

# Review of allergen analytical testing methodologies: Assessment of the variability of analysis and potential to inter-convert data

An expert consultation with Mr. Mark Sykes and Mr. Dominic Anderson on the variability of analysis between methods for the same allergen. The tables referenced in this section are appended in Appendix 2.

This section of the study concerns the variability of analysis between methods for the same allergen. Although regarding tree nut-allergens, much testing is progressing to LC-MS/MS methods, the majority of food testing laboratories world-wide continue to use ELISA kits which are specific not only to the allergen/matrix combination being tested but also specific to the design by the kit manufacturer. There is a fundamental lack of equivalence and reproducibility between different ELISA kit manufacturers (and sometimes between different types of kit for a single allergen by the same manufacturer). Batch-to-batch variation due to variations in kit-supplied buffers and calibration standards exists. Also, the antibodies used to detect the allergen and the composition of calibration standards differs between manufacturers for kits detecting the same allergen, leading to variations in PT data depending on the method applied. The potential to convert units of measurement in order to compare data between methods for the same allergen is discussed. The inter-conversion of units of measurement such as measurements of genomic sequence, peptide or single protein into a measurement of total food allergen is essential to evaluate and compare methods. Furthermore, there is a variability within the ELISA kit analysis (repeatability precision). This is evidenced by the data for homogeneity testing for proficiency test samples. This part of the project interrogates several years' worth of homogeneity data and proficiency test data for a range of allergens analysed by ELISA kit with a view to determining an objective measure of precision that is actually present in real-world data.

## 7.1. Background introduction

A proficiency test (PT) is the performance assessment of a laboratory by interlaboratory comparison. (FAPAS, 2023) The assessment is in the form of a z- score;

$$z = (X - X_{pt})/\sigma_p$$

X is the laboratory's result,  $X_{pt}$  is the assigned property value for the analysis in question,  $\sigma_p$  is the standard deviation for proficiency assessment.  $X_{pt}$  is usually derived from the consensus of participants' results. The successful operation of a PT depends on the test sample being sufficiently homogeneous between samples (such that each participant in the PT receives effectively the same sample). The assessment by z-score must be fit-for-purpose, defined by the

calculation of the assigned value  $X_{pt}$  and the application of an appropriate value for  $\delta p$ .

In the majority of PTs (of any kind in the food sector), there is no observable method dependency and no mandated method; participants in the PT are free to use their normal, routine method and all PT results are treated as equivalent. The principal exception is for allergens, where the vast majority of participants use ELISA kits and different ELISA kits provide significantly different results. (Owen et al., 2009; Sykes et al., 2012; Hamide et al., 2019)

In this situation, participants in a PT are still free to use their normal, routine ELISA method but their result must be submitted against the ELISA kit they have used. Results from the PT are then treated as groups corresponding to specific ELISA kits, with their separate assigned values.

Homogeneity data and assigned values are considered fit-for-purpose if they meet statistical criteria for acceptable precision, dependent on the value of  $\delta p$ . In allergens PTs, the value of  $\delta p$  is set as 25% relative to the assigned value. This relative standard deviation was derived as a result of examination of early Fapas® allergens PTs. (Owen et al., 2009)

Homogeneity data are acquired as a set of 20 data points; 10 samples analysed in duplicate. By analysing each sample in duplicate, this provides a measure of the repeatability precision of the analysis, prior to its use in assessing the sample-to-sample variance. An analysis which is imprecise might not provide sufficient evidence of acceptable homogeneity. The acceptable analytical precision is determined by the ratio  $s_{an}/\delta p$  where  $s_{an}$  is the observed analytical standard deviation. The critical value is 0.5, above which the analytical imprecision could compromise the power of the statistical test to detect inhomogeneity. (Fearn, T. et al., 2001)

The PT assigned values are derived from the consensus of results corresponding to each allergen/ELISA kit combination. The fitness-for-purpose of the consensus is determined by the ratio of the uncertainty  $u$  of the consensus to the value of  $\delta p$ , where the critical value is normally 0.3 (ISO, 2015) but, for practical purposes of rounding data, a critical value up to 0.35 is acceptable. (Fapas, 2023) A ratio above the critical value risks compromising the usefulness of the z-scores but assessments will sometimes be provided for information only where the critical value is slightly exceeded, in order to provide some guidance to participants.

This study against Section 7 of the project has investigated Fapas® allergens PTs for an up-to-date summary of the current state of allergens testing by ELISA kit.

