E

UPOV

TC-EDC/Jan12/4
ORIGINAL: English
DATE: December 16, 2011

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

ENLARGED EDITORIAL COMMITTEE

Geneva, January 11 and 12, 2012

REVISION OF TGP/8 "TRIAL DESIGN AND TECHNIQUES USED IN THE EXAMINATION OF DISTINCTNESS, UNIFORMITY AND STABILITY"

Document prepared by the Office of the Union

1. The purpose of this document is to report on the developments with respect to document TGP/8 "Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability" (document TGP/8/2) based upon the approach approved by the Technical Committee (TC) at its forty-seventh session, held in Geneva from April 4 to 6, 2011 (see document TC/47/26 "Report on the Conclusions", paragraphs 72 to 74), and the discussions at the Technical Working Parties (TWPs) at their sessions in 2011.

2. The following abbreviations are used in this document:

CAJ:	Administrative and Legal Committee
TC:	Technical Committee
TC-EDC:	Enlarged Editorial Committee
TWA:	Technical Working Party for Agricultural Crops
TWC:	Technical Working Party on Automation and Computer Programs
TWF:	Technical Working Party for Fruit Crops
TWO:	Technical Working Party for Ornamental Plants and Forest Trees
TWPs:	Technical Working Parties
TWV:	Technical Working Party for Vegetables

3. The structure of this document is as follows:

I. BACKGROUND

4. At its meeting on January 8, 2009, the Enlarged Editorial Committee (TC-EDC) noted that there were a number of sections within document TGP/8/1 Draft 1 for which development had not yet started, or for which substantial further development would be required. At the same time, the TC-EDC noted that there were a number of important sections within TGP/8 that were well-established and could already provide useful guidance. Therefore, the TC-EDC proposed that the TC should be invited to consider the adoption of a first version of document TGP/8 (document TGP/8/1) without the sections of that document that would require further substantial development. The TC-EDC also noted that the identification of well-established text within document TGP/8 would justify translation of those sections. With regard to the sections of document TGP/8 that would not be included in the first version of document TGP/8 (document TGP/8/1), the TC-EDC proposed that those sections should continue to be developed without delay and should be incorporated into document TGP/8 by means of a revision of document TGP/8 (document TGP/8/2) at the earliest opportunity.

5. The TC at its forty-fifth session, held in Geneva from March 30 to April 1, 2009, agreed that document TGP/8/1 should be scheduled for adoption in 2010 on the basis of the content included in document TGP/8/1 Draft 12. The TC further agreed that, at the same time, separately from consideration of the draft of document TGP/8/1, the sections omitted from document TGP/8/1 Draft 12, as reproduced in document TC/45/14, Annex I, should continue to be developed without delay and should be incorporated into document TGP/8 by means of a revision of document TGP/8/1 (i.e. document TGP/8/2) at the earliest opportunity (see documents TC/45/5 "TGP Documents" paragraph 24 and TC/45/16, "Report", paragraph 136).

6. The Technical Committee at its forty-sixth session in Geneva from March 22 to 24, 2010, agreed that, subject to agreement by the Administrative and Legal Committee (CAJ), document TGP/8/1 Draft 15, as amended by the TC, should be put forward for adoption by the Council at its forty-fourth ordinary session, to be held in Geneva on October 21, 2010. The TC noted that the French, German and Spanish translations of the original English text would be checked by the relevant members of the Editorial Committee prior to submission of the draft of document TGP/8/1 to the Council.

7. The CAJ, at its sixty-first session, held in Geneva on March 25, 2010, agreed that document TGP/8/1 Draft 15, as amended by the TC, should be put forward for adoption by the Council at its forty-fourth ordinary session, to be held in Geneva on October 21, 2010.

8. At its forty-fourth ordinary session in Geneva on October 21, 2010, the Council of UPOV adopted document TGP/8/1 "Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability" on the basis of document TGP/8/1 Draft 16.

9. The Technical Committee (TC) at its forty-seventh session held in Geneva, from April 4 to 6, 2011 considered document TC/47/20. (see document TC/47/26 "Report on the Conclusions", paragraph 72).

10. The TC noted the comments made by the TWPs at their sessions in 2010, with regard to document TGP/8, as set out in document TC/47/20, paragraphs 18 and 24. It agreed that the text of TGP/8/1 "Trial Design and Techniques Used in the Examination of Distinctness, Uniformity and Stability", Part II, should be amended in a future revision as follows:

(a) 1. The GAIA Methodology, Section 1.3.1.1, should be amended to clarify that there is an assumption that the length of panicle is used as a characteristic;

(b) 5: Pearson's Chi-Square Test Applied to Contingency Tables, Section 5.5 (4) should be amended to read: "(4) Always use Yates Correction for determining the chi-square test with only one degree of freedom."

11. The TC agreed the workplan for the development of TGP/8/2, as presented in Annex XV to this document.

II. DISCUSSIONS ON THE REVISION OF DOCUMENT TGP/8 AT THE TECHNICAL WORKING PARTIES DURING THEIR SESSIONS IN 2011

12. At their sessions in 2011, the TWPs considered the sections for further development of TGP/8. The proposals of the TWPs are included in the relevant annexes.

[Annexes follow]

TGP/8 PART I: DUS TRIAL DESIGN AND DATA ANALYSIS

	Comments of the TWPs in 2011	
General	The TWC received a presentation by Mrs. Sally Watson (United Kingdom) on "Cyclic Planting of Established Varieties to Reduce Trial Size". The TWC agreed that the text should be included in TGP/8 Part I in a new section on the reduction of the size of the trials.	TWC

New Section 2 - Data to be recorded (Drafter: Mr. Uwe Meyer (Germany)

2. DATA TO BE RECORDED

	Comments of the TWPs in 2011	
General	The TWA agreed that this document contains valuable information and should therefore be included in TGP/8.	TWA
	The TWC agreed that a new version of the document be prepared for discussion with the view of its incorporation into TGP/8	TWC
	The TWV and the TWF agreed that the information provided in Annex I should be included in document TGP/8.	TWV TWF
	The TWO agreed to replace the term "crop expert" with "DUS expert"	TWO

2.1 Introduction

Document TGP/9 Examining Distinctness, sections 4.4 and 4.5 provide the following guidance on the type of observation for distinctness in respect to the type of characteristic and the method of propagation of the variety:

"4.4 Recommendations in the UPOV Test Guidelines

"The indications used in UPOV Test Guidelines for the method of observation and the type of record for the examination of distinctness, are as follows:

"Method of observation

"M: to be measured (an objective observation against a calibrated, linear scale e.g. using a ruler, weighing scales, colorimeter, dates, counts, etc.);

"V: to be observed visually (includes observations where the expert uses reference points (e.g. diagrams, example varieties, side-by-side comparison) or non-linear charts (e.g. color charts). "Visual" observation refers to the sensory observations of the expert and, therefore, also includes smell, taste and touch.

"Type of record(s)

"G: single record for a variety, or a group of plants or parts of plants; "S: records for a number of single, individual plants or parts of plants

"For the purposes of distinctness, observations may be recorded as a single record for a group of plants or parts of plants (G), or may be recorded as records for a number of single, individual plants or parts of plants (S). In most cases, "G" provides a single record per variety and it is not possible or necessary to apply statistical methods in a plant-by-plant analysis for the assessment of distinctness.

"4.5 Summary

"The following table summarizes the common method of observation and type of record for the assessment of distinctness, although there may be exceptions:

	Type of expression of characteristic		
Method of propagation of the variety	QL	PQ	QN
Vegetatively propagated	VG	VG	VG/MG/MS
Self-pollinated	VG	VG	VG/MG/MS
Cross-pollinated	VG/(VS*)	VG/(VS*)	VS/VG/MS/MG
Hybrids	VG/(VS*)	VG/(VS*)	**

* Records of individual plants only necessary if segregation is to be recorded.
** To be considered according to the type of hybrid."

2.2 Types of expression of characteristics

2.2.1 Characteristics can be classified according to their types of expression. The following types of expression of characteristics are defined in the General Introduction to the Examination of Distinctness, Uniformity and Stability and the Development of Harmonized Descriptions of New Varieties of Plants, (document TG/1/3, the "General Introduction", Chapter 4.4):

2.2.2 Qualitative characteristics" (QL) are those that are expressed in discontinuous states (e.g. sex of plant: dioecious female (1), dioecious male (2), monoecious unisexual (3), monoecious hermaphrodite (4)). These states are self-explanatory and independently meaningful. All states are necessary to describe the full range of the characteristic, and every form of expression can be described by a single state. The order of states is not important. As a rule, the characteristics are not influenced by environment.

2.2.3 "Quantitative characteristics" (QN) are those where the expression covers the full range of variation from one extreme to the other. The expression can be recorded on a one-dimensional, continuous or discrete, linear scale. The range of expressions is divided into a number of states for the purpose of description (e.g. length of stem: very short (1), short (3), medium (5), long (7), very long (9)). The division seeks to provide, as far as practical, an even distribution across the scale. The Test Guidelines do not specify the difference needed for distinctness. The states of expression should, however, be meaningful for DUS assessment.

