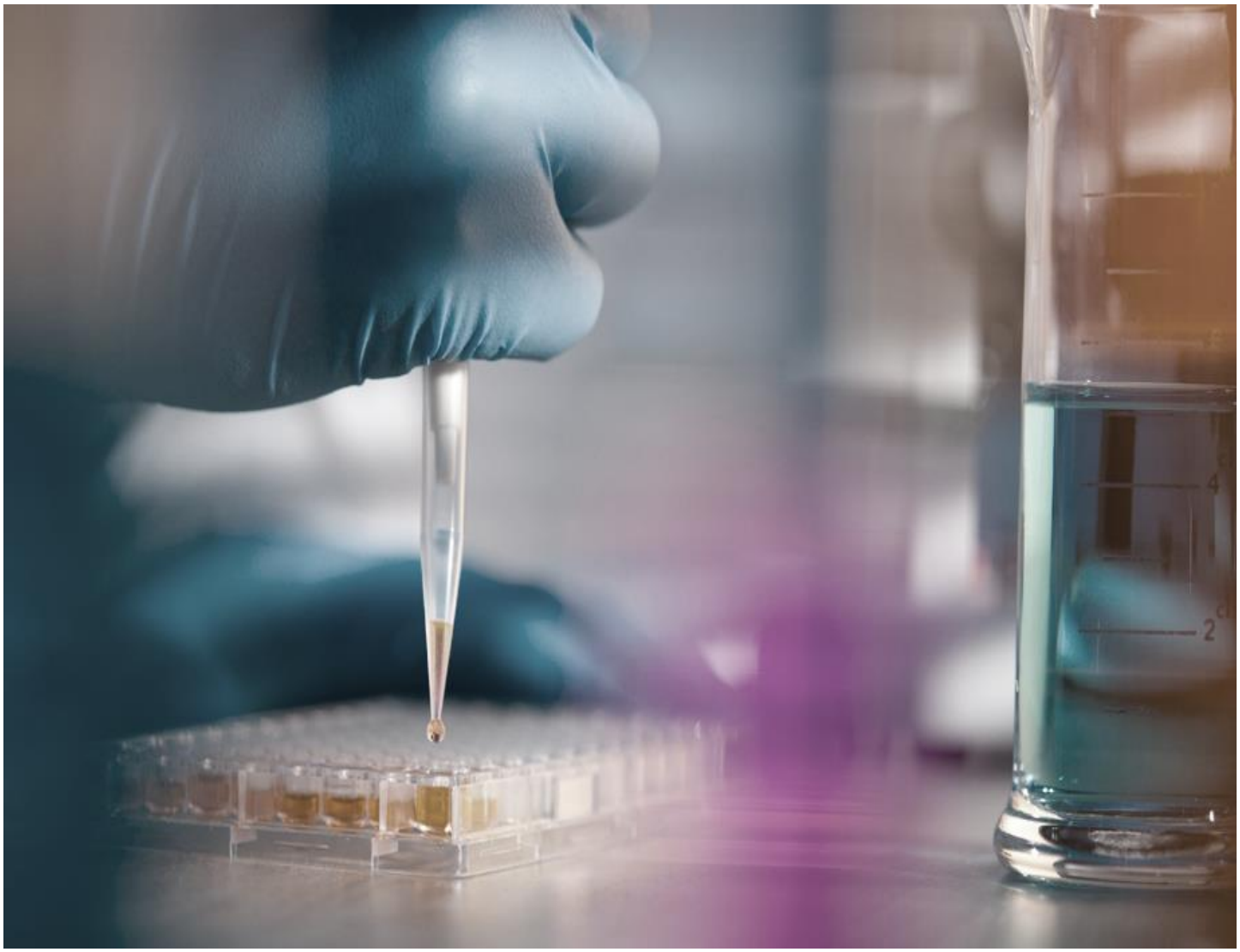

ORGANISATION OF A PROFICIENCY TEST FOR FOOD ALLERGENS

Panagiota Alamenou

RIKILT
WAGENINGEN UR

ORGANISATION OF A PROFICIENCY TEST FOR FOOD ALLERGENS

Author: Panagiota Alamenou

Registration number: 880904-009-050

Thesis code: TOX80436

Supervisors: Dr. Ir. Monique Bremer (RIKILT)

Dr. Ir. Ans Punt (Toxicology)

RIKILT Wageningen UR,

Wageningen, April 2015

This master thesis report is especially dedicated to my family, every member of which supported me in all possible ways to conduct my MSc studies in Food Safety.

Panagiota Alamenou

Contents

ABSTRACT

Food related allergies comprise a significant public health issue nowadays. Avoidance of food that contains allergen(s) is the only preventive measure for sensitive individuals, and therefore food labelling is of great significance. Food allergen labelling has been rendered obligatory by the legislation. In order foodstuffs to comply with this legislation, reliable methods for the detection and quantification of food allergens are needed, the reliability and accuracy of which can be assessed through validation studies and participation in Proficiency Tests. The number of Proficiency Tests for food allergens that were organised so far is limited compared to the needs for accurate detection of allergens. In this study an effort is made towards an organisation of a new Proficiency Test that focuses on five important allergens, namely almond, gluten, peanut, milk and soy. The main aim of this study is the investigation of the current status of detection and quantification of these five allergens performed by analytical laboratories worldwide.

To achieve this, 22 laboratories all over the world were asked to analyse three different materials of specific allergen content with methods of their choice. The materials were prepared by spiking baby food based on mixed cereals. The homogeneity of the three materials was ensured through homogeneity testing. The stability testing resulted in detection of consequential instability for almond, gluten and skim milk powder. The participants applied only immunochemical methods, mainly commercial ELISA kits. The statistical evaluation of the results collected showed that 73% and 83% of the quantitative results reported for almond and gluten respectively were satisfactory. For peanut, skim milk powder and soy flour the uncertainty of the assigned values was very high to be used for evaluation of the laboratories performance. For these allergens the results reported were grouped according to the different ELISA kits used. Correct classification of spiked samples was acquired in 100% of almond, 94% of gluten and peanut, 78% of skim milk powder and 53% of soy flour analyses.

Although the majority of the qualitative results were satisfactory, even at low concentrations, quantification of the allergens proved to be more difficult. For all the five allergens of interest a large variation regarding the quantitative results was noted, mainly depended on the commercial brand of the ELISA kits used. Therefore, the current allergen detection methods need to be further developed and harmonised. In daily practice, the majority of the analysis of these five allergens is routinely performed with commercially available ELISA kits. Gluten analysis seems that can be performed successfully, even quantitatively, and the labelling of “gluten-free” and “very low gluten” foodstuffs seems that can be successfully determined by the methods currently available.

1. INTRODUCTION

1.1 Food related allergies and intolerances

Food related allergies comprise a significant public health issue for modern societies. According to the definition adopted and used by EFSA in the respective scientific opinion report, food allergy is "an adverse health effect arising from a specific immune-mediated response that occurs reproducibly on oral exposure to a given food, which can be mediated by food-specific IgE antibodies, by cellular mechanisms or by both". Immune, IgE-mediated food allergies are characterised by rapid manifestation of mild to severe allergic reactions (mostly within two hours after the patient has orally been exposed to the responsible allergen), with symptoms involving various systems in the organism (respiratory, digestive, cutaneous or cardiovascular). As far as immune, non-IgE (cell)-mediated food allergies are concerned, they differ from IgE-mediated food allergies in that mainly digestive system is affected and the manifestation of the symptoms is delayed (2-48 hours after the patient has orally been exposed to the responsible allergen) (EFSA NDA Panel, 2014). Food intolerances (or non-immune-mediated adverse reactions to food) are caused by either absence or shortage of metabolic enzymes (e.g. lactose intolerance), or intake of pharmacologically active substances in high doses with foods (e.g. caffeine) leading to histamine or tyramine release (Hamer, 2013, EFSA NDA Panel, 2014).

The amount of the various allergens that can result in allergic reaction is difficult to be determined for every sensitive individual. In controlled studies the reported doses of the allergenic food or ingredient that are able to trigger adverse reactions were found to range from micrograms to milligrams, or even grams. So far, avoidance of food that contains the allergenic compound(s) is the unique preventive measure (EFSA NDA Panel, 2014).

1.2 Prevalence of food related allergies and intolerances

A general overview reported through a comprehensive bibliographic study states that ''food allergy affects more than 1% to 2% but less than 10% of the population'' (Sicherer & Sampson, 2014). Other sources in literature present percentages for food allergy of 5-7% and up to 4% in young children and adults, respectively, in the countries of West Europe. (Alexopoulos, Kakoulides, Lampi, 2014). According to the Food Allergen Labeling and Consumer Protection Act of 2004 (Public Law 108-282, Title II), in the United States 2% of the adults and 5% of infants and young children are estimated to suffer from food allergy.

The prevalence of food allergy is difficult to be determined. In literature the reported numbers are highly heterogeneous and there is uncertainty in the acquisition of accurate percentages. The difficulties in recording the realistic prevalence of food allergy worldwide can be attributed to various reasons. On the one hand, the different studies published so far may base their results on different applied methodologies, target populations and their age, geographical areas studied or even different definitions for allergy (EFSA NDA Panel, 2014, Rona et al., 2007, Sicherer & Sampson, 2014). Results, for instance, vary depending on which methods are used to characterise an individual as allergic to food, i.e. self-reported food allergy or confirmed diagnosis by a food challenge. In a recent study the self-reported food allergy in Europe was estimated to be 17.3 % and 5.9 % the lifetime and point prevalence respectively, but only 0.9 %, in both adults and children when there was a confirmation with a food challenge. On the other hand, a lack is reported in the existence of relevant studies, e.g. regarding prevalence of food allergy in specific geographical areas (EFSA NDA Panel, 2014).

In addition to the above, the prevalence of the food related allergies is influenced by a variety of environmental, as well as individual factors, like age and sex. More specifically, the geographical variations that play a role in food allergy prevalence are connected to varying genetic factors and other local factors. Such factors are for instance differences in consumption and eating habits or in exposure to allergenic foods. Cultural differences and general socioeconomic factors can also be implicated as differentiating factors for prevalence (EFSA NDA Panel, 2014).

1.3 Coeliac disease

Coeliac disease, is a non IgE-mediated condition (Hamer, 2013) which is described as "an autoimmune adverse reaction to food triggered by the ingestion of gluten and related to prolamins found in wheat, barley and rye" (EFSA NDA Panel, 2014). Gluten proteins are responsible for the damage caused in the intestine and the subsequent malabsorption of nutrients which leads to a series of serious, lifelong pathologies for the patients (Hamer, 2013). In Europe and the western world the prevalence estimated for coeliac disease is around 0.5 to 1 % (EFSA NDA Panel, 2014).

For coeliac disease, life-long adoption of a gluten-free diet is the only treatment available. This approach often results in alleviation of the disease's clinical symptoms within days or weeks, although recovery at histological level is a long-term process and can last for months or even years (Green & Cellier, 2007). Elimination of gluten from patients' diet is not always found to have the expected results. In 7 to 30% of the cases patients that follow a gluten-free

diet do not report alleviation of symptoms (Green & Cellier, 2007, Pantaleoni et al., 2014). According to research findings, persistent clinical symptoms can be caused by both non-food related and food-related reasons. The first category includes small intestinal bacterial overgrowth (Tursi et al., 2003), co-morbidities, such as thyroid diseases, or another gastrointestinal disorder such as irritable bowel syndrome (Paarlahti et al., 2013). In the second category, dietary non adherence (up to 60% of coeliac disease patients are partially non-adherent) (Hall et al., 2009), unintended consumption of 'hidden gluten' in processed foods (Hollon et al., 2013), consumption of food products contaminated with gluten during growing, packaging and transportation processes (Culliford & Green, 2003), and presence of another food intolerance such as lactose intolerance (Murray et al., 2004) are included.

1.4 Allergenic foods

Food allergens can be defined as proteins, peptides or carbohydrate moieties that can elicit an allergic reaction. A complex food product may contain one or more allergenic proteins (EFSA NDA Panel, 2014). Any food is possible to contain substances that can be potentially allergenic for any individual. Factors that contribute to this is the fact that the immunological responses to various food constituents are in many cases unpredictable for every individual, as well as the fact that there is redundancy and phylogenetic conservation of proteins among the different species (Kizis & Siragakis, 2014).

There are many different food allergens of animal, as well as of plant origin. There are eight main foods or food groups, i.e. fish, Crustacean shellfish, milk, eggs, tree nuts, peanuts, wheat, and soybeans, often mentioned as "the big 8", that are responsible for 90% of the caused food allergies according to the US Food Allergen Labelling and Consumer Protection Act of 2004. Moreover, there are results of some studies demonstrating that specific foodstuffs, i.e. cow's milk, egg, peanut, various nuts and fish account for the 75 % of allergic reactions in children. Foodstuffs that are widely implicated in adults' food allergies (up to 50 % of allergic reactions) are peanuts and various nuts, fruits belonging to the Rosaceae family, like apples, Apiaceae's family vegetables, like celery, and fruits of the latex group, like kiwi (EFSA NDA Panel, 2014).

1.5 Background of legislation on food allergens

As already mentioned, avoidance of food that contains the allergenic compound(s) is the only preventive measure for sensitive individuals. For this reason, food labelling is of great significance for individuals suffering from food allergy (Monaci & Visconti, 2010). The

importance of food labelling, including declaration of food allergens, and the effort towards it, is reflected on the European legislation. Regulation (EU) No 1169/2011 is important as it provides guidance on the presentation of food information in order to allow consumers to make informed choices. According to Article 9, paragraph 1(c) of the latter Regulation, ingredients of foods or processing aids that can induce allergic reactions or intolerances to consumers are included in the list of mandatory food information. A detailed list of the specific ingredients and processing aids that should be obligatory declared as allergens on the food labels is given in Annex II of the same Regulation. Further instructions regarding indication of allergens are given with Article 21.

Another relevant act of European legislation is Directive 2000/13/EC on labelling, presentation and advertising of foodstuffs. A number of amendments were issued for Directive 2000/13/EC particularly for allergen related issues. Amendments of importance are Directive 2003/89/EC, in which a list of allergenic foods is included in Annex IIIa, and Directive 2007/68/EC, which is amending Annex IIIa and introducing the up-to-date list of allergenic ingredients that should be clearly declared on the label and the exceptions.

The interest for protection of the allergic consumers has a worldwide character. Outside Europe there is also legislation and guidelines that regulate food allergens labelling. The US Food Labeling and Consumer Protection Act which provides a list of nine major allergens, the FSANZ Food Standards Code implemented in Australia and New Zealand, which contains a similar list, as well as the Food Sanitation Law in Japan are examples of the effort made worldwide on the management of food allergens (Lacorn & Immer, 2010, Lampidonis & Siragakis, 2014). As already discussed above, the prevalence of different food allergies can vary for the various geographical areas. Consequently, the allergens that are considered major and need to be declared in the labels can also be different for every geographical area or country. This is reflected in every country's legislation on food labelling, too. In Table 1 the allergens that are of particular interest within the legislative framework for some countries are presented (Lampidonis & Siragakis, 2014).

Table 1: Major food allergens in different countries' legislation (Lampidonis & Siragakis, 2014).

Allergenic foods	European union	Australia/New Zealand	Canada	China	Hong Kong	Japan	Korea	Mexico	United States
Wheat/cereals	x	x	x	x	x	x[a]	x[b]	x	x
Eggs	x	x	x	x	x	x	x	x	x
Milk	x	x	x	x	x	x	x	x	x
Peanut	x	x	x	x	x	x	x	x	x
Fish	x	x	x	x	x		x[b]	x	x
Crustaceans	x	x	x	x	x	x[a]	x[b]	x	x
Soy	x	x	x	x	x		x	x	x
Tree nuts	x	x	x	x	x			x	x
Sesame	x	x	x						
Shellfish/molluscs	x		x						
Mustard	x		x						
Celery	x								
Lupine	x								
Other						x[a]	x[b]		

[a]Shrimp and crab are the only crustaceans listed. Grains include wheat and buckwheat. 'Other' includes foods for which labelling is recommended but not required: abalone, squid, salmon roe, oranges, kiwifruit, beef, walnuts, salmon, mackerel, soybeans, chicken, bananas, pork, matsuake mushrooms, peaches, yarns, apples and gelatin. [b]Mackerel is the only fish, crab and shrimp are the only crustaceans listed. Grains include wheat and buckwheat. 'Other' includes pork, peaches, and tomatoes.

All the legislative acts mentioned above are indicative of the importance of the explicit declaration of food allergens on the food labels. Although a great effort is made, there are two major deficiencies in the European legislation regarding labelling of food allergens. First, except for gluten, no threshold levels are currently determined in legislation for the different allergens in order to facilitate the food allergens labelling. Regarding labelling of foodstuffs for gluten, REGULATION (EC) No 41/2009 and REGULATION (EU) No 828/2014 define limits for two labelling categories, i.e. "very low gluten" and "gluten-free". The gluten content of foodstuffs that bear the labelling "very low gluten" shall not exceed 100 mg/kg in the final food for consumption, while for "gluten-free" foods the limit is lowered to 20 mg/kg. However, there are not similar limits defined for the labelling of other allergens. This deficiency of limits leads quite often to the presence of the unnecessary precautionary

labelling on the foodstuffs and consequently to confusion and limitation of food choice options for sensitive individuals. Second, legislation on food labelling regulates only the intentionally added allergenic compounds as ingredients of the foods. The allergen traces of mg/kg level that often come from cross-contamination throughout the food chain and comprise an important source, are not considered in the legislation (Lacorn & Immer, 2010, Monaci & Visconti, 2010).

To overcome the deficiencies described above, food industry can optionally use VITAL (Voluntary Incidental Trace Allergen Labelling) approach, an initiative started by the Australian Food and Grocery Council and being more and more widely adopted. VITAL is a risk-based methodology that allows food manufacturers to assess the risk from cross-contamination with allergens and decide on the appropriate form of labelling of foodstuffs. The decisions are based in two Action levels, which are concentrations determined based on scientific threshold levels (LOAEL) acquired from literature. In Action level 1, the concentration of allergens is low and no precautionary labelling is required. In Action level 2, precautionary labelling is the key as allergen proteins are present in a significant concentration. The two Action levels are calculated from the Reference Dose and the Reference Amount/Serving Size according to the following equation: **Transition between Action levels = Reference Dose × (1000 / Reference Amount/Serving Size)** (http://allergenbureau.net/).

Nowadays, a great effort is made, in which different parties are involved, towards the determination of the threshold concentrations of allergenic foods or ingredients below which the sensitive individuals are not in danger of developing an allergic reaction. The main interest of this effort has been shifted to the development of the methodology for food allergens risk assessment. The different methodologies that are proposed so far include the risk assessment based on the no observed adverse effect level (NOAEL) approach, the Bench Mark Dose (BMD) and Margin of Exposure (MoE) approach, as well as risk assessment using probabilistic models. The latter seems so far to be the most promising (Spanjersberg et al., 2007, Madsen et al., 2009, Lacorn & Immer, 2010, EFSA NDA Panel, 2014).

1.6 Current methods for food allergen detection

Compliance of foodstuffs with food labelling legislation, and especially with declaration of food allergens, has led to the need for reliable methods for the detection and quantification of food allergens. Currently, there are different approaches available for the detection of allergenic ingredients, based on the targeted allergen, the food matrix involved, as well as the processing treatments applied. Therefore, no single method exists which can be applied in all

purposes (Poms et al., 2007, EFSA NDA Panel, 2014). The available methods target different structures, i.e. either directly a specific allergenic protein (or a number of proteins), or indirectly a marker that is indicative for the presence of allergenic ingredients. These markers include DNA fragments and other proteins specific for the presence of the target allergen (EFSA NDA Panel, 2014, Schubert-Ullrich et al., 2009). The methods available for detection of food allergens can be either qualitative or quantitative. For an analytical method to be reliable, well-defined reference materials, as well as an effective protocol of recovery of allergenic proteins from the material to be analysed are basic requirements. Additionally, other requirements that need to be met are sensitivity, specificity, accuracy and precision (repeatability, reproducibility). Further parameters that should be also taken into account for the selection of the suitable method are matrix effects, and level and type of food processing (EFSA NDA Panel, 2014).

Detection methods aimed at specific allergenic proteins are mainly immunological (mainly ELISA) and physicochemical methods (notably Mass Spectrometry (MS)). Regarding the methods that target specific DNA fragments, the analysis is based on their amplification. For this purpose, PCR method and the use of specific primers are employed (EFSA NDA Panel, 2014).

The majority of the food allergen analysis is routinely performed by using commercially available immunological kits. The detection of allergenic proteins with immunological methods is based on the use of antibodies (Monaci & Visconti, 2010, EFSA NDA Panel, 2014). There are two main classes of antibodies that are utilised against allergenic proteins, i.e. monoclonal and polyclonal antibodies. The former target one single epitope and can recognize specific antigenic sequences, while the latter are able to bind multiple epitopes of the antigenic proteins. The ability of polyclonal antibodies to target multiple epitopes renders them more tolerable to possible changes of the allergens and therefore more suitable for the detection of allergens in processed foodstuff. On the other hand, polyclonal antibodies are responsible for more cross-reactivity phenomena, which lowers the specificity of a method (EFSA NDA Panel, 2014, Hamer, 2013, Monaci & Visconti, 2010). The immunological methods for allergen detection that are currently used include: Enzyme-linked immunosorbent assays (ELISAs), lateral flow devices (LFDs), dipsticks, rocket immuno-electrophoresis (RIE), dot-immunoblotting (dot-blot), protein microarrays and protein biosensors (EFSA NDA Panel, 2014, Monaci & Visconti, 2010).

The most widely used immunological method for the detection of allergenic ingredients in foods is ELISA. There are two commercially available principle set-ups for ELISA; the sandwich and the competitive (direct and indirect competitive) (EFSA NDA Panel, 2014,

Hamer, 2013, Monaci & Visconti, 2010). The fact that ELISA is the method of preference can be attributed to its sensitivity (some ELISA kits are very sensitive with low LODs), specificity and stability for the detection of allergenic proteins. Moreover, ELISAs are easy to use, relatively inexpensive, specialised personnel is not required to perform them, while many analysis can be carried out simultaneously. However, ELISA methods present also some significant shortcomings that should be always taken into account. The different commercial kits are based on different extraction protocols, different calibration procedures, the use of antibodies of varying quality, and therefore the results acquired are dependent on the various commercial brands. Other limitations of ELISA kits are the possible matrix effects that can influence the final results, the insufficient extraction of allergenic proteins, cross-reactivity with various food components or other allergens that can lead to false positive results, and the difficulty in acquiring reproducible results (EFSA NDA Panel, 2014, Monaci & Visconti, 2010).

An alternative, emerging method of allergen identification and quantification in foods is Mass Spectrometry. This method focuses on either intact protein analysis or use of accurate peptide mass. As sequencing of complete proteins is not easy, the focus is mainly on individual peptides. Currently is used only for confirmation purposes, supplementary to other methods. When combined with other techniques, such as chromatography, for the preliminary separation of the proteins, and with databases for the subsequent allergen identification, MS can be a reliable approach. For sufficient identification of allergenic proteins with MS, scanned peptide spectra are being statistically identified with specific protein-sequence databases. An important advantage this method shows is the fact that it is not affected by cross reactions, since it targets certain allergenic proteins/peptides. However, MS-based methods are still developing and the available instrumentation is not optimal yet. Other disadvantages for these methods are the high cost, the specialised staff needed for application, as well as the long time for sample preparation (Costa et al., 2014, EFSA NDA Panel, 2014, Monaci & Visconti, 2009, Monaci & Visconti, 2010).

DNA-based methods, as they are indirect methods, detect the gene that encodes for the allergenic protein rather that the allergenic protein itself. This type of detection of allergens in food can be used as substitute method in cases of low expected levels of the target allergenic proteins or complementary to other methods. When the conventional PCR method is employed the results are qualitative, while with RT-PCR the determination of allergens can be quantitative, sometimes with challenging LODs. Other DNA-based methods used are the DNA microarrays and the DNA biosensors. One of the advantages of DNA-based detection methods is that it can be applied in analysis of processed foodstuff, since DNA structure is more stable compared to that of the proteins. However, in cases of food matrices that are

highly processed degradation of DNA can influence the results. Further strong points of these methods include the possibility of simultaneous detection of multiple food allergens in the same assay, and the well-defined extraction protocols available. PCR methods have also the advantage of high selectivity upon the selection of the suitable primers. A significant weak point of the DNA-based methods is that the results are dependent on the presence or not, the amount and quality of DNA material to encode for the targeting allergen in processed food (EFSA NDA Panel, 2014, Monaci & Visconti, 2010).

Other methods for allergen detection purposes in food are also applied, for example, the SDS-PAGE method, which is based on previous separation of targeted proteins. However, these methods are currently used only to minor extent, mainly for research purposes (EFSA NDA Panel, 2014).

1.7 Proficiency Testing

A large variety of commercial ELISA test kits for allergen detection and quantification in foods is currently available, and new methods are continuously being developed. The reliability and accuracy of the analytical results produced by all these kits, need to be assessed and ensured. This can be achieved through validation studies for these detection methods and participation in Proficiency Tests (Schubert-Ullrich et al., 2009). Validation of analytical methods used along with participation in Proficiency Tests, are significant elements of laboratory quality assurance. Laboratories should perform these methods under a well-established quality assurance/quality control system, as well as prove competence for their results. To prove their competence laboratories often participate in accreditation procedures according to International Standards Organization (ISO) 17025. The ISO/IEC 17025:2005 accreditation is a requirement for laboratories which are entrusted with the official control of food and feeds, according to Article 12 of EU Regulation (EC) 882/2004. Participation in existing suitable proficiency testing schemes is included Among the ISO 17025 requirements for the respective laboratories.

In the field of food allergen analytical detection methodologies validation is of significance not only for the developers, but also for the end users. Validation contributes to acquisition of reliable, repeatable, precise and accurate results by laboratories when using a specific method of detection (Monaci & Visconti, 2010). For the majority of the commercially available ELISA kits for the detection of food allergens in foods, validation data on their performance are not sufficient and well-established. For this reason, validation testing should be applied

for every commercial ELISA kit, before its final use, in order to determine its suitability for analysis of specific materials (Schubert-Ullrich et al., 2009).

