COLLABORATIVE STUDY FOR THE ESTABLISHMENT OF A RAT BIOASSAY FOR INACTIVATED POLIOMYELITIS VACCINE

Project leader: David Wood[1]; Statistician: Alan Heath[1]

1. INTRODUCTION

For many years a test in chicks or guinea pigs has been specified in the Ph. Eur.[2] as the assay for IPV (*Poliomyelitis vaccine (inactivated), 1999:0214*). However, a WHO collaborative study of immunogenicity assays of IPV found that these tests are poorly standardised (Wood and Heath 1995). An alternative test in rats (van Steenis et al. 1981, Bevilacqua et al. 1996) was found to be less variable than the test in chicks or guinea pigs. After an international consultation, the Ph. Eur. Commission, through the Biological Standardisation Programme of the EDQM, initiated a project to evaluate a rat *in vivo* bioassay. This study was performed under the aegis of the Biological Standardisation Programme and supported by the Council of Europe and the European Commission.

The neutralising antibody test for poliovirus requires the use of live poliovirus. For historical reasons many laboratories use wild-type strains of poliovirus. As the goal of eradicating poliomyelitis due to wild-type polioviruses is in sight (WHO, 1999a), laboratories that use wild polioviruses will become an important potential source of accidental reintroduction of wild virus into a community. To minimise this risk WHO have developed a global action plan that requires increased biosafety containment of wild type polioviruses (WHO, 1999b). The attenuated Sabin vaccine strains of poliovirus, on the other hand, will not require increased laboratory containment until vaccination ceases completely. Therefore a standard neutralising antibody assay was proposed that used the Sabin strains of poliovirus to assay neutralising antibodies in rat sera.

On the basis of published work (van Steenis et al. 1981, Bevilacqua et al. 1996) and the promising preliminary results of this project the rat assay was introduced into the Ph. Eur. as an alternative to chicks or guinea pigs. This report provides a detailed collaborative evaluation of the rat bioassay and suggests ways in which the assay specifications may be defined in more detail.

2. AIM OF THE STUDY

The collaborative study was performed with the aim to establish the transferability of the method to new laboratories, to establish specifications for the neutralising antibody test and, in particular, to establish the most appropriate method of statistical analysis of results.

(1) NIBSC, South Mimms, Potters Bar, Herts EN6 3QG, UK.

(2) Abbreviations: $CCID_{50}$: dose infecting 50% of cell cultures; ED_{50}: Effective Dose causing 50% effect; BRP: Biological Reference Preparation; DTP: Diphtheria Tetanus Pertussis; EDQM: European Directorate for the Quality of Medicines; GCV: Geometric Coefficient of Variance; GMT: Geometric Mean Titre; I: In-house neutralising test; i.m.: intra-muscular; IPV: Inactivated Poliomyelitis Vaccine; IS: International Standard; mIU: milli International Unit; NED: Normal Equivalent Deviate; NIBSC: National Institute for Biological Standards and Control; OMCLs: Official Medicines Control Laboratories; Ph. Eur: European Pharmacopoeia (Pharmacopée Européenne); S: candidate standard neutralising test; WHO: World Health Organisation.

3. PARTICIPANTS

Eight laboratories from seven countries, including both manufacturers (4 laboratories) and national Official Medicines Control Laboratories (OMCLs) participated. Of the eight laboratories only three had previous experience of an assay for IPV in rats. Participants are listed at the end of the report.

4. MATERIALS AND METHODS

4.1. MATERIALS

Participants were provided with two antigens. One was the Ph. Eur. BRP for IPV[(1)] (Tummers 1996). This is a commercially produced trivalent IPV with an assigned antigen content of 430, 95 and 285 D-antigen units per ml for poliovirus types 1, 2 and 3 respectively (Wood et al. 1997). The second antigen was coded 90/716. This is a commercially produced trivalent IPV previously studied in WHO studies of immunogenicity (Wood and Heath 1995) and antigenicity (Wood et al. 1995). The mean antigen content was found to be 38, 8 and 30 D-antigen units per ml for poliovirus types 1, 2 and 3 respectively. The human dose formulation of IPV is commonly considered to be 40, 8 and 32 D-antigen units per ml.

Participants were also provided with the 2nd IS for anti-poliovirus antibodies, code number 66/202 (Wood and Heath 1992). This is a human serum pool with an assigned activity of 25 IU of type 1, 50 IU of type 2 and 5 IU of type 3 anti-poliovirus antibodies.

4.2. RAT BIOASSAY

A candidate standard rat bioassay method was specified in the study protocol. Laboratories were asked to use this method unless they had previously established a rat bioassay, in which case they were asked to use their own established method. The candidate standard method used specific pathogen free female Wistar rats weighing between 175-250 g. The animals were shipped in filtered crates and participants were requested to inoculate the animals within 7-14 days of receipt and to take steps to minimise the risks of intercurrent infection occurring during the test, which is considered to influence the immunogenicity of IPV in rats (Minor 1990). Each dilution of IPV was to be inoculated into a group of 10 animals, with 0.5 ml i.m. per animal. Four out of five dilutions from a two-fold dilution series were to be inoculated per preparation. The penultimate dilution was not inoculated into animals to enable an end-point to be obtained for poliovirus type 2 which previously had been found to be more immunogenic in rats than poliovirus types 1 or 3 (Bevilacqua et al. 1996). Animals were to be bled after 21 ± 1 days and serum stored at – 20 ºC or colder. Details of the rat bioassays used by participants are given in Table 1.

Table 1 — *Details of the rat bioassay used by the participants*

Laboratory	Rat strain	Weight (min/max) (g) on receipt	Range of days from from receipt to inoculation	Days (min/max) from inoculation to bleeding
1*	Wistar	175-250	7-14	20/20
2	Wistar	Around 200	2-11	21/21
3	Wistar	Not given	4-18	20/27
4	Wistar RIV-tox	131-333	7-20	21/21
5	Wistar	Not given	38	21/21
6	Wistar	175-200	7	21/21
7	Wistar RIV-tox	175-200	19-33	21/21
8	Sprague-Dawley	220-239	7-9	21/21

* Laboratory 1 used an inoculum volume of 0.2 ml per rat.

(1) EDQM catalogue number P2160000.

4.3. NEUTRALISING ANTIBODY TEST

Participants were requested to assay all sera for neutralising antibodies to all three poliovirus serotypes using an established in-house neutralising antibody test. Details of the in-house neutralising antibody tests used are given in Table 2. All sera were to be retested with a candidate standard neutralising antibody test that used a 100 $CCID_{50}$ challenge of each of the Sabin polioviruses and Hep2C indicator cells. The neutralising conditions were to be 3 hours at 35-37 ºC followed by 18 hours at 2-8 °C and the assay was to be read after 7 days incubation at 35 °C.

Table 2 — *Details of in-house neutralising antibody tests*

Laboratory*	Neutralisation conditions (temp. in °C/hrs)	Cell line	Assay temperature and duration (°C/days)
2	37/3 + 5/18	Hep2	36/7
4	35.5/3 + 4-8/18	Vero	35.5/7
5	37/5 + 5/20	Vero	37/7
6	36.5/3 + 2-8/overnight	Vero	36.5/5-7
7	37/5	Vero	37/5
8	37/20	Not given	37/7

* Laboratories 1 and 3 did not have an in-house method.

4.4. STUDY DESIGN

Participating laboratories were requested to perform three independent rat bioassays. Each assay included the materials coded Ph. Eur. BRP and 90/716. Laboratories were provided with a suggested dilution series to use, but requested to evaluate the dose-response of the first test and adjust the dilution series for tests 2 and 3 if necessary. Participants were requested to assay all sera for poliovirus neutralising antibodies by (a) the laboratories routine in-house method and (b) the candidate standard neutralising antibody test. Laboratories were also requested to calibrate both neutralising antibody tests in mIU, using the 2nd IS for anti-poliovirus types 1, 2, 3 antibodies. A code number, allocated at random, refers to participating laboratories. Assays analysed using data from the candidate standard neutralising test or the in-house neutralising test are referred to as S or I respectively, and have been treated as independent assays for the purposes of this study.

4.5. STATISTICAL APPROACH

Both the probit method and the parallel line regression model for quantitative data have previously been proposed as appropriate methods for analysis of rat bioassay data (Ph. Eur. 1999:0214, van Steenis et al. 1981, Bevilacqua et al. 1996). For data that is measured on a quantitative scale rather than a binary (yes/no) scale, the parallel line model is often the most efficient method. Reducing quantitative data to a binary response may reduce the information content. However, the parallel-line method requires certain assumptions regarding the nature of the statistical distribution of the data, in addition to the requirements of linearity and parallelism, for the approach to be valid. In particular, the data are usually assumed to follow a normal distribution, with a constant variance at each dose level. When considering a method of analysis for use in a batch release context, particular attention must be given to the validity criteria to be applied. A batch release assay will often be performed in isolation, and the validity tests give a necessary check on the reliability of the results. The data from this study have been analysed using both the probit method and the parallel line method. The validity criteria specified in the Ph. Eur. for both methods have been applied (Ph. Eur. Chapter 5.3.).

4.6. VALIDITY CRITERIA - PROBIT METHOD

The probit method can be applied to qualitative (binary) response data. The requirements for an assay to give a statistically valid fit to the probit model are:
a) the overall assay dose-response is significant;
b) there are no significant deviations from parallelism;
c) there are no significant deviations from linearity.

For this study, an overall linearity test was used, rather than individual tests of linearity for each preparation. The usual tests for significance of dose-response, linearity and parallelism were applied, using an in-house probit program developed at NIBSC.

4.7. VALIDITY CRITERIA - PARALLEL LINE ASSAY

The parallel line method can be applied to quantitative response data. The requirements for an assay to give a statistically valid fit to the parallel model are:
a) the responses to each treatment (dose) group are normally distributed;
b) the variances of the responses to each treatment group are homogeneous;
c) the overall assay dose-response is significant;
d) there are no significant deviations from parallelism;
e) there are no significant deviations from linearity.

The response measure may be transformed to better approximate the above conditions. For this study, an overall linearity test was used, rather than individual tests of linearity for each preparation.

The usual tests for significance of dose-response, linearity and parallelism were applied, using an in-house parallel-line program developed at NIBSC. The normality of distributions was tested using the Shapiro-Wilks test, as suggested by the Ph. Eur. (General Chapter 5.3.), using the SAS statistical package . The test was performed for each treatment (dose) group, giving a p-value which was converted to a single degree of freedom NED. The NEDs were squared and summed over all treatment (dose) groups within an individual assay to give an overall chi-square test of normality. The homogeneity of variance for each assay was tested using Bartlett's test, as suggested by the Ph. Eur. (General Chapter 5.3.), using a program implemented in the SAS statistical package. Prior to implementation of the test, any treatment (dose) groups with zero variance were excluded.

