Inter-laboratory collaborative trial of real-time PCR method phase 2: results

3.1 Results returns and exceptions

The deadline for results returns was extended to 15 July 2022, to allow some laboratories to submit results. One laboratory was sent a reminder two days before the deadline. All 15 laboratories returned results by the extended deadline.

One laboratory (006) reported difficulty with some of the analyses and requested additional reagents so that they could complete the work. A new set of reagents (but not test samples) was dispatched to this laboratory.

One laboratory (011) reported a delayed delivery of the original samples and that the samples eventually arrived at ambient temperature, having lost all the dry ice. Since it was anticipated that a re-send of samples would encounter a similar transit time, the laboratory was advised to continue with the analysis of the original samples.

Additional comments were received at the time of results submission from two laboratories. These were:

Lab 006; The 30 DNA samples and reagents/consumables were delivered with 2 ice packs on 20.06.2022. However, all items were at room temperature upon arrival.

Lab 014; Work carried out using MA0201 practical instructions. Pork Plate 3 was repeated using newly opened Mastermix as the first run failed on PCR Efficiency.

No additional modification to the data analysis was necessary as a result of these comments.

3.2 Results and Data analysis

Results were downloaded from the database in csv format and imported into MS Excel. Data were filtered by sample type-replicate combinations against the known sample identifier code in Table 3. A template software tool (MS Excel) was used to analyse the data, one spreadsheet per each of the 15 sample types. The data analysis follows that of the IUPAC Protocol [9]. The following steps were used in sequence to identify non-valid data and their subsequent exclusion from the calculations:

- non-detects (both replicates excluded if only one replicate was a non-detect)
- Cochran's outlier (where the variance of the replicates is exceptionally large)
- Grubb's outlier (where both replicates are statistical outliers)
- secondary Cochran's outlier (if further identified following initial outlier removal)

The number n of valid paired data points for the statistical output was never less than 12. A majority of the data sets (9 of 15) had n = 14, one dataset had n = 15, and five data sets had n =

12.

Initial observations of the data showed two unusual occurrences. Lab 010 had submitted results that were almost exactly correspondent to the intended spike levels of the samples. Lab 012 results were mostly outliers (in fact, only two of the 15 sample types were not outliers). Both laboratories were contacted about their results with no indication of the nature of the problem, so as not to affect any bias. Lab 010 confirmed that their results were only available to a single significant figure, hence there was no reason to exclude their results from the data analysis. Lab 012 confirmed that they had followed the instructions exactly. Since their results were mostly already excluded as statistical outliers, no further action was necessary.

All raw data are provided in Appendix 2. The associated Youden plots are presented in Appendix 3, graphically showing the outlier data. The associated mean and range plots are presented in Appendix 4.
Appendix 2 additionally provides the analysis performance data per laboratory for each plate assay (calibration coefficient of determination R2 and PCR efficiency).

The principal outputs of the data analysis are the repeatability and reproducibility precisions, and the critical differences. In an analytical chemistry collaborative trial in the food sector, the acceptability of the observed precisions would ordinarily be measured against the precision predicted by the Horwitz equation. Thus, a Horwitz ratio (HorRat) of observed standard deviation to Horwitz-predicted standard deviation between 0.5 and 2 would be considered acceptable. In the case of this study, the use of the Horwitz equation is not appropriate, since the method in question is an amplification technique and the sample types are extracted DNA (so the whole method is not the subject of the trial). In the place of the HorRat, this study substitutes the 'target ratio' with the same indicator limits of 0.5 to 2, but the Horwitz-derived standard deviation is replaced with the relative standard deviation (RSD) of 25% (EURL-GMFF), which was verified as appropriate from the original method validation study [5, 6]. The critical difference value is the difference in result at 95% confidence at which a sample may be out of specification at a legal limit.
The summary data are presented in Table 4.

Table 4, principal outputs of the collaborative trial statistics.