2.2.4 In the case of "pseudo-qualitative characteristics" (PQ) the range of expression is at least partly continuous, but varies in more than one dimension (e.g. shape: ovate (1), elliptic (2), circular (3), obovate (4)) and cannot be adequately described by just defining two ends of a linear range. In a similar way to qualitative (discontinuous) characteristics – hence the term "pseudo-qualitative" – each individual state of expression needs to be identified to adequately describe the range of the characteristic.

2.3 Types of scales of data

2.3.1 The possibility to use specific procedures for the assessment of distinctness, uniformity and stability depends on the scale level of the data which are recorded for a characteristic. The scale level of data depends on the type of expression of the characteristic and on the way of recording this expression. The type of scale may be nominal, ordinal, interval or ratio.

2.3.2 Data from qualitative characteristics

2.3.2.1 Data results from qualitative characteristics are nominal scaled data without any logical order of the discrete categories. They result from visually assessed (notes) qualitative characteristics.

Examples:

Type of scale	**Example**	**Example number**
nominal	Sex of plant	1
nominal with two states	Leaf blade: variegation	2

For description of the states of expressions, see Table 6.

2.3.2.2 A nominal scale consists of numbers which correspond to the states of expression of the characteristic, which are referred to in the Test Guidelines as notes. Although numbers are used for designation there is no inevitable order for the expressions and so it is possible to arrange them in any order.

	Comments of the TWPs in 2011	
2.3.2.2	The TWC to replace the term "inevitable" by "logical" in the second sentence.	TWC

2.3.2.3 Characteristics with only two categories (dichotomous characteristic) are a special form of nominal scaled characteristic.

	Comments of the TWPs in 2011	
2.3.2.3	The second line to read as follows "form of a nominal scaled characteristic."	TWC

2.2.2.4 The nominal scale is the lowest classification of the scales (Table 1). Few statistical procedures are applicable for evaluations (section 2.3.8 *[cross ref.]*).

2.3.3 Data from quantitative characteristics

2.3.3.1 Data results from quantitative characteristics are metric (ratio or interval) or ordinal scaled data.

2.3.3.2 Metric scaled data are all data which are recorded by measuring or counting. Weighing is a special form of measuring. Metric scaled data can have a continuous or a discrete distribution. Continuous metric data result from measurements. They can take every value out of the defined range. Discrete metric data result from counting.

Examples

Type of scale	**Example**	**Example number**
Continuous metric	Plant length in cm	3
Discrete metric	Number of stamens	4

For description of the states of expression, see Table 6.

2.3.3.3 The continuous metric scaled data for the characteristic "Plant length" are measured on a continuous scale with defined units of assessment. A change of unit of measurement e.g. from cm into mm is only a question of precision and not a change of type of scale.

2.3.3.4 The discrete metric scaled data of the characteristic "Number of stamens" are assessed by counting (1, 2, 3, 4, and so on). The distances between the neighboring units of assessment are constant and for this example equal to 1. There are no real values between two neighboring units but it is possible to compute an average which falls between those units.

2.3.3.5 In biometrical terminology, metric scales are referred to as quantitative scales or cardinal scales. Metric scales can be subdivided into ratio scales and interval scales.

	Comments of the TWPs in 2011	
2.3.3.5	To delete the first sentence.	TWC

2.3.3.6 Ratio scale

2.3.3.6.1 A ratio scale is a metric scale with a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Ratio scaled data may be continuous or discrete.

The absolute zero point:

2.3.3.6.2 The definition of an absolute zero point makes it possible to define meaningful ratios. This is a requirement for the construction of index numbers (e.g. the ratio of length to width).

	Comments of the TWPs in 2011	
2.3.3.6.2	To read as follows: "The definition of an absolute zero point makes it possible to define	TWC

	meaningful ratios. This is a requirement for the construction of indexes, which are the combination of at least two characteristics (e.g. the ratio of length to width). In the General Introduction, this is referred to as a combined characteristic (see document TG/1/3, section 4.6.3)."	

An index is the combination of at least two characteristics. In the General Introduction, this is referred to as a combined characteristic (see document TG/1/3, section 4.6.3).

2.3.3.6.3 It is also possible to calculate ratios between expressions of different varieties. For example, in the characteristic 'Plant length' assessed in cm, there is a lower limit for the expression which is '0 cm' (zero). It is possible to calculate the ratio of length of plant of variety 'A' to length of plant of variety 'B' by division:

Length of plant of variety 'A' = 80 cm
Length of plant of variety 'B' = 40 cm
Ratio = Length of plant of variety 'A' / Length of plant of variety 'B'
= 80 cm / 40 cm
= 2.

2.3.3.6.4 So it is possible in this example to state that plant 'A' is double the length of plant 'B'. The existence of an absolute zero point ensures an unambiguous ratio.

2.3.3.6.5 The ratio scale is the highest classification of the scales (Table 1). That means that ratio scaled data include the highest information about the characteristic and it is possible to use many statistical procedures (section 2.3.8 *[cross ref.]*).

2.3.3.6.6 The examples 3 and 4 (Table 6) are examples for characteristics with ratio scaled data.

2.3.3.7 Interval scale

2.3.3.7.1 An Interval scale is a metric scale without a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Interval scaled data may be distributed continuously or discretely.

	Comments of the TWPs in 2011	
2.3.3.7.1	To replace the term "expressions" by "units" in the second sentence.	TWC

2.3.3.7.2 An example for a discrete interval scaled characteristic is 'Time of beginning of flowering' measured as date which is given as example 5 in Table 6. This characteristic is defined as the number of days from April 1. The definition is useful but arbitrary and April 1 is not a natural limit. It would also be possible to define the characteristic as the number of days from January 1.

2.3.3.7.3 It is not possible to calculate a meaningful ratio between two varieties which should be illustrated with the following example:

	Comments of the TWPs in 2011	
2.3.3.7.3	The second line to read “is illustrated by the following example”	TWC

Variety ‘A’ begins to flower on May 30 and variety ‘B’ on April 30

Case I) Number of days from April 1 of variety ‘A’ = 60
Number of days from April 1 of variety ‘B’ = 30

$$\text{Ratio}_{I} = \frac{\text{Number of days from April 1 of variety ‘A’}}{\text{Number of days from April 1 of variety ‘B}} = \frac{\text{60 days}}{\text{30 days}} = 2$$

Case II) Number of days from January 1 of variety ‘A’ = 150
Number of days from January 1 of variety ‘B’ = 120

$$\text{Ratio}_{II} = \frac{\text{Number of days from January 1 of variety ‘A’}}{\text{Number of days from January 1 of variety ‘B}} = \frac{\text{150 days}}{\text{120 days}} = 1.25$$

$\text{Ratio}_{I} = 2 > 1.25 = \text{Ratio}_{II}$

2.3.3.7.4 It is impossible to state that the time of flowering of variety ‘A’ is twice that of variety ‘B’. The ratio depends on the choice of the zero point of the scale. This kind of scale is defined as an “Interval scale”: a metric scale without a defined absolute zero point.

	Comments of the TWPs in 2011	
2.3.3.7.4	To replace “impossible” by “incorrect” in the first.	TWC

2.3.3.7.5 The interval scale is lower classified than the ratio scale (Table 1). Fewer statistical procedures can be used with interval scaled data than with ratio scaled data (section 2.3.8 *[cross ref.]*). The interval scale is theoretically the minimum scale level to calculate arithmetic mean values.

	Comments of the TWPs in 2011	
2.3.3.7.5	To read as follows: “The interval scale is lower classified than the ratio scale (Table 1). At the interval scale, no useful indexes can be formed such as ratios. The interval scale is theoretically the minimum scale to calculate arithmetic mean values.”	TWC

2.3.3.8 Ordinal scale

2.3.3.8.1 Discrete categories of ordinally scaled data can be arranged in an ascending or descending order. They result from visually assessed (notes) quantitative characteristics.

Example:

Type of scale	Example	Example number
ordinal	Intensity of anthocyanin	6

For description of the states of expressions, see Table 6

2.3.3.8.2 An ordinal scale consists of numbers which correspond to the states of expression of the characteristic (notes). The expressions vary from one extreme to the other and thus they have a clear logical order. It is not possible to change this order, but it is not important which numbers are used to denote the categories. In some cases ordinal data may reach the level of discrete interval scaled data or of discrete ratio scaled data (section 2.3.8 *[cross ref.]*).

	Comments of the TWPs in 2011	
2.3.3.8.2	The third sentence to read: "It is not important which numbers are used to denote the categories."	TWC

2.3.3.8.3 The distances between the discrete categories of an ordinal scale are not exactly known and not necessarily equal. Therefore, an ordinal scale does not fulfil the condition to calculate arithmetic mean values, which is the equality of intervals throughout the scale.

2.3.3.8.4 The ordinal scale is lower classified than the interval scale (Table 1). Less statistical procedures can be used for ordinal scale than for each of the higher classified scale data (section 2.3.8 *[cross ref.]*).

2.3.4 Data from pseudo-qualitative characteristics

2.3.4.1 Data results from pseudo-qualitative characteristics are nominal scaled data without any logical order of all discrete categories. They result from visually assessed (notes) qualitative characteristics.

