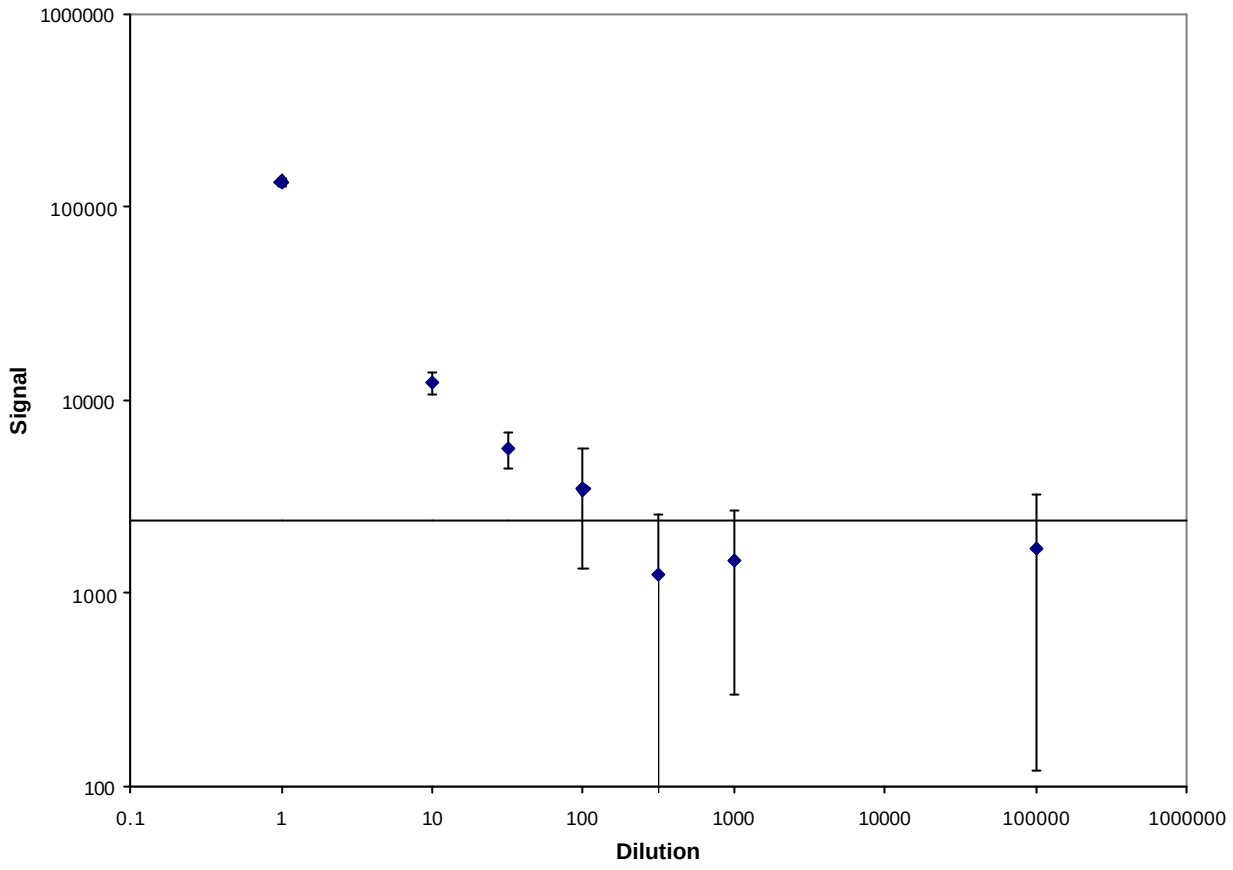


Fig 16: Dilution series (a) prepared by IRMM (negative controls are shown as a 1:100000 dilution)

Summarising the above the detection limit of the test appears to correspond to a 10^{-2} dilution (even if two high results would be considered as outliers) or equal to $10^{-2.25}$ based on interpolation of the means of all data.

7.5 Test E MRC Prion Unit, Imperial College, UK

The test is an immunoassay using two different monoclonal antibodies directed against two epitopes presented in a highly unfolded state of bovine PrP^{Sc}. The first biotinylated antibody is used for the extraction of PrP^{Sc} out of an enriched fraction, the second mAB is ruthenium tagged thus allowing the detection of bound PrP^{Sc} in a electrochemiluminescence reaction. This reaction is recorded with the ORIGEN system. Homogenates are prepared in a ribolyser, followed by solubilisation in detergents, Proteinase K treatment, centrifugation steps, delipidation and a stringent unfolding step of PK resistant PrP. The actual format (single test vials) does not yet allow a high throughput screening but the method could easily be upscaled. Results are achieved in 6 hours.

Decision criteria (cut off definition)

The cut-off value is defined as the mean of three negative controls (induplicate) per run plus three times the standard deviation. The definition is based on experimental data achieved on a series of measurements on negative samples. No zone identifying borderline cases was defined, both negative and positive populations are separated strictly by the single cut-off value. This cut-off is calculated for each experimental series.
Only one homogenate per sample is measured in duplicate and the mean value thereof is considered for classification. All positive samples are subjected to a second measurement to exclude any possible mistake in labelling and identification of the samples.

Sensitivity and Specificity

	True positive	True negative	Totals
Test Positive	48	0	48
Test negative	0	152	152
Totals	48	152	200

	%	95% Confidence Interval
Sensitivity	100	93.8%
Specificity	100	98.0%

Repeated samples

The first two series of measurements comprising 90 samples were not analysed since incomplete mixing of a detergent in the digestion buffer by an inexperienced technician led to incomplete digestion of the negative controls. These assays were deemed to have failed as the assay cut-off's were >100,000 and the samples were

retested. Nevertheless on the first failed run, all 32 samples that were processed by coworkers using a different aliquot of the digestion buffer could have been judged correctly as positive or negative simply on basis of the distribution of the raw data (highest 'negative' value 7072, lowest 'positive' value 674354, theoretical cut off would have been 209162). Likewise all samples were correctly identified in the second run, despite the assay failure. When these series were repeated all samples were correctly identified.

Two samples (IRMM 1934 and IRMM 1944) were swapped and therefore retested. This happened by wrong labelling of a test tube during the transfer of an aliquot of homogenate for testing from the mastertube. These samples were identified by staff of the test developer after comparing their results to those of the first run that were not considered. They were correctly assigned positive (IRMM 1934) and negative (IRMM 1944) upon retesting starting from the homogenate in the mastertube.
As foreseen in the testing scheme all positive samples from the first round were retested and correctly confirmed. The mean of the values relative to cut-off was 79.7 (SD 26.9).

Samples with false classification

All samples were correctly classified.

Data distribution

The test allows a clear separation of positive and negative data populations. The six highest negative values have a reading relative to cut off between 0.764 and 0.905, whereas the six lowest positives have a reading relative to cut off of 8.77 to 20.06. The maximum reading relative to cut off has a value of 212.59.

Highest Negative Mean rel. to cut off	**IRMM number**		**Lowest Positive** Mean rel. to cut off	**IRMM number**
0.90	664		8.77	124
0.89	264		11.35	714*
0.85	1634		17.0	1704*
0.83	774		17.21	1674
0.82	214		19.13	94
0.76	944		20.06	1604

* weak positive test samples

The samples giving readings relative to cut off of 11.35 (IRMM 714) and 17 (IRMM 1704), respectively, belong to the three samples that resulted in consistently low results in the evaluation of 1999. The third weak positive sample had a factor relative to cut-off of 38 (IRMM 64).

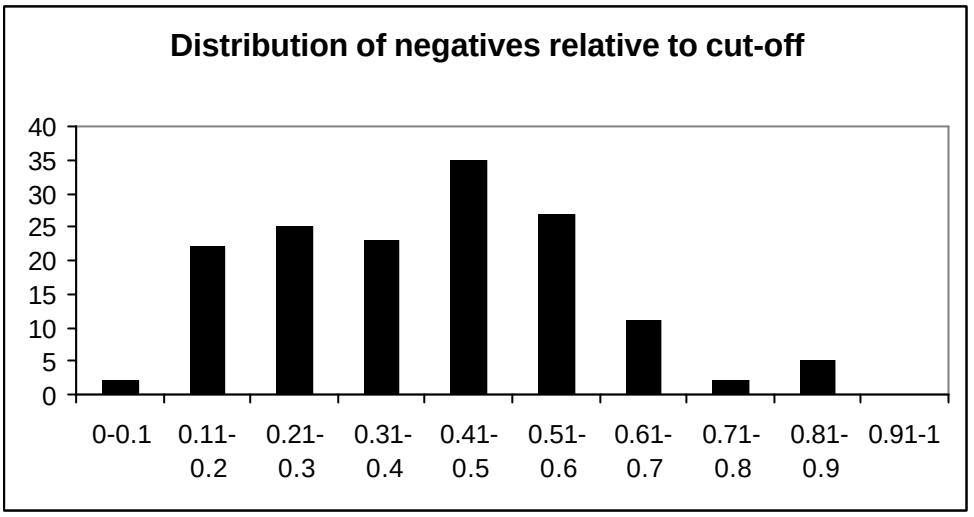


Fig 17: Test results for negative samples, relative to the cut off fixed at 1.

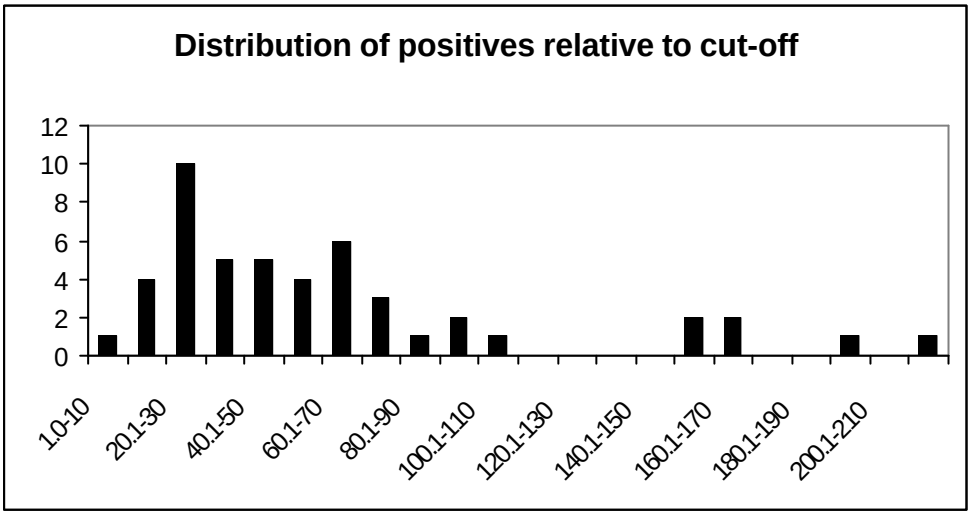


Fig 18. Test results for 48 positive samples relative to the cut off, fixed at 1.

Average of negative and positive data populations

	Average relative to cut off	SD
Positive	62.95	49.8
Negative	0.41	0.18

Dilution Series

Two different dilution series were tested to determine the detection limit of the test.

(a) Dilutions of homogenates in the range of 10^{-1} to 10^{-3} prepared at IRMM in 2001.

To measure the series **(a)** the homogenates were resuspended in 5% sucrose to achieve a 20% w/v solution as was required for further processing. Since the standard protocol includes the possibility of sample freezing after homogenisation no commutability problems were expected. All series (a) homogenates were processed according to the standard protocol. Two aliquots of each dilution step were tested in duplicate in two independent measurements resulting in mean 1 and mean 2 values.

Dilution	**Positive score**	**Mean 1 of readings relative to cut off**	**SD**	**Mean 2 of readings relative to cut off**	**Standard deviation**	**Factor**
undiluted	2/2	73.41		32.41		2.26
10 -1	8/8	22.53	4.2	9.77	0.79	2.30
10 -1.5	8/8	4.97	1.24	2.53	0.16	1.96
10 -2	7/8	2.01	1.13	1.49++	0.71	1.35
10 –2.5	2/8	0.97+	0.29	1.17+	0.57	1.21
10 -3	0/8	0.77	0.22	0.84	0.11	1.09
Negative*	*2/10	0.92/0.72*	0.47/0.18*	0.797/0.72*	0.188/0.09	1.00

* In both measurements one negative homogenate sample (IRMM 21626) was found to be positive with values just above cut off (14609 with cut off 8503 and 26445 with cut off 24021, respectively).

