Wageningen Academic Publishers

Validation of a qualitative R5 dip-stick for gluten detection with a new mathematical-statistical approach

K.A. Scherf[1], S. Uhlig[2], K. Simon[2], K. Frost[3], P. Koehler[1*], T. Weiss[4] and M. Lacorn[4]

[1]Deutsche Forschungsanstalt für Lebensmittelchemie, Leibniz Institut, Lise-Meitner-Straße 34, 85354 Freising, Germany; [2]QuoData GmbH, Obere Hauptstraße 36, 85354 Freising, Germany; [3]QuoData GmbH, Prellerstraße 14, 01309 Dresden, Germany; [4]R-Biopharm AG, An der neuen Bergstraße 17, 64297 Darmstadt, Germany; peter.koehler@tum.de

Received: 23 November 2015 / Accepted: 1 March 2016

RESEARCH ARTICLE

Abstract

Specific and sensitive analytical methods are needed to guarantee the safety of gluten-free products for coeliac disease patients at the threshold of 20 mg gluten/kg. Immuno-chromatographic assays, usually available in dip-stick or lateral-flow format, provide rapid qualitative results indicating the presence or absence of gluten. In September 2015, an AOAC expert review panel accepted the qualitative immuno-chromatographic dip-stick system RIDA®QUICK Gliadin as an AOAC Official Method℠ of Analysis First Action 2015.16 for detection of gluten in corn-based food products. One part of the required validation was an international collaborative study with 18 participants who analysed 8 different processed and non-processed samples with different gluten concentrations. Since the official statistical evaluation at AOAC does not provide a calculation of: (1) the concentration and its prediction interval from which a user expects a positive result; and (2) what a negative result means, a new statistical approach was used. The method for modelling the probability of detection (POD) curve and confidence intervals of the LOD95 used a generalised linear mixed effects model together with a 4-parameter sigmoid curve to describe the variability of the POD curves for randomly chosen laboratories beyond those that participated in the study. The results confirmed that, in case of a positive result, a non-processed sample contains a minimum of 4.54 mg gluten/kg (prediction range 3.59-5.74 mg/kg) while a processed sample contains more than 6.29 mg gluten/kg (prediction range 3.88-10.20 mg/kg). Therefore, the method is able to identify negative samples clearly below the threshold of 20 mg gluten/kg which is a prerequisite for gluten testing in food.

Keywords: AOAC Appendix N, gluten, R5, qualitative testing, statistical evaluation

1. Introduction

The consumption of wheat, rye or barley may lead to gluten-related disorders such as wheat allergy, non-coeliac gluten sensitivity and coeliac disease (CD) with estimated prevalences of 0.5-4%, 1-6% and 1% in Caucasian populations, respectively (Lionetti *et al.*, 2015; Scherf *et al.*, 2016; Schuppan *et al.*, 2015; Sicherer and Sampson, 2014). CD, one of the most common chronic digestive disorders, may be defined as a 'small intestinal immune-mediated enteropathy precipitated by exposure to dietary gluten in genetically predisposed individuals' (Ludvigsson *et al.*, 2013). In the context of CD, gluten is the protein fraction from wheat, rye, barley, oats or their crossbred varieties and derivatives thereof, to which some persons are intolerant and that is insoluble in water and 0.5 mol/l NaCl (Codex Standard 118-1979; FAO/WHO, 2015). Gluten proteins are the storage proteins of cereal grains and can be divided into the alcohol-soluble prolamin fraction and the alcohol-insoluble glutelin fraction, which is only soluble after addition of reducing and disaggregating agents. The prolamin content of gluten is generally taken as 50% (Codex Standard 118-1979; FAO/WHO, 2015). Although oats is part of the Codex definition of gluten, this crop is considered safe for the vast majority of persons intolerant to gluten, if it is not contaminated with other gluten-containing cereals (Thompson, 2003). Despite extensive research into alternative therapies, the only known effective

treatment for CD is a lifelong gluten-free diet (Plugis and Khosla, 2015). CD patients are advised not to exceed a total daily intake of 20 mg gluten and this level is reflected by the regulatory threshold of 20 mg gluten/kg for making a gluten-free claim on a product (EC, 2009; FAO/WHO, 2015; FDA, 2013; Health Canada, 2012).

Specific and sensitive analytical methods are necessary to ensure quality control and compliance testing and immunochemical methods are most commonly used in routine analysis (Scherf and Poms, 2016). An example is described in Bugyi *et al.* (2013). The sandwich ELISA based on the R5 monoclonal antibody (Valdes *et al.*, 2003) for the detection of intact gluten was laid down as a Codex Alimentarius type 1 method for the analysis of gluten (Codex Standard 234-1999; FAO/WHO, 2013) and has been adopted as official or approved method both by the AOAC International (Immer and Haas-Lauterbach, 2012) and the AACC International (Koehler *et al.*, 2013a). A competitive R5 ELISA for the detection of partially hydrolysed gluten has also been validated by both standardisation organisations (Koehler *et al.*, 2013b; Lacorn and Weiss, 2015). Raised against rye ω-secalins, the R5 antibody primarily recognises the epitope QQPFP, which is present in wheat gliadins, rye secalins and barley hordeins and part of many CD-toxic or -immunogenic peptides (Kahlenberg *et al.*, 2006; Osman *et al.*, 2001; Tye-Din *et al.*, 2010).

The immuno-chromatographic dip-stick system RIDA®QUICK Gliadin based on the R5 antibody provides rapid, qualitative results indicating the presence or absence of gluten. It may be used as a swab test of potentially contaminated surfaces and production sites and to check for gluten contamination of raw or processed materials. The samples are extracted either by 60% ethanol for raw materials or by Cocktail (patented; WO/2002/092633) (Garcia *et al.*, 2005) for processed foods. The resulting supernatant is analysed within 5 min followed by visual evaluation of the test and control lines on the dip-stick. The system was developed to detect gluten clearly below the threshold of 20 mg/kg and shows no high-dose hook effect.