Horse in processed beef

Data type	Result	Result	Result	Result	Result
Intended level, %w/w	0.1	0.5	1	3	10
Mean, %w/w	0.175	0.823	1.73	3.30	10.5
n	15	12	14	12	14
RSDr, %	19.9	10.4	6.90	5.77	10.1
Target ratio r	1.21	0.630	0.418	0.350	0.611
RSDR, %	34.9	21.4	21.1	11.0	12.5
Target ratio R	1.39	0.857	0.844	0.441	0.501
CD95	0.11	0.33	0.70	0.67	2.14

Pork in processed beef

Data type	Result	Result	Result	Result	Result
Intended level, %w/w	0.1	0.5	1	3	10
Mean, %w/w	0.133	0.617	1.38	2.44	7.35
n	14	14	14	12	14

Data type	Result	Result	Result	Result	Result
RSDr, %	16.2	11.9	16.3	6.96	10.8
Target ratio r	0.983	0.720	0.988	0.422	0.654
RSDR, %	32.6	19.9	28.4	17.5	17.3
Target ratio R	1.31	0.798	1.13	0.700	0.691
CD95	0.08	0.22	0.71	0.81	2.26

Pork in raw beef

Data type	Result	Result	Result	Result	Result
Intended level, %w/w	0.1	0.5	1	3	10
Mean, %w/w	0.105	0.619	1.26	3.21	10.2
n	14	14	14	12	14
RSDr, %	24.0	21.0	19.2	8.60	8.76
Target ratio r	1.45	1.28	1.16	0.521	0.531
RSDR, %	35.8	25.4	22.7	17.4	9.27
Target ratio R	1.43	1.01	0.908	0.696	0.371
CD95	0.07	0.25	0.45	1.04	1.39

The intended level is % w/w of raw meats combined prior to extraction. n is the number of paired data points following outlier or non-compliant data removal. RSDr is relative standard deviation of repeatability r. RSDR is relative standard deviation of reproducibility R. CD95 is the critical difference at 95% confidence. Target ratio is observed RSD to target of 25%.

The repeatability precision r target ratio varies from 0.350 to 1.45 and is less than 0.5 in only three instances (horse in processed beef at nominal 1% and 3% levels, and pork in processed beef at nominal 3% level). Clearly, no target ratio r exceeds the upper limit of 2. The repeatability precision is therefore generally compliant [9] with a target RSDr of 25% and, in 11 of 15 sample types, the target ratio r is less than 1.0, indicating a high degree of repeatability precision. Furthermore, the observed RSDr is never more than 24.0% at the lowest level of 0.1% w/w, which is in keeping with the minimum performance requirement of 25% as laid out by ENGL [10].

The reproducibility precision R target ratio varies from 0.371 to 1.43 and is less than 0.5 in only two instances (horse in processed beef at nominal 3% level, and pork in raw beef at nominal 10% level). No target ratio R exceeds the upper limit of 2. The reproducibility precision is therefore generally compliant [9] with a target RSDR of 25% and, in 10 of 15 sample types, the target ratio R is less than 1.0, indicating a high degree of reproducibility precision. The observed RSDR ranges between 32.6% and 35.8% at the 0.1%w/w level and is never more than 28.4% at the higher preparation levels. This is in keeping with the minimum performance requirement of 35% for reproducibility as laid out by ENGL [10] over the whole dynamic range and is less than the suggested 50% acceptable RSDR at levels less than 0.2% w/w.

The 0.1%, 0.5% and 1% levels were prepared from the verified 3% preparations. Section 2.2 describes the mitigation actions taken to ensure correct preparation of the 3% levels prior to preparation of the lower levels and the additional verification steps taken. The data in Table 4 show the upper maximum bias of the 3% w/w levels to be 10% relative (horse in processed beef and pork in raw beef). This would not be sufficient to account for the potential bias observed in the lower prepared levels, i.e. the method bias is not due to propagation of inaccurate 3% w/w preparation. The bias observed in pork in processed beef at the 3% w/w level is -0.56% w/w, whereas the mean observed levels for the lower preparations are apparently positively biased. The observed method bias for pork in processed beef at the lower levels is not attributable to propagation of inaccurate 3% w/w preparation.