Examples:

Type of scale	Example	Example number
nominal	Shape	7
nominal	Flower color	8

For description of the states of expressions, see Table 6.

2.3.4.2 A nominal scale consists of numbers which correspond to the states of expression of the characteristic, which are referred to in the Test Guidelines as notes. Although numbers are used for designation there is no inevitable order for all of the expressions. It is possible to arrange only some of them in an order.

2.3.4.3 The nominal scale is the lowest classification of the scales (Table 1). Few statistical procedures are applicable for evaluations (section 2.3.8 *[cross ref.]*).

2.3.5 The different types of scales are summarized in the following table.

Table 1: Types of expressions and type of scales

Type of expression	Type of scale	Description	Distribution	Data recording	Scale Level
QN	ratio	constant distances with absolute zero point	Continuous	Absolute Measurements	High
			Discrete	Counting	
	interval	constant distances without absolute zero point	Continuous	Relative measurements	↑
			Discrete	Date	
	ordinal	Ordered expressions with varying distances	Discrete	Visually assessed notes	↑
PQ or QL	nominal	No order, no distances	Discrete	Visually assessed notes	Low

2.3.6 Scale levels for variety description

The description of varieties is based on the states of expression (notes) which are given in the Test Guidelines for the specific crop. In the case of visual assessment, the notes from the Test Guidelines are usually used for recording the characteristic as well as for the assessment of DUS. The notes are distributed on a nominal or ordinal scale (see Part I: section 4.5.4.2 *[cross ref.]*). For measured or counted characteristics, DUS assessment is based on the recorded values and the recorded values are transformed into states of expression only for the purpose of variety description.

	<u>Comments of the TWPs in 2011</u>	
2.3.6	The TWO considered that paragraph “2.3.6 Scale levels for variety description” should be revised to reflect the use of notes and measurements for the examination of distinctness, as set out in TGP/9 “Examining Distinctness”.	TWO

2.3.7 Relation between types of expression of characteristics and scale levels of data

2.3.7.1 Records taken for the assessment of <u>qualitative characteristics</u> are distributed on a nominal scale, for example “Sex of plant”, “Leaf blade: variegation” (Table 6, examples 1 and 2).

2.3.7.2 For <u>quantitative characteristics</u> the scale level of data depends on the method of assessment. They can be recorded on a metric (when measured or counted) or ordinal (when visually observed) scale. For example, “Length of plant” can be recorded by measurements resulting in ratio scaled continuous metric data. However, visual assessment on a 1 to 9 scale may also be appropriate. In this case, the recorded data are ordinal scaled because the size of intervals between the midpoints of categories is not exactly the same.

Remark: In some cases visually assessed data on metric characteristics may be handled as measurements. The possibility to apply statistical methods for metric data depends on the precision of the assessment and the robustness of the statistical procedures. In the case of very precise visually assessed quantitative characteristics the usually ordinal data may reach the level of discrete interval scaled data or of discrete ratio scaled data.

2.3.7.3 A <u>pseudo-qualitative type of characteristic</u> is one in which the expression varies in more than one dimension. The different dimensions are combined in one scale. At least one dimension is quantitatively expressed. The other dimensions may be qualitatively expressed or quantitatively expressed. The scale as a whole has to be considered as a nominal scale (e.g. "Shape", "Flower color"; Table 6, examples 7 and 8).

2.3.7.4 In the case of using the off-type procedure for the assessment of uniformity the recorded data are nominally scaled. The records fall into two qualitative classes: plants belonging to the variety (true-types) and plants not belonging to the variety (off-types). The type of scale is the same for qualitative, quantitative and pseudo-qualitative characteristics.

2.3.7.5 The relation between the type of characteristics and the type of scale of data recorded for the assessment of distinctness and uniformity is described in Table 2. A <u>qualitative characteristic</u> is recorded on a nominal scale for distinctness (state of expression) and for uniformity (true-types vs. off-types). <u>Pseudo-qualitative characteristics</u> are recorded on a nominal scale for distinctness (state of expression) and on a nominal scale for uniformity (true-types vs. off-types). <u>Quantitative characteristics</u> are recorded on an ordinal, interval or ratio scale for the assessment of distinctness depending on the characteristic and the method of assessment. If the records are taken from single plants the same data may be used for the assessment of distinctness and uniformity. If distinctness is assessed on the basis of a single record of a group of plants, uniformity has to be judged with the off-type procedure (nominal scale).

<u>Table 2: Relation between type of characteristic and type of scale of assessed data</u>

Procedure	Type of scale	Distribution	Type of characteristic		
			Qualitative	Pseudo-qualitative	Quantitative
Distinctness	ratio	Continuous	No	No	**<u>Yes</u>**
		Discrete	No	No	**<u>Yes</u>**
	interval	Continuous	No	No	**<u>Yes</u>**
		Discrete	No	No	**<u>Yes</u>**
	ordinal	Discrete	No	No	**<u>Yes</u>**
	nominal	Discrete	**<u>Yes</u>**	**<u>Yes</u>**	No
Uniformity	ratio	Continuous	No	No	**<u>Yes</u>**
		Discrete	No	No	**<u>Yes</u>**
	interval	Continuous	No	No	**<u>Yes</u>**
		Discrete	No	No	**<u>Yes</u>**
	ordinal	Discrete	No	No	**<u>Yes</u>**
	nominal	Discrete	**<u>Yes</u>**	**<u>Yes</u>**	**<u>Yes</u>**

2.3.8 Relation between method of observation of characteristics, scale levels of data and recommended statistical procedures

2.3.8.1 Established statistical procedures can be used for the assessment of distinctness and uniformity considering the scale level and some further conditions such as the degree of freedom or unimodality (Tables 3 and 4).

2.3.8.2 The relation between the expression of characteristics and the scale levels of data for the assessment of distinctness and uniformity is summarized in Table 6.

Table 3: Statistical procedures for the assessment of distinctness

Type of scale	Distribu-tion	Observation method	Procedure and further Conditions	Reference document
ratio	continuous	MS MG (VS) [1]	COYD Normal distribution, df >=20 long term LSD Normal distribution, df<20 2 out of 3 methods (LSD 1%) Normal distribution, df>=20	TGP/9
	discrete			
interval	continuous			
	discrete			
ordinal	discrete	VG VS	See explanation for QN characteristics in TGP/9 sections 5.2.2 and 5.2.3, See explanation for QN characteristics in TGP/9 section 5.2.4	TGP/9 TWC/14/12
nominal	discrete	VG (VS) [2]	See explanation for QL and PQ characteristics in TGP/9 sections 5.2.2 and 5.2.3	TGP/9

1) see remark in section 2.3.3.8.2 *[cross ref.]*
2) normally VG but VS would be possible

	<u>Comments of the TWPs in 2011</u>	
Table 3	In the column "Procedure and further Conditions" - Subject to agreement, to refer to the new recommendation on the number of degrees of freedom - to replace "Long term LSD" by Long term COYD" - to replace "2 out of 3 method" by "2 x 1% Method"	TWC

Table 4: Statistical procedures for the assessment of uniformity

Type of scale	Distribu-tion	Observa-tion method	Procedure and Further Conditions	Reference document
ratio	continuous	MS	COYU Normal distribution	TGP/10
	discrete	MS	2 out of 3 method $(s^2_c<=1.6s^2_s)$ Normal distribution	
interval	continuous	VS	LSD for untransformed percentage of off-types	
	discrete			
ordinal	discrete	VS	threshold model	TWC/14/12
nominal	discrete	VS	off-type procedure for dichotomous (binary) data	TGP/10

2.4 Different levels to look at a characteristic

2.4.1 Characteristics can be considered in different levels of process (Table 5). The characteristics as expressed in the trial (type of expression) are considered as process level 1. The data taken from the trial for the assessment of distinctness, uniformity and stability are defined as process level 2. These data are transformed into states of expression for the purpose of variety description. The variety description is process level 3.

Table 5: Definition of different process levels to consider characteristics

Process level	Description of the process level
1	characteristics as expressed in trial
2	data for evaluation of characteristics
3	variety description

From the statistical point of view, the information level decreases from process level 1 to 3. Statistical analysis is only applied in level 2.

2.4.2 Sometimes for crop experts it seems that there is no need to distinguish between different process levels. The process level 1, 2 and 3 could be identical. However, in general, this is not the case.