\+ One of the four replicates was above cut off (1.38 and 2.02, respectively).

++ The mean of one sample had a reading relative to cut off of 0.98, with one replicate just above and a second below cut off (24315 and 22935 with cut off 24021).

The negative samples have a higher ratio relative to the cut off than it is the case with the slices (0.72 compared to 0.41).

(b) Dilutions of homogenates prepared freshly on site with the same mouse titrated positive brain homogenate used to produce series (a) (another sub-batch not containing any buffer) and a pool of 5 negative brain stem material originating from areas anterior to the obex. Equal amounts of the five brain stems (about 5 g) were homogenised, collected and mixed with the positive material in serial dilution steps. The test developer opted not to test lower dilutions and started the serial 1:3 dilutions from an initial 1:10 dilution. Each dilution was aliquoted into 4 subsamples and further divided into two or three aliquots (1:270, 1:810, 1:2430), respectively. Each aliquot was measured in duplicate. The samples were coded and analysed according to the standard protocol.

Series (b) prepared according to test developer's procedure:

Dilution	positive score	Mean of readings relative to cut off	Standard deviation
Undiluted		Not measured	-
1:10	8/8	16.23	6.07
1:30	8/8	6.43	1.88
1:90	8/8	2.06	0.293
1:270	10/12	1.134	0.22
1:810	4/12	0.956	0.154
1:2430	4/12	0.95	0.223
Negative controls	0/8	0.68	0.166

The average of the negative homogenates is clearly below cut off and corresponds to the reading obtained for the neagtive controls used to determine the cut off. However, compared to the sliced samples this mean is about 1.84 times higher.
The test has with a high probability the potential to detect at least a 1:270 dilution. 10 out of 12 samples of the 1:270 dilution had values above cut off. The two samples below cut off had both one replicate with values close to cut off (0.97 and 0.89 realtive to cut off, respectively).
5 readings of the aliquots in the 1:810 and 1:2430 dilutions were ranging from 1.003 to 1.08 relative to cut off, whereas 3 reached levels of 1.3 to 1.34 relative to cut off.

7.6 Summary Tables

Tables 2 and 3 below present a summary of the results obtained for the parameters of Sensitivity, specificity and detection limits. However, in selecting a test for a specific use, other factors such as ease of use and speed of sample throughput, should also be taken into account.

Table 1: Sensitivity and specificity of each test using the predetermined cut-off points.

	Test A	Test B	Test C	Test D	Test E
Sensitivity	97.9 %	100 %	97.9	100 %	100 %
Specificity	100 %	99.3 %	100	100 %	100 %

Table 2: Number of homogenate samples prepared by IRMM scoring positive (above cut-off) at each dilution level.

	Test A	Test B	Test C	Test D	Test E
Titre $10^{3.1}$	1/1	1/1	1/1	1/1	2/2
10^{-1}	4/4	1/4	4/4	4/4	8/8
$10^{-1.5}$			$1^{1}(4^{2})/4$	$2(2^{3})/4$	8/8
$10^{-2.0}$			$0^{1}(4^{2})/4$	$1(2^{4})/4$	7/8
$10^{-2.5}$				$0(1^{4})/4$	2/8
$10^{-3.0}$				$0(1^{4})/4$	0/8
Negative				$0(1^{4})/5$	2/10

1 Test with 5 % tissue homogenate and 1 hour digestion
2 Test with 10 % tissue homogenate and 2 hours digestion
3 Borderline positive cases normally to be repeated but clearly above cut off
4 Borderline positive cases normally to be repeated

Table 3: Number of homogenate samples prepared according to test developer's protocol scoring positive (above cut-off) at each dilution level.

	Test A	Test B	Test C	Test D[4]	Test E
Titre $10^{3.1}$	-	1/1	-	-	-
1:3	4/4	4/4	4/4	-	-
1:9	4/4	4/4	4/4	-	8/8 1:10
1:27	4/4		4/4	-	8/8 1:30
1:81	4/4 visually by trained persons (3)/4 CCD[1]		4/4	-	8/8 1:90
1:243			$4^{2}(2^{3})/4$	-	10/12 1:270
1:729			$4^{2}(1^{3})/4$		4/12 1:810
1:2181			$3^{2}(0^{3})/4$		4/12 1:2430
Negative			$1^{2}(0^{3})/4$	-	0/8

1 Read 3 times 'dubious' with CCD camera
2 On basis of cut off defined as 5 times negative control, one negative sample suspect
3 on basis of the "elevated cut off" (8 times negative control)
4 measurements not performed

7.7 Comparison of dilution series with 1999 evaluation [1]

In order to allow comparison between the evaluation carried out in 1999 with the evaluation summarised in this report diluted homogenates used in 1999 have been provided to various test developers. They were analysed in parallel to the series (a) produced by IRMM in 2001 with the remainders of the positive pool already used in 1999 and following the same production method. Because of shortage of storage space the diluted homogenates used in 1999 were stored at –20 °C whereas the raw materials for the production of the dilution series produced in 2001 have been stored all the time at –70 °C.
All test developers of the current evaluation were asked to perform analysis of samples produced in 1999 except ID Lelystad because it is a qualitative test and Prionics because of the commutability problems with homogenates produced by IRMM. Also the test developers of the 1999 evaluation [1] (except Prionics because the Western blot gives only qualitative results) were asked to analyse samples of the dilution series produced by IRMM in 1999 and 2001. This work is going on. So far results from BioRad could be obtained.

PerkinElmer Life Sciences

	Result in %	Standard deviation
Undiluted positive produced in 2001	95.4	0.1
Undiluted positive produced in 1999	91.7	1.3
10^{-1} dilution produced in 2001	7.6	0.5
10^{-1} dilution produced in 1999	5.9	0.6

UCSF

	Result in cpm	Standard deviation
Undiluted positive produced in 2001	124987	6672
Undiluted positive produced in 1999	163789	18196
10^{-1} dilution produced in 2001	12318	1647
10^{-1} dilution produced in 1999	14828	2016

Within the measurement uncertainty the dilution series produced in 1999 and 2001 agree well for the PerkinElmer Life Sciences test, the UCSF test and the BioRad Platelia test (see below). With the exception of the PerkinElmer Life Sciences test there seems to be a trend of the homogenates produced in 1999 giving somewhat higher readings although they have undergone some changes in matrix properties due to drying and possibly oxidation effects (material was more difficult to dissolve and showed mainly at the surface changes in colour). From this data it can be concluded that the homogenates have been stable with regard to the detectable amount PrP^{sc}.

BioRad Platelia

	Result expressed relative to cut off R3+0.21	Standard deviation
Undiluted positive produced in 2001	14.45	0.06
Undiluted positive produced in 1999	17.18	0.29
10^{-1} dilution produced in 2001	14.24	2.26
10^{-1} dilution produced in 1999	16.27	1.45
$10^{-1.5}$ dilution produced in 2001	8.20	3.08
$10^{-1.5}$ dilution produced in 1999	12.08	0.82
10^{-2} dilution produced in 2001	3.18	0.85
10^{-2} dilution produced in 1999	6.26	2.04
$10^{-2.5}$ dilution produced in 2001	0.85	0.18
$10^{-2.5}$ dilution produced in 1999	1.53	0.58

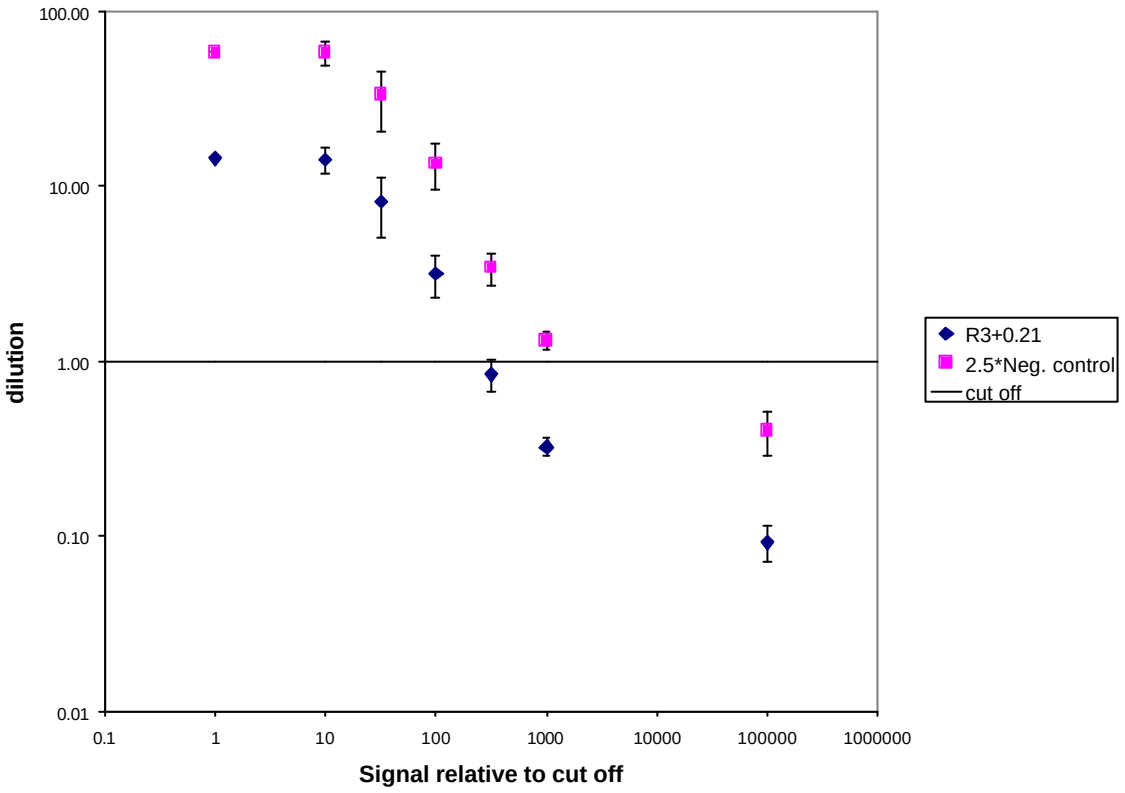


Fig 19: Dilution series (a) prepared by IRMM (negative controls are shown as a 1:100000 dilution)

The curve obtained with dilution series (a) is equivalent to what was obtained in the 1999 evaluation [1] indicating also good agreement between the series produced in 1999 and 2001 and therefore comparability of the two evaluations.

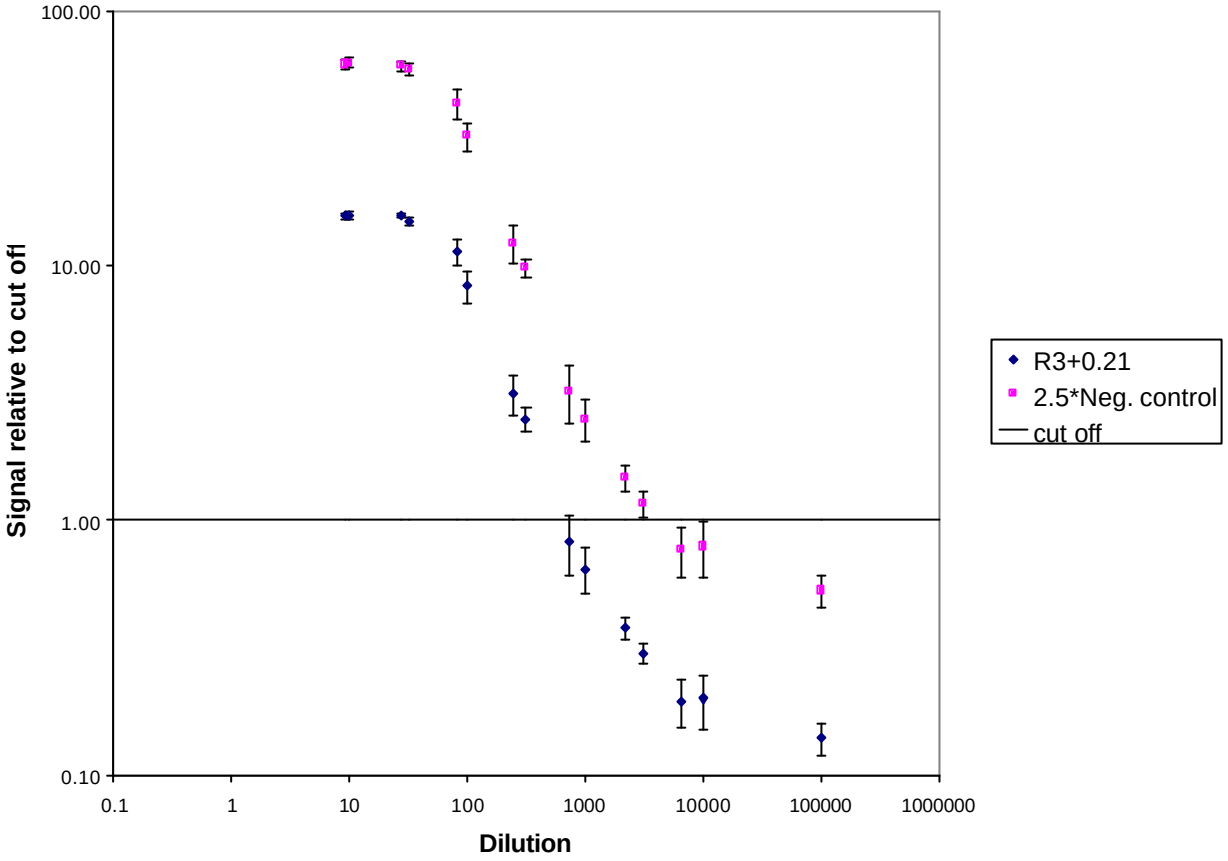


Fig 20: Dilution series (b) prepared according to test developer's procedure (negative controls are shown as a 1:100000 dilution)

Whereas the dilution series (a) gives a detection limit around a $10^{-2.5}$ dilution is the detection limit as determined with series (b) at around 10^{-3} dilution. The difference of a factor of three has been already observed with Test A, but has not been confirmed by test B. With test C the situation was not clear because of the elevated baseline signal for the negative homogenates.