Proficiency testing (PT) can be defined as an external and independent quality control procedure which aims at the objective assessment of the general performance of a participating laboratory. In a PT, the quality of an analytical result produced by a participating laboratory is evaluated against independently set criteria. The participants of a PT are asked to analyse an external quality check sample, with methods of their choice in order to simulate the procedures routinely applied. The providers of PT schemes convert afterwards the participants' results into scores in order to evaluate their performance. PT schemes can evaluate the quality system of a laboratory in total, as they provide unbiased standards and give the opportunity of comparison of analytical results with those produced by other laboratories. Other competencies of a laboratory, besides the ability to produce reliable and accurate results, that are assessed in a Proficiency Testing are for instance, the ability to report results within a given time scale or in specified units. All these are important aspects of quality for an analytical laboratory (Thompson et al., 2006, Owen & Gilbert, 2009, Monaci & Visconti, 2010).

A reliable comparison of the Proficiency Testing participants' results can be made on condition that all the latter receive identical materials for testing (Owen & Gilbert, 2009). However, it is difficult to achieve homogeneous materials for use in Proficiency Testing schemes. The knowledge on preparation of homogeneous, low level samples in bulk is generally limited. As a result, when this material is distributed to the different participants the produced results may vary slightly. For this reason, this variation should be kept at negligibly small levels. This can be proved through homogeneity testing schemes (Fearn & Thompson 2001, Thompson et al., 2006). The International Harmonized Protocol for Proficiency Testing of Analytical Laboratories (Thompson et al., 2006) and ISO 13528 (ISO 13528:2005(E), 2005) are mainly used as basis for the design of the homogeneity testing of the materials prepared for Proficiency Testing.

Apart from homogeneity, the materials prepared and distributed in a Proficiency Testing scheme should also present a sufficient level of stability, at least during the period in which the assigned value has to be valid. The period of interest is often the time between the materials preparation and the final date for participants to report the results (Thompson et al., 2006). For stability testing a comparison between two different storage conditions of the material is needed, i.e. storage of the material under conditions that ensure maximum stability, for example under low temperature or low light exposure, and storage under

conditions simulating the practices followed in reality that can lead to decomposition of the allergens (Thompson et al., 2006, Thompson, 2008).

Except for the difficulties in achieving sufficient homogeneity and stability for the PT samples, there are more problems encountered so far in relation to organisation of Proficiency Tests. Lack of reference methods and reference allergenic materials, spiking levels of allergens that are too high to resemble real-life samples in a realistic way, multimodality of the reported data which are often not normally distributed, as well as dependence of the reported results on the applied method of detection, are the most important (Monaci & Visconti, 2010, Scharf et al., 2013, Owen & Gilbert, 2009, Sykes et al., 2012, Lacorn & Immer, 2010). All the above problems are factors contributing to difficulties in interpretation of the results that are acquired from a Proficiency Test for food allergens and consequently in evaluation of laboratories.

Reference materials are materials or substances with sufficiently homogeneous and well-defined properties. They are used for method development and validation purposes to prove that the test kits used for allergen analysis are reliably detecting them and identify possible pitfalls. In this way, the performance of a method is assessed, the comparability of measurements acquired by different laboratories is promoted and the confidence of the analytical results is increased. Although reference materials are of significance for allergen analysis, a lack in their development is reported. So far, there are some major providers (e.g. IRMM in Belgium and NIST in USA) and great effort to develop reliable reference materials (Monaci & Visconti, 2010, Tomoskozi et al., 2014). Currently, there is only one reference material designed especially for allergen purposes (NIST SRM 8445, spray-dried whole egg). There is also a series of various reference materials available, but not developed for allergen analysis purposes, that can be used. Further problems are also encountered with the reference materials, such as the fact that the ELISA kits mainly target specific protein sequences, whilst reference materials are based on a complex protein mixture, which leads to the need of suitable conversion factors. However, the main problem related to reference materials is the fact that they are not stable. Reference materials are produced in batches, which can differ significantly. Moreover, their quality might depend on various external factors and vary with for example the crop, or the annual climatic conditions, which affect the protein composition of a food. Therefore, a standard, repeatable composition cannot be easily ensured for a reference material (Tomoskozi et al., 2014).

The number of Proficiency Tests for food allergens that were organised so far is limited compared to the needs for accurate detection of allergens. However, Proficiency Tests have been conducted for a variety of matrices, as well as for different targeting allergens. What is

more, either spiked, or incurred samples that are more realistic have been tested. All these studies performed so far report in general problems related to allergens analysis in food. There are difficulties with the multimodality of the data reported by the participants and their statistical evaluation. Another common issue reported is the fact that quantitative results acquired in a PT are strongly dependent on specific factors like the different commercial ELISA kit brands and the analyte of interest. The need for development of reliable reference materials for the different allergens is, also, underline in order comparable and reproducible results to be produced (Alexopoulos et al., 2014, Köppel et al., 2014, Monaci & Visconti, 2010, Owen & Gilbert 2009, Scharf et al., 2013, Sykes et al., 2012). .

1.8 Aim of the study

The number of Proficiency Tests that were organised so far is not sufficient compared to the needs for reliable detection of allergens. Moreover, the problems encountered in relation to organisation of Proficiency Tests are various and important and result in difficulties in interpretation of the PT results and, therefore, in evaluation of laboratories. Taking all the above into consideration, new Proficiency Tests need to be organised to acquire deeper knowledge on how specific problems can be overcome, to better identify the essential parameters that can lead to a successfully organised Proficiency Test and to have an in depth insight in the current status of allergen testing worldwide.

In this study an effort is made towards an organisation of a new Proficiency Test for food allergens with participants from different continents. The interest of this Proficiency Test is focused on five important allergens, namely almond, gluten, peanut, milk and soy. The main aim of this test is the investigation of the current status of detection and quantification performed from analytical laboratories worldwide for these five allergens. In addition to this, the general performance of the laboratories on reporting results for food allergen analysis, i.e. their quality assurance system, is assessed. The effectiveness of the commonly applied methods for allergen analysis is discussed. The compliance of the levels of allergens that laboratories can measure with the limits determined in legislation for allergen labelling is also investigated. In general, other points, such as sources of unexpected or insufficient results and various difficulties encountered are also discussed.

2. MATERIALS AND METHODS

2.1 Materials

This Proficiency Test focuses on five food allergens, namely almond (AL), gluten (GL), peanut (PN), skim milk powder (SMP) and soy flour (SF). The products used for every of these allergens were purchased by local retailers and they are the following:

- For almond: Roasted almonds, 1 × "Amandelen geroosterd", 175g, Hoogvliet
- For gluten: Wheat flour, 1 × "Patent tarwebloem", 500g, Albert Heijn
- For peanut: Unsalted, raw peanuts, 1 × "Pinda's ongezouten", 235g, Duyvis
- For skim milk powder: Skim milk powder, 1 × "elk, Magere Melkpoeder", 240g, Campina
- For soy flour: Soy flour, 1 × "Sojameel", 200g, JOANNUSMOLEN

The material selected as basis for the preparation of the PT samples (blank matrix) was organic baby food based on mixed cereals (mainly maize and buckwheat):

- 40 × "BIOBIM Banana", mixed cereals, 200g, BIOBIM (in total 8kg "BIOBIM Banana" were used)

This material was selected as it was described to be free from milk/lactose, peanuts, almonds, wheat/gluten (http://www.glutenvrijemarkt.com/biobim-banana.html), four out of five allergens of interest.

The ELISA kits that were used for homogeneity and stability testing of the materials prepared for this PT are:

- Almond ELISA Kit (BioFront Technologies)
- Alertox ELISA Allergen-Almond (Biomedal Diagnostics)
- RIDASCREEN® Gliadin (R-Biopharm)
- GlutenTox ELISA Sandwich (Biomedal Diagnostics)
- RIDASCREEN®FAST Peanut (R-Biopharm)
- RIDASCREEN®FAST Milk (R-Biopharm)
- RIDASCREEN®FAST Soya (R-Biopharm)

Materials and equipment as described in the instruction manual of every ELISA kit were also used. Luminex MAGPIX was also used, with materials and equipment as described in the in-

house protocol developed at RIKILT (SOP A 1217: FOOD AND FOOD PRODUCTS – Determination of peanut, hazelnut and soy – Luminex).

Further materials and devices used are:

- GRINDOMIX GM200, Retsch), for cryogenic grinding
- La Moulinette, Type DPA1, Moulinex, for "BIOBIM Banana" grinding
- Sartorius LE4202S scale, for weighing
- Plastic, 60ml, screw-cap closed containers, for storing the sub-portions of the three materials prepared for the PT
- Sticker labels, for identification of containers of the sub-portions of the three materials prepared for the PT

2.2 Methods

2.2.1. Sample preparation

Almonds and peanuts before use were subjected to cryogenic grinding with addition of liquid nitrogen, to acquire a fine powder. The total amount of "BIOBIM Banana" was subjected to grinding before use to avoid phase separation in the final samples as the baby food was a mixture of soft flakes and fine powder.

With the materials above, three different test materials (A, B, C), containing 3 of the 5 allergens each, in different combinations, were prepared for the needs of this PT. The composition and the target amounts of allergens of the 3 materials are presented in Table 2. The three PT materials were all prepared by spiking the blank baby food matrix with the allergens in the form of powder. The spiking/target amounts were selected based on the VITAL System described in the Introduction. The action levels for almond, gluten, peanut, skim milk powder and soy flour were calculated for a serving size of 25g which is a prescribed quantity of the baby food for 8-month-old babies (according to the instructions attached to "BIOBIM Banana" package, 1 soupspoon of "BIOBIM Banana" weighs 3g, and the dosage to prepare 200cc cream for an 8-month-old baby is 7-8 soupspoons). Based on the VITAL Action Level 2 (requirement for precautionary statement) (http://allergenbureau.net/) calculated for a serving size of 25g (Table 3) and the VITAL Typical Protein Reference Table (http://allergenbureau.net/wp-content/uploads/2013/11/VITAL_Allergenic_Protein_Levels_10_11_08.pdf) the action levels in mg/kg whole compound for the five allergens (Table 4) were calculated. For gluten

the limit 20mg/kg which defines the gluten-free products was taken into account. Having the latter as basis the final spiking levels were decided, which were concentrations two times higher, the same, half or much lower compared to Action level for the whole compound. In this way a range of different concentrations was tested.

Table 2: Composition and target amount of allergens of materials prepared for the Proficiency Test.

	Target amount (mg/kg)				
Material	Almond	Gluten/wheat	Peanut	Skim milk powder	Soy flour
A	20	-	5	20	-
B	10	113 (wheat)/ 10 (gluten)	-	-	25
C	-	340 (wheat)/ 30 (gluten)	15	5	-

Table 3: Action levels (protein in mg/kg) for the five allergens for Reference amount/Serving size of 25g, as calculated by VITAL.

	Action level (mg/kg) where Reference amount/Serving size is 25g	
Allergen	Action level 1	Action level 2
Almond	<4mg/kg	≥4mg/kg
Gluten-containing cereals (total)	<20mg/kg	≥20mg/kg
Milk	<4mg/kg	≥4mg/kg
Peanut	<8mg/kg	≥8mg/kg
Soy	<20mg/kg	≥40mg/kg

Table 4: Action levels in mg/kg whole compound for the five allergens for Reference amount/Serving size of 25g, as calculated by VITAL.

Allergen	Action level (mg/kg)
Almond	20
Peanut	32
Skim milk powder	11
Soy flour	111

The concentrations of the 3 different combinations were finally achieved through serial dilutions. The first step of dilution was done with a pestle and mortar. At this step the materials were mixed in quantities as presented in Table 5. A small quantity of blank matrix was initially spread in the mortar to avoid sticking of allergenic compounds in the inner walls of the mortar. Then a quantity of the allergenic compounds is added, which is followed by addition of equal amount of blank matrix. Afterwards 8.5g of every of the three 50g - materials produced were diluted stepwise in 1691.5g blank matrix to reach the final concentration for every allergen. In the end 1.7kg of each of the three materials were acquired which were then subjected to further homogenisation under cryogenic conditions with liquid nitrogen, according to in-house standard operating procedures.

Table 5: First step dilution for the preparation of the PT materials.

Concentration	Material		
	A	B	C
Almond (mg)	200	100	-
Peanut (mg)	50	-	150
Wheat flour (mg)	-	1130	3400
Skim milk powder (mg)	200	-	50
Soy flour (mg)	-	250	-
Blank matrix (g)	49.5	48.5	46.4
Total amount (g)	50	50	50

2.2.2. Sample identification

The three materials prepared were weighed and divided in sub-portions. The sub-portions were stored in plastic, 60ml containers, and coded with sticker labels. The coding of the sub-portions was done randomly through a website application developed at RIKILT for Proficiency Testing needs. The sub-portions were used for the needs of homogeneity and stability testing of the three materials, as well as for testing by the participating laboratories.

- For homogeneity testing, 60 containers (20 containers for every material = 10 containers + 10 spare containers) of 6g each, with the coding "**PT RIKILT Food Allergens 2014** A homogeneity 01-20", "**PT RIKILT Food Allergens 2014** B homogeneity 01-20" and "**PT RIKILT Food Allergens 2014** C homogeneity 01-20", were prepared in total and stored at room temperature.
- For stability testing, 120 containers (24 containers for every allergen = 12 containers + 12 spare containers) of 6g each, with the coding "**PT RIKILT Food Allergens 2014** A stability 01-48", "**PT RIKILT Food Allergens 2014** B stability 01-48", "**PT RIKILT Food Allergens 2014** C stability 01-24", were prepared in total. Containers with the coding "**PT RIKILT Food Allergens 2014** A stability 01-24", "**PT RIKILT Food Allergens 2014** B stability 01-24" and "**PT RIKILT Food Allergens 2014** C stability 01-12" were stored at -80^{o}C, while with coding "**PT RIKILT Food Allergens 2014** A stability 25-48", "**PT RIKILT Food Allergens 2014** B stability 25-48", "**PT RIKILT Food Allergens 2014** C stability 13-24" at room temperature.
- For participating laboratories, 105 containers (35 containers for every material) of 20g each were prepared in total and stored at room temperature. The coding of the samples finally delivered to the participants is "**PT RIKILT Food Allergens 2014** Allergens/2014/ , with the code, as presented in Table 6 in Supplements, in the end of the coding.

2.2.3. Homogeneity testing

The International Harmonized Protocol for Proficiency Testing of Analytical Laboratories (Thompson et al., 2006) and ISO 13528 (ISO 13528:2005(E), 2005) were followed for homogeneity testing. According to the latter, the between sample standard deviation (s_S) and the within-sample standard deviation (s_W) are compared with the standard deviation for proficiency assessment (σ_P), which is 25% of the grand mean of the homogeneity data (Owen & Gilbert, 2009). The suitability of the method applied for homogeneity testing is assessed

with the criterion $s_W < 0.5*\sigma_P$. A material is considered adequately homogeneous if the criterion $s_S < 0.3*\sigma_P$ is met. To assess these requirements, ten containers of every of the three materials were analysed in duplicate and in random order. The containers were randomly selected for this testing (Thompson et al., 2006). Homogeneity of material A and C was determined based on analysis of peanut, while of material B on analysis of almond. The homogeneity of every material was based on the analysis of only one of the three spiked allergens. The homogeneity of the other allergens was not tested based on the assumption that homogeneity of peanut and almond was considered adequate to ensure the sufficient homogeneity of the respective materials (Bremer & Elbers, 2014).

Luminex MAGPIX detection system was applied for peanut analysis for homogeneity testing of material A, according to the in-house protocol developed at RIKILT (SOP A 1217 : FOOD AND FOOD PRODUCTS – Determination of peanut, hazelnut and soy – Luminex). Almond ELISA Kit (BioFront Technologies) was applied for almond analysis for homogeneity testing of material B, according to the instructions provided in the kit. Luminex MAGPIX detection system was applied for peanut analysis for homogeneity testing of material C, according to the in-house protocol developed at RIKILT (SOP A 1217 : FOOD AND FOOD PRODUCTS – Determination of peanut, hazelnut and soy – Luminex). The results of the homogeneity testing were calculated in Microsoft Excel and GraphPad Prism 5.0.The coding of the containers from which material was used for homogeneity testing of the three materials is presented in Table 7 in Supplements. The statistical analysis of the homogeneity data acquired for every material was performed by a Microsoft Excel statistical tool used at RIKILT for assessing homogeneity of materials for Proficiency Test purposes based on ISO 13528:2005(E), 2005, Fearn & Thompson and 2001, Thompson, 2000.

2.2.4. Sample distribution

Twenty-two laboratories registered in total for participation in this Proficiency Test, of which 10 from Europe, 9 from America, 2 from Asia and 1 from Oceania. Each participant was assigned a random laboratory code, with a website application developed at RIKILT for Proficiency Testing needs. The laboratory codes given are:

- PT194, PT206, PT210, PT211, PT212, PT213, PT214, PT215, PT216, PT217, PT218, PT219, PT220, PT221, PT222, PT223, PT224, PT225, PT226, PT227, PT228, PT229.

The participants were asked beforehand to indicate which allergenic compounds are in their scope. A set of three randomly selected containers, one for every of the three materials prepared, packed in a carton box, was shipped by courier to every participant on the 11th of November 2014. The codification of the samples delivered to every participant is presented in Table 6 in the Supplements. Apart from the samples, all the participants received an Instruction letter (Document 1, Supplements), an acknowledgement of receipt form (Document 2, Supplements), as well as a form to report the methods applied for every allergen. Moreover, the participants received instructions via e-mail on how to report the final results.

2.2.5. Stability testing

The stability testing was also performed according to The International Harmonized Protocol for Proficiency Testing of Analytical Laboratories (Thompson et al., 2006) and ISO 13528 (ISO 13528:2005(E), 2005). Based on the latter, the occurrence of 'consequential instability' was determined. Consequential instability occurs when the average value of the samples stored at room temperature is more than $0.3\sigma_H$ below the average value of the samples stored at -80 °C. In such a case, the instability has a significant influence on the calculated z-scores. In addition, it was determined whether a statistically significant instability occurred by using a Students t-test (ISO 13528:2005(E), 2005, Thompson et al., 2006, Bremer & Elbers, 2014).

For the stability testing 24 containers of material A, 24 containers of material B, as well as 12 containers of material C were randomly selected and stored at -80^{o}C, as described above, on the 17th of November 2014. This was done based on the assumption that the allergenic compounds are sufficiently stable at this storage temperature. The remaining containers were stored at room temperature during the whole Proficiency Testing procedure. Official stability testing was performed for all the five allergenic compounds included in the Proficiency Testing separately. The stability of peanut and skim milk powder was tested in material A, of almond and soy flour in material B, and of gluten in material C, based on the spiking level of the allergens in every material and the quantification range of every kit. In every official stability testing six containers from each material that were stored at -80^{o}C and six containers from each material that were stored at room temperature were randomly selected and analysed for the respective allergenic compound in a single run of analysis. All the ELISA kits used for the official stability testing of the allergens were performed according to the instructions provided in every kit. Luminex MAGPIX detection system was applied for peanut analysis according to the in-house protocol developed at RIKILT (SOP A 1217: FOOD AND FOOD PRODUCTS – Determination of peanut, hazelnut and soy – Luminex).

The results of the stability testing were calculated in Microsoft Excel and GraphPad Prism 5.0. The statistical analysis of the stability testing data acquired for every allergen was performed by a Microsoft Excel statistical tool used at RIKILT for assessing stability of materials for Proficiency Test purposes.

- Stability testing for almond:

The stability of almond was tested with two different commercial ELISA kits, namely Almond ELISA Kit (BioFront Technologies), and Alertox ELISA Allergen-Almond (Biomedal Diagnostics). On the 4th of December 2014, 17 days after the samples were distributed to the participants, official stability test for almond with the Almond ELISA Kit (BioFront Technologies) was performed. The coding of the containers from which material was used for this testing is:

For the samples stored at -80°C: PT RIKILT Food Allergens 2014 B stability 07, PT RIKILT Food Allergens 2014 B stability 09, PT RIKILT Food Allergens 2014 B stability 20, PT RIKILT Food Allergens 2014 B stability 14, PT RIKILT Food Allergens 2014 B stability 22, PT RIKILT Food Allergens 2014 B stability 04.

For the samples stored at room temperature: PT RIKILT Food Allergens 2014 B stability 34, PT RIKILT Food Allergens 2014 B stability 44, PT RIKILT Food Allergens 2014 B stability 46, PT RIKILT Food Allergens 2014 B stability 32, PT RIKILT Food Allergens 2014 B stability 42, PT RIKILT Food Allergens 2014 B stability 30.

On the 5th of December 2014, 18 days after the samples were distributed to the participants, official stability test for almond with the Alertox ELISA Allergen-Almond (Biomedal Diagnostics) kit was performed. The coding of the containers from which material was used for this testing is the same as described for analysis with the Almond ELISA Kit (BioFront Technologies).

- Stability testing for gluten:

On the 12th of December 2014, 25 days after the samples were distributed to the participants, official stability test for gluten with the RIDASCREEN® Gliadin (R-Biopharm) kit, was performed. The coding of the containers from which material was used for this testing is:

For the samples stored at -80°C: PT RIKILT Food Allergens 2014 C stability 08, PT RIKILT Food Allergens 2014 C stability 03, PT RIKILT Food Allergens 2014 C stability 10, PT RIKILT Food Allergens 2014 C stability 01, PT RIKILT Food Allergens 2014 C stability 12, PT RIKILT Food Allergens 2014 C stability 05.

For the samples stored at room temperature: PT RIKILT Food Allergens 2014 C stability 18, PT RIKILT Food Allergens 2014 C stability 24, PT RIKILT Food Allergens 2014 C stability 17, PT RIKILT Food Allergens 2014 C stability 23, PT RIKILT Food Allergens 2014 C stability 14, PT RIKILT Food Allergens 2014 C stability 19.

For gluten, the GlutenTox ELISA Sandwich (Biomedal Diagnostics) kit was also applied two different times, for a preliminary testing to assess the suitability of the kit and for the stability testing of gluten.

- Stability testing for peanut:

The stability of peanut was tested with two different methods, i.e. the ELISA RIDASCREEN®FAST Peanut (R-Biopharm) kit, and the Luminex MAGPIX. On the 1st of December 2014, 14 days after the samples were distributed to the participants, official stability test for peanut with the Luminex MAGPIX method was performed. The coding of the containers from which material was used for this testing is:

For the samples stored at -80°C: PT RIKILT Food Allergens 2014 A stability 11, PT RIKILT Food Allergens 2014 A stability 24, PT RIKILT Food Allergens 2014 A stability 12, PT RIKILT Food Allergens 2014 A stability 23, PT RIKILT Food Allergens 2014 A stability01, PT RIKILT Food Allergens 2014 A stability 09.

For the samples stored at room temperature: PT RIKILT Food Allergens 2014 A stability28, PT RIKILT Food Allergens 2014 A stability 40, PT RIKILT Food Allergens 2014 A stability 41, PT RIKILT Food Allergens 2014 A stability 34, PT RIKILT Food Allergens 2014 A stability 25, PT RIKILT Food Allergens 2014 A stability 47.

On the 17th of December 2014, 30 days after the samples were distributed to the participants, official stability test for peanut with the RIDASCREEN®FAST Peanut (R-Biopharm) kit was performed. The coding of the containers from which material was used for this testing is:

For the samples stored at -80°C: PT RIKILT Food Allergens 2014 A stability 11, PT RIKILT Food Allergens 2014 A stability 01, PT RIKILT Food Allergens 2014 A stability 24, PT RIKILT Food Allergens 2014 A stability 06, PT RIKILT Food Allergens 2014 A stability 20, PT RIKILT Food Allergens 2014 A stability 05.

For the samples stored at room temperature: PT RIKILT Food Allergens 2014 A stability 47, PT RIKILT Food Allergens 2014 A stability 39, PT RIKILT Food Allergens 2014 A stability 29, PT RIKILT Food Allergens 2014 A stability 34, PT RIKILT Food Allergens 2014 A stability 41, PT RIKILT Food Allergens 2014 A stability 46.

- Stability testing for skim milk powder:

On the 15th of December 2014, 28 days after the samples were distributed to the participants, official stability test for skim milk powder with RIDASCREEN®FAST Milk (R-Biopharm), was performed. The coding of the containers from which material was used for this testing is:

For the samples stored at -80°C: PT RIKILT Food Allergens 2014 A stability 20, PT RIKILT Food Allergens 2014 A stability 06, PT RIKILT Food Allergens 2014 A stability 19, PT RIKILT Food Allergens 2014 A stability 07, PT RIKILT Food Allergens 2014 A stability 17, PT RIKILT Food Allergens 2014 A stability 05.

For the samples stored at room temperature: PT RIKILT Food Allergens 2014 A stability 26, PT RIKILT Food Allergens 2014 A stability 43, PT RIKILT Food Allergens 2014 A stability 32, PT46 RIKILT Food Allergens 2014 A stability 30, PT RIKILT Food Allergens 2014 A stability 38, PT RIKILT Food Allergens 2014 A stability 45.

- Stability testing for soy flour:

On the 8th of December 2014, 21 days after the samples were distributed to the participants, official stability test for soy flour with RIDASCREEN®FAST Soya (R-Biopharm), was performed. The coding of the containers from which material was used for this testing is:

For the samples stored at -80°C: PT RIKILT Food Allergens 2014 B stability 16, PT RIKILT Food Allergens 2014 B stability 11, PT RIKILT Food Allergens 2014 B stability 02, PT RIKILT Food Allergens 2014 B stability 23, PT RIKILT Food Allergens 2014 B stability 01, PT RIKILT Food Allergens 2014 B stability 18.

For the samples stored at room temperature: PT RIKILT Food Allergens 2014 B stability 45, PT RIKILT Food Allergens 2014 B stability 29, PT RIKILT Food Allergens 2014 B stability 25, PT RIKILT Food Allergens 2014 B stability 38, PT RIKILT Food Allergens 2014 B stability48, PT RIKILT Food Allergens 2014 B stability 37.

2.2.6. Blank matrix measurements

Randomly selected samples of blank matrix, "BIOBIM Banana", were analysed to determine their content of almond, gluten, peanut, milk protein and soy protein. The content of almond was measured by using two ELISA kits, namely the Almond ELISA Kit (BioFront Technologies), and the Alertox ELISA Allergen-Almond (Biomedal Diagnostics). The almond content was tested in two different measurements that took place in two different days with each of the two kits. Regarding gluten, one single measurement for determination of its concentration was performed, with the RIDASCREEN® Gliadin (R-Biopharm) kit. For peanut concentration two measurements with Luminex MAGPIX were performed in two different days. Peanut concentration was also estimated with the RIDASCREEN®FAST Peanut (R-Biopharm) in a single measurement. Milk protein concentration was determined through a single measurement of three samples with the RIDASCREEN®FAST Milk (R-Biopharm) kit. Soy protein content was estimated with the RIDASCREEN®FAST Soya (R-Biopharm) kit, in two different measurements that took place in two different days for this specific purpose.

