FOODINTEGRITY

Ensuring the Integrity of the European food chain

613688: Collaborative Project

Seventh Framework Programme
KBBE.2013.2.4-01: Assuring quality and authenticity in the food chain

Deliverable: D4.4

Title: Guideline for validating qualitative methods including statistical analysis.

Author(s): Noelia Tena, Inmaculada Romero, Ramón Aparicio-Ruiz, Ramón Aparicio, Diego L. García-González with the contributions of Food Integrity Work Package 4 Partners.

Beneficiary(s):

Date of preparation: 12.12.2016

Period covered: 01.01.2014 -31.12.2016

Status: version 1

Dissemination level		
PU	Public	X
PP	Restricted to other participants	
RE	Restricted to a group specified by the consortium	
CO	Confidential, only members of the consortium	

The project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 613688.

TABLE OF CONTENTS

Deliverable 4.4
Guideline for validating qualitative methods including statistical analysis

1. Rationale for building a guideline for validating qualitative methods

This deliverable is focused on building a guideline for validating qualitative methods. In particular, we centred our work on techniques with a non-targeted approach that often provide binary responses when they are applied for classification purposes. Some of the new methods proposed in olive oil analysis are based on procedures that provide a binary response. Regardless the fundamental of the method (targeted or non-targeted methods), methods with a qualitative response (e.g. yes/no, authentic/non authentic) require specific parameters for evaluating the performance of the method. However, most of the effort is commonly made in the evaluation of the methods with a quantitative perspective exclusively, while the qualitative perspective is sometimes ignored. The elaboration of a general guideline could be useful to establish the basic steps that are required to take in evaluating qualitative methods. The evaluation of qualitative methods with binary response is in fact a mathematical evaluation, and the restricted vision from instrumentation should be avoided. However, a guideline of this kind is presented in the frame of an analytical work for an easier understanding by the analyst.

Although non-targeted techniques do not necessarily have to be applied as a qualitative method, these techniques are often applied to classify samples according to their chemical characteristics. A major part of WP4 has focused on testing non-targeted techniques for a specific classification task: identifying the geographical provenance of virgin olive oils (VOOs) and improving the geographical traceability. In the first meetings held in this regards, a considerable discussion was centred on what is the actual meaning of geographical traceability. Thus, this term can refer to multiple tasks such as the identification of cultivars with particular geographical relevance, climate regions, countries, counties, and protected designations of origin. Prior to doing any work, and thinking on validation of a selected methodology, it is of paramount important to define the objective of the classification task. The validation is intended to check the quality of the method fit for a particular purpose, so this purpose should be defined beforehand. In this particular case, the definition of the purpose was achieved by means of discussion and consensus. The discussion among partners and stakeholders resulted in the main objective of classifying European and non-European virgin olive oils, which was also consulted with the WP leader of WP10. This classification was considered interesting enough for consumer protection and a major authenticity of the product and it was easy to understand by all the partners and it was feasible to implement in the sampling.

All the techniques used in this project were non-targeted techniques. They were:

- NMR.
- IRMS.
- FTIR.
- Raman Spectroscopy.
- NIR.
- U-HPLC-HRMS/MS analysis.
- FAST gas-chromatography.

One of the objectives of the work package is to reinforce rapid methods to solve current safety/quality demands of olive oil industry, with a particular interest in geographical provenance from new producing countries. One of the requirements to reinforce non-targeted techniques is to increase the trust of the olive oil actors (mainly regulatory bodies, producers and consumers) on these techniques and this should be done by testing them with a proper validation procedure. In general terms, the reliability of non-targeted techniques has been affected by the following problems to be solved or improved:

1) Lax statistical analysis, sometimes with overoptimistic results that are not checked with external samples, and difficult to reproduce.
2) Tentative or partial studies that use a biased set of samples that do not represent the variability of the universe of discernment.
3) Selection of a wrong biomarker or unknown biomarkers, which prevent a full chemical understanding of the method.
4) Development of application based on already solved or illogical kinds of adulteration (e.g. refined oil in VOO, crude hazelnut oil in VOO)

These four points can be solved by means of collaborative studies that promote the discussion between groups and the inter-comparison of different analytical approaches. Thus, a better understanding of the methodology is required prior to any validation procedure. For that reason, each one of the partners was asked to produce a protocol in order to understand the output of each analytical technique. Most of the qualitative methods consist of quantitative methods with a qualitative response. This qualitative response (e.g. binary response) is obtained with multivariate statistical analysis. That means that these statistical tools should be also tested when validating since they form part of the methodology. For that reason, it was considered very important to harmonize the data treatment as much as possible. PLS, as supervised multivariate analysis, was suggested to the partners as a common statistical tool to promote the inter comparison of results. PLS requires a

database, and in fact, most of the qualitative methods being suggested for olive oil authenticity are based on databases. For that reason, in order not to interfere in the validation work, a special care was taken in the sample selection.

In the course of the project, it was also necessary to discuss extensively the meaning of each one of the parameters to be calculated as well as the importance of them. One of the major difficulty in validation of qualitative methods is the lack of tradition in the validation of those methods in the regulatory bodies. In the particular case of virgin olive oil, all the standard methods (Figure 1.1) are properly validated by the International Olive Oil Council (IOC). However, the validation of methods with qualitative response is not so common. Thus, for example, the validation of the decision trees that consists in several criteria that results in a "genuine/non-genuine" answer is still under discussion in the IOC and in the expert groups on olive oil in the EU.