7.8 Impact of homogenisation on test signal

The production of the samples used for this investigation is summarised in chapter 6. The cutting scheme is described in Annex 3. The impact of homogenisation on three tests giving quantitative readings and being based on different sample pretreatment and detection principles has so far been evaluated. The results are summarised below. S in the tables stands for tissue slices whereas H stands for homogenised sample. The corresponding slice/homogenate pairs are on the same lines. The average of the measurements has been always put in the middle columns to allow comparison between the two results whereas the corresponding individual results have been put into columns to the right and left respectively of the columns in the middle. The cutting scheme is shown in Annex 3.

Prionics Check LIA test

Two different digestions (dig 1and dig.2) were carried out. From digestion one four replicates per sample were performed whereas from digestion one only two replicates were performed. The data are raw data and are not corrected for cut off because of the similarity of the signals of the positive control samples on the three microtitre plates.

	Dig. 1	Dig. 1	Dig. 2			Dig. 1	Dig. 1	Dig. 2	
	Controls + 102058 - 162	Controls + 110325 - 184	Controls + 104313 - 312	Mean	Mean	Controls + 102058 - 162	Controls + 110325 - 184	Controls + 104313 - 312	
Brain 1									
S13	633162	657773	805221	798702	649436	668468	614467	611698	H14
	907867	826808	961382			692856	659270	649857	
H16	387461	388119	240153	378307	260169	268882	245568	201470	S15
	470263	409737	374110			284273	281897	278926	
Brain 2									
S1	880552	822311	915575	982583	1427382	1383895	1328450	1260655	H2
	1141274	977145	1158638			1455412	1520547	1615333	
H4	1799916	1720348	1708144	1859564	1488782	1443614	1263025	1366019	S3
	2079745	1996464	1852764			1542065	1596634	1721337	
S5	1557723	1453119	1527063	1623361	2123930	2128071	1947674	2104980	H6
	1740019	1580396	1881843			2290983	2208066	2063806	
Brain 3									
S1	373261	387418	341028	394656	364340	351808	360851	232015	H2
	403521	398880	463829			458821	440665	341881	
H4	675492	633610	467804	648820	699870	625104	619153	996790	S3
	759176	663892	692948			697443	683161	577569	
S5	675702	660968	794747	787848	874879	919021	740279	815851	H6
	851876	770913	972879			969120	865145	939860	
Brain 4									
S13	252452	224458	159784	235878	343811	321489	293144	298706	H14
	275380	238407	264786			423896	381439	344191	
H16	193297	158590	125610	177351	124649	123717	119401	79110	S15
	195172	199780	191654			162112	143610	138906	

UCSF

The homogenisation of the tissue slices (about 1 g) was not possible in one homogenisation step. However the test developer was asked to merge the different homogenised sub-samples of each slice in order to make sure that the investigation is not biased by further sampling effects. Samples with marked with an R (e.g. H8R) had to be repeated because of discordant reading of the first measurement.

	Controls	Ratio D/N	Norm. Difference D-N	Mean	Mean	Norm. Difference D-N	Ratio D/N	Controls	
Brain 1									
H12	+ 160105 - 947	31.6 42.8	92042 111189	101616	134829	145175 124483	61.4 60.9	+ 160105 - 947	S11
Brain 2									
H8	+ 160105 - 947	41.7 47.7	136039 108986	122513	160301	163952 156649	67.7 73.6	+ 160105 - 947	S7
S9	+ 160105 - 947	81.3 72.8	175039 124279	149659	160394	177641 143146	57.6 45.5	+ 160105 - 947	H10
H12	+ 160105 - 947	37 59.4	84980 124283	104632	71349	131601 11096	65.8 8.7	+ 160105 - 947	S11
					103248	145724 60771	147.8 63.6	+ 163246 - 344	S11R
Brain 3									
H8	+ 244443 - 536	30 47.8	88682 147620	118151	67994	99717 36271	60.4 30.4	+ 160105 - 947	S7
H8R	+ 163246 - 344	70.6 52.3	92965 74006	83486	81744	92994 70493	90.2 70.4	+ 163246 - 344	S7R
S9	+ 244443 - 536	150.7 50.6	194709 71750	133230	132290	116986 147593	39.4 51.9	+ 244443 - 536	H10
S9R	+ 163246 - 344	83.6 93.5	93568 166362	129965					
H12	+ 244443 - 536	47.9 56.2	138327 153250	145789	121395	133799 108990	82.2 58.3	+ 244443 - 536	S11
Brain 4									
H8	+ 244443 - 536	66.4 67.3	139351 123735	131543	88438	87179 89696	91.2 77.4	+ 244443 - 536	S7
S9	+ 244443 - 536	147.8 94	169610 112741	141176	132764	115293 150234	39.6 59.7	+ 244443 - 536	H10
H12	+ 244443 - 536	55.9 71.4	152033 143876	147955	155461	176308 134614	136.1 116.7	+ 244443 - 536	S11

BioRad Platelia

In order to be able to quantify and compare the signals for the slices and corresponding homogenates the samples had to be diluted. In the table below the corresponding dilutions are next to each other. The results are expressed as signals relative to the cut off (R3+0.21 AU).

	Dil.			Mean	Ratio R3+0.21	Ratio R3+0.21	Mean			Dil.	
Brain 1											
H8	1/32	3.448	3.493	3.471	15.72	14.55	3.211	3.177	3.245	1/32	S7
	1/100	2.880	2.809	2.845	12.89	6.69	1.477	1.588	1.365	1/100	
	1/316	1.846	1.708	1.777	8.05	2.11	0.465	0.477	0.453	1/316	
S9	1/32	3.242	3.241	3.242	14.68	15.27	3.371	3.393	3.348	1/32	H10
	1/100	2.372	2.037	2.205	9.99	10.96	2.419	2.498	2.339	1/100	
	1/316	0.741	0.726	0.734	3.32	3.89	0.859	0.873	0.844	1/316	
Brain 2											
S1	1/10	3.699	3.806	3.753	17.00	12.83	2.833	2.703	2.962	1/10	H2
	1/32	2.747	2.637	2.692	12.19	9.79	2.161	2.087	2.234	1/32	
	1/100	0.701	0.776	0.739	3.35	2.50	0.553	0.499	0.606	1/100	
	1/316	0.181	0.170	0.176	0.80	1.35	0.299	0.302	0.295	1/316	
H4	1/1	3.270	3.638	3.454	15.27	15.13	3.423	3.424	3.421	1/1	S3
	1/10	2.605	2.690	2.648	11.90	7.41	1.649	1.721	1.576	1/10	
	1/32	1.053	0.930	0.992	4.46	3.25	0.724	0.735	0.713	1/32	
	1/100	0.254	0.244	0.249	1.12	0.73	0.162	0.161	0.163	1/100	
	1/316	0.108	0.090	0.099	0.44	0.26	0.057	0.057	0.057	1/316	
S5	1/10	3.706	3.801	3.754	16.87	16.06	3.573	3.646	3.500	1/10	H6
	1/32	2.338	2.252	2.295	10.31	11.69	2.601	2.524	2.678	1/32	
	1/100	0.807	0.718	0.763	3.43	3.31	0.737	0.714	0.759	1/100	
	1/316	0.198	0.187	0.193	0.87	0.78	0.173	0.164	0.181	1/316	
Brain 3											
S13	1/1	3.370	3.624	3.497	15.46	16.58	3.751	3.443	4.058	1/1	H14
	1/10	0.954	1.017	0.986	4.43	4.79	1.067	1.163	0.970	1/10	
	1/32	0.340	0.315	0.328	1.47	1.07	0.238	0.250	0.226	1/32	
	1/100	0.089	0.091	0.090	0.40	0.41	0.091	0.097	0.085	1/100	
	1/316	0.049	0.047	0.048	0.22	0.21	0.047	0.046	0.047	1/316	
H16	1/1	3.382	3.479	3.431	15.16	8.82	1.995	2.105	1.884	1/1	S15
	1/10	1.292	1.417	1.355	6.09	5.42	1.207	1.161	1.253	1/10	
	1/32	0.388	0.341	0.365	1.64	1.58	0.351	0.359	0.343	1/32	
	1/100	0.097	0.104	0.101	0.45	0.50	0.112	0.107	0.116	1/100	
	1/316	0.053	0.040	0.047	0.21	0.15	0.033	0.034	0.032	1/316	
Brain 4											
H8	1/100	3.256	3.325	3.291	14.84	14.87	3.298	3.334	3.262	1/100	S7
	1/316	1.452	1.539	1.496	6.74	7.33	1.625	1.503	1.747	1/316	
S9	1/32	3.210	3.242	3.226	14.55	15.44	3.424	3.492	3.356	1/32	H10
	1/100	2.162	2.049	2.106	9.49	15.68	3.476	3.455	3.497	1/100	
	1/316	0.421	0.416	0.419	1.89	6.19	1.373	1.277	1.468	1/316	
H12	1/10	3.436	3.271	3.354	15.12	17.21	3.816	3.706	3.925	1/10	S11
	1/32	2.343	2.143	2.243	10.11	12.82	2.842	2.726	2.958	1/32	
	1/100	0.886	0.809	0.848	3.82	5.02	1.113	1.057	1.169	1/100	
	1/316	0.274	0.242	0.258	1.16	1.29	0.286	0.270	0.302	1/316	

The slices could also in this case not be homogenized in one piece and therefore the test developer was asked to merge the homogenates of the sub-samples belonging to a slice in order to make sure that sub-sampling effects can be excluded.

Enfer Ltd.

Slices and homogenates from three different brains were tested in this series. The tests were performed with reagents from two different production batches in parallel by two independent technicians. The mean of the results obtained for each slice and homogenate by both technicians is presented.
SD= standard deviation. S= slice. H= homogenate.

