Study report: Proficiency testing study on a WHO/TAL-trained methodology

High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) for the determination of the total and free saccharide content in *Haemophilus influenzae* type b (Hib) liquid combined vaccines.

1. Introduction

Current *Haemophilus influenzae* type b (Hib) vaccines are made by conjugating the capsular polysaccharide (poly ribosyl-ribitol-phosphate; PRP) to a carrier protein. This conjugated form will induce a T-dependent B-cell response in infants and will result in immune memory [1, 2].

When the native form of Hib polysaccharide is used, the polysaccharide is usually linked covalently to tetanus toxoid (TT-PRP). Some vaccines use an oligosaccharide form instead; for these vaccines, the oligosaccharide is linked to a non-toxic variant of the diphtheria toxin, the cross-reacting material CRM_{197} (CRM-PRP) [3]. The Hib glycoconjugate component can be combined with different vaccine antigens such as diphtheria (D), tetanus (T), whole-cell pertussis (wP) or acellular pertussis (aP), hepatitis B (HepB) and inactivated polio vaccine (IPV). Combination with any of these antigens, as well as with adjuvants, preservatives and other excipients, can interfere with analysis of the critical parameters indicative of Hib vaccine quality and efficacy, specifically, molecular size distribution and the total and free (unconjugated) saccharide content [1].

Total and free saccharide content are two parameters that must be checked routinely by manufacturers and by national control laboratories (NCLs) as required by WHO in the guideline for production and control of Hib vaccines [4] and by the *European Pharmacopoeia* [5].

Pentavalent DTwP-HepB-Hib liquid vaccine is used globally and has been designated a high-priority vaccine by the WHO Prequalification Team (WHO/PQT). Quality control testing of these vaccines is performed pre-and post-prequalification [6] by WHO-contracted laboratories all of which are national, official medicines control laboratories. Furthermore, post-licensing, the vaccines are undergoing official batch release testing by the authority with regulatory oversight responsibility before release to markets.

Study Co-ordinator: Dr Ute Rosskopf – WHO / Department Regulation & Prequalification, CH
Scientific Advisor: Dr Christina von Hunolstein – Istituto Superiore di Sanità (ISS), I
Statistician: Dr Andrea Gaggioli – Istituto Superiore di Sanità (ISS), I

The proficiency testing studies (PTSs) are performed to offer NCLs and manufacturers an opportunity to assess and measure their laboratory performance, based on an inter-laboratory comparison. In general, the scope of the PTS is to cover the overall performance of a laboratory, and includes the whole process starting with reception and storage of the samples, experimental work, calculations, interpretation and transcription of the data, and reporting of the results and conclusions in the recording sheets provided by study coordinators.

Once the initial test report has been submitted by a laboratory, that report cannot be modified even if the laboratory discovers a failure of one of the process steps. However, comments from the laboratories can be added to the draft final report, but tables, figures and conclusions will not be modified, unless the data submitted by the laboratories have been misinterpreted at ISS and/or clear mistakes have been made at ISS.

2. Aim of the study

This PTS was organised and coordinated by the Technical Assistance & Laboratory Services (TAL) - Vaccines, Group (subsequently renamed the Laboratory Networks and Services (LNS)) within the Regulatory System Strengthening (RSS) Team, the Regulation of Medicines and Other Health Technologies (RHT) Unit, the Essential Medicines and Health Products (EMP) Department in the Health Systems and Innovation (HIS) Cluster of the World Health Organization (WHO), and was done in collaboration with the Unit of Biological and Biotechnological Products, Bacterial Vaccine Section of the Istituto Superiore di Sanità (ISS). The aim of this PTS was to assess the proficiency of the participating laboratories to quantify the total and free (unconjugated) polysaccharide (PRP) content of the *Haemophilus influenzae* type b (Hib) component in different liquid pentavalent vaccine (DTwP-HepB-Hib) presentations when tested according to a method using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). The methodology was designed to determine the PRP content in liquid vaccine combinations containing a whole-cell pertussis component [7].

3. Participation

Establishment of a quality management system compliant with quality standard ISO/IEC17025: "General requirements for the competence of testing and calibration laboratories" is a requirement for laboratories in charge of vaccines' quality control "to demonstrate that they operate competently and are able to generate valid results" [8]. Participation in proficiency testing schemes is one essential requirement for laboratories adhering to this standard.

WHO/TAL invited those NCLs and manufacturers' quality control laboratories that had participated either in the hands-on training courses [9] and/or in the previously performed WHO collaborative study [10]. A total number of 21 laboratories registered for participation in the study; of these, twelve are NCLs and nine are quality control laboratories from vaccine manufacturers. A list of participants in alphabetical order by country is given in section 9. Herein, they are referred to by an arbitrarily allocated code number (1-21), not related to the order of listing.

4. PTS Design

4.1 Composition of the panel of test samples

Each laboratory received a panel of three pentavalent vaccines (containing D, T, wP, HepB and Hib (DTwP-HepB-Hib) active components) (Table 1). Two of the test vaccines were WHO prequalified DTwP-HepB-Hib liquid formulated vaccines, while the third test vaccine was especially prepared for the study and contained a sub-potent PRP content and a high free, unconjugated PRP content. All the vaccines were fully liquid formulations that included aluminium phosphate as adjuvant. Their content of total and free polysaccharide was not revealed to the study participants. The vaccines samples were kindly donated by different manufacturers. The samples were shipped in appropriate packaging to protect the samples during transit and under temperature-controlled conditions. Participants were asked to check the contents of the package immediately and return the completed sample arrival report to WHO indicating any problems regarding the condition of the test materials or accompanying documents. Upon receipt of the samples, participants were asked to store the vaccines at 5 °C ± 3 °C until use.

4.2 Study design and reporting of results

The shipment of the samples from WHO Headquarters in Geneva to the participating laboratories was initiated in January 2019. Following the receipt of the samples, the participants were asked to submit their test data within two months after receipt of the samples as indicated in the study protocol. Due to delays in shipments (e.g., custom clearances) and test performance, the latest test report was received in December 2019.

Participants were requested to test the vaccine panel according to the analytical protocol provided; this protocol was the same as that distributed during the hands-on training courses and for the previously performed collaborative study [10].

Participants were requested to quantify the total and free PRP content of each vaccine sample in three independent runs (i.e., 3 separate testing days) by using a vaccine pool freshly prepared on each test day. Calculation of both polysaccharide contents was based on the current in-use calibration curve (i.e., either the WHO 2nd International Standard for *Haemophilus influenzae* polysaccharide PRP or the ribitol reference standard). Determination of the total PRP, according to the analytical protocol, does not require any pre-treatment of the vaccine sample. To test for free PRP, however, the vaccine sample must undergo several steps of preparation. The steps include pre-treatment with 5 mM phosphate buffer, pH 6.8 [10], centrifugation to eliminate the adjuvant, and application of the supernatant to a SPE C4 wide pore cartridge. The cartridge permeate was collected to recover the free PRP. Hydrolysis of the permeate (to test free PRP) and the vaccine sample (to test total PRP), was performed by adding 50 µL 6N HCl to all samples and then incubating them for 2 hours at 100 °C. In addition to the test sample, samples include the positive control, the system suitability test (SST) sample, and 1 mL of each dilution point of the calibration curve. The samples then were cooled for 10 min at 5 ± 3 °C and 400 µL of 1 M NaOH was added. Each sample was then appropriately diluted, filtered and analysed by HPAEC-PAD.

The analytical protocol defined the chromatographic conditions to be followed. The laboratories were asked to complete a form reporting the characteristics of their HPAEC equipment, details of the ribitol used as a positive control (i.e., percent purity, moisture content, diluents, and storage time and temperature), the system suitability criteria in place, any deviations from the analytical protocol, any difficulties encountered, and any observations regarding the study protocol. An electronic data reporting sheet was provided to record the experimental data, specifically, the content, in µg per single human dose (µg/shd) of total and free saccharide (indicated in the report as total and free PRP). It was requested to report test results using two decimal places, which was applicable as well to the reporting of the free saccharide content as a percentage of the total PRP. Data from laboratories were received by WHO/TAL and ISS between February 2019 and December 2019. The original deadline date was extended to allow the acceptance of data that could not be sent on time because of the heavy workload or late sample receipt.

4.3 Statistical methods

See Appendix 1 for more details on the statistical evaluation of the PTS data.

5. Results and Statistical Analysis

5.1 Study test results

A total of twenty-one laboratories reported their assay results, but the data from only nineteen (19) laboratories were analysed in this report. One laboratory was excluded because it sent the results after the extended deadline (December 2019) and another laboratory was excluded because it provided results from only one test run.

Vials for PTS-1 sample were shipped to 18 laboratories only, due to limited supply of that sample. All 19 laboratories, included in the data analyses reported below, carried out three assays (test runs) for each vaccine in accordance with the study protocol.

Tables 2 A-C show a complete listing of values reported by the laboratories. The tables present the results per individual run, the geometric mean (GM) of the three runs, and the rounded GM for each test sample. The results for free saccharide content are also reported as a percentage of the total saccharide content. Grey shaded cells indicate results that were declared non-valid according to the outcome for the system suitability test. These results were retained in the analyses since they were in line with the results from the other laboratories.

In figures 1-4, results are reported using two different ways of plotting the total and free PRP content for the three PTS samples. Figures 1-3 plot, for each sample and each test, the GM and the individual test results for each of the 19 laboratories. Figures 4A-4F report the GMs for the 19 laboratories in a dot plot format (similar to the histogram plot). The central horizontal line (figures 1-3) and the vertical line in the dot plots (figures 4 A-F) represent the consensus values (see section 5.3).

5.2 Precision of the method

Precision of the method was calculated for each PTS sample for both parameters (total and free PRP content) as inter-laboratory precision, intra-laboratory precision and reproducibility. Data for the three PTS samples are reported in tables 3A, 3B and 3C.

5.3 Consensus values

Consensus values were obtained based on an analysis of the raw data and an evaluation of the data distributions.

The hypothesis of normality was assessed by means of the Shapiro-Wilk Test. This is considered the most powerful normality test when there is a small sample size, as in this case of 19 laboratories. If the p-value of the Shapiro-Wilk test is less than the classical threshold of 0.05, the results indicate a significant deviation from the assumption of normality whereas Shapiro-Wilk p-values greater than 0.05 indicate no significant deviation from normality. However, for the purposes of this report, a more conservative approach was used. Specifically, a distribution was considered to be normal if normality was consistent with visual inspection of exploratory graphs of the data and if the Shapiro-Wilk p-values were greater than 0.5. When these conditions were met, the classical calculation of consensus value and standard deviation was used instead of a robust calculation. This conservative approach was adopted because the classical estimator of consensus value is very sensitive even to small deviations from normality. A complete list of the robust estimators is provided in Appendix 2 (table A-1). When the Shapiro-Wilk value was < 0.5, the consensus value was based on the robust estimator defined in "ISO 13528:2015 Statistical methods for use in proficiency testing by inter-laboratory comparison" [11].

The uncertainty of measurement [12] associated with the consensus values was determined by pooling the uncertainty of the mean of the n_{Lab} laboratory means: i.e., the standard error: (S_{Lab}/sqrt(n_{Lab})), and the uncertainty associated with the homogeneity of the vials (i.e., S_{Run}, the within-Lab standard deviation). Other potential uncertainty sources to be included in the uncertainty budget are considered negligible (i.e., less than 1/5 of the major contributor). The consensus values for all samples (derived from the geometric means of the laboratories listed in table 2) are reported in figures 3, 4 and 5, where the expanded uncertainty (U) is shown for the consensus value of all six parameters, i.e., total and free saccharide content for the three PTS samples.

In particular, the details on the approach used for the different samples are the following:

- **PTS -1 sample, total PRP content:**

The data distribution can be considered to be normal (Shapiro-Wilk: p-value=0.927); no anomalous values were detected. Therefore, the classical calculation (i.e., arithmetic mean and SD), is used for the consensus value and the standard deviation (which equalled 1.41 µg/shd).

Consensus value: 10.5 ± 1.7 µg/shd (k=2); relative U = 16.2% (k=2)

- PTS -1 sample, free PRP content:
Although the data distribution does not significantly deviate from normality (Shapiro-Wilk: p-value=0.375), a slight positive asymmetry is observed. Therefore, the robust estimate obtained using Algorithm A [11] was considered to be the consensus value and the robust estimate of the standard deviation (which equalled 0.529 µg/shd) was considered as the consensus standard deviation.
***Consensus value**: 1.176 ± 0.337 µg/shd (k=2); relative U = 28.7% (k=2)*

- PTS -2 sample, total PRP content:
Although the data distribution does not deviate from normality (Shapiro-Wilk: p-value=0.108), a slight left asymmetry and an anomalous value were observed. Therefore, the robust estimate obtained by using Algorithm A [11] was considered to be the consensus value and the robust estimate of the standard deviation (which equalled 0.748 µg/shd) was considered to be the consensus standard deviation.
***Consensus value**: 5.835 ± 1.375 µg/shd (k=2); relative U = 23.6% (k=2)*

- PTS -2 sample, free PRP content:
The data distribution significantly deviates from normality (Shapiro-Wilk: p-value=0.020). Visual inspection, in this case, reveals that the distribution can be considered bimodal; this means that a true consensus was not really obtained. In fact, two different groups of laboratories were observed, as illustrated in the figure of Appendix 2 (figure A-1). The minor mode makes a considerable contribution to the area of the kernel; therefore, two discrepant populations are represented in the participants' results. Without additional independent information (e.g., further details of the participants' analytical methods and product label claim), it is not possible to determine which of these modes is the correct one.

In this situation, two alternatives can be considered:
1) Do not attempt to determine a consensus value, and do not report for this parameter any individual laboratory performance scores. In this scenario, a repetition of the PTS would be recommended.
2) Determine the consensus value based on the principle that the higher mode is the most probable true value.

The PTS provider decided to determine a consensus value obtained by the result of the Huber robust estimator (robust standard deviation equal to 1.22 µg/shd) because it is a more suitable estimator when the data distribution is not unimodal (i.e., since it gives more weight to the higher mode).
***Consensus value**: 3.171 ± 0.6 µg/shd (k=2); relative U = 18.9% (k=2)*

- PTS -3 sample, total PRP content:

The data distribution can be assumed to be normal (Shapiro-Wilk: p-value=0.976); no anomalous values are detected. Therefore, the classical calculation (i.e., arithmetic mean and SD), was used for the consensus value and the standard deviation (which equalled 1.146 µg/shd).

Consensus value: 8.19 ± 1.28 µg/shd (k=2); relative U = 15.6% (k=2)

- PTS -3 sample, free PRP content:

The data distribution can be assumed to be normal (Shapiro-Wilk: p-value=0.510); no anomalous values were detected. Therefore, the classical calculation (i.e.: arithmetic mean and SD), was used for the consensus value and the consensus standard deviation (which equalled 0.607 µg/shd).