In addition to the difficulty described above, the high number of methods with a qualitative response (a binary response in most of the cases) that are being proposed for complex authentication tasks are increasing the interest in validating properly qualitative methods. The reason of this higher number of methods is the fact that authenticity of virgin olive oil has been perfected for years, and the current authenticity tasks are related with more sophisticated issues such as distinguishing virgin olive oils by geographical locations (i.e., neighbouring PDOs) or virgin olive oil spiked with tailored oils with a chemical composition that is similar to VOOs. The only strategy for solving this authenticity issues is analysing a large data set that requires supervised statistical procedures that provide a nominal response, and therefore a validation procedure for qualitative methods should be implemented.

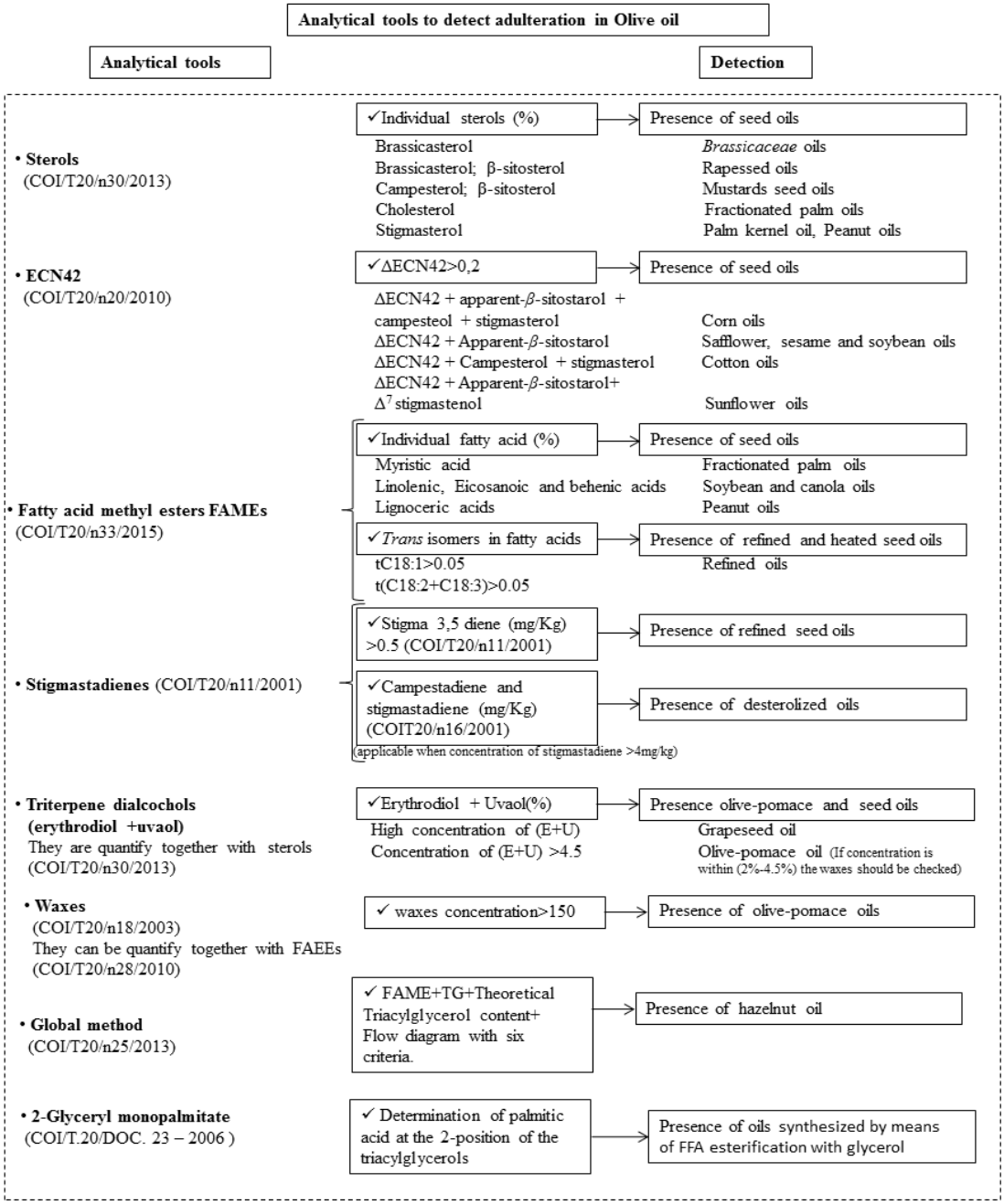

Figure 1.1. Standard analytical tools for virgin olive oil authentication.

2. Characteristic aspects in validation of non-targeted techniques with a binary response: A case study focused on virgin olive oil

A non-targeted method is typically sensitive to multiple chemical species and therefore it is necessary to couple the results (e.g. absorbance values) with a statistical model that classifies the samples into two (or more) groups. In complex classification tasks, such as the identification of geographical provenance of virgin olive oils, the objective of the method is not to identify and quantify a specific chemical compound, but to characterize the whole chemical profile. Therefore, it

is not possible to apply a simple calibration of the method for a well-defined purpose. For that reason, it is necessary to carry out the development and validation of the method to ensure the lack of false positive and negatives when it is implemented under a regular basis.

Considering that non-targeted techniques can be applied to many uses, the peculiarities of the validation process should be adapted to each application and basic guidance is still necessary when a new method is evaluated. This guidance has to be flexible enough to be adapted to each case, since qualitative analysis in food authenticity is usually a complex task in which methods of very different nature and fundamentals are used and tested. Nevertheless, some common actions can be defined for all the application areas.