Brain 1	**Batch1**	**SD**	**Batch 2**	**SD**	**Batch 1**	**SD**	**Batch 2**	**SD**	
S 1	**1596.17**	2.828	**2329.16**	502.75	**273.79**	13.64	**53.68**	58.25	**H 2**
	3278.31	1449.24	**3353.41**	610.00	**88.62**	120.14	**63.38**	65.66	
S 3	**240.22**	95.95	**962.25**	40.23	**7.61**	2.34	**55.7**	20.30	**H 4**
	231.59	24.62	**252.75**	4.62	**12.85**	3.69	**16.67**	0.11	
S 5	**306.92**	24.81	**848.56**	226.41	**103.29**	10.15	**492.96**	61.09	**H 6**
	352.82	125.85	**364.81**	96.72	**90.63**	4.53	**176**	47.38	
Brain 2									
S 1	**1211.67**	12.02	**3605.66**	90.51	**196.17**	65.62	**805.11**	80.11	**H 2**
	1446.6	711.92	**3322.22**	225.91	**236.85**	81.39	**252.85**	19.13	
S 2	**1257.17**	135.76	**4167.66**	237.58	**165.97**	22.48	**1466.66**	159.80	**H 4**
	994.3	358.84	**2410.89**	1365.60	**471.35**	111.23	**907.2**	244.83	
Brain 3									
S 1	**1165.62**	376.95	**2718.16**	1103.79	**828.07**	564.41	**2032.16**	449.01	**H 2**
	1957.95	22.40	**1286.19**	78.39	**1330.09**	95.51	**1023.75**	123.27	
S 2	**1387.67**	81.31	**2675.16**	200.11	**186.77**	34.50	**879.76**	59.25	**H 4**
	1843.93	124.09	**2232.48**	762.37	**254.24**	28.51	**514.45**	370.20	
S 5	**1845.67**	103.94	**5172.16**	408.00	**436.87**	54.58	**737.71**	230.16	**H 6**
	2738.85	325.29	**3580.11**	1104.96	**407.98**	14.22	**647.17**	169.13	
S 13	**303.77**	81.03	**1918.16**	84.14	**161.97**	27.86	**104.56**	6.22	**H 14**
	489.27	49.47	**917.55**	92.04	**101.42**	22.21	**138.14**	46.48	
S 15	**124.12**	18.59	**411.46**	155.42	**135.34**	33.30	**163.36**	18.52	**H 16**
	213.83	93.40	**398.19**	132.24	**102.39**	21.59	**159.65**	50.58	

Summary

There are strong indications that homogenisation if done under harsh conditions (higher mechanical forces, higher temperature etc.) can have a severe impact on the integrity of the samples and consequently on the test signals. One test developer obtained for instance a dilution series containing 10 % tissue in PBS buffer. The raw materials were the same as for the production of series (a) used for the assessment of the detection limits. The homogenates in PBS buffer where produced by turraxing 3 minutes each dilution level with the consequence that the signal dropped to approximately 30 % of the value obtained with the dilution series (a).
Preparation of 1 in 10 diluted homogenates applying 3 minutes turraxing has led to negative results with two approved BSE post mortem tests. The fact that homogenates prepared with the same starting materials but containing 80 % of a buffer provided with one of the tests and having been turraxed 45 seconds gave clearly positive readings with the corresponding test is making the impact of different homogenisation procedures on the performance of the tests obvious.
Therefore parameters like temperature, type of equipment used, speed and dimensions, volume of the sample and its temperature seem to play an important role.
The experiment shows for the tests applied that losses in test signal observed at other occasions when producing homogenates are not due the homogenisation as such possibly causing a biochemical reaction because of the distortion of the original cell structure. The reason for losses of signal during homogenisation is most probably of mechanical nature (essentially shear forces occurring during homogenisation disrupting PrP^{sc} aggregates).
Comparison of the readings for the homogenates and the corresponding slices shows clearly that the homogenisation procedure applied (45 seconds turraxing with 20 % of a 5% sucrose buffer) did not influence the UCSF/CDI, the BioRad Platelia and the Prionics Check LIA tests evaluated but had a very significant impact on the Enfer test. In case of the latter test the readings for homogenates were decreased by a factor of up to 40 (see brain 1 slice 1 compared to homogenate 2). The influence of the homogenisation is very variable and is e.g. minor for brain 3 if slice 1 and homogenate 2 are compared but can be, even within the same brain, up to a factor of 20.
Consequently the influence of the same homogenisation procedure can be very different on BSE post mortem tests. The assessment of analytical sensitivity on basis of diluted homogenates would lead to a discrimination of e.g. the Enfer test in cases as described above.
The results also show that there can be steep gradients of detectable PrPsc in the brain stems. Neighbouring slices of brain 1 and 4 in the set of data from Prionics show a difference in signal of a factor of 2 although these slices are only 3 mm away from each other. In brain 3 a factor of three over 3 slices (distance 6 mm) could be detected. Similar observations were made with the other tests (brain 3 and 4 with the UCSF test, brain 2 in the BioRad Platelia test). The data also clearly show large differences in the signals from one brain to the other.
These results also serve to underline the importance of correct sampling.

7.9 Influence of storage temperature on sample quality

Some information on the stability of homogenates and tissue slices can be derived form the data obtained during the evaluation.

Homogenates

For the production of dilution series (a) the same homogenate was used as for the preparation of the equivalent series in 1999. Whereas the homogenate used to produce series (a) in 2001 was stored at –70 °C since 1999 after addition of 20 % of a 5 % sucrose solution (series (a) homogenate 1 of 2001) some samples of the dilution series produced and used in 1999 were kept at –20 °C since 1999 (homogenate 1 of 1999). Both homogenates 1 therefore contain 20 % of a 5 % sucrose buffer.
In spring 2001 another sub-batch of the same pool used in 1999 was delivered by VLA. The material has been always kept at –70 °C (series (b) homogenate 2 (100 %)). This material has been used by test developer's to produce dilution series (b) according to their own procedure. Part of the material has been treated in the same way as in 1999 i.e. 20 % of a 5 % sucrose buffer were added and the mixture was turraxed for 45 seconds as it has been done in 1999 (homogenate 2 (80 %)).
Below the results with the PerkinElmer Life Science, the UCSF and the BioRad Platelia tests are summarised.

PerkinElmer Life Sciences

	Series (a) homogenate 1 of 2001		Homogenate 1 of 1999		Homogenate 2 (80 %)		Series (b) homogenate 2 (100%)	
Units/ml	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Soluble PrP	1.1	0.1	1.7	0.4	188.4	14.5	285.3	5.7
Insoluble PrP	22.9	1.2	18.6	1.0	75.7	7.0	117.8	5.9
Insoluble PrP / total PrP	95.4	0.1	91.7	1.3	28.6	1.7	29.2	0.8

UCSF

	Series (a) homogenate 1 of 2001		Homogenate 1 of 1999		Homogenate 2 (80 %)		Series (b) homogenate 2 (100%)	
Cpm	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Undiluted	134987	6672	163788	18196	131810	14762	151867	5398

BioRad Platelia

	Series (a) homogenate 1 of 2001		Homogenate 1 of 1999		Homogenate 2 (80 %)		Series (b) homogenate 2 (100%)	
Dilution	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Undiluted	14.45	0.06	17.18	0.29	-	-	-	-
1/10	15.34	0.00	15.28	0.36	16.97	0.68	15.81	0.61
1/32	14.00	0.05	13.86	0.17	14.47	0.13	14.91	0.60
1/100	6.23	0.37	6.41	0.04	7.21	0.23	8.31	1.22
1/316	1.98	0.05	2.25	0.08	2.25	0.07	2.50	0.25

With the BioRad and the UCSF test all four different sub-batches of the homogenate look the same. The storage of the samples of “homogenate 1 of 1999” at –20 °C since 1999 did not have any effect on the test results. Also the turraxing and addition of 20 % sucrose buffer to the “series (b) homogenate 2 (100 %)” did not have an adverse effect. The two previous observations hold also true for the PerkinElmer Life Sciences test. The two series (a) and series (b) homogenates respectively look very similar. However there are differences between these two series which cannot be explained by storage temperature or addition of buffer. The question is whether these results are not caused by an analytical artefact. Otherwise the differences between the homogenates 1 and 2 can only be explained by already existing differences upon receipt of the different sub-batches (e.g. transport conditions but all samples arrived in good state) or by the probably longer processing time of homogenates 1. The homogenate 1 was used to produce the dilution series in 1999. It has been processed during one day only and was put back in the freezer after some hours.

Tissue samples

During the evaluation process PerkinElmer Life Sciences, Prionics and BioRad received sets of sample slices which have been stored since 1999 at –20 °C. The data are summarised below.

PerkinElmer Life Sciences

	Samples stored at –70 °C			Samples stored at –20 °C		
	Mean	SD	No of samples	Mean	SD	No of samples
Positives			48			48
% insoluble PrP / total PrP	36.8	19.4		36.6	20.4	
Soluble PrP	96.5	51.3		79.9	22.9	
Insoluble PrP	79.3	84.2		66.3	64.4	
Negatives			152			151
% insoluble PrP / total PrP	4.86	1.26		3.8	1.3	
Soluble PrP	86.8	28.1		92.9	27.9	
Insoluble PrP	4.49	1.89		3.6	1.6	

Within the variability of the individual samples there is no statistically significant difference between the different fractions. A trend for samples stored at higher temperature to give lower results for the insoluble PrP fraction in negative samples and lower results for the soluble and insoluble PrP fraction in positive samples can be observed. These lower results lead to somewhat lower results for the % insoluble PrP versus total PrP fraction in negative samples and compensate themselves in positive samples leading in average to equivalent % insoluble PrP versus total PrP values. Although the distribution of the negative and positive samples is much alike in samples stored at –70 °C compared to –20 °C an impact on low positive samples cannot be excluded because an overlap of the positive and negative populations was observed in samples stored at –20 °C. Three samples out of 48 fell below cut off and were lower than the two highest negatives (7.4 and 7.8 %).

Prionics Check LIA test

Results relative to cut off		No of sampl.	-70 °C		-20 °C	
			Mean	SD	Mean	SD
Standard protocol, 10 % homogenates, 1 hour digestion	Pos	47	609	649	1645	1228
	Neg	153	0.20	0.12	2.01	1.48
5 % homogenates, 1 hour digestion	Pos	47	-	-	1276	1212
	Neg	153	-	-	0.18	0.03
10 % homogenates, 2 hours digestion	Pos	47	-	-	1993	1685
	Neg	153	-	-	0.44	1.15

The samples stored at the different temperatures were measured at different occasions and therefore the previously mentioned problems with the substrate do not apply for the samples stored at –20 °C.

With the standard testing protocol the negative samples stored for two years at –20 °C gave a high background signal. Only 43 out of 153 negative samples were below cut off. The highest negative sample gave a signal relative to cut off of 8.3 whereas the lowest was at 0.42. No obvious explanation could be found for this effect but it is clearly related to sample quality and not to the test as such. Fresh samples from Switzerland tested the day before as negative and kept in a fridge over night gave the

expected results around 0.2 units relative to the cut off. Also when these samples were frozen they gave the same results. Therefore the effect is not caused by freezing only but probably by drying effects during prolonged storage. As it has been the case with the homogenates (see chapter 7.3) part of the normal PrP^c may not be digested. The fraction is estimated to be lower than 0.1 % of the PrP^c. The positive samples were however clearly above cut off the lowest being 90 times above cut off (i.e. 450 times the mean of the negative control). By shifting the cut off to e.g. 75 times instead of 5 times the mean of the negative control a 100 % correct identification would therefore be possible.
The test developer proposed to homogenise less tissue in the initially used buffer i.e. to put instead of 10 % only 5 % tissue into the homogenisation buffer. This measure lead to much lower signals for the negative samples. The distribution of the negative samples was however remarkably narrow (lowest negative 0.09, highest 0.30 units relative to cut off) which indicated according to the test developer overdigestion of the samples posing a risk that low contaminated samples would become false negative. The lowest positive was however 54 times above cut off and under the conditions applied a very clear separation between the negative and positive sample population and consequently 100 % correct sample classification could be achieved.
In order to minimise the risk of overlooking weak positive samples the tissue concentration was increased again to 10 % but the incubation time for the protease K digestion was prolonged to 2 hours. Under these conditions the distribution of the negative samples looked more normal but some negatives were again above cut off (highest negative 13.8, lowest 0.11 units relative to cut off). Again a clear separation of the negative and positive populations could be achieved since the lowest positive sample gave 122 units relative to cut off. By shifting the cut off upward 100 % correct identification could be achieved.
The data mentioned above clearly demonstrate that the test is susceptible to storage conditions of samples. On the other side they also support the statement made in paragraph 7.3 that the sample having given a false negative result must have been a sample which contained exceptionally low amounts of PrPsc. Otherwise similar observations should have been made with the set stored at –20 °C especially under the less optimal conditions described above giving nevertheless a very clear separation between positive and negative data populations.
The data reiterate the need for proper quality control of test kit batches because even the third run with the samples stored at –70 °C was probably done under sub-otimal conditions which can be derived from the significantly lower mean for the positive samples which was quite constant under the conditions applied to the samples stored at –20 °C.

BioRad Platelia

Signal relative to R3+0.21	Mean	SD
Samples stored at –20 °C	0.12	0.02
Samples stored at –70 °C	0.14	0.06

20 negative samples each stored over two years at –20 °C and –70 °C respectively were provided to BioRad. No difference was found between those samples.

