INTERNATIONAL OLIVE COUNCIL

COI/T.20/Doc. No 17/Rev 1
June 2023

ENGLISH
Original: ENGLISH

Príncipe de Vergara, 154 – 28002 Madrid – España Telef.: +34 915 903 638 Fax: +34 915 631 263 e-mail: iooc@internationaloliveoil.org - http://www.internationaloliveoil.org/

INTERNAL QUALITY CONTROL GUIDELINES FOR SENSORY LABORATORIES

Introduction

This document is a guideline for the complete quality control for sensory laboratories undertaking the analysis of virgin olive oils. It includes a broad range of procedures. As some are time-consuming, it is not compulsory to apply all procedures; the panel leader can select the most appropriate procedures that will ensure the competence of tasters and the panel, and will prove that results are reliable.

Two worksheets for calculating the indexes included in these guidelines and for their illustration in charts are attached to these guidelines. The instructions for the use of the two worksheets are included in APPENDIX I. The use of these worksheets serves the function of sensory laboratories but is not compulsory.

Contents

1. METHODS OF INTERNAL QUALITY CONTROL IN SENSORY ANALYSIS

Since the measuring instrument in sensory analysis is the group of tasters and the results depend on its members, the performance of each individual taster and the whole panel should be monitored on a regular basis. Internal quality control in a sensory laboratory must therefore ensure that the panel and each sensory assessor are checked. The effectiveness of monitoring the performance of the panel and each taster depends on the method used for internal quality control and the appropriate processing of the results.

Some of the quality control procedures are:

(a) replicate the analysis in a specific percentage of the total number of samples at adequate intervals.

(b) analyse reference materials and characterised materials as part of the quality control system.

As a guideline, the frequency of quality check may be at least 9% of all samples analysed.

1.1. Replicate analysis

A sample to be replicated will be taken from those to be analysed or from those already analysed in previous days.

According to COI/T.20/Doc. No 15, a maximum of 12 samples shall be tasted per day. The minimum frequency for internal quality control should therefore be every 11 samples (9% of the samples analysed). However, the recommended frequency is to check the samples every tasting day.

If the replicate is performed every tasting day, the monitoring frequency is shown in the table below. It varies depending on the number of samples analysed per day (≥9% of all samples analysed).

Number of samples per day	Frequency of check (*)
4 (= 3+1)	1/3 = 33 %
5 (= 4+1)	1/ 4 = 25 %
6 (= 5+1)	1/6 = 20%
7 (= 6+1)	1/5 = 17%
8 (= 7+1)	1/7 = 14 %
9 (= 8+1)	1/8 = 13 %
10 (= 9+1)	1/9 = 11%
11 (= 10+1)	1/10 = 10%
12 (= 11+1)	1/11 = 9%

(*) % of duplicate samples, in relation to the total number of samples.

If the sensory panel has not been active for some time (for example, holidays or long breaks), the internal quality checks must be performed immediately before analysing samples. The replicate samples must cover the widest possible range of intensities of fruitiness and defects. Their position in the sessions should be random.

Although replicating samples has the advantage that it does not require special samples, its main disadvantages are that it only gives information on random errors (it evaluates the precision of both panel and tasters) and it does not check the correct classification of a sample.

1.2. Analysis of reference materials and characterised materials

At least one reference material will be analysed each month (except for the months when no sample is analysed). These materials will be CRM, if they exist. If they are not available, the remaining samples from proficiency tests should be used. In the absence of the mentioned samples, the laboratory will prepare a sufficient number of samples for quality checks, which will be characterised by comparison with at least three accredited panels. The criteria for assigning reference values of defect and/or fruity attribute should be defined and documented by the panel leader and reported in the standard operating procedure (SOP).

The range of the samples used as reference materials will be varied in order to cover different classes of virgin olive oil, intensities and attributes, over the course of a year. The laboratory should define the shelf-life of the reference material.

The main advantage of this method is that the results obtained by analysing reference materials or characterised materials could be used to monitor the trueness of the panel and each individual taster. On the other hand, certified or secondary reference materials are difficult to use in the sensory tests, due to the large quantity required for carrying out organoleptic assessment and the changes that occur in the organoleptic characteristics of a sample during storage.

2. CHECKING THE PERFORMANCE OF EACH TASTER

The panel leader shall set the minimum levels of precision and trueness in relation to the tasters, in order to keep their qualification. Additional requirements may be defined, such as the minimum attendance to panel sessions.

A taster's performance must be checked over time using different types of samples and product categories, as well as the psychophysiological stages they may undergo.

The technique for checking taster performance is based on the use of a set of samples analysed double-blind. From the results of these analyses, the precision number (PN) and deviation number (DN) are calculated. These indicators are needed because the taster's performance consists of two different factors, namely:

- Deviation from itself when analysing the same sample at two different times;
- Deviation from the group (the panel) during the same session.

To measure these differences, the PN and DN can be used. They must be analysed together, and they are defined as follows:

$$Precision\ Number\ (PN) = \frac{\sum_{i=1}^{n}(x_{i,1} - x_{i,2})^2}{n} \leq 2.0$$

where $x_{i,1}$ and $x_{i,2}$ are the values given by the taster to the first and second assessment of a duplicated sample, and *n* is the number of differences ($x_{i,1}$-$x_{i,2}$), corresponding to the number of duplicated samples analysed.

$$Deviation\ Number\ (DN) = \frac{\sum_{i=1}^{n}(x_{i,1} - \bar{x}_{i,1})^2}{n} \leq 2.0$$

where $x_{i,1}$ is the value of the first replicate given by the taster, $\bar{x}_{i,1}$ is the value of the median of the replicate considered, and n is the number of differences ($x_{i,1}$ - $\bar{x}_{i,1}$), i.e. the number of duplicated samples analysed.

For the DN, only the value from one of the two replicates shall be used, to avoid the bias that could be added in the calculation. Therefore, the panel leader must indicate in the SOP which of the two replicates will be used.

The number of samples analysed in duplicate should be between 6 and 10, depending on the frequency of the panel's analysis.

To find the direction of the differences of the DN, control charts shall be used.

The DN can also be used to check other aspects of performance (for example, to check the deviation of the taster, or the panel, to the reference value of characterised sample/reference material).

As the limit value for these numbers is 2.0, the maximum allowed deviation for the taster is 1.4 (=2*0.7), on average. For example:

$$\text{PR or DN} = \frac{1.4^2 + 1.4^2 + 1.4^2 + 1.4^2 + 1.4^2 + 1.4^2}{6} = 1.96$$

2.1. Checking the taster's precision

Precision is the closeness of agreement between independent test values. A precision assessment involves estimating repeatability and in-laboratory reproducibility or intermediate precision. In the sensory method, the precision (repeatability and in-laboratory intermediate precision) of tasters is determined by using the replicate analysis. The repeatability of each taster is checked by comparing the score of the intensities given by the taster when analysing a sample in duplicate.

Intermediate precision may be checked over time by means of the PN, which takes into account the intensities given by the taster to a set of duplicate samples, between 6 and 10 (12-20 analysed samples in total) as described in section 1.1. Alternatively, the intermediate precision of each taster can be measured over time using the same index but analysing the same sample on different days. To do so, samples are prepared for tasting as double-blind samples by the tasters within a maximum of six months, depending on the attributes. If possible, these samples should be representative of the categories tested most often by the laboratory. The samples must be properly stored to guarantee that their characteristics remain unchanged.

Table 1. Indicator of taster precision, when several numbers of replicated samples have been analysed.