5. RESULTS

5.1. DATA RECEIVED

Eight laboratories returned data. Laboratories 2, 4, 5, 6 and 7 returned data from both in-house and candidate standard neutralising tests. Laboratories 1 and 3 returned data from the candidate standard neutralising test only, and did not have a separate in-house method. Laboratory 8 returned data from an in-house neutralising test only. Laboratory 8 stated that although there were some differences between their in-house method and the candidate standard method, the use of the IS (66/202) to calibrate the test made these differences quite irrelevant. All participants returned data from three independent bioassays as requested, with the exception of laboratory 5 which returned data from two; they omitted the data from the first of their three assays due to the rats being too heavy. All participants used the recommended four dose groups for each preparation, although some made minor changes to the actual doses used.

Hence, there were 37 individual sets of assay data for analysis, consisting of 74 individual dose-response curves (2 preparations per assay) with four dose levels each, for each of the

three polio types giving a total of 222 dose response curves. The data from the in-house and standard neutralisation tests from the same laboratory are not independent however, being tests conducted on the same sera. All laboratories except 4 and 5 returned calibrations of their neutralising tests against the IS, for each individual assay. Laboratory 4 returned a geometric mean for the six assays (3 each with in-house and standard neutralising tests). Laboratory 5 failed to obtain neutralisation with their initial dilution of the IS and was thus unable to provide any data.

5.2. CALIBRATION OF NEUTRALISATION TESTS

Participants were asked to convert a titre of 8 in their neutralising antibody assay(s) into mIU. To do this the IS was titrated in each assay. The assigned potency of this material is 25.000 mIU of antibody to type 1, 50.000 mIU of antibody to type 2, and 5.000 mIU of antibody to type 3 poliovirus (Wood and Heath 1992). Thus if a laboratory obtained an end point titre of 256, 256, and 64 to types 1, 2 and 3 respectively, then a titre of 8 would be equivalent to 781 ($8/256 \times 25.000$), 1.562 ($8/256 \times 50.000$), and 625 ($8/64 \times 5.000$) mIU of antibody to types 1, 2 and 3 poliovirus respectively.

Geometric means of the participants' mIU values over all assays were calculated, keeping the results for the in-house and standard neutralisation tests separate where possible. The exact method of calculation used by the participants was not always clear, however. There is a possible query with the results provided by laboratory 6, as they appear to have quoted assigned potencies for the IS 10-fold lower than the correct unitages. Their calculated figures, which are in close agreement to those of other laboratories, are used below.

The geometric means of neutralising antibody in mIU corresponding to a titre of 8 are given in Table 3. The figures represent a variation in sensitivity between laboratories and methods of around 5-fold (maximum/minimum) for types 1 and 2, and around 10-fold for type 3. Inter-laboratory agreement was better with the standard method than with the in-house methods. A comparison of the sensitivity of the in-house and standard methods was only possible within three laboratories (2, 6 and 7) where 66/202 had been titrated with both methods. For laboratories 6 and 7, the figures in mIU for a titre of 8 were higher for the standard method than for the in-house method for all three polio types. However the extent of the differences was not large, the standard method figures varying from 1.2-fold to 2.5 fold higher than the in-house method figures. For laboratory 2, the in-house and standard methods gave identical figures for types 1 and 2, but for type 3 the in-house test gave a figure over 4-fold higher than the standard test.

Table 3 — *Calibration of neutralising antibody tests (geometric means). Neutralising antibody in mIU corresponding to a titre of 8.*

Laboratory	Method	Type 1 (mIU)	Type 2 (mIU)	Type 3 (mIU)
1	S	131	282	57
2	I	248	688	240
2	S	248	688	55
3	S	137	344	105
4	I + S	287	440	109
6	I	81	132	47
6	S	147	290	114
7	I	109	162	86
7	S	263	415	104
8	I	63	126	25
Max/Min	S	263/137 (2)*	688/290 (2)*	114/55 (2)*
Max/Min	I	248/63 (4)*	688/126 (5) *	240/25 (10)*

S = candidate standard neutralising test; I = in-house neutralising test.

* Figures in brackets are the fold-differences, rounded to the nearest integer, between maximum (Max) and minimum (Min) values.

5.3. CORRELATION OF NEUTRALISATION TEST CALIBRATIONS AGAINST THE IS 66/202 WITH OBSERVED TITRES IN RATS

The IS 66/202 is a human serum pool. If poliovirus neutralising antibodies in this preparation are qualitatively similar to neutralising antibodies in rat sera from an IPV test, a correlation should be observed between the calibration in mIU of the neutralising antibodies and the overall level of observed titres in rat sera for a fixed dose of a particular vaccine across laboratories.

The GMTs obtained for the 1:10 dilution of the Ph. Eur. BRP were calculated for each laboratory and polio type, and are given in Table 4. Laboratory 7 used a different set of dilutions for their assays, which did not include a 1:10 dilution of Ph. Eur. BRP, and so can not be included in this table. There is more variation between laboratories when results are expressed in GMT, with ranges of around 9-fold for type 1, 28-fold for type 2 and 34-fold for type 3, than variation between laboratories when results are expressed in mIU. This is consistent with a previous study of the poliovirus neutralising antibody test (Wood and Heath 1992). The variation between laboratories was identical for the standard and in-house methods for GMTs of poliovirus type 1, but for types 2 and 3 the variation was greatest with the standard method (Table 4). The GMTs across laboratories are type 1 - 12.7; type 2 - 575.5 and type 3 - 35.2 which confirms poliovirus type 2 as the most immunogenic antigen in IPV when inoculated into rats.

Table 4 — *Geometric mean titre for 1:10 dilution of Ph. Eur. BRP*

Laboratory	Method	Type 1	Type 2	Type 3
1	S	15.1	236.4	7.5
2	I	19.2	256.1	36.9
2	S	15.5	175.3	68.9
3	S	40.2	389.6	20.4
4	I	4.1	357.3	19.0
4	S	5.1	528.3	32.8
5	I	4.6	1844.2	24.9
5	S	20.8	4913.1	166.3
6	I	32.4	1520.6	254.7
6	S	16.5	484.4	17.9
8	I	7.5	440.6	22.7
Max/Min	S	40/5 (8)*	4913/175 (28)*	166/7.5 (22)*
Max/Min	I	32/4 (8)*	1844/256 (7)*	254/19 (12)*

Laboratory 7 did not use a 1:10 dilution.

* Figures in brackets are fold-differences, rounded to the nearest integer, between maximum (Max) and minimum (Min) titres.

The correlations with the calibration of neutralisation tests (mIU for titre of 8) are shown in Figures 1-3 for polio types 1-3 respectively. It can be seen that there is no clear correlation for any of the polio types (Spearman correlation coefficient not significantly different from zero). Laboratory 4 only provided a calibration in mIU as a mean of their in-house and standard tests, and these figures thus appear twice in the plots, against the observed titres for Ph. Eur. BRP from standard and in-house tests. Analysis was repeated using the 1:20 dilution of Ph. Eur. BRP, and also with overall mean titres across the entire assay, but with similar results (data not shown).

Figure 1 — *Correlation of geometric mean titres (1:10 Ph. Eur. BRP) and mIU for* ***poliovirus type 1***

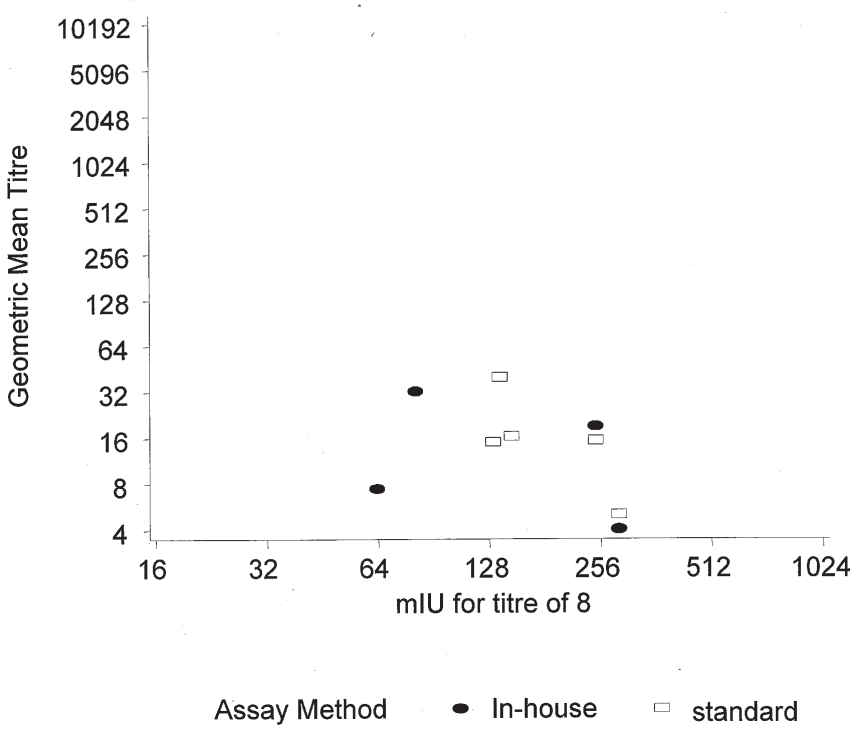


Figure 2 — *Correlation of geometric mean titres (1:10 Ph. Eur. BRP) and mIU for* ***poliovirus type 2***