To validate the suitability of the R5 dip-stick for qualitative gluten detection in raw and processed foods an international collaborative study was set up with 18 participating laboratories. As the first study of its kind in the field of food analysis, the general outline followed the AOAC guidelines for validation of qualitative binary chemistry methods (AOAC, 2013, Appendix N). Because this guideline only describes a 'reproducibility of probability of detection (POD) values' and not the actual reproducibility of concentrations, this paper uses a new statistical approach published by Uhlig *et al.* (2015a). This new approach enables the prediction of a range of concentrations where the POD is 0.95 and thus equivalent to a limit of detection (LOD) for quantitative methods. The term LOD95 is used for this concentration throughout the entire paper.

2. Materials and methods

Study design

The result of a (true) qualitative method is always either a 'yes' or 'no' answer. The main challenge for the validation of such a method is the low amount of information per sample after analysis compared to a quantitative method (AOAC, 2002, Appendix D; Sykes *et al.*, 2014), which always results in a number for the measurement signal of each sample. Therefore, a high number of replicates have to be analysed for qualitative methods. From these results, a POD is calculated for each sample as the ratio between the number of positive results and the total number of replicates per concentration. For a blank sample the POD should be zero while a POD=1 is expected for a positive sample. Within a small range of concentrations, results between 0 and 1 will be obtained, e.g. POD=0.6 for 6 positive samples out of 10 analyses samples. This concentration range can be described as the fractional range of the method.

Before analysing the blind-coded samples, each participant was asked to perform checks for contamination of surfaces and buffers in their laboratories and to become familiar with the test method by using commercially available check samples with known concentrations. Both procedures are necessary because the qualitative nature of the result makes a later check for sample mix-up or improper performance during extraction or testing very difficult or even impossible. Additionally, a training video on how to accurately conduct the test procedure was provided to each participant.

The collaborative study was split into two parts (A and B) to prevent mix-up of samples and procedures due to different extraction procedures for processed (part B) and non-processed samples (part A). The total number of 40 samples (10 replicates for each concentration) per part was a compromise between the number of replicates and the number of concentration levels on the one hand and the number of samples which a participant could manage within an acceptable time on the other hand. This compromise was partly compensated for by the high number of 18 participants (see 'Acknowledgements'). Each blind-coded sample was extracted once and analysed according to the test kit instruction. In total, 80 samples had to be analysed by each laboratory. Each sample had to be marked positive or negative or invalid. In case of an invalid result (missing control line or incomplete target line), retesting of the sample was requested. All results obtained by visual inspection had to be recorded in a ready-to-use Excel sheet (Microsoft, Redmond, WA, USA). The final data from the laboratories were sent to the study coordinator.

Collaborative study samples

The following samples were prepared for part A of the collaborative study:
1. corn flour, containing gluten at 1.8 mg/kg;
2. corn flour, containing gluten at 4.8 mg/kg;
3. corn flour, containing gluten at 11.0 mg/kg;
4. corn flour, containing gluten at 18.8 mg/kg.

All concentrations reported in the manuscript are given as mg gluten/kg and were determined using the quantitative ELISA RIDASCREEN® Gliadin R7001 (R-Biopharm AG, Darmstadt, Germany) (AOAC Official Method℠ of Analysis First Action 2012.01, AACCI approved method 38-50.01, and type 1 method according to CODEX Standard 234-1999; AACCI, 2013; AOAC, 2012; FAO/WHO, 2013). This ELISA is also based on the R5 monoclonal antibody, which is the only antibody that was accepted by the Codex Alimentarius Commission in 2008 as part of the type 1 method. The corn flour samples 2 to 4 were prepared by mixing a naturally contaminated corn flour sample with the 'gluten-free' corn flour sample 1.

The following samples were prepared for part B of the collaborative study:
5. cookie (processed), containing gluten at 0.4 mg/kg;
6. corn snack (processed), containing gluten at 6.4 mg/kg;
7. corn snack (processed), containing gluten at 13.3 mg/kg;
8. corn snack (processed), containing gluten at 47.1 mg/kg.

Considerations for the choice of the gluten concentrations of the samples were that samples at the lower and higher side of the LOD should be provided and also clearly positive samples. In part A (ethanol extraction) samples with gluten concentrations of 1.8 and 4.8 mg/kg fulfilled the first requirement. Samples with 11 and 18.8 mg gluten/kg were expected to be tested positive by all laboratories. In part B (Cocktail extraction) differently diluted samples were present in the test tube (factor 4) compared to part A. Thus, the gluten concentrations 0.4 and 6.4 mg/kg were on the lower side of the LOD and the sample containing 13.3 mg gluten/kg was comparable to sample 2 of the ethanol extraction. Sample 8 was aimed as a clearly positive sample.

The processed snack samples 6 to 8 were prepared by mixing a snack sample containing wheat gluten (spiked at 100 mg/kg before processing) with a 'gluten-free' snack sample as described by Koehler *et al.* (2013a). The processed sample 5 was a gluten-free cookie from a local retailer. All materials were prepared by grinding in a knife mill and subsequently a disk mill to ensure all materials passed a 40 mesh screen. The complete sample was mixed for 2 h in a vertical mixer/flat beater, sieved through a 40 mesh screen using a vibratory sieve shaker and then mixed again. Samples were aliquoted into foil pouches at an amount of 0.7 g for processed samples and 2.8 g for non-processed samples and stored at 4 °C.

Homogeneity was tested using the quantitative sandwich ELISA RIDASCREEN® Gliadin (R7001; R-Biopharm AG). The determination of homogeneity was done according to the IUPAC recommendations for proficiency tests (Thompson *et al.*, 2006). The standard deviation s_p was derived from the Horwitz equation to calculate a deviation that is dependent on the concentration. According to this commonly used procedure all samples with gluten concentrations above the limit of quantification (LOQ) of the method type 1 ELISA system (samples 2 to 4 and samples 6 to 8) showed acceptable homogeneities. Samples 1 and 5 showed results clearly below the LOQ of the R5 ELISA system. The concentrations of these blank samples were estimated by extrapolation of the calibration curve (Koehler *et al.*, 2013a).

The samples were marked with a laboratory-specific letter (A to W), an 'E' for ethanol extraction (part A) or a 'C' for Cocktail (patented) extraction (part B) and a randomised number from 1 to 40. Each laboratory obtained its own coding (different randomised numbers for each laboratory).