2.2.7. Validation testing

The validation testing was performed for the following commercial ELISA kits: Almond ELISA Kit (BioFront Technologies), Alertox ELISA Allergen-Almond (Biomedal Diagnostics), RIDASCREEN® Gliadin (R-Biopharm), GlutenTox ELISA Sandwich (Biomedal Diagnostics), RIDASCREEN®FAST Peanut (R-Biopharm), RIDASCREEN®FAST Milk (R-Biopharm) and RIDASCREEN®FAST Soya (R-Biopharm). Within this preliminary validation testing, every test kit was checked with material prepared for the samples to be distributed to the participants of Proficiency Testing for analysis, as a kind of reference material, since the latters are of a known allergen concentration and they were found to be homogeneous. A range of different concentrations of every material were tested, always based on the quantification range of the ELISA kit and the given concentration of every allergen in the respective sample. The dilution scheme tested with every kit is presented in Tables 8, 9, 10, 11, 12 and 13 in the Supplements.

2.2.8. Results collection

The deadline for submitting the results was 2 weeks after receipt of the samples. The participating laboratories were asked to submit their final results in a web application used by RIKILT for reporting Proficiency Test results.

2.2.9. Statistical analysis of results of the Proficiency Testing

For the statistical evaluation of the quantitative results of the study the International Harmonized Protocol for the Proficiency Testing of Analytical Laboratories (Thompson et al., 2006), elaborated by ISO 13528 (ISO 13528:2005(E), 2005), were applied. The insights on robust statistics published by the Analytical Methods Committee (Analytical Methods Committee, 1989 (1), Analytical Methods Committee, 1989 (2)) were also considered. Application of robust statistics shows considerable advantages, since all values are taken into consideration. The outlying observations are retained, but they are given less weight. Moreover, when robust statistics are applied, the data do not have to show normal distribution, which is expected in a Proficiency Test (Analytical Methods Committee, 1989 (1), Analytical Methods Committee, 1989 (2)).

The final evaluation of the results was based on the calculation of the assigned value, the uncertainty of the assigned value, the standard deviation for proficiency assessment and the z-scores.

2.2.9.1. Assigned value (X)

The assigned value (X) was determined by using robust statistics (Analytical Methods Committee, 1989 (1), Analytical Methods Committee, 1989 (2), ISO 13528:2005(E), 2005). The assigned value in this Proficiency Test is a consensus value. The robust mean of the reported results of all participants, calculated from an iterative process that starts at the median of the reported results using a cut-off value depending on the number of results, was used as the assigned value (Analytical Methods Committee, 1989 (1), ISO 13528:2005(E), 2005, Thompson et al., 2006).

2.2.9.2. Uncertainty of the assigned value (u)

The uncertainty of the assigned value is calculated since its influence is important on the evaluation of the laboratories. In case that the uncertainty of the assigned value is high, this will lead to a high uncertainty of the calculated z_a-scores for the participating laboratories. In such a case, the evaluation could indicate unsatisfactory method performance for the participants, but without any cause within the laboratory. Therefore, there is a risk that illegitimate conclusions are drawn for the performance of the participating laboratories from the calculated z_a-scores if the uncertainty of the assigned value is not taken into account. The calculation of the uncertainty of the assigned value (the robust mean) is based on the estimation of the standard deviation of the assigned value and the number of values used for the calculation of the assigned value (Thompson et al., 2006), as presented in the following equation:

$$u = 1.25 * \frac{\hat{\sigma}}{\sqrt{n}}$$

where:

u = uncertainty of the assigned value;

n = number of values used to calculate the assigned value;

$\hat{\sigma}$ = the estimate of the standard deviation of the assigned value resulting from robust statistics.

According to ISO 13528 (ISO 13528:2005(E), 2005) the uncertainty of the assigned value (u) is negligible and therefore does not have to be included in the statistical evaluation if:

$$u \leq 0.3\sigma_P$$

where:

u = the uncertainty of the assigned value;

σ_P = standard deviation for proficiency assessment.

In case the above criterion for the uncertainty of the assigned value is not fulfilled, the uncertainty of the assigned value should be taken into account for the evaluation of the performance of the participants regarding the accuracy. In case the uncertainty is > $0.7\sigma_P$, the calculated z-scores should not be used for evaluation of laboratories' performance and are presented for information only.

2.2.9.3. Standard deviation for proficiency assessment (σ_P)

The use of the Horwitz equation, which is normally used in Proficiency Testing in general, to estimate the standard deviation for proficiency assessment of ELISA-based methods is not assessed as fit for purpose for food allergens. Therefore a target standard deviation for proficiency assessment (σ_P) of 25% was applied as appropriate in this Proficiency Test (ISO 13528:2005(E), 2005, Owen & Gilbert, 2009):

$$\sigma_P = 0.25*c$$

where:

σ_P = expected standard deviation in proficiency tests;

c = concentration of the analyte (mg/kg).

2.2.9.4. Performance assessment of participants with regard to the accuracy

In order to illustrate the performance of the participating laboratories with regard to the accuracy, z_a-scores are calculated according to ISO 13528 (ISO 13528:2005(E), 2005). The main idea of the z_a-scores application is that all the acquired scores by the participating laboratories become comparable. (Thompson et al., 2006). According to ISO guidelines, z_a-scores are classified as presented in Table 14:

Table 14: Classification of z_a-scores.

$\|z_a\| \leq 2$	**Satisfactory**
$2 < \|z_a\| < 3$	**Questionable**
$\|z_a\| \geq 3$	**Unsatisfactory**

In case the calculated uncertainty of the assigned value complies with the criterion described in §2.2.9.2, the uncertainty is negligible. In such a case the accuracy z-score is calculated from the following equation:

$$z_a = \frac{\bar{x} - X}{\sigma_P} \qquad \text{(Equation I)}$$

where:

z_a = accuracy z-score;

$\bar{x}$ = the average result of the laboratory;

X = assigned value;

σ_P = standard deviation for proficiency assessment.

However, if the uncertainty of the assigned value does not comply with the criterion mentioned in §2.2.9.2, it can have an influence on the evaluation of the laboratories. According to ISO 13528, in such a case no z-scores can be calculated if a consensus value is used as the assigned value. However, evaluation of the participating laboratories is of main importance, as it is justifying the participating laboratories' effort. Therefore in this case, the uncertainty is taken into account and the accuracy z-score is calculated as follows (ISO 13528:2005(E), 2005):

$$z'_a = \frac{\bar{x} - X}{\sqrt{\sigma_P^2 + u^2}} \qquad \text{(Equation II)}$$

where:

z'_a = accuracy z-score taking into account the uncertainty of the assigned value;

$\bar{x}$ = the average result of the laboratory;

X = assigned value;

σ_P = standard deviation for proficiency assessment;

u = uncertainty of the assigned value.

In the case that consequential instability of the proficiency test materials is observed, this can influence the evaluation of the laboratory performance. Therefore, in such a case the consequential instability is considered for the calculation of z-scores. Because instability only regards one side of the confidence interval (a decrease of the concentration) this correction only applies to the lower 2s limit and results in an asymmetrical confidence interval. In the case of a consequential instability the accuracy z-score for the laboratories that reported an

amount below the assigned value is corrected for this instability by using the following equation:

$$z_{ai} = \frac{\bar{x} - X}{\sqrt{\sigma_P^2 + \Delta^2}} \qquad \text{(Equation III)}$$

where:

z_{ai} = accuracy z-score taking into account the instability of the assigned value;

$\bar{x}$ = the average result of the laboratory;

X = assigned value;

σ_P = standard deviation for proficiency assessment;

Δ = difference between average concentration of compound stored at -70°C or at room temperature.

Lastly, there are cases in which the uncertainty of the assigned value does not comply with the criterion in §2.2.9.2, as well as consequential instability is observed. In such cases the z'_a-score for the laboratories that reported an amount below the assigned value is corrected for this instability by applying the following equation:

$$z'_{ai} = \frac{\bar{x} - X}{\sqrt{\sigma_P^2 + \Delta^2 + u^2}} \qquad \text{(Equation IV)}$$

where:

z'_{ai} = accuracy z-score taking into account the uncertainty and instability of the assigned value;

$\bar{x}$ = the average result of the laboratory;

X = assigned value;

σ_P = standard deviation for proficiency assessment;

Δ = difference between average concentration of compound stored at -70°C or at room temperature;

u = uncertainty of the assigned value.

3. RESULTS

3.1 Blank matrix measurements

The blank baby food matrix, which was used for the preparation of the three materials for this Proficiency Testing, was tested for the presence of the five allergens used for spiking (almond, gluten, peanut, skim milk powder, soy flour). An overview of the results for the content of blank baby food matrix in the different allergens analysed is given in Table 15. Almond was detected in low concentrations, approximately 0.5mg/kg, which can be characterised as traces. Regarding gluten, it was detected in a low (traces), though considerable concentration. The soy flour content was found to be in high, considerable level. The results for peanut were either zero or of very low concentration, while for skim milk powder the concentrations found were very low.

Table 15: Blank matrix ("BIOBIM Banana") content in different allergens.

Allergen	Method	Measurement (mg/kg)		Average (mg/kg)	Lowest standard of the kit (mg/kg)	+/-
		1st	2nd			
Almond	Almond ELISA Kit (BioFront Technologies)	0.5	0.2	0.4	0.2	+ (traces)
	Alertox ELISA Allergen-Almond (Biomedal Diagnostics)	0.9	0	0.5	0.5	+ (traces)
Gluten	RIDASCREEN® Gliadin (R-Biopharm)	3.7	-	3.7	5.0 (gluten)	- (traces)
Peanut	Luminex MAGPIX	0.01	0.1 0.1	0.05	-	-
	RIDASCREEN®FAST Peanut (R-Biopharm)	0	-	0	2.5	-
Soy flour	RIDASCREEN®FAST Soya (R-Biopharm)	14.4	13.0	13.7	6.4 (soy flour)	+
Skim Milk powder	RIDASCREEN®FAST Milk (R-Biopharm)	0.5 0.6 0.5	-	0.5	7.0 (skim milk powder)	-

3.2 Validation testing results

3.2.1. Validation testing for ELISA kits for detection of almond

Two different ELISA kits were used for the purposes of detection of almond in this PT, the Almond ELISA Kit (BioFront Technologies), and the Alertox ELISA Allergen-Almond (Biomedal Diagnostics). For Almond ELISA Kit (BioFront Technologies) the validation testing results are presented in Figure 1. The different dilutions of material A and B tested were detected with sufficient differentiation by the kit. The dilutions prepared based on both materials differed by 50%. The detected almond concentrations obtained for the respective dilutions differ also by approximately 50% (Table 8, Supplements). These differences are obvious also in Figure 1, in which the distances among either the red or the yellow diamond-shaped dots are almost equal. Moreover, the results for both materials tested fell within the linear part of the calibration curve created for this kit. This is depicted in Figure 1, in which the yellow and red coloured diamond-shaped dots, that represent the different dilutions of material A and B, lie mainly in the linear part of the trend line. For Alertox ELISA Allergen-Almond (Biomedal Diagnostics) kit the results are similar with the ones for the previous kit (Table 9, Supplements). The different dilutions tested based on material B were detected with sufficient differentiation by the kit, and the results fell within the linear part of the calibration curve of the kit, as depicted with the red diamond-shaped dots in Figure 2.

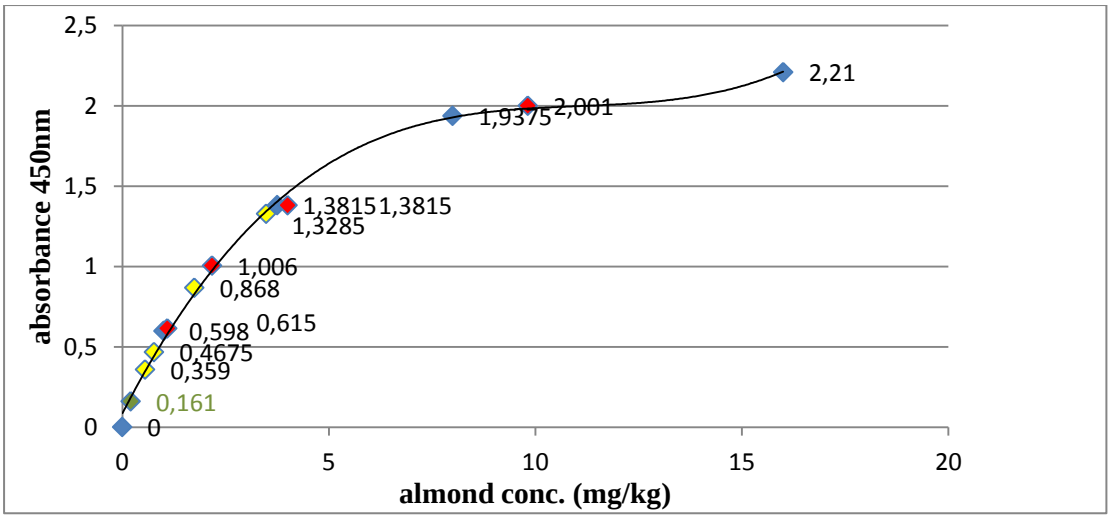

Figure1: Results of validation testing for Almond ELISA Kit (BioFront Technologies). The standard curve was constructed by analysing the standards included in the kit (blue diamond-shaped dots). The red diamond-shaped dots represent the results of almond dilutions of material A, the yellow the results of almond dilutions of material B tested, and the green represents the result of the lowest concentration standard of the kit.

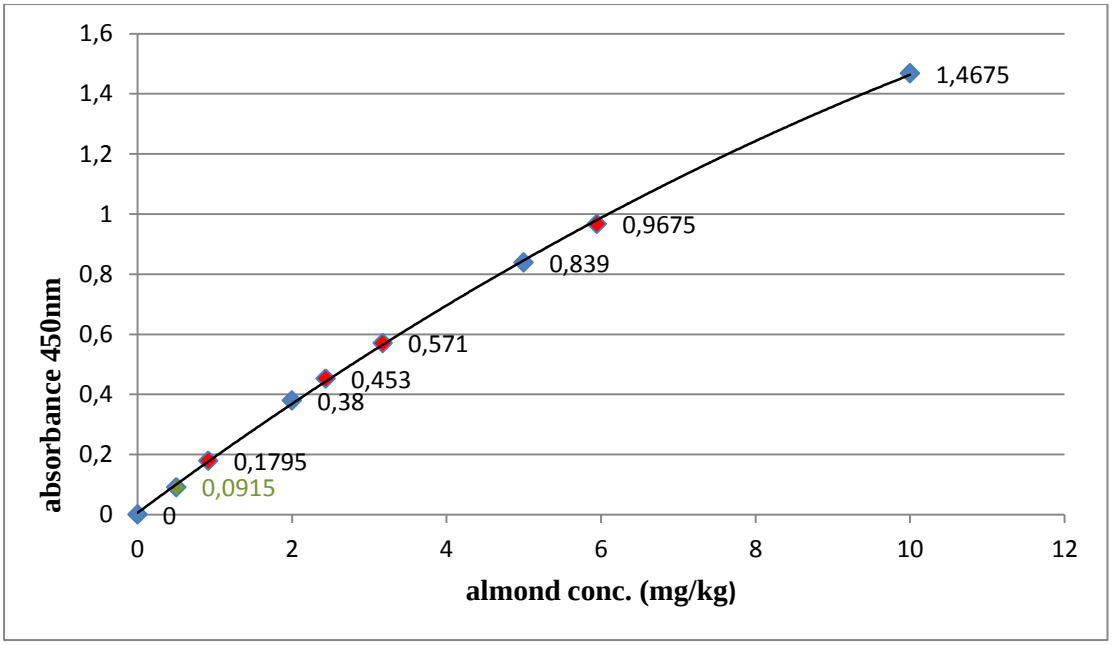

Figure 2: Results of validation testing for Alertox ELISA Allergen-Almond (Biomedal Diagnostics) kit. The standard curve was constructed by analysing the standards included in the kit (blue diamond-shaped dots). The red diamond-shaped dots in the graph represent the results of almond dilutions of material B tested, and the green the result of the lowest concentration standard of the kit.

3.2.2. Validation testing for ELISA kit for detection of gluten

For the purposes of detection of gluten the RIDASCREEN® Gliadin (R-Biopharm) kit, was selected. The results obtained show that this ELISA kit sufficiently differentiated the different dilutions of material C tested (Table 10, Supplements), as explained above for almond. The results of the dilutions tested (red diamond-shaped dots) fell within the linear part of the calibration curve of the kit (Figure 3).

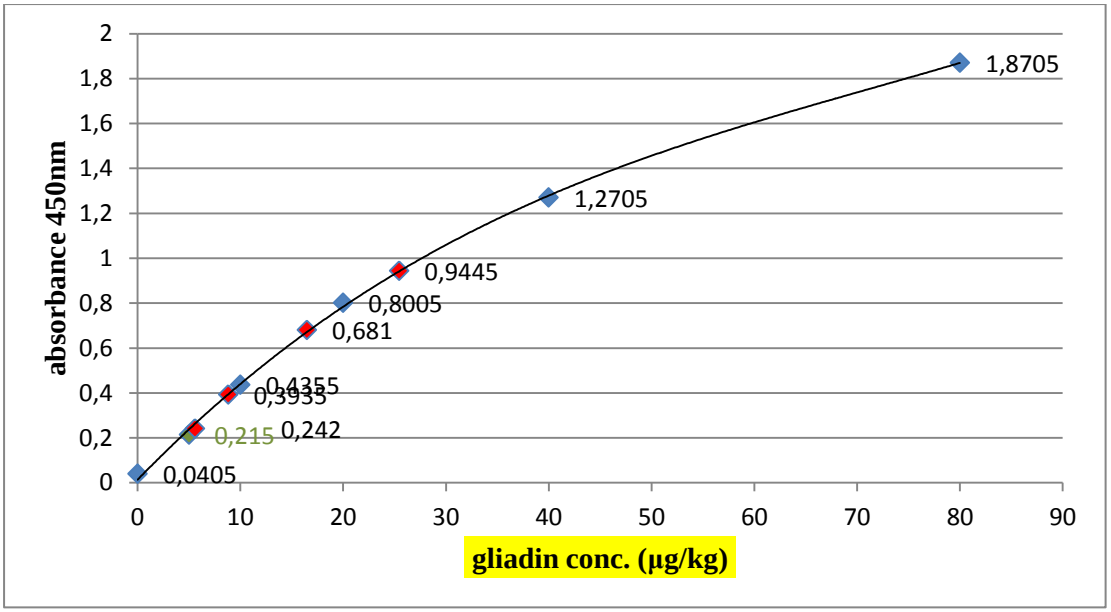

Figure 3: Results of validation testing for RIDASCREEN® Gliadin (R-Biopharm) kit. The standard curve was constructed by analysing the standards included in the kit (blue diamond-shaped dots). The red diamond-shaped dots in the graph represent the results of gluten dilutions of material C tested, and the green represents the result of the lowest concentration standard of the kit.

For the purposes of gluten detection, except for the RIDASCREEN® Gliadin (R-Biopharm) kit, the GlutenTox ELISA Sandwich (Biomedal Diagnostics) kit was also used. This kit was applied for gluten detection two different times. The results acquired from the first measurement showed that the kit was problematic in detection and quantification of gluten. More specifically, the absorbance values acquired for the gliadin standards at 450nm (Table 16) were much lower (approximately ten times lower for the highest standard) compared to the ones expected according to the instruction leaflet provided by the supplier. The results of the second measurement were similar to the ones reported for the first (Table 16). For this reason, further calculation of results for the samples tested and validation testing were not possible.

Table 16: O.D. 450nm values for gliadin standard solutions from two different measurements with GlutenTox ELISA Sandwich (Biomedal Diagnostics).

GlutenTox gliadin standards (ng/ml)	Expected mean O.D. 450nm	Reported mean O.D. 450nm	
		1st measurement	2nd measurement
1.56	0.29	0.07	0.07
3.12	0.42	0.08	0.08
6.25	0.75	0.11	0.08
12.5	1.22	0.16	0.13
25	1.97	0.24	0.19

3.2.3. Validation testing for ELISA kit for detection of peanut

RIDASCREEN®FAST Peanut (R-Biopharm) kit, was selected for detection of peanut in this Proficiency Testing. For this kit, validation testing took place with samples from material A. The results are presented in Figure 4. As described for almond, the different dilutions of peanut tested were detected with sufficient differentiation (Table 11, Supplements) and the results fell within the linear part of the calibration curve of the kit. However, two of the dilutions tested gave results that fall at the lowest part of the calibration curve with values lower than the one acquired for the lowest standard of the kit. The results for the undiluted concentration were sufficiently within the linear part.

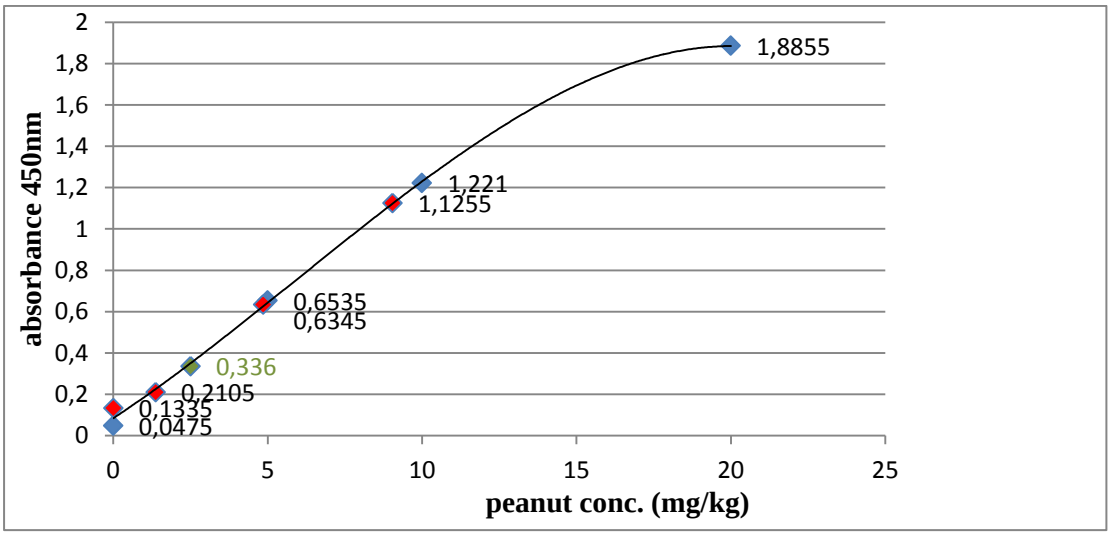

Figure 4: Results of validation testing for RIDASCREEN®FAST Peanut (R-Biopharm) kit. The standard curve was constructed by analysing the standards included in the kit (blue diamond-shaped dots). The red diamond-shaped dots in the graph represent the results of peanut dilutions of material A tested, and the green represents the result of the lowest concentration standard of the kit.

3.2.4. Validation testing for ELISA kit for detection of skim milk powder

RIDASCREEN®FAST Milk (R-Biopharm), was the kit of choice for detection of milk in this Proficiency Testing. For this, validation testing was performed with samples from material A. The results, as presented in Figure 5, show that RIDASCREEN®FAST Milk (R-Biopharm) kit can sufficiently detect the range of different dilutions tested (Table 12, Supplements). However, it is important to notice that the results acquired fall at the lowest part of the calibration curve and give an absorbance lower compared to the lowest standard provided in the kit. Only the undiluted sample of material A fell in the linear part.

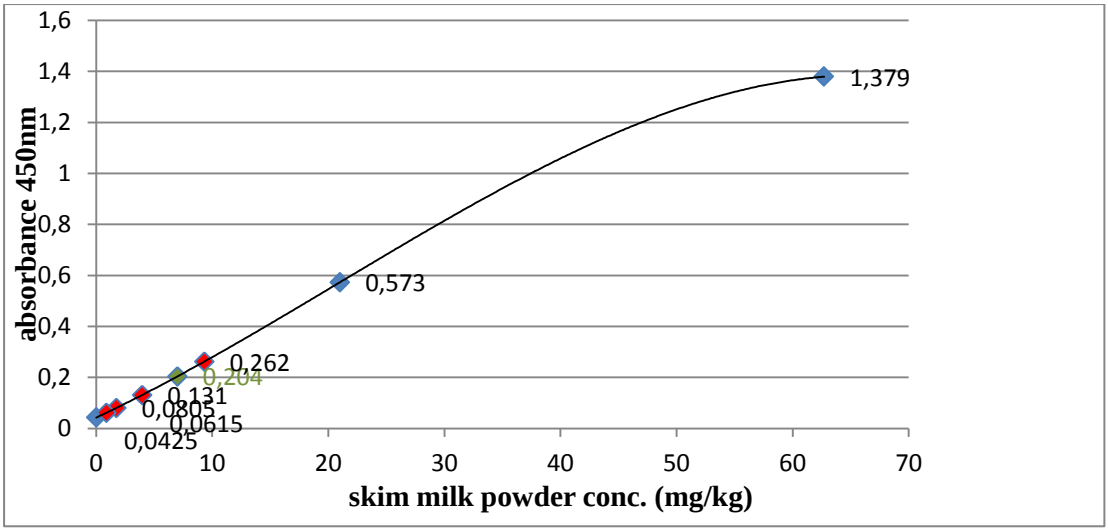

Figure 5: Results of validation testing for RIDASCREEN®FAST Milk (R-Biopharm) kit. The standard curve was constructed by analysing the standards included in the kit (blue diamond-shaped dots). The red diamond-shaped dots in the graph represent the results of skim milk powder dilutions of material A tested, and the green represents the result of the lowest concentration standard of the kit.

3.2.5. Validation testing for ELISA kits for detection of soy flour

RIDASCREEN®FAST Soya (R-Biopharm) kit, was selected for detection of soy. For this kit, validation testing took place with samples from material B. Figure 6 presents the results found. Most of the different dilutions tested were detected with sufficient differentiation by the kit, except for the lowest dilution tested (Table 13, Supplements). The results for the two lowest dilutions prepared fell in the linear part of the calibration curve of the kit, but lower compared to the lowest standard. The rest dilutions were sufficiently quantified within the linear part of the curve.