8. References

[1] J. Moynagh, H. Schimmel, The evaluation of tests for the diagnosis of transmissible spongiform encephalopathy in bovines, http://europa.eu.int/comm/food/fs/bse/bse12_en.html
[2] J. Moynagh, H. Schimmel, Tests for BSE evaluated, Nature 100, 105 (1999)

9. Acknowledgements

This evaluation would not have been possible without the assistance of the following;

The expert group who advised on the selection of participants and on the operation of the evaluation; Prof. J. Almond, Dr. D. Heim, Dr. N. Hunter, Dr E. Sprengers.

Those who supplied samples; the Central Veterinary Laboratory, Weybridge, UK and Ministry of Agriculture and Forestry, New Zealand.

The participants who contributed a considerable amount of time and effort.

The Veterinary Services of Switzerland, the United Kingdom, the Netherlands and the United States of America.

Annex 1. Protocol

PROTOCOL

for the Evaluation of Tests for BSE

Objective

The evaluation process is designed to evaluate three parameters;

1. Diagnostic Specificity
2. Diagnostic Sensitivity
3. Test detection limits

General

The tests will be evaluated using the tissue type for which they were developed. However, it will also be extremely useful to obtain information on the behaviour of each test on standardised tissues. This would remove variability due to any possible tissue sampling bias where there may be differences in PrP distribution within tissues and would indicate whether there are inherently different detection levels in different tissues.

Consequently the diagnostic specificity and sensitivity tests will be evaluated on the tissue types that would be used whereas the test detection limits will be evaluated on specially prepared homogenised samples.

Any differences in performance between raw and homogenised samples will give valuable information, particularly concerning the inherent repeatability of the test and problems arising from uneven distribution of PrP in the tissues.

It is likely that homogenised samples will be finely minced. The addition of buffer may be incompatible with some test methods and prompts a decline in PrP levels, so will be avoided.

Parameters

1. Diagnostic Specificity

Diagnostic Specificity may be defined as the proportion of uninfected reference animals that test negative in the assay.
The diagnostic specificity will be evaluated by testing approx. 150 samples of known negative tissue. The tissue, as required for the test, will be supplied in frozen form.

2. Diagnostic Sensitivity

Diagnostic Sensitivity may be defined as the proportion of known infected reference animals that test positive in the assay;

It is estimated that diagnostic sensitivity and the distance between the data populations between negative and positive samples can be evaluated by testing approx 50 samples of tissue from confirmed positive animals. The tissue, as required for the test, will be supplied in frozen form.

3. Test Detection Limits

Test detection limits will be a major determinant in the assessment of the capability of the selected tests to detect the presence of preclinical BSE.

The test detection limit may be defined as the smallest detectable amount of the analyte. Because of the nature of the assay, this determination may be relative.

This parameter will be assessed by supplying each candidate with specially prepared samples made up from minced central nervous tissue, containing positive tissue diluted in negative tissue. The samples will be presented for testing in a number of dilutions ranging down to 10^{-3}.

To achieve standardisation and external comparability, several samples which has been titrated by mouse assay will be included in the batch sent to each candidate

Origin of Samples

Positive samples will come from bovines naturally infected with BSE which showed clinical signs of the disease. In each case the disease is confirmed by histology or other appropriate method.

Negative samples will come from healthy animals of similar age to the above, originating in an area generally accepted as presenting minimal risk of BSE being present.

Sample Material

The sample materials will consist of brainstem. The majority of samples analysed will consist of pieces of whole tissue derived from a single animal. A substantial number of samples for testing will be presented in minced form. The exact determination of the tissues to be used will, of necessity, correspond to the tissues for which each test has been developed. The provision of suitable tissue will depend on its availability.

Of necessity, samples used to assess test detection limits will be minced, but in this case may contain material from several animals. If diluted, the dilution will be with negative brain tissue.

Because of the valuable nature of the sample material, participants will be presented with the smallest size of tissue material compatible with the test under evaluation.

Preparation of samples

Sample materials will be prepared by the Joint Research Centre, Geel, Belgium. The samples for testing will be blinded and identified by code. The code shall be kept confidential.

Sample Delivery

Batches of samples will be despatched to the participants by the Joint Research Centre, Geel, Belgium who will co-ordinate the evaluation exercise. On receipt of the samples, participants must test the samples and report the results within a set time period. Depending on the logistics samples may be delivered in a single batch or over a period of time.

Retention of sample

A portion of each sample supplied for testing will be retained in Geel. Fixed portions of all negative samples will also be available.

Sample Storage

Samples will be stored at –70°C, following collection and in Geel. Where such facilities are unavailable, participants will be asked to arrange to obtain access to such facilities. If this proves to be impossible, the material may be stored at –20°C and in such cases must be tested it as soon as possible.

Contract

All participants will be required to agree a contract covering the practical aspects of the evaluation exercise. This will cover areas such as confidentiality, reporting arrangements, test timing, security etc.

Security

The following measures will be taken to minimise the risk of scientific fraud.

1. On the spot visits will take place to the participating laboratories when testing is underway.
2. Participants will be required to notify the opening/defrosting of each package and to report results each evening.

3. All original laboratory results and raw data must be submitted; in the case of western blot photographs must be supplied. Laboratory workbooks will be open for inspection.

Safety

Participants must have authorisation from the Competent Authority to work with BSE positive material and to dispose of same.

Reporting of results

Results must be reported on a standard form, initially by fax on the evening of the day the testing was carried out,

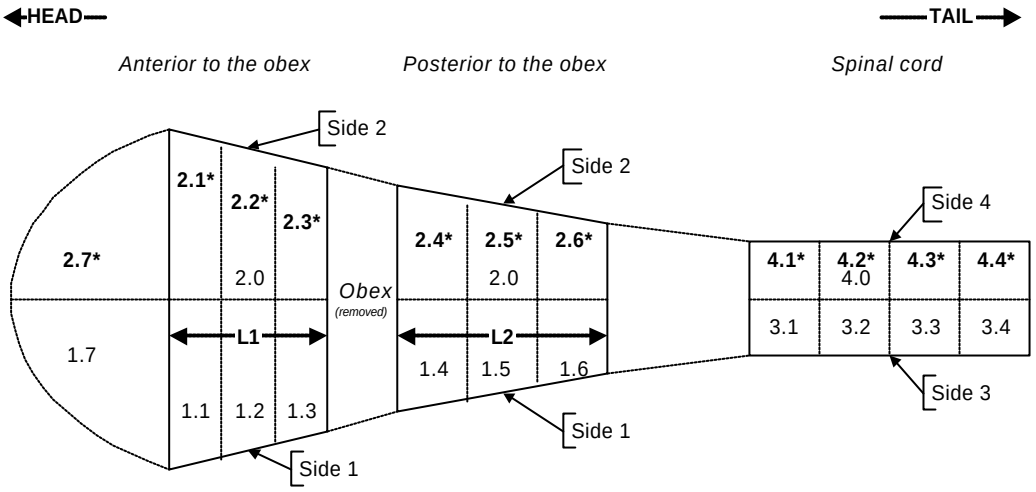
In addition, all original laboratory results must be submitted; in the case of western blots or other blot types photographs must be supplied. Laboratory workbooks will be open for inspection.

Location of samples on microtitre plates and blots must also be noted.

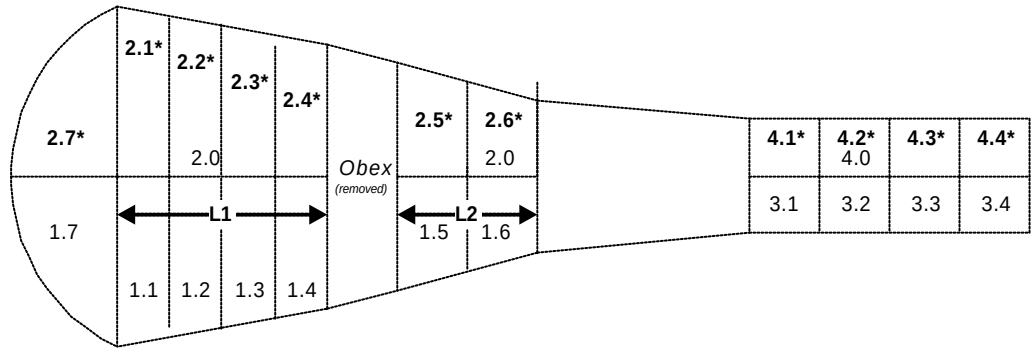
False Positives

Following testing, participants will be required to retain all unused tissue materials until authorised to destroy them. False positive samples will be subjected to follow up examination on retained samples on a case by case basis.

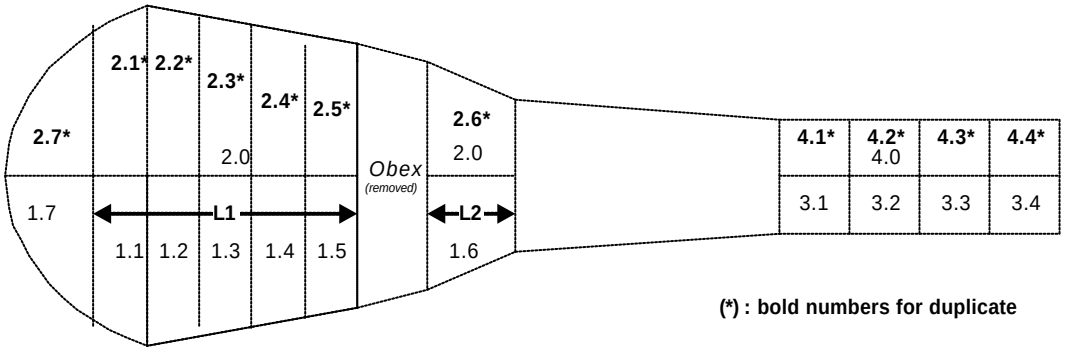
Cutting scheme no. 1 :



Cutting scheme no. 2 :



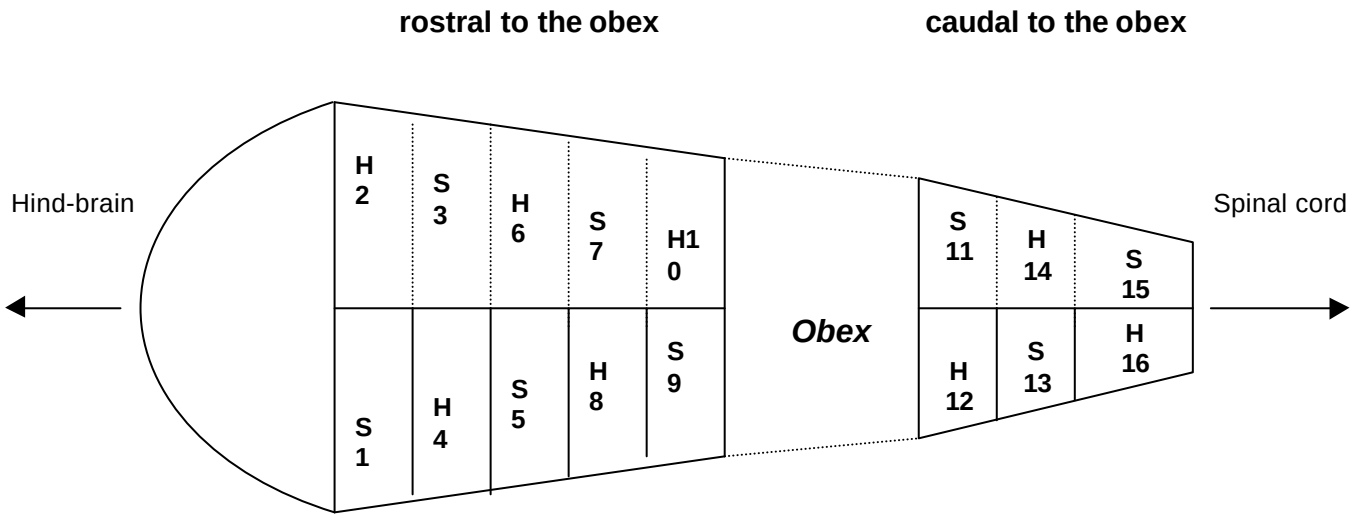
Cutting scheme no. 3 :