Consensus value: 1.22 ± 0.37 µg/shd (k=2); relative U = 30.3% (k=2)

5.4 Z-score

The performance of the participating laboratories was evaluated using the performance indicator called Z-score (see Appendix 2 for additional details). Z-score should be within the range > -2 to <+2 to declare the performance of a single laboratory as satisfactory. Table 4 reflects the Z-scores that the different laboratories achieved for each PTS sample and parameter.

The limitations induced by the high uncertainty associated with the consensus values for some of the tested parameters are not negligible, and therefore, the information content of the Z-scores will be reduced correspondingly. This means that truly questionable and unsatisfactory Z-scores could be not detected in this study. This certainly is the case for the measurements of the free PRP content for the 3 test samples; each of these showed consensus values with an associated expanded relative uncertainties (U) higher than 25%. These high U values were the result of a low reproducibility (for PTS-1 and PTS-3) or of the bimodal distribution (PTS-2). When a high uncertainty is observed, ISO 13528 ([11]) and Eurachem [13] suggest the use of Modified Z-score (i.e., z' score = $\frac{x-C}{\sqrt{u^2+\sigma^2}}$ with C: consensus value; u: standard uncertainty; and σ: standard deviation for proficiency assessment), while Thomson *et al.* [14] offers a slightly different recommendation. In any case, Modified Z-scores tend towards values similar to standard Z-scores (data not shown).

In figures 6A-6F (for each sample and each parameter) the Z-score per laboratory in a size order within the Z- score range are reported i.e., from negative to positive value. When the bars of Z-scores are systematically on either the positive side or negative side, a systematic bias is indicated. If the bars extend more than 2 on both sides, a random bias is indicated. However, due to the variability of the results, evaluation based only on scoring is considered to be of limited value. Therefore, it did not seem appropriate to draw the conclusions regarding the proficiency of the participants' based only on the classical criteria as presented in Appendix 1.

Similarly, it is recommended to use the Z-scores that are presented here with caution for purposes outside of this report.

5.5 Overall assessment of performance

Six different parameters have been analysed in this study: two test results (i.e., total and free PRP content) for each of the three samples. To give an overview on the performance of each participating laboratory, two different tools for data presentation have been used. The first tool is based on the Youden Plot [15]; the second is based on an attempt to pool the six different Z-scores into unique scores, specifically, Rolling Performance Indicators (RPIs) [13, 14].

5.5.1 Youden Plots

Youden plots are a graphical way to represent the results obtained by the laboratories and, at the same time, to provide a critical assessment of the performances [15]. In a single graph both test results obtained for a given test sample, i.e. the total and free PRP contents, are reported. Therefore, this figure presents a combined assessment of the two parameters measured on the same sample.

The Youden plots reported in figures 7-9 are based on the raw data (in µg/shd) for each of the 3 tested samples; the scales of the axes for these plots were adapted for parameters that are not similar (i.e., total and free PRP).

Each point in the plot represents the results of one laboratory; the total PRP result is plotted on the horizontal axis and the free PRP result is plotted on the vertical axis. The centre of the circles is called the Manhattan median; this centre point represents the median values of the two test results (see Appendix 3 for additional details on construction of the Youden plots). Points outside the outer circle indicate large total error. Points that lie near the 45-degree reference line, but far from the centre of the circles, indicate large systematic error. Points that lie far from the 45-degree line indicate large random error. Thus, the plot also provides a graphical indication of the likely types of error.

Laboratories appearing in the 1st quadrant (consistently high results) or in the 3rd quadrant (consistently low results) exhibit systematic errors or bias, i.e., the further the point is from the centre of the circles, the greater is the error. If random errors are small, the points would be close to the 45° line. The lengths of perpendiculars drawn from the points to the 45° line are directly related to the random errors. Ideally, all observations should be placed on the diagonal line and inside the inner circle. Laboratories within the inner circle have a normal random variation, while laboratories outside the outer circle but near the diagonal line exhibit systematic bias.

5.5.2. Rolling Performance Indicators (RPIs)

RPIs are usually computed to assess the performance of participants over a successive series of PTS studies, i.e., consecutive rounds of testing (13, 14). They can be used to assess bias and precision (systematic and random error) by pooling different results in terms of Z-scores; in case of this study for the six different parameters. There are various RPIs which can be calculated based on the obtained Z-scores: Rescaled Sum of the Z-scores, Sum of the Squared Z-scores, Rescaled Sum of the Squared Z-scores, and Absolute Rescaled Sum of the Squared Z-scores (for details, see Appendix 4).

All the above-mentioned indicators give a unique score for each of the participants. There is a temptation to determine a single indicator of merit, that can summarize the overall performance of each laboratory within this study, based on six different tested parameters.

Although it is recognized that such RPIs may have specific applications (provided that they are based on sound statistical principles and issued with proper cautionary notice), every RPI can have its limitations (e.g., the loss of the Z-scores signs). Moreover, in this specific case, any RPI can be biased by the high uncertainty associated with the consensus values of the free PRP parameters, as highlighted above in section 5.4.

For this reason and others, the provision of combined scores is usually not recommended for inclusion in a study report for participants. Therefore, it was decided to include this information only in Appendix 4.

6. Conclusions

6.1 General comments

There was a large participation in the study both by NCLs as well as manufacturers' quality control laboratories. The goal of this PTS was reached as it provided an assessment of the proficiency of the participating laboratories to quantify the total and free (unconjugated) PRP content of the Hib component in different liquid pentavalent vaccine (DTwP-HepB-Hib) presentations when tested according to a method using HPAEC-PAD. This goal was achieved despite the limitations of the reported Z-scores (see § 5.4). For none of the participating laboratories could an unsatisfactorily performance be observed in terms of Z-scores for all six tested parameters, even though there was limited validity. In fact, no laboratory obtained Z-scores > 3 for all the 6 parameters, and therefore all the ARSSZ scores are lower than 9 (see Appendix 4).

The determination of the free PRP content was more variable for all samples than the determination of the corresponding total PRP content. This means that poor reproducibility was observed for this parameter as illustrated by the results reported in the tables for the method precision (tables 3A - 3C). Thus, for the total PRP results, the partition between intra- and inter-laboratory variabilities is as expected (i.e., inter-lab variances account for 74, 65 and 78 percent of the total variance). However, the inter-laboratory portion of the total variability represents 96 - 98 % of the total variability for the free PRP results. The observed higher variability may be related to the additional sample preparation and processing that is required for the free PRP determination whereas the determination of the total PRP requires only a dilution of the sample under test.

A PTS is performed once a method has been shown to be suitable for the intended purpose. The validity of this method has already been evaluated in small studies and in a published collaborative study [10]. It is assumed that the WHO method had been validated in-house by each participating laboratory prior to participation in the PTS. Participation in this proficiency testing scheme provided the laboratories with an objective means of assessing and demonstrating the reliability of the data being produced.

6.2. Individual laboratory assessment and comments

According to the consideration given in section 5.4, this assessment was based not only on criteria described in Appendix 1, but also on examination of Youden Plots. A decision was made to report "partially unsatisfactory" proficiency to those participants that showed a clear systematic error, and a "questionable" proficiency to those laboratories with sporadic random error. As already stated in section 5.4, the comments below are intended for information only.

Lab 1: for information only: a large intra-lab variability was observed for the total PRP content for PTS-2 sample (GCV = 31.9%). This is due to a higher value in the first run; it does not affect the overall good performance.

Lab 2: Partially unsatisfactory: a systematic bias was observed. The laboratory also tends to underestimate the true value for all the tested parameters as visible in the Youden plots (note that this is also illustrated by the high negative RPIs).

Lab 3: The laboratory was in the lower mode group for the free PRP content of PTS sample number 2.

Lab 4: No comment.

Lab 5: No comment.

Lab 6: There was a very slight tendency to underestimate the saccharide contents, both for total and free PRP. However, the size of the negative bias was very small. The bias was slightly higher for the free PRP measurements (see lower mode on free PRP for PTS-2).

Lab 7: Questionable: a questionable Z-score was observed for the free PRP content of the sample PTS-1. The laboratory tends to systematically underestimate (slightly) the total PRP content and, at the same time, to overestimate (slightly) the content of the free PRP.

Lab 8: No comment.

Lab 9: There was a slight tendency to underestimate the saccharide contents, both for total and free PRP. However, the size of the negative bias was very small. The bias was slightly higher for the free PRP measurements. The lab was in the lower mode group regarding the free PRP content of the PTS-2 sample.

Lab 10: No comment.

Lab 11: No comment.

Lab 12: No comment.

Lab 13: There was a very slight tendency to underestimate the saccharide contents, both for total and free PRP. However, the size of the negative bias was very small. The bias was slightly higher for the free PRP measurements. The lab was in the lower mode group regarding the free PRP content of the PTS-2 sample.

Lab 14: No comment.

Lab 15: No comment.

Lab 16: No comment.

Lab 17: No comment.

Lab 18: Partially unsatisfactory: a systematic bias was observed. The laboratory tended to overestimate the true value for all the tested parameters as illustrated in the Youden plots and as highlighted by the high positive RPIs (taking into account invalid test results). The data were analysed by both including and excluding invalid test results. There were no significant changes in the final performance and assessment observed.

Lab 19: The assessment was based on only 4 parameters since the PTS-1 sample was not provided due to a limited number of samples. There was a very slight tendency to overestimate the saccharide contents. However, the size of the positive bias was very small. The bias was slightly higher for the total PRP contents, which was contrary to what was the common trend for the other laboratories.

7. References

1. Schneerson R, Barrera O, Sutton A *et al.* Preparation, characterization, and immunogenicity of *Haemophilus influenzae* type b polysaccharide protein conjugates. *J Exp* 1980; **152**:361–6.

2. Zepp F, Schmitt HJ, Kaufhold A *et al.* Evidence for induction of polysaccharide specific B-cell-memory in the 1st year of life: plain *Haemophilus influenzae* type–PRP (Hib) boosters children primed with a tetanus-conjugate Hib-DTaP-HBV combined vaccine. *Eur J Pediatr* 1997; **156**:18–24.

3. Frasch CE. Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. *Vaccine* 2009; **27**:6468–6470.

4. Recommendations for the production and control of *Haemophilus influenzae* type b conjugate vaccines. WHO TRS 897, 2000.

5. *Haemophilus influenzae* type b conjugate vaccine. Monograph 1219. *Ph.Eur.* Suppl. 10. Strasbourg, France; Council of Europe; 2020.
6. WHO Procedure for assessing the acceptability, in principle, of vaccines for purchase by United Nations Agencies. WHO Technical Report Series 978, Annex 6. Geneva: World Health Organization, 2010.

7. Efficient means to determine Hib component in liquid vaccine presentations, for different vaccines, produced by different manufacturers. Available from: https://www.who.int/immunization_standards/vaccine_quality/Study_Report_Vaccines_May2014.pdf?ua=1; last visited 25/06/2020

8. ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories; 2017.

9. World Health Organization. *Supporting countries in quality assurance for pentavalent vaccines. First WHO-ISS hands-on laboratory training course held in Rome, Italy*. Geneva: World Health Organization; 2015. Available from: http://www.who.int/immunization_standards/vaccine_quality/hib_component_determination/en/ ; last visited 25/06/2020

10. Rosskopf U, Daas A, Terao E, von Hunolstein C. Collaborative study on saccharide quantification of the *Haemophilus influenza*e type b component in liquid vaccine presentations. *Pharmeur Bio Sci Notes* 2017; 2017:44-68.

11. ISO/13528:2015: Statistical methods for use in proficiency testing by interlaboratory comparisons.

12. ISO Guide 98-3, Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement (GUM)

13. Eurachem: Selection, use and interpretation of proficiency testing (PT) schemes by laboratories. 2000.

14. M. Thompson *et al.* IUPAC International - The International harmonized protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report) - Pure Appl. Chem., 2006; 78, No.1:145–196.

15. Youden WJ. Graphical diagnosis of inter-laboratory test results. Industrial Quality Control.1959; 15: 24-28.

8. Acknowledgements

The authors give sincere thanks to:

- All participating laboratories for their valuable contribution to this study.
- The following vaccine manufacturers (in alphabetical order) for their donation of the vaccine samples, especially the preparation of a pentavalent vaccine exclusively for the purpose of this study:
  - Biological E. Limited, India
  - PT Bio Farma (Persero), Indonesia
  - Serum Institute of India Limited, India.
- NIBSC for donation of the WHO 2nd IS for Hib PRP Polysaccharide (12/306).
- The following WHO colleagues for the preparative work and shipment of the study samples: Mr Kamel Aboudi, Mrs Tomislava Bouquet, Mr Raimundo Gomes da Silva Arbenz and colleagues, Mr Daniel Hollies as well as Ms Louise Els (consultant).
- Bruce Meade for editorial review of the report.

9. List of participants

Country	Name	Institution
Bangladesh	Dr Mohammad Mainul AHASAN	Incepta Vaccine Ltd.
Belgium	Dr Lorenzo TESOLIN Dr Laure CUIGNET Dr Adeline GAZAN	Sciensano Quality of Vaccines and Blood Products Biochemistry team
China	Dr Qiang YE Dr Maoguang LI	National Institute for Food and Drug Control (NIFDC) Division of Respiratory Bacterial Vaccines
France	Dr Olivier GARINOT	French National Agency for Medicines and Health Products Safety (ANSM) Laboratory Controls Division
Germany	Dr Wolf Hagen HOLTKAMP	Paul-Ehrlich Institut (PEI) Section 3/1 Product testing of immunological medicinal products
India	Dr Dipankar DAS	Bharat Biotech International Ltd.
India	Dr Srinivas KOSARAJU	Biological E. Limited
India	Dr Ravikumar RAMPA	Cadila Health Care Limited Zydus Vaccine Division
India	Mr Jaipal MEENA Mr Ade Ajay KUMAR	National Institute of Biologicals (NIB)
India	Dr Deepak MAHAJAN	Panacea Biotec Ltd.
India	Dr Sunil GAIROLA	Serum Institute of India PVT LTD
India	Dr Radhakrishnam Raju MANTENA	Shantha Biotechnics Limited (A Sanofi Company)
Indonesia	Dra Togi Junice HUTADJULU Ms Elisabeth ARISETININGSIH	National Quality Control Laboratory of Drug and Food (NQCLDF) National Agency of Drug and Food Control
Indonesia	Dr Dori UGIYADI Ms Lin SUSANTI	PT Bio Farma (Persero)

Italy	Dr Cristina PEZZELLA Dr Luisa RALLI	Istituto Superiore di Sanità (ISS)
Mexico	Ms Imelda Rocío GUZMÁN CERVANTES Ms Laura MUNGUÍA Dr Guillermo VEGA RODRIGUEZ	Commission for Analytical Control and Expansion of Coverage (CCAYAC) Federal Commission for the Protection from Sanitary Risks (COFEPRIS)
Republic of Korea	Ms Helen KIM Dr ByoungKook HYUN	LG Chem Ltd.
Republic of Korea	Dr Naery LEE Dr Sun bo SHIM	National Institute of Food and Drug Safety Evaluation (NIFDC) Vaccines Division
South Africa	Dr Ruan ELLS	National Control Laboratory for Biological Products (NCLBP)
Thailand	Dr Supaporn PHUMIAMORN Mrs Wereyarmarst JAROENKUNATHUM	Institute of Biological Products (IBP), Department of Medical Sciences, Ministry of Public Health
United Kingdom	Dr Frank GAO Dr Barbara BOLGIANO Mrs Karena Burkin	National Institute for Biological Standards and Control (NIBSC) Bacterial Division

Appendix 1 – Statistical Analysis

All results were collected at ISS and the data were then analysed using the software IBM SPSS 25.0 and MS Excel 16.0. An assessment of intra-laboratory precision is provided for each of the sample vaccines and for both the total and the free saccharide measures. For laboratory assessment, Z-scores were intended to be employed.