Extraction for non-processed samples (part A)

One g of each sample was weighed in a vial and 10 ml of 60% (v+v) ethanol solution were added. Vials were mixed thoroughly for at least 30 s on a vortex mixer and centrifuged at 2,500×*g* (room temperature; 20-25 °C) for 10 min. Resulting supernatants were diluted by mixing 50 µl of supernatant with 500 µl sample diluent in test tubes (both contained in the test kit). Then, the diluted solutions were analysed by the dip-stick test.

Extraction for processed samples (part B)

Each sample was weighed at an amount of 0.25 g in a vial and 2.5 ml of Cocktail (R7006; R-Biopharm AG) was added. The vial was closed and mixed well (vortex) to suspend the sample. After incubation at 50 °C for 40 min in a water bath, the samples were cooled down to room temperature (20-25 °C) and 7.5 ml 80% (v+v) ethanol was added. The vials were closed and shaken up-side down by a rotator for 1 h at room temperature (20-25 °C). Afterwards the vials were centrifuged at 2,500×*g* at room temperature (20-25 °C) for 10 min. By diluting 50 µl supernatant with 500 µl sample diluent in the test tubes (both contained in the test kit), the extracts were ready for dip-stick testing.

Dip-stick procedure

The immuno-chromatographic dip-stick system RIDA®QUICK Gliadin (R7003; R-Biopharm AG) consists of different zones (Figure 1). Analyte in the sample solution

will be 'chromatographed' above the 'maximum line' and reacts with the R5-antibody coupled to a red latex microsphere. The 'maximum line' indicates the maximal liquid level of the diluted sample extract to the user. The 'result window' contains a small band of immobilised R5 antibody (red line after positive reaction) and a second line that turns blue when the reaction was valid. Results are read visually only. Generally, the higher the analyte level in the sample the stronger the red colour of the test band will be until the maximum intensity of the colour is reached.

The dip-stick was placed vertically into the test tube which was filled with the diluted sample extract. Afterwards the stick was taken out after exactly 5 min (±10 s). The result was evaluated using the evaluation card (part of the test kit). If two coloured bands (test band in red and control band in blue) were visible in the result window (Figure 1) after 5 min, the sample was positive for gluten. If only the blue control band was visible in the result window (Figure 1) after 5 min, the sample was negative for gluten. If no bands occurred after 5 min, the test was invalid and had to be repeated using a new dip-stick.

3. Results and discussion

Collaborative study results

All participants reported that all control samples were evaluated in the expected way and no contamination was observed in their laboratories. None of the participants reported an invalid result.

The resulting raw data sets are shown in Table 1 (ethanol extraction) and Table 2 (Cocktail (patented) extraction). Each participant performed ten replicates for each sample. For ethanol extraction, the results were remarkably consistent and 14 out of 18 laboratories showed no detection of the blank sample and detection of all contaminated samples. Besides these 14 laboratories, only one laboratory assigned two false positive blank samples. The remaining three laboratories found one false negative for the low concentration (4.8 mg gluten/kg) and just one laboratory found two false-negatives for the medium concentrated sample with a gluten concentration of 11.0 mg/kg. False-positive results can be explained by the fact that homogeneity testing for blank samples is challenging, because gluten contaminations can occur very infrequently at very low levels and are therefore not detectable by the quantitative ELISA method. At these low concentrations, some degree of inhomogeneity is possible and, therefore, a few false-positives (2 out of 180 samples) could be expected from this point of view.

When using the Cocktail (patented) extraction procedure, the extract was diluted fourfold higher compared to the ethanol extract. Consequently, it was not surprising that the low concentrated sample with a concentration of 6.4 mg gluten/kg (corresponding to 15 ng/ml in the Cocktail (patented) extract) showed a higher variation compared to the ethanol extraction of a sample with a concentration of 4.8 mg gluten/kg (corresponding to 44 ng/ml in ethanol extract). The study coordinator decided to exclude laboratory B for the statistical data treatment since it was obvious from the initial raw data (Excel sheet sent to the study coordinator) that a blank sample had been mixed up with a sample containing the highest concentration of gluten. Nevertheless, 9 out of 17 laboratories reported all blank samples as negative and all contaminated samples as positive. Only one laboratory found two false-positive results. This rate of 2 out of 170 is comparable to the rate of the ethanol extraction procedure. It is interesting to see that for the low concentrated sample (6.4 mg/kg) laboratories could be separated into two groups reporting either 70 up to 100% positive detection or only 0 to 10% positive results. This is one reason why an alternative statistical approach besides the AOAC Appendix N was necessary (AOAC, 2013). Because the variation between laboratory specific rates of detection was far from being normal (Figure 2 and 3), the alternative approach was calculated as described below. Statistics according to Appendix N

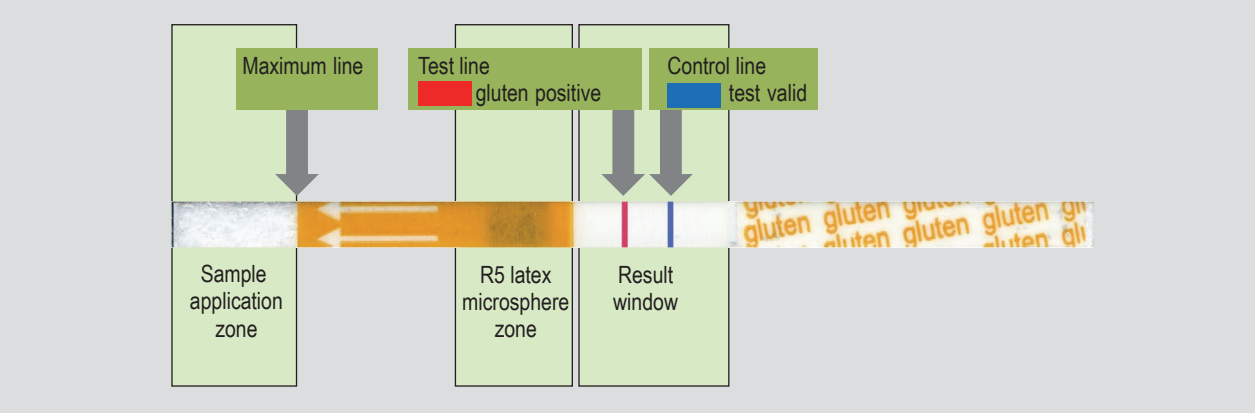

Figure 1. Schematic presentation of the dip-stick procedure and the subsequent interpretation of possible results.