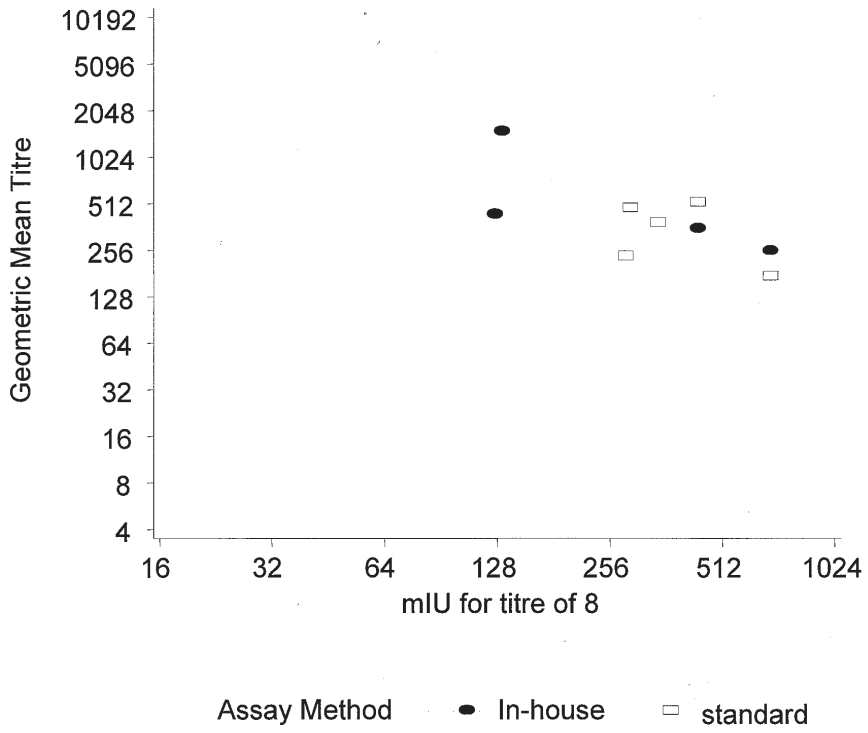
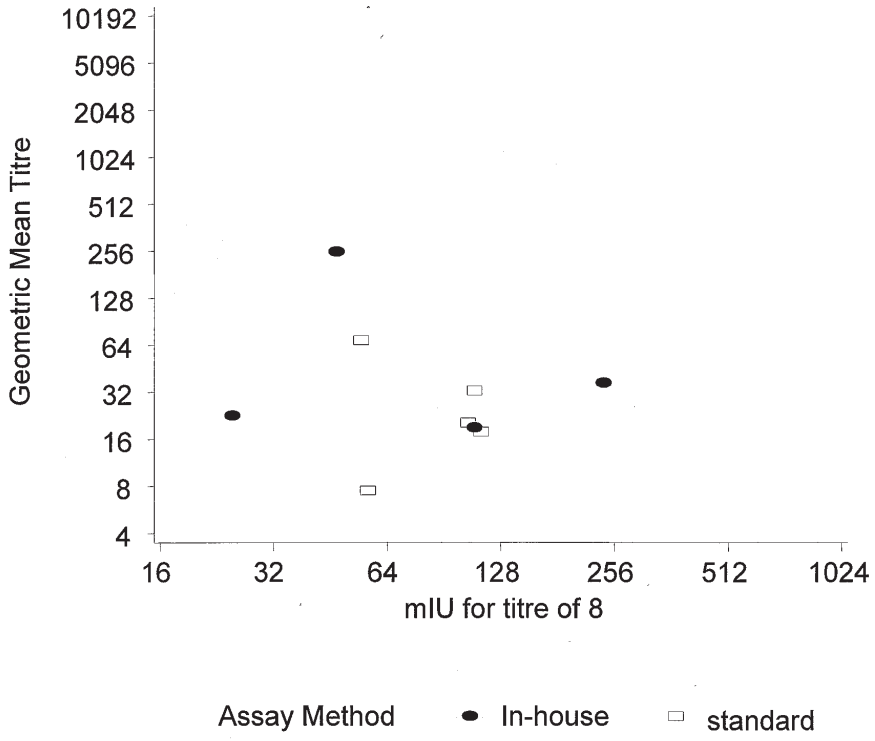


Figure 3 — *Correlation of geometric mean titres (1:10 Ph. Eur. BRP) and mIU for* ***poliovirus type 3***



The above analysis could be confounded by differences in sensitivity of rats between laboratories. Laboratories 2 and 6 provided data for the same sera tested with neutralising antibody tests of differing sensitivity. Table 5 shows that, at least for poliovirus types 2 and 3, there is no clear correlation between sensitivity of neutralising antibody tests as shown by calibration in mIU and titration of rat sera. It can therefore be concluded that the calibration of neutralisation test results against the IS 66/202 will not improve agreement in observed antibody levels from rat tests between laboratories.

Table 5 — *Correlation between calibration in mIU and observed titres for rat sera*

Laboratory	Method	Type 1		Type 2		Type 3	
		Titre of 8 in mIU	GMT of Ph. Eur. BRP 1/10	Titre of 8 in mIU	GMT of Ph. Eur. BRP 1/10	Titre of 8 in mIU	GMT of Ph. Eur. BRP 1/10
2	I	248	19.2	688	256.1	240	36.9
	S	248	15.5	688	175.3	55	68.9
6	I	81	32.4	132	1520.6	47	254.7
	S	147	16.5	290	484.4	114	17.9

5.4. CHOICE OF CUT-OFF FOR THE PROBIT METHOD

To apply the probit method to the rat bioassay data, it is necessary to define a positive or negative response. This reduces to the choice of a cut-off value for each polio type, where any observed titre greater than the cut-off is considered positive. The most obvious definition would be for any observed titre greater than 2 to be considered positive. However, with the current data this would lead to too many positive responses, and an inability to fit a dose-response for many assays. The range of titres observed for the same dose of Ph. Eur. BRP

between laboratories (Table 4) suggests that a single cut-off for observed titre would not be suitable for all laboratories. The results of the calibration of neutralisation tests in mIU also suggest that specifying the cut-off in mIU will not improve the suitability of a single cut-off for all laboratories.

Individual cut-offs were thus calculated for each laboratory, tailored to the range of titres observed by the laboratory. For each laboratory and method (in-house or standard neutralisation tests) the GMT for all three assays at each dose of the Ph. Eur. BRP was calculated, for each polio type. A cut-off was calculated as the mid-point (on a log scale) between the minimum and maximum, rounded down to the nearest observable titre on a doubling scale (so for example a mid-point of 5.7 would be rounded to 4; a mid-point of 12.6 to 8). An individual rat was classified as having a positive response if the observed titre was strictly greater than the cut-off. The calculated cut-offs are shown in Table 6. The overall geometric mean cut-offs across all labs were type 1 - 5.2; type 2 - 109.1; type 3 - 10.4. The variation in cut-offs between laboratories was 8-fold for types 1 and 3, and 16-fold for type 2 and the difference between the standard and in-house methods were very similar (Table 6).

Table 6 — *Calculated cut-offs for each laboratory and method. Individual rats defined as positive responder if observed titre strictly greater than cut-off.*

Laboratory	Method	Type 1	Type 2	Type 3
1	S	4	64	4
2	I	8	32	16
2	S	4	32	16
3	S	8	64	4
4	I	2	64	4
4	S	2	64	8
5	I	4	256	8
5	S	8	512	32
6	I	8	256	32
6	S	4	64	4
7	I	16	512	32
7	S	8	256	16
8	I	4	64	8
Max/Min	S	8/2 (4)*	512/32 (16)*	32/4 (8)*
Max/Min	I	16/2 (8)*	512/32 (16)*	32/4 (8)*

* Figures in brackets are fold-differences between maximum (Max) and minimum (Min) cut-offs.

5.5. Assay validity with the probit method

The probit method was applied to the data using the calculated cut-offs to define positive and negative response. Of a total of 111 assays, the NIBSC probit program failed to converge or fit a model in two cases. These were the in-house and standard neutralising antibody results from lab 2, assay 3, type 2. In one further case (lab 5, assay 2, type 2) the program failed to obtain chi-square tests of parallelism or linearity, although a potency estimate was obtained.

Of the remaining assays, all but 2 had a significant dose-response at the 5 % level. Considering the parallelism and linearity as well, 92 assays (83 % of the total 111 assays) were statistically valid using 5 % significance levels. This comprised 29/37 for type 1; 30/34 for type 2; and 33/37 for type 3. If the parallelism and linearity are tested at a 1% significance level, these figures rise to 98 (88 % of the total 111 assays) overall, comprising 32/37 for type 1; 32/34 for type 2 and 34/37 for type 3. One assay (lab 8, assay 3, type 1, in-house) that was assessed to be statistically valid by the above criteria had zero responders for all dose levels of sample 90/716 and so no potency estimate could be obtained. The overall validity figures are summarised in Table 7.

Table 7 — *Assay validity using 5 % or 1 % significance level. Percentage of the total 111 assay data sets satisfying the validity criteria.*

Validity Criteria	Probit (5 %)	Parallel Line (1) (5 %)	Parallel Line (2) (5 %)	Probit (1 %)	Parallel Line (1) (1 %)	Parallel Line (2) (1 %)
Normality + Homogeneity of Variance	-	10.8 %	28.8 %	-	20.7 %	52.3 %
Dose Response, Linearity and Parallelism	82.9 %	55.0 %	34.2 %	88.3 %	71.2 %	47.7 %
All	82.9 %	8.1 %	11.7 %	88.3 %	16.2 %	25.2 %

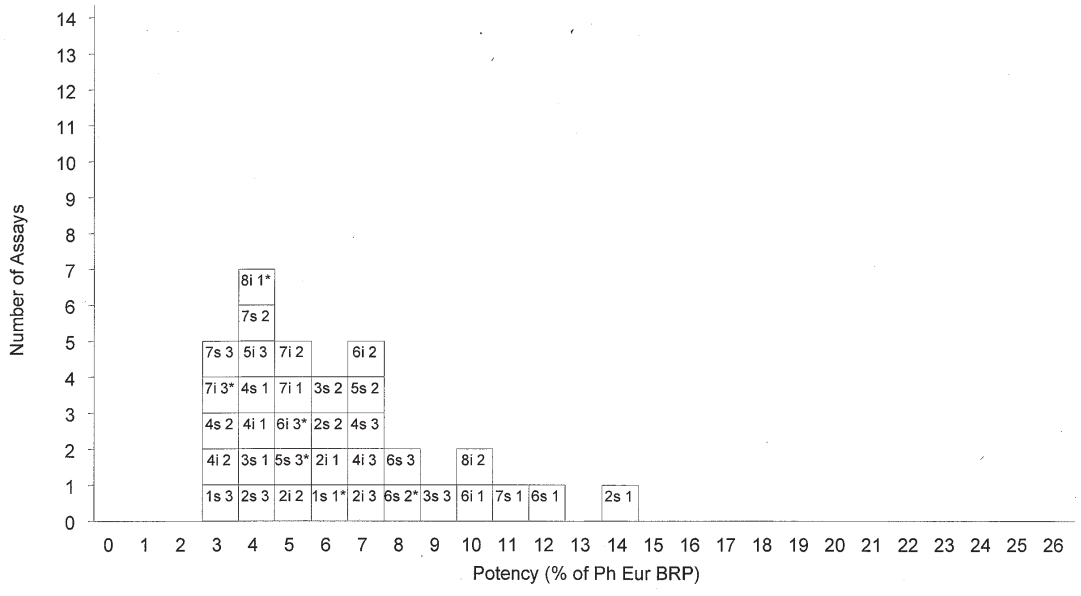
(1) Excluding dose groups with all responses identical.

(2) Excluding dose groups with greater than 50 % of responses identical.

5.6. POTENCY ESTIMATES FROM THE PROBIT METHOD

The potencies of 90/716 relative to the Ph. Eur. BRP calculated from individual assays are displayed in histogram form in Figures 4-6 for polio types 1-3 respectively. The two assays that did not have a significant overall dose-response (lab 1, assay 2, type 1, standard and lab 5, assay 2, type 1, in-house) are excluded from the plots, but other invalid assays (non-linear or non-parallel) are included.

Figure 4 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the probit method for **poliovirus type 1**. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*



The statistically invalid assays do not appear to give results any more discrepant than the other assays. The spread of estimate is greater for type 2 than for types 1 and 3. In particular, there are two assays with very high estimates for type 2. In one of these (lab 1, assay 2) there was not a significant dose-response for sample 90/716, although the overall dose-response for the assay was significant. There is no obvious explanation for the other discrepant assay.