<table>
<tr><th colspan="2">Field of application: taster</th></tr>
<tr><td colspan="2">Frequency: when the number of duplicate samples is between 6 and 10.</td></tr>
<tr><td>Estimation of Precision</td><td>Precision Number (PN_t) of the taster
$$PN_t = \frac{\sum(x_{a1} - x_{a2})^2}{n}$$
where:
❖ PN_t is the precision number of one taster, for a specific attribute (a defect, the fruity attribute or the classified attribute).
❖ x_{a1}, x_{a2} are the intensities given by the taster to a specific attribute in the first and second assessment of the replicated sample (for a defect, the fruity attribute or the classified attribute)
❖ n is the number of duplicate samples tested (example: one duplicate sample, n=1 / six samples tested in duplicate, n=6).
Criteria of acceptance: $PN_t \leq 2.0$

If the PN_t value is above 2.0, training should be arranged for the taster.</td></tr>
</table>

Notes:

1. The sensory laboratory can use either:
    - one Precision Number for each taster <u>for the classified attribute identified by the panel</u> (fruity for EVOO and predominant defect for other categories), or
    - <u>one for the defects and another for the fruity attribute</u>, separately. In any case, the laboratory should keep fully documented records.
2. Since the PNs are used to check how the taster repeats their own assessments, it is possible to calculate these numbers with the attribute score with greater intensity, according to the taster (classified or higher defect/fruity found by the taster), and not to the panel. The selected option must be previously defined in the operative protocol.
3. When the taster evaluation is performed with duplicate samples, the PN must be calculated with the DN, at the same time and with the same samples, since both give an indication of the analytical behaviour of the taster. They are strictly linked and should not be studied separately or at different times. These two indicators must agree: if one falls outside acceptable limits, this indicates poor performance.
4. <u>Warning limit</u> = optionally, a warning limit may be defined, so, when indices are between 1 and 2, the panel leader should study the possible causes and, if necessary, perform preventive actions to improve the taster's performance. It will not be necessary to exclude the taster from the panel since the indices are lower than 2.
5. The tables below include the necessary calculations to estimate the cumulative PN, in order to facilitate the work of sensory laboratories (optionally, these results can be expressed to two decimal places).

Table 2.a. Example of calculations for PN of the taster for predominant defect and fruity.

Intensity given by the taster				(Difference)^2	
Predominant Defect		Fruity		Predominant Defect	Fruity
1st test	2nd test	1st test	2nd test		
X_{D11}	X_{D12}	X_{F11}	X_{F12}	$(X_{D11} - X_{D12})^2$	$(X_{F11} - X_{F12})^2$
X_{D21}	X_{D22}	X_{F21}	X_{F22}	$(X_{D21} - X_{D22})^2$	$(X_{F21} - X_{F22})^2$
X_{D31}	X_{D32}	X_{F31}	X_{F32}	$(X_{D31} - X_{D32})^2$	$(X_{F31} - X_{F32})^2$
X_{D41}	X_{D42}	X_{F41}	X_{F42}	$(X_{D41} - X_{D42})^2$	$(X_{F41} - X_{F42})^2$
……	……	……	……	…….	…….
X_{Dn1}	X_{Dn2}	X_{Fn1}	X_{Fn2}	$(X_{Dn1} - X_{Dn2})^2$	$(X_{Fn1} - X_{Fn2})^2$
				SUM D	SUM F
				PN_{dt} = SUM D / n	PN_{ft} = SUM F / n

Table 2.b. Example of calculation of PN with six duplicate samples (n=6), for a given attribute, in "batch mode".

Intensity given by the taster		(Difference)^2	Calculations
1st test	2nd test		
X_{11}	X_{12}	$(X_{11}- X_{12})^2$	PN_t = SUM(1-6) / 6
X_{21}	X_{22}	$(X_{21}- X_{22})^2$	
X_{31}	X_{32}	$(X_{31}- X_{32})^2$	
X_{41}	X_{42}	$(X_{41}- X_{42})^2$	
X_{51}	X_{52}	$(X_{51}- X_{52})^2$	
X_{61}	X_{62}	$(X_{61}- X_{62})^2$	
X_{71}	X_{72}	$(X_{71}- X_{72})^2$	PN_t = SUM(7-12) / 6
X_{81}	X_{82}	$(X_{81}- X_{82})^2$	
X_{91}	X_{92}	$(X_{91}- X_{92})^2$	
X_{101}	X_{102}	$(X_{101}- X_{102})^2$	
X_{111}	X_{112}	$(X_{111}- X_{112})^2$	
X_{121}	X_{122}	$(X_{121}- X_{122})^2$	
X_{131}	X_{132}	$(X_{131}- X_{132})^2$	PN_t = SUM(13-18) / 6
X_{141}	X_{142}	$(X_{141}- X_{142})^2$	
X_{151}	X_{152}	$(X_{151}- X_{152})^2$	
X_{161}	X_{162}	$(X_{161}- X_{162})^2$	
X_{171}	X_{172}	$(X_{171}- X_{172})^2$	
X_{181}	X_{182}	$(X_{181}- X_{182})^2$	

Table 2.c. Example of calculation of PN with six duplicate samples (n=6), for a given attribute, in "continuous mode".

Intensity given by the taster		(Difference)^2	Calculations	(Difference)^2	Calculations	(Difference)^2	Calculations
1st test	2nd test						
X_{11}	X_{12}	$(X_{11}- X_{12})^2$	PN_t =SUM(1-6)/6				
X_{21}	X_{22}	$(X_{21}- X_{22})^2$		$(X_{21}- X_{22})^2$	PN_t=SUM(2-7)/6		
X_{31}	X_{32}	$(X_{31}- X_{32})^2$		$(X_{31}- X_{32})^2$		$(X_{31}- X_{32})^2$	PN_t=SUM(3- 8)/6
X_{41}	X_{42}	$(X_{41}- X_{42})^2$		$(X_{41}- X_{42})^2$		$(X_{41}- X_{42})^2$	
X_{51}	X_{52}	$(X_{51}- X_{52})^2$		$(X_{51}- X_{52})^2$		$(X_{51}- X_{52})^2$	
X_{61}	X_{62}	$(X_{61}- X_{62})^2$		$(X_{61}- X_{62})^2$		$(X_{61}- X_{62})^2$	
X_{71}	X_{72}			$(X_{71}- X_{72})^2$		$(X_{71}- X_{72})^2$	
X_{8-1}	X_{8-2}					$(X_{81}- X_{82})^2$	

Table 2.d. Example of calculation of PN with 6 duplicate samples (n=6), in "batch mode".

Sample			**Median of the panel for one attribute**	
	1st test	**2nd test**	**1st test**	**2nd test**
M1	2.6	2.9	3.1	3.4
M2	4.3	3.9	4.5	4.1
M3	1.8	2.2	2.5	2.7
M4	6.2	5.7	6.0	6.3
M5	3.5	3.1	3.8	3.4
M6	0.9	1.6	1.4	1.7

$$PN_t = \frac{(2.6-2.9)^2 + (4.3-3.9)^2 + (1.8-2.2)^2 + (6.2-5.7)^2 + (3.5-3.1)^2 + (0.9-1.6)^2}{6} = 0.22$$

2.2. Checking the taster's trueness

In addition to evaluating the precision of each taster, their trueness must also be evaluated. Trueness is the closeness of agreement between the average value of a large series of measurements and the accepted reference value or the "true value." Systematic error (bias) is a measure of accuracy.

The **trueness of tasters** is determined by using the analysis of **reference materials** or **characterised materials** (if the sample to be used is clearly defined). Since the reference materials are not included in each session of the panel, this estimation does not assure a **continuous control of the performance of the taster**, so this calculation is just a complement of the previous one (2.1. Checking the taster's precision). In the same manner, the performance of the tasters with respect to the panel over time could be included as well, by using the replicate analysis.

Table 3.a. Estimator of single taster "trueness" (deviation from the panel median) using DN.

Field of application: taster
Frequency: 9% of analysed samples in case of replicate analysis. The frequency should be once per month when analysing reference material, depending on its availability.