2.4.3 Understanding the need for process levels

2.4.3.1 The crop expert may know from UPOV Test Guidelines or his own experience that, for example, ‘Length of plant’ is a good characteristic for the examination of DUS. There are varieties which have longer plants than other varieties. Another characteristic could be ‘Variegation of leaf blade’. For some varieties, variegation is present and for others not. The crop expert has now two characteristics and he knows that ‘Plant length’ is a quantitative characteristic and ‘Variegation of leaf blade’ is a qualitative characteristic (definitions: see Part I: section 2.2.3 to 2.2.2 *[cross ref.]* below). This stage of work can be described as **process level 1.**

2.4.3.2 The crop expert then has to plan the trial and to decide on the type of observation for the characteristics. For characteristic 'Variegation of leaf blade', the decision is clear. There are two possible expressions: 'present' or 'absent'. The decision for characteristic 'Plant length' is not specific and depends on expected differences between the varieties and on the variation within the varieties. In many cases, the crop expert will decide to measure a number of plants (in cm) and to use special statistical procedures to examine distinctness and uniformity. But it could also be possible to assess the characteristic 'Plant length' visually by using expressions like 'short', 'medium' and 'long', if differences between varieties are large enough (for distinctness) and the variation within varieties is very small or absent in this characteristic. The continuous variation of a characteristic is assigned to appropriate states of expression which are recorded by notes (see document TGP/9, section 4)*[cross ref.]*. The crucial element in this stage of work is the recording of data for further evaluations. It is described as **process level 2**.

	Comments of the TWPs in 2011	
2.4.3.2	To review the text to avoid the inference that statistics are always applied to characteristics observed by measurement, which does not take into account MG observations (e.g. Section 2.4.3.2)	TWO

2.4.3.3 At the end of the DUS test, the crop expert has to establish a description of the varieties using notes from 1 to 9 or parts of them. This phase can be described as **process level 3**. For 'Variegation of leaf blade' the crop expert can take the same states of expression (notes) he recorded in process level 2 and the three process levels appear to be the same. In cases where the crop expert decided to assess 'Plant length' visually, he can take the same states of expression (notes) he recorded in process level 2 and there is no obvious difference between process level 2 and 3. If the characteristic 'Plant length' is measured in cm, it is necessary to assign intervals of measurements to states of expressions like 'short', 'medium' and 'long' to establish a variety description. In this case, for statistical procedures, it is important to be clearly aware of the relevant level and to understand the differences between characteristics as expressed in the trial, data for evaluation of characteristics and the variety description. This is absolutely necessary for choosing the most appropriate statistical procedures in cooperation with statisticians or by the crop expert.

4.5 Summary

The following table summarizes the common method of observation and type of record for the assessment of distinctness, although there may be exceptions:

	Type of expression of characteristic		
Method of propagation of the variety	QL	PQ	QN
Vegetatively propagated	VG	VG	VG/MG/MS
Self-pollinated	VG	VG	VG/MG/MS
Cross-pollinated	VG/(VS*)	VG/(VS*)	VS/VG/MS/MG
Hybrids	VG/(VS*)	VG/(VS*)	**

* Records of individual plants only necessary if segregation is to be recorded.
** To be considered according to the type of hybrid."

The following sections consider the data in relation to the type of record and type of trial design:

2.2 Side-by-side visual comparison[a]
2.2.1 When distinctness is assessed by side-by-side visual comparison, uniformity is assessed by off-types. In these cases, the trial design is a single plot, there is a single record per variety, which is obtained from visual observations of a group of plants or part of plants (VG), which provide notes (see sections 1.6.1.6 and 1.6.2) *[cross ref.]*.

2.3 Notes/Single variety records[b]
2.3.1 When distinctness is assessed by notes/single variety records, uniformity is assessed by off-types. In these cases, the trial design consists of single plots[c]. There is a single record per variety which is obtained from visual observation of a group of plants or part of plants (VG), providing a note, or a measurement of a group of plants or parts of plants (MG) (see sections 1.6.1.6 and 1.6.2) *[cross ref.]*.

2.4 Variety mean/statistical analysis of records of group of plants [variety mean statistical analysis of records of group data][d]
2.4.1 In general, when distinctness is assessed, for at least some characteristics, by a variety mean or by statistical analysis of groups of plants, uniformity is assessed by off-types. In these cases, the trial design is replicate plots (see sections 1.6.1.7 and 1.6.3.2) *[cross ref.]*.

2.4.2 Records from visual observation of a group of plants or part of plants provide notes which belong to qualitative scale data. It is important to note that, in general, it is not possible to calculate means with qualitative scale data (see section 2.5.4.2) *[cross ref.]*.

2.5 Statistical analysis of individual plant data

2.5.1 Introduction

2.5.1.1 When distinctness is assessed, for at least some characteristics, by statistical analysis of individual plant data, uniformity is assessed by standard deviation for relevant characteristics.

2.5.1.2 In order to understand how statistical analysis can be appropriate to trial data it is necessary to answer the following questions:

1. What is a characteristic?
2. What is a process level?
3. What is a scale level of a characteristic?
4. What is the influence of the scale level on the :
    - planning of a trial,
    - recording of data,
    - determination of distinctness and uniformity and
    - description of varieties.

2.5.2 Different levels to look at a characteristic
2.5.2.1 Introduction
2.5.2.1.1 Characteristics can be considered in different levels of process (Table 1). The characteristics as expressed in the trial (type of expression) are considered as process level 1.

The data taken from the trial for the assessment of distinctness, uniformity and stability are defined as process level 2. These data are transformed into states of expression for the purpose of variety description. The variety description is process level 3.

Table 1: Definition of different process levels to consider characteristics

Process level	Description of the process level
1	characteristics as expressed in trial
2	data for evaluation of characteristics
3	variety description

From the statistical point of view, the information level decreases from process level 1 to 3. Statistical analysis is only applied in level 2.

2.5.2.1.2 Sometimes for crop experts it seems that there is no need to distinguish between different process levels. The process level 1, 2 and 3 could be identical. However, in general, this is not the case.

2.5.2.2 Understanding the need for process levels
2.5.2.2.1 The crop expert may know from UPOV Test Guidelines or his own experience that, for example, 'Length of plant' is a good characteristic for the examination of DUS. There are varieties which have longer plants than other varieties. Another characteristic could be 'Variegation of leaf blade'. For some varieties, variegation is present and for others not. The crop expert has now two characteristics and he knows that 'Plant length' is a quantitative characteristic and 'Variegation of leaf blade' is a qualitative characteristic (definitions: see Part I: section 2.5.3.2 to 2.5.3.4 *[cross ref.]* below). This stage of work can be described as **process level 1.**

2.5.2.2.2 The crop expert then has to plan the trial and to decide on the type of observation for the characteristics. For characteristic 'Variegation of leaf blade', the decision is clear. There are two possible expressions: 'present' or 'absent'. The decision for characteristic 'Plant length' is not specific and depends on expected differences between the varieties and on the variation within the varieties. In many cases, the crop expert will decide to measure a number of plants (in cm) and to use special statistical procedures to examine distinctness and uniformity. But it could also be possible to assess the characteristic 'Plant length' visually by using expressions like 'short', 'medium' and 'long', if differences between varieties are large enough (for distinctness) and the variation within varieties is very small or absent in this characteristic. The continuous variation of a characteristic is assigned to appropriate states of expression which are recorded by notes (see document TGP/9, section 4*)[cross ref.]*. The crucial element in this stage of work is the recording of data for further evaluations. It is described as **process level 2**.

2.5.2.2.3 At the end of the DUS test, the crop expert has to establish a description of the varieties using notes from 1 to 9 or parts of them. This phase can be described as **process level 3**. For 'Variegation of leaf blade' the crop expert can take the same states of expression (notes) he recorded in process level 2 and the three process levels appear to be the same. In cases where the crop expert decided to assess 'Plant length' visually, he can take the same states of expression (notes) he recorded in process level 2 and there is no obvious difference between process level 2 and 3. If the characteristic 'Plant length' is measured in cm, it is necessary to assign intervals of measurements to states of expressions like 'short', 'medium' and 'long' to establish a variety description. In this case, for statistical procedures, it is important to be clearly aware of the relevant level and to understand the differences between characteristics as expressed in the trial, data for evaluation of characteristics and the variety description. This is absolutely necessary for choosing the most appropriate statistical procedures in cooperation with statisticians or by the crop expert.

2.5.3 Types of expression of characteristics

2.5.3.1 Characteristics can be classified according to their types of expression. The consideration of the type of expression of characteristics corresponds to process level 1. The following types of expression of characteristics are defined in the General Introduction to the Examination of Distinctness, Uniformity and Stability and the Development of Harmonized Descriptions of New Varieties of Plants, (document TG/1/3, the "General Introduction", Chapter 4.4):

2.5.3.2 Qualitative characteristics" are those that are expressed in discontinuous states (e.g. sex of plant: dioecious female (1), dioecious male (2), monoecious unisexual (3), monoecious hermaphrodite (4)). These states are self-explanatory and independently meaningful. All states are necessary to describe the full range of the characteristic, and every form of expression can be described by a single state. The order of states is not important. As a rule, the characteristics are not influenced by environment.

2.5.3.3 "Quantitative characteristics" are those where the expression covers the full range of variation from one extreme to the other. The expression can be recorded on a one-dimensional, continuous or discrete, linear scale. The range of expressions is divided into a number of states for the purpose of description (e.g. length of stem: very short (1), short (3), medium (5), long (7), very long (9)). The division seeks to provide, as far as practical, an even distribution across the scale. The Test Guidelines do not specify the difference needed for distinctness. The states of expression should, however, be meaningful for DUS assessment.