The validation of a non-targeted technique for a qualitative analysis (e.g. binary answer) pursues assuring the consistency between the intermediate results yielded by the technique (e.g. Intensity signal) and the final required answer (e.g. European/non-European virgin olive oils). Thus, even though some techniques may have excellent values of analytical quality parameters in quantitative applications (e.g. precision), they may fail in providing a binary answer and therefore a new approach is needed to evaluate this ability. In this regards, the validation includes the evaluation of the whole process including the steps of building a training set, carrying out the instrumental analysis and implementing the statistical processing. In general terms, the validation procedure is intended to determine the false positive and false negative rates at different concentration levels of the analyte and all the subsequent parameters described in the following section (e.g. specificity and sensitivity). In other words, the objective is to evaluate model boundaries in which, after applying multivariate statistical analysis, the sample can be defined as belonging to one class or another (Figure 2.1).

In order to examine this boundary zone, it can be said that it is convenient to frame the validation process within a specific experimental approach for a better comprehension of the analysis and to evaluate better that the “fitness for purpose” required in validation is fulfilled. With a generalization purpose, the following steps could be remarked, which have been those carried out within FoodIntegrity project in WP4:

1) **Definition of the objective:** As it was already mentioned, a well-defined objective is critical since the false positives and negatives are defined according to this general objective. Thus, for example, in FoodIntegrity project, the general objective of geographical traceability was already defined in the DOW. Further discussion was carried out among the partners for defining this objective in depth before starting the analytical work. Thus, geographical traceability is a general term that means many features such as distinguishing samples by country, continent, county,

PDO, etc. The final selection of the objective (European virgin olive oils vs. Non-European virgin olive oils) permitted to address the selection of the samples and the best methodologies to be tested. The definition of the objective needs to be achieved in a consensual and decisive manner and it should be simple enough to be converted into a question that will result in the binary answer.

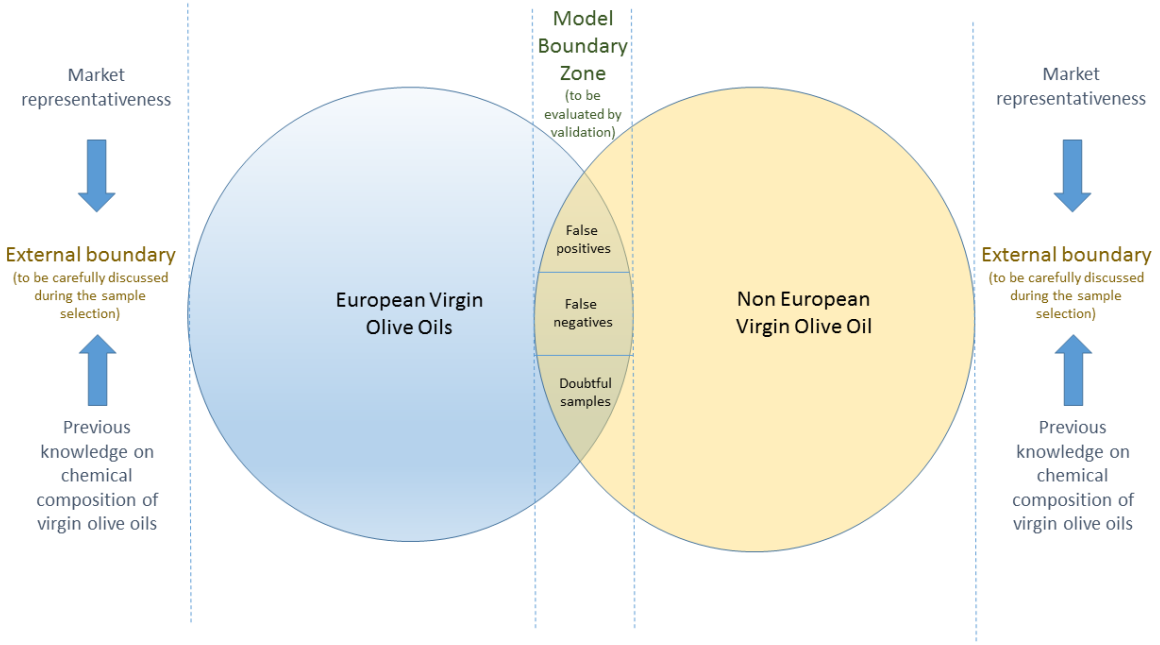

Figure 2.1. Scheme of the boundaries defined in a case study of geographical identification of virgin olive oil (European and non-European) carried out in FoodIntegrity project.

2) **Sample selection for training set:** It also requires discussion between the partners taking part in the validation study to get a compromise between the number of samples of each class, the required number for getting a robust model, and the feasibility of sampling with the allocated budget. Thus, for example, some techniques may require more samples than others, or some of them may already have some databases from previous studies. In the case of techniques with existing databases, these techniques should be evaluated only with the common data set to obtain comparable results.

In sample selection it is important to establish the membership of the selected samples to one of the groups with no doubt. In the case of applications where this is not possible because there are not well accepted criteria for doing so, then a "temporary label" can be established to be revised, if needed, in further steps. In FoodIntegrity project, for example, we have selected samples directly from producers, some of them being stakeholders, to make sure that the virgin olive oils were from inside/outside Europe with no mixing with oils from other sources.

The ability of each sample to reflect the property under study (geographical provenance in our case) should be guaranteed by collecting samples with a large diversity of properties (mainly pedoclimatic effects and cultivars in our application). The sample collection should consider the whole universe of possible samples. However, sometimes it is not affordable. In this case, it is very important to pay attention to the external boundaries as defined in Figure 2.1. That is to say, it is important to consider which types of samples are included within each class, with their own variability. Thus, there are many cultivars (olive tree varieties) of virgin olive oils, some of them being very minor in production, that can be (or not) included in the model. Thus, in general terms, two criteria to consider in sample selection is the previous knowledge in chemical variability of the samples (in olive oil that knowledge was high, as stated in Deliverable 4.1) and the representativeness of each sample in the real market at world level.