(*) : bold numbers for duplicate

Annex 2. Cutting scheme for samples

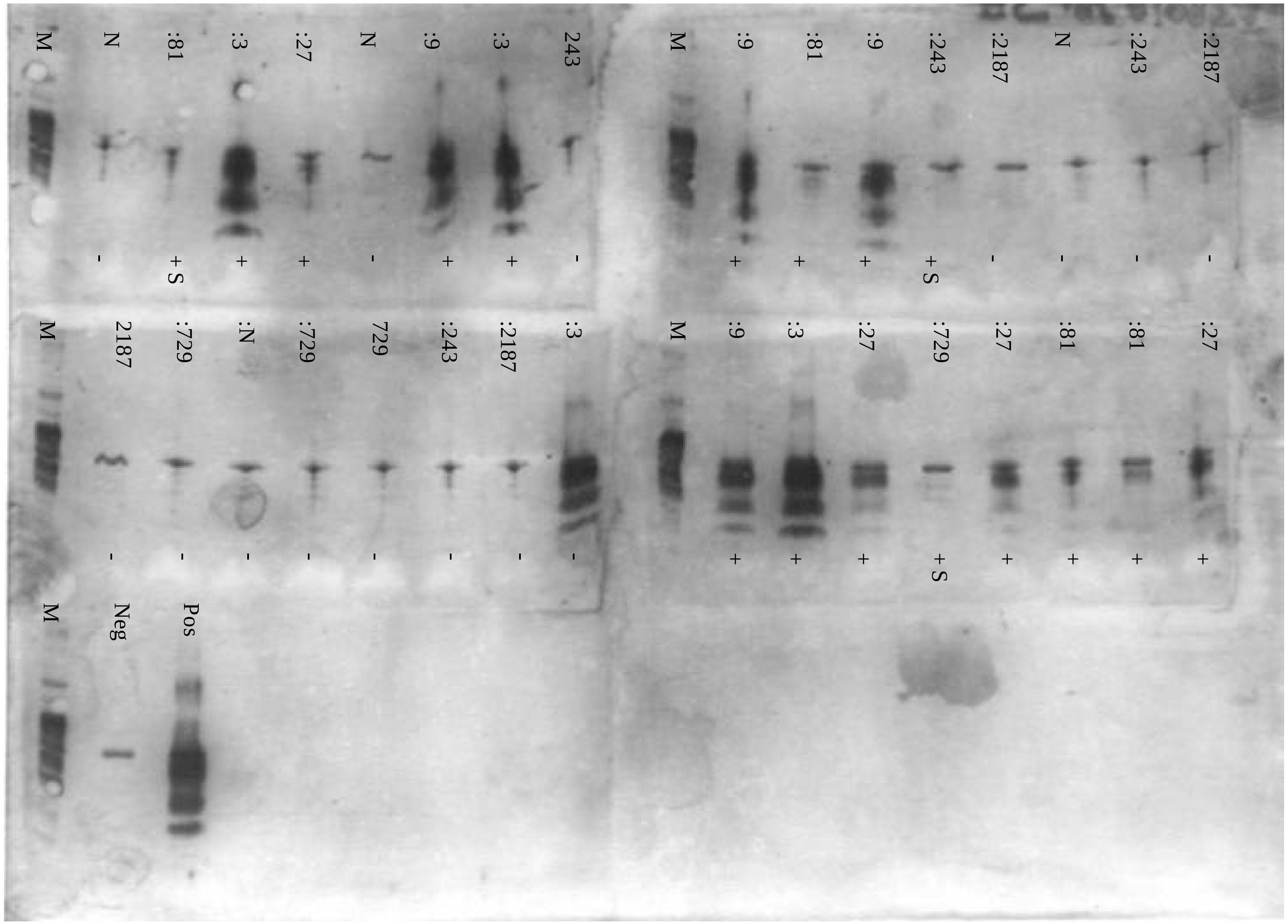
Annex 3. Cutting scheme for samples for investigation of the impact of homogenisation



H = Homogenised with Omni-turrax

S = Kept as slice

Annex 4: Prionics Check Western Blot with homogenate series (b)



Annex 5. Comments by the participants

All Participants were given the opportunity to comment on any aspect of the evaluation exercise and in particular on the results of the evaluation on their own test. These comments are set out in this Annex

A. Comments by ID Lelystad

a) The ID-Lelystad TSE-test is designed to be a simple, fast and reliable assay. BSE as well as scrapie can be detected. For the performance of this test no specialized laboratory equipment is required. Visual comparison of pairs of dots on photographs yields qualitative test results. The evaluation with a CCD-camera produces more quantitative results. The test procedure containing a single antibody incubation step, can be completed within 6 hours. No centrifugation is needed. The microtiter format of the assay lends itself to automation. In the present evaluation no false positives were detected and samples were determined with a high chemical sensitivity.

b) The method shows a unique property towards other non-Western blot TSE-tests sofar in use: false positives by incomplete digestion are not possible since even *non-digested* confirmed BSE-negative control samples remain negative by visual inspection and below the cut-off level of 0.015 in the CCD-camera measurement. This is visualized with the CCD-camera results in Figure 1 below. These undigested control samples are located on all 7 blots in position H12. In addition, all *digested* negative control samples scored negative (blot-positions G10-G12, H10, H11).

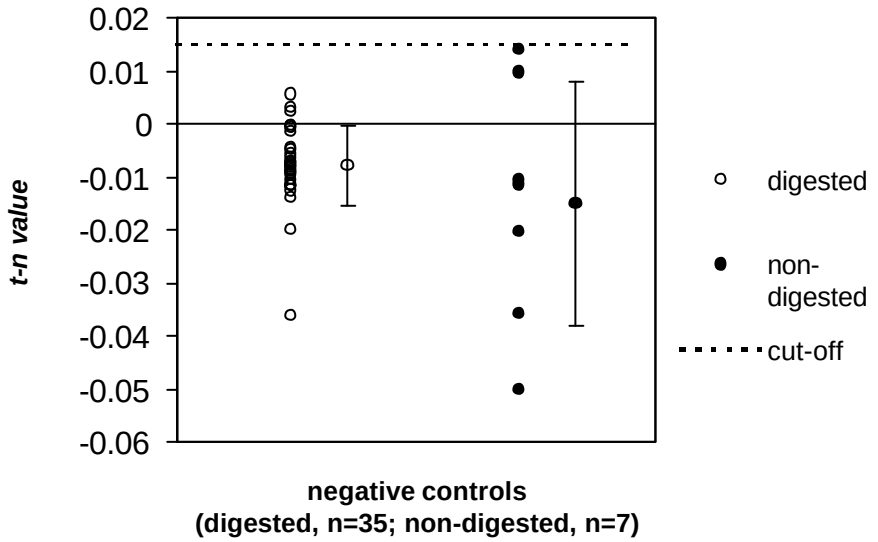


Figure 1: *t-n* values of all negative control samples used in the 7 blots by CCD-camera measurement. Values for each individual sample are represented by the symbols. For digested and non-digested samples also the means ± SD are shown.

c) A cut-off level for the CCD-camera system was empirically obtained from preceding series of confirmed BSE-negative samples. The cut-off position at 0.015 on the (*t*-*n*)-axis was chosen which is around the (*t*-*n*)-means plus 2 standard deviations of these series. The use of CCD-camera readings allows a reproducible standard way of measuring large numbers of samples; 'dubious' samples should be further looked at visually. In Figure 2 below, the frequency distribution of CCD-camera measurements of 175 out of 200 samples that could be processed by the accompanying software, is presented. Sample 2874 (arrow) is the one which at repetition scored positive (*see also paragraph of report: 'Samples with false classification').* Dashed line indicates the position of the cut-offpoint of 0.015. Apart from sample 2874 (*t*-*n*=0.0064; average of duplicate), the highest reading of other negative-scoring samples was 0.0031; the lowest positive scoring sample was 0.0525.

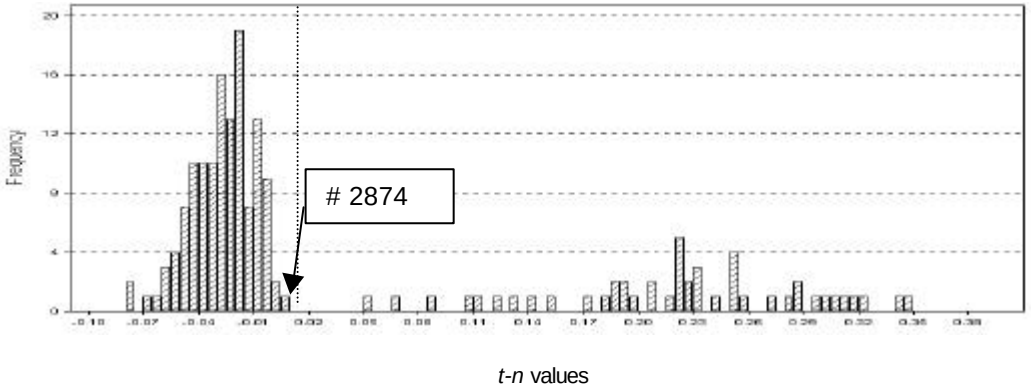


Figure 2: Frequency distribution of CCD-camera measurements of 175 out of 200 samples which could be processed by the accompanying software. Sample 2874 (arrow) is the one which at repetition scored positive (*see report- paragraph "Samples with false classification").* The vertical dashed line indicates the position of the cut-offpoint of 0.015. Apart from sample 2874 (*t*-*n*=0.0064; average of duplicate), the highest reading of other negative-scoring samples was 0.0031; the lowest positive scoring sample was 0.0525.

With respect to the CCD-camera measurements, the data further show that the choice of the cut-off is close to being optimum. Nevertheless, the fact that the only false negative sample (2874) has also the highest reading of all negative-scoring samples suggests that the cut-off point for the CCD-camera evaluation could be slightly reduced. This is further substantiated by the fact that in dilution series (b) the 1/81 dilution could not be scored positive by the CCD-camera evaluation (3 times 'dubious' and 1 time negative) while being scored positive by visual inspection (4 out of 4). A slightly reduced cut-off point increases the sensitivity of the CCD-camera measurement. If a slightly reduced cut-off point is decided on, then it should be with a provision for retesting samples giving results between the reduced cut-off point (*e.g.* 0.005) and the now used cut-off point (0.015). In this way sample 2874 would not have scored false negative. However, such a decision will be influenced by other factors such as the number of samples in practice giving results between the reduced and the now used cut-off point that

could be retested, and the consequences on how to treat two results on the same brainstem ('dubious'-'dubious', 'dubious'-negative, 'dubious'-positive).
The method for elaboration the CCD-camera recording requires some additional solutions in software use, as is illustrated for the cases of 25 negative samples surrounded by an unusual number of strongly positive samples, which the software was not yet able to interpretate. However, such cases are in practice very unusual. The present CCD-camera system reads intensities of sample spots from photographic X-ray films exposed after chemoluminescent immunostaining of the N- and T-membranes. In addition, more sophisticated CCD-camera systems are available to directly measure chemoluminescence intensities arising from the enzymatic reaction on the membrane (no photographic film needed). In any case, visual inspection of photographic films is also quite convenient for large scale practice.

d) Only sample 2874 is described as false negative and should be considered a borderline case. This sample was read both visually and with the camera negative. However, sample 2874 was the only one out of the 200 samples-set interpreted differently between the two persons: person 1 judged both duplicates positive and person 2 both duplicates negative. For scientific interest it was asked by IRMM to repeat the measurement of sample 2874 having been tested false negative.The repetition was made blindly in a series of 9 samples surrounding the false negative sample starting from the same left-over tissue. This set is represented in the Table to illustrate that the differences within duplicates are small.