Table 1. Numbers of positive samples detected in part A of the collaborative study using the R5 dip-stick after ethanol extraction (data by each of the 18 participating laboratories); each laboratory obtained 10 blinded replicates for each concentration level.

Laboratory code	Gluten (mg/kg)			
	1.8	4.8	11.0	18.8
	Sample 1 (negative)	Sample 2 (low)	Sample 3 (medium)	Sample 4 (high)
A	0	10	10	10
B	0	10	10	10
D	0	10	10	10
E	0	10	10	10
F	0	10	10	10
G	0	10	10	10
H	0	10	10	10
I	0	9	10	10
L	0	10	10	10
M	0	9	8	10
N	0	10	10	10
O	0	10	10	10
P	0	10	10	10
R	0	10	10	10
S	0	9	10	10
T	0	10	10	10
U	0	10	10	10
W	2	10	10	10

are claimed to be based on ISO 5725-2 (ISO, 1994), which requires approximate normal distribution (clause 1.4). In case of a bimodal distribution, as in the present case, the interpretation of results is difficult and not recommended.

A descriptive way to show the results for both parts of the collaborative study is given in Figure 2 (ethanol extraction) and Figure 3 (Cocktail (patented) extraction). In these figures, the POD is plotted for each sample. It has to be noted that only 10% increments are possible for the POD in these figures, because only 10 replicates were analysed by each participant. The area of the circles is equivalent to the number of laboratories indicated by the number next to the circles that reported this POD.

New statistical approach

In the course of the AOAC Official Method^{SM} of Analysis process (AOAC, 2015) only a basic statistical evaluation was performed which only dealt with POD values and their variability. From this kind of statistical evaluation it is not possible to calculate: (1) the concentration and its prediction interval from which a user expects a positive result; and (2) what a negative result means.

For this reason, a new statistical approach was used in this case to describe the POD by means of a mathematical-statistical model as a function of the concentration and to derive LOD95 concentrations (Uhlig *et al.,* 2015a). The method for modelling POD curves and prediction intervals of the LOD95 concentrations uses a generalised linear mixed effects model (GLMM) together with a 4-parameter sigmoid curve to describe the variability of the POD curves for randomly chosen laboratories. A 4-parameter curve fitting is often used for ELISA calibration curves. The 4-parameter sigmoid curve is given by:

$$POD_{ij} = \frac{A - D}{1 + (\chi_j / C\gamma_i)^B} + D$$

where i denotes the laboratory (i=1, …, 18); χ_j denotes the concentration of level j (j=1, …, 4); A, B, C, and D are model parameters, which will be estimated globally (across all laboratories); γ_i denotes the random relative variation of the global parameter C due to laboratory i.

It is assumed that the parameters A (lowest possible value), B (slope), and D (highest possible value) are the same for all laboratories. The parameter $C\gamma_i$ describes the location of the individual inflection point of the curve for laboratory i; for A=0.00 and D=1.00 it corresponds to the concentration at which a POD of 50% is reached for this laboratory. This

Table 2. Numbers of positive samples detected in part B of the collaborative study using the R5 dip-stick after Cocktail (patented) extraction (data by each of the 18 participating laboratories); each laboratory obtained 10 blinded replicates for each concentration level.

Laboratory code	Gluten (mg/kg)			
	0.4	6.4	13.3	47.1
	Sample 5 (negative)	Sample 6 (low)	Sample 7 (medium)	Sample 8 (high)
A	2	7	10	10
B[1]	1	10	10	9
D	0	9	10	10
E	0	1	10	10
F	0	10	10	10
G	0	10	10	10
H	0	10	10	10
I	0	9	10	10
L	0	8	10	10
M	0	10	10	10
N	0	10	10	10
O	0	10	10	10
P	0	10	10	10
R	0	10	10	10
S	0	0	10	10
T	0	9	10	10
U	0	1	10	10
W	0	10	10	10

[1] Data set of laboratory B was not included in the statistical calculation, because two samples were apparently exchanged due to their direct vicinity during testing.

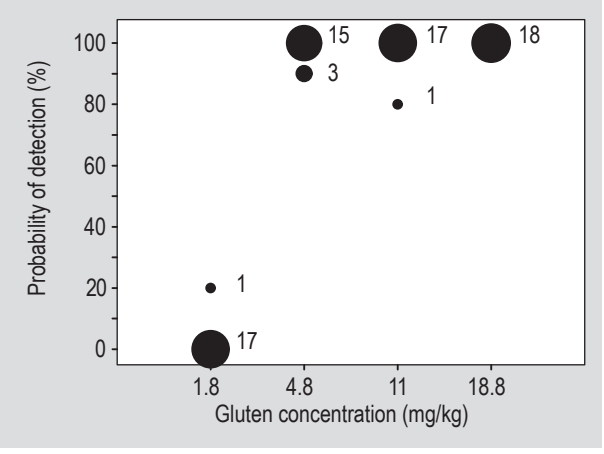

Figure 2. Probability of detection (POD) observed by each of 18 participating laboratories for samples extracted with ethanol (part A) between 1.8 and 18.8 mg gluten/kg gluten. Number stated at each circle means number of laboratories with the same POD. Areas of circles are proportional to number of laboratories.