The requested information for each sample has to be recorded according to the defined objective (e.g. if the oils was collected in a European olive mill or not). However, some other information may be needed in further steps of the study. For example, a question arose about if the oil was filtered or not, because it seems to affect the results in some cases. For that reason, it is advisable to have an open communication line with the sources (e.g. suppliers) of the samples.

The collection of samples for the training set should be a reproducible procedure so it could be repeated at any time. However, working with complex matrices that depends on many factors may hinder this purpose. In validation works, reference materials that encompass all the basic properties of the defined groups (e.g. chemical characteristics that European oils have and non-European oils do not have or vice versa) are always desirable. However, in qualitative analysis with non-targeted techniques, where the analytical tasks are highly complex, that is hardly managed and it is considered to be hardly challenging. For that reason, and to give support to the validation procedure, detailed information about how the sampling was done is highly recommended.

3) **Information gathering on non-targeted techniques:** Prior to starting the validation, a discussion about which analytical techniques will be tested is required. One of the main difficulty of testing and comparing different non-targeted techniques is that they have quite diverse physical-chemical fundamentals, type of delivered results and different way of data processing. Concerning this diversity, it is advisable to collect information of the protocols carried out in each case with a double purpose:

   - To understand the working principle of the methods and to devise a validation procedure that is easily adaptable to every case.

- To predict if the method has enough sensitivity or is adequate to measure those chemical markers characterizing the properties that define the binary response, and therefore, to decide whether the method is eligible to be tested. This prediction is often difficult to carry out because in many cases it is not possible to know beforehand what are the chemical markers. However, it is possible to evaluate the adequacy of the method considering its application to the defined objective. Thus, for example, in geographical traceability, those methods measuring chemical markers that are labile (phenols) would be less appropriate than others that are sensitive to non-labile markers (sterols, fatty acids, etc.).

In FoodIntegrity-WP4, the partners provided a protocol of their methods for this information gathering.

4) **Consensus on data processing:** When validating a method based on non-targeted techniques, actually the associated multivariate statistical processing is also under evaluation. In consequence, in order to obtain comparable information between the methods being tested, it is recommended to establish a common statistical procedure. Thus, the result of sample allocation in the defined classes will depend of the technique used rather than the statistical processing. However, once again a flexibility is necessary in this regards because participants should be able to apply, in addition, other statistical processing to improve the results. It is important to note that the classification tasks in food authenticity are often quite complex and in some circumstances the high complexity of data requires using advanced statistical tools for obtaining satisfactory results. If statistical tools other than the suggested for all the participants are used, it is recommended to provide the description of these new methods.

In FoodIntegrity project we suggested to use PLS as the common supervised statistical procedure. We also suggested to apply a cross-validation (prior to the validation with blind samples), when possible, to get more information about the performance of the classification model.

In other to avoid any variability between partners not associated to the methodology, we even suggested to use the same software for all of them. We identify that the validation of the statistical processing, as part of the whole method, is one of the requirements to be carried out in validation processes of non-targeted methods in qualitative results.

5) **Reporting quality parameters in qualitative analysis:** At this stage, the quality parameters have to be provided by each participant in the validation. This is a critical step because each partner has to understand how to calculate each one of the parameters. There are several quality

parameters described in international documents (López et al., 2015, Valcárcel et al., 2003), but the most useful and practical are those derived from the number of false positives and false negatives: Rate of false positives, rate of false negatives, reliability, sensitivity, selectivity, efficiency, Youden's index, and likehood ratio. All of them are described in section 3 of this document. On the other hand, the chemometric model often produces a continues score prior to classification and these scores can be interpreted as posterior class membership probabilities. Thus, although the final answer is binary, information from probability is also important to assess method performance.

6) **Verification and blind samples:** Once the most relevant spectral features (e.g. spectral bands) are selected and the mathematical model is built with the training samples, and some quality parameters are tentatively determined, the methods are again tested with two new sample sets: verification set and a set of blind samples (Figure 2.2.). The verification samples are applied to check the model in external samples as well as to readjust the model if necessary. In complex applications such as the geographical traceability it may be convenient to include these samples in the model and to rebuild the classification equation. The classification algorithms typically search for the coefficients that provide the lowest root mean square error in the classification. Depending on the methods, more samples may be needed as well as an overtraining of the model may result in an overfitting that can lead to false positives/negatives. Thus, if verification samples are included in the model, a significant improvement should be determined in the classification reliability. Once the classification model is trained, the performance of the designed model is tested with blind samples that has to be of similar characteristics of those used in the training set.
7) **Consensus report on quality parameters:** After determining all the quality parameters from the blind samples, a discussion is also convenient to highlight those variables in techniques and methods that can be improved. Non-targeted techniques are typically affected by many variables (e.g. presentation of samples with a high diversity of accessories) and more experience is often needed in each specific application. Therefore, methods are susceptible for improvement for future works and projects. In general, the development of applications in non-targeted techniques follows the steps described in Figure 2.3.