Deviation Number (DN_t)

Calculation with replicate samples

$$DN_t = \frac{\sum(x_i - Me_i)^2}{n}$$

Calculation with reference materials

$$DN_t = \frac{\sum(x_i - TMe_i)^2}{n}$$

where:

- DNt is the deviation number of a taster "t", for a specific attribute (predominant defect, fruity attribute or classified attribute), in the first (1) or second (2) replicate of the duplicate sample "i".
- xi is the intensity score given by the taster for the predominant defect, fruity attribute or classified attribute in the first (1) or second assessment (2) of the duplicate sample "i",
- Mei is the value of the panel median for the predominant defect, fruity attribute or classified attribute in the first or second assessment of the duplicate sample.
- TMei is the reference value (assigned value) of the reference material.
- The selected duplicate for the calculation (first or second) must be previously defined in the operative protocol.
- *n* is the number of differences considered for the calculation (example: for six duplicate samples or six reference materials n=6)

Criteria of acceptance: $DN_t \leq 2.0$

If the DN_t value is above 2.0, training should be arranged for the taster.

Table 3.b. Estimator of single taster "trueness" in terms of deviation from all panels participating in the reference material certification.

Field of application: taster
Frequency: once per month depending on the availability of reference materials.

Taster's z-score

$$\text{z-score}_t = \frac{(x - TMe)}{SD}$$

where:

- x is the intensity given by a taster "t", for a specific attribute (predominant defect, fruity attribute or classified attribute), in the assessment of sample,
- TMe is the value of reference sample (assigned value) for the attribute (either the predominant defect, the fruity attribute or the classified attribute)
- SD is the standard deviation of all values of the laboratories participating in the certification process of the material, for the predominant defect, the fruity attribute or the classified attribute, or the standard deviation of the method (±0.7).

Criteria of acceptance:
Warning limits: $\text{z-score}_t = \pm 2.0$, and action limits: $\text{z-score}_t = \pm 3.0$.

If the z-score_t is out of action limits, training should be arranged for the taster.

Notes:

6. As for the estimators for precision, the sensory laboratory can calculate either:
    - one index for classified attribute determined by the panel (fruity for EVOO and defect of a higher intensity – predominant defect – for other categories), or
    - one for the defects and another for the fruity attribute separately.

In any case, the laboratory should keep fully documented records.

7. Unlike the PN, the attribute chosen to calculate the DN, with replicate samples, should be the attribute the panel used to classify the sample, and not the taster (classified or higher defect/fruity found by the panel), because this index measures how the taster deviates from the panel.

8. When the taster is evaluated using duplicate samples, the DN shall be calculated together with the PN and they must agree at the same time (see note 3).

9. The DN can also be calculated with any sample of the tasting day, not duplicated. Here, two conditions shall be considered: (i) the level of control must not be lower than 9% of analysed samples; and (ii) the selected sample for DN calculation must be clearly pre-defined in the operative protocol. For example, if a reference material is included in the tasting session, the DN can be calculated with the score of the taster and the median of the panel, given for that reference material. The analysis of the duplicate sample can be therefore avoided that day.

10. The calculation can be performed in batch mode or in continuous mode, and two decimal digits can be used.

11. A useful system to check taster's performance is to include, from time to time, one or more reference samples (clearly defined, pre-tested oils), as explained in ISO 13299:2016. Studying the individual variance of the scores obtained by each taster for these check samples shows whether the tasters' performance is consistent over time, by checking the correspondent F value. Likewise, the use of the variance of the mean values obtained by the panel is a useful indicator to understand whether the panel has consistent results over time.

12. Analysis of variance (ANOVA) can easily be carried out with software available commercially or as freeware.

Table 4. Example of calculation of the DN with six duplicate samples, in batch mode, selecting the second assessment (*), n=6.

Sample	Score for one attribute, given by the taster		Median of the panel for one attribute	
	1st test	2nd test	1st test	2nd test
M1	2.6	2.9	3.1	3.4
M2	4.3	3.9	4.5	4.1
M3	1.8	2.2	2.5	2.7
M4	6.2	5.7	6.0	6.3
M5	3.5	3.1	3.8	3.4
M6	0.9	1.6	1.4	1.7

(*) The same calculation may be performed with the *first* assessment of the duplicate sample.

$$DN_t = \frac{(2.9-3.4)^2+(3.9-4.1)^2+(2.2-2.7)^2+(5.7-6.3)^2+(3.1-3.4)^2+(1.6-1.7)^2}{6} = 0.17$$

2.3. Checking the taster's competence (sample classification and intensity evaluation)

The above estimation of trueness only considers the values given by the tasters when assessing a reference material. However, the organoleptic method is both qualitative and quantitative, since it classifies samples based on the median of the predominant defect and the presence or not of the fruity attribute. Tasters can therefore be checked using a similar procedure applied by the IOC to evaluate the results of panel proficiency tests.

This check evaluates the performance of the tasters on one day only, without considering possible changes over time. The recommended procedure is presented below.

The competence of a taster could be checked by using the results of each taster in the most recent interlaboratory proficiency testing. If this is not possible (for example, when evaluating a new taster), the taster's competence can be checked through organoleptic analysing of samples with known and reliable data (CRM, samples from proficiency tests or characterised samples).

The samples selected for the competence check should belong to different categories and have defined reliable data (category statistically significant at 95% confidence level, median of predominant defect and/or fruity, standard deviation – not robust standard deviation – or upper and lower confidence limit for the predominant defect and fruity). The taster's score will be set to 1 if they have correctly classified the sample and the intensity of the predominant defect for the virgin and lampante categories and of fruitiness for the extra virgin category. The taster will give a score for each sample, which should fall between the upper and lower confidence limit or according to the acceptance criteria of the taster's z-score. If any of the preceding cases do not apply, the taster's score will be set to 0.

The taster's score is evaluated by considering either:

(a) The z-score limit of 2*SD (where SD is the standard deviation and not the robust standard deviation of all values of the laboratories participating in the certification process of the material, or the standard deviation of the method (±0.7)); or
(b) The upper and lower confidence limits of the material. This criterion is stricter than using the z-score.

The panel leader can choose the most appropriate criterion for the laboratory. The median taster score is calculated for each taster. If it is 1, the taster is considered competent for carrying out organoleptic tests. If the score is 0, when this is not the case for other tasters, then training is required for this taster.

An example of the evaluation of taster's competence is reported as follows.

Frequency: the taster's competence should be evaluated whenever the taster participates in inter-laboratory testing and at least once a year.

Table 5. Example of calculation of taster's competence based on the intensities of the perceived attributes and the classification of a sample.

		sample 1	sample 2	sample 3
Taster's results				
	Classification	extra	virgin	lampante
	Taster's score	fruity 3.9	defect 2.0	defect 8.0
Reliability data of the samples				
	Classification	extra	virgin	lampante
	Median	fruity 4.3	defect 1.0	defect 6.1
	Upper limit	5.2	1.3	7.2
	Lower limit	3.4	0.6	4.9
	2*SD	2.4	1.0	2.5
Evaluation of the taster				
Option a (z-score)	**z-score**	-0.33	+2.00	+1.52
	taster's score	1	1	1
	Median of scores=1 ⇨ **THE TASTER IS COMPETENT**			
Option b	**taster's score**	1	0	0
	Median of scores=0 ⇨ **THE TASTER IS NOT COMPETENT**			

3. CHECKING THE PANEL'S PERFORMANCE

When checking the performance of each taster, the precision and trueness of the values obtained from the whole panel can be checked as well.

3.1. Checking the panel's precision

The panel's precision can be estimated during the procedure of replicate analysis for the assessment of a single tasters' precision.

The performance of the panel may be checked every tasting day by means of replicate analysis, calculating the normalised error "En" as reported in COI/T.20/Doc. No 15 (section 10.5) and Table 7.a below. "En" determines whether the two results of a duplicate analysis are homogeneous or statistically acceptable. The tasting day is considered "valid" if the "En" value of the replicate sample is correct. This is known as "validation of the tasting day".

The repeatability of the panel is evaluated by comparing the pair of medians obtained by analysing a sample in duplicate.

The precision of the panel is checked by comparing the pairs of medians obtained by analysing several duplicate samples between 6 and 10 (in total 12-20 samples analysed).

The panel leader should keep a record of the historical performance of the panel in an appropriate database as well as in tabulated form.