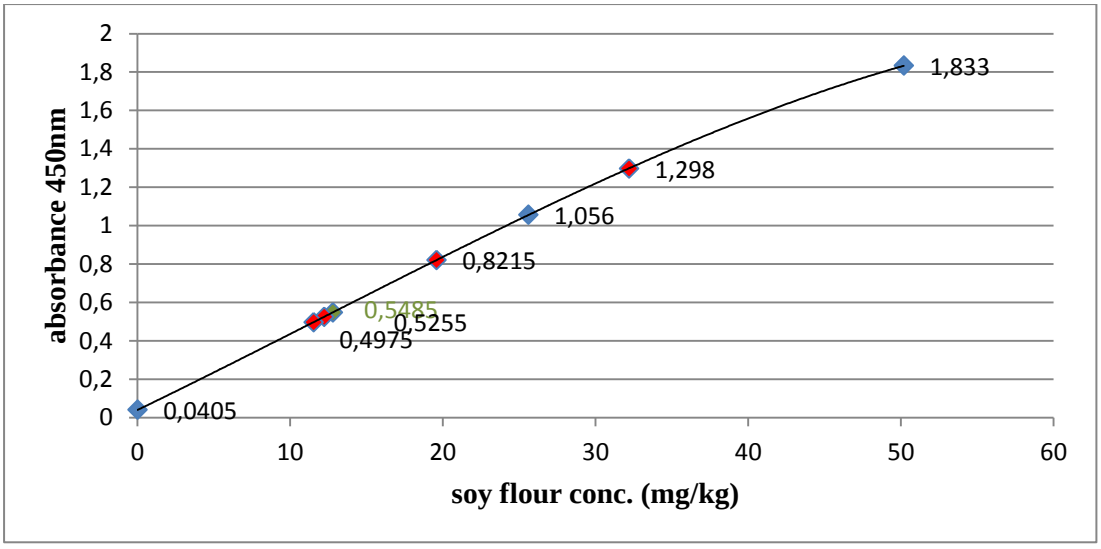

Figure 6: Results of validation testing for RIDASCREEN®FAST Soya (R-Biopharm) kit. The standard curve was constructed by analysing the standards included in the kit (blue diamond-shaped dots). The red diamond-shaped dots in the graph represent the results of soy flour dilutions of material B tested, and the green represents the result of the lowest concentration standard of the kit.

3.3 Homogeneity testing results

Luminex MAGPIX, applied for homogeneity testing of materials A and C, was assessed as suitable for application in the homogeneity testing, since the criterion $s_w<0.5* \sigma_p$ was met. For Almond ELISA Kit (BioFront Technologies), applied for homogeneity testing of material B, the same criterion was not fulfilled. Material A was found to be sufficiently homogeneous for peanut (Table 17). Material B was found to be sufficiently homogeneous for almond (Table 18). For both materials A and B the criterion $s_S< 0.3*\sigma_P$ was met. Based on the assumption described in "Materials and methods" for homogeneity testing, material A was considered homogeneous also for skim milk powder and almond, and material B was considered homogeneous also for soy flour and gluten. Consequently, materials A and B could be used as material for analysis by the participants in this Proficiency Testing.

On the other hand, the results acquired for material C initially prepared were not satisfactory (Table 19). It is obvious that there were huge differences in peanut concentrations found and based on the statistical analysis the samples tested were not sufficiently homogeneous. Therefore, material C was assessed as not sufficiently homogeneous for peanut and consequently for skim milk powder and gluten. This material is not acceptable to be used for analysis by the participants in this Proficiency Testing and a second effort was made to prepare a new material C. Among the results acquired from the homogeneity study for the new material C (Table 20), there is a single outlying pair with individual outlying results, i.e.

for the sample with the coding HOM C3 tested the results were: Extraction a= 35mg/kg peanut, Extraction b= 30mg/kg peanut. In this pair the results are discordant with one another and according to Cochran's variance test are characterised as outliers. These discordant results should be excluded from the respective statistical evaluation (Thompson et al., 2006). When these outlying results are excluded there are no outliers anymore and material C was assessed as sufficiently homogeneous for peanut. Material C was considered homogeneous also for skim milk powder and gluten. Finally, material C was acceptable material for analysis for the participants in this Proficiency Testing.

Table 17: Statistical evaluation of homogeneity data of material A based on peanut analysis.

Sample coding	Peanut (mg/kg) Extraction a	Peanut (mg/kg) Extraction b
HOM A1	10	10
HOM A2	9	9
HOM A3	9	10
HOM A4	10	9
HOM A5	10	10
HOM A6	9	9
HOM A7	9	8
HOM A8	9	9
HOM A9	9	9
HOM A10	8	9
Grand mean	9.20	
Cochran's test		
C	0.250	
C_{crit}	0.602	
$C < C_{crit}$?	NO OUTLIERS	
Target s=0.25*grand mean	2.300	
s_x	0.54	
s_w	0.45	
s_s	0.43	
Critical= 0.3*target s	0.69	
S_s<critical?	ACCEPTED	
S_w<0.5*target s?	ACCEPTED	

s_x =standard deviation of the sample averages, s_w =within-sample standard deviation, s_s =between-sample standard deviation

Table 18: Statistical evaluation of homogeneity data of material B based on almond analysis.

Sample coding	Almond (mg/kg) Extraction a	Almond (mg/kg) Extraction b
HOM B1	5	5
HOM B2	4	5
HOM B3	4	6
HOM B4	4	5
HOM B5	4	6
HOM B6	4	5
HOM B7	5	4
HOM B8	5	4
HOM B9	5	4
HOM B10	5	5
Grand mean	4.70	
Cochran's test		
C	0.286	
C_{crit}	0.602	
$C < C_{crit}$?	NO OUTLIERS	
Target s=0.25*grand mean	1.175	
s_x	0.26	
s_w	0.84	
s_s	0.00	
Critical= 0.3*target s	0.35	
S_s<critical?	ACCEPTED	
S_w<0.5*target s?	NOT ACCEPTED	

s_x =standard deviation of the sample averages, s_w =within-sample standard deviation, s_s =between-sample standard deviation

Table 19: Statistical evaluation of homogeneity data of the first material C prepared based on peanut analysis.

Sample coding	Peanut (mg/kg) Extraction a	Peanut (mg/kg) Extraction b
HOM C1	86	89
HOM C2	29	28
HOM C3	30	3
HOM C4	25	24
HOM C5	66	65
HOM C6	53	52
HOM C7	39	35
HOM C8	92	106
HOM C9	43	43
HOM C10	24	23
Grand mean	49.15	
Cochran's test		
C	0.863	
C_{crit}	0.602	
$C < C_{crit}$?	OUTLIERS DETECTED	
Target s=0.25*grand mean	12.288	
s_x	26.79	
s_w	3.37	
s_s	26.68	
Critical= 0.3*target s	3.69	
S_s<critical?	NOT ACCEPTED	
S_w<0.5*target s?	ACCEPTED	

s_x =standard deviation of the sample averages, s_w =within-sample standard deviation, s_s =between-sample standard deviation

Table 20: Statistical evaluation of homogeneity data of the final material C prepared based on peanut analysis.

Sample coding	Peanut (mg/kg) Extraction a	Peanut (mg/kg) Extraction b
HOM C1	33	33
HOM C2	35	35
HOM C3	*	*
HOM C4	34	33
HOM C5	31	31
HOM C6	31	32
HOM C7	31	30
HOM C8	28	29
HOM C9	29	31
HOM C10	32	34
Grand mean	31.85	
Cochran's test		
C	0.676	
C_{crit}	0.602	
$C < C_{crit}$?	NO OUTLIERS	
Target s=0.25*grand mean	7.963	
s_x	1.92	
s_w	1.36	
s_s	1.66	
Critical= 0.3*target s	2.39	
S_s<critical?	ACCEPTED	
S_w<0.5*target s?	ACCEPTED	

s_x =standard deviation of the sample averages, s_w =within-sample standard deviation, s_s =between-sample standard deviation
*single outlying pair with individual outlying results (Extraction a= 35mg/kg peanut, Extraction B= 30mg/kg peanut)

3.4 Stability testing results

The results of stability testing for almond performed with the Almond ELISA Kit (BioFront Technologies) and their statistical evaluation are presented in Table 21. Consequential instability, no statistic instability and no significant difference in standard deviation were detected for almond in material B in measurements with this specific kit. Stability testing of almond performed with the Alertox ELISA Allergen-Almond (Biomedal Diagnostics) kit, resulted in detection of consequential instability, no statistic instability and significant difference in standard deviation (Table 22). Both ELISA kits applied for almond gave concordant results, i.e. that there is consequential instability for almond. Stability testing of gluten with RIDASCREEN® Gliadin (R-Biopharm) kit resulted in detection of consequential instability, statistic instability, as well as no significant difference in standard deviation in

material C (Table 23). The stability of peanut was tested by using the ELISA kit RIDASCREEN®FAST Peanut (R-Biopharm) and the Luminex MAGPIX. Results of stability testing for peanut in material A with both methods applied (Tables 24, 25) include no consequential, no statistic instability and no significant difference in standard deviation. The stability of skim milk powder, tested with the ELISA kit RIDASCREEN®FAST Milk (R-Biopharm) in material A, resulted in consequential instability, but no statistic instability and no significant difference in standard deviation (Table 26). The stability of soy flour was tested with the RIDASCREEN®FAST Soya (R-Biopharm) kit in material B. No consequential and no statistic instability, as well as no significant difference in standard deviation were detected (Table 27).

Table 21: Statistical evaluation of stability data for almond in material B (Almond ELISA Kit (BioFront Technologies) results).

Storage temperature	-80°C	RT
Time at RT (days)		17
Calculated amounts (mg/kg)	5	5
	5	4
	6	5
	5	6
	5	4
	5	4
Average amount (mg/kg)	5.2	4.7
n	6	6
Standard deviation (mg/kg)	0.408	0.816
Difference		0.5
$0.3*\sigma_P$		0.388
Consequential difference? **Diff < $0.3*\sigma_P$**		Consequential instability detected
t		1.34
t_{crit}		2.23
Statistical difference? $t < t_{crit}$		No statistic instability detected

Table 22: Statistical evaluation of stability data for almond in material B (Alertox ELISA Allergen-Almond (Biomedal Diagnostics) kit results).

Storage temperature	-80°C	RT
Time at RT (days)		18
Calculated amounts (mg/kg)	7	5
	7	7
	6	5
	6	5
	13	7
	8	5
Average amount (mg/kg)	7.8	5.7
n	6	6
Standard deviation (mg/kg)	2.639	1.033
Difference		2.17
$0.3*\sigma_P$		0.588
Consequential difference? Diff < $0.3*\sigma_P$		Consequential instability detected
t		1.87
t_{crit}		2.23
Statistical difference? $t < t_{crit}$		No statistic instability detected

Table 23: Statistical evaluation of stability data for gluten in material C (RIDASCREEN® Gliadin (R-Biopharm) kit results).

Storage temperature	-80°C	RT
Time at RT (days)		25
Calculated amounts (mg/kg)	35	26
	31	23
	26	22
	27	27
	29	25
	30	26
Average amount (mg/kg)	29.7	24.8
n	6	6
Standard deviation (mg/kg)	3.204	1.941
Difference		4.83
$0.3*\sigma_P$		2.225
Consequential difference? Diff < $0.3*\sigma_P$		Consequential instability detected
t		3.16
t_{crit}		2.23
Statistical difference? $t < t_{crit}$		Statistic instability detected

Table 24: Statistical evaluation of stability data for peanut in material A (Luminex MAGPIX results).

Storage temperature	-80°C	RT
Time at RT (days)		14
Calculated amounts (mg/kg)	8	9
	9	9
	10	9
	9	8
	9	9
	9	8
Average amount (mg/kg)	9,0	8,7
n	6	6
Standard deviation (mg/kg)	0,632	0,516
Difference		0,33
$0.3*\sigma_P$		0,675
Consequential difference? Diff < $0.3*\sigma_P$		No consequential instability detected
t		1,00
t_{crit}		2,23
Statistical difference? $t < t_{crit}$		No statistic instability detected

Table 25: Statistical evaluation of stability data for peanut in material A (RIDASCREEN®FAST Peanut (R-Biopharm) kit results).

Storage temperature	-80°C	RT
Time at RT (days)		30
Calculated amounts (mg/kg)	12	10
	11	11
	12	11
	11	10
	10	10
	10	11
Average amount (mg/kg)	11,0	10,5
n	6	6
Standard deviation (mg/kg)	0,894	0,548
Difference		0,50
$0.3*\sigma_P$		0,825
Consequential difference? Diff < $0.3*\sigma_P$		No consequential instability detected
t		1,17
t_{crit}		2,23
Statistical difference? $t < t_{crit}$		No statistic instability detected

Table 26: Statistical evaluation of stability data for skim milk powder in material A (RIDASCREEN®FAST Milk (R-Biopharm) kit results).

Storage temperature	-80°C	RT
Time at RT (days)		28
Calculated amounts (mg/kg)	10	7
	9	8
	10	10
	14	9
	7	7
	7	11
Average amount (mg/kg)	9,5	8,7
n	6	6
Standard deviation (mg/kg)	2,588	1,633
Difference		0,83
$0.3*\sigma_p$		0,713
Consequential difference? Diff < $0.3*\sigma_p$		Consequential instability detected
t		0,67
t_{crit}		2,23
Statistical difference? $t < t_{crit}$		No statistic instability detected

Table 27: Statistical evaluation of stability data for soy flour in material B (RIDASCREEN®FAST Soya (R-Biopharm) kit results).

Storage temperature	-80°C	RT
Time at RT (days)		21
Calculated amounts (mg/kg)	29	23
	23	27
	25	26
	26	22
	22	22
	25	22
Average amount (mg/kg)	25,0	23,7
n	6	6
Standard deviation (mg/kg)	2,449	2,251
Difference		1,33
$0.3*\sigma_p$		1,875
Consequential difference? Diff < $0.3*\sigma_p$		No consequential instability detected
t		0,98
t_{crit}		2,23
Statistical difference? $t < t_{crit}$		No statistic instability detected

3.5 Results of the Proficiency Testing

3.5.1. Methods applied by the participants

The total number of the laboratories that registered for participation in this Proficiency Testing is twenty-two. Only immunochemical methods, mainly ELISA, were applied by the participants. A general overview of the methods applied by the participants is given in Table 28 in Supplements.

For the detection of almond, sixteen out of twenty-two laboratories reported the application of one method for analysis. Fifteen of them reported quantitative and one (PT229) qualitative result for almond, but all applied ELISA kits. In total, seven different ELISA kits, from different commercial brands were used.

Seventeen laboratories reported applied methods for gluten measurement. Only one laboratory (PT225) applied three different methods for detection of gluten. Seventeen of the participants measured for gluten applied ELISA methods and reported quantitative results. The participant with the code PT225 reported also qualitative results by applying a lateral flow device. Ten different ELISA kits, from different commercial brands were used, as well as one lateral flow device.

Sixteen of the twenty-two laboratories reported applied methods for peanut. Only one method was applied for peanut detection by every participant. Sixteen of the participants reported quantitative and one (PT229) qualitative result for peanut, but all applied ELISA methods. In total, seven different ELISA kits, from different commercial brands were used.

For skim milk powder, sixteen of the twenty-two participating laboratories reported methods for detection. PT206 applied two different methods and PT225 applied four different methods, while the rest laboratories only one. Sixteen of the laboratories used ELISA methods and reported quantitative results. Only PT225 reported also qualitative results by using two different lateral flow devices. Twelve different ELISA kits, from different commercial brands were used, as well as two different lateral flow devices from the same brand.

Fourteen laboratories reported applied methods for the detection of soy flour. Only one method was applied by every participant and all participants used ELISA methods. All the laboratories reported quantitative results. In total, five different ELISA kits, from different commercial brands were used.

One participant (PT218) did not report the applied methods used for detection of the allergens, but did report quantitative results for gluten, peanut, skim milk powder and soy flour. This participant did not test for almond.

3.5.2. Results of the participants

An overview of the participants' results for materials A, B and C is presented in Tables 29, 30 and 31, respectively, in Supplements. More specifically, the results for the different allergens in mg/kg, as reported by the laboratories, as well as the calculated z-scores are included in these Tables. Moreover, for the available z-scores graphical representations are included for every allergen. For the allergens that z-scores could not be calculated an effort was made to assess the results based on the ELISA kit applied by the participants. A table presenting the false negative results (Table 32, Supplements) is also included.

3.5.2.1. Almond

In material A, false negative results were not reported by the participants. The lowest value reported is 1.6mg/kg and the highest 23mg/kg. The assigned value calculated is 14.2mg/kg, with a robust standard deviation of 3.4mg/kg which is lower than the target standard deviation of 3,5mg/kg. The uncertainty of the assigned value is significant, 1.1mg/kg, and the criterion $u \leq 0.3*\sigma_P$ is not met, therefore this uncertainty is considered in the statistical evaluation. Consequential instability of 0.5mg/kg was observed for almond during storage for 17 days with the Almond ELISA Kit (BioFront Technologies) (Table 21). Although stability of almond was measured with two different ELISA kits, the result of the Almond ELISA Kit (BioFront Technologies) is going to be used for the assessment of the results, since this kit was also used for the homogeneity testing and by many of the participants. Because the assigned value is of 14.2mg/kg level and the instability of 0.5mg/kg was calculated for concentrations of approximately 5.5mg/kg almond, a linear extrapolation was considered necessary. Therefore, the instability that was taken into account for the statistical evaluation of almond results in material A was 1mg/kg. Taking into account the uncertainty of the assigned value and the consequential instability calculated, z'ai-scores were reported as presented in Table 29 in the Supplements. Two of the results were questionable (PT219 and PT226) and one was unsatisfactory (PT222) (Table 29, Figure 7).

In material B, false negative results were not reported by the participants. The lowest value reported is 0.6mg/kg and the highest 20.8mg/kg. The assigned value calculated is 6.1mg/kg, with a robust standard deviation of 3.3mg/kg which is more than 2 times higher compared to

the target standard deviation of 1.5mg/kg. The uncertainty of the assigned value is significant, 1mg/kg, and the criterion $u \leq 0.3*\sigma_P$ is not met, therefore this uncertainty is considered in the statistical evaluation. Consequential instability of 0.5mg/kg was observed for almond during storage for 17 days with the Almond ELISA Kit (BioFront Technologies) (Table 21). Because the assigned value is of 6.1mg/kg level and the instability of 0.5mg/kg was calculated for concentrations of approximately 5.5mg/kg almond, linear extrapolation, as described previously, was not considered necessary. Therefore, the instability that was taken into account for the statistical evaluation of almond results in material B was 0.5mg/kg. Taking into account the uncertainty of the assigned value and the consequential instability calculated, z'ai-scores were reported as presented in Table 30 in the Supplements. Three of the results were questionable (PT194, PT211, PT222) and two were unsatisfactory (PT219, PT226) (Table 30, Figure 10).

Material C was not spiked with almond. Three positive results (PT194=1.33mg/kg, PT226=1mg/kg, PT229= +) were reported.

3.5.2.2. Gluten

Eighteen laboratories carried out quantitative analysis for gluten in every of the three materials. Material A was not spiked with gluten. Seven positive results (PT212=5.64mg/kg, PT213=5.3mg/kg, PT214=6mg/kg, PT215=82mg/kg, PT219=traces, PT220=10.4 mg/kg, PT226=10.1 mg/kg) were reported for gluten in this material. PT225 reported a negative qualitative result by using a lateral flow device.

In material B, PT225 reported also a positive qualitative result by using a lateral flow device. One false negative result was reported (PT218). The lowest value reported is 6.2mg/kg and the highest 466mg/kg. The assigned value calculated is 13.2mg/kg, with a robust standard deviation of 6.5mg/kg which is 2 times higher compared to the target standard deviation of 3.3mg/kg. The uncertainty of the assigned value is significant, 1.9mg/kg, and the criterion $u \leq 0.3*\sigma_P$ is not met, therefore this uncertainty is considered in the statistical evaluation. Consequential instability of 4.83mg/kg was observed for gluten during storage for 25 days with the RIDASCREEN® Gliadin (R-Biopharm) kit (Table 23). Because the assigned value is of 13.2mg/kg level and the instability of 4.83mg/kg was calculated for concentrations of approximately 30mg/kg gluten, linear extrapolation, as described previously, was considered necessary. Therefore, the instability that was taken into account for the statistical evaluation of gluten results in material B was 2.5mg/kg. Taking into account the uncertainty of the assigned value and the consequential instability calculated, z'ai-scores were reported as

presented in Table 30 in the Supplements. Three of the results (PT215, PT220, and PT226) were unsatisfactory (Table 30, Figure 11).

In material C, PT225 reported also a positive qualitative result by using a lateral flow device. One false negative result was reported (PT218). The lowest value reported is 18.9mg/kg and the highest 1678mg/kg. The assigned value calculated is 33.2mg/kg, with a robust standard deviation of 12.8mg/kg which is higher compared to the target standard deviation of 8.3mg/kg. The uncertainty of the assigned value is significant, 3.8mg/kg, and the criterion $u \leq 0.3*\sigma_P$ is not met, therefore this uncertainty is considered in the statistical evaluation. Consequential instability of 4.83mg/kg was observed for gluten during storage for 25 days with the RIDASCREEN® Gliadin (R-Biopharm) kit (Table 23). Because the assigned value is of 33.2mg/kg level and the instability of 4.83mg/kg was calculated for concentrations of approximately 30mg/kg gluten, linear extrapolation, as described previously, was not considered necessary. Therefore, the instability that was taken into account for the statistical evaluation of gluten results in material C was 4.83mg/kg. Taking into account the uncertainty of the assigned value and the consequential instability calculated, z'ai-scores were reported as presented in Table 31 in the Supplements. Three of the results (PT215, PT220, and PT226) were unsatisfactory (Table 31, Figure 13).

3.5.2.3. Peanut

In material A, two false negative results were reported by the participants (PT216, PT221). The lowest value reported is 1.96mg/kg and the highest 12mg/kg. The assigned value calculated is 6.2mg/kg, with a robust standard deviation of 3.9mg/kg which is more than two times higher than the target standard deviation of 1.6mg/kg. The uncertainty of the assigned value is very high, 1.3mg/kg, more than $0.7* \sigma_P$ (1.12mg/kg) and therefore the calculated z-scores should not be used for the evaluation of the performance of participating laboratories. Consequential instability was not observed during storage of either 14 or 30 days, with both methods applied, as described in stability testing results (Tables 24, 25). Based on the above, z'a-scores were calculated, which are presented for information only (Table 29, Figure 8) and not for evaluation because there is a risk that the laboratories will receive questionable or unsatisfactory z-scores due to inaccuracy of the assigned value and not as a result of the laboratory performance.

Material B was not spiked with peanut. In material B, false positive results were not reported for peanut.

In material C, false negative results were not reported. The lowest value reported is 2.84mg/kg and the 38.5highest mg/kg. The assigned value calculated is 18.5mg/kg, with a robust standard deviation of 11.4mg/kg which is more than two times higher compared to the target standard deviation of 4.6mg/kg. The uncertainty of the assigned value is very high, 3.7mg/kg, more than 0.7* σ_P (3.22mg/kg) and therefore the calculated z-scores should not be used for the evaluation of the performance of participating laboratories. Consequential instability was not observed during storage of either 14 or 30 days of storage, with both methods applied, as described in stability testing results (Tables 24, 25). Based on the above, z'a-scores were calculated, which are presented for information only (Table 31, Figure 14) and not for evaluation because there is a risk that the laboratories will receive questionable or unsatisfactory z-scores due to inaccuracy of the assigned value and not as a result of the laboratory performance.

3.5.2.4. Skim milk powder

In material A, laboratory PT206 applied two different methods; however PT206 reported one single value. PT225 reported two positive qualitative results by using two different lateral flow devices. One false negative result was reported (PT206). The lowest value reported is 0.5mg/kg and the highest 30.6mg/kg. The assigned value calculated is 8.7mg/kg, with a robust standard deviation of 6.7mg/kg which is more than three times higher than the target standard deviation of 2.2mg/kg. The uncertainty of the assigned value is very high, 2.1mg/kg, more than 0.7* σ_P (1.54mg/kg), and therefore the calculated z-scores should not be used for the evaluation of the performance of participating laboratories. Consequential instability of 0.83mg/kg was observed for skim milk powder during storage for 28 days with the RIDASCREEN®FAST Milk (R-Biopharm) kit (Table 26). Because the assigned value is 8.7mg/kg and the instability of 0.83mg/kg was calculated for concentrations of approximately 9mg/kg skim milk powder, linear extrapolation, as described previously, was not considered necessary. Therefore, the instability that was taken into account for the statistical evaluation of skim milk powder results in material A was 0.83mg/kg. Based on the above, z'ai-scores were calculated, which are presented for information only (Table 29, Figure 9) and not for evaluation because there is a risk that the laboratories will receive questionable or unsatisfactory z-scores due to inaccuracy of the assigned value and not as a result of the laboratory performance.

Material B was not spiked with skim milk powder. PT225 reported two negative qualitative results by using two different lateral flow devices. One false positive result was reported in this material, by laboratory PT213 (1.7mg/kg).

In material C, laboratory PT206 applied two different methods; however PT206 reported one single value. PT225 reported two positive qualitative results by using two different lateral flow devices. Six false negative results were reported (PT206, PT211, PT220, PT224, PT227, PT228). The lowest value reported is 0.6mg/kg and the highest 6.6mg/kg. The assigned value calculated is 3.1mg/kg, with a robust standard deviation of 2.8mg/kg which is more than three times higher than the target standard deviation of 0.8mg/kg. The uncertainty of the assigned value is very high, 1.1mg/kg, more than 0.7* σ_P (0.56mg/kg), and therefore the calculated z-scores should not be used for the evaluation of the performance of participating laboratories. Consequential instability of 0.83mg/kg was observed for skim milk powder during storage for 28 days with the RIDASCREEN®FAST Milk (R-Biopharm) kit (Table 26). Because the assigned value is 3.1mg/kg and the instability of 0.83mg/kg was calculated for concentrations of approximately 9mg/kg skim milk powder, linear extrapolation, as described previously, was considered necessary. Therefore, the instability that was taken into account for the statistical evaluation of skim milk powder results in material C was 0.3mg/kg. Based on the above, z'ai-scores were calculated, which are presented for information only (Table 31, Figure 15) and not for evaluation because there is a risk that the laboratories will receive questionable or unsatisfactory z-scores due to inaccuracy of the assigned value and not as a result of the laboratory performance.