As part of the serial preparation process, the 3% mixture underwent three times more mincing than the 10% mixture; the lower percentage mixture also underwent three times further mincing to ensure homogeneous mixtures. For the processed matrices, the pork was cooked first; initial trial and error of the cooking process for the 10% and 3% mixtures may have caused more DNA degradation than the subsequent processed samples (although the beef DNA would have similarly been affected, so the ratio should be unaltered to this degree). One further possibility is that the beef overall underwent more mincing than the contaminant meats, so the apparent positive bias might in fact be proportionally greater beef DNA degradation. There is also the potential for differences in extractability of DNA from different species which could be the subject of a different study in which beef is the contaminant in pork (for example).

Further analysis was undertaken (data not shown) of the trend between the mean results plotted against the intended target levels, for both the collaborative trial data and the homogeneity data (linear unweighted determination of slope and intercept). Across the full collaborative trial calibration range, the intercept is about 0.3% w/w for processed matrices and 0.1% w/w for the raw pork matrix. The intercept is negligible for the lower concentration range 0.1-1% w/w (but with a more positive slope), therefore the observed break point is between the 1% and 3% levels. The homogeneity data trend is similar to the processed horse and processed pork at the lower part of the range but more similar to the full range for raw pork in raw beef. The data don't quite support the premise that the sample production method derived from the 3% target level material may cause an over-estimation of contamination due to the extended processing, since the change in slope should be observed between the 3% and 10% levels, not the 1% and 3% levels. However, the slope of the pork in processed beef being <1 supports the possibility of over-processing of this initial material at the higher levels. This is particularly noticeable in the homogeneity data for this matrix where there is a clear distinction between the top two levels and lower three levels. In real-world samples, the propagation effects observed here would not be expected to occur, hence the apparent positive bias may not be attributable to the PCR method nor to the ability of the laboratories in the trial. However, this is difficult to demonstrate at low levels unless the method of production of the test materials is radically altered which, in turn, may compromise their homogeneity.

The critical difference value (CD95) is the difference in result at 95% confidence at which a sample may be out of specification at a legal limit, i.e. whether the measurement uncertainty of a result may affect the interpretation of that result. At levels between 0.1% and 1%, the CD95 values are in keeping with these prepared levels for potential contamination detection. At the higher levels of 3% and 10%, the CD95 values are lower, as might be expected of adulteration levels. The principal discrepancy observed in the data is at the 1% target level in processed matrices, in which the CD95 values are 0.70 and 0.71 (horse and pork, respectively). For assumed regulatory levels between 0.5% and 1% there would be expected to be some potential for overlap of analytical results due to measurement uncertainty. However, since the CD95 values in the same processed matrices at the 3% levels are not too dissimilar (0.67 and 0.81, horse and pork, respectively) there remains the possibility that CD95 values are anomalies worth further investigation on a larger scale. Ultimately, further interpretation of the CD95 values would depend on decisions of regulatory limits in use.

The performance data of coefficient of determination R^2 and PCR efficiency are mostly within the acceptance criteria of the method SOP, i.e. R^2 ?0.98 and efficiency within the range 85 – 115%. One assay for each of two laboratories (9 and 14) had R^2 <0.98 and laboratory 15 had three assays with R^2 <0.98 (one pork and two horse). Laboratory 15 had a higher rate of outlier data than the other laboratories (with the exception of laboratory 12). Since the outliers were detected, any compromised calibration from laboratory 15 has not significantly affected the outcome of the study. The PCR efficiency for laboratory 2 was slightly <85% for two assays and laboratory 12 had low PCR efficiency for all six assays, five of which were <85%. The calibration performance for laboratory 12 was very good so the low PCR efficiency doesn't correlate with the outlier data from this laboratory, which are very high, often by more than a factor of two.