The formulae used for the assessment of repeatability and intermediate precision of the panel are reported as follows.

Table 6.a. Estimation of panel precision by Normalised Error.

Field of application: panel	
Frequency: every 11 tests (9% of all the samples analysed) or every tasting day (≥9% of all the samples analysed).	
Estimation of Repeatability	Normalised Error (En) $En = \frac{\lvert Me_1 - Me_2 \rvert}{\sqrt{U_1{}^2 + U_2{}^2}}$ where: ❖ En is the normalised error of the panel for a specific attribute (predominant defect, fruity attribute or classified attribute). ❖ Me_1 and Me_2 are the medians obtained by the panel for a specific attribute (predominant defect, fruity attribute or classified attribute) in the first and second assessment of a sample, respectively. ❖ U_1 and U_2, are the respective expanded uncertainties calculated as $c*s_1$ and $c*s_2$, with c=1.96 for a 95% probability, being s_1 and s_2 the experimental robust standard deviation values of the medians Me_1 and Me2, respectively, for the predominant defect, fruity attribute or classified attribute. Occasionally, it could be considered as the maximum error allowed by the method or the standard deviation of the method (±0.7). Criterion of acceptance: En ≤ 1.0

Table 6.b. Estimation of panel precision when several duplicated samples are available.

Field of application: panel	
Frequency: when several replicate samples between 6 and 10 have been analysed.	
Estimation of Panel Intermediate Precision	Precision Number for the panel (PN_p) $PN_P = \frac{\sum(Me_1 - Me_2)^2}{n}$ where ❖ PN_p is the indicator of the consistency of a panel "p", for a specific attribute (predominant defect, fruity attribute or classified attribute), in evaluating a sample in duplicate. ❖ Me_1, Me_2 are the medians of the panel for a specific attribute (predominant defect, fruity attribute, or classified attribute) in the first and second assessment of the duplicated sample, respectively. ❖ n is the number of duplicated samples taken into account (example: one duplicate sample, n=1 / six duplicate samples, n=6). Criteria of acceptance: $PN_p \leq 2.0$ **If PN_p is above 2.0, training should be arranged for the whole panel.**

Note: the notes describing the control of each taster's precision should also be applied for the control of the panel's precision. In the current section, the word "taster" is substituted by the word "panel".

3.2. Checking the panel's "trueness"

Checking the panel's "trueness" is the object of the laboratory's external quality control. Nevertheless, the trueness of the panel can also be estimated when analysing reference materials or characterised materials for assessing taster trueness. The formulae used to estimate the panel's "trueness" are as follows.

Table 7.a. Estimators of panel's trueness by using DN on data obtained from reference material (or characterised samples).

Field of application: panel
Frequency: once per month depending on the availability of reference materials

Deviation Number of a panel (DN_p)

$$DN_p = \frac{\sum(Me_i - TMe_i)^2}{n}$$

where:

- ❖ DN_p is the deviation number of a panel "p", for a specific attribute (predominant defect, fruity attribute or classified attribute).
- ❖ Me_i is the median value of the panel for the attribute (predominant defect, fruity attribute or classified attribute) in the assessment of sample i.
- ❖ TMe_i is the value of the reference sample i (training sample), for the attribute (predominant defect, fruity attribute or classified attribute).
- ❖ n is the number of reference samples analysed (example: for six reference materials, n=6).

Criterion of acceptance: $DN_p \leq 2.0$

If DNp is above 2.0, training should be arranged for the whole panel.

Table 7.b. Estimators of panel's trueness by using z-score on data obtained from reference material.

Field of application: panel
Frequency: once per month depending on the availability of reference material.

z-score for a panel

$$\text{z-score}_p = \frac{(Me_p - TMe)}{SD}$$

where:

- ❖ Me_p is the median obtained by a panel "p", for a specific attribute (predominant defect, fruity attribute or classified attribute) in the assessment of the reference sample.
- ❖ TMe is the value of the reference sample (assigned value), for a specific attribute (predominant defect, fruity attribute or classified attribute).
- ❖ SD is the standard deviation of all values of the laboratories participating in the certification process of the material, for the predominant defect and for the fruity attribute, or in general, the standard deviation corresponding to the TMe. Occasionally, the maximum error allowed by the method or the standard deviation of the method (±0.7) could be considered.

Criterion of acceptance:

Warning limits: z-score$_p$ = ± 2.0, and action limits: z-score$_p$ = ± 3.0.

If z-score$_p$ is out of action limits, training should be arranged for the panel.

Note: the notes describing the control of each taster's "trueness" should also be applied for the control of the panel's "trueness". In the current section, the word "taster" is substituted by the word "panel".

4. QUALITY CONTROL CHARTS IN SENSORY ANALYSIS

The quality control charts used in analytical laboratories are a control mechanism to determine whether the analytical procedure to be followed is "in statistical control", i.e. if the results produced are continuously within control limits.

In sensory analysis, the changes in performance of each taster and the whole panel shall be checked over time. To do so, the values obtained during the procedures for the performance check of each taster and the panel should be placed in quality control charts, as part of the internal quality control. **The quality charts facilitate the monitoring of the performance of each taster and panel over time**.

The laboratory should define corrective actions to perform when a result falls outside the limits, or several consecutive results are obtained at the same side (positive or negative) of the central value, but within the limits, since this may indicate systematic error (bias).

4.1. Quality control charts for indices based on replicate analysis.

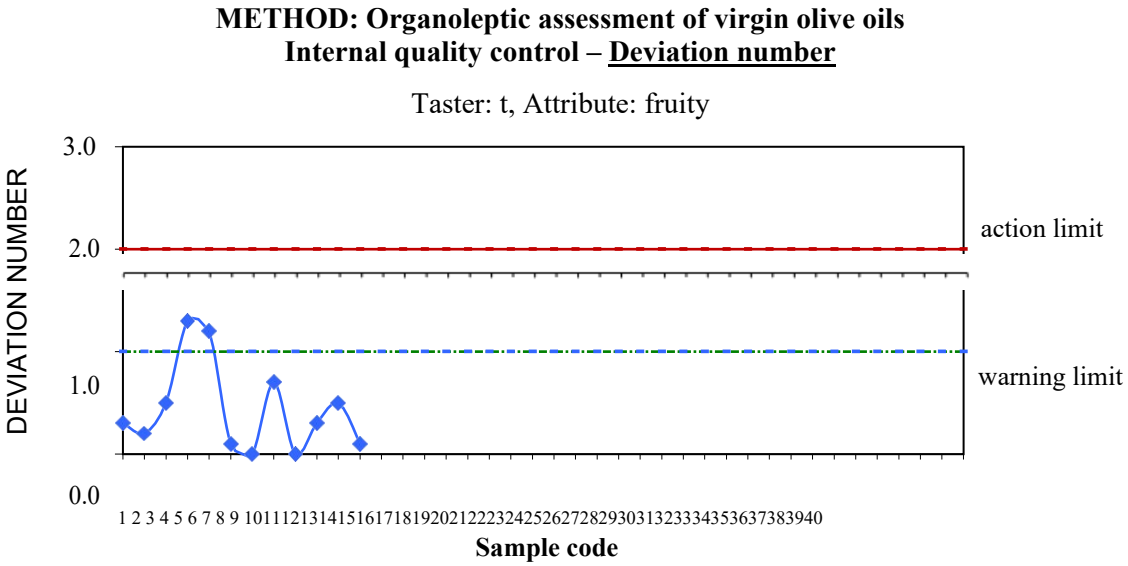
As described in paragraphs 2.1, 2.2 and 3.1, the indexes based on replicate analysis are the precision and deviation numbers of tasters and the normalised error and precision numbers of the panel.

Given that they are always positive numbers, their control chart could be a “trend chart”. In this group, the DNs of panels and tasters based on the analysis of reference materials should be included, since they are also always positive. **"Trend charts" can be used to illustrate experimental results when quality control is based on the assessment of conformity by performing duplicate measurements of a sample**. The minimum value of these indices is zero (0) and the maximum value is two (2) except for the normalized error where the maximum value is 1. In both cases, the “x” axis intersects the “y” axis at 0.