## 7.2. PT Data set

Summary data from Fapas® allergens PTs were collated from January 2018 to November 2022. The PT data include the matrix, allergen, ELISA kits used, consensus assigned values  $X_{pt}$  per ELISA kit, uncertainty ( $u$ ) of the consensus and  $\delta p$  values. Summary data were also collated for these PTs from the homogeneity testing in the form of the mean value,  $\bar{x}$ , observed analytical standard deviation for repeatability ( $s_{an}$ ) and the ratio  $s_{an}/\delta p$ . The PT data in total summarized 14295 individual submitted results (although this includes qualitative results, so the total quantitative results would be approximately half this number, i.e. about 7000 individual submitted results). Earlier data sets do not include results not assigned to a specific ELISA kit, so the overall number of submitted results to the PTs will be in excess of 7000. The homogeneity data numbered 133 sets of data, of which each set comprises 20 individual data points, i.e. a total of 2660 individual data points. In addition, comments relating to the use of PCR methods were collated (25 comments).

NB. The data cover a period of about five years. During this period, some ELISA kits have been either discontinued, revised, replaced or entirely new kits have come on the market. The data

reflects as accurately as possible the ELISA kits that were contemporary at the time of the PT.

The majority of allergens PTs comprise two test materials, labelled A and B. Superficially, one will be 'blank' (i.e. not spiked with allergen) and the other spiked with allergenic ingredient. This adds challenge to the participants in correctly identifying the spiked sample. Some PTs, however, will have both samples spiked but at different levels. Some PTs contain only one test material (usually the processed sample matrices). Therefore, the data sets include PT references with A, B or no specific material reference.

Part-way through the five-year period of this data collection, a change was made to how participants could report results. The earlier PTs required participants to enter their result and then choose the ELISA kit from a drop-down menu. The later PTs required participants to enter their result directly against a list of ELISA kits. The change in reporting was an efficiency improvement in how Fapas® was able to handle the data. Hence, the earlier PTs list only the ELISA kits against which assigned values could be calculated but later PTs additionally list all ELISA kits, regardless of whether an assigned value could be calculated.

The median number of registrations in the PT data studied was 57.5, of which a median of 54 (94%) submitted results. Participants in PTs fail to submit results by the deadline for various reasons, including equipment failure, unexpected staff resource unavailable, business priorities coming before PT samples. For the PT data set in question, this includes the principal periods of the global coronavirus pandemic and, while most food testing laboratories continued operating as essential services, they would have been compromised in staff availability. The 94% of results submission for allergens PTs therefore represents a very high proportion.

The majority (53%) of participants were from the EU, 15% were UK laboratories and 32% were rest-of-world (comprising North America, South America, Middle East Africa, Asia-Pacific). The UK laboratories comprise official control laboratories and third-party testing laboratories.

Appendix 2 presents the list of PTs included in the study, their registration numbers and results returns, and the broad geographic location of laboratories.

### **7.3. Section 7b. Short review of applicable analysis of existing Fapas® proficiency testing data**

This part of the study benefits from the use of large volumes of Fapas® data which are not publicly available free of charge. The data from the last five years' worth of Fapas® proficiency test reports will be reviewed. The allergens proficiency tests attract between 30 and 150 participating laboratories per test, of which there are 25- 30 such tests per year. Participants are assessed on the basis of the correct detection of allergen (qualitatively) and on the accuracy of quantitation of the detected allergen. Results are assessed separately according to the ELISA kit (or alternative method) used. There are unequal sub-populations of laboratories using different ELISA kits, so the most popular kits are more likely to be performance assessed than the less popular kits (due to insufficient numbers of data points). For processed foods in particular, the less popular or available kits tend to be those most suited to denatured allergenic proteins. Although the majority of laboratories are routinely using ELISA kits, a few laboratories report results from LC-MS/MS or PCR methods. This section of the study describes the effect of the different populations of data and what that means for the interpretation of allergen analyses.

#### **7.3.1. PT Data analysis and interrogation**

### 7.3.1.1. Homogeneity data

The mean ratio  $s_{an}/\sigma_p$  was calculated for each allergen and an overall mean calculated (Appendix 2). The mean ratios varied between 0.277 (lupin) and 0.567 (?-LG), with an overall mean of 0.432. The critical value is 0.5 for the normal acceptability of repeatability precision. The mean ratios of two other analytes (gluten and milk) exceeded the critical ratio (at 0.516 for gluten and 0.549 for milk). Of the individual data sets, 48 (36%) exceeded the critical value of 0.5. Of the ?-LG data, 5 of 6 data sets exceeded the critical value (but mean concentrations tend to be low for ?-LG). Hence, the overall premise of using 25% RSD for  $\sigma_p$  appears to be appropriate in evaluating homogeneity data with regard to the risk of not detecting true inhomogeneity. However, the data are also clearly showing that the repeatability limit of ELISA kits has been reached. The mean ratio  $s_{an}/\sigma_p$  is only just being maintained (overall mean of 0.432 is just less than the critical value of 0.5), so the repeatability of ELISA kits has not improved. If the repeatability had improved, we would expect to observe a much lower  $s_{an}/\sigma_p$  ratio.