3.5.2.5. Soy flour

Fifteen laboratories carried out quantitative analysis for soy flour in all three materials. Material A was not spiked with soy flour. Two positive results were reported in this material, by laboratories PT211 (12.6ppm), PT228 (9.1ppm).

In material B, seven false negative results were reported (PT206, PT213, PT214, PT216, PT218, PT220, and PT223). The lowest value reported is 2mg/kg and the highest 38.5mg/kg. The assigned value calculated is 16.2mg/kg, with a robust standard deviation of 17.3mg/kg which is more than four times higher than the target standard deviation of 4.1mg/kg. The uncertainty of the assigned value is very high, 8.2mg/kg, more than 0.7* σ_P (2.87mg/kg), and therefore the calculated z-scores should not be used for the evaluation of the performance of participating laboratories. Consequential instability was not observed for soy flour during storage for 21 days with the RIDASCREEN®FAST Soya (R-Biopharm) kit (Table 27). Based on the above, z'a-scores were calculated, which are presented for information only (Table 30, Figure 12) and not for evaluation because there is a risk that the laboratories will receive questionable or unsatisfactory z-scores due to inaccuracy of the assigned value and not as a result of the laboratory performance.

Material C was not spiked with soy flour. Two positive results were reported in this material, by laboratories PT211 (12.5ppm), PT228 (14.1ppm).

3.5.2.6. Grouping of results based on ELISA kits

The uncertainty of the assigned value calculated for peanut, skim milk powder and soy flour in the respective spiked materials was found to be very high, more than 0.7* σ_P. In such a case, the calculated z-scores should not be used for the evaluation of the performance of participating laboratories, but are presented for information only. As an alternative approach, the results reported from participants for these three allergens were also grouped by the different ELISA kits used (Figures 16, 17, 18, 19, 20). In these Figures, only quantitative results, >0mg/kg, were included;

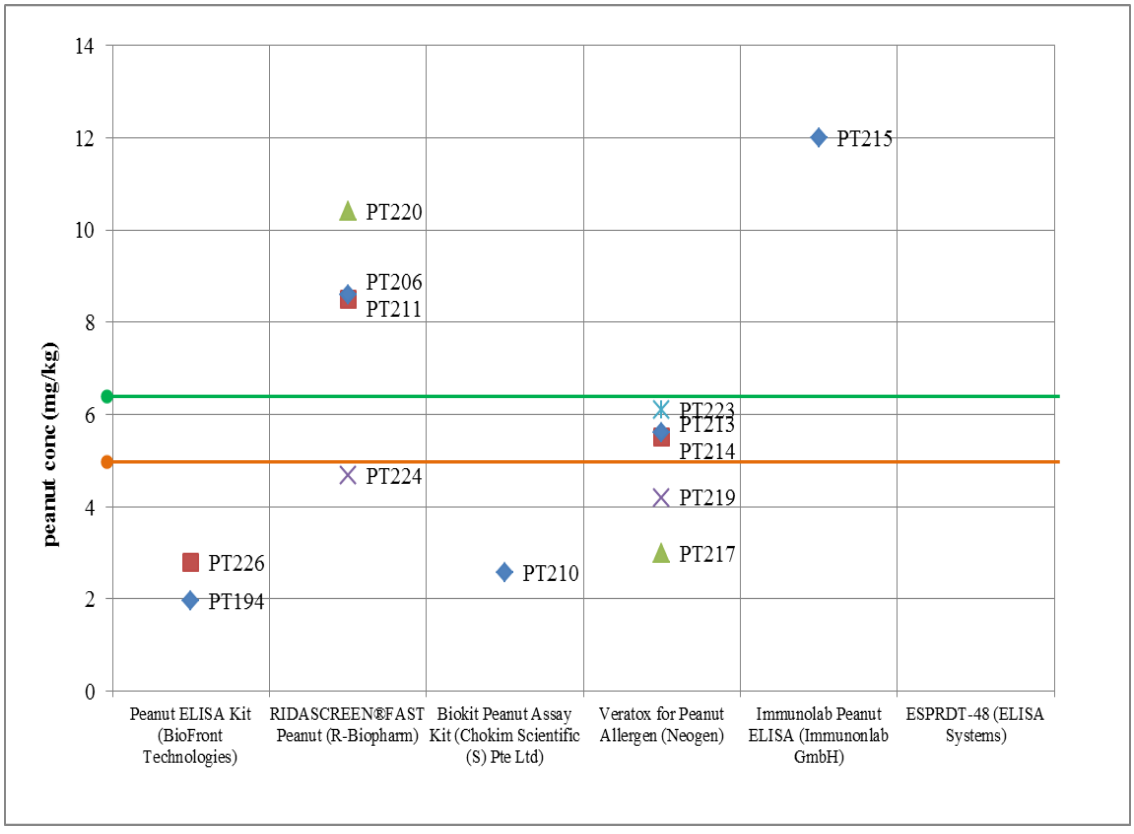

Figure 16: Graphical representation of results reported for peanut in material A, against the different kits used by the participants. The green line represents the assigned value (6.2mg/kg), and the orange line the target level (5mg/kg).

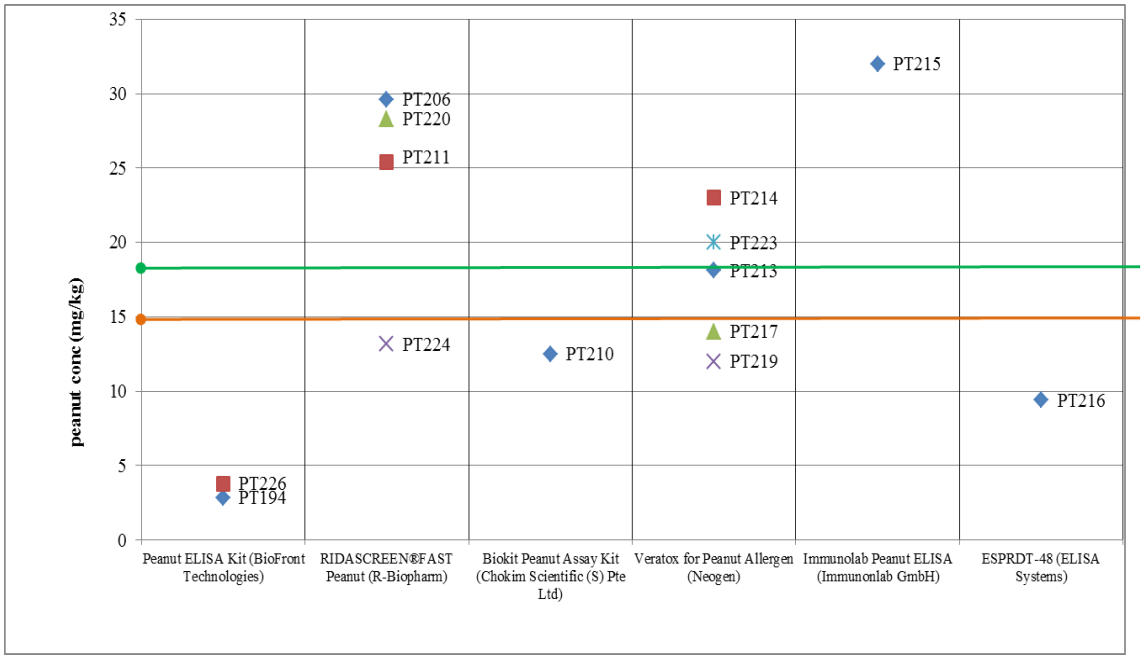

Figure 17: Graphical representation of results reported for peanut in material C, against the different kits used by the participants. The green line represents the assigned value (18.5mg/kg), and the orange line the target level (15mg/kg).

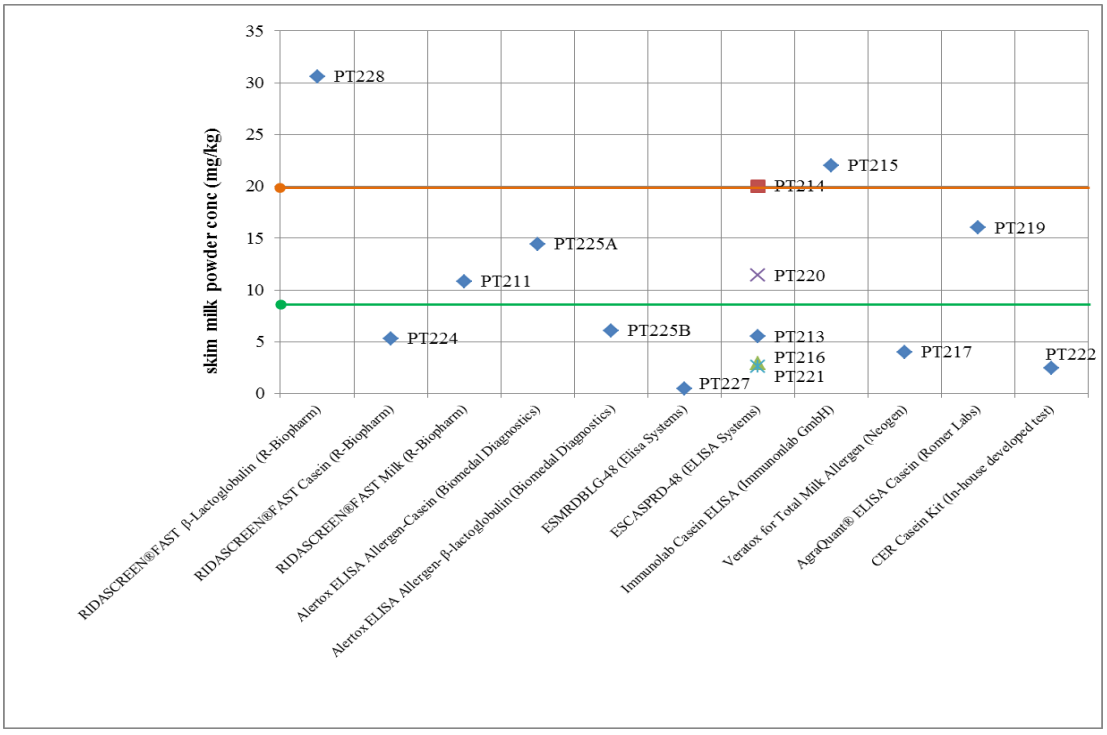

Figure 18: Graphical representation of results reported for skim milk powder in material A, against the different kits used by the participants. The green line represents the assigned value (8.7mg/kg), and the orange line the target level (20mg/kg).

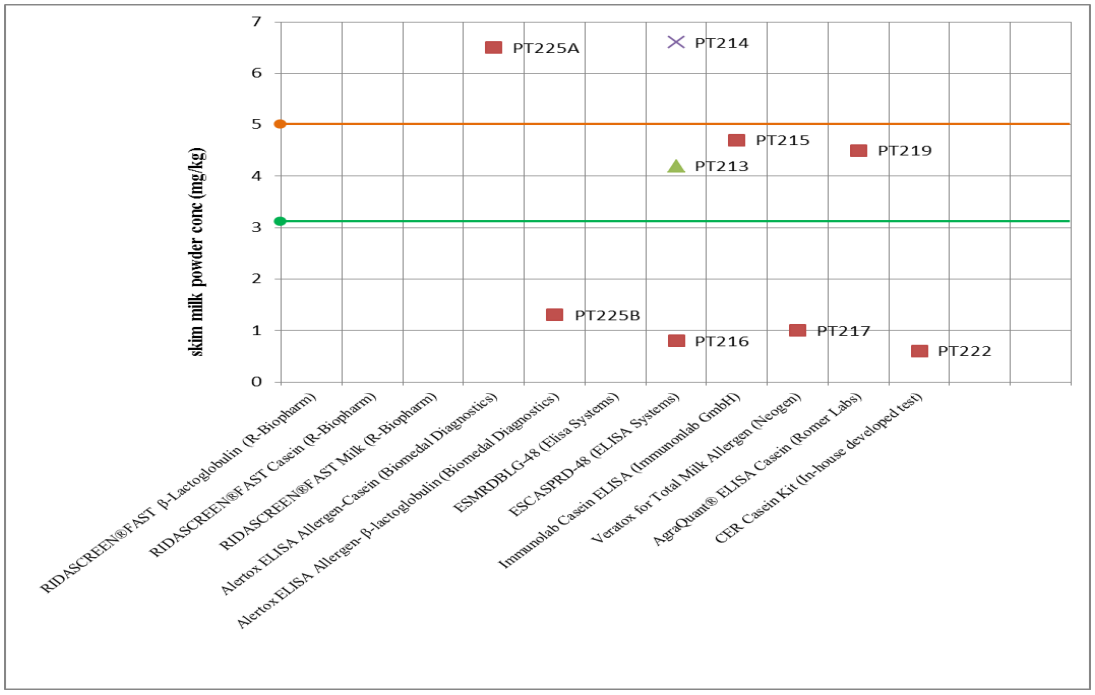

Figure 19: Graphical representation of results reported for skim milk powder in material C, against the different kits used by the participants. The green line represents the assigned value (3.1mg/kg), and the orange line the target level (5mg/kg).

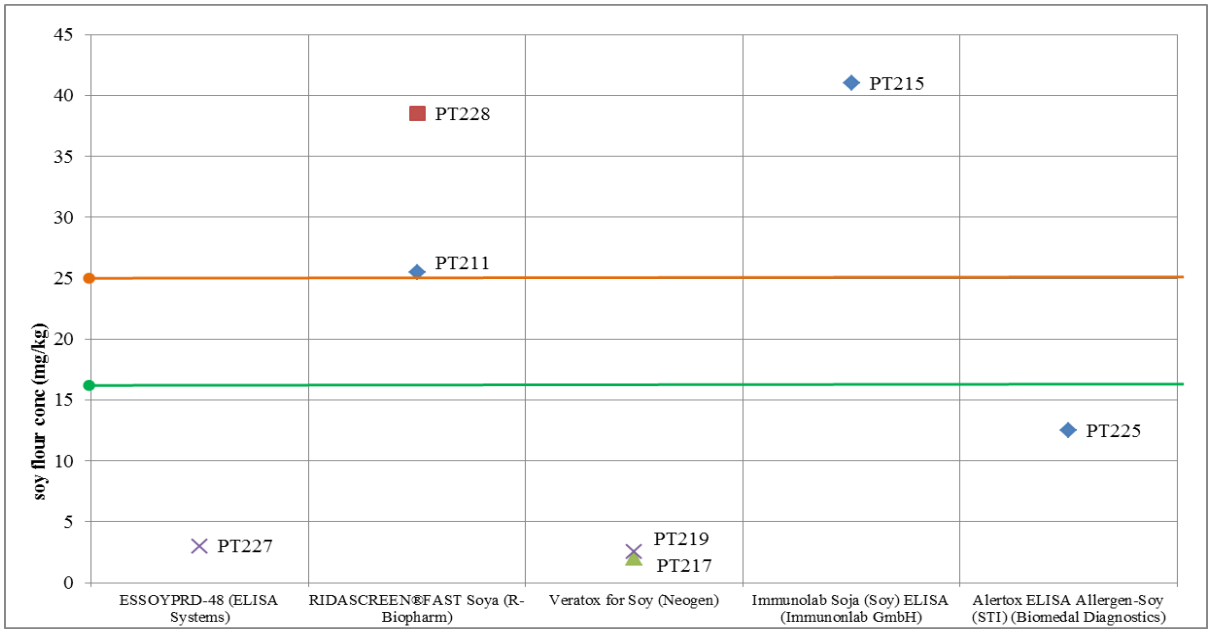

Figure 20: Graphical representation of results reported for soy flour in material B, against the different kits used by the participants. The green line represents the assigned value (16.2mg/kg), and the orange line the target level (25mg/kg).

4. DISCUSSION

4.1 Blank matrix measurements

Based on the results presented in § 3.1, "BIOBIM Banana", that was used as the basis for the preparation of the Proficiency Testing samples, is not free from all the five allergenic compounds used for spiking. The obtained results should be assessed as positive or negative depending on the LOD (level of detection) of every kit. The LOD is determined by testing different, multiple matrices. This means that multiple testing of the blank matrix used with every kit was needed. Due to the fact that this is a difficult and time consuming procedure, in this study the positive and negative results were assessed based on the lowest standard provided in the kit. Whether the blank matrix is positive or negative for the presence of one of the five allergens spiked, was decided based on the comparison of the result acquired for the allergen concentration and the concentration of the lowest standard provided by the respective ELISA kit. This is because the lowest standard in every kit is the starting point for reliable quantification for the kit. When necessary, the concentration of the lowest standards was converted in the unit that the final results for the respective allergens were reported; in gluten from gliadin according to the kit instructions, in soy flour from soy protein according to the kit instructions, and in skim milk powder from milk protein according to VITAL (0.28mg of skim milk powder contain 0.1mg milk protein, therefore the final values were all multiplied by the factor 2.8).

Almond was detected in low concentrations (traces). The results for almond measured by both kits, apart from the one negative result, can be considered as positive, since the respective result acquired for each of them was equal, or higher compared to the lowest standard of every kit (Table 15). The respective absorbance measured for them was also higher compared to that acquired for the lowest standard. Both kits used for the measurements have high sensitivity and can detect traces of almond, as the lowest almond standards they provide are of very low concentration. Especially the Almond ELISA Kit (BioFront Technologies) includes the lowest standard of 0.2mg/kg almond, which is a very low concentration for detection of almond among the commercially available ELISA test kits (Schubert-Ullrich et al., 2009). Regarding gluten, the RIDASCREEN® Gliadin (R-Biopharm) kit used detected a low, still considerable concentration. However, the gluten concentration detected cannot be reliably quantified as the result acquired is lower compared to the lowest standard of the kit (Table 15). Therefore, the result is negative, but as the 3.7mg/kg gluten detected in a considerable concentration, it should be considered later on for

the discussion of the results acquired for gluten in this Proficiency Testing. The soy flour content was found to be in high, considerable level with the specific kit applied, with a definite positive result.

The results for peanut and skim milk powder were negative, and we consider that blank matrix is free from these allergens based on our measurements. For peanut, all the results obtained both with Luminex MAGPIX and the RIDASCREEN®FAST Peanut (R-Biopharm) kit are considered negative. The MFI values for these samples of blank matrix were quite close to the MFI's acquired for the 0mg/kg standard when Luminex MAGPIX method was applied, while the measurement with the ELISA kit gave a clear negative result. For skim milk powder, the result is considered negative, as the result acquired for these samples was quite close to the one found for the 0mg/kg standard and lower compared to the one of the lowest (2.5ppm milk protein) standard.

On the packaging of "BIOBIM Banana" there is a precautionary labelling according to which the product "can contain traces of almond". Therefore, this precautionary labelling is in consensus with our results from two different kits, and almond traces can be detected in the matrix that was not spiked with allergens. What is more, on the label of this material the indication "gluten free" is included. This is acceptable according to the limits set in the European legislation regarding gluten content of foodstuffs and its labelling, as the 3.70 mg/kg of gluten that were detected in our measurement is lower than 20mg/kg set by legislation for "gluten-free" labelling.

The traces of almond and gluten, as well the high level of soy flour content of blank matrix should be taken into account for the discussion of the results reported by the participants and especially for the assessment not only of positive results for materials that were not spiked, but also for the false negative results reported. False positive results, i.e. positive results reported in materials that do not contain the respective allergens, are not a very significant problem when reported. This is because these results can be afterwards confirmed or not by further testing with screening methods. On the other hand, false negative results, i.e. negative results reported in materials that do contain the respective allergens, consist the real problem as they are not indicative for further investigation. In general, it would be easier and more clear-cut to draw conclusions based on the results of the Proficiency Testing, if the infant food based on mixed cereals was free from all the five allergens of interest. However, the use of a material like this is useful since it better resembles real samples and the situations encountered in a routine basis (Owen & Gilbert, 2009, Thompson et al., 2006).

4.2 Validation testing

All the ELISA kits that were chosen and subjected to validation testing were accepted as suitable for use in this PT. These kits could sufficiently differentiate the range of the different dilutions and concentrations prepared and tested for every allergen. Therefore, the kits can be able to detect the possible changes in allergen concentration of the three materials prepared for the PT, important for the homogeneity and stability testing. Moreover, the majority of the results acquired fell within the linear part of the calibration curve and within the limits of quantification of the respective kits. This indicates that the kits can sufficiently quantify the concentration of the allergen of interest in the form and level spiked in the original, undiluted samples, and with the specific background matrix used for the preparation of the samples.

However, some of the findings of the validation testing are interesting to be discussed. To begin with, in some cases the expected concentration of the allergens was not detected. For almond and skim milk powder the detected concentrations with the kits used were lower, approximately 50%, of the targeted (Tables 8, 9, 12), while for peanut and soy flour were higher (almost 2 times the targeted concentration) (Tables 11, 13). This can be explained by the fact that the target allergen concentration was not possibly reached during the spiking procedure. Therefore, lower or higher amounts of the allergens could finally be spiked compared to the targeted ones. This under or over-estimation of allergen concentrations detected can be also attributed to existing differences between the materials used as calibrators in the kit and the material used for spiking the samples of PT. Moreover, differences between the protein sequences that the antibodies of the ELISA kits target and the spiked allergenic material can be implicated (van Eckert et al., 2010). However, in some cases, reference materials are used as calibrators in the kits, when available, which embody the wider possible kind variety of an allergen in order to overcome this problem (Lacorn & Immer, 2010), but this is not the case for all the kits used. The background material used for the preparation of the PT samples could also play a role in under or over-estimation of allergen concentrations by causing matrix effects. Another possible reason to explain this phenomenon is the effects that the extraction procedure and the recovery level can have on the later binding of allergens to the antibodies, which is dependent on every kit (Abbott et al., 2010, Monaci & Visconti, 2010).

Another remark that should be made is related to the appropriateness of the dilution scheme applied for the validation testing of the kits. For peanut, skim milk powder and soy flour the results for some of the dilutions tested fell at the lowest part of the calibration curve, with O.D. values lower compared to the ones acquired for the lowest standard of the kits (Figures 4, 5, 6). In those cases the dilutions prepared and tested (Tables 11, 12, 13) were out of the

quantification range of the kit, and the quantification is not reliable. On the other hand, the undiluted samples fell within the quantification range and could be analysed without significant problems. Therefore, the selection of those dilution schemes indicates that samples that contain the respective allergens in very low concentrations, out of the quantification range of those kits, cannot be reliably analysed. What is more, it is important that for the selection of the kits for analysis of specific samples, the quantification limits of the kits and the expected allergen concentrations are taken into account.

For the purposes of gluten detection, two different kits were initially selected, as already mentioned above. The GlutenTox ELISA Sandwich (Biomedal Diagnostics) kit, which was applied for gluten detection in two different analyses, gave unexpected results for the standards included in the kit. These results were further investigated and the supplier was contacted. Finally, the conclusion reached was that the gliadin standard solutions provided with the kit were not stable under the delivering conditions of the kit. For the above reason this kit was excluded as options for detection and quantification of gluten in this PT as not reliable and the RIDASCREEN® Gliadin (R-Biopharm) kit was finally used. Through this experience, it is obvious that the step of validation testing, before the final use of ELISA kits, is very important and it should be applied in order various failures related to them to be avoided early.

4.3 Homogeneity testing

The homogeneity of every material was tested based on detection and quantification of only one allergen and on the assumption that homogeneity test of peanut and almond was considered adequate to prove the homogeneity of the respective materials. Therefore, the homogeneity of the other allergenic compounds spiked in the materials was not examined. The procedure of analysis for homogeneity is of high cost and the more allergens that are tested, the more expensive it gets. For this reason, the homogeneity testing was limited to one allergen for every material. However, this limitation results in smaller number of samples tested, which can probably be insufficient. What is more, homogeneity (or heterogeneity) of a material is often uneven, and when the number of samples tested is small, there is a risk that these samples may not be representative of the homogeneity level of a material (Fearn & Thompson, 2001, Thompson, 2008).

Homogeneity was ensured for all the three different materials prepared for the Proficiency Testing based on the results presented in § 3.3. The criteria set for comparison of the within-sample standard deviation (s_w) and the between-sample standard deviation (s_s) with the

standard deviation for proficiency assessment (σ_p) (s_w<0.5* σ_p and s_s<0.3* σ_p respectively) were fulfilled for every of the three materials, before their distribution for analysis to the participants. Only material C that was initially prepared was not found to be homogeneous according to the statistical analysis (Table 19), in which the within-sample standard deviation was unacceptably high. For this reason, a second attempt of preparing material C was made, which was successful and the material prepared was finally homogeneous (Table 20). The Luminex MAGPIX method which was used to test the homogeneity of materials A and C was found to be suitable for application, as s_w<0.5* σ_p was met. Only for Almond ELISA Kit (BioFront Technologies), used for homogeneity testing of material B, this criterion was not fulfilled. However, in Table 18 it can be seen that the results acquired for both sets of ten samples tested for homogeneity in material B do not defer in a significant level, i.e. the within-sample standard deviation for every set is not significant. The fact that this criterion is not met can be attributed to either small differences in the results due to the analysis procedure, or to the statistical analysis of this specific application used. Therefore, this was not taken into consideration and the method used was not assessed as inappropriate for use, based also on the validation results for this ELISA kit.

Generally, homogeneity of bulk materials prepared for a Proficiency Test is difficult to be achieved despite the best efforts made and the knowledge available in literature for preparing homogeneous bulk materials is limited (Fearn & Thompson 2001). The inhomogeneous material C that was initially prepared can be due to mistakes made during its preparation, and more specifically to insufficient mixing of the materials with the pestle and mortar at the first steps, or later when every material was further homogenized with liquid nitrogen. In general, information related to failures in the procedure of preparing bulk, homogeneous samples for Proficiency Tests and their causes is difficult to be found in literature.

4.4 Stability testing

Stability was examined for every of the five allergens spiked separately as already described. In order to determine the stability of every allergen, the materials were stored at two different conditions, namely at room temperature and at -80^{o}C, and a comparison was made based on the allergen's concentration in these temperatures. Storage at room temperature was applied in order to simulate realistic conditions often encountered in normal practice during distribution or storage of the materials, also by the participating laboratories. Regarding storage of materials at -80^{o}C, it was chosen based on the assumption that this temperature supports maximum stability conditions for all the compounds used (Thompson et al., 2006, Thompson, 2008).