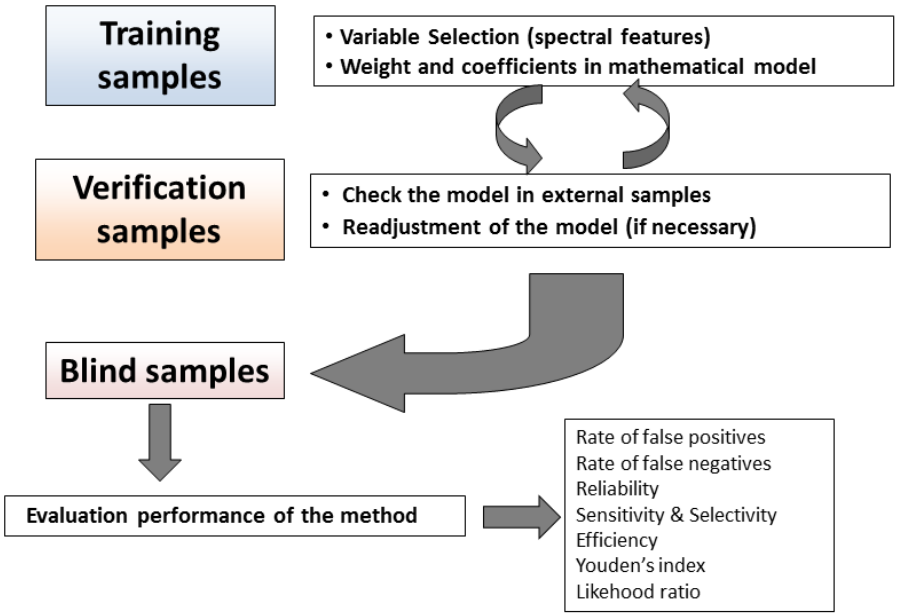

Figure 2.2. Classification of the samples divided in the three sets and the relationship between them.

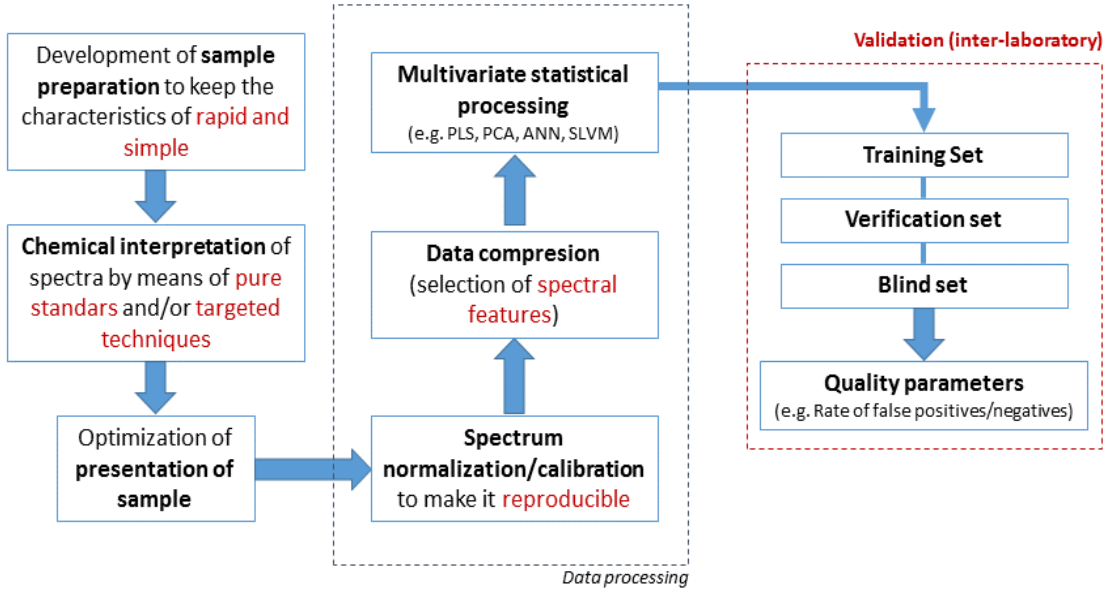

Figure 2.3. Basic steps in developing applications with non-targeted techniques.

3. Analytical quality parameters studied in the project

All the non-targeted methods proposed were evaluated with the analytical quality parameters described below considering the required binary response: identification of European vs. non-European virgin olive oils. The errors in qualitative methods are measured in terms of "false positives" and "false negatives". From a strict analytical perspective, the numbers of false positives and false negatives depend on the occurrence of the chemical markers that characterize each one of the groups and/or the required concentration to belong to each one of the groups. Thus, a "false

positive" (e.g. sample classified as a European virgin olive oil while it is not European) is obtained when a "yes" response is delivered by the statistical processing while the analysed sample does not contain those chemical markers (or at least not at the required concentration) that characterize the samples clustered in the "yes" answer (European oils). Accordingly, a false negative is wrongly assigned in the "no" samples (e.g. non-European oils) while the chemical markers associated to the "yes" answer are present and at the required concentration. The false positives and negatives arise because the classification task require a large number of chemical compounds being measured and they may interfere between them as well as with others in the food matrix. In non-targeted methods, where sample preparation is not commonly applied, interferences between analytes in the sample play a major role in the measurement and these interferences are difficult to know or predict if the sample is not studied by other methods. For that reason, in order to interpret better why false negatives and false positives are obtained it is strongly recommended to have further information about the chemical assignments of the selected bands or spectral features.

1) Rate of false positives:

It corresponds to the following formula:

$$\text{Rate of false positives} = \frac{FP}{TN + FP}$$

Where:
FP is the number of false positives, and
TN is the number of true negatives.

From a statistical point of view, a false positive is defined as error of the first kind (type 1 error) where rejection of the null hypothesis is given even though it is true. A false positive has to be also defined in regards to the specific binary response, from the viewpoint of the specific application. Thus, in our case a false positive is defined as a non-European virgin olive oil that is classified as European. This parameter can be also expressed as percentage.

2) Rate of false negatives:

It is defined as follows:

$$\text{Rate of false negatives} = \frac{FN}{TP + FN}$$

Where
FN is the number of false negatives, and
TP is the number of true positives.

A false negative is an error of the second kind (type 2 error) associated to the acceptance of a null hypothesis even though it is false. In our application, a false negative would be a European virgin olive oil that is classified as non-European.