The index value is indicated on the vertical axis and the sample code is indicated on the horizontal axis (or date of the analysis), to ensure traceability.

Some models are illustrated below, including the criteria for chart interpretation. Each laboratory should define the criteria for implementing preventive and corrective actions.

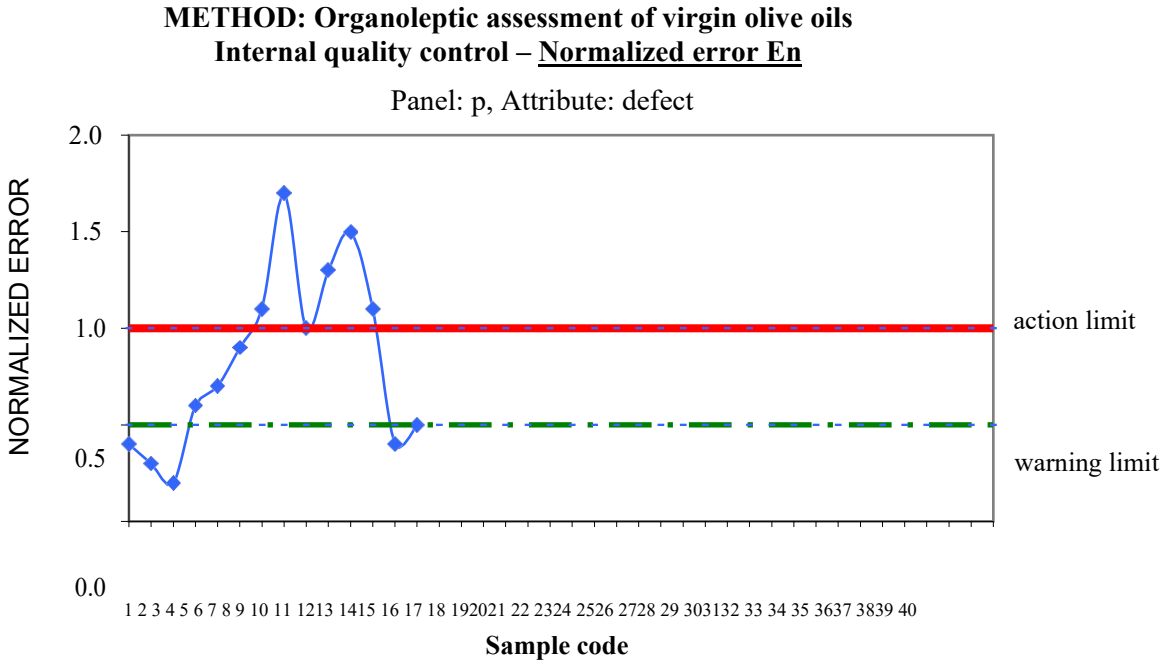
Figure 1. Quality control chart for single taster DN for the fruity attribute.



Criteria

1. One (at least) from 5 consecutive blue points must lie under the dotted line.
2. if a blue point is above the red line, the taster is out of control.
3. If 5 or more consecutive blue points lie between the red and dotted lines, there is a trend for the taster to be out of control.

Figure 2. Quality control chart for the Normalised Error of the panel, for defects.



Criteria

1. One (at least) from 5 consecutive blue points must lie under the dotted line.
2. If a blue point is above the red line, the analytical procedure is out of control.
3. If 5 or more consecutive blue points lie between the red and dotted lines, there is a trend for the analytical procedure to be out of control.

4.2. Quality control charts for indices based on analysis of reference material

As mentioned above in paragraphs 2.2. and 3.2., the main indices based on the analysis of reference materials are the z-score and the DN of the taster and the panel.

- Deviation Number

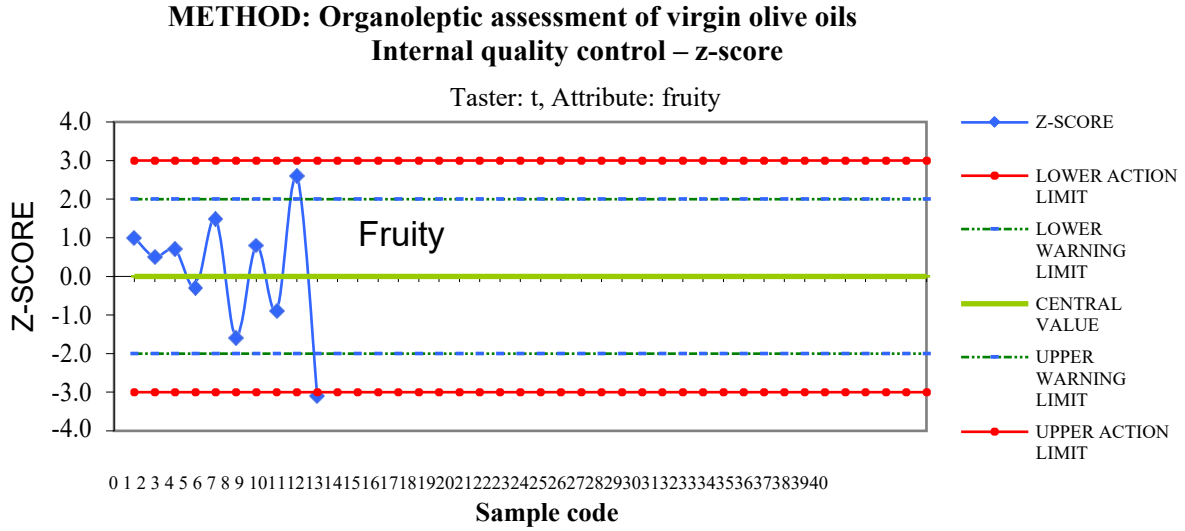
The graphs are designed as explained in section 4.1.

- z-score graphs

This index can have positive or negative values, the central value being zero, the warning limits for the index ±2, and the action limits ±3. The laboratory should define the corrective and preventive actions that will be performed whether a result is outside the limits, or several consecutive results are obtained at the same side (positive or negative) of the central value (bias). The same chart can be used by the sensory laboratory for the graphic representation of its z-score from its participation in the interlaboratory proficiency tests (external quality control). They are very useful to evaluate the "trueness" of the panel over time.

An example of the graph and some criteria for its interpretation are presented below.

Figure 3. Quality control chart for z-score of a single taster, for the fruity attribute.



Criteria

1. If a blue point (z-score) is under or above the red lines, the taster is out of control.
2. If 2 consecutive blue points lie between red and dotted lines, the taster is out of control.
3. If 10 consecutive blue points lie in the same side between the green and dotted lines, the taster is out of control.
4. If 7 consecutive blue points lie in the same side between the green and dotted lines, there is a trend for the taster to be out of control.
5. If one from 20 consecutive blue points lie between the dotted and red lines, the taster is within control.

4.3. Quality control charts of quality control samples

A control sample should be treated exactly as any other sample. Quality control charts are used to graphically represent the results of analysing control samples over time; they are known as $\bar{x}$ chart.

As mentioned in paragraph 1.2, it is difficult to use certified or secondary reference materials in sensory analysis. However, these samples can be prepared and refrigerated in 150 mL bottles for one year. The frequency these samples are used could be the same as reference materials (once per month) or every 20 unknown samples. The results of the analysis of these quality control samples should be recorded in a $\bar{x}$chart, in which the vertical axis represents the median of fruitiness or the defect, and the horizontal axis represents the date of the analysis or the sample code. These charts could be double, to illustrate both fruity and negative sensory attributes (fruity to the positive axis, defect to the negative axis).