Table: CCD-camera evaluation of repeated samples.

sample	initial test				repeat test			
	replicate 1	replicate 2	mean	result	replicate 1	replicate 2	mean	result
2784	0.3627	0.2764	0.3196	BSE	0.2060	0.1757	0.1909	BSE
3334	no result[1]	no result[1]	no result[1]	no result[1]	-0.0281	-0.0109	-0.0195	NEG
2394	0.1754	0.1867	0.1810	BSE	-0.0241	-0.0246	-0.0244	NEG[2]
3194	no result[1]	no result[1]	no result[1]	no result[1]	-0.0039	0.0008	-0.0016	NEG
2874	0.0066	0.0062	0.0064	NEG	0.0843	0.0429	0.0636	BSE[3]
3044	no result[1]	no result[1]	no result[1]	no result[1]	-0.0130	-0.0113	-0.0122	NEG
3304	-0.0557	-0.0392	-0.0475	NEG	-0.0111	-0.0102	-0.0106	NEG
2634	-0.0312	-0.0267	-0.0289	NEG	-0.0152	-0.0107	-0.0129	NEG
3864	-0.0462	-0.0294	-0.0378	NEG	-0.0020	-0.0043	-0.0031	NEG

[1] No result because light spots could not be transformed by the software. [2] Wrong due to lack of remaining sample and probably only white matter. [3] Repeat test weakly positive

An interesting phenomenon appeared during the repetition when a clearly positive sample in the first run became negative (sample number 2394). In this case by shortage of remaining tissue only 0.16 g of brainstem (possibly only white matter) was subjected to homogenisation. It demonstrates the importance of taking sufficiently large samples reducing the risk of missing areas with accumulated PrPsc (grey matter) and of being very critical in taking sub-samples.

e) To further get insight in the reproducibility of the test, the measurements were carried out both on basis of remaining spotting dilution, remaining digest, remaining homogenate and remaining tissue. The remaining digest, remaining homogenate and remaining tissue turned then out positive, but the signal was still

weak. The remaining spotting dilution was again classified by person 1 as positive (both duplicates positive) and by person 2 as negative (both duplicates negative). This shows the reproducibility of the test even with weak samples. The fact that sample 2874 has been tested first negative and when repeated - on basis of a different aliquot of the same homogenate - positive is reflecting the weak positivity of this sample. Additional confirmation of sample 2874 was received with the standard Prionics-Check Western blot system where the sample exhibited a single weak band and was judged dubious; according to the criteria of that test such a result would not be considered positive (characteristic banding pattern absent i.e. three bands with lowest band less intense). It should be noted that the sample slice originated from one of the three positive samples tending to give low readings in the 1999 evaluation.

f) There appears to be a difference in detection limits as determined by the dilution series (**a**) of homogenates prepared by IRMM, and (**b**) of homogenates prepared freshly on-site with the same mouse titrated positive brain homogenate used to produce series (**a**) (another sub-batch without addition of any buffer). This is probably due to a minor compatibility problem with diluted samples of series (**a**) or a variation in stability and homogeneity. The undiluted homogenates of series (**a**) and (**b**) have the same *t-n* value by CCD-camera measurement of around 0.3.

B. Comments by PE Biosystems

The avoidance of PK as the discriminating reagent in the assay allows for the differential extraction method to be potentially much more than a 'yes-no' test. While one can understand the desire to have a 'simple' result, the disease is not a 'yes-no' disease. TSEs are characterised by prolonged incubation periods, during which time the animals are 'infected' but do not manifest the characteristic clinical symptoms. This pre-clinical stage of the disease is likely to be characterised by an increasing accumulation and deposition of aggregated PrP but at low concentrations. It is here that the 'sensitivity' of the diagnostic becomes of paramount importance. Any immunodiagnostic, certainly one that aspires to be quantitative or semi-quantitative, must inevitably be more sensitive than a 'yes-no' test. Furthermore, clinical diagnostic tests are most often characterised by reference ranges and probabilities of diagnosis.

The second point that is worth noting from the results obtained in various blind trials is that as soon as any arbitrary cut-off is established in a method that is not a 'yes-no' test, the potential for false positives and false negative results become inevitable. For example, in October 1999, the VLA set up a blind trial of the method. The data is shown in the following Table.

Results from the VLA blind trial

TYPE OF SAMPLE	Number of tissues	Mean %	SD %	Range of values
NZ NEGATIVES	100	3.96	1.63	1.1 to 9.8
CVL BSE HISTO NEGATIVE	18	5.78	2.58	3.2 to 12.2
CVL BSE HISTO POSITIVE	47	51.84	21.71	17.1 to 89.5
OTMS ROSTRAL HISTO NEGATIVE	95	6.79	3.63	2.9 to 19.1
OTMS CAUDAL HISTO NEGATIVE	95	6.37	2.13	2.5 to 14.3
OTMS ROSTRAL HISTO POSITIVE	23	41.58	22.83	14.4 to 81.4
OTMS CAUDAL HISTO POSITIVE	23	42.53	17.51	15.5 to 82.8

Although, an initial arbitrary cut-off of 8% was established, 31 samples in the blind trial gave values between 8% and 10%. In addition, 21 samples gave results that were greater than 10%. The overall mean of these 52 samples was 10.2% with a SD of 2.25%. None of these 52 samples could be confirmed as positive by histology. The possibility that these samples, collected from the UK national herd as part of the

OTMS scheme, were taken from animals in the pre-clinical stage of BSE is worthy of consideration. The average age of cattle displaying the clinical signs of BSE is between 4 and 4.5 years. Animals over 30 months but younger than 4 years may well be in the pre-clinical stage of the disease *if* they are infected.

In the latest EU trial conducted at the end of August 2001, the same arbitrary threshold of 8% was set. Accordingly, one sample (8.1%) was above this threshold and was classified as a false positive. A more statistically relevant threshold, however, was 8.7% (i.e. three times the standard deviation above the negative mean of 4.8%). At this threshold, 100 % classification was obtained. It is interesting to note that this is almost exactly the same statistical threshold that was determined for the 100 NZ negative tissues in the VLA blind trial that was conducted in October 1999 (Mean 3.96%, SD 1.63, Threshold 8.9%).

Nevertheless, if there has to be a threshold, there may be some epidemiological rationale in setting a lower threshold in order to obtain at least small number of false positives. The reason for this is that any positive result will be repeated whereas a false negative remains totally hidden and ignored.

Geoff Barnard
December 2001

C. Comments by Prionics A.G.

Prionics-Check-LIA Test-Design
The Prionics-Check LIA is intended for use in the detection of the disease-specific prion protein in bovine brain tissues to assess BSE-disease in cattle independent of origin and breed. The LIA-procedure uses an integrated process of homogenization, proteinase K digestion and detection using microtiter plates without any change of format. Detection follows a solid-phase sandwich-ELISA-format with two monoclonal antibodies and horse radish peroxidase combined with a luminescence producing substrate as signal system. Cut-off calculation is based on test inherent controls. The kit contains any reagent used to performing the test and in the pack insert all accessories are defined.

Sample material is defined as brain stem, medulla oblongata (Obex-region) from cattle of different origin/breed and age from over 12 months and the storage conditions of sample material are clearly described in the pack insert as keeping the samples up to 5 days at 2-8°C or two freeze/thawing cycles or storage at –70°C and avoiding freeze-drying.

Influence of sampling / reference material criteria
Sensitivity and specificity were evaluated using tissue slices caudal and rostral to the obex region of the medulla. Samples from the obex were not available for this study as this region had been formalin-fixed for initial diagnosis. Earlier studies (e.g. Acta Neuropathologica, 98:437-443, and studies conducted at the U.K. Veterinary Laboratory Agency, Weyebridge) indicated inhomogeneous distribution of PrP^{Sc} outside the obex leading to non-perfect correlation of diagnosis based on obex or non-obex tissue respectively. Nevertheless, the EU-evaluation of BSE tests carried out in 1999 as well as the present study demonstrate that most (but not all) samples from BSE-positive animals are grouped within a certain range of PrP^{Sc} levels and therefore correlate to the obex samples used for the original diagnosis. To ensure an optimal correlation between screening- and confirmatory tests for BSE surveillance, many EU-countries have in the meantime adopted a sampling scheme in which one side of the obex is used for a first rapid test leaving tissue from the opposite side of the obex for confirmatory analysis.

In the present evaluation using tissue slices caudal and rostral to the obex region, Prionics-Check LIA produced a large separation of the signals for positive and negative samples indicating excellent discrimination. Two sets of samples stored at –20°C (partially freeze dried) or –70°C were analysed in this evaluation. The set of samples stored for a prolonged time at –20°C gave clearly separate readings for all negative and positive samples, despite the fact that the partial freeze drying of the samples had negative effects. In the second set of samples, stored at –70°C, a sample with exceptionally low PrP^{Sc} concentrations was observed. This sample (#5634) was found positive after retesting of additional tissue. The results illustrate the intrinsic difficulty of determining the performance of a diagnostic test with non-standardized samples. With sampling outside the obex, it will be inevitable that rare substandard samples have to be excluded from the calculation of the true

diagnostic sensitivity. In light of the separation of all positives and negatives in the first set of samples, and in light of the excellent discrimination factor of more than 1:100 on an average between positive and negative samples, as well as considering the high analytical sensitivity of the assay (1:240 above cutoff, 1:2000 above the negative) sample #5634 of the second sample has to be considered such a sample to be excluded from the calculation of representative diagnostic sensitivity.

Currently, no optimal reference material has been defined to evaluate prion tests. The current method using brain slices has the intrinsic problem that the content of PrP^{Sc} in a given slice can not be predetermined. An alternative to brain slices is being developed by Prof. Dr. Habermehl, Gesellschaft f. Biotechnologische Diagnostik mbH, Berlin for ring trials and has been used successfully with the LIA (see below Field performance).

Influence of sample quality

Storage conditions

The specification of the test allows only the use of freshly prepared tissue or tissue stored under conditions avoiding freeze drying. During development of the test it could be shown that especially digestion is influenced by freeze drying often leading to higher signals. This could be also confirmed during the evaluation and by adapting the digestion conditions also this type of material could be determined with a 100% of correct identification. The test therefore is robust and flexible enough to cope with partially freeze-dried specimens.

Pre-processed samples (homogenates)

In order to assess the analytical sensitivity of the test expressed as a factor of dilution prepared homogenised samples of IRMM were used (series a). As already mentioned the Prionics-Check LIA is designed for freshly prepared tissue which has to be homogenized in the test-included homogenization-buffer. The homogenization conditions initial to testing are very sensitive to changes, e.i. prehomogenized material in matrices other then the Prionics-buffer alters the conditions of the whole test. Therefore it is not astonishing producing incorrect results at first time. But also in this case only a slight adaption of digestion conditions (prolonging time) was leading to an detection limit of 1/200 with the IRMM-Material showing the flexibility of the test to be adapted to material other then routine testing. Using the freshly prepared samples from frozen tissue-slices (series b) no problems arose and an excellent detection limit of 1/240 above cutoff (1:2000 above background) was found.

Influence of procedure

Correct homogenization is essential for carrying out the test as already mentioned in the chapter above "pre-processed samples". If using fresh samples, the buffer included and keeping to the instructions the procedure works reliable and reproducible (chapter 7.7).

The same is valid for the digestion procedure which was shown very reproducible during evaluation where two series of digestion confirmed the robustness of this process-step (see chapter 7.7).

Analytical sensitivity
Analytical sensitivity is defined as a factor of dilution referred to homogenate of one brain sample diluted in negative homogenate of five brains of healthy cattle. An excellent detection limit of at least 1/240 above cutoff (and 1:2000 above negative sample value) was found (series b) but one has to take into consideration that also this material was stored for two years and that comparing these results to that in the study of 1999 is only allowed under reservations.

Diagnostic sensitivity / specificity
Excluding the undefined sample 5634 (see above "Influence of sampling"), a sensitivity and specificity of 100% was found. The present study demonstrates that most samples from BSE-positive animals are grouped within a certain range of PrP^{Sc} levels, with only a few samples lying above or below. 80% of the negative sample group at a cut-off of 0.2 showing a discrimination-factor of at least 1/100 on an average (see fig. 7). With Prionics-Check LIA, a large separation of the signals for positive and negative samples was observed indicating excellent discrimination.

Other validation sudies
Another validation study of the LIA has been performed under the supervision of the **Swiss Veterinary Authorities** using 1400 positive and negative samples and yielding 100% sensitivity and specificity. These results as well as the comparison of Prionics Check Western and LIA on sheep samples are being published (J. Virol. Methods, in press).

Field performance
Despite all validation studies, we consider the **field performance** of any BSE test to be the most important criterion. Field testing with the Prionics Check LIA has therefore been undertaken over the last 6 months in a number of laboratories in Europe. At the State Institute for Consumer Protection and Food Safety, Hannover, Germany, (Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Veterinärinstitut Hannover, Dr. Brigitte Thoms) the Prionics Check LIA was sucessfully tested in comparison to a previously evaluated BSE test. During the same field application test, a series of blinded positive and negative standardized homogenates were successfully analyzed to demonstrate the sensitivity and specificity of the test under routine conditions (provided by Prof. Dr. Habermehl, Gesellschaft f. Biotechnologische Diagnostik mbH, Berlin in collaboration with BMBF and Prof. Dr. M. Groschup as part of a study to produce standardized BSE material for ring trials). Furthermore, the LIA assay protocol has been implemented on a Hamilton robotic station and tested with 100 % sensitivity and specificity for 928 negative samples (Swiss surveilance and New Zealand negatives) and 28 positive samples (ascertained BSE cases) in the lab of SQTS (Swiss Quality Testing Services, Courtepin, Switzerland). These results

clearly show the suitability of the Prionics Check LIA for routine surveillance. Further tests in private as well as offical European government laboratories are ongoing (results available upon request).