The rates of false positives and false negatives can be regarded as a quantitative measure of the performance of the method. Their analyses go beyond the quantitative figures and often lead to discussion about why these misclassifications appear, how can be solved, and, considering the specific application, if a kind of error is preferred over the other. Thus, in food authenticity, for example, it is usually relevant to decide whether, in case of error, it is better to define a genuine sample as adulterated, or an adulterated sample as genuine. It would depend on criteria linked to the specific application (e.g. toxicological impact of the adulteration).

3) Reliability:

It is a determination that encompasses the number of false positives and false negatives and therefore it assesses on the global performance of the method and it can be regarded as the percentage of correct answers. It is defined by the following equation:

Reliability= 100-%False Positive Rate-%False Negative Rate

This equation is related to the certainty of identification and points out the unreliability zone of the method through the probability level.

4) Sensitivity:

Sensitivity is defined as the ratio between the number of true positive tests (TP) and the sum of them and the number of false negatives:

$$\text{Sensitivity}=\frac{TP}{TP+FN}$$

This sensitivity has to be differentiated from the analytical sensitivity (change in response of a measuring instrument divided by the corresponding change in the stimulus). Thus, this parameter, also called diagnostic sensitivity, is rather associated to the probability of obtaining a true positive. The sensitivity is related to the critical concentration value of the chemical markers that decide that a sample belongs to one of the groups. In other words, it can be associated to the ability of the method to classify samples containing similar concentrations of compounds into the correct group.

5) Selectivity/Specificity:

The selectivity is defined with the following equation:

$$\text{Selectivity} = \frac{TN}{TN + FP}$$

This parameter is also different from the analytical selectivity (the ability of a particular analytical method to distinguish between the target analyte and the potentially interfering substances). Thus, the so called diagnostic selectivity is defined as the ratio of true negatives (TN) and total negatives. It can be associated to a measurement of the degree to which other substances in the matrix affect the classification if the classification is explained by a chemical marker that positives samples have and negative samples do not have, which is not always the case. For that reason, in some classifications tasks the interpretation of sensitivity and selectivity is not so straightforward. As a general approach, selectivity and sensitivity can be regarded as a degree of the certainty in the identity of each group.

6) Efficiency:

Efficiency, together with Youden's Index and the likelihood ratio, also inform on the overall suitability and it provides a measurement of a combination of true responses:

$$\text{Efficiency} = \frac{TP + TN}{TP + TN + FP + FN}$$

In consequence, this parameter is directly related to selectivity and sensitivity.

7) Youden's index:

The mathematical expression of Youden's index expresses a combination of selectivity and sensitivity.

Youden's index = 100 × (Sensitivity + Selectivity - 1)

8) Likehood ratio:

This parameter is a combination of the information related to the misclassification (both false positives and false negatives).

$$\text{Likehood ratio: } \frac{1 - \text{False Negative Rate}}{\text{False Positive Rate}}$$

9) Other parameters:

The participants should be encouraged to provide a single response for each parameter instead of several values corresponding to different variables (e.g. associated to several types of statistical processing). In addition to these parameters, other performance parameters commonly evaluated in quantitative analysis are recommended since they are sometimes useful to interpret the number of false positives and negatives. One of them is the repeatability (precision), expressed as relative standard deviation or %RSD (of the whole spectra and from different spectral features selected for the classification model).

In the case of two methods with similar values in all these parameters, other parameters of the statistical analysis can be used for an inter-comparison. Thus, for example, Mahalanobis distance can serve as comparative criterion.

Another aspect of paramount importance is the detection of outliers before applying any statistical treatment since abnormal values strongly influence the final result. In this application the detection of outliers has to be achieved by means of both univariate and multivariate statistical analysis.

Robustness, which is defined for quantitative analysis, is also an important parameter to be tested in qualitative analysis by modifying some of the most relevant variables and checking the influence on the number of false positives and negatives. That is a critical part in non-targeted techniques since different commercial brands use different approaches to present the sample, to manipulate the spectra or to treat the signal and therefore sometimes it is difficult to reproduce the same results just using different instruments with the same basic variables.

4. Conclusions

In the course of the work carried out within WP4 in terms of methods for geographical identification we have made a reflection about all the difficulties that we have faced when carrying out a performance evaluation of non-targeted methods being applied in a classification task and when interpreting the information that is available at the moment concerning the validation of qualitative methods. Although there is information available about which parameters and procedures have to be followed, the reality is that the implementation of a validation procedure in a specific application leads to discussions and raise questions and doubts between partners about which steps have to be taken. These questions mean that there is a gap between the theoretical approach of the validation of qualitative methods and the practical implementation in a specific

application, mainly due to the imprecision in the definitions of concepts and in the order of the steps to carry out. The experiences gained in previous European projects (e.g. MEDEO project, G6RD-2000-00440) in which qualitative methods have been validated with similar procedures were very helpful and contributed to solve these doubts. The thoughts and ideas collected in this process have served to describe a procedure of 7 points in which we have described the most relevant aspects and critical issues to discuss between the participants (see section 2). This procedure can be useful in the field of regulatory bodies because the methods with qualitative response are starting to be proposed for being accepted as standard methods and the regulatory bodies may have questions about how to guarantee the fit for purpose and the robustness of the methods.