Z-scores are performance scores for proficiency assessment. Z-scores evaluate the difference between each participant's result and the assigned value (in this study, a consensus value based on the results from the participating labs will be used). This difference is then compared as a ratio to the overall (consensus) standard deviation. Therefore, an assigned value and standard deviation for proficiency assessment are necessary; both were derived from the results of the participating laboratories (i.e., consensus values).

According to the ISO 13528:2005 (E) there are several recognised ways to establish the consensus value and the standard deviation for proficiency assessment in a proficiency testing scheme (e.g., algorithm A [11]).

Once the consensus value and the standard deviation have been determined, the Z-scores for each laboratory are then calculated, one score for each of the six reported parameters (i.e., total and free PRP for each of the three samples):

$$\mathrm{Z-score} = \frac{\mathrm{Lab's\ Result} - \mathrm{Consensus\ Value}}{\mathrm{Consensus\ Standard\ Deviation}}$$

For the purposes of performance assessment, the following classification is commonly adopted:
$|Z| \leq 2.00$ Satisfactory result
$2.00 < |Z| \leq 3.00$ Questionable result
$|Z| > 3.00$ Unsatisfactory result

Uncertainty associated with the consensus values is also reported. However, as already stated in the text of this report, non-negligible uncertainty exists with respect to how close the consensus values are to the true values, mainly for the free PRP parameters.

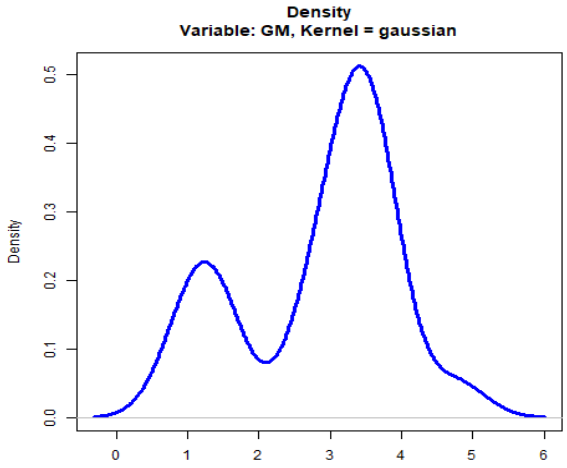
The final results of the study also are presented in the tables and graphs (e.g., histograms, descriptive individual plots, and Youden Plots) provided in this report.

Appendix 2 – Robust Estimators

Table A-1. Consensus Values: Robust Estimators *versus* Mean Value

Sample			Robust Estimators					Mean Value
		Shapiro-Wilk p-value	Huber's M-Estimator[a]	Tukey's Biweight[b]	Hampel's M-Estimator[c]	Andrews' Wave[d]	Alg_A[e] (ISO 13528)	
PTS-1	Total PRP	*0.927*	10.483	10.534	10.517	10.534	10.543	**10.502**
	Free PRP	*0.375*	1.173	1.113	1.167	1.107	**1.176**	1.209
PTS- 2	Total PRP	*0.108*	5.786	5.790	5.783	5.795	**5.835**	5.886
	Free PRP	*0.020*	**3.171**	3.380	3.194	3.380	2.857	2.856
PTS- 3	Total PRP	*0.976*	8.116	8.093	8.145	8.091	8.173	**8.191**
	Free PRP	*0.510*	1.187	1.162	1.190	1.162	1.200	**1.220**
All results (except the p-values) are in µg / shd. In bold are the adopted consensus values								
a. The weighting constant is 1.339.								
b. The weighting constant is 4.685.								
c. The weighting constants are 1.700, 3.400, and 8.500								
*d. The weighting constant is 1.340*pi.*								
e. The weighting constant is 1.483								
Another robust estimator, i.e., the Median, is used and shown in the Youden Plots figures								

Figure A-1. Kernel Density for free PRP in PTS-2 Sample



Appendix 3 - Youden Plot Details

The Youden plots provide another view of the performance, using an approach which emphasizes bias. One Youden plot is created for each PTS sample (figures 7-9). They are constructed as follows: a vertical line and a horizontal line are drawn through the median values. The axes of the plot are not drawn on the same scale; instead, one standard deviation on the x-axis has the same length as one standard deviation on the y-axis.

A horizontal median line is drawn parallel to the x-axis so that there are as many points above the line as there are below it. A second median line is drawn parallel to the y-axis so that there are as many points on the left as there are on the right of this line. Outliers (according to the Tukey approach) are not used in determining the position of the median lines. The intersection of the two median lines is called the Manhattan median.

Analogous to the 45-degree reference line in the original Youden plot, a reference line is drawn which in this case represents a constant ratio of the two parameters.

Two circles are drawn that should include 95% and 99% of the laboratories, respectively, if individual constant errors could be eliminated (95% and 99% coverage probabilities).

Therefore, the radii of the two circles are calculated based on the coverage probabilities of 95% and 99%. Specifically, the radius of the circles is obtained by multiplying the quantity $\sqrt{-2\ln(p)}$ (with p equal to 0.05 and 0.01 for the inner and the outer circle, respectively) by the averaged intra-lab standard deviation.

Appendix 4 - Rolling Performance Indicators Approach

A list of four types of scores combination (that may be useful to assess a sequence of Z-scores) is given below (Z-score is simply denoted with z; n is the number of the tested parameters):

- RSZ - Rescaled Sum of the Z-Scores; RSZ = $\frac{\sum_{i=1}^{n} z_i}{\sqrt{n}}$
- SSZ – Sum of the Squared Z-Scores; SSZ = $\sum_{i=1}^{n} z_i^2$
- RSSZ – Rescaled Sum of the Squared Z-Scores, RSSZ = $\frac{\sum_{i=1}^{n} z_i^2}{n}$
- ARSSZ – Absolute Rescaled Sum of the Squared Z-Scores, ARSSZ = $\frac{\sum_{i=1}^{n} |z_i| * z_i}{n}$

In this report, only the RSZ and ARSSZ are presented (see table A-2 below), since they are considered the most informative and statistically sound indicators in this context. However, it is emphasized that there are limitations and weaknesses in any indicator that combines Z-scores from dissimilar parameters.

Table A-2: RSZ and ARSSZ Results

Lab	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
RSZ	1.35	-4.72	-1.55	0.24	-1.72	-2.61	1.43	0.52	-2.24	1.72	-0.55	-1.73	-2.08	1.36	0.08	1.71	0.76	5.31	1.66
ARSSZ	0.54	-3.85	-0.67	0.04	-0.60	-1.40	1.33	0.11	-0.96	0.82	-0.06	-0.69	-0.90	0.43	0.02	0.53	0.17	5.33	1.16

Note: for RSZ the shading means the same as reported in table 4: for ARSSZ the shading denotes results higher than |3|, denoting partially unsatisfactory performance

The RSZ "can" be interpreted as a single Z-score, i.e., it is expected to be zero-centered with Variance = 1. This indicator has the useful property of demonstrating a persistent bias or trend, so that a sequence of satisfactory results [e.g.: 1.4, 1.8, 1.1, 1.5, 1.7, 1.7] would provide statistically significant RSZ equal to 3.0, even though each individual Z-score is less than 2. As is obvious, the main deficiency is the lost information due to large Z-scores of opposite signs.

The ARSSZ has the advantage of avoiding the cancellation of large Z-scores of opposite signs (the signs are maintained), but is less sensitive to small biases. If an ARSSZ is lower than 9, it means that there is not a systematic unsatisfactory performance (i.e., when all the six Z-scores ≥ |3|).

It is recommended that these combinations of scores should not be misused; there is a danger that such an approach can be misinterpreted or abused by non-experts, especially outside the context of the individual scores.

It is especially emphasized that there are limitations and weaknesses in any approach that combines Z-scores from dissimilar analyses. If a single score out of several produced by a laboratory were outlying, the combined score may well be not outlying. In some respects, this is a useful feature, in that a lapse in a single analysis is down weighted in the combined score. However, there is a danger that a laboratory may be consistently yield a faulty value for only a single parameter, and thus frequently report an unacceptable value for that analysis in all the tested samples. This factor may well be obscured by the combination of scores.

TABLES

Table 1. Overview of test panel composition

Vaccine sample code	Doses per container	Quantity of provided vials	Specification for total and free saccharide content
PTS-1	5-dose	12	Total: 8-12 mcg/0.5 mL Free: < 20% of total PRP
PTS-2	10-dose	6	not applicable (inferior quality)
PTS-3	10-dose	6	Total: 8-12 mcg/0.5 mL Free: < 30% of total PRP

Table 2A. PTS-1 sample: Individual results reported by the participating laboratories

Lab	Total PRP - µg / shd (shd - 0.5 ml)					Free PRP - µg / shd (shd - 0.5 ml)					Free PRP - % of Total PRP				
	Run 1	Run 2	Run 3	GM	GCV (%)	Run 1	Run 2	Run 3	GM	GCV (%)	Run 1	Run 2	Run 3	GM	GCV (%)
1	14.15	11.42	11.33	**12.23**	**12.66**	1.23	1.26	1.09	**1.19**	**7.78**	8.69	10.98	9.63	**9.72**	**11.76**
2	6.09	8.94	7.46	**7.41**	**19.38**	0.46	0.35	0.37	**0.39**	**14.52**	7.39	3.91	4.96	**5.23**	**33.02**
3	10.04	8.47	9.8	**9.41**	**9.22**	1.51	1.38	1.2	**1.36**	**11.62**	15.04	16.29	12.24	**14.42**	**14.83**
4	11.68	10.74	11.04	**11.15**	**4.28**	1.16	1.12	1.11	**1.13**	**2.33**	9.93	10.41	10.09	**10.14**	**2.40**
5	9.23	9.29	8.66	**9.06**	**3.88**	0.91	0.90	0.85	**0.89**	**3.66**	9.83	9.62	9.82	**9.76**	**1.22**
6	9.9	9.65	9.71	**9.75**	**1.33**	0.54	0.51	0.7	**0.58**	**17.00**	5.45	5.28	7.21	**5.92**	**17.27**
7	10.03	10.43	10.33	**10.26**	**2.04**	2.42	2.64	1.94	**2.31**	**15.98**	24.13	25.28	18.78	**22.54**	**16.09**
8	11.76	10.92	10.48	**11.04**	**5.84**	1.21	1.19	1.2	**1.20**	**0.83**	10.29	10.86	11.41	**10.84**	**5.17**
9	9.97	9.58	9.95	**9.83**	**2.25**	0.58	0.52	0.56	**0.55**	**5.58**	5.79	5.44	5.64	**5.62**	**3.13**
10	12.43	12.68	12.48	**12.53**	**1.05**	1.24	1.19	1.25	**1.23**	**2.64**	9.96	9.33	10.00	**9.76**	**3.89**
11	10.33	9.97	10.47	**10.25**	**2.53**	0.97	1.01	1.15	**1.04**	**9.00**	9.00	10.00	11.00	**9.97**	**10.06**
12	9.35	8.78	8.62	**8.91**	**4.26**	1.17	1.05	1.08	**1.10**	**5.62**	12.51	11.97	12.56	**12.34**	**2.67**
13	9.71	9.92	9.38	**9.67**	**2.82**	0.60	0.62	0.61	**0.61**	**1.64**	6.13	6.25	6.52	**6.30**	**3.15**
14	11.94	11.5	10.86	**11.42**	**4.78**	1.87	1.73	1.61	**1.73**	**7.50**	15.58	15.04	14.73	**15.11**	**2.84**
15	9.48	10.76	10.62	**10.27**	**6.97**	1.39	1.57	1.39	**1.45**	**7.04**	14.67	14.57	13.03	**14.07**	**6.66**
16	11.45	11.06	11.69	**11.40**	**2.80**	1.63	1.72	1.59	**1.65**	**4.02**	14.24	15.55	13.6	**14.44**	**6.82**
17	11.9	12.68	10.09	**11.50**	**11.83**	1.16	1.22	1.08	**1.15**	**6.19**	9.74	9.58	10.66	**9.98**	**5.75**
18	12.24	14.69	12.03	**12.93**	**11.10**	2.29	2.15	2.15	**2.20**	**3.64**	19.00	15.00	18.00	**17.25**	**12.43**
19	na	na	na	**na**	**na**	na	na	na	**na**	**na**	na	na	na	**na**	**na**
		Overall	**GM**	**10.41**			**Overall**	**GM**	**1.09**			**Overall**	**GM**	**10.49**	
			GCV (%)	**13.91**				**GCV (%)**	**50.82**				**GCV (%)**	**42.03**	

Table 2B. PTS-2 sample: Individual results reported by the participating laboratories