Conclusion

As a conclusion from all these validation and field studies we can only reiterate the excellent sensitivity and specificity, and - in light of the **simple handling procedure** (only pipetting steps) - emphasize that the Prionics Check LIA has been designed as a simple ELISA system which can easily be handled by any diagnostic laboratory.

D. Comments by UCSF

December 7, 2001

IND(UCSF) Comments on Evaluation of CDI by IRMM

Jiri G. Safar[1,2] and Stanley B. Prusiner[1,2,3]*

[1]Institute for Neurodegenerative Diseases, Departments of [2]Neurology and of [3]Biochemistry and Biophysics, University of California, San Francisco, CA

***To whom correspondence should be addressed: University of California, Institute for Neurodegenerative Diseases, Box 0518, San Francisco, CA 94143-0518. Tel: (415) 476-4482;**

General Comments. The conformation dependent immunoassay (CDI) evaluated by EC is the first fully automated and scaled-up version of a manual test for the diagnosis of BSE in bovine (Bo) brain tissue. The CDI is based upon measuring the immunoreactivity of an epitope in PrP that is exposed in PrP^{C} but buried in PrP^{Sc}. Because the CDI does not require protease digestion to remove PrP^{C}, both protease sensitive and protease resistant PrP^{Sc} are measured. Thus, the CDI is different from all other immunoassays currently validated for the detection of bovine PrP^{Sc}.

Incorporating newly developed high-affinity recombinant antibody fragments (Fabs) reacting with residues 95-105 of BoPrP, the CDI can detect PrP^{Sc} in BSE-infected bovine brains. Our observations indicate that the automated CDI is a robust diagnostic system for the identification of prion-contaminated tissue in mammals.

The performance characteristics of the automated CDI for the detection of PrP^{Sc} in BSE-infected brainstem homogenates include:

1. Validated flow-through of 376 tests per 10 hours is readily scalable to 1410 tests per 10 hours while maintaining the already validated assay specificity and sensitivity,
2. Automated sample tracking, quality control and data evaluation,
3. High diagnostic sensitivity for the detection of PrP^{Sc} in brainstems collected postmortem with no false positive or false negative results,
4. Signals from positive BSE samples are up to 1000 times greater than those recorded for negative samples,
5. Detection limit approximates the sensitivity of bovine bioassays for BSE prions,
6. Potential for rapid prion strain typing in large numbers of samples.

CDI and transgenic mouse bioassays exhibit similar sensitivity. As evidenced in the EC validation study and according to our ongoing experiments, the concentration of PrP^{Sc} in BSE-infected brain homogenates vary up to eight-fold due to tissue sampling of the bovine brainstems, up to three-fold due to the exposure of the homogenate to room temperature for 4 hours, and may diminish up to 30% after one freezing/thawing cycle. Additional factors include the denaturation of PrP^{Sc} during homogenization and dilution as also mentioned in the EC evaluation. All the above factors have a cumulative impact on the measurement of the PrP^{Sc}. Thus, the only correct conclusion about the analytical sensitivity of the detection method can be drawn from direct comparison with bioassays using the same tissues and the same protocols for tissue processing.

To compare the sensitivity of the automated CDI with that of bioassay, we prepared homogenates from the medulla of a Hereford bull with BSE (case PG31/90). Three separate titration series performed in parallel in Tg(BoPrP$^{+/+}$)*Prnp*$^{0/0}$4092 mice gave an average end-point titer of $10^{6.9}$ ID_{50} units/g of

brain tissue. This finding compares with titers of $10^{3.1}$ mouse LD_{50} units/g of BSE-infected brain tissue titrated i.c. or i.p. in RIII mice (reported in the 8 July, 1999 report of the European Commission) on validation of new assays for BSE (Deslys, Comoy et al. 2001) and 10^6 ID_{50} units/g reported for end-point titration in cattle (Wells, Hawkins et al. 1998). These data indicate that Tg(BoPrP) mice are perhaps as much as ~10 times more sensitive than cattle and nearly 10,000 times more sensitive than standard, non-transgenic RIII mice to infection with BSE prions. A calibrated incubation time bioassay for BSE prions in Tg(BoPrP)*Prnp*$^{0/0}$ mice is shown in **Fig. 1A.** The calibration curve was used to estimate the infectivity titer of additional BSE brain samples we obtained previously from VLA, U.K.

Our results indicate that Bo PrP^{Sc} in less than 20 µg of BSE-infected tissue per well upon serial dilution can be detected by CDI. A 50% positive detection rate for the CDI was found after diluting the sample ~$10^{4.6}$-fold **(Fig. 1B)**. Likewise, a 50% positive detection rate was observed for BSE prions measured by end-point titration in Tg(BoPrP)*Prnp*$^{0/0}$after diluting the sample ~10^4-fold. Thus, both the CDI and the bioassay in Tg(BoPrP)*Prnp*$^{0/0}$ mice exhibit similar sensitivities for BSE prions.

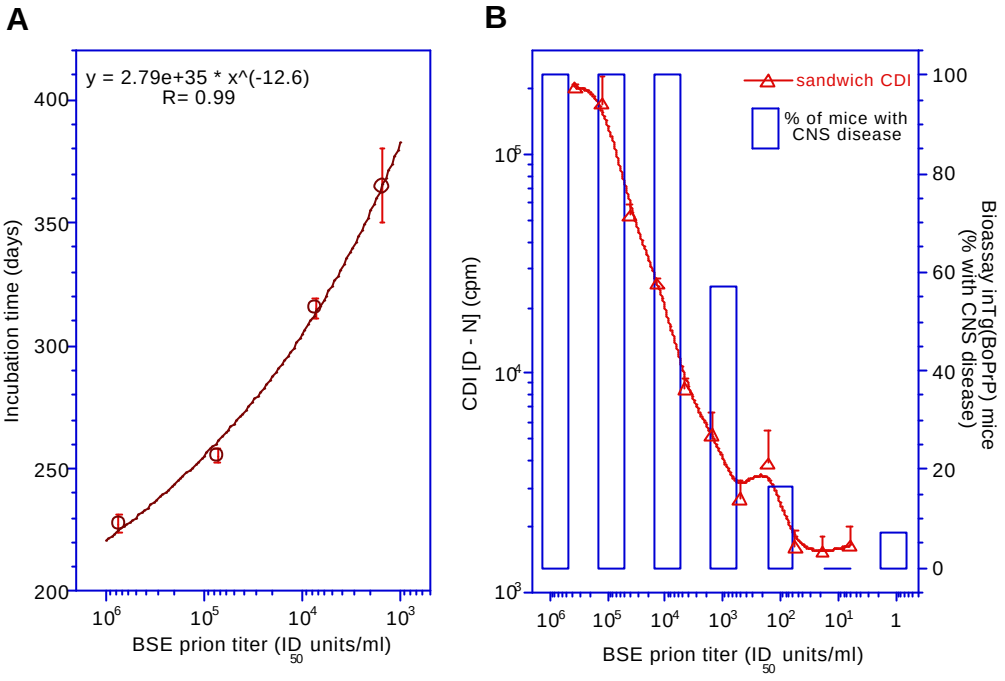
References

Deslys, J. P., E. Comoy, et al. (2001). “Screening slaughtered cattle for BSE.” Nature **409**: 476-478.

Safar, J., H. Wille, et al. (1998). “Eight prion strains have PrP^{Sc} molecules with different conformations.” Nat. Med. **4**(10): 1157-1165.

Wells, G. A. H., S. A. C. Hawkins, et al. (1998). “Preliminary observations on the pathogenesis of experimental bovine spongiform encephalopathy (BSE): an update.” Vet. Rec. **142**: 103-106.

Figure 1. Correlation between CDI and Tg(BoPrP) bioassay for quantification of BSE prions. **(A)** Inverse exponential relationship between titers of BSE prions and incubation times in Tg(BoPrP$^{+/+}$)*Prnp*$^{0/0}$4092 mice. The data points are the average ± standard error of the mean (SEM) calculated from three independent end-point titrations. **(B)** Direct relationship between CDI detection of bovine PrP^{Sc} and BSE prion titers measured in Tg(BoPrP$^{+/+}$)*Prnp*$^{0/0}$4092 mice. The 10% (for bioassay) or 5% (for CDI) BSE-infected brain homogenates were serially diluted 10-fold for bioassay or 3-fold for CDI. The (D – N) value of CDI is directly proportional to the concentration of PrP^{Sc} (Safar, Wille et al. 1998). The percentage of mice displaying CNS disease at each BSE sample dilution used in the bioassay was calculated from three independent end-point titrations.


A
y = 2.79e+35 * x^(-12.6)
R= 0.99
Incubation time (days)
400
350
300
250
200
10^6
10^5
10^4
10^3
BSE prion titer (ID_{50} units/ml)
B
sandwich CDI
% of mice with CNS disease
CDI [D - N] (cpm)
10^5
10^4
10^3
100
80
60
40
20
0
Bioassay inTg(BoPrP) mice (% with CNS disease)
10^6
10^5
10^4
10^3
10^2
10^1
1
BSE prion titer (ID_{50} units/ml)

E. Comments by MRC Prion Unit, Imperial College

Summary

This test was developed in an academic laboratory and while it is already 100% specific and sensitive for BSE cases, and detects considerable dilutions of such material, it is anticipated that full optimisation within a commercial diagnostics environment would lead to further gains in sensitivity. As the evaluators note, the test should be easily adaptable to a high-throughput format. It is also anticipated that in the course of such development, a considerable reduction in the current assay time (four hours for sample processing and two hours for immunoassay) could be achieved. The test sample used in the current protocol might also be scaled up to provide further sensitivity. Although we have used IGEN's ORIGEN electrochemiluminescence technology for PrP detection, other assay platforms could be readily utilised.

A combination of the UK Foot and Mouth Disease epidemic and our limited resources precluded extensive field testing (the UK Meat Hygiene Service required us to remove whole cattle heads from the local abattoir prior to dissecting brainstem samples). Although we did not provide an assay 'grey-zone', this would be defined during field testing of the fully commercialised assay. We envisage that all positive samples and samples in this 'grey-zone' would be retested.

BSE prions appear to be a lethal human pathogen. While relatively low sensitivity screening tests for BSE at an abattoir level may be of value in disease surveillance, epidemiological research and in detecting late pre-clinical cases that might otherwise have entered human food, the key need is for a test that can detect BSE prions in tissues of animals early in the incubation period, ideally at less than 30 months of age. The protection of the EU and other potentially exposed populations is paramount. This study has attempted to evaluate tests by comparing their ability to detect increasing dilutions of tissue from clinically affected BSE cattle. While this allows comparisons between tests to be made using the same materials, it would also be helpful to assay tissues obtained from experimentally BSE-infected cattle at various time points during the prolonged incubation period which may give data that can more accurately guide both epidemiological studies and public health protection.

Detailed comments

a) *Brain slice testing*
The current assay, although a prototype, is robust and the initial problem with the first series of measurements on days 1 and 2 was due to technical error related to unforeseen illness in a key technician causing assay controls from one workstation to be unacceptably high (although, as is stated in the evaluation report, discrimination between positives and negative brain slices remained good). The problem was fully recognised and the coded samples re-tested without incident.

b) *IRMM dilution series*
While we appreciate the difficulties for IRMM providing serial dilutions in a form that was suitable for all tests, the 80% homogenates provided were difficult to work

with and we question whether adequate mixing of spiked BSE brain tissue was actually possible. We also noted that the diluted ribolysed material was much more particulate than similar homogenates prepared from fresh or frozen brain slices. An apparent overall reduction in our assay sensitivity with these samples may be attributable to inadequate homogenisation of these particulates that were not aspirated from the ribolysed tubes and therefore not assayed.

The single false positive sample (IRMM 21626) gave ECL counts that were significantly above the assay cut-off on two independent runs with minimal replicate variability. It is possible that the samples were mixed up (possibly with the false negative 1:100 sample), due to technical error.

c) *Test Developer's dilution series*
We welcomed the opportunity to prepare our own dilution series under supervision and noted that the assay sensitivity approached that of dilution series we prepared from normal and BSE brain tissue prior to this evaluation.