Consequential instability was detected for almond, gluten and skim milk powder, but not for peanut and soy flour. Therefore, the samples prepared for the PT and contain almond, gluten and skim milk powder are not considered sufficiently stable. On the other hand, samples prepared for the PT and contain peanut and soy flour are considered sufficiently stable. The decrease reported for almond concentration in the respective materials was of approximately 10% calculated with the use of Almond ELISA Kit (BioFront Technologies) for storage at room temperature that lasted 17 days (Table 21). The concentration of gluten decreased by approximately 16%, as measured with RIDASCREEN® Gliadin (R-Biopharm) kit, for storage at room temperature for 25 days (Table 23). The decrease reported for skim milk powder concentration is 8% during storage at room temperature for 28 days, as measured with RIDASCREEN®FAST Milk (R-Biopharm) kit (Table 26). Based on these percentages, it can be seen that the calculated instability for almond, gluten and skim milk powder is only small. It is more likely that this is not a real instability and is probably caused by differences in allergen concentrations due to variation in analyses performed. Such variance in results from allergens analysis is expected since the allergens are present in relatively low concentrations. However, the value of instability calculated for almond, gluten and skim milk powder was incorporated in the statistical analysis of the results acquired by the participants and influenced in this way the calculation of respective z'ai-scores.

4.5 Methods applied in the Proficiency Test by the participants

All the twenty-two participants that took part in this Proficiency Test applied only immunochemical methods. The large majority of them used commercially available ELISA kits for the detection and quantification of the different allergens. For detection and quantification of almond, peanut and soy flour the participating laboratories applied only ELISA methods. Only for gluten and skim milk powder, although the large majority used ELISA kits, one laboratory applied lateral flow devices as already described. Therefore, it is obvious that currently the majority of the analysis of these five allergens is routinely performed with commercially available ELISA kits. This is in consensus with reports in literature according to which the majority of the food allergen analyses is routinely performed with commercially available immunological kits, mainly based on ELISA methodology (Monaci & Visconti, 2010, EFSA NDA Panel, 2014). Other immunochemical methods, and more specifically the lateral flow devices that were applied in this Proficiency Test, are also used in practice for qualitative detection of these allergens but in a much lower extent. The participants that used quantitative ELISA methods reported mainly quantitative results with some few exceptions that used them only for qualitative detection of the allergens. Interesting is also the fact that no one of the participants applied Mass

spectrometry or DNA-based methods, which shows that these methods are not currently applied in routine analyses.

Out of the five allergens that were spiked in the three materials, 82% of the participating laboratories reported results for gluten, 73% for almond, peanut and skim milk powder, and 68% for soy flour. Based on these percentages, we can say that analysis for gluten is one of the most common allergen analysis carried out currently by the laboratories. On the other hand, analysis for soy seems to be less common in practice, although with a small difference in percentage. Almond, peanut and skim milk powder is also among the common allergens analysed by the laboratories worldwide. Regarding the variety of the methods that were reported and used for the detection and quantification of the five allergens, only for gluten and skim milk powder there were one and two laboratories respectively to use more than one method. Generally, it seems that every laboratory that carries out analysis for one of these allergens focuses on the use of one immunochemical method, without having further methods for confirmation, comparison, or other purposes. As for the number of the different brands of ELISA kits applied by the participants, the largest reported for skim milk powder and gluten, while the smallest for soy flour. It seems that gluten is an allergen of great interest, not only for laboratories to perform analyses, but also for producers to develop kits. This does not seem to be the case for soy.

4.6 Results of the participants

4.6.1. Remarks for every allergen

4.6.1.1. Almond

The great majority of the laboratories reported quantitative results for almond and only one reported qualitative result, all using ELISA kits. 80% of the quantitative results reported for almond in material A and 67% in material B were satisfactory. According to these percentages, in material B the satisfactory results are less compared to material A. This can be possibly explained by the fact that almond was spiked in a lower concentration (exactly half) in material B compared to material A. Consequently, we can see that the lower the concentration spiked the less satisfactory results are obtained by the participants. Overall, detection and quantification of almond can be characterised successful, since 73% of the quantitative results reported for almond in spiked materials were satisfactory. The qualitative detection with the ELISA kit was also successful, as 100% of the qualitative results that

laboratory PT229 reported for spiked materials were satisfactory. However, since only one laboratory reported qualitative results, the number of analyses is not statistically relevant.

For materials A and B that were spiked with almond, participants did not report false negative results. Therefore, correct classification of spiked samples, i.e. detection of the presence of allergen, was acquired in 100% of almond analyses. However, in material C which was not spiked with almond, there were three positive results reported. Two of them were quantitative, but not more than 1.3 mg/kg almond, and the third was qualitative. The three participants that reported these results applied the same kit, Almond ELISA Kit (BioFront Technologies) (Table 28). No one of the rest of the participants used this specific commercial kit. As already mentioned, the blank infant food matrix was found to contain almond in low concentrations (0.4mg/kg) by using the same Almond ELISA Kit (BioFront Technologies). Therefore, these positive results cannot be safely assessed as false positive results, as the three participants were able to detect the almond traces present in the blank matrix of the not spiked material, but only by using a specific kit. Based on the above, it seems that only the Almond ELISA Kit (BioFront Technologies) detects traces of almond in the blank matrix and material C that was not spiked with almond. This ELISA kit shows a very low LOD and range of quantification compared to the other kits available, which might account for the results obtained.

In the previous Proficiency Test for allergens in food which was organised last year at RIKILT (Bremer & Elbers, 2014), although the materials prepared were not spiked with almond some of the participants still reported presence of almond up to 2mg/kg by using various ELISA kits. However, the matrix material used was different and it is not known whether the blank matrix contains almond in any level. In a different interlaboratory trial for allergens spiked in meat products, matrices that are more complex, the results acquired for almond were also satisfactory, although the results included ELISA and PCR methods (Köppel et al., 2014). Therefore, comparison of the results acquired for almond from this study with the studies above is quite difficult.

4.6.1.2. Gluten

For gluten, all the results reported by using ELISA kits were quantitative. Only PT225 that used a lateral flow device reported also qualitative results. 83% of the quantitative results reported for gluten in material B and 83% in material C were satisfactory. In the case of gluten, the lower concentration spiked in material B compared to material C, did not lead to smaller percentages of satisfactory results. Overall, detection and quantification of gluten was successful, since 83% of the quantitative results reported for gluten in spiked materials were

satisfactory. The qualitative detection with the lateral flow device used by PT225 was also successful, as 100% of the qualitative results reported for spiked materials were satisfactory. For both materials B and C, which were spiked with gluten, one false negative result was reported. Therefore, correct classification of spiked samples, i.e. detection of the presence of allergen, was acquired in 94% of gluten analyses. However, it would not be safe to compare and draw conclusions on the effectiveness of the two different methods for detection of gluten, as the number of participants used lateral flow devices is not statistically relevant. What is more, lateral flow devices are methods used for qualitative detection of allergens, while ELISA kits are mainly used for quantification.

Regarding material A, this was not spiked with gluten but several positive results were reported. We already know, based on our measurements of the blank matrix, that this contains gluten in low, though considerable concentration (3.7mg/kg). Some of the positive results reported for material A are close to this level, while the rest are more or less higher. All these seven participants that reported the positive results used different ELISA kits, but the positive results that are close to our measurements were acquired by using the RIDASCREEN® Gliadin (R-Biopharm) kit, as the one we applied. The rest kits reported higher values. These results cannot be described as false positive, as it is possible that these laboratories were able to detect gluten already present in the blank matrix.

The assigned value of gluten calculated for materials B and C was found to be higher compared to the target concentration of spiking by approximately 3.5mg/kg. This difference could be possibly attributed to the fact that the participants detected the gluten present at blank matrix at this concentration.

In literature a number of Proficiency Tests organised for gluten is reported, presenting results that are in consensus with the results found in this PT. The data obtained so far from other PTs show differences in the reported levels of gluten for the different commercial ELISA kits applied (Australian Government, National Measurement Institute, 2012, Scharf et al., 2013). However, detection and quantification of gluten could be reliably performed, by most of the ELISA kits. What is more, in these PTs 98% of the participants detected the gluten positive samples, while false negative results were reported only when complex matrices, like pastry, were implicated (Scharf et al., 2013).

4.6.1.3. Peanut

All the participants used commercial ELISA kits for detection and quantification of peanut. However, only one of them (PT229) reported finally qualitative results. Due to very high

uncertainty of the assigned value calculated for peanut, z'a-scores were calculated only for information. Therefore, it was not possible to classify the results as satisfactory, questionable, or unsatisfactory. Correct classification of spiked with peanut samples was obtained in 94% of peanut analyses. There were no false positive results for material B which was not spiked with peanut. Moreover, false negative results reported only in material A by two participants, but not in material C which was also spiked with peanut. The spiked peanut concentration in material A was 5mg/kg, three times lower compared to material C, which is a low level close to the lower limit of detection of the ELISA kits. For this reason, it is expected to have false negative results when such low concentrations are tested. The two participants that reported the false negative results they were the only ones to use the ESPRDT-48 (ELISA Systems) kit. Although the two results acquired are not sufficient to draw conclusions, there is an indication that this specific kit is not able to detect peanut in such low concentration.

In both material A and C the assigned value calculated for peanut was found to be higher compared to the target spiking level, which may indicate that some ELISA kits used tend to overestimate the concentration of peanut present. More specifically, mainly two ELISA kits, RIDASCREEN®FAST Peanut (R-Biopharm) and Veratox for Peanut Allergen (Neogen) seem to report results that overestimate the peanut present. It seems that when raw peanuts are used for spiking, extraction of allergenic proteins is easier and the recovery rate is better compared to heat-treated peanuts (Jayasena et al., 2015). On the other hand, there are also two other ELISA kits applied by the participants, namely the Peanut ELISA Kit (BioFront Technologies) and the ESPRDT-48 (ELISA Systems), the use of which gave underestimated peanut results.

Very high uncertainty for peanut results was also reported in the previous Proficiency Test organised at RIKILT last year, but only in one of the materials prepared. In that study, large variation was also observed in the quantitative peanut results (Bremer & Elbers, 2014). Therefore, no significant changes are reported from the previous study for peanut. In a different interlaboratory trial for allergens spiked in the complex matrix of meat products conducted, the results acquired for peanut were satisfactory, but the results include ELISA and PCR methods (Köppel et al., 2014).

4.6.1.4. Skim milk powder

The majority of the participants analysed for skim milk powder reported quantitative results by using ELISA kits. Only participant PT225 used also two different lateral flow devices and reported qualitative results. The uncertainty of the assigned value calculated for skim milk powder was found to be very high and the calculated z'a-scores are used only for

information. Therefore, it was not possible to classify the results as satisfactory, questionable, or unsatisfactory. Correct classification of samples spiked with skim milk powder was obtained in 78% of skim milk powder analyses. There was one false positive result in material B which was not spiked with skim milk powder. Moreover, in material A there was only one participant to report false negative result, while in material C which was also spiked with skim milk powder six participants gave false negative results. The spiked skim milk powder concentration in material A was 20mg/kg, four times higher compared to material C, with 5mg/kg. The latter is a low level of spiking, close to the lower limit of detection of the ELISA kits. False negative results were expected for skim milk powder in this material since such low concentration was tested. The given false negative results were acquired by various kits and not by specific kits.

Participant PT225 apart from the ELISA kits used also two different lateral flow devices. The qualitative results reported by both of them correctly classified the presence, as well as the absence of skim milk powder in the respective materials. However, since both of these devices were used only once by one participating laboratory, their effectiveness for detection of skim milk powder can be neither assessed nor compared to the ELISA kits.

A large variation is reported for the quantitative results especially in material A. The values reported range from 0.5 to 30.6mg/kg skim milk powder, while the assigned value is 8.7mg/kg. It is interesting to mention that a considerable number of different ELISA kits used for detection and quantification of skim milk powder gave results that are much lower compared to the target spiking level in materials A and C (Tables 29, 31, Supplements). This is an indication that many of the commercially available kits probably underestimate the concentration of skim milk powder truly present in a sample, even when the spiking level is sufficiently high, like in material A.

In the previous Proficiency Test for allergens in food which was organised last year at RIKILT (Bremer & Elbers, 2014) milk was not examined and the materials prepared were not spiked with milk. However, the participating laboratories reported results for presence of milk that were ranging from 0.45 to 179.69mg/kg milk. In this case the above results indicate overestimation of presence of milk, although the matrix used for the preparation of samples is different from the present Proficiency test and it is unknown whether the blank matrix contains milk at any level. Further information on results acquired from other Proficiency Test was not easily found in literature.

In general, the different commercial ELISA kits available for detection of milk target different proteins, caseins, β-lactoglobulin, or total milk proteins. Although every kit provides a conversion factor for converting the respective targeted proteins to milk

concentration, differences can be observed with the use of the various kits based also on the proteins present in the sample tested and the specificity of the kit's antibodies.

4.6.1.5. Soy flour

All the fifteen laboratories that analysed for soy used only ELISA kits and reported only quantitative results. The uncertainty of the assigned value calculated for soy flour was found to be very high and the calculated z'a-scores are used only for information. Therefore, it was not possible to safely classify the results as satisfactory, questionable, or unsatisfactory, as mentioned for peanut and skim milk powder. Correct classification of spiked samples was obtained in 53% of soy analyses, which is a very low percentage compared to the ones calculated above for the other allergens. In materials A and C, which were not spiked with soy flour, two positive results were found respectively. These positive results reported the presence of soy flour at an average concentration of 12mg/kg. According to our measurements performed for the blank matrix of the samples, the latter contains soy flour at an average level of 13.5mg/kg. However, these results of the two participants and our results were acquired by applying the same ELISA kit, namely the RIDASCREEN®FAST Soya (R-Biopharm) kit. No other participant used this specific kit in this PT. Therefore, those results can be attributed either to overestimation of soy presence from this specific kit or to inability of other kits used to detect soy flour present in our matrix. Moreover, overestimation can be caused from matrix effects or cross-reactivity. Another interesting finding is that in material B, only two of the participants (PT215, PT228) reported higher concentrations of soy flour, close to the ones expected based on the target spiking level and the soy content of the blank matrix.

In material B, spiked with soy flour, there were seven participants to report false negative results, which is a considerable number. The participants that reported the false negative results used mainly two specific kits, namely the Veratox for Soy (Neogen) and the ESSOYPRD-48 (ELISA Systems). The rest of the participants that used one of those two kits reported positive result, but still very low (<5.8mg/kg). There is an indication that these two kits are not able to sufficiently detect soy, spiked as soy flour in the infant food matrix, even at the considerable level of 25mg/kg.

In general, there is again a large variation among the quantitative results reported for soy, which are ranging in material B from 2 to 38.5mg/kg soy flour, with an assigned value of 16.2mg/kg which is quite low. Moreover, in case soy content of blank matrix was correctly detected and quantified, it seems that there is an underestimation of soy concentration present by many of the kits used in this PT, as described above.

In Proficiency Test organised for allergens in food in 2014 by RIKILT, the uncertainty of the assigned value calculated for soy was also very high, not allowing classification of results and use of z-scores for evaluation of the participants. Moreover, for soy large variation in the quantitative results was reported (Bremer & Elbers, 2014). In an international interlaboratory trial performed recently, the results for soy were also unsatisfactory due to very high measurement uncertainty. Based on these results, the study concludes that methods available for soy must be improved (Köppel et al., 2014). In a review of the PTs performed from 2006 to 2011, for soy multimodal distribution of the results was reported. What is more, the ELISA methods evaluated were not found to be suitable for detection of soy protein, and all the ELISA test kits applied reported a considerable high number of false negative results, findings similar with ours. The underestimation of the soy content from the applied ELISA kits was also discussed. However, all these results were acquired for matrices that are more complex compared to the matrix we used (Scharf et al., 2013).

4.6.2. General remarks

The uncertainty of the assigned value calculated for analysis of peanut, skim milk powder and soy flour was found to be very high. More specifically, the criterion $u \leq 0.3\sigma_P$ is not met and the uncertainty is even $>0.7\sigma_P$. For this reason, the calculated z-scores cannot be used for evaluation of the performance of the participating laboratories, but they are used for information only. In such a case there is a risk that participating laboratories is possible to receive questionable or unsatisfactory z-scores due to inaccuracy in the determination of the assigned value and not as a result of unsatisfactory performance within the laboratory analysis (ISO 13528:2005(E), 2005). According to ISO 13528:2005, the uncertainty of the assigned value is affected by the method used for its determination and also, in case that it is derived based on measurements from several laboratories, by the number of laboratories and, perhaps, by other factors.

In some cases, in materials that were spiked with the different allergens, the calculated assigned value was found to be either lower or higher compared to the target concentration of spiking. This shows that either the target concentration was not achieved, or that there is a general trend of underestimation or overestimation, respectively, of the concentration of the allergen present in a sample, by the ELISA kits used.

Another interesting point, observed for all the allergens included in this PT, is the large variation in the quantitative results. Even in the cases that detection and quantification of allergens was performed at a sufficient level (almond, gluten), a large variation is observed in the reported quantitative results. This can be easily seen in Tables 29, 30 and 31

(Supplements). In the respective materials spiked with allergens, the range among the lowest and the highest values reported is often wide, and there are values that differ significantly from the calculated assigned value. This variation made the statistical evaluation of the results impossible in some cases.

All the above results acquired could be attributed to the variety of the commercial ELISA kits used for detection and/or quantification of an allergen. All these kits show differences in the effectiveness of the extraction protocols, the calibration materials and the specificity of the antibodies employed by every commercial kit (van Eckert et al., 2010, Scharf et al., 2013).

For the evaluation of the results reported from participants for peanut, skim milk powder and soy flour, for which z-scoring is not reliable, an alternative approach was applied. The results were plotted against the different ELISA kits used (Figures 16, 17, 18, 19, 20). In this way an effort was made to facilitate the comparison of the performance of the laboratories that used the same ELISA kit, and consequently the performance of these kits. Every participant can assess the result produced by comparing it with the results produced by other laboratories, as well as with the assigned value and the target level of the spiking. The most important point that all these figures support is that even when the same ELISA kit is used for detection and/or quantification of a specific allergen by different participants, there is large variation in the results. What is more, for many of the kits, it is difficult to assess their effectiveness since only one or two participants applied them, number that is not sufficient and representative (Sykes et al., 2012). For the same reason, even a separate evaluation for every kit was not possible (Scharf et al., 2013). Another interesting point is that, some of the participants although they closely detected the target spiking level, they finally obtained an unsatisfactory or questionable z-score (always for information only) (e.g. PT214 for skim milk powder in material A, Figure 18). This is because the assigned value is calculated based on the results that participants report.

However, the results are not only dependent on the ELISA kits used, but also on the performance of the laboratories applying them. Application of the same kit by different participants for detection and/or quantification of a specific allergen in the same material resulted in very different results. There were also other problems observed during results handling by the participants, such as no conversion to the form of allergen asked, use of kits for allergens outside the scope of this PT (e.g. hazelnut or egg), and reporting the results for the three materials in the wrong order. All these failures and mistakes are also important aspects of quality and fall within the general quality assessment of the participating laboratories (Owen & Gilbert, 2009).

4.6.3. Compliance of allergen detection levels with limits set by legislation

As it was already discussed above, gluten could be successfully detected and quantified, since 83% of the quantitative results reported for gluten in spiked materials were satisfactory and 100% of the qualitative results reported for spiked materials were satisfactory. According to REGULATION (EC) No 41/2009, the gluten content of foodstuffs that bear the labelling "very low gluten" shall not exceed 100 mg/kg in the final food for consumption, while for "gluten-free" foods the limit is lowered to 20 mg/kg. In material B, which was spiked with approximately 10mg/kg gluten (in order to resemble "gluten-free" labelled foodstuffs based on the legislation limit), 83% of the quantitative results and 100% of the qualitative results were satisfactory. Therefore, it seems that labelling of "gluten-free" and "very low gluten" foodstuffs can be successfully determined by the methods currently available. According to data found in literature, the tolerable gluten concentration varies among patients with coeliac disease. However, a daily gluten intake of <10mg is assessed as unlikely to cause significant histological abnormalities based on the findings acquired so far (Ludvigsson et al., 2014). This means that gluten concentration of 10mg/kg, that is a critical threshold for patients, can be detected at a satisfactory level by the current commercial methods. However, the limit set by legislation is two times higher (20mg/kg gluten) than the critical threshold. Since this study shows that even 10mg/kg gluten can be satisfactorily detected by the current methods used, this might be an indication towards examining the possibility of lowering the legislation limit that defines the gluten-free foodstuffs.

For the rest of the allergens, there are not similar limits defined for their labelling in legislation (Lacorn & Immer, 2010). This is probably because the amount of the various allergens that can result in allergic reaction is difficult to be determined for every sensitive individual as already mentioned. The doses of the allergenic food or ingredient reported so far, that are able to trigger adverse reactions, were found to range from micrograms to milligrams, or even grams (EFSA NDA Panel, 2014). Consequently, the current labelling of the rest of the allergens cannot be really reliable and informative for the allergenic consumers.

5. CONCLUSIONS

In conclusion, the presence of almond, gluten and peanut can successfully be detected, even when low concentrations are targeted. On the contrary, skim milk powder and soy flour cannot be successfully detected. Quantification of the allergens proved to be more difficult. In general, for all the five allergens of interest a large variation regarding the quantitative results was noted, mainly depended on the commercial brand of the ELISA kits used. Moreover, even when the same ELISA kit was used for detection and/or quantification of a specific allergen by different participants, there was large variation in the reported results. These are important indications that the current allergen detection methods need to be further developed towards their harmonisation and production of accurate, reliable results. What is more, there is a dire need for improvements related to the commercial ELISA kits available, as well as to the way they are applied by the laboratories in practice.

In this Proficiency Test gluten was analysed by the highest number of laboratories. Therefore, analysis for gluten is one of the most common allergen analysis carried out currently by the laboratories. Gluten analysis can be performed successfully, even quantitatively. Labelling of "gluten-free" and "very low gluten" foodstuffs seems that can be successfully determined by the methods currently available. It seems also that gluten is an allergen of great interest, not only for laboratories to perform analyses, but also for producers to develop kits. On the other hand, analysis for soy seems to be less common in practice, possibly due to unsatisfactory results acquired with the available kits as seen in this PT. Almond, peanut and skim milk powder are also among the common allergens analysed by the laboratories worldwide, but still with problems in their analyses.

In daily practice, the majority of the analysis of these five allergens is routinely performed with commercially available ELISA kits, as reflected by the application of only immunochemical methods by the participants. Other immunochemical methods, and more specifically the lateral flow devices, are also applied in practice for qualitative detection of these allergens but in a much lower extent. Mass spectrometry or DNA-based methods were not applied by the participants in this Proficiency Test. Hence, it can be concluded that these methods did not find their way in to application in routine of the laboratories yet.

In addition to the above, validation testing for every of the ELISA kits that are applied for allergen analysis is of significance and should be preceded of every testing. Lastly, difficulties in achieving the sufficient homogeneity of bulk materials for Proficiency Test needs, as well as problems with consequential instability for some of the allergens remain as significant issues.

6. RECOMMENDATIONS FOR FURTHER RESEARCH

In the field of detection and quantification of food allergens there are still a lot of aspects for further research. What is more, the number of Proficiency Tests for food allergens organised so far is limited, as already mentioned. For this reason, new Proficiency Tests need to be organised that focus on a greater variety of matrices and food allergens that were not investigated so far. The allergens, for which unsatisfactory results were acquired so far, like soy and milk, should also continue to be included in the organised Proficiency Tests.

Regarding the target levels that should be examined in new Proficiency Tests, they should cover a wide range of concentrations, as the doses of the allergenic foods or ingredients reported so far, that are able to trigger adverse reactions, were found to range from micrograms to even grams. In this way, reliable quantification of a range of different concentrations can be tested by the methods currently applied. In low target levels many problems are reported so far with detection of allergens. Hence, it would be of interest if the target levels of the allergens are as low as possible, in order to determine whether presence of allergens in traces can be detected and/or quantified by the available methods. Target levels should also continue to be based on scientific thresholds reported in literature, in order to more closely resemble the doses of allergens needed to trigger allergic reactions. This will also help towards setting of limits in legislation for labelling foodstuffs as “free” or not, as the ones already set for gluten, in order to make labelling of more allergens more reliable and helpful for allergic consumers.

There are also more aspects of Proficiency Tests that need further investigation and improvement. A larger number of laboratories should be encouraged to participate in organised Proficiency Tests for food allergens, in order to acquire larger and statistically relevant number of results from different methods. Moreover, the different types of detection methods should also be included in such studies for comparison reasons. The causes that lead to inhomogeneity and instability of the samples that are prepared for Proficiency Tests should also be further investigated, and new methodologies that will help to overcome them should be developed. Regarding the type of samples that are prepared for a PT, incurred samples are preferable for investigation against spiked samples, as the former are more realistic. Incurred samples can help to better investigate the difficulties in allergen recovery and detection caused by processing of foodstuffs.

Another significant problem is the large variation of the reported quantitative results, which is mainly caused by the application of ELISA kits of different commercial brands. To overcome this problem, harmonisation of the different commercial kits needs to be achieved.

Development of new reference materials, specially developed for allergen analysis purposes, as well as of reference methods would be the most important and effective approaches towards this direction.

Finally, it seems that gluten is an allergen of great interest, not only for laboratories to perform analyses, but also for producers to develop kits. Moreover, the current methods available for gluten analysis can meet the target concentrations set in EU legislation. However, the limit set by legislation for "gluten-free" foodstuffs is two times higher compared to the critical threshold for patients currently reported in literature. Therefore, there might be an indication towards examining the possibility of lowering the legislation limit that defines the gluten-free foodstuffs.