2.5.3.4 In the case of "pseudo-qualitative characteristics" the range of expression is at least partly continuous, but varies in more than one dimension (e.g. shape: ovate (1), elliptic (2), circular (3), obovate (4)) and cannot be adequately described by just defining two ends of a linear range. In a similar way to qualitative (discontinuous) characteristics – hence the term "pseudo-qualitative" – each individual state of expression needs to be identified to adequately describe the range of the characteristic.

2.5.4 Types of scales of data

The possibility to use specific procedures for the assessment of distinctness, uniformity and stability depends on the scale level of the data which are recorded for a characteristic. The scale level of data depends on the type of expression of the characteristic and on the way of recording this expression. The type of scale may be quantitative or qualitative.

2.5.4.1 Quantitatively scaled data (metric or ordinal scaled data)

2.5.4.1.1 Introduction

2.5.4.1.1.1 Quantitatively scaled data are all data which are recorded by measuring or counting. Weighing is a special form of measuring. Quantitatively scaled data can have a continuous or a discrete distribution. Continuous data result from measurements. They can take every value out of the defined range. Discrete quantitative data result from counting.

Examples

Quantitatively scaled data	Example	Example number
- continuous	Plant length in cm.	1
- discrete	Number of stamens	2

For description of the states of expression, see Table 6.

2.5.4.1.1.2 The continuous quantitatively scaled data for the characteristic “Plant length” are measured on a continuous scale with defined units of assessment. A change of unit of measurement e.g. from cm into mm is only a question of precision and not a change of type of scale.

2.5.4.1.1.3 The discrete quantitatively scaled data of the characteristic “Number of stamens” are assessed by counting (1, 2, 3, 4, and so on). The distances between the neighboring units of assessment are constant and for this example equal to 1. There are no real values between two neighboring units but it is possible to compute an average which falls between those units.

2.5.4.1.1.4 In biometrical terminology, quantitative scales are referred to as metric scales or cardinal scales. Quantitative scales can be subdivided into ratio scales and interval scales.

2.5.4.1.2 Ratio scale

2.5.4.1.2.1 A ratio scale is a quantitative scale with a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Ratio scaled data may be continuous or discrete.

The absolute zero point:

2.5.4.1.2.2 The definition of an absolute zero point makes it possible to define meaningful ratios. This is a requirement for the construction of index numbers (e.g. the ratio of length to width). An index is the combination of at least two characteristics. In the General Introduction, this is referred to as a combined characteristic (see document TG/1/3, section 4.6.3).

2.5.4.1.2.3 It is also possible to calculate ratios between the expression of different varieties. For example, in the characteristic ‘Plant length’ assessed in cm, there is a lower limit for the expression which is ‘0 cm’ (zero). It is possible to calculate the ratio of length of plant of variety ‘A’ to length of plant of variety ‘B’ by division:

[TWC Chairperson: To review if this paragraph is relevant for DUS testing]

Length of plant of variety ‘A’ = 80 cm
Length of plant of variety ‘B’ = 40 cm
Ratio = Length of plant of variety ‘A’ / Length of plant of variety ‘B’
= 80 cm / 40 cm
= 2.

2.5.4.1.2.4 So it is possible in this example to state that plant ‘A’ is double the length of plant ‘B’. The existence of an absolute zero point ensures an unambiguous ratio.

2.5.4.1.2.5 The ratio scale is the highest classification of the scales (Table 2). That means that ratio scaled data include the highest information about the characteristic and it is possible to use many statistical procedures (section 2.5.7 *[cross ref.]*).

2.5.4.1.2.6 The examples 1 and 2 (Table 6) are examples for characteristics with ratio scaled data.

2.5.4.1.3 Interval scale

2.5.4.1.3.1 An Interval scale is a quantitative scale without a defined absolute zero point. There is always a constant non-zero distance between two adjacent expressions. Interval scaled data may be distributed continuously or discretely.

2.5.4.1.3.2 An example for a discrete interval scaled characteristic is 'Time of beginning of flowering' measured as date which is given as example 6 in Table 6. This characteristic is defined as the number of days from April 1. The definition is useful but arbitrary and April 1 is not a natural limit. It would also be possible to define the characteristic as the number of days from January 1.

2.5.4.1.3.3 It is not possible to calculate a meaningful ratio between two varieties which should be illustrated with the following example:

Variety 'A' begins to flower on May 30 and variety 'B' on April 30

Case I) Number of days from April 1 of variety 'A' = 60
Number of days from April 1 of variety 'B' = 30

$$\text{Ratio}_I = \frac{\text{Number of days from April 1 of variety ‘A’}}{\text{Number of days from April 1 of variety ‘B}} = \frac{60 \text{ days}}{30 \text{ days}} = 2$$

Case II) Number of days from January 1 of variety 'A' = 150
Number of days from January 1 of variety 'B' = 120

$$\text{Ratio}_{II} = \frac{\text{Number of days from January 1 of variety ‘A’}}{\text{Number of days from January 1 of variety ‘B}} = \frac{150 \text{ days}}{120 \text{ days}} = 1.25$$

$\text{Ratio}_I = 2 > 1.25 = \text{Ratio}_{II}$

2.5.4.1.3.4 It is impossible to state that the time of flowering of variety 'A' is twice that of variety 'B'. The ratio depends on the choice of the zero point of the scale. This kind of scale is defined as an "Interval scale": a quantitative scale without a defined absolute zero point.

2.5.4.1.3.5 The interval scale is lower classified than the ratio scale (Table 2). Fewer statistical procedures can be used with interval scaled data than with ratio scaled data (see Part I: section 4.5.7 *[cross ref.]*). The interval scale is theoretically the minimum scale level to calculate arithmetic mean values.

2.5.4.2 Qualitatively scaled data

Qualitatively scaled data are data which can be arranged in different discrete qualitative categories. Usually they result from visual assessment. Subgroups of qualitative scales are ordinal and nominal scales.

2.5.4.2.1 Ordinal scale

2.5.4.2.1.1Ordinally scaled data are qualitative data of which discrete categories can be arranged in an ascending or descending order. They result from visually assessed (notes) quantitative characteristics.

Example:

Qualitative data	**Example**	**Example number**
- ordinal	Intensity of anthocyanin	3

For description of the states of expressions, see Table 6.

2.5.4.2.1.2 An ordinal scale consists of numbers which correspond to the states of expression of the characteristic (notes). The expressions vary from one extreme to the other and thus they have a clear logical order. It is not possible to change this order, but it is not important which numbers are used to denote the categories. In some cases ordinal data may reach the level of discrete interval scaled data or of discrete ratio scaled data (section 4.5.6 *[cross ref.]*).

2.5.4.2.1.3 The distances between the discrete categories of an ordinal scale are not exactly known and not necessarily equal. Therefore, an ordinal scale does not fulfil the condition to calculate arithmetic mean values, which is the equality of intervals throughout the scale.

2.5.4.2.1.4 The ordinal scale is lower classified than the interval scale (Table 2). Less statistical procedures can be used for ordinal scale than for each of the higher classified scale data (see Part I: section 4.5.7 *[cross ref.]*).

2.5.4.2.2 Nominal scale

2.5.4.2.2.1 Nominal scaled qualitative data are qualitative data without any logical order of the discrete categories. They result from visually assessed (notes) pseudo-qualitative and qualitative characteristics.

Examples:

Qualitative data	**Example**	**Example number**
- nominal	Sex of plant	4
- nominal with two states	Leaf blade: variegation	5

For description of the states of expressions, see Table 6.

2.5.4.2.2.2 A nominal scale consists of numbers which correspond to the states of expression of the characteristic, which are referred to in the Test Guidelines as notes. Although numbers are used for designation there is no inevitable order for the expressions and so it is possible to arrange them in any order.

2.5.4.2.2.3 Characteristics with only two categories (dichotomous characteristic) are a special form of nominal scales.

2.5.4.2.2.4 The nominal scale is the lowest classification of the scales (Table 2). Few statistical procedures are applicable for evaluations (section 4.5.7 *[cross ref.]*).

2.5.4.2.2.5 The different types of scales are summarized in the following table.

Table 2: Types of scales and scale levels

[TWC Chairperson: To modify the table for consistency with the subsequent paragraphs]

<table>
<tr><th colspan="2">Type of scale</th><th>Description</th><th>Distribution</th><th>Data recording</th><th>Scale Level</th></tr>
<tr><td rowspan="4">quantitative data (measured or counted)</td><td rowspan="2">ratio</td><td rowspan="2">constant distances with absolute zero point</td><td>Continuous</td><td>Absolute Measurements</td><td rowspan="2">High</td></tr>
<tr><td>Discrete</td><td>Counting</td></tr>
<tr><td rowspan="2">interval</td><td rowspan="2">constant distances without absolute zero point</td><td>Continuous</td><td>Relative measurements</td><td rowspan="2">↑</td></tr>
<tr><td>Discrete</td><td>Date</td></tr>
<tr><td>qualitative data (visually observed QN)</td><td>ordinal</td><td>Ordered expressions with varying distances</td><td>Discrete</td><td>Visually assessed notes</td><td>↑</td></tr>
<tr><td>qualitative data (visually observed notes without logic order from PQ or QL)</td><td>nominal</td><td>No order, no distances</td><td>Discrete</td><td>Visually assessed notes</td><td>Low</td></tr>
</table>

2.5.4.2.2.6 From the statistical point of view a characteristic is only considered at the level of data which has been recorded, whether for analysis or for describing the expression of the characteristic. Therefore, characteristics with quantitative data are denoted as quantitative characteristics and characteristics with ordinal and nominal scaled data as qualitative characteristics.