In olive oil chemistry, some previous experiences have shown that methods with qualitative response (genuine/adulterated) should be removed from the standards if they were not properly validated according to a better defined objective and a validation procedure intended for qualitative applications. A guideline and a validation scheme to be understood by everyone are important not only for the working groups in regulatory bodies, but also by the end users that have to rely on the performance of the methods and have to understand how these methods have been previously tested. Additionally, a validation procedure for qualitative methods has to be understood by those administrations in charge of the global trade and imposing fines in the case of samples not meeting with the requirements defined in the standards. In this case, a method that is proven to be robust and it is recognized by everyone by means of a well-accepted and understood validation procedure is necessary to avoid possible conflicts in interpreting the results. A guideline is also important to understand that a certain flexibility is needed with the aim that a validation procedure can be adapted to each non-targeted technique or application. Thus, a guideline would help analysts to choose the best validation procedure that guarantees the robustness of the method for a well-defined objective.

The question that a validation procedure should answer is how the performance of a method is when dealing with a well-defined binary response. Any deviation from this question would need further validation testing. Sometimes the evaluation of the performance of a method needs to follow a holistic approach in which not only the quality parameters are examined independently, but also a balance of all types of errors has to be considered as well as other characteristics inherent to the method (e.g. cost of the techniques).

If possible, the quality parameters have to be compared with existing methods in the current standard. In the case of the methods for the geographical provenance, which has been the application studied after the amendment of WP4 at the beginning of the project, there is not any

standard method to compare with. However, there is extensive information about the precision values (relative standard deviations for repeatability and reproducibility) of other methods analysing minor components that can be used for comparison purposes. These values are described in Deliverable 4.3. The precision values of these methods (based on targeted techniques) are sometimes higher than 10%, which may induce us to think that those methods should not be suggested as standards. Table 4.1, however, shows the precision values for some non-targeted methods described in Deliverable 4.3. The precision values are easily better in non-targeted techniques compared with chromatographic methods. For example, the determination of fatty acid ethyl esters, one of the latest methods that have been proposed by IOC for olive oil assessment and it is under provisional approval. As an example, it can be said that in NIR spectroscopy, as shown by DLO, the average repeatability (%RSD) is 0.14% (Tabla 4.1). Other techniques tested by partners (NMR, Raman, IRMS, FTIR) show %RSD lower than 5%. Thus, in terms of precision, non-targeted techniques would be appropriate to be suggested as standard method.

Table 4.1. IOC values of analytical quality parameters (precision for repeatability) for the determination of ethyl esters concentration (mg/kg).

Partners	Method/Technique	RSD_r(%) (Repeatability)
	Fatty acid ethyl esters concentration	2.41-28.83[1]
FERA	^{13}C-NMR	3.60
DLO	NIR	0.14
CRA-W	Raman	2.66
FEM	IRMS	1.00
CSIC	CaF_2-FTIR	1.50
VSCHT	U-HPLC–HRMS/MS	11.00

NOTE: RSDr(%), relative standard deviation in repeatability, [1], A range is shown since this value can vary between different categories of olive oils (source: COI/T.20/Doc. No 31, November 2012).

In general terms, the work carried out in a collaborative study for evaluating the performance of qualitative methods should be established in the context of comparison with standard methods, if they exist, and the possible improvements to obtain better quality parameters. Thus, Figure 4.1 shows the previous and further steps of a collaborative trial.

The performance of the non-targeted methods studied in this work package are susceptible to be improved since these methods are complex in their technical characteristics and in the data treatment, and small changes in the protocol can have some unexpected changes in the values of the analytical quality parameters. Therefore, a small change in the variables of the method should lead again to a new validation study. If this validation is repeated, once again, a clear validation scheme is needed to get results that can be compared with previous ones.

The information included in this guideline could be distributed as a first approach to regulatory bodies (mainly the Sub-Group of Olive Oil of European Union and International Olive Council). A publication in a peer review journal is being prepared as well. A description of the steps being taken in the validation of a qualitative method can improve the acceptance of these methods by analysts and experts in regulatory bodies and can promote the use of non-targeted techniques as routine and reliable analytical tools.

It is clear that a harmonised protocol applicable for the validation of all methods with qualitative response is required. Further to the commissioning of this olive oil specific application, FoodIntegrity has commissioned a new WP (18) dedicated to achieving the goal.