REFERENCES

Abbott, M., Hayward, S., Ross, W., Godefroy, S.B., Ulberth, F., Van Hengel, A.J., Roberts, J., Akiyama, H., Popping, B., Yeung, J.M., Wehling, P., Taylor, S., Poms, R.E., Delahaut, P. (2010). Validation Procedures for Quantitative Food Allergen ELISA Methods: Community Guidance and Best Practices. *Journal of AOAC International, 93 (2)*, 442-450

Alexopoulos, C., Kakoulides, E., Lampi, E. (2014). Proficiency schemes for food allergen testing. In G. Siragakis, & D. Kizis (Eds), *Food Allergen Testing, Molecular, Immunochemical and Chromatographic Techniques* (pp. 237-265). UK: JohnWiley&Sons,Ltd

Analytical Methods Committee (1989 (1)). Robust statistics - How not to reject outliers, Part 1. Basic concepts. *Analyst, 114*, 1693-1697

Analytical Methods Committee (1989 (2)). Robust statistics - How not to reject outliers, Part 2. Inter-laboratory trials. *Analyst, 114*, 1699-1702

Australian Government, National Measurement Institute (2012). PROFICIENCY STUDY AQA 12-10 ALLERGENS IN FOOD. (http://www.measurement.gov.au/publications/proficiencystudyreports/Pages/default.aspx)

Bremer, M.G.E.G., & Elbers, I.J.W. (2014). Proficiency test for allergens in food. Wageningen, RIKILT Wageningen UR, Confidential RIKILT report 2014.509.

Costa, J., Ansari, P., Mafra, I., Oliveira, M.B.P. P., Baumgartner, S. (2014). Assessing hazelnut allergens by protein- and DNA-based approaches: LC-MS/MS, ELISA and real-time PCR. *Journal of Analytical and Bioanalytical Chemistry*, *406*, 2581–2590, doi:10.1007/s00216-014-7679-x

Culliford, A.N., & Green, P.H. (2003). Refractory sprue. *Current gastroenterology reports, 5(5)*, 373-378

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2014. Scientific Opinion on the evaluation of allergenic foods and food ingredients for labelling purposes. *EFSA Journal* 2014;12(11):3894, 286 pp. doi:10.2903/j.efsa.2014.3894

Fearn, T., & Thompson, M. (2001). A new test for 'sufficient homogeneity'. *Analyst*, *126*, 1414–1417, doi:10.1039/b103812p

Green, P.H., & Cellier, C. (2007). Celiac disease. *New England Journal of Medicine. 357(17)*, 1731-1743, doi:10.1056/NEJMra071600

Hall, N.J., Rubin, G., & Charnock, A. (2009). Systematic review: adherence to a gluten-free diet in adult patients with coeliac disease. *Journal of Alimentary pharmacology & therapeutics, 30(4),* 315-330, doi:10.1111/j.1365-2036.2009.04053.x

Hamer R. (2013). Biochemical aspects of allergies and intolerances. In Wierenga, P.A., Savelkoul, H., Hamer, R., Wichers, H., Gilisen, L., & van der Goot, A.J. (Eds.), *Reader "Food-related allergies and intolerances" FCH-21806* (pp. 4-92). Wageningen University.

Hollon, J.R., Cureton, P. A., Martin, M. L., Puppa, E. L. L., & Fasano, A. (2013). Trace gluten contamination may play a role in mucosal and clinical recovery in a subgroup of diet-adherent non-responsive celiac disease patients. *BMC gastroenterology, 13(1),* 40, doi:10.1186/1471-230X-13-40

ISO 13528:2005(E) (2005). Statistical methods for use in proficiency testing by interlaboratory comparisons, First edition 2005-09-01

Jayasena, S., Smits, M., Fiechter, D., de Jong, A., Nordlee, J., Baumert, J., Taylor, S.L., Pieters, R.H., Koppelman, S.J. (2015). Comparison of six commercial ELISA kits for their specificity and sensitivity in detecting different major peanut allergens. *Journal of Agricultural and Food Chemistry, 63(6),* 1849-1855, doi:10.1021/jf504741t

Kizis, D., & Siragakis, G. (2014). Introduction. In G. Siragakis, & D. Kizis (Eds), *Food Allergen Testing, Molecular, Immunochemical and Chromatographic Techniques* (pp. 1-9). UK: JohnWiley&Sons,Ltd

Köppel, R., Rentsch, J., Ruf, J., Eugster, A., Graf, C., Felderer, N., Pietsch, K., Ilg, E. (2014). Results of an International Interlaboratory Trial to Determine Twelve Allergens Using Real-time PCR- and ELISA-based Assays. *Chimia, 68 (10),* 721–725, doi:10.2533/chimia.2014.721

Lacorn, M., & Immer, U. (2010). Standardization in allergen determination, *Journal of Accreditation and Quality Assurance., 15,* 207-216, doi:10.1007/s00769-009-0592-x

Lampidonis, A., & Siragakis, G. (2014). Allergens in the food industry: customer and legislation demands. In G. Siragakis, & D. Kizis (Eds), *Food Allergen Testing, Molecular, Immunochemical and Chromatographic Techniques* (pp. 197-212). UK: JohnWiley&Sons,Ltd

Ludvigsson, J.F., Bai, J.C., Biagi, F., et al. (2014). Diagnosis and management of adult coeliac disease: guidelines from the British Society of Gastroenterology. *Gut, 63*, 1210–1228, doi:10.1136/gutjnl-2013-306578

Madsen, C.B., Hattersley, S., Buck, J., Gendel, S.M., Houben, G.F., Hourihane, J.O'B., Mackie, A., Mills, E.N.C., Nørhede, P., Taylor, S.L., Crevel, R.W.R. (2009). Approaches to risk assessment in food allergy: Report from a workshop ''developing a framework for assessing the risk from allergenic foods". *Journal of Food and Chemical Toxicology, 47 (2)*, 480–489, doi:10.1016/j.fct.2008.12.001

Mena, M.C., Lombardía, M., Hernando, A., Méndez, E., Albar, J.P. (2012). Comprehensive analysis of gluten in processed foods using a new extraction method and a competitive ELISA based on the R5 antibody. *Talanta, 91*, 33–40, doi:10.1016/j.talanta.2011.12.073

Monaci, L., & Visconti, A. (2009). Mass spectrometry-based proteomics methods for analysis of food allergens. *Journal of TrAC Trends in Analytical Chemistry, 28(5)*, 581–591, doi:10.1016/j.trac.2009.02.013

Monaci, L., & Visconti, A., (2010). Immunochemical and DNA-based methods in food allergen analysis and quality assurance perspectives. *Journal of Trends in Food Science & Technology, 21*, 272-283

Murray, J.A., Watson, T., Clearman, B., & Mitros, F. (2004). Effect of a gluten-free diet on gastrointestinal symptoms in celiac disease. *The American journal of clinical nutrition, 79(4)*, 669-673.

Owen, L., & Gilbert, J. (2009). Proficiency testing for quality assurance of allergens methods. *Journal of Analytical and Bioanalytical Chemistry, 395*, 147–153, doi:10.1007/s00216-009-2739-3

Paarlahti, P., Kurppa, K., Ukkola, A., Collin, P., Huhtala, H., Mäki, M., Kaukinen, K. (2013). Predictors of persistent symptoms and reduced quality of life in treated coeliac disease patients: a large cross-sectional study. *BMC gastroenterology, 13(1)*, 75, doi: 10.1186/1471-230X-13-75

Pantaleoni, S., Luchino, M., Adriani, A., Pellicano, R., Stradella, D., Giuseppe Ribaldone, D., Sapone, N., Isaia, G. C., Di Stefano, M., & Astegiano, M. (2014). Bone mineral density at diagnosis of celiac disease and after 1 year of gluten-free diet. *The Scientific World Journal*, Volume 2014, Article ID 173082, http://dx.doi.org/10.1155/2014/173082

Poms, R. E., Klein, C. L., & Anklam, E. (2007). Methods for allergen analysis in food: a review. *Food Additives & Contaminants, 21(1)*, 1-31, doi:10.1080/02652030310001620423

Rona, R. J., Keil, T., Summers, C., Gislason, D., Zuidmeer, L., Sodergren, E., Sigurdardottir, S. T., Lindner, T., Goldhahnb, K., Dahlstrom, J., McBride, D., Madsen, C. (2007). The prevalence of food allergy: A meta-analysis. *Journal of Allergy and Clinical Immunology, 120*, 638-646, doi:10.1016/j.jaci.2007.05.026

Scharf, A., Kasel, U., Wichmann, G., Besler, M., (2013). Performance of ELISA and PCR Methods for the Determination of Allergens in Food: An Evaluation of Six Years of Proficiency Testing for Soy (*Glycine max* L.) and Wheat Gluten (*Triticum aestivum* L.), *Journal of Agricultural and food chemistry 61*, 10261-10272, doi:10.1021/jf402619d

Schubert-Ullrich, P., Rudolf, J., Ansari, P., Galler, B., Führer, M., Molinelli, A., Baumgartner, S. (2009). Commercialized rapid immunoanalytical tests for determination of allergenic food proteins: an overview. *Journal of Analytical and Bioanalytical Chemistry, 395(1)*, 69-81, doi:10.1007/s00216-009-2715-y

Sicherer, H. Scott, & Sampson, A. Hugh (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *Journal of Allergy and Clinical Immunology, 133*, 291–307.e5, doi:10.1016/j.jaci.2013.11.020

Spanjersberg, M.Q.I., Kruizinga, A.G., Rennen, M.A.J., Houben, G.F. (2007). Risk assessment and food allergy: the probabilistic model applied to allergens. *Journal of Food and Chemical Toxicology, 45 (1)*, 49–54, doi:10.1016/j.fct.2006.07.018

Sykes, M., Anderson, D., Parmar, B. (2012). Normalisation of data from allergen proficiency tests. *Journal of Analytical and Bioanalytical Chemistry, 403*, 3069-3076, doi:10.1007/s00216-012-5780-6

Thompson, M. (2008). Instability and heterogeneity: a new approach needed! *Journal of Accreditation and Quality Assurance, 13*, 581–584, doi:10.1007/s00769-008-0432-4

Thompson, M., (2000). Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing, *Analyst, 125*, 385-386, doi:10.1039/b000282h

Thompson, M., Ellison, S.L., & Wood, R. (2006). The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories, (IUPAC Technical Report). *Journal of Pure Appl. Chem., 78 (1)*, 145–196, doi:10.1351/pac200678010145

Tomoskozi, S., Torok, K., Bugyi, Z., Hajas, L. (2014). Reference materials for food allergen testing. In G. Siragakis, & D. Kizis (Eds), *Food Allergen Testing, Molecular, Immunochemical and Chromatographic Techniques* (pp.215-233). UK: JohnWiley&Sons,Ltd

Tursi, A., Brandimarte, G., & Giorgetti,G. (2003). High prevalence of small intestinal bacterial overgrowth in celiac patients with persistence of gastrointestinal symptoms after gluten withdrawal. *The American journal of gastroenterology, 98(4)*, 839-843, doi:10.1111/j.1572-0241.2003.07379.x

van Eckert, R., Bond, J., Rawson, P., Klein, Ch.L., Stern, M., Jordan, T.W. (2010). Reactivity of gluten detecting monoclonal antibodies to a gliadin reference material. *Journal of Cereal Science, 51 (2)*, 198–204, doi:10.1016/j.jcs.2009.11.012

WEB REFERENCES

http://allergenbureau.net/, last accessed: 20/01/2015

http://allergenbureau.net/wp-content/uploads/2013/11/VITAL_Allergenic_Protein_Levels_10_11_08.pdf, last accessed: 15/03/2015

https://www.moniqa.eu/node/1193, last accessed: 10/02/2015

SUPPLEMENTS

Table 6: Codification of the samples shipped to every participant.

Laboratory code	Material A	Material B	Material C
PT194	735	184	623
PT206	325	718	879
PT210	371	918	912
PT211	408	114	446
PT212	613	818	633
PT213	673	825	199
PT214	420	938	696
PT215	742	487	874
PT216	367	100	527
PT217	794	392	413
PT218	559	300	585
PT219	900	550	506
PT220	296	161	627
PT221	270	791	442
PT222	779	749	766
PT223	215	829	988
PT224	676	522	883
PT225	747	375	619
PT226	262	122	572
PT227	401	360	974
PT228	976	639	508
PT229	953	298	402

Table 7: Coding of the samples used in homogeneity testing of materials A, B and C.

Material A		Material B		Material C	
Code for analysis	Container code	Code for analysis	Container code	Code for analysis	Container code
HOM A1	PT RIKILT Food Allergens 2014 A homogeneity 01	HOM B1	PT RIKILT Food Allergens 2014 B homogeneity 20	HOM C1	PT RIKILT Food Allergens 2014 C homogeneity 14
HOM A2	PT RIKILT Food Allergens 2014 A homogeneity 02	HOM B2	PT RIKILT Food Allergens 2014 B homogeneity 17	HOM C2	PT RIKILT Food Allergens 2014 C homogeneity 16
HOM A3	PT RIKILT Food Allergens 2014 A homogeneity 03	HOM B3	PT RIKILT Food Allergens 2014 B homogeneity 13	HOM C3	PT RIKILT Food Allergens 2014 C homogeneity 09
HOM A4	PT RIKILT Food Allergens 2014 A homogeneity 04	HOM B4	PT RIKILT Food Allergens 2014 B homogeneity 05	HOM C4	PT RIKILT Food Allergens 2014 C homogeneity 02
HOM A5	PT RIKILT Food Allergens 2014 A homogeneity 05	HOM B5	PT RIKILT Food Allergens 2014 B homogeneity 06	HOM C5	PT RIKILT Food Allergens 2014 C homogeneity 10
HOM A6	PT RIKILT Food Allergens 2014 A homogeneity 06	HOM B6	PT RIKILT Food Allergens 2014 B homogeneity 16	HOM C6	PT RIKILT Food Allergens 2014 C homogeneity 17
HOM A7	PT RIKILT Food Allergens 2014 A homogeneity 07	HOM B7	PT RIKILT Food Allergens 2014 B homogeneity 15	HOM C7	PT RIKILT Food Allergens 2014 C homogeneity 01
HOM A8	PT RIKILT Food Allergens 2014 A homogeneity 08	HOM B8	PT RIKILT Food Allergens 2014 B homogeneity 08	HOM C8	PT RIKILT Food Allergens 2014 C homogeneity 12
HOM A9	PT RIKILT Food Allergens 2014 A homogeneity 09	HOM B9	PT RIKILT Food Allergens 2014 B homogeneity 02	HOM C9	PT RIKILT Food Allergens 2014 C homogeneity 19
HOM A10	PT RIKILT Food Allergens 2014 A homogeneity 10	HOM B10	PT RIKILT Food Allergens 2014 B homogeneity 14	HOM C10	PT RIKILT Food Allergens 2014 C homogeneity 08

Document 1: Instruction letter sent to the participants with the samples of the Proficiency Test.

Dear participant,

Thank you very much for your interest in the proficiency study for the analysis of allergens in food.
Hereby I send you a parcel containing three randomly coded samples. Each sample consists of approximately 20 grams of baby cereal. The samples may contain one or more of the following allergens (in alphabetical order):

Almond
Gluten
Milk powder (skim)
Peanut
Soy flour

Please fill out the accompanied 'acknowledgement of receipt form' and return it immediately upon receipt of the samples, preferably by e-mail.

Instructions:
- *After arrival store the samples at room temperature (20ºC- 25 ºC).*
- *Homogenize the samples before analysis according to your laboratory's procedures.*
- *Please analyze the samples according to your routine methods and make use of your own reference standards.*
- *Carry out a **single analysis** for each sample.*
- *The deadline for this test is **November 28th 2014.***
- *Please use the web application for entering your **analysis results only**. For description of the method, please fill out attached excel file. Also please ignore all other questions on details of methods etc. (https://crlwebshop.wur.nl/apex/f?p=307:1000).*
- *Your username is: …*
- *Your password is: …*
- *Your lab code to enter this proficiency test is: …*
- *For entering results:*
    - *Please describe your method, e.g. PCR/ELISA/MS/dipstick and in-house or commercial (supplier!) in the excel file attached, not in the webap.*
    - *For quantitative methods:*
        - *please report your results as follows:*
          mg/kg almond
          mg/kg gluten
          mg/kg skim milk powder
          mg/kg peanut
          mg/kg soy flour
          *If you obtain protein concentrations, e.g. 1 mg/kg almond **protein**, please clearly indicate this and if possible please convert this result to allergen concentration, i.e. mg/kg almond. Please state the conversion factor in the description of the method. If the kit insert does not provide a conversion factor we suggest to use the VITAL typical protein reference table, http://allergenbureau.net/wp-content/uploads/2013/11/VITAL_Allergenic_Protein_Levels_10_11_08.pdf*

If you do not detect the presence of an allergen, please report "not detected".

- *For qualitative methods:*
  - *Please report "detected" or "not detected".*

Please contact me if you have any questions or need any assistance.

Kind regards,
Dr. Ir. Monique Bremer
Monique.Bremer@wur.nl

Document 2: Acknowledgement of receipt form sent to the participants with the samples of the Proficiency Test.

Acknowledgement of receipt
Proficiency test for Food Allergens

Please fill out this form and return it immediately upon receipt of the samples, preferably by e-mail, to:

Monique.Bremer@wur.nl
or
Fax: +31 317 417717

RIKILT
WAGENINGEN UR

Laboratory code:

Hereby, we confirm the receipt of the following samples:

	Sample Code
1	*Food allergens/2014/*
2	*Food allergens/2014/*
3	*Food allergens/2014/*

☐ *The samples were in a good condition*
☐ *The samples were not in good condition, namely:*

Date of receipt (dd-mm-yy):…….. *Signature:……..*

Table 8: Information of validation testing for Almond ELISA Kit (BioFront Technologies).

Material	Dilution scheme	Expected almond concentration (mg/kg)	Detected almond concentration (mg/kg)
Material A	1g material A	20	9.9
	0.5g blank matrix +0.5g material A	10	3.8
	0.75g blank matrix +0.25g material A	5	2.2
	0.875g blank matrix +0.125g material A	2.5	1.1
Material B	1g sample B	10	3.5
	0.5g blank matrix +0.5g material B	5	1.8
	0.75g blank matrix +0.25g material B	2.5	0.8
	0.875g blank matrix +0.125g material B	1.25	0.5

Table 9: Information of validation testing for Alertox ELISA Allergen-Almond (Biomedal Diagnostics) kit.

Material	Dilution scheme	Expected almond concentration (mg/kg)	Detected almond concentration (mg/kg)
Material B	0.5g material B	10	6.0
	0.25g blank matrix +0.25g material B	5	3.2
	0.375g blank matrix +0.125g material B	2.5	2.4
	0.4375g blank matrix +0.0625g material B	1.25	0.9

Table 10: Information of validation testing for RIDASCREEN® Gliadin (R-Biopharm) kit.

Material	Dilution scheme	Expected gluten concentration (mg/kg)	Detected gluten concentration (mg/kg)
Material C	0.25g material C	30	25.4
	0.125g blank matrix + 0.125g material C	15	16.5
	0.1875g blank matrix + 0.0625g material C	7.5	8.8
	0.21875g blank matrix + 0.03125g material C	3.75	5.6

Table 11: Information of validation testing for RIDASCREEN®FAST Peanut (R-Biopharm) kit.

Material	Dilution scheme	Expected peanut concentration (mg/kg)	Detected peanut concentration (mg/kg)
Material A	1g material A	5	9.0
	0.5g blank matrix +0.5g material A	2.5	4.9
	0.75g blank matrix +0.25g material A	1.25	1.4
	0.875g blank matrix +0.125g material A	0.625	0

Table 12: Information of validation testing for RIDASCREEN®FAST Milk (R-Biopharm) kit.

Material	Dilution scheme	Expected skim milk powder concentration (mg/kg)	Detected skim milk powder concentration (mg/kg)
Material A	1g material A	20	9.3
	0.5g blank matrix +0.5g material A	10	4.0
	0.75g blank matrix +0.25g material A	5	1.7
	0.875g blank matrix +0.125g material A	2.5	0.9

Table 13: Information of validation testing for RIDASCREEN®FAST Soya (R-Biopharm) kit.

Material	Dilution scheme	Expected soy flour concentration (mg/kg)	Detected soy flour concentration (mg/kg)
Material B	1g material B	25	32.2
	0.5g blank matrix +0.5g material B	12.5	19.6
	0.75g blank matrix +0.25g material B	6.25	12.2
	0.875g blank matrix +0.125g material B	3.125	11.5

Table 28: Overview of the applied methods by the participants of the PT.

Lab code	Almond	Gluten	Skim milk powder		Peanut	Soy flour
PT194	Almond ELISA Kit (BioFront Technologies)	-	-		Peanut ELISA Kit (BioFront Technologies)	-
PT206	RIDASCREEN®FAST Mandel / Almond (R-Biopharm)	RIDASCREEN® Gliadin (R-Biopharm)	RIDASCREEN®FAST Casein (R-Biopharm)	RIDASCREEN® β-Lactoglobulin (R-Biopharm)	RIDASCREEN®FAST Peanut (R-Biopharm)	ESSOYPRD-48 (ELISA Systems)
PT210	-	RIDASCREEN Gliadin (Bioscience Diagnostics Pte Ltd)	-		Biokit Peanut Assay Kit (Chokim Scientific (S) Pte Ltd)	-
PT211	RIDASCREEN®FAST Mandel / Almond (R-Biopharm)	RIDASCREEN® Gliadin (R-Biopharm)	RIDASCREEN®FAST Milk (R-Biopharm)		RIDASCREEN®FAST Peanut (R-Biopharm)	RIDASCREEN® FAST Soya (R-Biopharm)
PT212	-	In house method analysis : Determination Gluten in Food / RIDASCREEN. Enzyme Immunoassay for the Quantitative Analysis of Gliadins and Corresponding Prolamin (R- Biopharm)	-		-	-
PT213	Veratox for Almond Allergen (Neogen)	RIDASCREEN® Gliadin (R-Biopharm)	ESCASPRD-48 (ELISA Systems)		Veratox for Peanut Allergen (Neogen)	Veratox for Soy (Neogen)
PT214	Veratox for Almond Allergen (Neogen)	RIDASCREEN® Gliadin (R-Biopharm)	ESCASPRD-48 (ELISA Systems)		Veratox for Peanut Allergen (Neogen)	Veratox for Soy (Neogen)
PT215	Immunolab Almond ELISA (Immunonlab GmbH)	Immunolab Gliadin/Gluten ELISA (Immunonlab GmbH)	Immunolab Casein ELISA (Immunonlab GmbH)		Immunolab Peanut ELISA (Immunonlab GmbH)	Immunolab Soja (Soy) ELISA (Immunonlab GmbH)
PT216	ESARD-48 (ELISA Systems)	ESGLISS-48 (ELISA Systems)	ESCASPRD-48 (ELISA Systems)		ESPRDT-48 (ELISA Systems)	ESSOYPRD-48 (ELISA Systems)
PT217	Veratox for Almond Allergen (Neogen)	RIDASCREEN® FAST Gliadin (R-Biopharm)	Veratox (Neogen)		Veratox for Peanut Allergen (Neogen)	Veratox for Soy (Neogen)
PT218						
PT219	RIDASCREEN®FAST Mandel / Almond (R-Biopharm)	RIDASCREEN® Gliadin (R-Biopharm)	AgraQuant® ELISA Casein (Romer Labs)		BioKits Peanut Assay Kit (Neogen)	Veratox for Soy (Neogen)

Lab code	Almond	Gluten			Skim milk powder				Peanut	Soy flour
PT220	ESARD-48 (ELISA Systems)	Aller-Tek™ Gluten ELISA (ELISA Technologies)			ESCASPRD-48 (ELISA Systems)				RIDASCREEN®FAST Peanut (R-Biopharm)	ESSOYPRD-48 (ELISA Systems)
PT221	ESARD-48 (ELISA Systems)	ESGLISS-48 (ELISA Systems)			ESCASPRD-48 (ELISA Systems)				ESPRDT-48 (ELISA Systems)	ESSOYPRD-48 (ELISA Systems)
PT222	CER Almond Kit (In-house developed test)	-			CER Casein Kit (In-house developed test)				-	-
PT223	Veratox for Almond Allergen (Neogen)	Veratox for Gliadin R5 (Neogen)			Veratox for Total Milk Allergen (Neogen)				Veratox for Peanut Allergen (Neogen)	Veratox for Soy (Neogen)
PT224	-	RIDASCREEN® Gliadin (R-Biopharm)			RIDASCREEN®FAST Casein (R-Biopharm)				RIDASCREEN®FAST Peanut (R-Biopharm)	-
PT225	Alertox ELISA Allergen-Almond (Biomedal Diagnostics)	INGEZIM Gluten (INGENASA)	GlutenTox ELISA Sandwich (Biomedal Diagnostics)	Gluten Tox Stick Plus (Biomedal Diagnostics)	Alertox ELISA Allergen-Casein (Biomedal Diagnostics)	Alertox ELISA Allergen- β-lactoglobulin (Biomedal Diagnostics)	Alertox Stick Casein (Biomedal Diagnosticks)	Alertox Stick β-lactoglobulin (Biomedal Diagnostics)	-	Alertox ELISA Allergen-Soy (STI) (Biomedal Diagnostics)
PT226	Almond ELISA Kit (BioFront Technologies)	Aller-Tek™ Gluten ELISA (ELISA Technologies)			-				Peanut ELISA Kit (BioFront Technologies)	-
PT227	-	Gluten-Check (Bio-Check (UK))			ESMRDBLG-48 (Elisa Systems)				-	ESSOYPRD-48 ELISA Systems)
PT228	-	-			RIDASCREEN®FAST β-Lactoglobulin (R-Biopharm)				-	RIDASCREEN® FAST Soya (R-Biopharm)
PT229	Almond ELISA Kit (BioFront Technologies)	-			-				Peanut ELISA Kit (BioFront Technologies)	-

Table 29: Results reported by participating laboratories for material A.