Lab	Total PRP - µg / shd (shd - 0.5 ml)					Free PRP - µg / shd (shd - 0.5 ml)					Free PRP - % of Total PRP				
	Run 1	Run 2	Run 3	*GM*	*GCV%*	Run 1	Run 2	Run 3	*GM*	*GCV%*	Run 1	Run 2	Run 3	*GM*	*GCV%*
1	9.17	5.48	5.23	***6.41***	***31.93***	3.67	3.61	3.69	***3.66***	***1.14***	40.46	65.88	70.60	***57.31***	***31.05***
2	3.53	4.57	3.81	***3.95***	***13.32***	0.79	0.96	1.02	***0.92***	***13.41***	22.38	20.79	26.77	***23.18***	***13.06***
3	6.41	4.36	5.45	***5.34***	***19.53***	1.69	1.60	1.13	***1.45***	***22.09***	26.37	36.70	20.73	***27.17***	***29.28***
4	6.17	5.82	5.92	***5.97***	***3.00***	3.24	3.26	3.29	***3.26***	***0.77***	52.45	56.03	55.51	***54.64***	***3.57***
5	5.12	5.16	5.16	***5.14***	***0.42***	3.04	3.10	3.14	***3.09***	***1.63***	59.23	60.57	60.86	***60.22***	***1.45***
6	4.99	5.06	5.53	***5.19***	***5.58***	0.92	1.12	1.12	***1.05***	***11.39***	18.44	22.13	20.25	***20.22***	***9.14***
7	5.82	5.32	5.56	***5.57***	***4.49***	3.55	3.75	3.72	***3.67***	***2.96***	60.90	70.37	66.83	***65.92***	***7.33***
8	6.49	5.90	6.03	***6.13***	***5.00***	2.86	2.85	2.93	***2.88***	***1.51***	44.14	48.35	48.55	***46.97***	***5.39***
9	5.51	4.92	5.50	***5.30***	***6.49***	1.46	1.34	1.47	***1.42***	***5.16***	26.43	27.21	26.68	***26.77***	***1.48***
10	6.70	6.77	6.94	***6.80***	***1.80***	3.54	3.60	3.42	***3.52***	***2.62***	52.83	53.18	49.25	***51.72***	***4.26***
11	5.44	5.57	5.81	***5.60***	***3.33***	3.10	2.89	3.19	***3.05***	***5.09***	57.00	52.00	55.00	***54.63***	***4.63***
12	5.15	4.91	4.89	***4.98***	***2.88***	2.69	2.55	2.56	***2.60***	***2.98***	52.03	52.01	52.30	***52.11***	***0.31***
13	5.70	5.42	5.33	***5.48***	***3.49***	1.24	1.33	1.29	***1.29***	***3.51***	21.71	24.55	24.06	***23.41***	***6.60***
14	6.52	6.24	5.49	***6.07***	***8.95***	3.7	3.41	3.18	***3.42***	***7.59***	56.75	54.48	57.92	***56.37***	***3.12***
15	5.27	6.14	5.82	***5.73***	***7.76***	2.95	3.69	3.17	***3.26***	***11.46***	55.97	60.06	54.36	***56.75***	***5.13***
16	6.34	6.23	7.07	***6.54***	***6.86***	3.78	3.60	3.72	***3.70***	***2.49***	59.62	57.78	52.62	***56.59***	***6.50***
17	6.41	6.83	5.36	***6.17***	***12.62***	3.75	3.73	3.27	***3.57***	***7.78***	58.41	54.59	60.86	***57.90***	***5.49***
18	9.06	8.38	8.66	***8.70***	***3.92***	4.80	4.67	4.83	***4.77***	***1.80***	53.00	56.00	56.00	***54.98***	***3.18***
19	6.60	7.12	6.59	***6.77***	***4.43***	3.47	3.78	3.81	***3.68***	***5.19***	52.64	53.16	57.77	***54.48***	***5.11***
		Overall	**GM**	**5.76**			**Overall**	**GM**	**2.54**			**Overall**	**GM**	**44.08**	
			GCV (%)	**16.23**				**GCV (%)**	**53.38**				**GCV (%)**	**41.42**	

Table 2C. PTS-3 sample: Individual results reported by the participating laboratories

Lab	Total PRP - µg / shd (shd - 0.5 ml)					Free PRP - µg / shd (shd - 0.5 ml)					Free PRP - % of Total PRP				
	Run 1	Run 2	Run 3	GM	GCV%	Run 1	Run 2	Run 3	GM	GCV%	Run 1	Run 2	Run 3	GM	GCV%
1	11.29	8.63	8.28	***9.31***	***16.95***	1.27	1.13	1.08	***1.16***	***8.57***	11.25	13.10	13.00	***12.42***	***8.59***
2	5.65	6.67	5.41	***5.89***	***11.08***	0.33	0.29	0.40	***0.34***	***16.29***	5.84	4.20	7.39	***5.66***	***28.96***
3	7.99	6.42	6.90	***7.07***	***11.18***	1.17	1.08	0.89	***1.04***	***14.45***	14.64	16.82	12.90	***14.7***	***13.33***
4	8.51	8.02	8.35	***8.29***	***3.03***	1.21	1.10	1.12	***1.14***	***5.07***	14.17	13.70	13.40	***13.75***	***2.81***
5	7.11	7.19	7.15	***7.15***	***0.60***	0.78	0.77	0.75	***0.77***	***2.00***	11.02	10.66	10.50	***10.72***	***2.47***
6	7.67	7.56	7.88	***7.70***	***2.10***	0.21	0.21	0.20	***0.21***	***2.82***	2.74	2.78	2.54	***2.68***	***4.85***
7	8.01	7.25	7.88	***7.71***	***5.35***	1.92	2.59	2.64	***2.36***	***18.00***	24.03	35.69	33.49	***30.62***	***21.48***
8	9.87	8.58	8.45	***8.94***	***8.58***	1.35	1.16	1.23	***1.24***	***7.66***	13.63	13.41	14.50	***13.84***	***4.12***
9	7.41	7.49	7.68	***7.53***	***1.84***	0.57	0.56	0.56	***0.56***	***1.02***	7.62	7.47	7.26	***7.45***	***2.43***
10	9.28	9.37	9.51	***9.39***	***1.23***	1.23	1.25	1.24	***1.24***	***0.81***	13.16	13.29	12.99	***13.15***	***1.15***
11	7.86	7.76	8.33	***7.98***	***3.78***	0.98	0.97	1.15	***1.03***	***9.48***	13.00	12.00	14.00	***12.97***	***7.72***
12	6.81	7.22	6.71	***6.91***	***3.88***	1.09	1.12	1.05	***1.09***	***3.24***	15.95	15.50	15.55	***15.67***	***1.57***
13	7.71	7.93	7.77	***7.80***	***1.45***	0.58	0.59	0.57	***0.58***	***1.72***	7.57	7.37	7.29	***7.41***	***1.94***
14	8.72	8.31	8.13	***8.38***	***3.59***	1.88	1.73	1.75	***1.79***	***4.51***	21.56	20.82	21.53	***21.30***	***1.98***
15	6.94	8.37	7.74	***7.66***	***9.43***	1.34	1.59	1.43	***1.45***	***8.65***	19.33	18.99	18.38	***18.90***	***2.56***
16	8.56	8.87	8.63	***8.69***	***1.86***	1.66	1.76	1.76	***1.73***	***3.38***	19.39	19.84	20.39	***19.87***	***2.52***
17	8.97	9.45	7.80	***8.71***	***9.95***	1.16	1.32	1.12	***1.20***	***8.45***	12.89	13,91	14.39	***13.72***	***5.64***
18	11.42	10.46	10.56	***10.80***	***4.82***	2.44	2.29	2.41	***2.38***	***3.36***	11.42	10.46	10.56	***10.80***	***4.82***
19	9.45	9.84	9.89	***9.72***	***2.49***	1.71	1.97	1.92	***1.86***	***7.55***	18.04	20.00	19.45	***19.15***	***5.34***
		Overall	**GM**	**8.12**			**Overall**	**GM**	**1.04**			**Overall**	**GM**	**12.38**	
			GCV (%)	**14.12**				**GCV (%)**	**69.84**				**GCV (%)**	**58.78**	

na (not applicable): laboratory 19 did not receive PTS-1 sample and therefore no results were reported.
GM: Geometric Mean.
For the calibration curve, Lab 1 and Lab 2 used the WHO 2nd IS for PRP
GCV: Geometric Coefficient of Variation (defined by sqrt(e^{ω} -1)*100), with ω = Sample Variance of the ln-transformed results). GCVs for each Lab are single measures of Intra-Lab Precision. These measures should be used assuming a Log-Normal distribution.
Grey shaded values indicate results that were declared non-valid according to the outcome of the system suitability test.

Table 3A. Method precision for PTS-1 sample

	Total PRP				Free PRP			
VAR Component	**VAR Estimate**	**S (µg/shd)**	**% of tot**	**RSD (%)**	**VAR Estimate**	**S (µg/shd)**	**% of tot**	**RSD (%)**
Inter-Lab (between)	1.769	S_{Lab}=1.330	74		0.277	S_{Lab}=0.527	96	
Intra-Lab (within)	0.630	S_{Run}=0.794	26	**7.6**	0.013	S_{Run}=0.114	4	**9.4**
Reproducibility	2.400	S_{Repr}=1.549	100	**14.8**	0.290	S_{Repr}=0.539	100	**44.5**

Table 3B. Method precision for PTS-2 sample

	Total PRP				Free PRP			
VAR Component	**VAR Estimate**	**S (µg/shd)**	**% of tot**	**RSD (%)**	**VAR Estimate**	**S (µg/shd)**	**% of tot**	**RSD (%)**
Inter-Lab (between)	0.810	S_{Lab}=0.900	65		1.189	S_{Lab}=1.090	98	
Intra-Lab (within)	0.430	S_{Run}=0.656	35	**11.2**	0.027	S_{Run}=0.165	2	**5.8**
Reproducibility	1.240	S_{Repr}=1.113	100	**18.9**	1.217	S_{Repr}=1.103	100	**38.6**

Table 3C. Method precision for PTS-3 sample

	Total PRP				Free PRP			
VAR Component	**VAR Estimate**	**S (µg/shd)**	**% of tot**	**RSD (%)**	**VAR Estimate**	**S (µg/shd)**	**% of tot**	**RSD (%)**
Inter-Lab (between)	1.206	S_{Lab}=1.098	78		0.367	S_{Lab}=0.606	96	
Intra-Lab (within)	0.346	S_{Run}=0.588	22	**7.2**	0.015	S_{Run}=0.0122	4	**10.0**
Reproducibility	1.552	S_{Repr}=1.246	100	**15.2**	0.382	S_{Repr}=0.618	100	**50.7**

VAR: Variance; S: Standard Deviation; RSD: Relative Standard Deviation

S_{Run}: Intra-Lab (or within Lab) S; it represents the "Variability among Runs + Repeatability of the Method" *It is obtained by the pooled S within the 3 Runs results, per each of the N Labs.*

S_{Lab}: Inter-Lab S; it represents the variability among the participating laboratories. It is obtained by $\sqrt{S_b^2 - \frac{S_{Run}^2}{Nrun}}$ with S_b: standard deviation between laboratories, and N_{run}: number of Runs (i.e., 3).

S_{Repr} is the standard deviation of "Reproducibility"; it is obtained by $\sqrt{S_{Run}^2 + S_{Lab}^2}$.

Table 4. Z-scores

Lab	PTS - 1		PTS - 2		PTS - 3	
	Total PRP	Free PRP*	Total PRP	Free PRP^	Total PRP	Free PRP*
1	1.23	0.03	0.76	0.40	0.98	-0.10
2	-2.19	-1.49	-2.53	-1.88	-2.02	-1.45
3	-0.77	0.34	-0.66	-1.43	-0.98	-0.30
4	0.46	-0.09	0.18	0.08	0.09	-0.13
5	-1.02	-0.55	-0.92	-0.06	-0.91	-0.74
6	- 0.53	-1.13	-0.86	-1.77	-0.43	-1.66
7	-0.17	2.15	-0.36	0.42	-0.42	1.88
8	0.38	0.05	0.40	-0.24	0.66	0.03
9	-0.47	-1.18	-0.71	-1.46	-0.58	-1.09
10	1.44	0.10	1.29	0.29	1.05	0.03
11	-0.17	-0.26	- 0.31	-0.10	-0.19	-0.31
12	-1.13	-0.15	-1.14	-0.48	-1.12	-0.21
13	-0.59	-1.07	-0.47	-1.57	-0.34	-1.05
14	0.66	1.05	0.31	0.21	0.17	0.94
15	-0.16	0.51	-0.14	0.07	-0.47	0.38
16	0.64	0.89	0.94	0.44	0.44	0.84
17	0.71	-0.05	0.45	0.33	0.46	-0.03
18	1.73	1.93	3.83	1.33	2.29	1.91
19	na	na	1.24	0.43	1.34	1.05

* interpretation of the Z-scores should be made carefully due to the high uncertainty of measurement associated with the consensus values.

^ interpretation of the Z-scores should be made carefully due to the bimodal distribution of data results

"Light grey" cells indicate $|Z| \geq 2$ and < 3; "Dark grey" cell indicates $|Z| \geq 3$.