2.5.5 Scale levels for variety description

The description of varieties is based on the states of expression (notes) which are given in the Test Guidelines for the specific crop. In the case of visual assessment, the notes from the Test Guidelines are usually used for recording the characteristic as well as for the assessment of DUS. The notes are distributed on a nominal or ordinal scale (see Part I: section 4.5.4.2 *[cross ref.]*). For measured or counted characteristics, DUS assessment is based on the recorded values and the recorded values are transformed into states of expression only for the purpose of variety description.

2.5.6 Relation between types of expression of characteristics and scale levels of data

2.5.6.1 Records taken for the assessment of qualitative characteristics are distributed on a nominal scale, for example “Sex of plant”, “Leaf blade: variegation” (Table 6, examples 4 and 5).

2.5.6.2 For quantitative characteristics the scale level of data depends on the method of assessment. They can be recorded on a quantitative (when measured) or ordinal (when visually observed) scale. For example, “Length of plant” can be recorded by measurements resulting in ratio scaled continuous quantitative data. However, visual assessment on a 1 to 9 scale may also be appropriate. In this case, the recorded data are qualitatively scaled (ordinal scale) because the size of intervals between the midpoints of categories is not exactly the same.

Remark: In some cases visually assessed data on quantitative characteristics may be handled as measurements. The possibility to apply statistical methods for quantitative data depends on the precision of the assessment and the robustness of the statistical procedures. In the case of very precise visually assessed quantitative characteristics the usually ordinal data may reach the level of discrete interval scaled data or of discrete ratio scaled data.

2.5.6.3 A pseudo-qualitative type of characteristic is one in which the expression varies in more than one dimension. The different dimensions are combined in one scale. At least one dimension is quantitatively expressed. The other dimensions may be qualitatively expressed or quantitatively expressed. The scale as a whole has to be considered as a nominal scale (e.g. “Shape”, “Flower color”; Table 6, examples 7 and 8).

2.5.6.4 In the case of using the off-type procedure for the assessment of uniformity the recorded data are nominally scaled. The records fall into two qualitative classes: plants belonging to the variety (true-types) and plants not belonging to the variety (off-types). The type of scale is the same for qualitative, quantitative and pseudo-qualitative characteristics.

2.5.6.5 The relation between the type of characteristics (process level 1) and the type of scale of data recorded for the assessment of distinctness and uniformity is described in Table 3. A qualitative characteristic is recorded on a nominal scale for distinctness (state of expression) and for uniformity (true-types vs. off-types). Pseudo-qualitative characteristics are recorded on a nominal scale for distinctness (state of expression) and on a nominal scale for uniformity (true-types vs. off-types). Quantitative characteristics are recorded on an ordinal, interval or ratio scale for the assessment of distinctness depending on the characteristic and the method of assessment. If the records are taken from single plants the same data may be used for the assessment of distinctness and uniformity. If distinctness is assessed on the basis of a single record of a group of plants, uniformity has to be judged with the off-type procedure (nominal scale).

Table 3: Relation between type of characteristic and type of scale of assessed data

Procedure	Type of scale (level 2)	Distribution	Type of characteristic (level 1)		
			Quantitative	Pseudo-qualitative	Qualitative
Distinctness	ratio	Continuous	✔		
		Discrete	✔		
	interval	Continuous	✔		
		Discrete	✔		
	ordinal	Discrete	✔		
	~~combined~~	~~Discrete~~		~~✔~~	
	nominal	Discrete		✔	✔
Uniformity	ratio	Continuous	✔		
		Discrete	✔		
	interval	Continuous	✔		
		Discrete	✔		
	ordinal	Discrete	✔		
	~~combined~~	~~Discrete~~	~~✔~~		
	nominal	Discrete	✔	✔	✔

2.5.7 Relation between method of observation of characteristics, scale levels of data and recommended statistical procedures
[TWC Chairperson: To update these paragraphs in accordance with any changes to documents TGP/7 and TGP/9]

2.5.7.1 Established statistical procedures can be used for the assessment of distinctness and uniformity considering the scale level and some further conditions such as the degree of freedom or unimodality (Tables 4 and 5).

2.5.7.2 The relation between the expression of characteristics and the scale levels of data for the assessment of distinctness and uniformity is summarized in Table 6.

Table 4: Statistical procedures for the assessment of distinctness

Type of scale	Distribu-tion	Observa-tion method	Procedure[1)] and further Conditions	Reference document
ratio	continuous	MS MG (VS) [1)]	COYD Normal distribution, df >=20 long term LSD Normal distribution, df<20 2 out of 3 methods (LSD 1%) Normal distribution, df>=20	TGP/9
	discrete			
interval	continuous			
	discrete			
ordinal	discrete	VG VS	See explanation for QN characteristics in TGP/9 sections 5.2.2 and 5.2.3, See explanation for QN characteristics in TGP/9 section 5.2.4	TGP/9 TWC/14/12
Combina-tion of ordinal or ordinal and nominal scales	discrete	VG (VS) [32]	See explanation for PQ characteristics in TGP/9 sections 5.2.2 and 5.2.3	TGP/9
nominal	discrete	VG (VS) [2)]	See explanation for QL characteristics in TGP/9 sections 5.2.2 and 5.2.3	TGP/9

1) see remark in section 4.5.6.2 *[cross ref.]*
2) normally VG but VS would be possible

Table 5: Statistical procedures for the assessment of uniformity

Type of scale	Distribu-tion	Observa-tion method	Procedure[1)] and Further Conditions	Reference document
ratio	continuous	MS	COYU Normal distribution 2 out of 3 method (s^2_c<=1.6s^2_s)) Normal distribution LSD for untransformed percentage of off-types	TGP/10
	discrete	MS		
interval	continuous	VS		
	discrete			
ordinal	discrete	VS	threshold model	TWC/14/12
Combina-tion of ordinal or ordinal and nominal scales	discrete		There is no case where uniformity is assessed on combined scaled data	
nominal	discrete	VS	off-type procedure for dichotomous (binary) data	TGP/10

Table 6: Relation between expression of characteristics and scale levels of data for the assessment of distinctness and uniformity

Example	Name of characteristic	Distinctness			Uniformity		
		Unit of assess-ment	Description (states of expression)	Type of scale	Unit of assess-ment	Description (states of expression)	Type of scale
1	Length of plant	cm	assessment in cm without digits after decimal point	ratio scaled continuous quantitative data	cm	assessment in cm without digits after decimal point	ratio scaled continuous quantitative data
					True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
2	Number of stamens	counts	1, 2, 3, ... , 40,41, ...	ratio scaled discrete quantitative data	counts	1, 2, 3, ... , 40,41, ...	ratio scaled discrete quantitative data
3	Intensity of anthocyanin	1 2 3 4 5 6 7 8 9	very low very low to low low low to medium medium medium to high high high to very high very high	ordinally scaled qualitative data (with an underlying quantitative variable)	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data
4	Sex of plant	1 2 3 4	dioecious female dioecious male monoecious unisexual monoecious hermaphrodite	nominally scaled qualitative data	True-type Off-type	Number of plants belonging to the variety Number of off-types	nominally scaled qualitative data

<table>
<tr><th rowspan="2">Example</th><th rowspan="2">Name of characteristic</th><th colspan="3">Distinctness</th><th colspan="3">Uniformity</th></tr>
<tr><th>Unit of assess-ment</th><th>Description (states of expression)</th><th>Type of scale</th><th>Unit of assess-ment</th><th>Description (states of expression)</th><th>Type of scale</th></tr>
<tr><td>5</td><td>Leaf blade: variegation</td><td>1
9</td><td>absent
present</td><td>nominally scaled qualitative data</td><td>True-type

Off-type</td><td>Number of plants belonging to the variety
Number of off-types</td><td>nominally scaled qualitative data</td></tr>
<tr><td rowspan="2">6</td><td rowspan="2">Time of beginning of flowering</td><td rowspan="2">date</td><td rowspan="2">e.g. May 21, 51st day from April 1</td><td rowspan="2">interval scaled discrete quantitative data</td><td>Date</td><td>e.g. May 21, 51st day from April 1</td><td>interval scaled discrete quantitative data</td></tr>
<tr><td>True-type

Off-type</td><td>Number of plants belonging to the variety
Number of off-types</td><td>nominally scaled qualitative data</td></tr>
<tr><td>7</td><td>Shape</td><td>1
2
3
4
5
6
7</td><td>deltate
ovate
elliptic
obovate
obdeltate
circular
oblate</td><td>combination of ordinal and nominal scaled discrete qualitative data</td><td>True-type

Off-type</td><td>Number of plants belonging to the variety
Number of off-types</td><td>nominally scaled qualitative data</td></tr>
<tr><td>8</td><td>Flower color</td><td>1
2
3
4
5
6
7
8
9
10</td><td>dark red
medium red
light red
white
light blue
medium blue
dark blue
red violet
violet
blue violet</td><td>combination of ordinal and nominal scaled discrete qualitative data</td><td>True-type

Off-type</td><td>Number of plants belonging to the variety
Number of off-types</td><td>nominally scaled qualitative data</td></tr>
</table>

	<u>Comments of the TWPs in 2011</u>	
Table 6	In the column "Type of Scale", to refer to the type of scale and distribution as in Table 4.	TWC

[Annex II follows]

TGP/8 PART I: DUS TRIAL DESIGN AND DATA ANALYSIS

New Section 3 - Control of variation due to different observers (Drafter: Mr. Gerie van der Heijden (Netherlands))

Notes

1. The TWC at its twenty-fifth session, held in Sibiu, Romania, from September 3 to 6, 2007, agreed that this section be developed on the basis of sections I and II of document TWC/25/12.