Lab code	Almond (mg/kg)	z'ai-score	Gluten* (mg/kg)	Peanut (mg/kg)	z'a-score	Skim milk powder (mg/kg)	z'ai-score	Soy flour* (mg/kg)
PT194	15.01	0.21		1.96	**-2.11**			
PT206	10.6	-0.94	<10	8.6	1.17	<5		<2.5
PT210			nd	2.58	-1.80			
PT211	21.3	1.90	nd	8.5	1.12	10.8	0.69	12.6
PT212			5.64					
PT213	14.7	0.13	5.3	5.6	-0.31	5.5	-1.03	nd
PT214	16	0.48	6	5.5	-0.36	20	**3.75**	nd
PT215	15	0.21	82	12	**2.85**	22	**4.41**	nd
PT216	6.8	-1.93	0	0		2.9	-1.86	0
PT217	13	-0.32	<10	3	-1.59	4	-1.51	<1
PT218	nt		nd	11.6	**2.65**	7.9	-0.26	nd
PT219	23	**2.36**	Traces	4.2	-1.00	16	**2.42**	nd
PT220	10.7	-0.91	10.4	10.4	**2.06**	11.4	0.89	nd
PT221	12.8	-0.37	nd< 2	nd <3.3		2.6	-1.95	nd < 2.9
PT222	1.6	**-3.27**				2.5	-1.99	
PT223	13	-0.32	nd	6.1	-0.06	nt		nd
PT224			nd	4.68	-0.76	5.29	-1.09	
PT225	15	0.21	A[1]:<5 / B[1]:<0.6 / C[1]:-			A[2]:<14.4 / B[2]:6.1 / C[2]: + / D[2]: +	A[2]:1.89 / B[2]:-0.83	<1
PT226	22.3	**2.17**	10.1	2.8	-1.69	nt		nt
PT227	nt		nd	nt		0.5	**-2.62**	nd
PT228	nt		nt	nt		30.6	**7.26**	9.1
PT229	+			+				

*materials were not spiked with gluten and/or soy flour. The respective results are presented for information only.
nd= not detected. nt= not tested
+: Positive qualitative result. -: Negative qualitative result
A[1]: INGEZIM Gluten (INGENASA) kit, B[1]: GlutenTox ELISA Sandwich (Biomedal Diagnostics) kit, C[1]: Gluten Tox Stick Plus (Biomedal Diagnostics)
A[2]: Alertox ELISA Allergen-Casein (Biomedal Diagnostics) kit, B[2]: Alertox ELISA Allergen- β-lactoglobulin (Biomedal Diagnostics) kit, C[2]: Alertox Stick Casein (Biomedal Diagnostics), D[2]: Alertox Stick β-lactoglobulin (Biomedal Diagnostics)
The numbers in bold indicate the questionable and unsatisfactory results

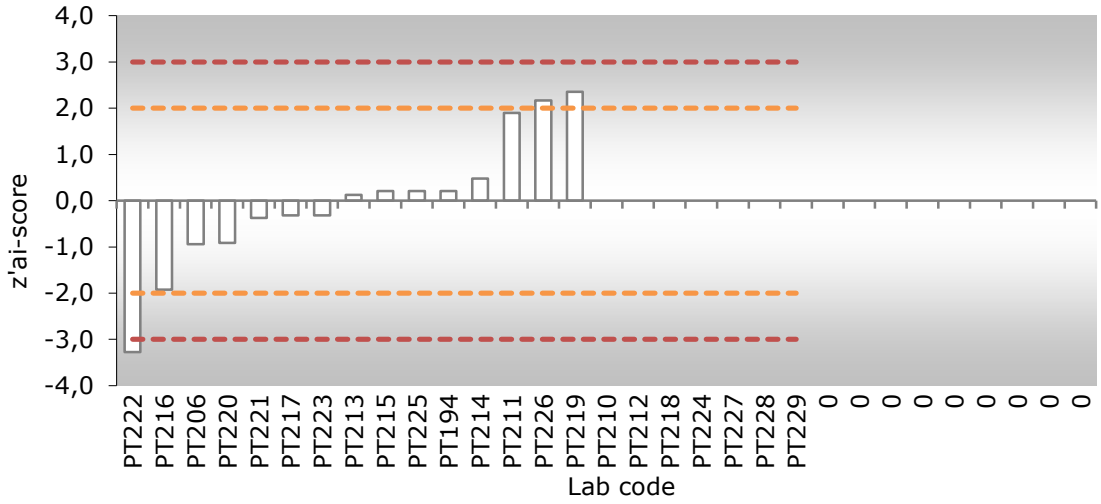

Figure 7: Graphical representation of z'ai-scores for almond in material A. The $X \pm 2\sigma_P$ and $X \pm 3\sigma_P$ lines (yellow and red respectively) are calculated according to the equation IV in § 2.2.9.4.

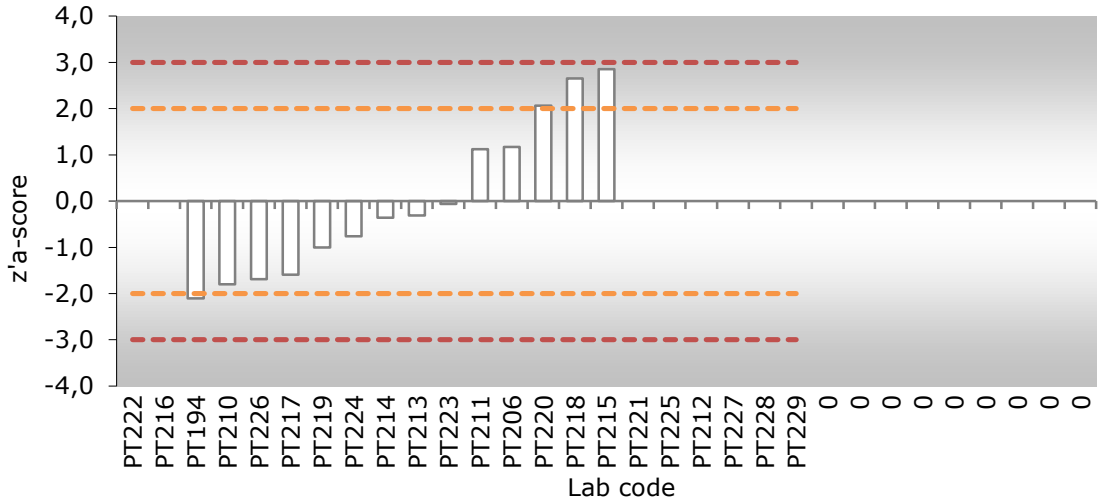

Figure 8: Graphical representation of z'a-scores for peanut in material A. The z'a-scores are presented for information only and not for evaluation of laboratories, as the uncertainty of the assigned value is very high. The $X \pm 2\sigma_P$ and $X \pm 3\sigma_P$ lines (yellow and red respectively) are calculated according to the equation II in § 2.2.9.4.

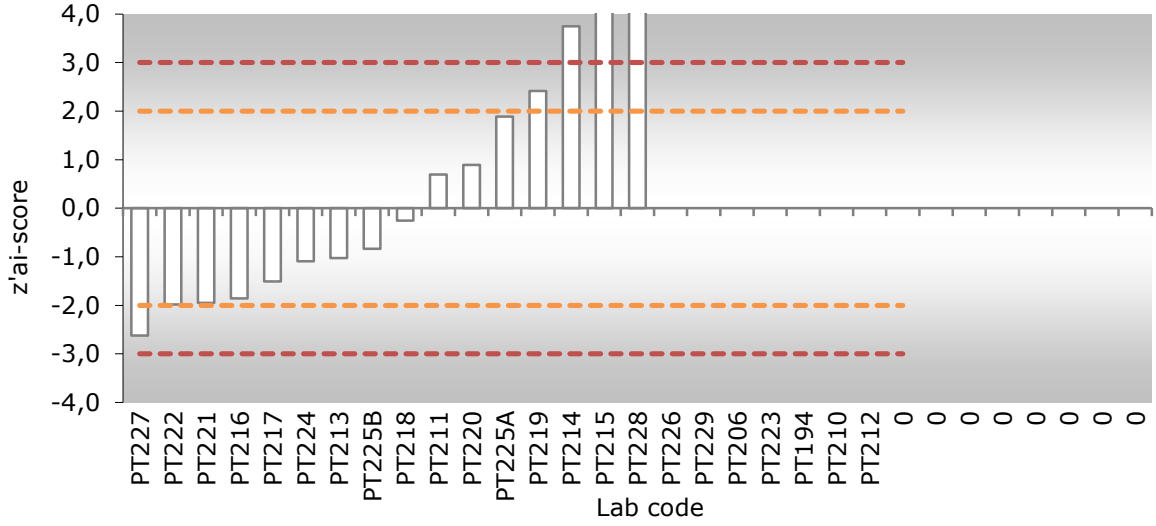

Figure 9: Graphical representation of z'ai-scores for skim milk powder in material A. The z'ai-scores are presented for information only and not for evaluation of laboratories, as the uncertainty of the assigned value is very high. The $X \pm 2\sigma_P$ and $X \pm 3\sigma_P$ lines (yellow and red respectively) are calculated according to the equation IV in § 2.2.9.4.

.

Table 30: Results reported by participating laboratories for material B.

Lab code	Almond (mg/kg)	z'ai-score	Gluten (mg/kg)	z'ai-score	Peanut* (mg/kg)	Skim milk powder* (mg/kg)	Soy flour (mg/kg)	z'a-score
PT194	2.23	**-2.04**			nd			
PT206	3.8	-1.22	18	1.27	<2.5	<5	<2.5	
PT210			12.14	-0.23	nd			
PT211	10.7	**2.47**	14.1	0.24	nd	nd	25.5	1.02
PT212			14.94	0.46				
PT213	6.3	0.09	17	1.00	nd	1.7	nd	
PT214	7.6	0.79	10	-0.70	nd	nd	nd	
PT215	8	1.01	466	**118.85**	nd	nd	41	**2.72**
PT216	4.1	-1.06	7	-1.36	0	0	0	
PT217	5	-0.59	16	0.74	<1	<1	2	-1.56
PT218	nt		nd		nd	nd	nd	
PT219	13	**3.71**	11	-0.48	nd	nd	2.5	-1.51
PT220	5.9	-0.12	73.9	**15.94**	nd	nd	nd	
PT221	5.3	-0.43	7	-1.36	nd < 3.3	nd < 0.5	pos < 5.8	
PT222	0.6	**-2.89**				nd		
PT223	7.8	0.90	7	-1.36	nd	nt	nd	
PT224			7.48	-1.25	nd	nd		
PT225	7.4	0.68	A^1:9 / B^1:12.5 / C^1:+	A^1:-0.92 / B^1:-0.15		A^2:<0.9 / B^2:<0.9 / C^2: - / D^2: -	12.5	-0.41
PT226	20.8	**7.93**	55.3	**11.06**	nd	nt	nt	
PT227	nt		6.2	-1.53	nt	nd	3	-1.45
PT228	nt		nt		nt	nd	38.5	**2.44**
PT229	+				-			

*materials were not spiked with peanut and/or skim milk powder. The respective results are presented for information only.

nd= not detected, nt= not tested

+: Positive qualitative result, -: Negative qualitative result

A^1: INGEZIM Gluten (INGENASA), B^1: GlutenTox ELISA Sandwich (Biomedal Diagnostics), C^1: Gluten Tox Stick Plus (Biomedal Diagnostics)

A^2: Alertox ELISA Allergen-Casein (Biomedal Diagnostics) kit, B^2: Alertox ELISA Allergen- β-lactoglobulin (Biomedal Diagnostics) kit, C^2: Alertox Stick Casein (Biomedal Diagnostics), D^2: Alertox Stick β-lactoglobulin (Biomedal Diagnostics)

The numbers in bold indicate the questionable and unsatisfactory results

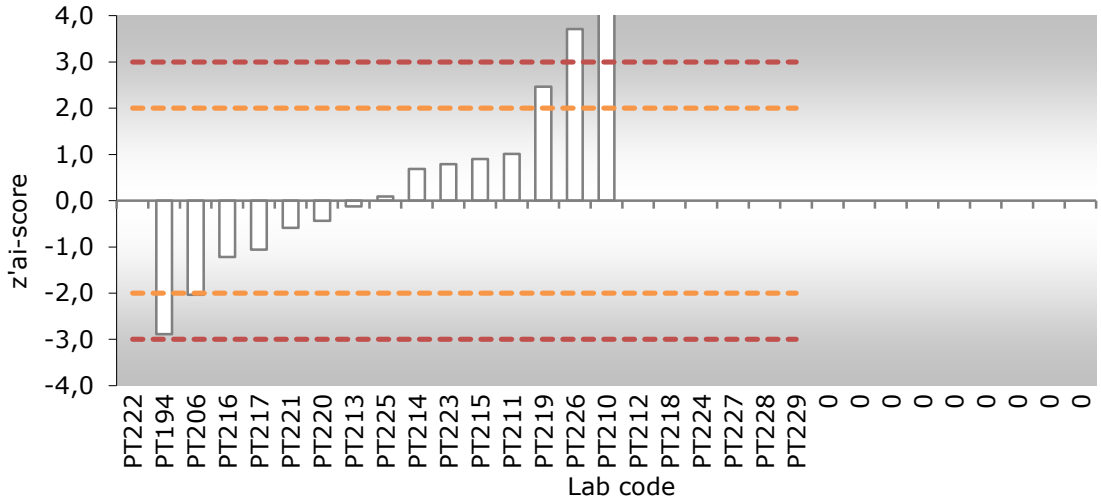

Figure 10: Graphical representation of z'ai-scores for almond in material B. The $X\pm2\sigma_p$ and $X\pm3\sigma_p$ lines (yellow and red respectively) are calculated according to the equation IV in § 2.2.9.4.

.

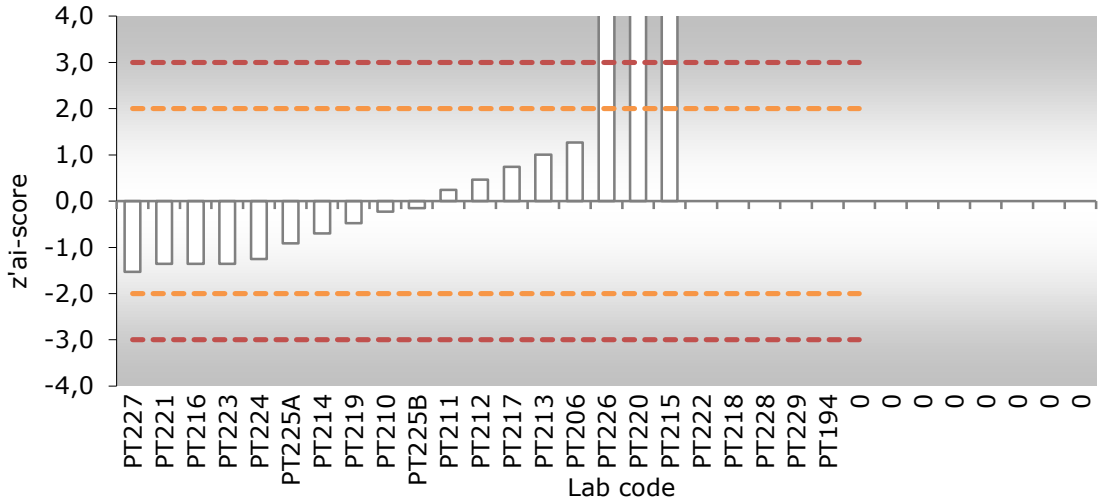

Figure 11: Graphical representation of z'ai-scores for gluten in material B. The $X\pm2\sigma_p$ and $X\pm3\sigma_p$ lines (yellow and red respectively) are calculated according to the equation IV in § 2.2.9.4.

.

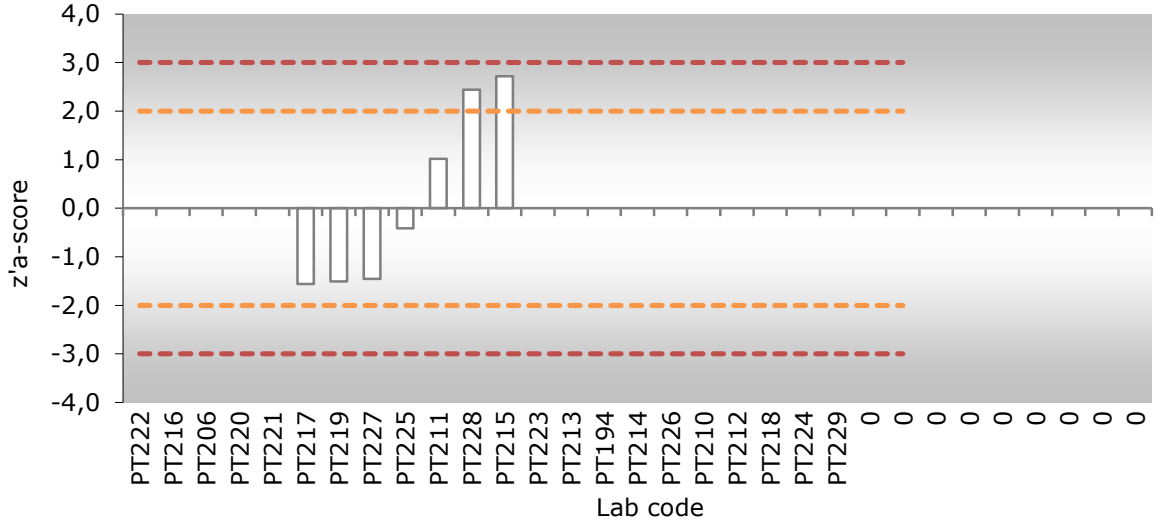

Figure 12: Graphical representation of z'a-scores for soy flour in material B. The z'a-scores are presented for information only and not for evaluation of laboratories, as the uncertainty of the assigned value is very high. The $X \pm 2\sigma_P$ and $X \pm 3\sigma_P$ lines (yellow and red respectively) are calculated according to the equation II in § 2.2.9.4.

Table 31: Results reported by participating laboratories for material C.

Lab code	Almond* (mg/kg)	Gluten (mg/kg)	z'ai-score	Peanut (mg/kg)	z'a-score	Skim milk powder (mg/kg)	z'ai-score	Soy flour* (mg/kg)
PT194	1.33			2.84	**-2.65**			
PT206	<2.4	33	-0.02	29.6	1.87	<5		<2.5
PT210		34.14	0.11	12.5	-1.02			
PT211	nd	27.3	-0.57	25.4	1.16	nd		12.5
PT212		35.72	0.28					
PT213	nd	48.1	1.64	18.1	-0.07	4.2	0.79	nd
PT214	nd	31	-0.21	23	0.76	6.6	**2.56**	nd
PT215	nd	1678	**180.60**	32	**2.28**	4.7	1.16	nd
PT216	0	22.3	-1.05	9.4	-1.54	0.8	-1.67	0
PT217	<1	30	-0.31	14	-0.77	1	-1.52	<1
PT218	nt	nd		38.5	**3.38**	1	-1.52	nd
PT219	nd	32	-0.11	12	-1.10	4.5	1.02	nd
PT220	nd	238	**22.49**	28.3	1.65	no detect		nd
PT221	nd< 1.2	22	-1.08	pos < 6.7		pos < 1.0		nd < 2.9
PT222	nd					0.6	-1.81	
PT223	nd	21	-1.18	20	0.25	nt		nd
PT224		40.65	0.82	13.17	-0.91	< LOD		
PT225	< 1	A^1:18.9 / B^1:21.9 / C^1: +	A^1:-1.38 / B^1:-1.09			A^2:6.5 / B^2:1.3 / C^2: + / D^2: +	A^2:**2.49**	B^2:-1.31 / <1
PT226	1	101.5	**7.50**	3.8	**-2.49**	nt		nt
PT227	nt	25	-0.79	nt		nd		nd
PT228	nt	nt		nt		nd		14.1
PT229	+			+				

*materials were not spiked with soy flour and/or almond. The respective results are presented for information only.
nd= not detected
nt= not tested
+: Positive qualitative result, -: Negative qualitative result
A^1: INGEZIM Gluten (INGENASA), B^1: GlutenTox ELISA Sandwich (Biomedal Diagnostics), C^1: Gluten Tox Stick Plus (Biomedal Diagnostics)
A^2: Alertox ELISA Allergen-Casein (Biomedal Diagnostics) kit, B^2: Alertox ELISA Allergen- β-lactoglobulin (Biomedal Diagnostics) kit, C^2: Alertox Stick Casein (Biomedal Diagnostics), D^2: Alertox Stick β-lactoglobulin (Biomedal Diagnostics)
The numbers in bold indicate the questionable and unsatisfactory results

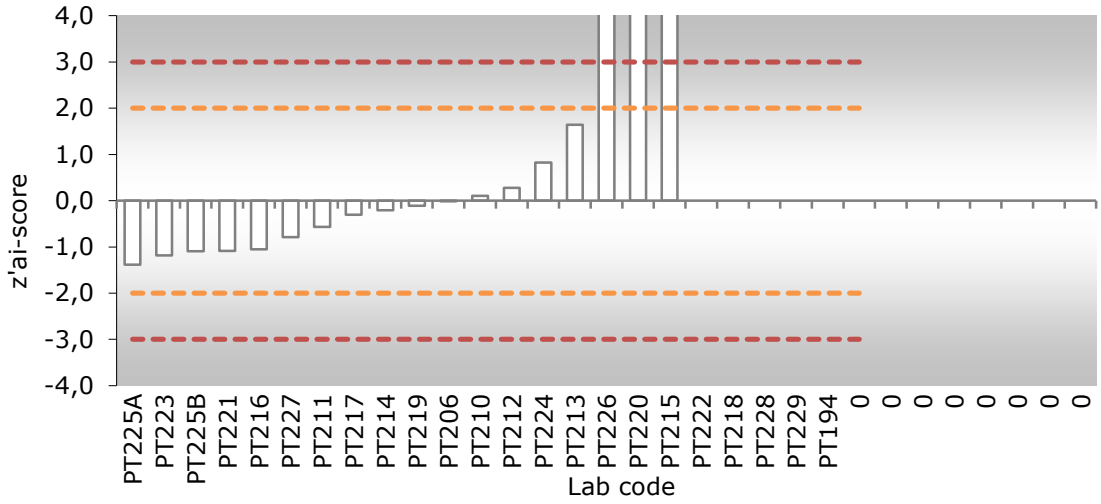

Figure 13: Graphical representation of z'ai-scores for gluten in material C. The $X \pm 2\sigma_P$ and $X \pm 3\sigma_P$ lines (yellow and red respectively) are calculated according to the equation IV in § 2.2.9.4.

.

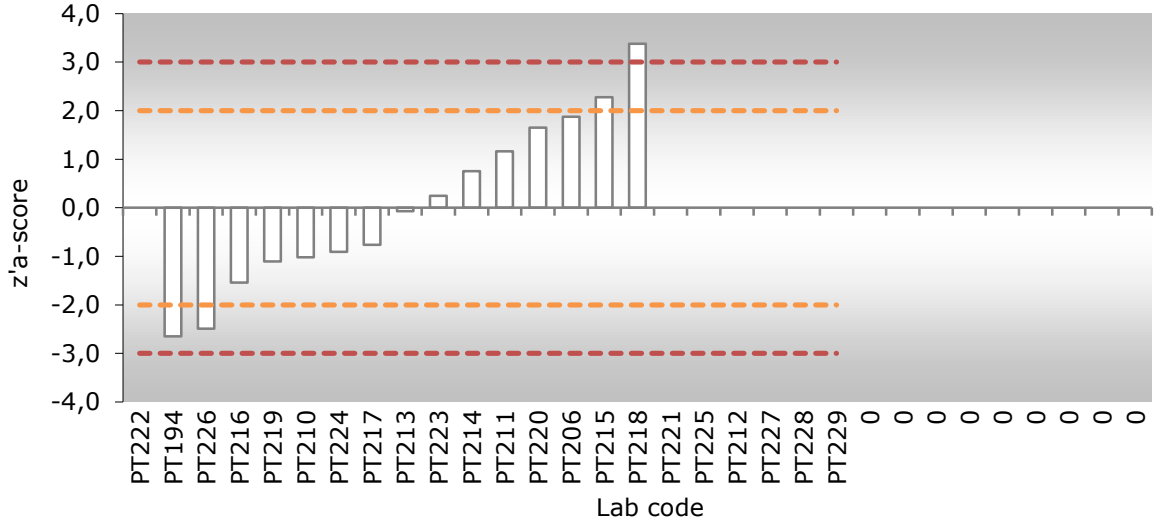

Figure 14: Graphical representation of z'a-scores for peanut in material C. The z'a-scores are presented for information only and not for evaluation of laboratories, as the uncertainty of the assigned value is very high. The $X \pm 2\sigma_P$ and $X \pm 3\sigma_P$ lines (yellow and red respectively) are calculated according to the equation II in § 2.2.9.4.

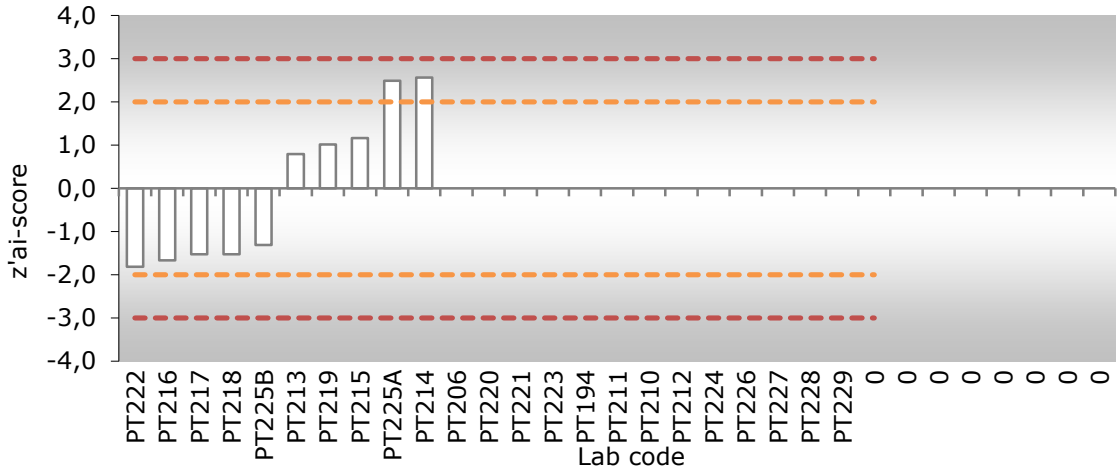

Figure 15: Graphical representation of z'ai-scores for skim milk powder in material C. The z'ai-scores are presented for information only and not for evaluation of laboratories, as the uncertainty of the assigned value is very high. The $X \pm 2\sigma_P$ and $X \pm 3\sigma_P$ lines (yellow and red respectively) are calculated according to the equation IV in § 2.2.9.4. .

Table 32: False negative results in materials which were spiked with respective allergens.

A

ALMOND	GLUTEN	PEANUT	SKIM MILK POWDER	SOY FLOUR
		FN: PT216 PT221	**FN**: PT206	

B

ALMOND	GLUTEN	PEANUT	SKIM MILK POWDER	SOY FLOUR
	FN: PT218			**FN:** PT206 PT213 PT214 PT216 PT218 PT220 PT223

C

ALMOND	GLUTEN	PEANUT	SKIM MILK POWDER	SOY FLOUR
	FN: PT218		**FN**: PT206 PT211 PT220 PT224 PT227 PT228	