Figure 5 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the probit method for **poliovirus type 2**. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*

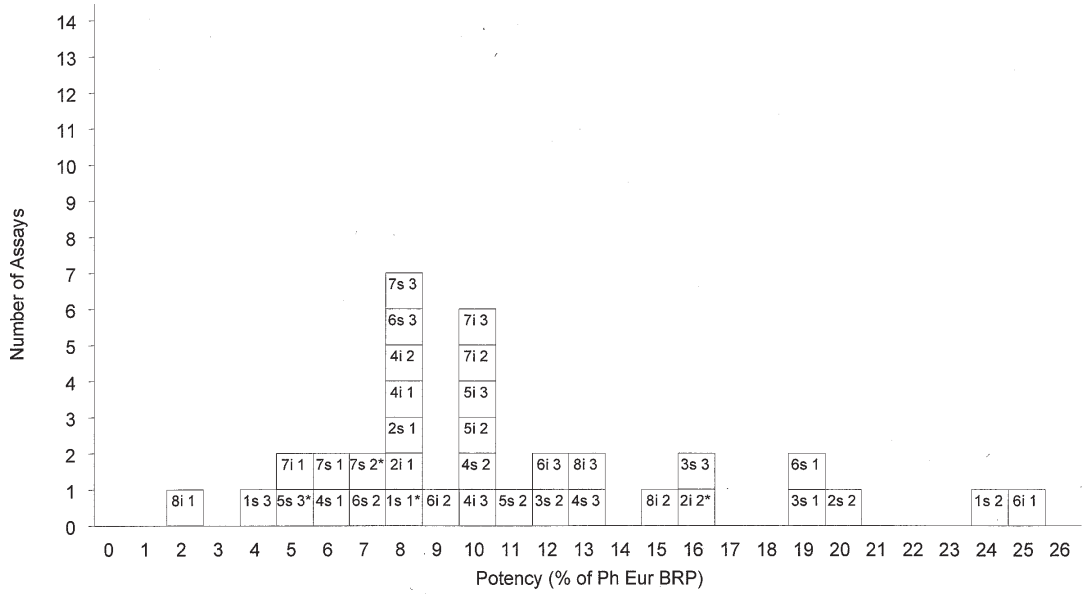
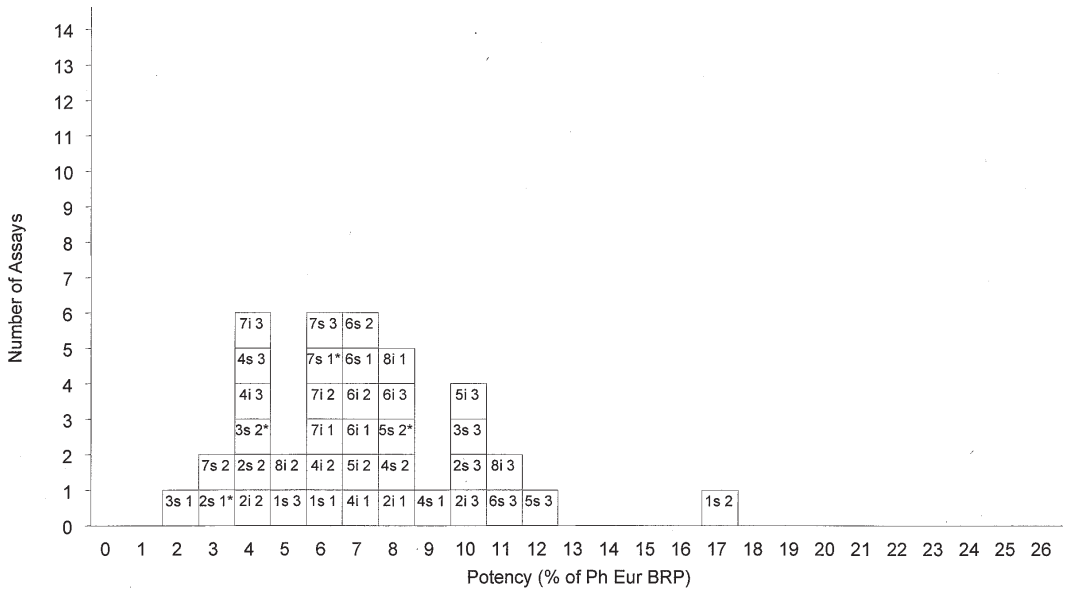


Figure 6 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the probit method for **poliovirus type 3**. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*



The overall geometric means, GCV, minimum, maximum and fold range for the three polio types are given in Table 8. The % GCVs illustrate the increased variability of the type 2 results, as does the fold-range of 11, compared to 5 or 8 for types 1 and 3. If the two assays with particularly high potency results were excluded, the fold-range for type 2 would also be closer to 8-fold. Inter-laboratory variation was similar for poliovirus type 1 whether the standard or in-house neutralising antibody assays were used, was greater for the in-house method for type 2, and greater for the standard method for type 3.

The assigned D-antigen units for the two materials were 430, 95, and 285 for Ph. Eur. BRP and 38, 8 and 30 for 90/716 for poliovirus types 1, 2 and 3 respectively. This would give theoretical potencies (D-antigen) for 90/716 relative to Ph. Eur. BRP of – 9.3 %, – 8.4 % and – 11.2 % for types 1, 2 and 3. The overall mean estimated from the rat assay (Table 8) were – 5.7 %, – 9.6 % and – 6.4 % for types 1, 2 and 3. This leads to a comparison of rat to D-antigen units of 61 %, – 114 % and – 57 % for types 1, 2 and 3. Preparation 90/716 had been stored at 2-8 °C for several years prior to this study. Previous experience has shown that there may not be good agreement between antigen content and immunogenicity of IPV preparations stored at 2-8 °C for long periods of time (Wood and Heath 1995, van Steenis et al. 1981, Bevilacqua et al. 1996). The antigenicity of 90/716 relative to the Ph. Eur. BRP should therefore be investigated.

Table 8 — *Distribution of potency estimates from probit analysis by neutralising assay method*

Polio Type	No./type of assays		Geometric Mean Potency (% of Ph. Eur. BRP)	% GCV	Minimum Potency (% of Ph. Eur. BRP)	Maximum Potency (% of Ph. Eur. BRP)	Fold-Range
Type 1	15	I	5.36	44.5	2.78	9.83	3.5
Type 1	19	S	5.91	60.1	2.64	14.19	5.4
Type 1	34	All	5.66	53.0	2.64	14.19	5.4
Type 2	16	I	9.52	69.5	2.36	25.47	10.8
Type 2	19	S	9.72	68.5	3.63	23.87	6.6
Type 2	35	All	9.63	67.7	2.36	25.47	10.8
Type 3	17	I	6.64	37.2	3.56	11.14	3.1
Type 3	20	S	6.24	70.4	2.27	17.24	7.6
Type 3	37	All	6.42	55.6	2.27	17.24	7.6

5.7. Assay validity for the parallel line method

For parallel line analysis data that were reported as, for example, < 2 were taken as 1.0. Data reported as greater than the upper limit of the dilution series used in the neutralising antibody test, for example, > 4096 were taken as 8192, i.e. the next 2-fold step. All analysis was based on log transformed titres, as inspection of plotted data indicated that this gave the best approximation to normality.

Initially the Shapiro-Wilks test for normality and the Bartlett test for homogeneity of variance were applied. Prior to applying these tests, any dose groups with zero variance were excluded as being clearly non-normal and to prevent problems with the application of the above tests. Only assays with a minimum of two doses per preparation were retained, leaving a total of 110 usable assays. Using both tests at a 5 % significance level, the numbers valid for normality were 20/110, for homogeneity 43/110, and for both 12/110 (10.8% of the total 111 assays). Using both tests at a 1% significance level, the numbers valid for normality were 30/110, for homogeneity 55/110, and for both 23/110 (20.7 % of the total 111 assays).

An alternative approach was to exclude dose groups that had more than 50 % of replicate responses identical. This left 96 usable assays (minimum of 2 doses per preparation). Using both tests at a 5 % significance level, the numbers valid for normality were 41/96, for homogeneity 66/96, and for both 32/96 (28.8 % of the total 111 assays). Using both tests at a 1 % significance level, the numbers valid for normality were 66/96, for homogeneity 80/96, and for both 58/96 (52.3 % of the total 111 assays).

These figures were recalculated, but with a requirement of a minimum of three doses per preparation, leaving only 65 usable assays. The numbers valid for both normality and homogeneity were 19/65 at 5 % (17.1 % of the total 111 assays) or 35/65 at 1% (31.5 % of the total 111 assays).

It can be seen that depending on the approach taken, and the significance level selected, anything between 50% to 90% of assays would be rejected on the criteria of normality and homogeneity of variance of the distribution of responses alone. To investigate if this was an artefact due to the range of dilutions used in the neutralising antibody test, the number of responses reported at the lower or upper limits was estimated. All laboratories used a lower limit of < 2 whereas different laboratories used different upper limits for their titrations. Laboratories 2, 4 and 6 appear to perform full titrations, with no > responses reported. Laboratory 7 reports a highest titre of 4096, with no > responses. However, from review of their data, it appears that 4096 is an upper limit, and should read greater than or equal to 4096. Of the other laboratories, the upper limits were: laboratory 1 - 362; laboratory 3 - 256; laboratory 5 - 2890 or 5790; laboratory 8 - 1024.

The overall percentages of responses that were reported at the lower or upper limits were:

	Vaccine	Lower	Upper
Type 1	90/716	51%	1%
	Ph. Eur. BRP	35%	2%
Type 2	90/716	7%	11%
	Ph. Eur. BRP	6%	12%
Type 3	90/716	33%	1%
	Ph. Eur. BRP	24%	2%

From these figures, it is clear that changing to full titrations for all laboratories would not give a great improvement to the normality of the distribution of responses as the upper limit of titration is not a problem, at least for the types 1 and 3. Some improvement for type 2 may occur. The main problem is at the lower limit of titration with a large number of antibody negative (< 2) responses. This could be solved by inoculation of more concentrated antigen doses into the animals. However the most concentrated dose of 90/716 (which is approximately human dose formulation) inoculated in the current study was undiluted material.

Final lots of IPV would thus need to be concentrated to extend the dose range but this is not recommended for routine tests. Inoculation of more concentrated doses of the Ph. Eur. BRP would be possible since the most concentrated antigen dose used in the current study was a 1/10 dilution. The overall dose response (for the probit approach) for Ph. Eur. BRP by laboratory and serotype is shown in Figures 13-15. The slope of the curves relative to the dilution step is reasonable and confirms the Ph. Eur. BRP as suitable as a reference material for this assay. Inoculation of 1/5, 1/10, 1/40 and 1/160 dilutions of Ph. Eur. BRP may be more appropriate in the future, however, to cover the relatively poor immunogenicity of polioviruses types 1 and 3 but good immunogenicity of poliovirus type 2.