2. The TWC at its twenty-sixth session agreed that Mr. Gerie van der Heijden (Netherlands) will consult his Naktuinbouw colleagues in the Netherlands to see if they could contribute a draft for this section.

3. The TWV at its forty-second session, held in Cracow, Poland, from June 23 to 27, 2008, noted that it had encouraged the development of that section and agreed that it should provide suitable text for aspects which were not adequately covered in document TWC/25/12.

	Comments of the TWPs in 2011	
General	The TWA noted the information provided in Annex II and recommended to replace the title of the first heading “Control of variation due to different observers” by “Minimizing the variation due to different observers” and to delete “and this procedure should preferably be described in ISO Guidelines” at the end of the paragraph on “Training”.	TWA
	The TWC agreed with the comments made by the TWA at its fortieth session .and agreed that Mr. Gerie van der Heijden (Netherlands) and Mr. Adrian Roberts (United Kingdom) should prepare a new document taking into account the information contained in document TWC/25/12 Rev. “Review of Test Design: Checking Levels of Quality (Revised)”.	TWC
	The TWV agreed that the information provided in Annex II, provided valuable information that should be included in document TGP/8. With regard to the proposal of the TWC that a new version of that guidance should be prepared taking into account the information contained in document TWC/25/12 Rev. “Review of Test Design: Checking Levels of Quality (Revised)”, it concluded that the volume of information provided in document TWC/25/12 Rev. would detract from the main purpose of the document and suggested that a cross-reference might be made to such information.	TWV

	The TWC agreed with the comments made by the TWA at its fortieth session .and agreed that Mr. Gerie van der Heijden (Netherlands) and Mr. Adrian Roberts (United Kingdom) should prepare a new document taking into account the information contained in document TWC/25/12 Rev. "Review of Test Design: Checking Levels of Quality (Revised)".	TWC
	The TWF considered information in Annex II, and agreed that it provided valuable information that should be included in document TGP/8, however it did not come to an agreement on how the section "Testing the calibration" should be handled. It concluded that a revision should go ahead in order to make it less prescriptive.	TWF

[DRAFT TEXT FOLLOWS]

Control of variation due to different observers

Variation in measurements or observations can be caused by many different factors, like the type of crop, type of characteristic, year, location, trial design and management, method and observer. Especially for visually assessed characteristics (QN/VG or QN/VS) differences between observers can be the reason for large variation and potential bias in the observations. An observer might be less well trained, or have a different interpretation of the characteristic. So, if observer A measures variety 1 and observer B variety 2, the difference measured might be caused by differences between observers A and B instead of differences between varieties 1 and 2. Clearly, our main interest lies with the differences between varieties and not with the differences between the observers. It is important to realize that the variation caused by different observers cannot be eliminated, but there are ways to control it.

Training

UPOV test guidelines try to harmonize the variety description process and describe as clearly as possible the characteristics of a crop and the states of expression. This is the first step in controlling variation and bias. However, the way a characteristic is observed or measured may vary per location or testing authority. Calibration manuals made by the local testing authority are very useful for the local implementation of the UPOV test guideline. Where needed these crop-specific manuals explain the characteristics to be observed in more detail, and specify when and how they should be observed. Furthermore they may contain pictures and drawings for each characteristic, often for every state of expression of a characteristic. The calibration manual can be used by new inexperienced observers but are also useful for more experienced or substitute observers, as a way to recalibrate themselves.
Training of new observers is essential for consistency and continuity of databases, both by means of calibration manuals, and by supervision and guidance of experienced observers. This should be done on a regular basis and this procedure should preferably be described in ISO guidelines.

	Comments of the TWPs in 2011	
Training	The TWA noted the information provided in Annex II and agreed that example varieties illustrating the range of expressions can also be a useful element in the training of experts (see paragraph 2 (Training)).	TWA

Testing the calibration

After training an observer, the next step is to test the performance of the observers in a calibration experiment. This is especially true for inexperienced observers who have to make visual observations (QN/VG characteristics). They should preferably pass a calibration test prior to making observations in the trial. But also for experienced observers, it is important to test themselves on a regular basis to verify if they still fulfil the calibration criteria.

A calibration experiment can be set up and analyzed in different ways. Generally it involves multiple observers, measuring the same set of material and assessing differences between the observers.

For observations made by measurement tools, like rulers (often QN/MS characteristics), the measurement is often made on an interval or ratio scale. In this case, the approach of Bland and Altman (1986) can be used. This approach starts with a plot of the scores of every pair of observers in a scatter plot, and compare it with the line y=x. This helps the eye gauging the degree of agreement between measurements. In a next step, the difference per object is taken and a plot is constructed with on the y-axis the difference between the observers and on the x-axis either the index of the object, or the mean value of the object. By further drawing the horizontal lines y=0, y=mean(dif) and the two lines y = mean(dif)+/- 2 x standard deviations, the bias between the observers and any outliers can easily be spotted. Test methods like the paired t-test can be applied to test for a significant deviation of the observer from another observer or from the mean of the other observers. By taking repeated measurements of the same object, one can use a more advanced test involving variance components. However, in many cases of QN/MS, a good and clear instruction usually suffices and variation or bias in measurements between observers is often negligible. If there is reason for doubt, a calibration experiment as described above can help in providing insight in the situation.

For the analysis of ordinal data (QN/VS or QN/VG characteristics), the construction of contingency tables between each pair of observers for the different scores is very instructive.

A test for a structural difference (bias) between two observers can be obtained by using the Wilcoxon Matched-Pairs test (often called Wilcoxon Signed-Ranks test).

To measure the degree of agreement the Cohen's Kappa statistic (Cohen, 1960) is often used. The statistic tries to accounts for random agreement:
κ = P(agreement) – P(e) / (1-P(e)), where P(agreement) is the fraction of objects which are in the same class for both observers (the main diagonal in the contingency table), and P(e) is the probability of random agreement, given the marginals (like in a Chi-square test).
If the observers are in complete agreement the Kappa value $\kappa = 1$. If there is no agreement among the observers, other than what would be expected by chance (P(e)), then $\kappa = 0$.

The standard Cohen's Kappa statistic only considers perfect agreement versus non-agreement. If one wants to take the degree of disagreement into account (for example with ordinal characteristics), one can apply a linear or quadratic weighted Kappa (Cohen, 1968).
If we want to have a single statistic for all observers simultaneously, a generalized Kappa coefficient can be calculated. Most statistical packages, including SPSS, Genstat and R (package Concord), provide tools to calculate the Kappa statistic.

As noted, a low κ-value indicates poor agreement and values close to 1 indicate excellent agreement. Often scores between 0.6-0.8 are considered to indicate substantial agreement, and above 0.8 to indicate almost perfect agreement. If needed, z-scores for kappa (assuming an approximately normal distribution) are available.

	Comments of the TWPs in 2011	
General	The TWO agreed that the Section "Testing the calibration" should be revised to reflect the likelihood that inexperienced observers would not be entrusted to make VG observations, whilst inexperienced observers might be entrusted to make MG and MS observations. The TWO agreed that the guidance on different types of training and calibration for DUS experts and for staff that would undertake specified measurements should be reflected in the document.	TWO

Trial design
If we have multiple observers in a trial, the best approach is to have one person observe one or more complete replications. In that case, the correction for block effects also accounts for the bias between observers. If more than one observer per replication is needed, extra attention should be given to calibration and agreement. In some cases, the use of incomplete block designs (like alpha designs) might be helpful, and an observer can be assigned to the sub blocks. In this way we can correct for the systematic difference between observers.

Example of Cohen's Kappa
In this example, there are three observers and 30 objects (plots or varieties).
The characteristic is observed on a scale of 1 to 6.
The raw data and their tabulated scores are given in the following tables.