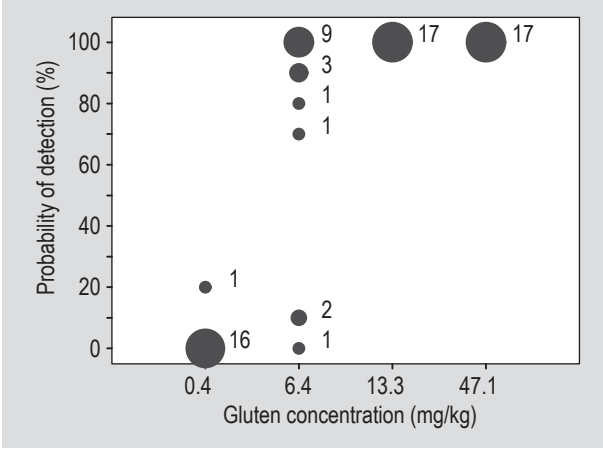

Figure 3. Probability of detection (POD) observed by each of 17 participating laboratories for samples extracted with Cocktail (patented) (part B) between 0.4 and 47.1 mg gluten/kg gluten. Number stated at each circle means number of laboratories with the same POD. Areas of circles are proportional to number of laboratories.

parameter is thus a direct measure of the performance of the specific laboratory. The parameter C corresponds to the performance of a typical laboratory. C is some kind of geometric mean of the laboratory specific values $C\gamma_i$ and

not the average over these values. The relative variation γ_i is modelled as a log-normally distributed random variable, i.e. $\ln(\gamma_i)$ is normally distributed with zero mean and standard deviation σ. This standard deviation describes the inherent variation of the laboratory performance in the hypothetical population. The approach is similar to the one described by Uhlig *et al.* (2015b) for PCR analysis, with the exception that another type of POD function is used, that does not take specific aspects of PCR amplification into account. The model reduces to a two parameter model when setting the starting probability A equal to 0 and the saturation probability D to 1. This would mean that there are no false detections in case of zero concentration and no false non-detections in case of very high concentrations.

From the data of part A (ethanol extraction), the common parameters were estimated as A=0.011, B=7.84, and D=0.989. The laboratory-specific parameter was estimated with a mean of C = 3.03 mg/kg. This yielded a 95% prediction interval for parameter $C\gamma_i$ of [2.39 mg/kg, 3.82 mg/kg] of a future laboratory i.

Figure 4 shows the POD curve of a laboratory with average performance (blue curve) together with the expected POD curves corresponding to laboratories at the lower (green) and upper (red) limits of the 95% prediction interval of parameter C. For example, a POD of 80% is reached by a laboratory with average performance at a concentration of about 3.6 mg/kg (blue curve), whereas a top-performing laboratory will reach this POD already at 2.9 mg/kg (green curve) and a bottom-performing laboratory will need a concentration of about 4.6 mg/kg (red curve).

To test the fit of the model, a Monte-Carlo simulation with 100,000 simulation runs was calculated. This approach is recommended by ISO for different purposes (ISO, 2015; ISO/IEC, 2008). In Figure 4, the black step-functions show the 95% prediction band for the rate of detection (ROD) assuming 10 replicates per laboratory. This prediction band was obtained by 100,000 random samples based on the calculated model. While the POD curves are continuous, the ROD graphs depend on the number of replicates, so that only RODs of 0, 0.1, 0.2, etc. are possible with 10 replicates. According to the definition of the prediction interval, about 95% of all data should be within the prediction band – otherwise the model would not be appropriate. About 5% of the data – no outliers – are expected to be out of the prediction band. Outliers were not identified or eliminated from the dataset. If more than 5% of the data are out of the prediction band, the model might be not appropriate.

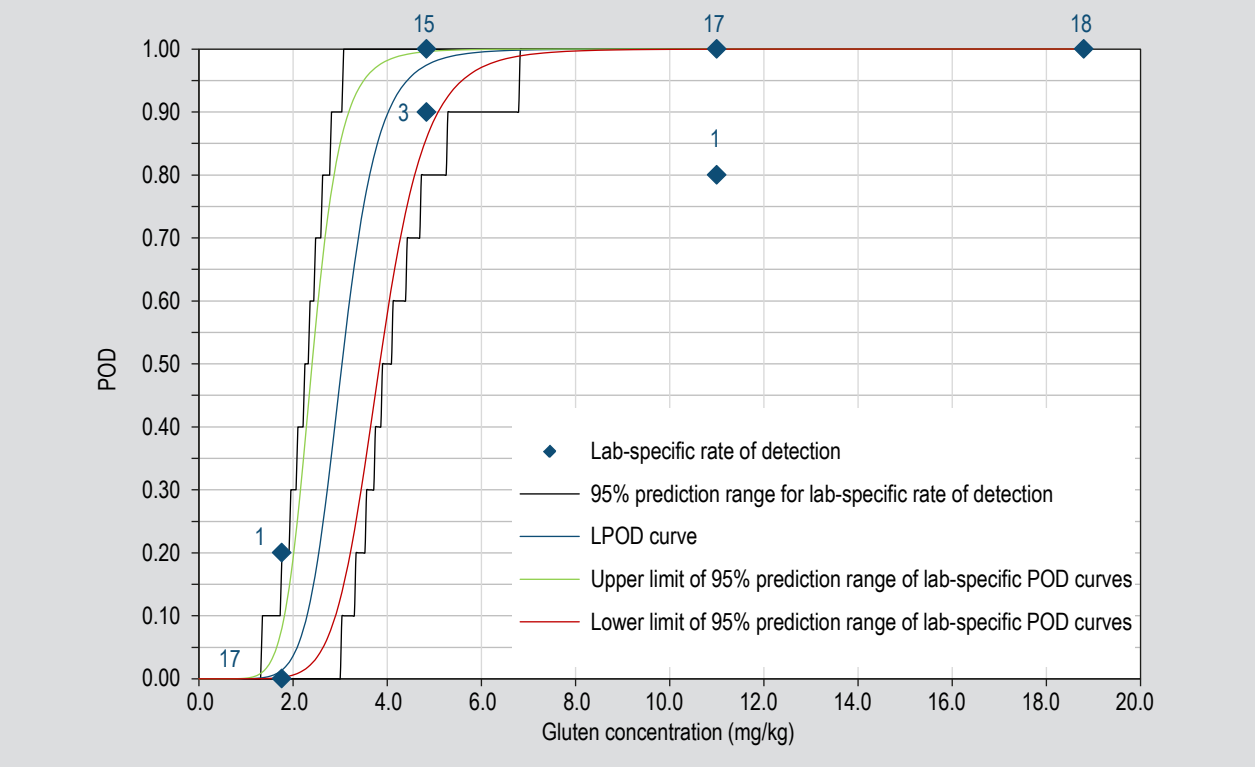

Figure 4. Performance statistics for part A (ethanol extraction): probability of detection (POD) curves for a laboratory with average performance (blue), top-performance (green), and bottom performance (red), calculated from the lower and upper limit of the 95% prediction interval for the parameter $C\gamma_i$. Numbers near the diamonds for the laboratory-specific detection rate indicate the number of laboratories. Black step functions show the results of the Monte-Carlo simulations.