FIGURES

Figure 1A. PTS-1 sample, descriptive plot for total PRP

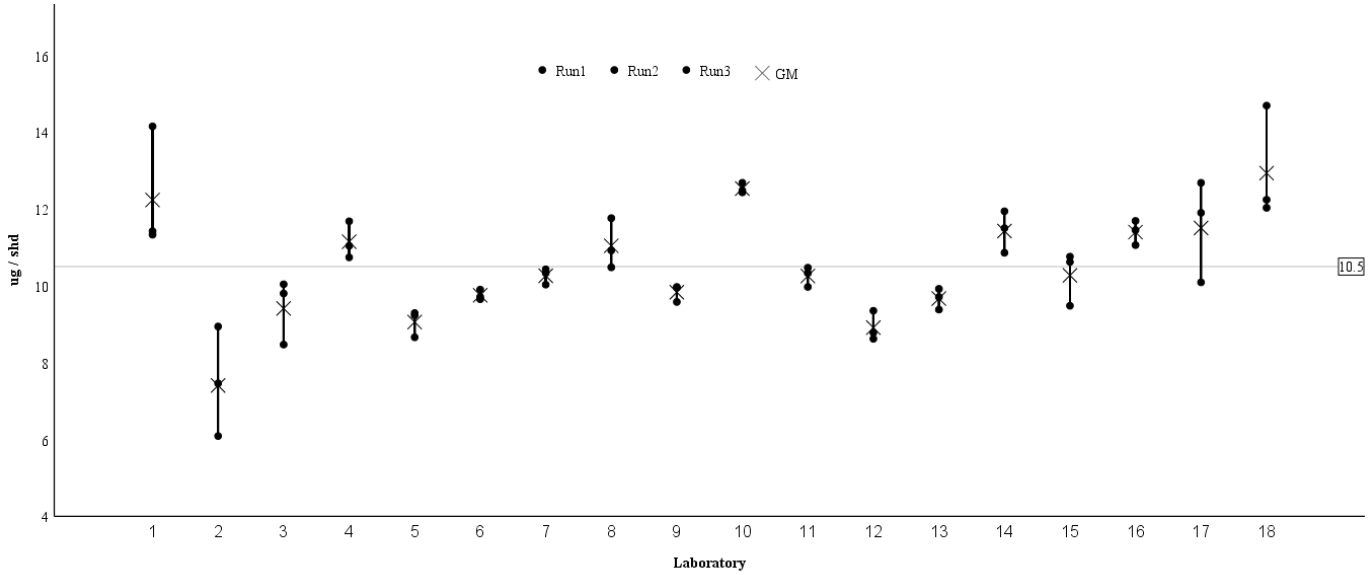


Figure 1B. PTS-1 sample, descriptive plot for free PRP

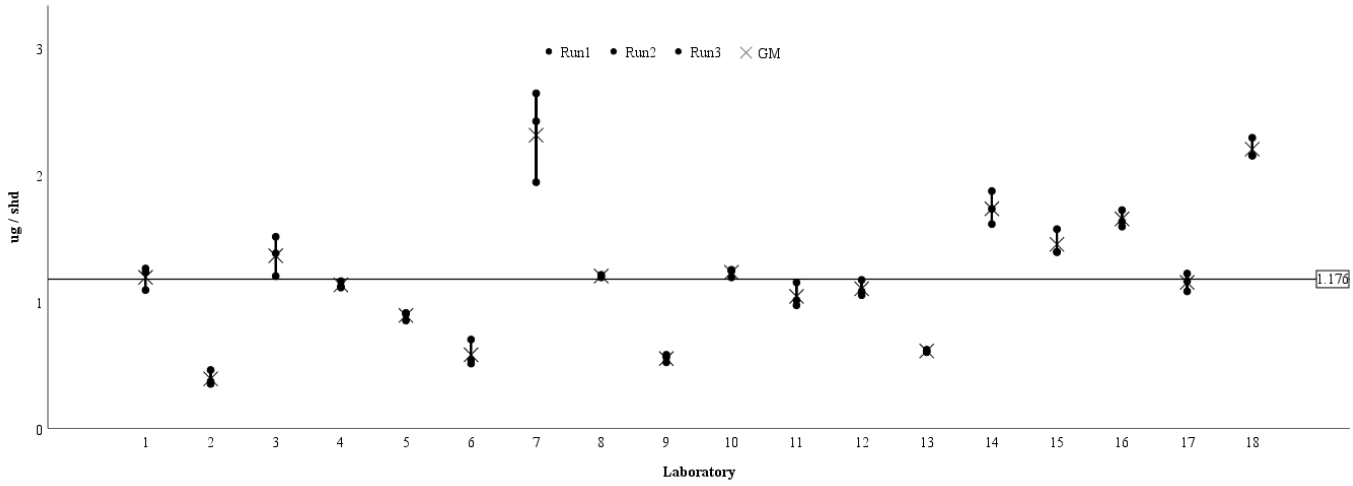


Figure 2A. PTS-2 sample, descriptive plot for total PRP

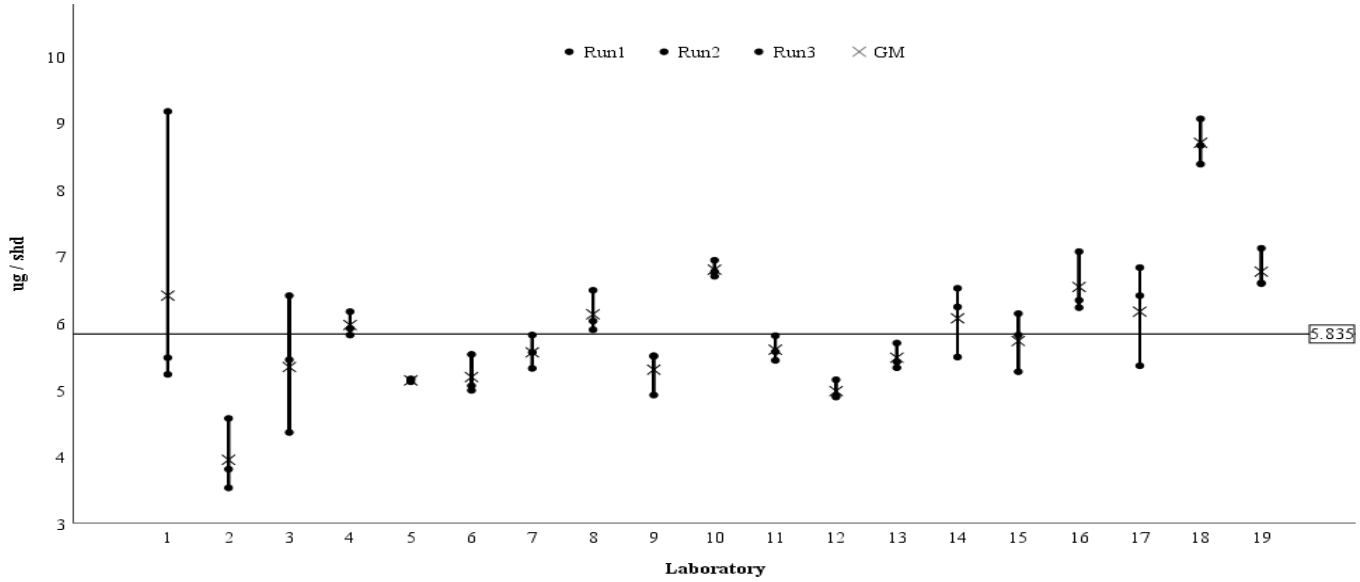


Figure 2B. PTS-2 sample, descriptive plot for free PRP

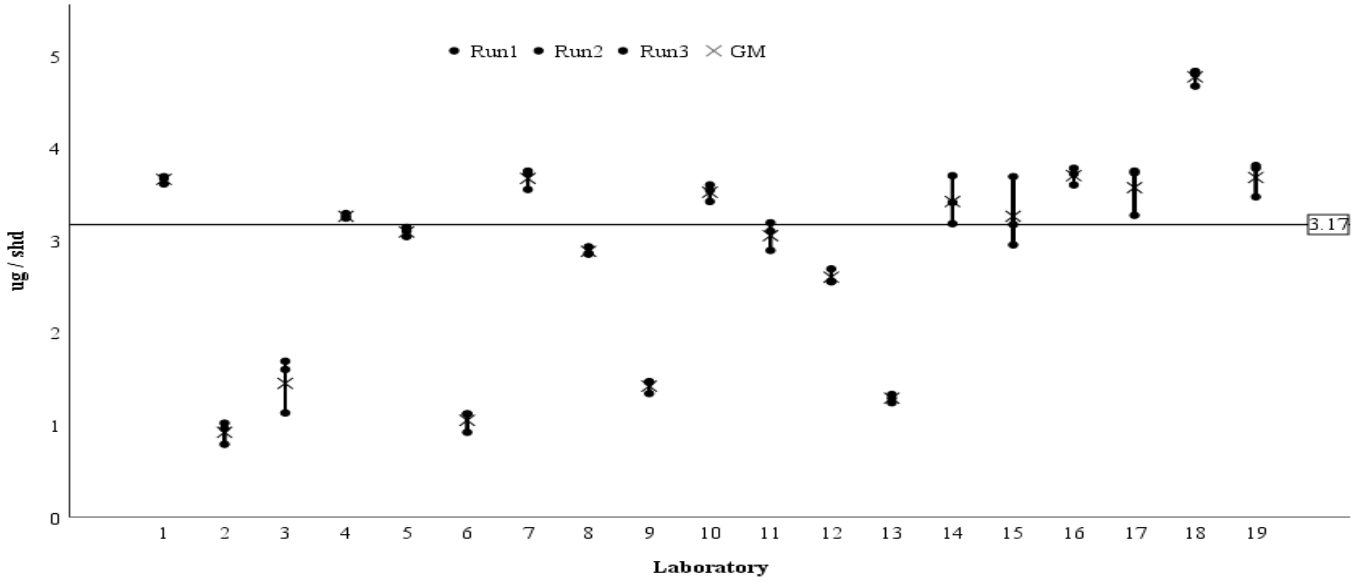


Figure 3A. PTS-3 sample, descriptive plot for total PRP

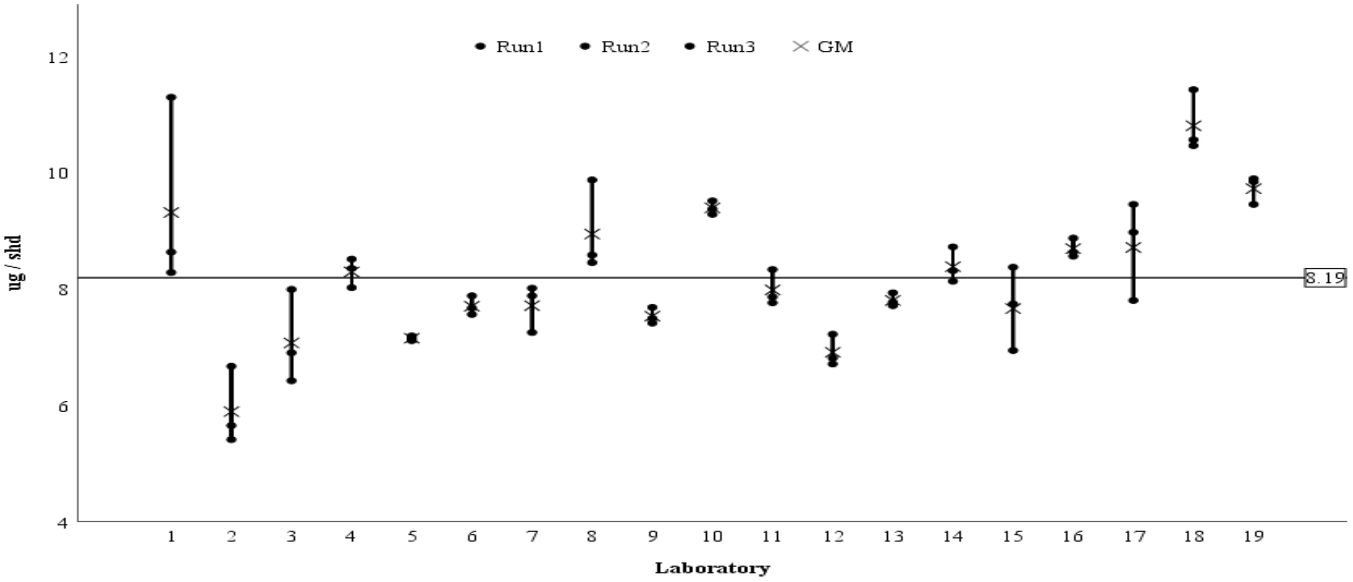
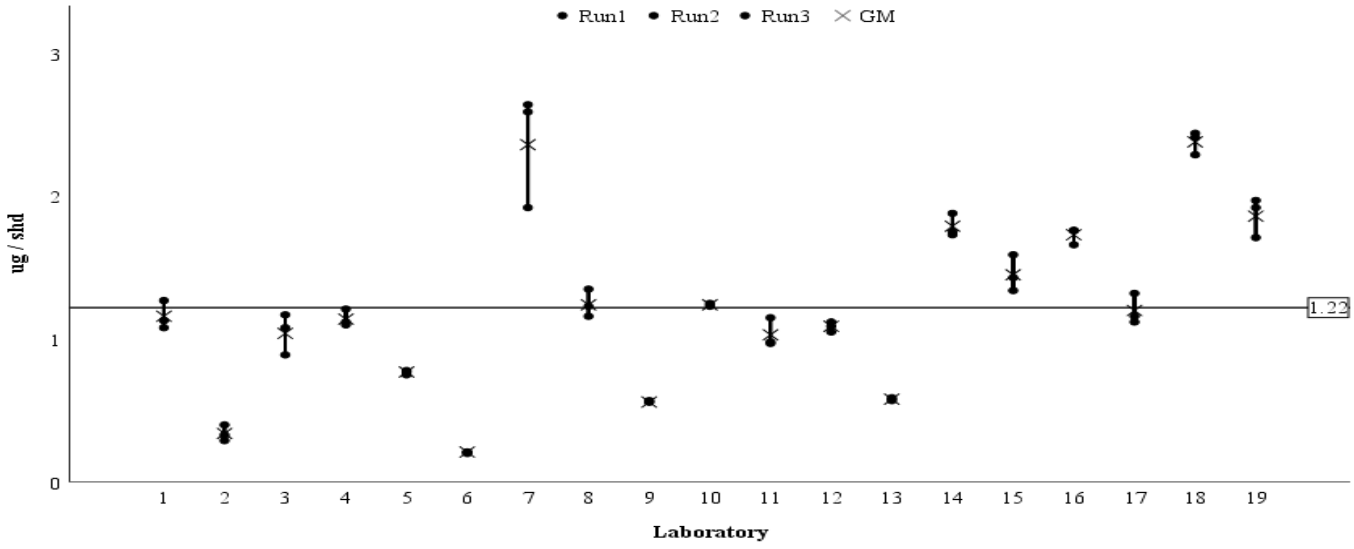


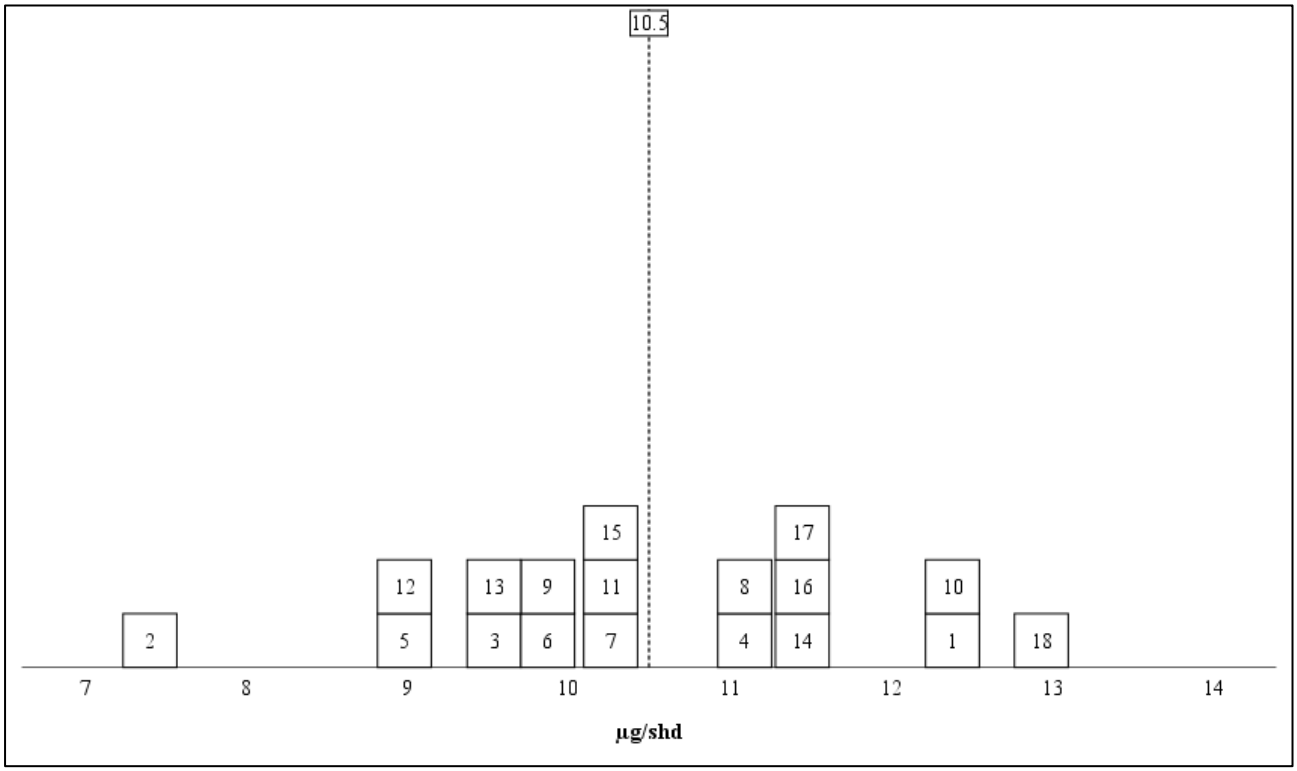
Figure 3B. PTS-3 sample, descriptive plot for free PRP