Moreover, in organoleptic assessment the correct intensity score and the correct classification of a sample should be checked. The following restrictions by category should also be adopted:

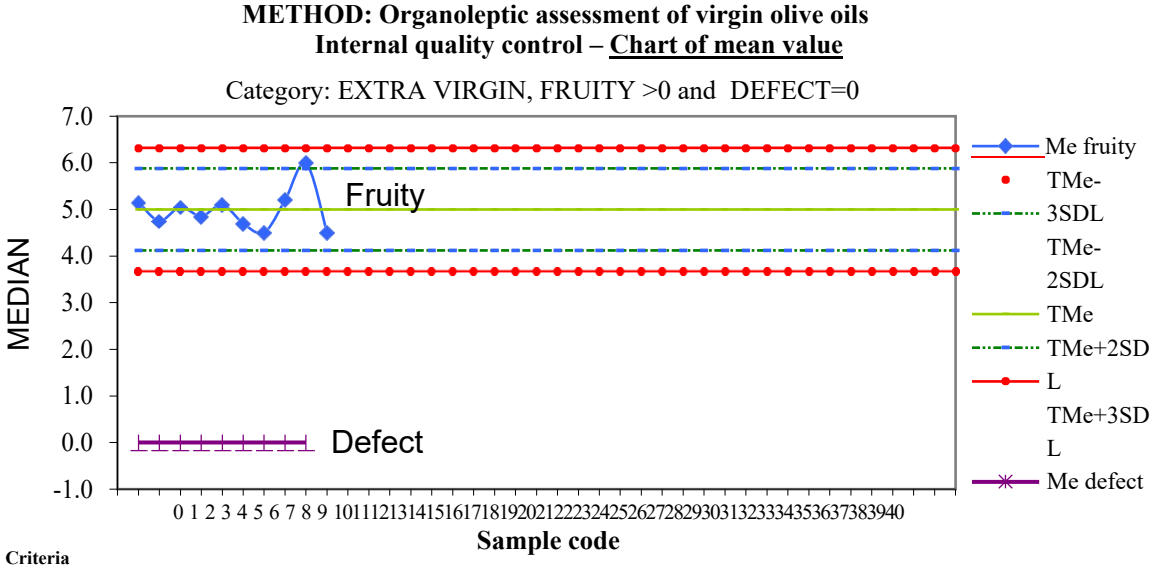
- Extra virgin: If defect >0, the analytical procedure is out of control.
- Virgin: If defect=0, the analytical procedure is out of control.
- Ordinary: If fruity>0 and defect≤3.5 or defect>6, the analytical procedure is out of control.

- Lampante: If defect≤6, the analytical procedure is out of control. In case that the category ordinary does not exist.
- Lampante: If fruity>0 and defect ≤3.5, the analytical procedure is out of control.

Some examples of quality control charts for each category are presented below, including examples of the criteria for interpreting the chart. In these charts:

- TMe is the "assigned value" of the quality control sample.
- SD_L is the standard deviation (not the robust standard deviation) determined when preparing the quality control sample or during the procedure to verify the method in the laboratory. The standard deviation of the method (± 0.7) could also be used.

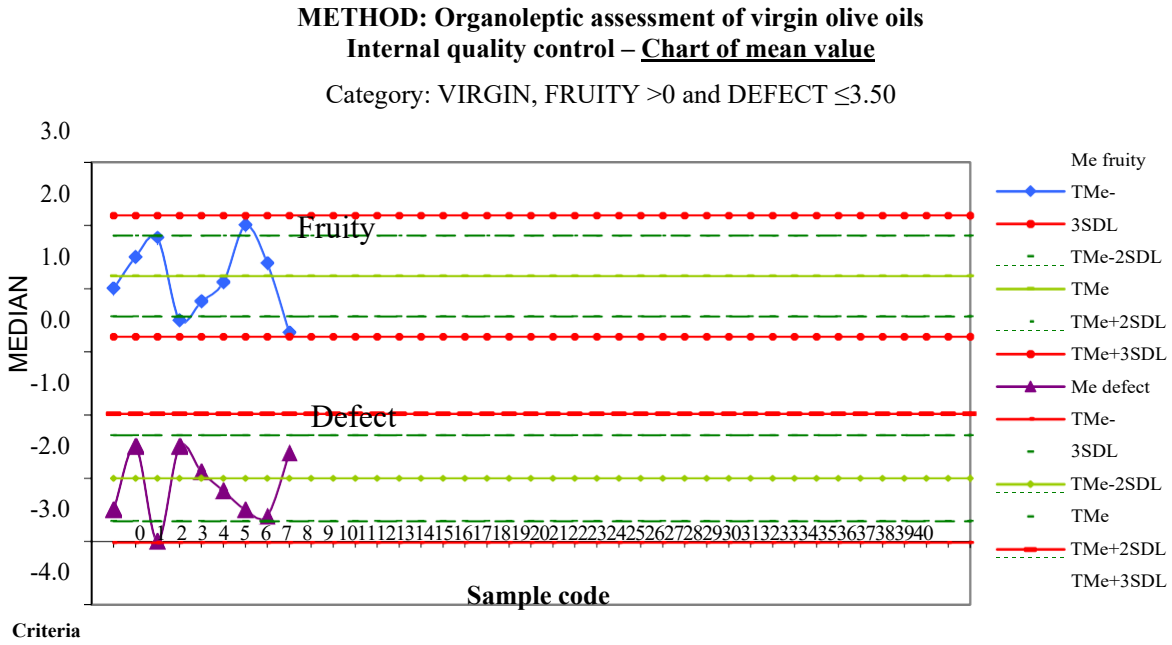
Figure 4. Example of quality control chart for the extra virgin category.



Criteria

1. If a violet point (defect) is >0, the analytical procedure is out of control.
2. If a blue point (fruity) is under or above the red line, the analytical procedure is out of control.
3. If 2 consecutive blue points (fruity) lie between red and dotted lines, the analytical procedure is out of control.
4. If 10 consecutive blue points (fruity) lie in the same side between the green and dotted lines, the analytical procedure is out of control.
5. If 7 consecutive blue points (fruity) lie in the same side between the green and dotted lines, there is a trend for the analytical procedure to be out of control.
6. If one from 20 consecutive blue points lie between the dotted and red lines, the analytical procedure is within control.

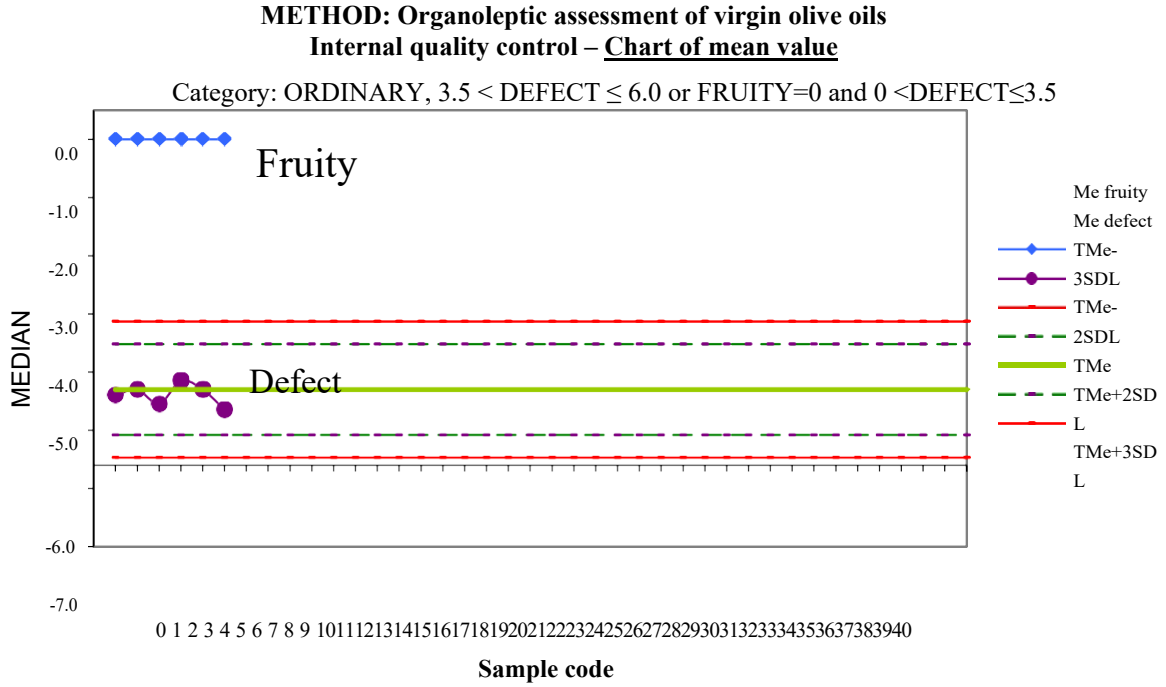
Figure 5. Example of quality control chart for the virgin category.