The assays were then assessed for the other validity requirements; a significant dose-response, parallelism and linearity. When data from dose groups with zero variance (all responses identical) were excluded, the figures were: 2 assays did not have a significant dose-response (5 % significance level). Of the remaining 108 assays, 61 were valid (at 5 %) in terms of linearity and parallelism, comprising 18/34 for type 1, 21/37 for type 2 and 22/37 for type 3. Using a 1 % level for testing linearity and parallelism the figures were 79/108 overall, 20/34 for type 1, 29/37 for type 2 and 30/37 for type 3. Considering the tests for

normality and homogeneity as well, only 9/108 assays (8.1 % of total 111 assays) were valid (at 5 %) for all criteria, or 18/108 (16.2 % of total 111 assays) at 1 %.

When data from dose groups with greater than 50 % identical responses were excluded, the figures were: 10 assays did not have a significant dose-response (5 % significance level). Of the remaining 86 usable assays, 38 were valid (at 5 %) in terms of linearity and parallelism, comprising 5/22 for type 1, 18/33 for type 2 and 13/31 for type 3. Using a 1 % level for testing linearity and parallelism the figures were 53/86 overall, 9/22 for type 1, 26/33 for type 2 and 18/31 for type 3. Considering the tests for normality and homogeneity as well, only 13/86 assays (11.7 % of total 111 assays) were valid (at 5 %) for all criteria, or 28/86 (25.2 % of total 111 assays) at 1%. The overall validity figures are summarised in table 7.

5.8. POTENCY ESTIMATES FROM THE PARALLEL LINE METHOD

The potencies of 90/716 relative to the Ph. Eur. BRP were calculated from the assay data. As before, the analysis was repeated using data where dose groups which had all identical responses were excluded, and data where dose groups that had greater than 50 % identical responses were excluded. Where there was not a significant overall dose-response, no potency estimate was calculated. The estimated potencies for polio types 1-3 are shown in figures 7-9, and figures 10-12, for the two methods of calculation described above. Assays that were invalid for parallelism or linearity are marked with a *.

The assays invalid for linearity or parallelism give estimated potencies generally in line with other assays. The extreme estimates in figures 9 and 10 were invalid for linearity or parallelism, but those in figures 8 and 11 were not. The overall geometric means, GCV, minimum, maximum and fold range for the three polio types are given in tables 9 and 10, for the two methods of calculation. There is little to choose between the two approaches to parallel line analysis in terms of final potency estimates overall, although some individual assays give quite different estimates depending on the method used. The exclusion of groups with more than 50% identical responses results in more assays with non-significant dose-responses. However, it does improve the number of assays that are acceptable in terms of normality or homogeneity of variance.

Table 9 — *Distribution of potency estimates from parallel line analysis by neutralising assay method. Dose groups with all identical responses excluded.*

Polio Type	No./type of assays		Geometric Mean Potency (% of Ph. Eur. BRP)	% GCV	Minimum Potency (% of Ph. Eur. BRP)	Maximum Potency (% of Ph. Eur. BRP)	Fold-Range
Type 1	15	I	4.77	50.0	1.63	7.87	4.8
Type 1	19	S	5.53	38.4	2.91	9.55	3.3
Type 1	34	All	5.18	44.0	1.63	9.55	5.9
Type 2	17	I	8.35	36.7	3.58	12.9	3.6
Type 2	20	S	9.00	47.8	4.78	24.9	5.2
Type 2	37	All	8.69	42.5	3.58	24.9	7.0
Type 3	17	I	6.60	23.1	4.47	9.63	2.2
Type 3	20	S	6.16	56.6	1.80	18.8	10.4
Type 3	37	All	6.36	42.7	1.80	18.8	10.4

Table 10 — *Distribution of potency estimates from parallel line analysis by neutralising assay method. Dose groups with greater than 50 % identical responses excluded.*

Polio Type	No./type of assays		Geometric Mean Potency (% of Ph. Eur. BRP)	% GCV	Minimum Potency (% of Ph. Eur. BRP)	Maximum Potency (% of Ph. Eur. BRP)	Fold-Range
Type 1	12	I	5.25	66.4	2.71	16.21	6.0
Type 1	10	S	5.98	27.7	4.64	10.91	2.4
Type 1	22	All	5.67	50.3	2.71	16.21	6.0
Type 2	16	I	8.49	37.2	3.58	12.91	3.6
Type 2	17	S	8.29	55.3	4.12	24.93	6.1
Type 2	33	All	8.39	46.1	3.58	24.93	7.0
Type 3	17	I	6.35	22.8	4.60	10.93	2.4
Type 3	14	S	5.39	37.8	2.72	8.16	3.0
Type 3	31	All	5.90	31.2	2.72	10.93	4.0

Figure 7 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the parallel line method for **poliovirus type 1**. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*

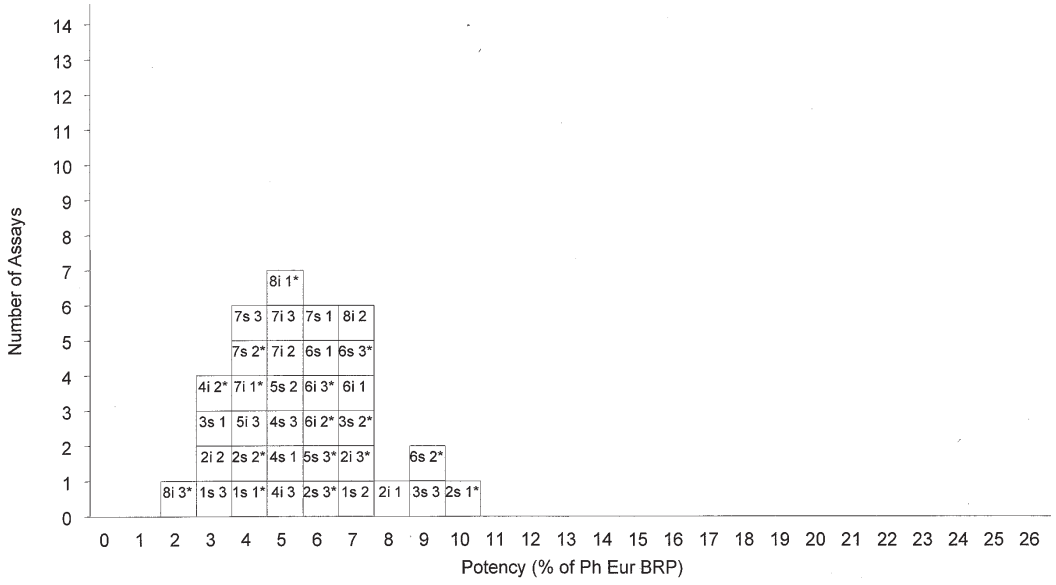


Figure 8 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the parallel line method for **poliovirus type 2**. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*

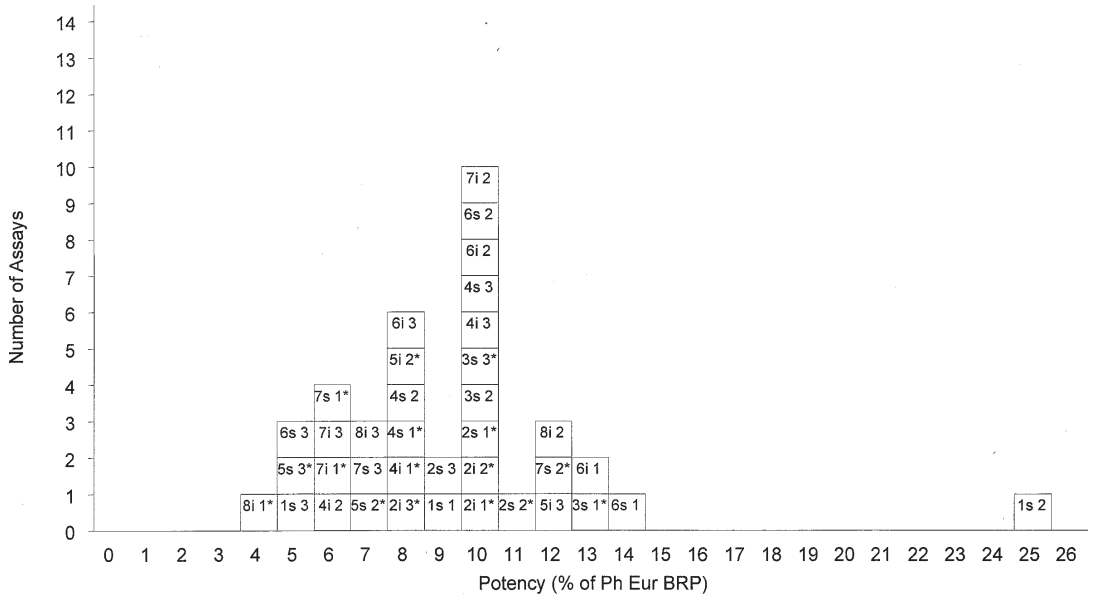


Figure 9 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the parallel line method for **poliovirus type 3**. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*

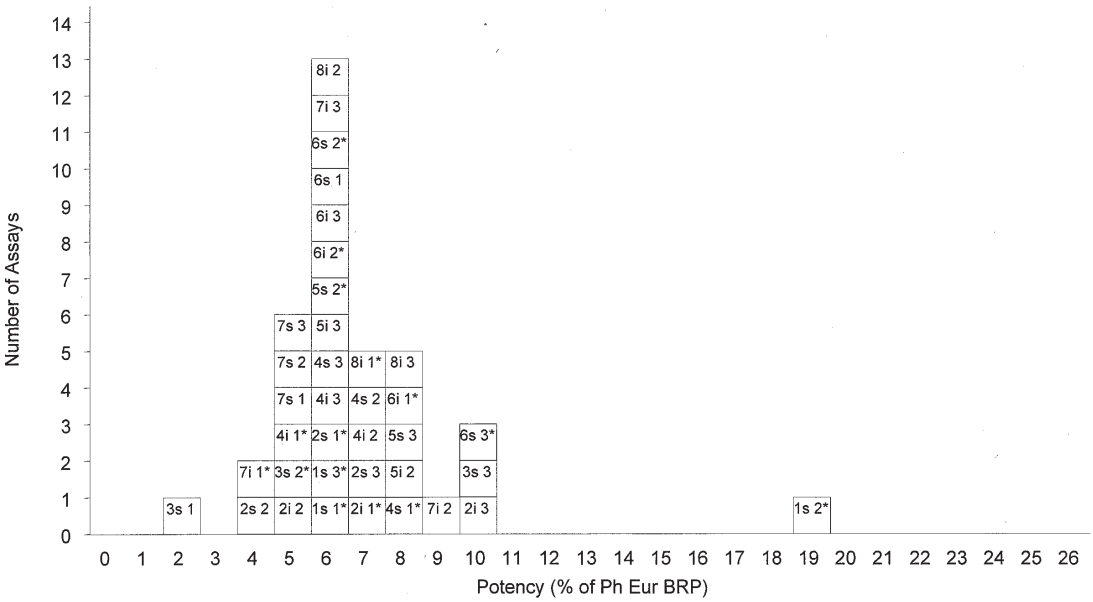


Figure 10 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the parallel line method but excluding assays with > 50 % identical responses for* ***poliovirus type 1****. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*