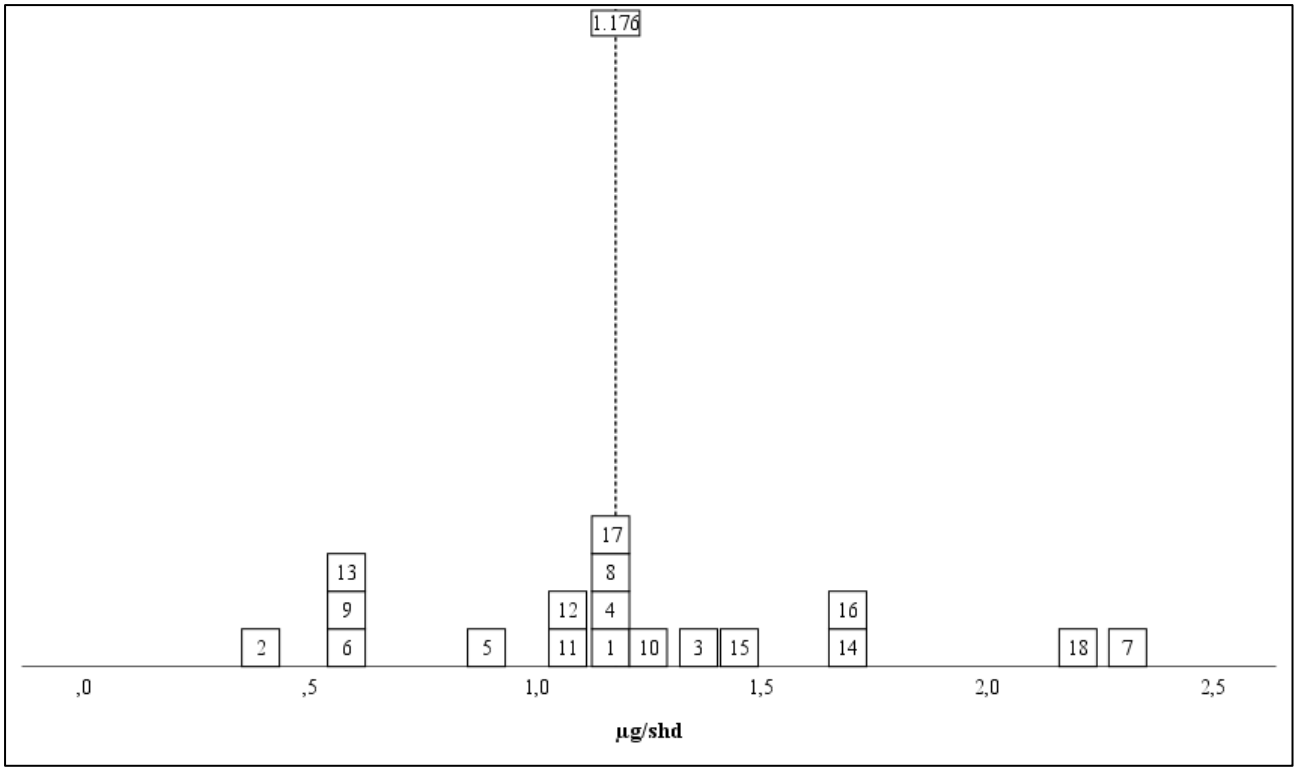
The central horizontal lines represent the consensus values.