Criteria

1. If a violet point (defect) is equal to zero, the analytical procedure is out of control.
2. If a blue (fruity) or a violet point (defect) is under or above the red line, the analytical procedure is out of control.
3. If 2 consecutive blue points (fruity) or violet points (defect) lie between red and dotted lines, the analytical procedure is out of control.
4. If 10 consecutive blue or violet points lie in the same side between the green and dotted lines, the analytical procedure is out of control.
5. If 7 consecutive blue points (fruity) lie in the same side between the green and dotted lines, there is a trend for the analytical procedure to be out of control.
6. If one from 20 consecutive blue or violet points lie between the dotted and red lines, the analytical procedure is within control.

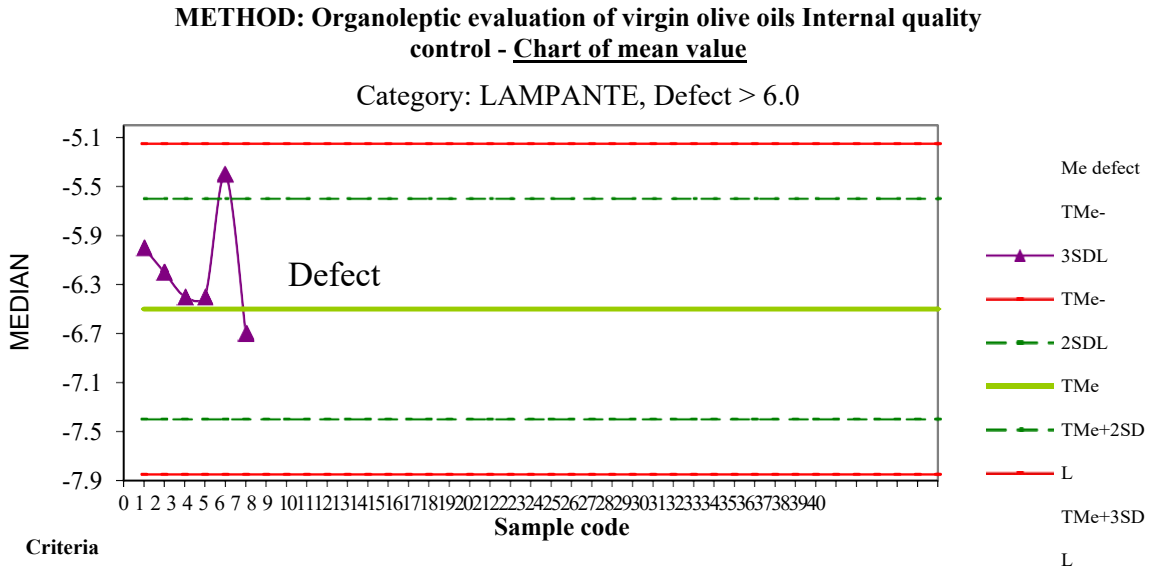
Figure 6. Example of quality control chart for the ordinary category.



Criteria

1. If a violet point (defect) is > -3.5 and a blue point (fruity) is > 0, the analytical procedure is out of control.
2. If a violet point is < -6, the analytical procedure is out of control.
3. If a violet point is above or under the red line, the analytical procedure is out of control.
4. If 2 consecutive violet points (defect) lie between red and dotted lines, the analytical procedure is out of control.
5. If 10 successive violet points (defect) lie in the same side between the green and dotted lines, the analytical procedure is out of control.
6. If 7 consecutive violet points (defect) lie in the same side between the green and dotted lines, there is a trend for the analytical procedure to be out of control.
7. If one from 20 consecutive violet points lie between the dotted and red lines, the analytical procedure is within control.

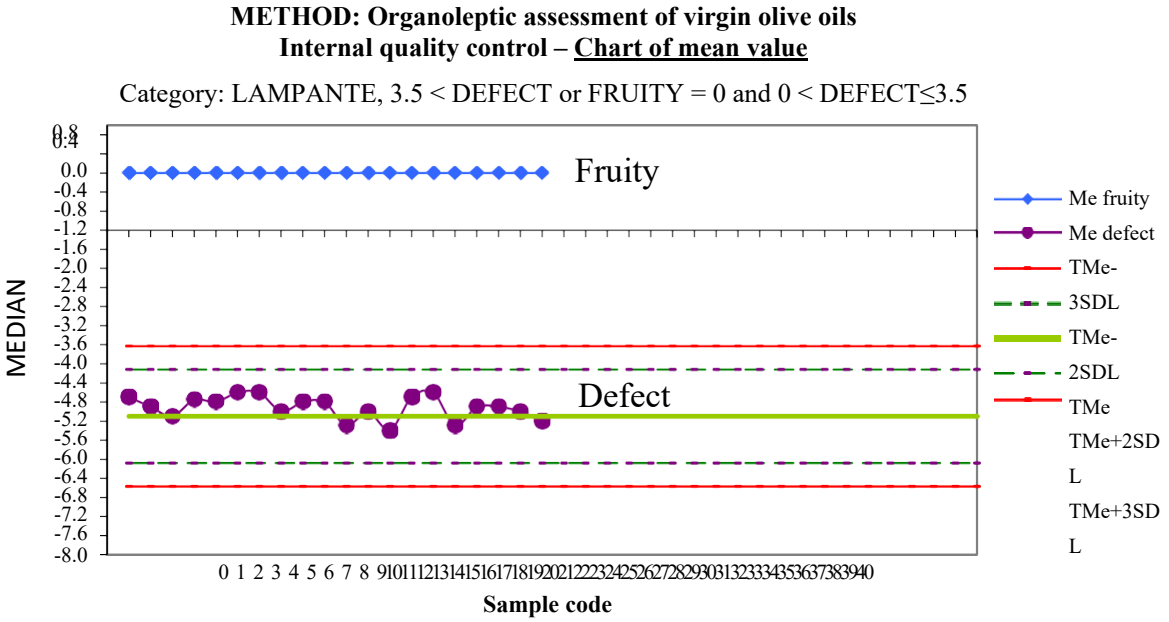
Figure 7. Example of quality control chart for the lampante category.



Criteria

1. If a violet point (defect) is > -6, the analytical procedure is out of control.
2. If a violet point is above or under the red line, the analytical procedure is out of control.
3. If 2 consecutive violet points (defect) lie between red and dotted lines, the analytical procedure is out of control.
4. If 10 consecutive violet points (defect) lie in the same side between the green and dotted lines, the analytical procedure is out of control.
5. If 7 consecutive violet points (defect) lie in the same side between the green and dotted lines, there is a trend for the analytical procedure to be out of control
6. If one from 20 consecutiveviolet points lie between the dotted and red lines, the analytical procedure is within control.

Figure 8. Example of quality control chart for the lampante category (in case that the ordinary category does not exist).



Criteria:

1. If a violet point (defect) is > -3.5 and a blue point (fruity) is > 0, the analytical procedure is out of control.
2. If a violet point is above or under the red line, the analytical procedure is out of control.
3. If 2 consecutive violet points (defect) lie between red and dotted lines, the analytical procedure is out of control.
4. If 10 consecutive violet points (defect) lie in the same side between the green and dotted lines, the analytical procedure is out of control.
5. If 7 consecutive violet points (defect) lie in the same side between the green and dotted lines, there is a trend for the analytical procedure to be out of control.
6. If one from 20 consecutive violet points lie between the dotted and red lines, the analytical procedure is within control.