From the GLMM (no calibration interval), the LOD95 (concentration at POD=0.95) was calculated as 4.54 mg/kg, with a 95% prediction interval of [3.59 mg/kg, 5.74 mg/kg]. In case the test is negative, it can be concluded that the actual (unknown) concentration lies between zero and the LOD95 of the respective laboratory, which is between the range of 3.59 and 5.74 mg/kg.

For part B (Cocktail (patented) extraction), the three common parameters were estimated as A=0.031, B=19.75, D=0.996, and the laboratory-specific parameter C was estimated as C = 5.40 mg/kg. This yielded a 95% prediction interval for parameter $C\gamma_i$ of [3.33 mg/kg, 8.76 mg/kg] of a future laboratory i. The overall LOD95 was 6.29 mg/kg with a 95% prediction interval of [3.88 mg/kg, 10.20 mg/kg].

Figure 5 shows the resulting POD curve for a laboratory with average performance (blue), and the POD curves for a top-performing laboratory (green) and a bottom-performing laboratory (red), calculated from the lower and upper limits of the 95% prediction interval for the parameter C. As for part A, a negative result means that the actual (unknown) concentration can be between zero and the LOD95 of the respective laboratory, which is within the prediction range of 3.88 and 10.20 mg/kg. The 95% prediction band for the ROD assuming 10 replicates per laboratory was again calculated using Monte-Carlo simulations (black step-functions). This prediction band was comparable to the one obtained for ethanol extraction.

4. Discussion and conclusions

The immuno-chromatographic method evaluated in this collaborative study was designed to detect gluten at levels clearly below the threshold of 20 mg/kg. A qualitative method to detect gluten will only result in a 'yes' or 'no' answer, but a user of this system needs to know with a given confidence: (1) what minimal concentration is present, if the result is positive; and (2) what maximum amount of gluten is may be present, if the result is negative. This was the reason why a new statistical approach beside the AOAC Appendix N (AOAC, 2013) was used. According to AOAC Appendix N, a mean value ('LPOD') is calculated for each concentration out of all single 'yes' or 'no' answers from the results of all participating laboratories. The influence of the results from an individual laboratory on the mean value is therefore not considered. It became clear from the presented data for qualitative gluten analysis, that the variation within a given laboratory was low whereas the differences between laboratories could be quite high. That is why even single laboratories with very high or very low sensitivities will influence the 'LPOD' in a statistically not

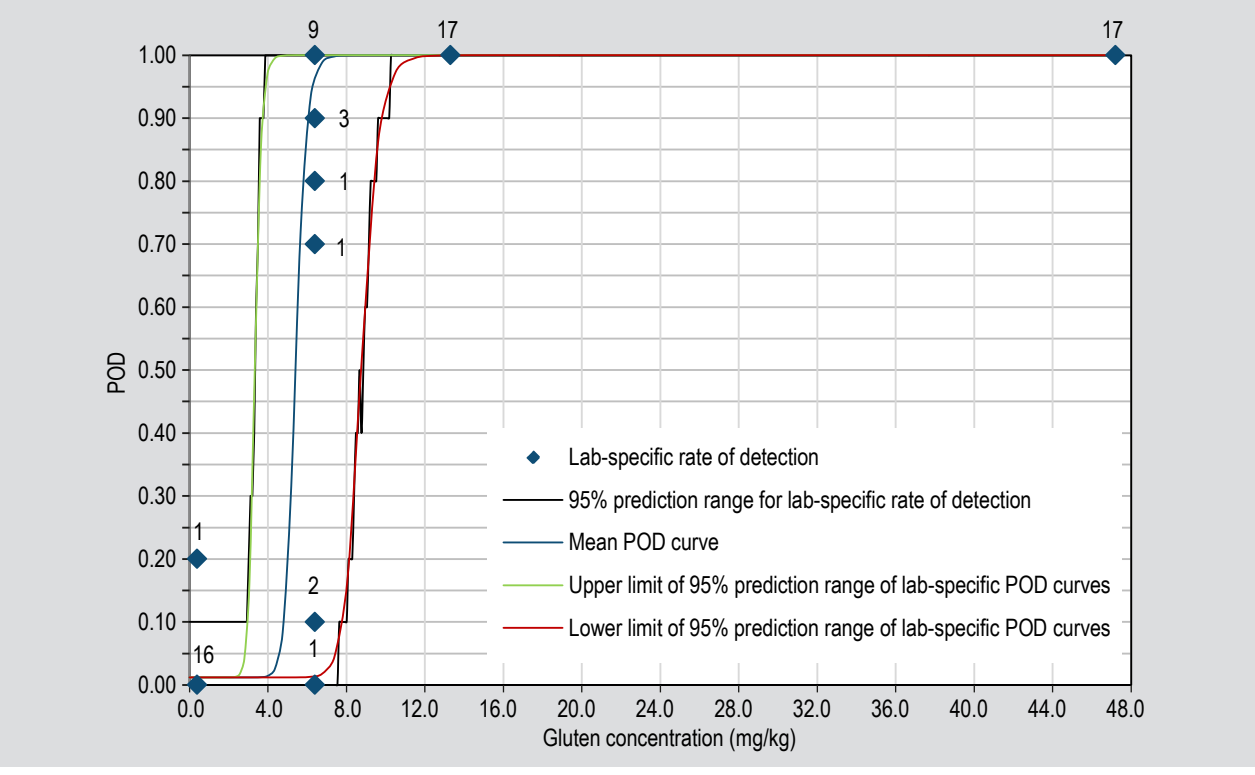

Figure 5. Performance statistics for part B (Cocktail (patented) extraction): probability of detection (POD) curves for a laboratory with average performance (blue), top-performance (green), and bottom performance (red) calculated from the lower and upper limit of the 95% prediction interval for the parameter $C\gamma_i$. Numbers near the diamonds for the lab-specific detection rate indicate the number of laboratories. Black step functions show the results of the Monte-Carlo simulations.