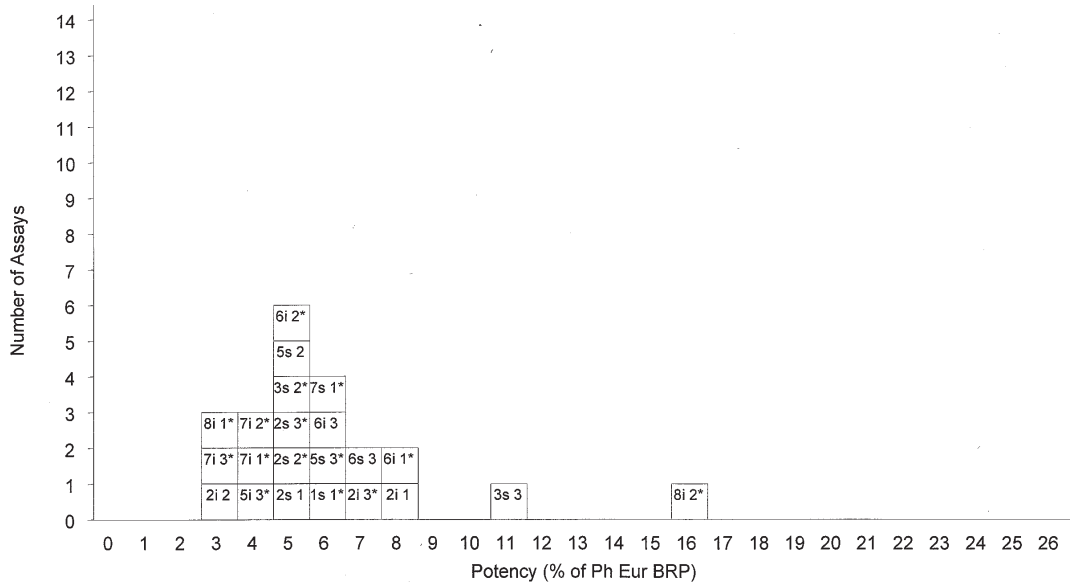


Figure 11 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the parallel line method but excluding assays with > 50 % identical responses for* ***poliovirus type 2****. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*

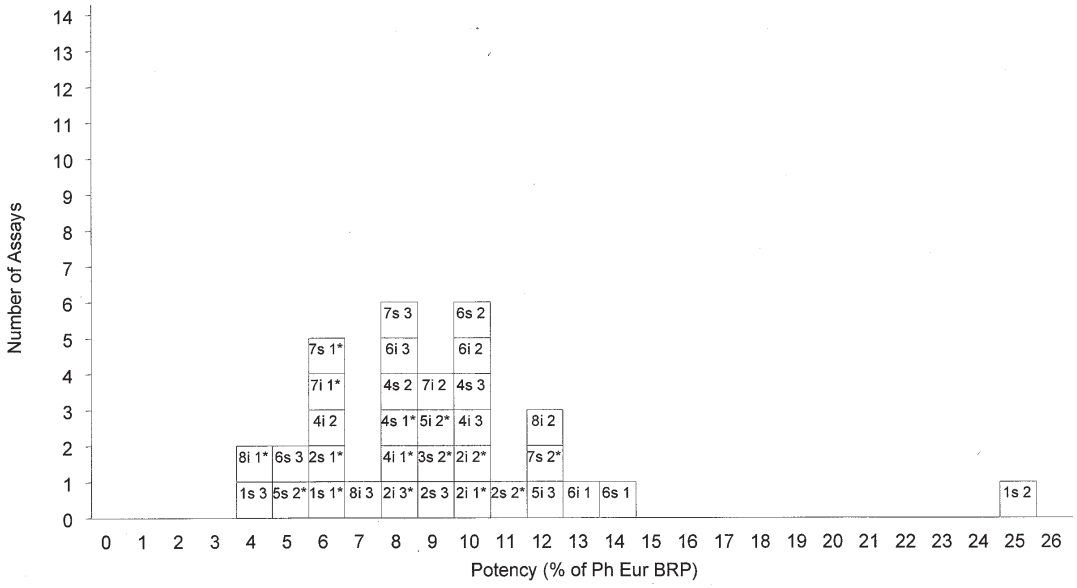


Figure 12 — *Potency estimates of 90/716 as a percentage of Ph. Eur. BRP from the parallel line method but excluding assays with > 50 % identical responses for* ***poliovirus type 3****. The boxes represent individual potency estimates and are labelled with the laboratory and method code and assay number. Statistically invalid assays (5 % significance level) are marked with a *.*

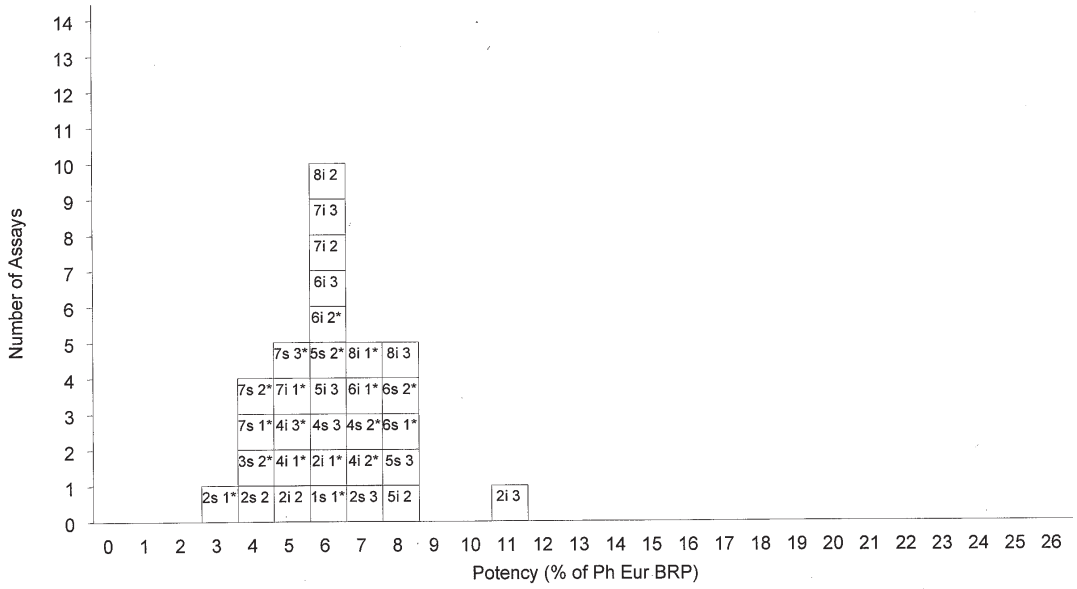


Figure 13 — *Overall dose response of Ph. Eur. BRP by laboratory for* ***poliovirus type 1***

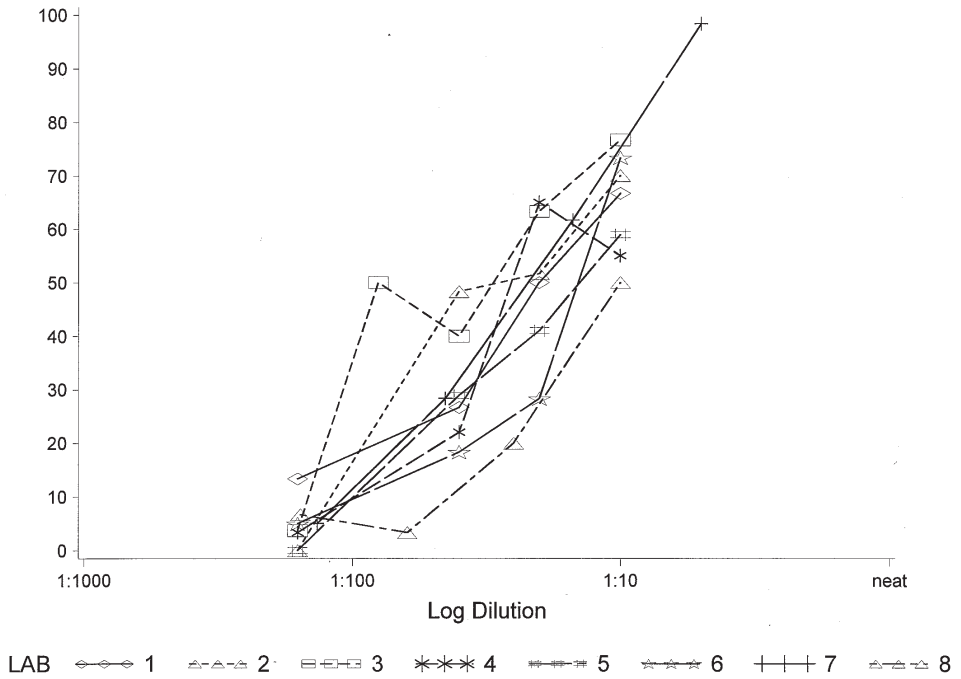


Figure 14 — *Overall dose response of Ph. Eur. BRP by laboratory for **poliovirus type 2***