APPENDIX I

WORKSHEETS FOR CALCULATING QUALITY CONTROL INDEXES INCLUDED IN THE DOCUMENT COI/T.20/Doc. Nº 17 AND FOR THEIR ILLUSTRATION IN CHARTS

Given that the indexes for the quality control of tasters and the whole panel included in the document COI/T.20/Doc. Nº 17 are numerous and complex, two worksheets are recommended for calculating these indexes. The purpose of these worksheets is to facilitate the function of sensory laboratories with easy-to-use tools. However, their use is not mandatory for sensory laboratories. Each sensory laboratory can use either the two worksheets recommended by the IOC or other worksheets developed by the laboratory itself to calculate the indexes.

In any case, and for the purpose of accreditation, the worksheets used should be verified by the sensory laboratory, comparing the results obtained by using the worksheets with those obtained from calculations made by hand.

Note: The two worksheets are attached to this document on the IOC website.

GENERAL INSTRUCTIONS FOR THE USE OF BOTH WORKSHEETS

To prevent users from making accidental errors when using the worksheets, both are locked. The password to unlock each sheet is **QCEFI2021.**

Quality control indexes included in COI/T20/Doc No.17 are divided in two main groups according to the field of their application: QC taster and QC panel. So, following this structure, two separate worksheets (Excel spreadsheets) are recommended:

1. QC TASTERS CALCS
2. QC PANEL CALCS

Note: The worksheet QC TASTERS CALCS should be applied to the data of each taster on the panel. So, each taster will have their own Excel spreadsheets and their performance will be monitored over time.

According to COI/T20/Doc No.17, the indexes for the quality control of tasters and panel and their illustration in charts are:

Quality control of	Analysis	Indexes	Charts
TASTER	Duplicate	1. Precision number 2. Deviation number	PN chart DN chart
	Reference or characterized samples	1. z-score 2. Deviation number	z-score chart DN chart
	Proficiency checks	SCOREct	
PANEL	Duplicate	1. Normalized error 2. Precision number	En chart PN chart
	Reference or characterized samples	1. z-score 2. Deviation number	z-score chart DN chart
	Quality control samples		$\bar{x}$ charts

The above indexes are calculated and all charts illustrated with the recommended worksheets.

In each worksheet, there are sub-sheets numbered 1, 2 and 3. Sheet number 1 is used to input data of duplicate analysis, sheet 2 is used to input data by analysis of reference or characterized samples and sheet 3 is used to either calculate SCOREct (worksheet for the taster) or for the charts of the quality control samples (worksheet for panel). The respective calculations can also be performed using these sheets.

At the upper part of each of these sheets, some instructions are written. In addition, there is the warning: "**Please, fill in only the cells with yellow color. The orange color cells include formulas.**" In these worksheets, the orange color cells are protected.

In all cases, the deviation number is calculated by taking into account the first duplicate. However, as it is referred in the instructions of worksheets, panel leader can easily use the second replicate, by changing the formula for its calculation. In the sheets included calculations, the formulas for the calculation of the various indexes are presented.

Each of the sheets 1 and 2 is accompanied by two other sheets (1a-1b and 2a-2b respectively), in which the charts are illustrated. These sheets are connected automatically with the respective sheet 1 or 2 and so you have **not to fill in anything** (relative note is written).

Correct calculation of the batch and continuous mode of the indexes DN and PN by using the two worksheets.

As mentioned in the COI/T20/Doc No.17:

"The sensory laboratory can use either

- one Precision Number and Deviation Number for the classified attribute identified by the panel (fruity for EVOO and predominant defect for other categories), or
- one for the defects and another for the fruity attribute, separately."

These worksheets calculate the indexes separately, so the number of values for fruity or defect is different depending on the category of samples.

To calculate the batch and continuous mode of the indexes using these worksheets, these are the following solutions:

1. The panel leader should proceed to this calculation when the number of tests is six for the fruity attribute or for the defect, changing the formula in the respective column.
2. The panel leader should calculate one index only for the classified attribute.
3. The sensory laboratory should use samples of VOO for quality control.
4. The creation of different sheets to calculate cumulative indexes (in batch or continuous mode).

DATA PROCESSING									
FRUITY					PREDOMINANT DEFECT				
Attribute	Normalized error fruity (En≤1.0)	PRECISION NUMBER fruity (≤2.00)	PNpf		Attribute	Normalized error defect (En≤1.0)	PRECISION NUMBER defect (≤2.00)	PNpd	
			BATCH MODE	CONTINUOUS MODE				BATCH MODE	CONTINUOUS MODE
FRUITY	1.40	0.04							
FRUITY	0.60	0.04							
					MUSTY	0.48	0.16		
FRUITY	0.60	0.01			RANCID	1.42	0.04		
FRUITY	0.36	0.04			MUDDY	0.89	0.04		
FRUITY	0.32	0.04	~~0.03~~	~~0.03~~	RANCID	0.72	0.04	~~0.05~~	~~0.05~~
FRUITY	0.49	0.04	0.04	0.04	MUDDY	0.31	0.01		~~0.05~~
FRUITY	0.44	0.04		0.04	RANCID	0.76	0.04	0.05	0.05

SPECIFIC INSTRUCTIONS FOR THE USE OF EACH WORKSHEET

The **worksheet QC TASTERS CALCS** consists of the following sheets:

- 1 TASTER DATA DOUBLE SAMPLES: In this sheet, data by duplicate analysis of each taster are entered and the indexes PN and DN are calculated.
  - ✓ 1a TASTER CHART PN: This sheet outlines the PN charts for fruity and defect.
  - ✓ 1b TASTER CHART DN: This sheet outlines the DN charts for fruity and defect.
- 2 TASTER DATA REF SAMPLES: In this sheet, data by using analysis of reference or characterised samples of each taster are entered and the indexes DN and z-score are calculated. Note that the **z-score is calculated taking the standard deviation of the method (0.7) into account.**
  - ✓ 2a TASTER CHART DN REF: This sheet outlines the DN charts for fruity and defect, when the taster analyses the reference sample.
  - ✓ 2b TASTER CHART Z-SCORE: This sheet outlines the z-score charts for fruity and defect.
- 3 SCOREct: In this sheet, the competence of a taster is evaluated by using the results of the taster in the last interlaboratory proficiency testing. That is, **this is the external quality control of the tasters**.

The **worksheet QC PANEL CALCS** consists of the following sheets:

- 1 PANEL DATA DOUPLE SAMPLES: In this sheet, data by duplicate analysis of panel are entered and the indexes PN and normalized error are calculated.
  - ✓ 1a PANEL CHART PN: This sheet outlines the PN charts for fruity and defect.
  - ✓ 1b PANEL CHART En: This sheet outlines the charts of normalized error for fruity and defect.
- 2 PANEL DATA REF SAMPLES: In this sheet, data by using analysis of reference or characterized samples of each taster are entered and the indexes DN and z-score are calculated. Note that the **z-score is calculated taking the standard deviation of the method (0.7) into account.**
  - ✓ 2a PANEL CHART DN: This sheet outlines the DN charts for fruity and defect, in case that the panel performs analysis of reference sample.
  - ✓ 2b PANEL CHART Z-SCORE: This sheet outlines the charts of z-score for fruity and defect.
- 3 X CHARTS are used to illustrate the results of the quality control samples used by the laboratory. So, this programme includes five sheets for the various categories, that is:
  - ✓ 3 X CHART EVOO: for extra virgin olive oil
  - ✓ 3 X CHART VOO: for virgin olive oil
  - ✓ 3 X CHART OVOO: for ordinary virgin olive oil
  - ✓ 3 X CHART LOO IOC: for lampante virgin olive oil according to the IOC TRADE STANDARD
  - ✓ 3 X CHART LOO EU: for lampante olive oil according to Reg. (EEC) 2568/91.