comprehensible way. As an example (assuming 10 replicates for a sample), 171 out of 180 results were reported positive. For the Appendix N statistical approach, it is irrelevant if the 9 negative results originate from one laboratory with 9 negative results or from 9 laboratories with just one negative result each. It is more than a small difference if one found 17 out of 18 laboratories with a POD of 1 in combination with only 1 laboratory out of 18 laboratories with a POD of 0.1 compared to 9 out of 18 laboratories with a POD of 1 in combination with 9 laboratories with a POD of 0.9. If all laboratories have a POD of 0.9 or more, it will be very unlikely that 8 or more test results out of 10 single tests of an unknown sample are negative, if the concentration of the sample is the same. Therefore, in this case, it could be concluded that the unknown sample has a lower concentration. This conclusion cannot be drawn if there are some laboratories with very low POD. This difference between the two cases is reflected in the laboratory standard deviation which will be much larger if laboratory-specific PODs are very different. A mathematically more comprehensive discussion of this aspect can be found in Uhlig *et al.* (2015a).

The resulting 'LPOD graph' is speculated to be the performance of a median or average laboratory by Appendix N, but it does not really represent the performance of a median performing laboratory or of any other laboratory. In case of a large laboratory standard deviation, the 'LPOD graph' can be misleading because it is – almost always – less steep than the POD curves of individual laboratories.

Also from a technical point of view, the calculated confidence intervals are incorrect: if the LPOD is below 0.15 or above 0.85, variability between laboratories is not taken into account, and hence the confidence intervals underestimate uncertainty (Uhlig *et al.,* 2011, 2013).

In contrast, the mean values and their prediction intervals resulting from this new way of calculation clearly show that the immuno-chromatographic dip-stick RIDA®QUICK Gliadin is fit for its intended use to detect samples as negative that are clearly below the Codex threshold of 20 mg gluten/kg. The approach is state-of-the-art in the area of statistics.

If a trained potential user works in a gluten-free laboratory and sets up a quality control plan using check samples, the results obtained with the described method will be comparable to the results of the participating laboratories.

Acknowledgements

We would like to thank Carrie Maune for preparation of the samples and Nathalie Widmann and Helene Bayer for homogeneity testing and blinding of the samples. We thank Clyde Don for his helpful contributions to this successful study. The participation of the following laboratories in the collaborative study is greatly acknowledged: Paulo Da Costa (Nestle Research Center, Lausanne, Switzerland), Guenther Augustin (Dr. Schär AG, Burgstall/Postal, Italy), Sandor Tömösközi (University of Technology and Economics, Budapest, Hungary), Roberto Lattanzio (Eurofins, Hamburg, Germany), Tuula Sontag-Strohm (University of Helsinki, Helsinki, Finland), Joe Baumert (FARRP, University of Nebraska, Lincoln, NE, USA), Mia Hallgren (National Food Agency, Uppsala, Sweden), Lukas Kraft (R-Biopharm AG, Darmstadt, Germany), Rupert Hochegger (AGES, Vienna, Austria), Gavin O'Connor (JRC-IRMM, Geel, Belgium), Andrew Flanagan (Public Analyst's Laboratory, Galway, Ireland), Fernando Chirdo (Facultad de Ciencias Exactas, La Plata, Argentina), Paul Wehling (General Mills, Golden Valley, MN, USA), Dan Thompson (Eurofins, Metairie, LA, USA), Terry Koerner (Health Canada, Ottawa, Canada), Janette Gelroth (AIB International, Manhattan, KS, USA), Helen Brown (Campden BRI, Chipping Campden, United Kingdom), and Peter Koehler (DFA, Leibniz Institut, Freising, Germany).

References

AACC International (AACCI), 2013. Approved methods of analysis, 11th Ed. Method 38-50.01. Gluten in corn flour and corn-based products by sandwich ELISA. AACCI, St. Paul, MN, USA.

AOAC International, 2002. AOAC official methods of analysis. Appendix D: guideline for collaborative study procedures to validate characteristics of a method of analysis. AOAC, Gaithersburg, MD, USA.

AOAC International, 2012. Gliadin as a measure of gluten in foods containing wheat, rye, and barley. AOAC official method of analysis 2012.01. AOAC, Gaithersburg, MD, USA.

AOAC International, 2013. AOAC official methods of analysis. Appendix N: ISPAM guidelines for validation of qualitative binary chemistry methods. AOAC, Gaithersburg, MD, USA.

AOAC International, 2015. Gluten in processed and non-processed corn products by qualitative R5 immunochromatographic dip-stick. AOAC official method of analysis 2015.16. AOAC, Gaithersburg, MD, USA.

Bugyi, Z., Török, K., Hajas, L., Adonyi, Z., Popping, B. and Tömösközi, S., 2013. Comparative study of commercially available gluten ELISA kits using an incurred reference material. Quality Assurance and Safety of Crops & Foods 5: 79-87.

European Commission (EC), 2009. Commission regulation (EC) No. 41/2009 of January 20, 2009, concerning the composition and labelling of foodstuffs suitable for people intolerant to gluten. Official Journal of the European Union 16: 3-5.

Food and Agriculture Organisation/World Health Organisation (FAO/WHO), 2013. Codex Standard 234-1999. Recommended methods of analysis and sampling. Codex Alimentarius Commission, FAO/WHO, Rome, Italy.

Food and Agriculture Organisation/World Health Organisation (FAO/WHO), 2015. Codex Standard 118-1979 (Revision 1, Amendment 2). Foods for special dietary use for persons intolerant to gluten. Codex Alimentarius Commission, FAO/WHO, Rome, Italy.

Food and Drug Administration (FDA), 2013. Gluten-free labeling of foods; final rule. Federal Register, Vol. 78, No. 150. Rules and Regulations, pp. 47154-47179. FDA, College Park, MD, USA. Available at: http://tinyurl.com/zks9ssa.

Garcia, E., Llorente, M., Hernando, A., Kieffer, R., Wieser, H. and Méndez, E., 2005. Development of a general procedure for complete extraction of gliadins for heat processed and unheated foods. European Journal of Gastroenterology and Hepatology 17: 529-530.

Health Canada, 2012. Health Canada's position on gluten-free claims. Health Canada, Ottawa, Ontario, Canada. Available at: http://tinyurl.com/hus37m2.