### 7.3.1.2. Proficiency Testing data

The PT data were separated by allergen determination corresponding to gluten, egg, milk, soya, tree nut, peanut and other (itself comprising celery, mustard, lupin, sesame, fish). The other category corresponds to more recently added PTs into the programme, so there are fewer data sets for these allergens. Some PTs combined more than one allergen.

The value of  $u/\sigma_p$  was calculated for PT data and summarised by allergen, with the mean and standard deviation calculated for the grouped data. The results of this analysis are shown in Appendix 2. The mean  $u/\sigma_p$  ranged between 0.228 and 0.328, so within the Fapas® critical value of 0.35. The standard deviations ranged between 0.109 and 0.208, not insignificant given the values of the means. Hence, the overall premise of using 25% RSD for  $\sigma_p$  appears to be appropriate in evaluating PT consensus assigned values with regard to the risk of compromising the usefulness of z-score performance assessments. However, the data are also clearly showing that the reproducibility limit of ELISA kits has been reached. If the reproducibility of ELISA kit use had improved, we would expect to observe much lower  $u/\sigma_p$  values.

The PT data were visually inspected for general trends and anomalies (data in Tables 4-10 (Appendix 2). Two trends become immediately obvious: the majority of results from R-Biopharm ELISA kits and the discrepancy of assigned values between different ELISA kits. The latter issue has been reported previously and the data in the current project simply demonstrate that the situation has not changed in more than 13 years. (Owen et al., 2009; Sykes et al., 2012; Hamide et al., 2019) It is recognised that some ELISA kits report against whole ingredient (e.g. peanut) and some against the protein component (e.g. peanut protein). There still remains a discrepancy in assigned values between kits that purport to report like-for-like. The former trend of over-representation by one ELISA kit manufacturer has been suspected for a long time (and known in individual Fapas® PT reports) but the data of the current project reinforces this view. In addition, it is also apparent in the data that, where there are multiple R-Biopharm ELISA kits represented in the PTs, one R- Biopharm kit will be distinctly more popular than others. For example, in gluten analysis, the R-Biopharm RIDASCREEN Gliadin (R7001) kit will have many more results submitted against it than the R-Biopharm RIDASCREEN Fast Gliadin (R7002) or R-Biopharm RIDASCREEN FAST Gliadin sensitive (R7051) kits. This is because the R-Biopharm RIDASCREEN Gliadin (R7001) kit, in combination with the Mendez Cocktail solution, is the Gold Standard method recommended by CODEX for the analysis of gluten (Lacorn, Dubois, et al., 2022b; R-Biopharm, 2022)

In addition to the issue of significantly different results being associated with different ELISA kit manufacturers, a further issue is evident in this data. This issue relates to the lack of comparability between ELISA kits from the same manufacturer. To provide one example, in PT

27316 (milk in infant soya formula) test material A, three R-Biopharm kits are represented: R-Biopharm RIDASCREEN Fast  $\gamma$ -Lactoglobulin (R4912), R-Biopharm RIDASCREEN Fast Casein (R4612) and R-Biopharm RIDASCREEN Fast Milk (R4652). The consensus assigned values were, respectively, 1.56 mg/kg, 31.0 mg/kg and 17.8 mg/kg. Clearly, the assigned value for casein should not be nearly 2x that of milk. This issue is particularly prevalent for egg and milk determinations.

### **7.3.1.3. PCR method comments**

The majority of laboratories represented in PTs continue to use ELISA as the primary method for allergens analysis. The relative lack of other methods is evident in that, of the data sets studied in this project, only 25 comments related to the use of other methods (24 using PCR and one using LC-MS/MS), out of thousands of PT results. Non-ELISA methods are used either as a primary method or as confirmation of an ELISA result or to verify the absence of an allergen that is not expected to be detected by ELISA. Most of the PCR kits were the Congen SureFood kit (19 responses), even though some comments refer to R-Biopharm as the kit, R- Biopharm is the distributor of the kit. Two PCR kits were described as 'in-house', two PCR kits were unknown and one kit was from ALScreen.