Figures 4 A-F. Dot plots of geometric means by laboratories

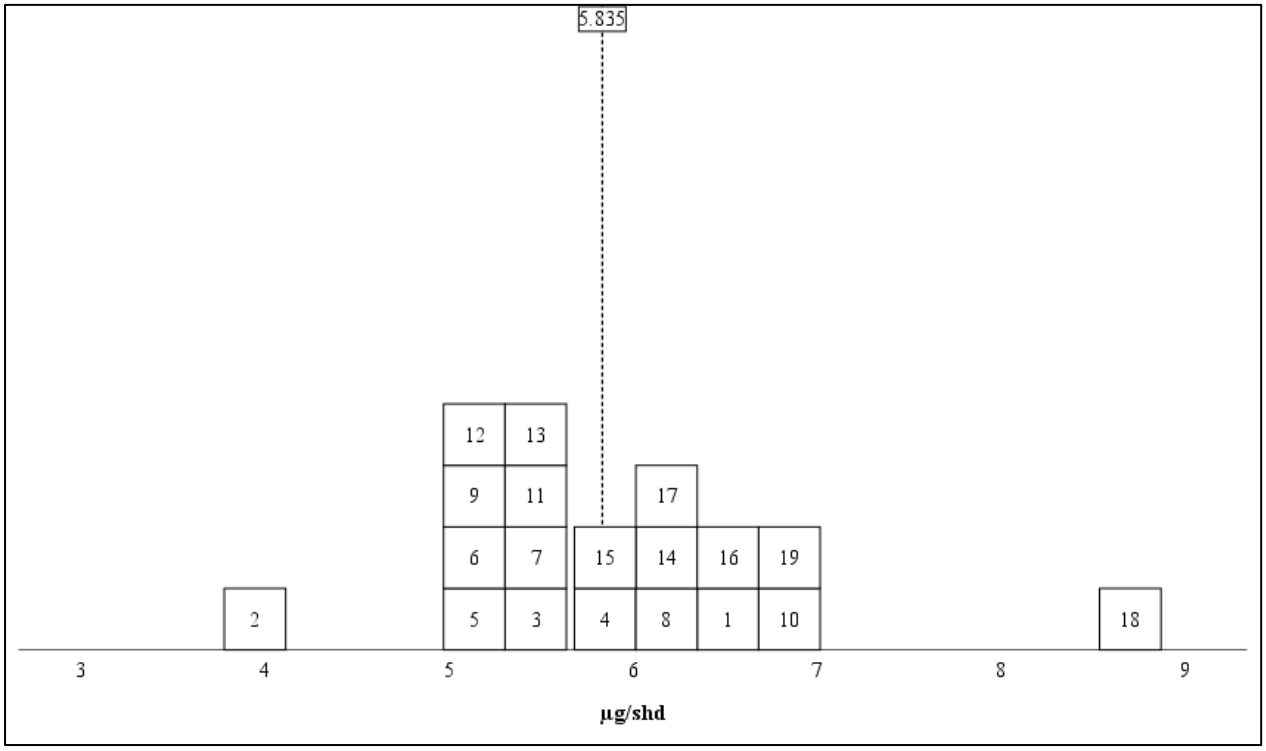
4A. PTS-1 sample, total PRP



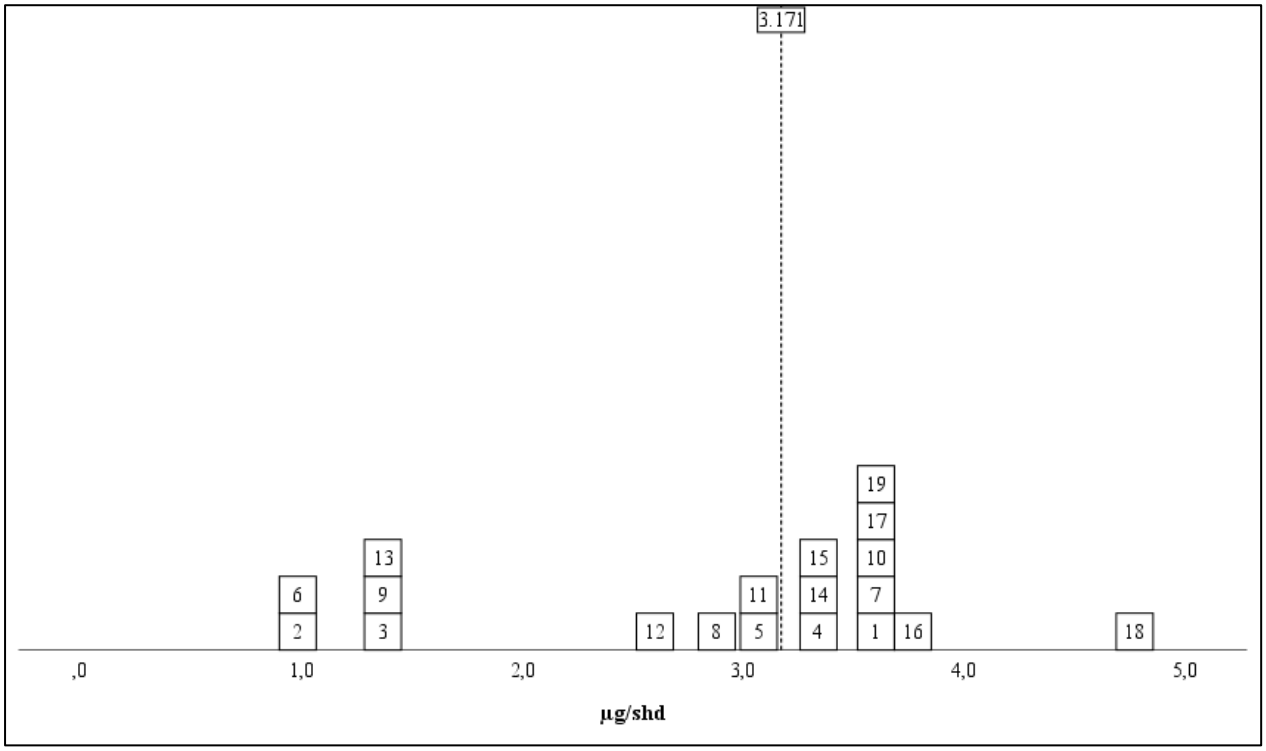
4B. PTS-1 sample, free PRP



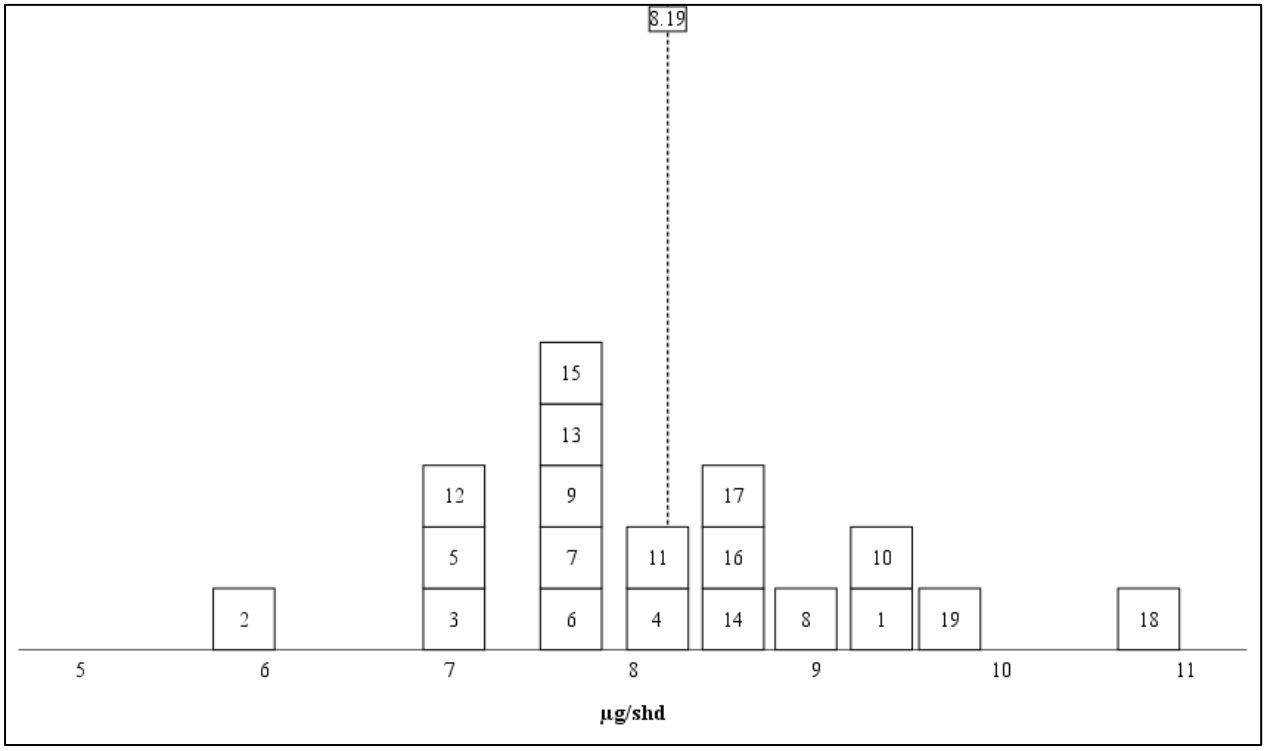
4C. PTS-2 sample, total PRP



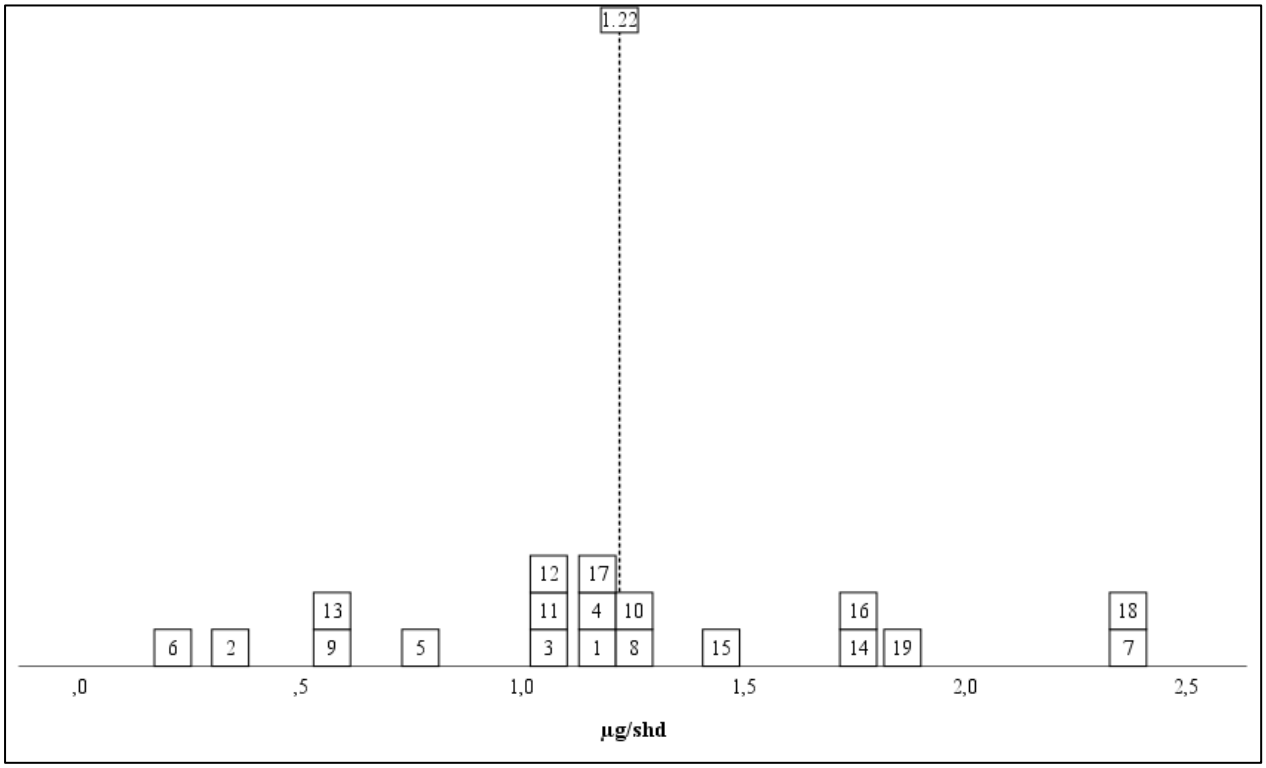
4D. PTS-2 sample, free PRP



4E. PTS-3 sample, total PRP

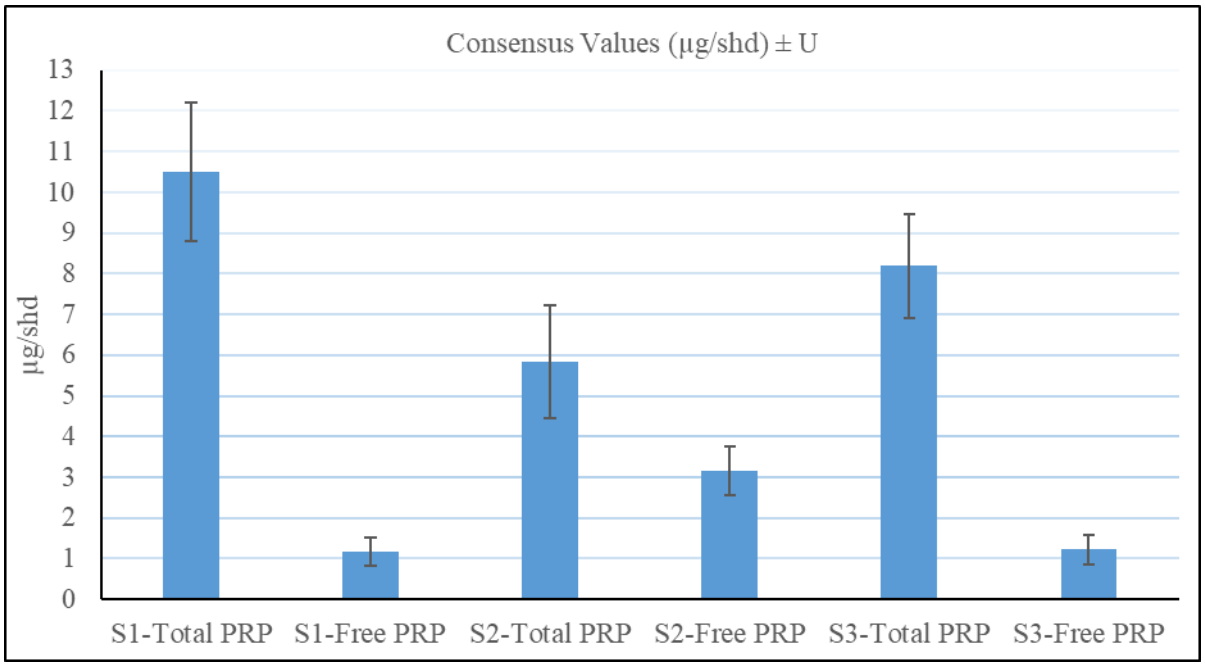


4F. PTS-3 sample, free PRP



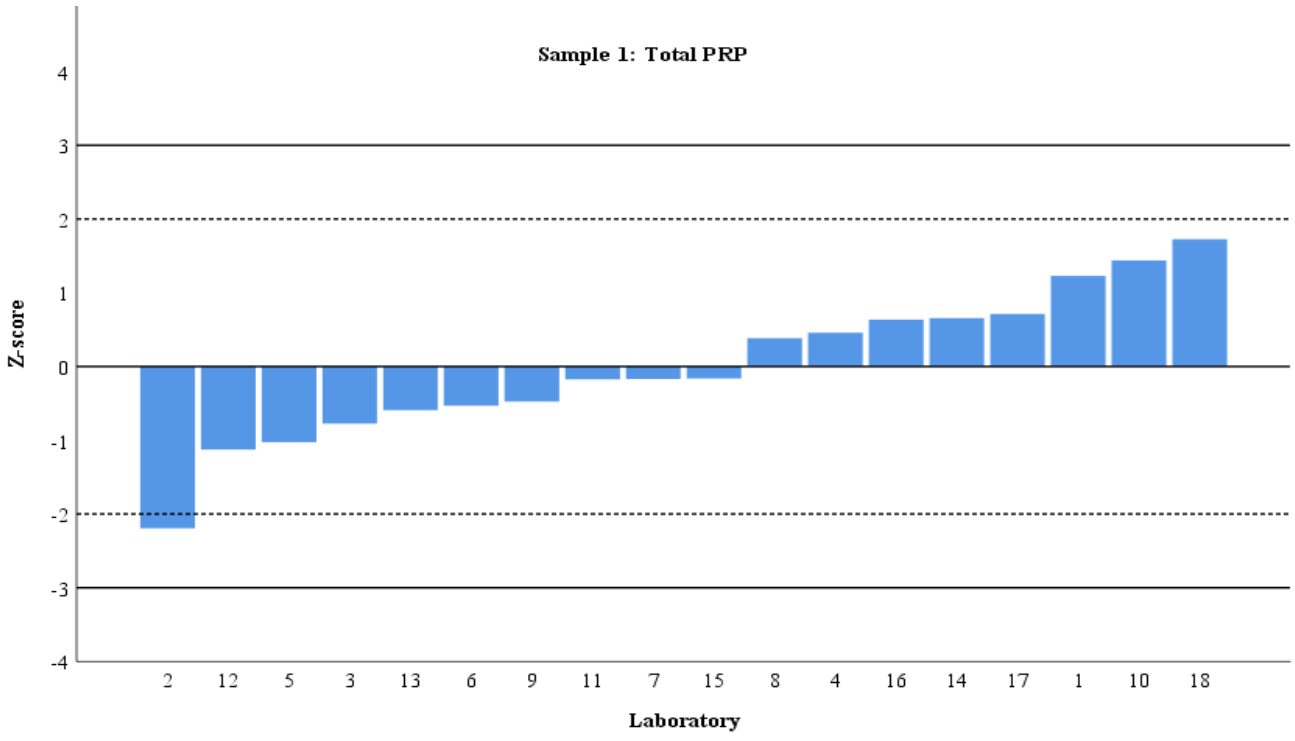
The vertical lines (Figures 4 A-F) represent the consensus values; numbers in the boxes indicate the code number of the laboratory.

Figure 5. Consensus Values for the six parameters; the error bars represent the Expanded Uncertainties (U; k=2)

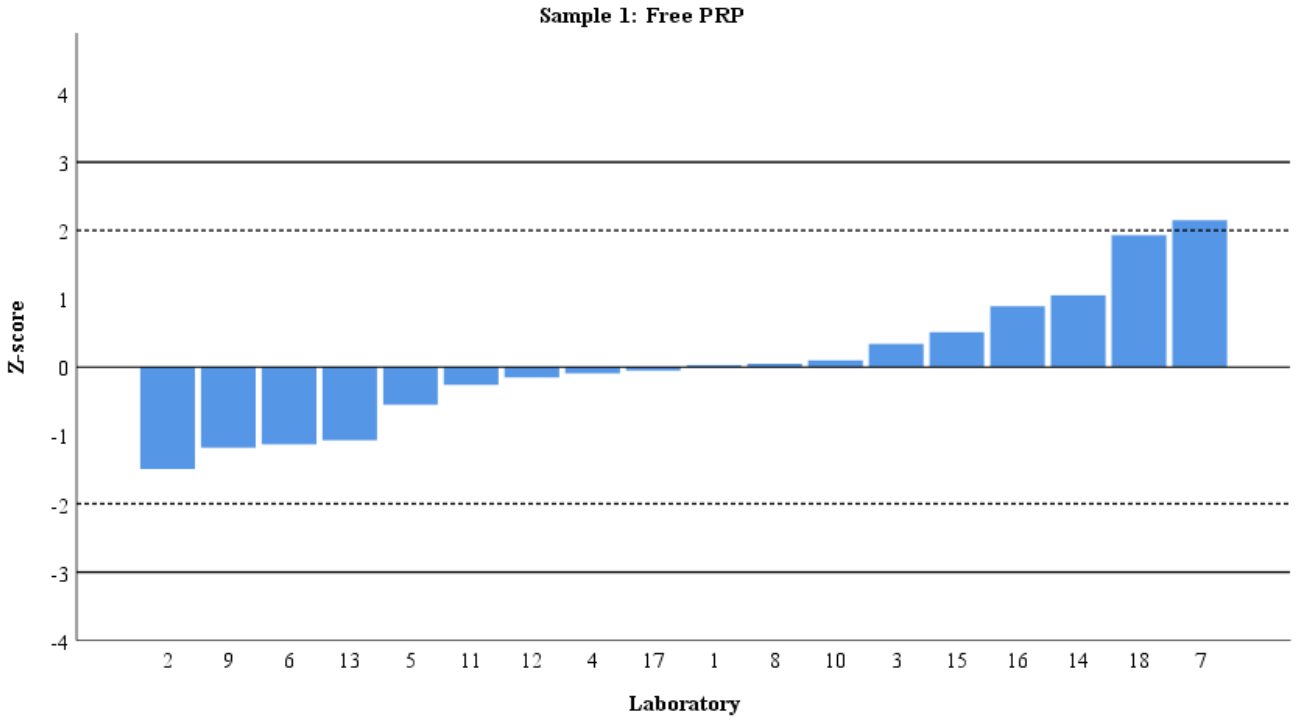


Figures 6 A-F. "Size-Ordered" Histogram Plots of Z-scores

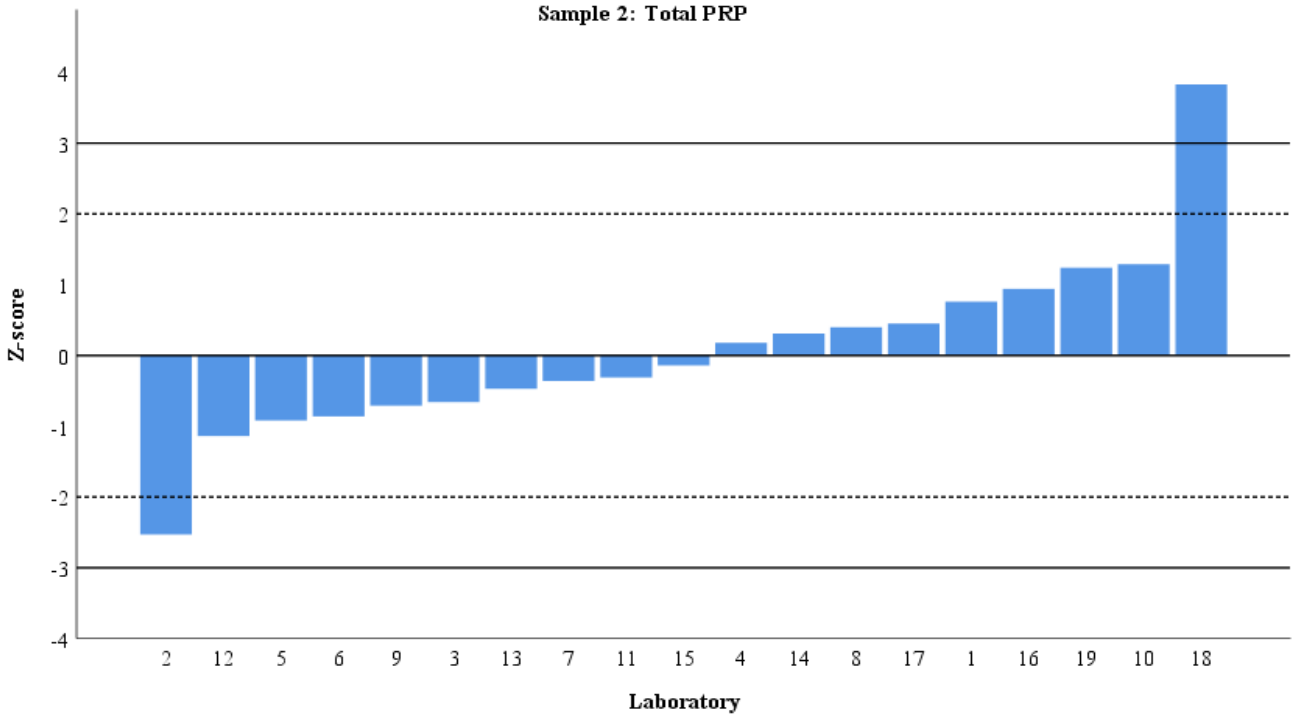
A)



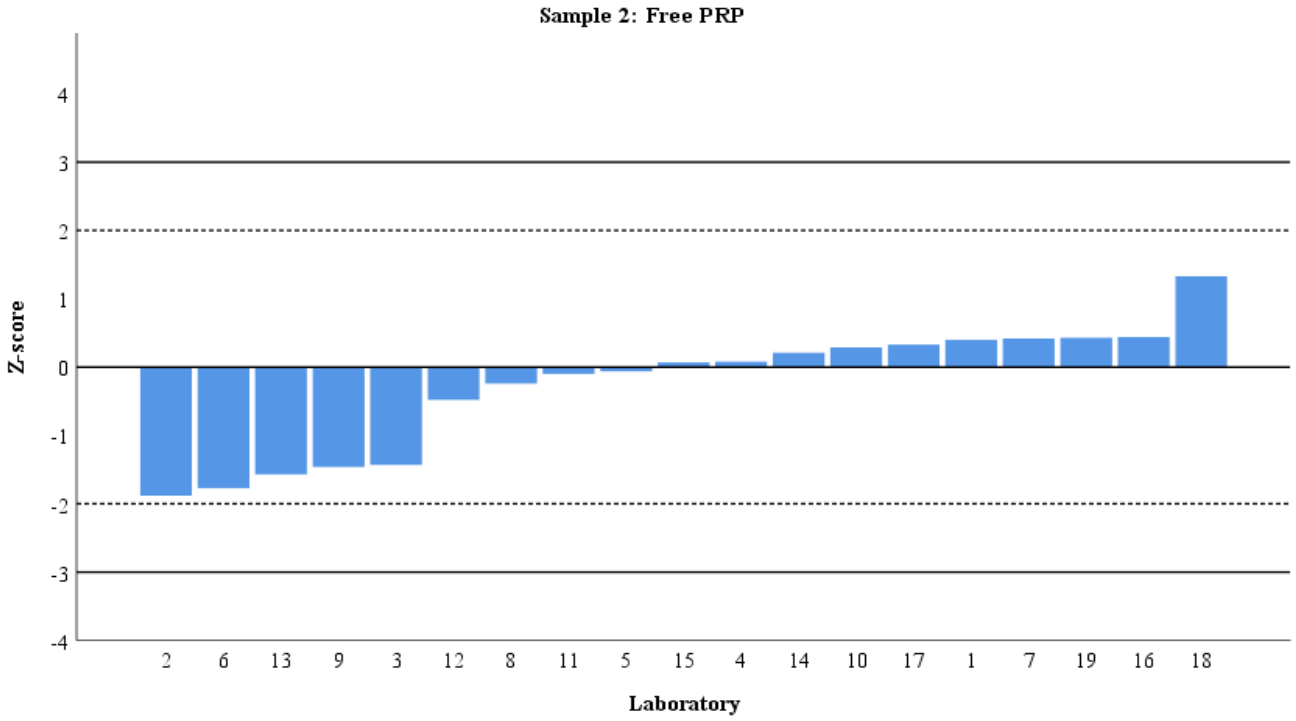
B)