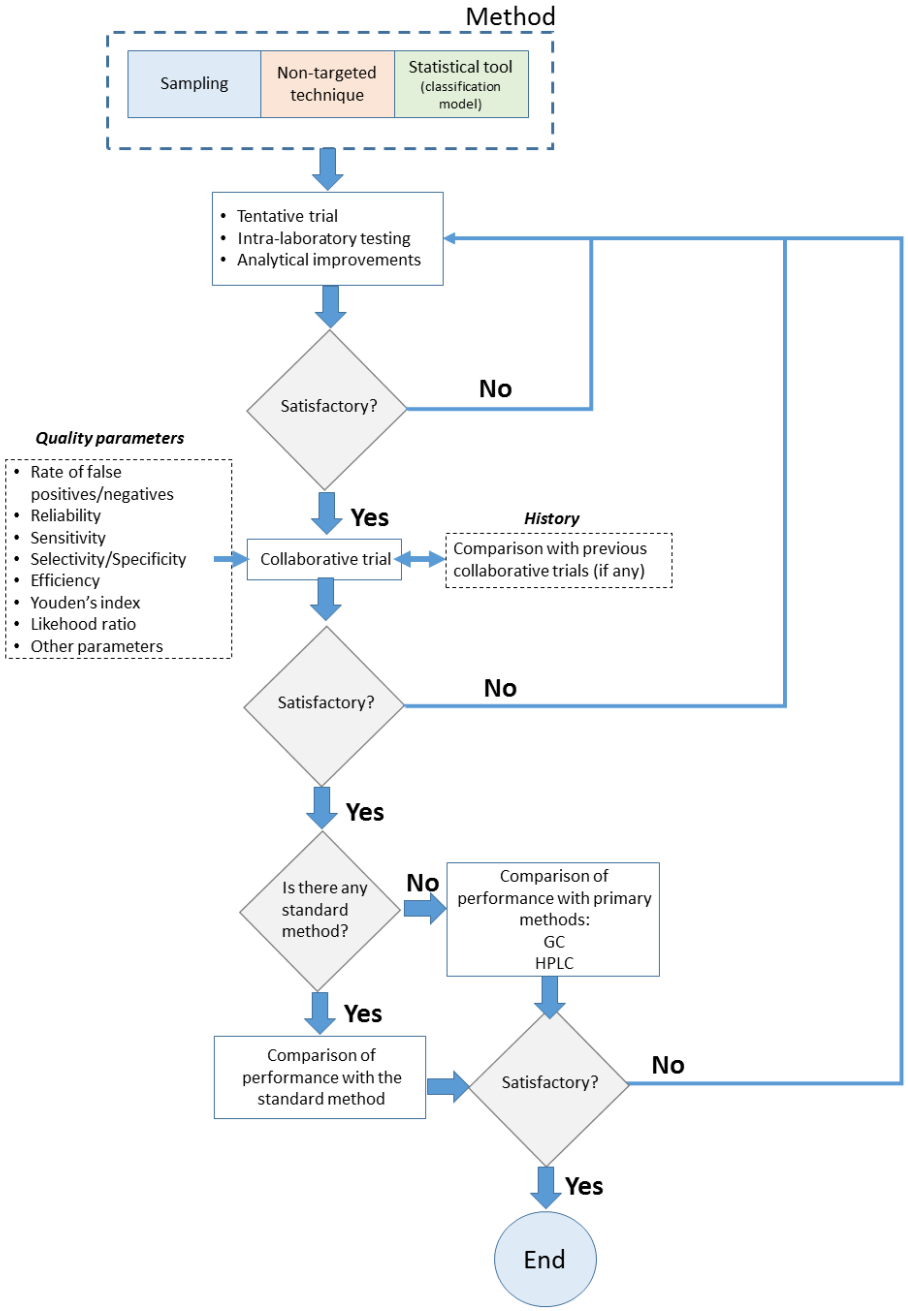

Figure 4.1. Scheme of previous and further steps of a collaborative trial.

5. List of WP4 Participants

1 Fera Science Ltd.
2 EUROFINS
3 JRC IRMM
6 QUB
7 SITEIA.UNIPR
10 CRA-W
13 DLO
14 VSCHT Praha
15 FEM
25 CSIC
38 UNIROMA1

Associated partner: CoopItalia.

6. References

(1) Valcarcel, M., Cardenas, S., Barceló, D., Buydens, L., Heydorn, K., Karlberg, B., Klemm, K., Lendl, B., Milman, B., Neidhart, B., Rios, A., Stephany, R., Townshend, A., & Zchunke, A. (2003) Accred. Qual. Assur. 8, 68–77.

(2) Ellison, S.L.R., Fearn, T. (2005). Characterising the performance of qualitative analytical methods: Statistics and terminology, Trends Anal. Chem., 24, 468-476.

(3) van de Voort, F.R., Ghetler, A., García-González, D.L., Li, Y.D. (2008). Perspectives on quantitative Mid-FTIR spectroscopy in relation to edible oil and lubricant analysis: Evolution and integration of analytical methodologies. Food Anal. Methods.1, 153-163.

(4) López M.I., Callao M.P., Ruisánchez I. (2015). A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. Anal. Chim. Acta, 891, 62-72.

(5) García-González, D.L., Viera-Macías, M., Aparicio-Ruiz, R., Morales, M.T., Aparicio, R. (2007). Validation of a method based on triglycerides for the detection of low percentages of hazelnut oil in olive oil by column liquid chromatography. J. AOAC Int., 90, 1346-1353.

(6) Aparicio, R. (2004). Final report project G6RD-CT2000-00440 (MEDEO). Commission of the European Communities (Growth Program), Brussels (Belgium).

(7) Valcárcel, M., Cárdenas, S., Barceló, D., Buydens, L., Heydorn, K., Karlberg, B, Klemm, K., Lendl, B., Milman, B., Neidhart, B., Rios, A., Stephany, R., Townshend A., Zschunke A. (2002). Metrology of Qualitative Chemical, Final Report Contract G6MA-CT2000-01012.

(8) García-González, D.L., Tena, N., Aparicio, R. (2012). Describing the chemical singularity of the Spanish protected designations of origin for virgin olive oils in relation to oils from neighbouring areas. Grasas Aceites, 63, 26-34.

(9) Riedl, J., Esslinger, S., Fauhl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. Anal. Chim. Acta, 885, 17-32.

(10) Tena, N., Wang, S.C., Aparicio-Ruiz, R., García-González, D.L., Aparicio, R. (2015). In-depth assessment of analytical methods for olive oil purity, safety, and quality characterization. J. Agric. Food Chem., 63, 4509-4526.

(11) Romero, I., García-González, D.L., Aparicio-Ruiz, R., Morales, M.T. (2015). Validation of SPME-GCMS method for the analysis of virgin olive oil volatiles responsible for sensory defects. Talanta, 134, 394-401.

(12) Aparicio, R., Conte, L.S., Fiebig, H.-J. Olive oil authentication. In: Handbook of Olive Oil. Analysis and Properties, Aparicio, R., Harwood, J. (Editors). 2° Edition, Springer, New York, 2013, pp. 739-768.

(13) Alewijn, M, van der Voet, H, van Ruth, S. (2016). Validation of multivariate classification methods using analytical fingerprints – concept and case study on organic feed for laying hens. J. Food Compos. Anal. 51, 15-23.