Immer, U. and Haas-Lauterbach, S., 2012. Gliadin as a measure of gluten in foods containing wheat, rye, and barley – enzyme immunoassay method based on a specific monoclonal antibody to the potentially celiac toxic amino acid prolamin sequences: collaborative study. Journal of AOAC International 95: 1118-1124.

International Organisation for Standardisation (ISO), 1994. ISO 5725-2:1994. Accuracy trueness and precision) of measurement methods and results – Part 2: basic method for the determination of repeatability and reproducibility of a standard measurement method. ISO, Geneva, Switzerland.

International Organisation for Standardisation (ISO), 2015. ISO 13528:2015. Statistical methods for use in proficiency testing by interlaboratory comparison. ISO, Geneva, Switzerland.

International Organisation for Standardisation/International Electrotechnical Commission (ISO/IEC), 2008. JCGM 101 – Evaluation of measurement data – Supplement 1 to the 'Guide to the expression of uncertainty in measurement' – Propagation of distributions using Monte Carlo method (guide 98-3-1). ISO, Geneva, Switzerland.

Kahlenberg, F., Sanchez, D., Lachmann, I., Tuckova, L., Tlaskalova, H., Méndez, E. and Mothes, T., 2006. Monoclonal antibody R5 for detection of putative celiac-toxic gliadin peptides. European Food Research and Technology 222: 78-82.

Koehler, P., Schwalb, T., Immer, U., Lacorn, M., Wehling, P. and Don, C., 2013a. AACCI Approved Methods Technical Committee report: collaborative study on the immunochemical determination of intact gluten using an R5 sandwich ELISA. Cereal Foods World 58: 36-40.

Koehler, P., Schwalb, T., Immer, U., Lacorn, M., Wehling, P. and Don, C., 2013b. AACCI Approved Methods Technical Committee report: collaborative study on the immunochemical determination of partially hydrolyzed gluten using an R5 competitive ELISA. Cereal Foods World 58: 154-158.

Lacorn, M. and Weiss, T., 2015. Partially hydrolyzed gluten in fermented cereal-based products by R5 competitive ELISA: collaborative study, first action 2015.05. Journal of AOAC International 98: 1346-1354.

Lionetti, E., Gatti, S., Pulvirenti, A. and Catassi, C., 2015. Celiac disease from a global perspective. Best Practice & Research Clinical Gastroenterology 29: 365-379.

Ludvigsson, J.F., Leffler, D.A., Bai, J.C., Biagi, F., Fasano, A., Green, P.H.R., Hadjivassiliou, M., Kaukinen, K., Kelly, C.P., Leonard, J.N., Lundin, K.E.A., Murray, J.A., Sanders, D.S., Walker, M.M., Zingone, F. and Ciacci, C., 2013. The Oslo definitions for coeliac disease and related terms. Gut 62: 43-52.

Osman, A.A., Uhlig, H.H., Valdes, I., Amin, M., Méndez, E. and Mothes, T., 2001. A monoclonal antibody that recognizes a potential coeliac-toxic repetitive pentapeptide epitope in gliadins. European Journal of Gastroenterology and Hepatology 13: 1189-1193.

Plugis, N.M. and Khosla, C., 2015. Therapeutic approaches for celiac disease. Best Practice & Research Clinical Gastroenterology 29: 503-521.

Scherf, K.A. and Poms, R.E., 2016. Recent advances in analytical methods for tracing gluten. Journal of Cereal Science 67: 112-122.

Scherf, K.A., Koehler, P. and Wieser, H., 2016. Gluten and wheat sensitivities – an overview. Journal of Cereal Science 67: 2-11.

Schuppan, D., Pickert, G., Ashfaq-Khan, M. and Zevallos, V., 2015. Non-celiac wheat sensitivity: differential diagnosis, triggers and implications. Best Practice & Research Clinical Gastroenterology 29: 469-476.

Sicherer, S.H. and Sampson, H.A., 2014. Food allergy: epidemiology, pathogenesis, diagnosis, and treatment. Journal of Allergy and Clinical Immunology 133: 291-307.

Sykes, M., Knaggs, M., Hunter, S., Leach, E., Eaton, C. and Anderson, D., 2014. Some selected discrepancies observed in food chemistry proficiency tests. Quality Assurance and Safety of Crops & Foods 6: 291-297.

Thompson, M., Ellison, S.L.R. and Wood, R., 2006. The international harmonized protocol for the proficiency testing of analytical chemistry laboratories. Pure and Applied Chemistry 78: 145-196.

Thompson, T., 2003. Oats and the gluten-free diet. Journal of the American Dietetic Association 103: 376-379.

Tye-Din, J.A., Stewart, J.A., Dromey, J.A., Beissbarth, T., Van Heel, D.A., Tatham, A., Henderson, K., Mannering, S.I., Gianfrani, C., Jewell, D.P., Hill, A.V., McCluskey, J., Rossjohn, J. and Anderson, R.P, 2010. Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. Science Translational Medicine 2: 41ra51.

Uhlig, S., Niewöhner, L. and Gowik, P., 2011. Can the usual validation standard series for quantitative methods, ISO 5725, be also applied for qualitative methods? Accreditation and Quality Assurance 16: 533-537.

Uhlig, S., Krügener, S. and Gowik, P., 2013. A new profile likelihood confidence interval for the mean probability of detection in collaborative studies of binary test methods. Accreditation and Quality Assurance 18: 367-372.

Uhlig, S., Frost, K., Colson, B., Simon, K., Made, D., Reiting, R., Gowik, P. and Grohmann, L., 2015a. Validation of qualitative PCR methods on the basis of mathematical-statistical modelling of the probability of detection. Accreditation and Quality Assurance 20: 75-83.

Uhlig, S., Simon, K., Frost, K., 2015b. Mathematical-statistical procedure for determining the concentration-dependent probability of detection of qualitative chemical methods on the basis of interlaboratory study data. Available at: http://tinyurl.com/zgqyqly.

Valdes, I., Garcia, E., Llorente, M. and Mendez, E., 2003. Innovative approach to low-level gluten determination in foods using a novel sandwich ELISA protocol. European Journal of Gastroenterology and Hepatology 15: 465-474.