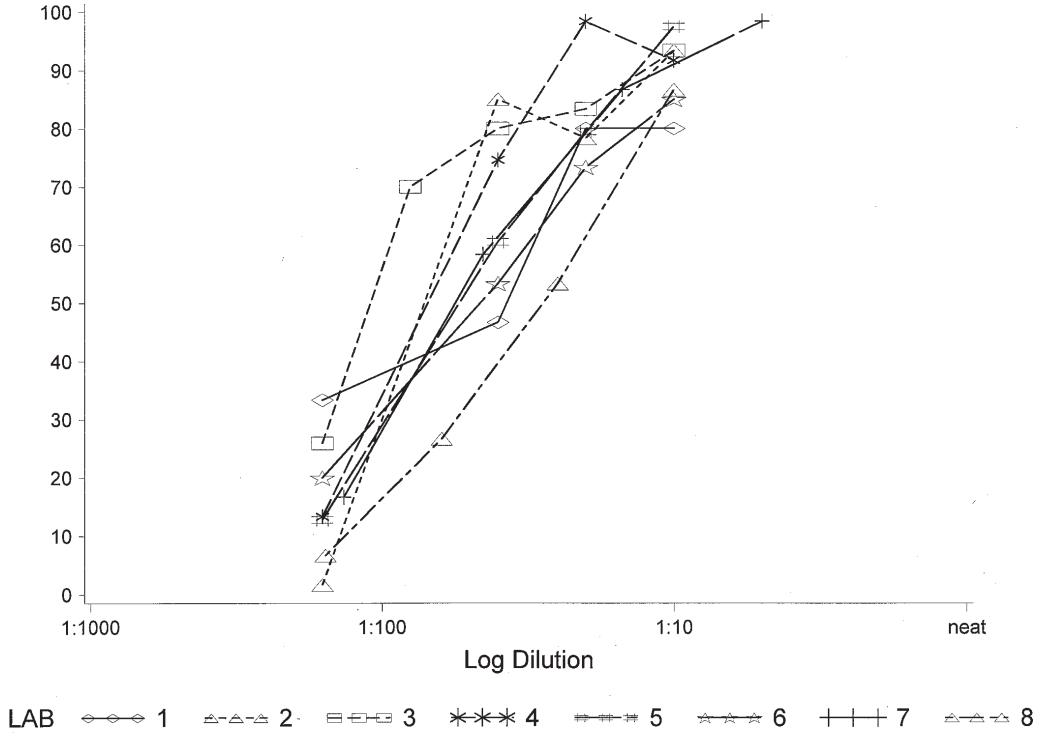
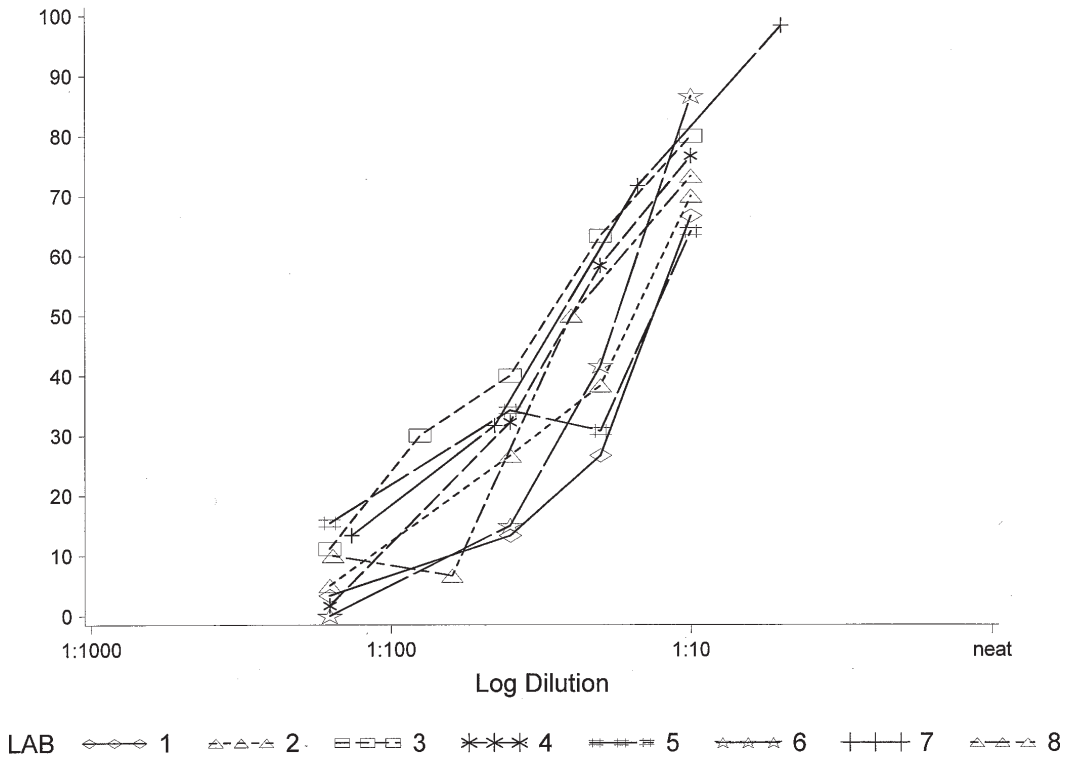


Figure 15 — *Overall dose response of Ph. Eur. BRP by laboratory for **poliovirus type 3***



5.9. COMPARISON OF PROBIT AND PARALLEL LINE APPROACHES

The biggest difference between the two approaches is the number of assays considered statistically valid. The numbers of valid assays are summarised in table 7. If the usual 5 % significance level is used, 83 % of assays can be analysed using the probit method and produce statistically valid assay data. In contrast only 8 % or 12 % of assays can be analysed by the parallel line method and produce statistically valid assays. Even if the requirements of normality and homogeneity of variance, which form part of the underlying statistical model and are specified in the Ph. Eur. (General Chapter 5.3.), are ignored, no more than 55 % of assays could be analysed using the parallel-line method. The nature of the response data does not appear suited to the usual parallel line analysis.

The estimates of the potency of 90/716 relative to the Ph. Eur. BRP have been calculated, ignoring the validity criteria of parallelism and linearity, and the distribution assumptions of normality and homogeneity of variance. Assays without a significant dose response could not provide a potency estimate. The estimates of potency vary considerably between assays whichever approach to analysis is employed. The overall geometric means (Tables 8, 9, 10) are similar. The probit method gives slightly higher overall means for polio types 1 and 2, but inspection of Figures 4-9 suggest that this is due to individual assays giving high estimates, rather than a consistent difference for the majority of the assays.

The % GCVs in Tables 8 and 9 indicate that there is slightly better agreement between estimates obtained from the parallel line analysis than from probit analysis. The improved agreement between assays using the parallel line method, although not great, indicates that the reduction of the data to a single positive/negative response does not make full use of the information present. However, both methods give levels of agreement between assays that are not unreasonable for *in vivo* assays with group sizes of 10 animals. For example, the requirement for the precision of the mouse potency test for Hepatitis A Vaccine (HAV) is a confidence interval of 33-300 % of the estimate (Ph. Eur. 1998:1107). Furthermore a recently completed WHO/EDQM collaborative study of HAV, that included immunogenicity assays, found % GCVs for inter-laboratory variability with a range of samples in the range of 42-119 %, comparable to that observed here (Tables 8, 9, 10).

The WHO collaborative study of immunogenicity assays of IPV (Wood and Heath 1995) found variations of greater than 100-fold in the ED_{50} results of the chick/guinea pig tests. In the same study, four laboratories used the rat test, and data were analysed using the probit method. Ranges in estimates across labs and assays were between 2-fold and 12-fold for the different preparations included in the study. The ranges observed in this current study are similar (5-fold to 11-fold) but based on more laboratories, some with less experience of the rat test. In this context, the performance of the rat test in the current study can be considered good.

5.10. VALIDITY CRITERIA

In addition to statistical validity criteria as used above, further validity criteria for the assay were examined. These were that the ED_{50} of the reference vaccine should be within the dose range inoculated into rats and that the confidence limits of the estimated relative potency should be within defined limits.

The following percentages of assays would satisfy a validity criterion of the ED_{50} of the reference vaccine within the dose range inoculated into rats: Type 1: 84 %; Type 2: 100 %; Type 3: 92 %.

However, the cut-offs used for defining a response were calculated as the mid-point for the observed data for the three assays for each laboratory. It is therefore to be expected that the ED_{50} will be within the range. The only cases where this did not occur were where there was a big difference in response between assays, or where there was a problem with getting a dose-response in an individual assay. This is therefore an important validity criterion, but the actual percentages of assays that would satisfy this criterion would need to be verified in future assays.

Using the probit approach, the following percentages of assays gave potency estimates within the suggested limits. For all laboratories:

	30-300 %	**33-300 %**	**50-200 %**
Type 1	76 %	68 %	3 %
Type 2	89 %	89 %	26 %
Type 3	70 %	65 %	22 %
Overall	78 %	74 %	17 %

Clearly the 50-200 % limit is too strict. Even with the 30-300 % limit attainable, overall figures from this study still give too many assays with limits too wide to be practical. However as techniques improve within individual laboratories the proportion of assays within the limits would be expected to improve. To evaluate this the above analysis was repeated but restricted to laboratories with previous experience of the assay (Laboratories 6, 7 and 8).

The proportion of assays within given confidence limits was as follows:

	30-300 %	**33-300 %**	**50-200 %**
Type 1	93 %	87 %	0 %
Type 2	87 %	87 %	13 %
Type 3	93 %	93 %	40 %
Overall	91 %	86 %	18 %

For type 1 there were several assays with lower limits between 30 % and 33 %, which explains the difference in figures. These data suggest that confidence limits of 30-300 % for this assay are attainable on a regular basis.

5.11. SINGLE DOSE ASSAYS

This study was not primarily designed to evaluate the rat test as a single dose assay. Nevertheless due to interest in this approach (Bevilacqua et al. 1996) the data were evaluated to investigate the potential of the rat test as a single dose assay. An estimate of potency requires that sufficient doses are included in the assay design to allow a reliable estimate of the slope of the dose-response curve to be obtained. Assay designs with a single dose of reference and test vaccines do not allow the estimation of potency. Designs using a range of doses for the reference, and a single dose for the test vaccine, have been suggested. For such a design, the choice of dose of test vaccine used is critical, particularly for an assay being analysed by the probit method. The dose of vaccine would need to produce a percentage of responders close to 50 % (i.e. the dose of test vaccine should be close to the ED_{50} for the test vaccine). If the percentage of responders was 0 % or 100 % it would be impossible to obtain estimates, and with response rates close to these extremes, any potency estimate would be very unreliable (due to low precision). The reduction in the numbers of animals used for such assays must be balanced against the associated drop in precision.

Although assays including a single dose of test and reference vaccine can not give estimates of potency, they can be used as screening assays, to demonstrate that a test vaccine is above

an acceptable level of potency. Such a procedure has been described for diphtheria and tetanus vaccines (Manual of laboratory methods). A dose of a reference vaccine is selected as representing an "acceptable" level of potency. If the single dose of test vaccine produces significantly higher responses (or not significantly lower responses) than the reference vaccine, then it is considered "acceptable". Pre-requisites of this test are agreement on the appropriate dose of reference, and a demonstration by the laboratory that they can obtain appropriate (linear and parallel) dose-response curves for test and reference vaccines. Assays of this type have the potential for producing considerable savings in the numbers of animals used, and may well be adequate for batch release testing where a history of consistent production exists.

The approach of single dose testing could be applied to the rat assay for IPV. Problems with the statistical analysis of responses resulting from the non-normal distribution can also be avoided, by applying established non-parametric tests such as the Wilcoxon rank test to the comparison of responses from test and reference vaccines. This has the potential to make better use of the data than the reduction to a binary "response/non-response" value necessary for the probit analysis proposed earlier.

Some of the data from this collaborative study have been analysed, to investigate the potential of this approach. The 1:10 dilution of the Ph. Eur. BRP has been taken as the "acceptable" dose of reference vaccine, and single doses of 90/716 have been tested against it. The neat, 1:2, and 1:4 dilutions of 90/716 were individually compared to the 1:10 dilution of the Ph. Eur. BRP. Analysis was by a Wilcoxon rank-sum test. The test vaccine was considered a "pass" if the responses were not significantly lower than those for the reference.