Sample 1: Free PRP
Z-score
4
3
2
1
0
-1
-2
-3
-4
2 9 6 13 5 11 12 4 17 1 8 10 3 15 16 14 18 7
Laboratory

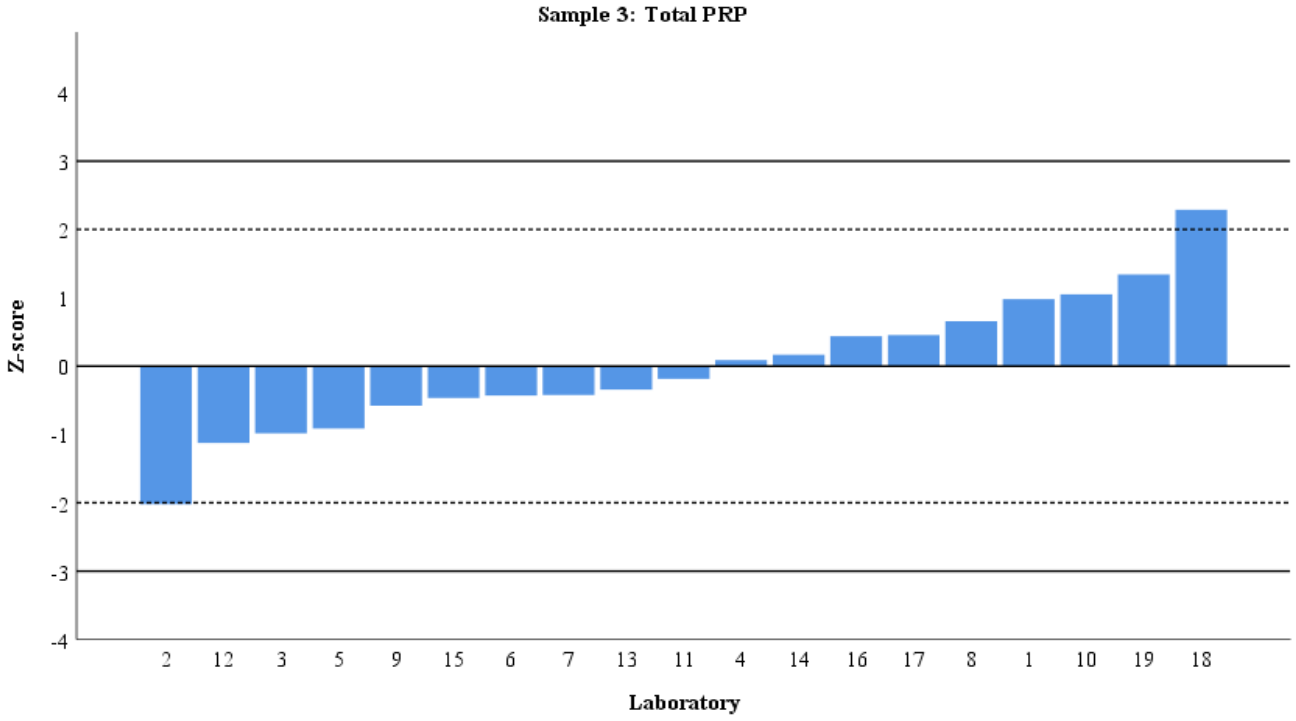
C)


Sample 2: Total PRP
Z-score
4
3
2
1
0
-1
-2
-3
-4
2 12 5 6 9 3 13 7 11 15 4 14 8 17 1 16 19 10 18
Laboratory

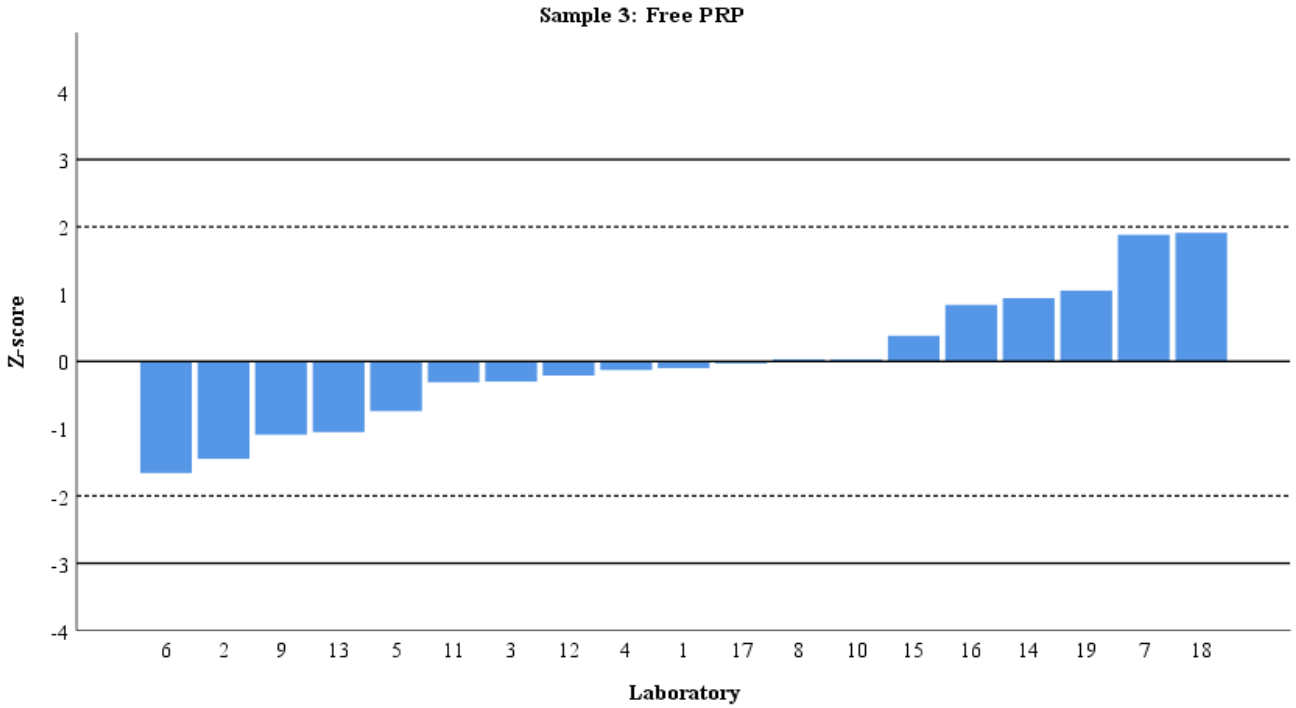
D)


Sample 2: Free PRP
Z-score
4
3
2
1
0
-1
-2
-3
-4
2 6 13 9 3 12 8 11 5 15 4 14 10 17 1 7 19 16 18
Laboratory

E)


Sample 3: Total PRP
Z-score
4
3
2
1
0
-1
-2
-3
-4
2 12 3 5 9 15 6 7 13 11 4 14 16 17 8 1 10 19 18
Laboratory

F)



The horizontal central line (at Z-score = 0) corresponds to the consensus values; the dotted lines represents the range from -2 to +2 and is considered a satisfactory Z-score. The solid line represents the range from -3 to +3 and is considered a questionable Z-score. A Z-score is considered unsatisfactory if it is outside the range from -3 to +3.

Figure 7. Youden Plot for PTS-1 sample, in µg/shd

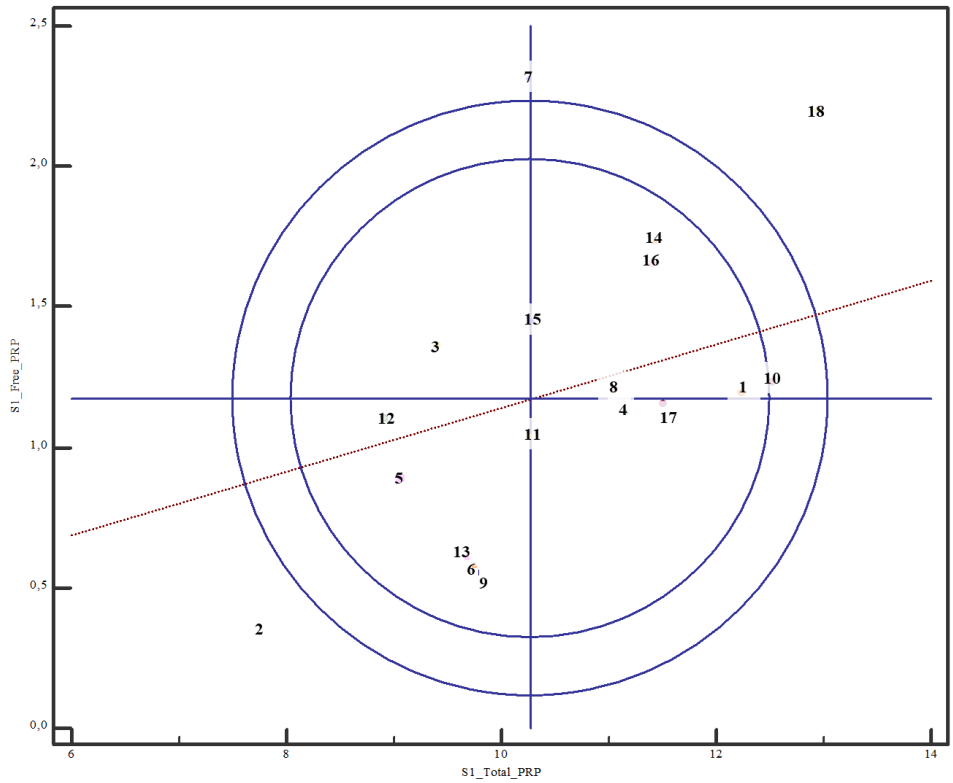


Figure 8. Youden Plot for PTS-2 sample, in µg/shd

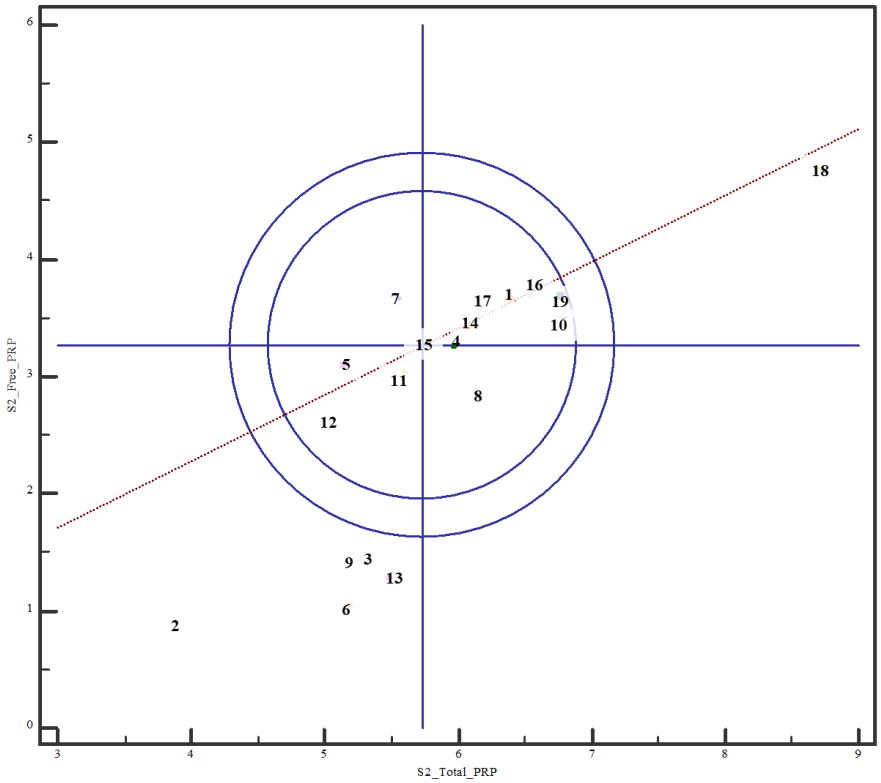
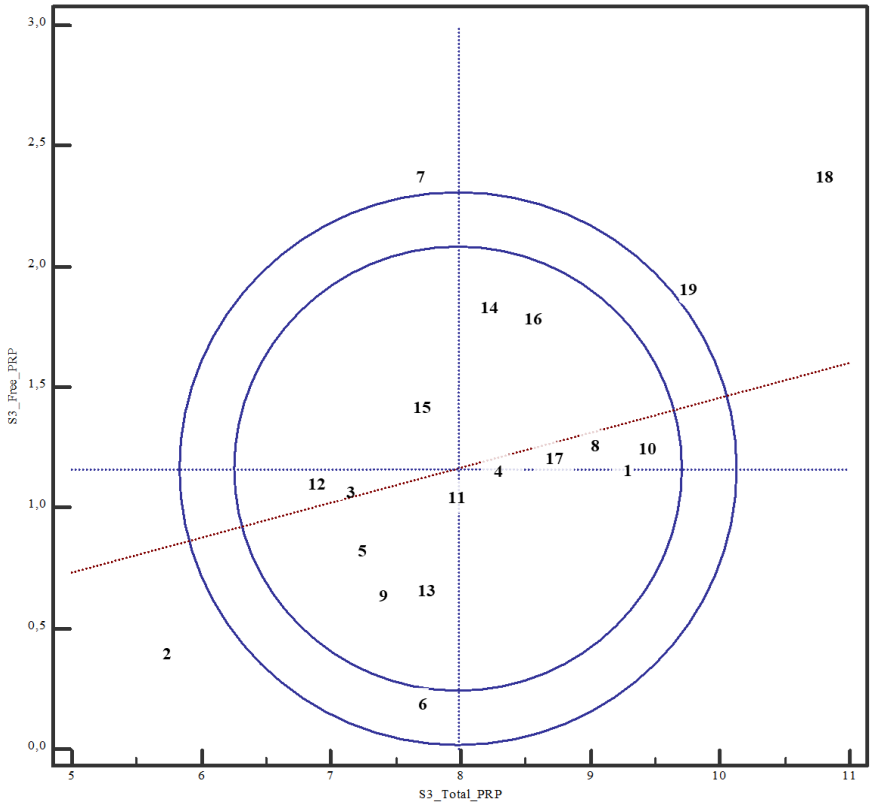


Figure 9. Youden Plot for PTS-3 sample, in µg/shd



Note: Numbers in the graphs indicate the code number of the laboratory.

***