Assays from all laboratories except 7 and 8 were analysed. These latter laboratories did not use a neat, 1:2, 1:4 dilution scheme for 90/716 and so could not be directly compared to other laboratories. The number of assays producing a "fail" for the test vaccine were:

	Type 1	**Type 2**	**Type 3**	**Overall**
90/716 - neat	10.7 %	3.6 %	10.7 %	8.3 %
90/716 - 1:2	60.7 %	32.1 %	60.7 %	51.2 %
90/716 - 1:4	67.9 %	78.6 %	82.1 %	76.2 %

These preliminary figures indicate that for increasing dilutions of 90/716 (representing sub-potent vaccine), increasing percentages of assays "fail" the test vaccine. The full 4-dose per vaccine assay only produced potency estimates with confidence intervals better than 33-300 % for 77 % of assays. It should also be borne in mind that two of the three experienced laboratories (labs 7 and 8) were not included in this analysis. Performance may improve as laboratories obtain more experience with the rat test.

The approach of a single-dose-screening assay has potential, but will require further development. Rather than the "not significantly lower" test, a "significantly higher" requirement may be preferable. This would require agreement on a dilution of the Ph. Eur. BRP giving minimum acceptable potency. Power calculations would also be required to select the number of animals to be tested at the single dose to give appropriate levels of probability of accepting or rejecting good or sub-potent vaccines.

6. DISCUSSION

The Ph. Eur. test in rats (Ph. Eur. 1999:0214) estimates potency by comparison of the regression curves for the vaccine and the reference vaccine. The parallel line method was not the most suitable regression model for data from this study. Statistically valid assay analysis,

as defined by the Ph. Eur. (Chapter 5.3.), could only be performed on around 10% of all assays with the parallel line method. Even ignoring the validity criteria for normality and homogeneity of variance, only 50-70 % (depending on the significance level used) of assays were statistically valid. This situation would not be acceptable for a general recommendation for a batch release assay. However if a laboratory demonstrates that they can consistently obtain valid analyses with the parallel line method for the products they routinely test, then they may wish to continue with this method, as long as it was approved by the competent authorities.

The probit method, on the other hand, was found to be an effective method of determining relative potency estimates. Although it does not make full use of the information present in the data, it allowed statistically valid assay analysis to be performed for 80-90 % of all assays performed. A previous study in a single laboratory also found that the probit method gave consistently more valid analyses than the parallel line method (Bevilacqua et al. 1996). It is therefore concluded that probit is the preferred method of statistical analysis at the present time for the rat bioassay. The cut-off, to define a positive or negative response, should be determined for each individual laboratory to allow analysis by the probit method. The cut-off calculated will be dependent on the antigen and dose levels used. The Ph. Eur. BRP should be used as a standard antigen for this purpose. The doses of Ph. Eur. BRP to be used in the assays to determine cut-off can be defined on the basis of this study. A consistent cut-off within a laboratory, for each polio type, is preferable to determining cut-offs on an individual assay basis. A minimum series of three tests should be used to determine cut-off levels and individual laboratories should then update their cut off values as further experience is accumulated.

The use of the Sabin strains of poliovirus in the neutralising antibody test gave similar results to the use of wild type strains. To facilitate the eventual global certification of the eradication of wild type polioviruses the use of the Sabin strains should be the method of choice in the future.

The calibration of neutralisation tests in mIU against the IS was not correlated with the actual titres in rat sera observed by different laboratories for the same dose of the Ph. Eur. BRP. An explanation for this observation may be that the IS is a pool of human serum collected from UK blood donors in the 1960's (Wood and Heath 1992). The majority of antibodies in the pool are likely to have been induced by natural infection with live poliovirus. The rat sera in this study were raised by immunisation with IPV. There are known to be differences in antigenic structure between live and formalin inactivated polioviruses (Ferguson et al. 1993). Furthermore it is known that the immunodominant antigenic site 1 of poliovirus type 3 is destroyed by trypsin treatment (Minor et al. 1986) and consequently is modified by intestinal trypsin (Roivainen and Hovi 1987). Therefore there are differences in the humoral immune response to IPV and live poliovirus infection (Roivainen and Hovi 1987). On the other hand clinical study data have shown that the immunogenicities of inactivated polioviruses in rats and man are similar (van Steenis et al. 1981) although additional data comparing human and rat responses would be helpful. One consequence of this observation is that to reduce the inter-laboratory variation in potency estimates of IPV, a standard rat serum raised against IPV may be necessary.

Further work is also recommended to evaluate the impact of the observed between assay variation in the rat bioassay. For this purpose a study that included a low potency antigen would be valuable. Also the present study tested plain IPV formulations. Nowadays IPV is often formulated in combination with other antigens, especially DTP. Results from one laboratory suggest that parallel dose responses are obtained in the rat bioassay for IPV in combination vaccines in comparison to an in-house reference material (Bevilacqua et al. 1996). An additional collaborative study is required to determine if this finding is applicable to the Ph. Eur. BRP.

7. CONCLUSION

The EDQM initiated a project to evaluate a rat *in vivo* bioassay for IPV. The aims were to establish the transferability of the method to new laboratories, to establish specifications for the neutralising antibody test used to assay sera from the test and, in particular, to establish the most appropriate method of statistical analysis.

Eight laboratories, including both manufacturers and national control laboratories, participated. A candidate standard rat bioassay method and, where available, an established in-house method were compared in 3 independent assays. Two antigens, the Ph. Eur. BRP and a plain IPV were included. Participants were requested to assay all sera for neutralising antibodies to all three poliovirus serotypes using an established in-house neutralising antibody test. Additionally a candidate standard neutralising antibody test that used a 100 $CCID_{50}$ challenge of each of the Sabin polioviruses and Hep2C indicator cells was specified and participants were requested to retest all sera with this method.

The results showed a variation in sensitivity between laboratories for neutralising antibody titres of around 5-fold (maximum/minimum) for types 1 and 2, and around 10-fold for type 3 when results were expressed as mIU. Inter-laboratory agreement was better with the standard neutralising antibody method than with the in-house methods. A correlation was not observed, however, between the calibration in mIU and the overall level of observed titres in rat sera for a fixed dose of a particular vaccine across laboratories.

Potency estimates in the rat bioassay are made by comparison of the regression curves for the vaccine and the reference vaccine. The parallel line method was not the most suitable regression model for data from this study since statistically valid assay analysis could only be performed on around 10% of all assays. The parallel line method would not be acceptable in a general recommendation for batch release purposes. The probit method, on the other hand, was found to be an effective method of determining relative potency estimates for 80-90 % of all assays performed.

The inter-laboratory variation of relative potency estimates when the probit method was used was greatest for type 2, with an 11-fold range, compared to 5- or 8-fold for types 1 and 3. If two assays with particularly high potency results were excluded, the range for type 2 was closer to 8-fold. Between laboratory variation was similar for poliovirus type 1 whether the standard or in-house neutralising antibody assays were used; was greater for the in-house method for type 2; and greater for the standard method for type 3.

In conclusion the rat bioassay was successfully transferred to a number of laboratories, a standard neutralising antibody test that uses the Sabin strains of poliovirus was shown to be suitable for the assay of rat sera, probit analysis was shown to be the most appropriate method of regression analysis for general recommendation for the assay and validity criteria for the assay were defined.

8. REFERENCES

Bevilacqua JM, Young L, Chiu SW et al. Rat immunogenicity assay of inactivated poliovirus. *Developments in Biological Standardisation*, **86**, 121 (1996).

Ferguson M, Wood DJ and Minor PD. Antigenic structure of poliovirus in inactivated vaccines. *Journal of General Virology*, **74**, 685 (1993).

Manual of Laboratory Methods - For testing of vaccines used in the WHO Expanded Programme on Immunization, **WHO/VSQ/97.04**, 162.

Minor PD. Summary report of a meeting on the estimation of the potency of inactivated poliovaccine. *Biologicals*, **18**, 243 (1990).

Minor PD, Ferguson M, Evans DMA et al. Antigenic structure of polioviruses of serotypes 1, 2 and 3. *Journal of General Virology*, **67**, 1283 (1986).

Roivainen M and Hovi T. Intestinal trypsin can significantly modify antigenic properties of polioviruses: implications for the use of inactivated poliovirus vaccines. *Journal of Virology*, **61**, 3749 (1987).

Tummers B. Collaborative study for the establishment of a BRP for inactivated poliomyelitis vaccine. *Pharmeuropa-BIO,* **96-2**, 81 (1996).

van Steenis G, van Wezel AL and Sekhuis VM. Potency testing of killed polio vaccine in rats. *Developments in Biological Standardisation*, **47**, 119 (1981).

Wood DJ and Heath AB. A WHO collaborative study of immunogenicity assays of inactivated poliovirus vaccines. *Biologicals*, **23**, 301 (1995).

Wood DJ and Heath AB. The second International Standard for anti-poliovirus sera types 1, 2 and 3. *Biologicals*, **20**, 203 (1992).

Wood DJ, Heath AB, Kersten GFA et al. A new WHO international reference reagent for use in potency assays of inactivated poliomyelitis vaccine. *Biologicals*, **25**, 59 (1997).

Wood DJ, Heath AB and Sawyer LA. A WHO collaborative study on assays of the antigenic content of inactivated poliovirus vaccines. *Biologicals*, **23**, 83 (1995).

World Health Organisation. Progress towards global poliomyelitis eradication, as of May 1999. *Weekly Epidemiology Records*, **74**, 165 (1999a).

World Health Organisation. *Global Action Plan for laboratory containment of wild polioviruses* (1999b, in press).

9. ACKNOWLEDGEMENTS

The organisers express their sincere thanks to the project leader, Dr. David Wood, and to all participants for their contributions to this study. The study was run by the EDQM with the internal number BSP021.

10. PARTICIPANTS

L Peyron, Pasteur Mérieux Serums et Vaccins,
1541 Avenue Marcel Mérieux, 69280 MARCY L'ETOILE, France.

E Ostergaard, Statens seruminstitut, Virus Vaccine Department,
Artillerivej 5, 2300 COPENHAGEN, Denmark.

M Duchêne, SB Bio, 89 rue de l'Institut, 1330 RIXENSART, Belgium.

F Fuchs, AFSSAPS, Unité de contrôle des médicaments immunologiques,
321, Avenue Jean Jaurès, 69007 LYON, France.

HJM van der Donk, RIVM,
Postbus 1, Antonie van Leeuwenhoeklaan 9, 3720 BA BILTHOVEN, The Netherlands.

I Clement, Pasteur Mérieux Connaught,
1755 Steeles Avenue West, North York, Ontario M2R 3T4, Canada.

DJ Wood, NIBSC, South Mimms, Potters Bar, Herts EN6 3QG, UK.

V Rhen, SBL Vaccin AB, 10521 STOCKHOLM, Sweden.