Variety	Observer 1	Observer 2	Observer 3
V1	1	1	1
V2	2	1	2
V3	2	2	2
V4	2	1	2
V5	2	1	2
V6	2	1	2
V7	2	2	2
V8	2	1	2
V9	2	1	2
V10	3	1	3
V11	3	1	3
V12	3	2	2
V13	4	5	4
V14	2	1	1
V15	2	1	2
V16	2	2	3
V17	5	4	5
V18	2	2	3
V19	1	1	1
V20	2	2	2
V21	2	1	2
V22	1	1	1
V23	6	3	6
V24	5	6	6
V25	2	1	2
V26	6	6	6
V27	2	6	2
V28	5	6	5
V29	6	6	5
V30	4	4	4

Scores for variety	1	2	3	4	5	6
V1	3	0	0	0	0	0
V2	1	2	0	0	0	0
V3	0	3	0	0	0	0
V4	1	2	0	0	0	0
V5	1	2	0	0	0	0
V6	1	2	0	0	0	0
V7	0	3	0	0	0	0
V8	1	2	0	0	0	0
V9	1	2	0	0	0	0
V10	1	0	2	0	0	0
V11	1	0	2	0	0	0
V12	0	2	1	0	0	0
V13	0	0	0	2	1	0
V14	2	1	0	0	0	0
V15	1	2	0	0	0	0
V16	0	2	1	0	0	0
V17	0	0	0	1	2	0
V18	0	2	1	0	0	0
V19	3	0	0	0	0	0
V20	0	3	0	0	0	0
V21	1	2	0	0	0	0
V22	3	0	0	0	0	0
V23	0	0	1	0	0	2
V24	0	0	0	0	1	2
V25	1	2	0	0	0	0
V26	0	0	0	0	0	3
V27	0	2	0	0	0	1
V28	0	0	0	0	2	1
V29	0	0	0	0	1	2
V30	0	0	0	3	0	0

The contingency table for observer 1 and 2 is:

O1\O2	1	2	3	4	5	6	Total
1	3	0	0	0	0	0	3
2	10	5	0	1	0	1	17
3	2	1	0	0	0	0	3
4	0	0	0	1	0	0	1
5	0	0	0	1	0	2	3
6	0	0	1	0	0	2	3
Total	15	6	1	3	0	5	30

The Kappa coefficient between observer 1 and 2, $\kappa(O1,O2)$ is calculated as follows:
$\kappa(O1,O2)$ = P(agreement between O1 and O2) – P(e) / (1 – P(e)) where
P(agreement) = (3+5+0+1+0+2)/30 = 11/30 ≈ 0.3667 (diagonal elements)
P(e) = (3/30).(15/30) + (17/30).(6/30) + (3/30).(1/30) + (1/30).(3/30) + (3/30).(0/30) + (3/30).(5/30) ≈ 0.1867. (pair-wise margins)
So $\kappa(O1,O2) \approx (0.3667-0.1867) / (1-0.1867) \approx 0.22$
This is a low value, indicating very poor agreement between these two observers. There is reason for concern and action should be taken to improve the agreement.

Similarly the values for the other pairs can be calculated: $\kappa(O1,O3) \approx 0.72$, $\kappa(O2,O3) \approx 0.22$. Observer 1 and 3 are in good agreement. Observer 2 is clearly different from 1 and 3 and probably needs additional training.

References

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20:** 37-46.
Cohen, J. (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, **70**(4): 213-220.
Bland, J. M. Altman D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet:* 307–310.

[Annex III follows]

ANNEX III

TGP/8 PART I: DUS TRIAL DESIGN AND DATA ANALYSIS

New Section 6 – Data processing for the assessment of distinctness and for producing variety descriptions (Drafters: experts from Finland, France, Germany, Japan, Kenya and the United Kingdom)

Notes

1. The TWC at its twenty-sixth session agreed that the information provided in documents TWC/26/15 and TWC/26/23, presented by Mr. Vincent Gensollen (France) and Mr. Uwe Meyer (Germany), respectively, and an oral presentation by Ms. Mariko Ishino (Japan) included in document TWC/26/15 Add. provided valuable guidance on data processing for the assessment of distinctness and for producing variety descriptions and noted that UPOV did not have guidance on that matter in the TGP documents. It agreed that a new section should be created in document TGP/8/1, Part I as "Data processing for the assessment of distinctness and for producing variety descriptions for producing variety descriptions" and that the methods used by France, Germany and Japan should be included in a new section in document TGP/8/1, Part II as "Methods for data processing for the assessment of distinctness and for producing variety descriptions.

2. The TWC at its twenty-seventh session agreed that experts from France, Germany, Italy, Japan, Kenya and United Kingdom to provide a short description of the principles underlying the detailed methods provided in Part II.

3. Mrs. Sally Watson (United Kingdom) to provide an example for Section 7.1

[DRAFT TEXT FOLLOWS]

6. DATA PROCESSING FOR THE ASSESSMENT OF DISTINCTNESS AND FOR PRODUCING VARIETY DESCRIPTIONS

See PART II, New Section 13

	Comments of the TWPs in 2011	
General	The TWA considered Annex III and recommended to combine new Section 6 "Data processing for the assessment of distinctness and for producing variety descriptions" (Annex III to document TWA/40/14) with new Section 13 "Methods for data processing for the assessment of distinctness and for producing variety descriptions" (Annex VIII to document TWA/40/14) and new Section "Guidance for the development of variety descriptions" (Annex XI to document TWA/40/14).	TWA

	The TWC recalled that the objective of this New Section 6 on PART I of TGP/8 was to present the principles for producing variety descriptions whilst New Section 13 was aimed to present the methods. The TWC expressed a preference to develop this section for Part I in TGP/8, describing the principles for producing variety descriptions, whereas New Section 13 "Methods for data processing for the assessment of distinctness and for producing variety descriptions" should reflect the methods to be included in Part II of TGP/8.	TWC
	The TWV considered Annex III in conjunction with Annex VIII of that document. It agreed that the information provided in Annex VIII was a very important first step in developing common guidance on data processing for the assessment of distinctness and for producing variety descriptions, but concluded that the information as presented in Annex VIII would not be appropriate for inclusion in document TGP/8. It agreed to propose that the Office of the Union be requested to summarize the different approaches set out in Annex VIII with regard to aspects in common and aspects where there was divergence. As a next step, on the basis of that summary, consideration could be given to developing general guidance.	TWV
	The TWO noted the information provided in Annex III in conjunction with Annex VIII of that document. It agreed that the section should include an example of an ornamental variety, with consideration of the number of notes for a quantitative characteristic.	TWO
	The TWF considered Annex III in conjunction with Annex VIII of that document. It agreed that the information provided in Annex VIII was a very important first step in developing common guidance on data processing for the assessment of distinctness and for producing variety descriptions, but concluded that the information as presented in Annex VIII would not be appropriate for inclusion in document TGP/8. It agreed to propose that the Office of the Union be requested to summarize the different approaches set out in Annex VIII with regard to aspects in common and aspects where there was divergence. As a next step, on the basis of that summary, consideration could be given to developing general guidance. The TWF agreed that the section should include an example of a fruit variety, with consideration of the number of notes for a quantitative characteristic.	TWF

[Annex IV follows]

TGP/8 PART I: DUS TRIAL DESIGN AND DATA ANALYSIS

New Section – Information of good agronomic practices for DUS field trials (Drafters: Mrs. Anne Weitz (European Union) and Argentina and France to contribute))

Comments: proposed by the TC at its forty-fifth session

[DRAFT TEXT FOLLOWS]

INFORMATION OF GOOD AGRICULTURAL PRACTICES FOR DUS FIELD TRIALS

	Comments of the TWPs in 2011	
General	The TWA considered Annex IV and recommended not to retain this section.	TWA
	The TWC took note of the information contained in Annex IV and the recommendation made by the TWA at its fortieth session.	TWC
	TWV agreed on the importance of employing good agronomic practice in the conduct of DUS trials and on the need to ensure that staff had the appropriate training and experience for conducting DUS trials. However, it concluded that it would not be desirable to seek to develop detailed guidance in document TGP/8.	TWV
	The TWO noted the importance of employing good agronomic practice in the conduct of DUS trials and on the need to ensure that staff had the appropriate training and experience for conducting DUS trials. However, it agreed that it would not be desirable to seek to develop detailed guidance in document TGP/8.	TWO
	The TWF agreed on the importance of employing good agronomic practice in the conduct of DUS trials and on the need to ensure that staff had the appropriate training and experience for conducting DUS trials. However, it concluded that it would not be desirable to seek to develop detailed guidance in document TGP/8.	TWF

1. Introduction

According to the Food and Agricultural Organization of the United Nations (FAO), Good Agricultural Practices are “practices that address environmental, economic and social sustainability for on-farm processes, and result in safe and quality food and non-food agricultural products” (Source: FAO COAG, 2003, GAP paper: http://www.fao.org/docrep/meeting/006/y8704e.htm)

The notion of Good Agricultural Practices (GAP) covers a rather wide field of activities. It forms the basis for a number of regulations, standards and codes applied by authorities, producers, retailers, consumers, NGOs, quality controllers etc.

For the purpose of this document, the information of good agricultural practices shall consider only field trials set up in the framework of DUS tests.

1.1 General Introduction

The General Introduction states that the DUS examination is based mainly on growing tests. The examination generates a description of the variety, using its relevant characteristics, by which it can be defined as a variety in terms of the Convention. Furthermore, the characteristics generated form the basis for the assessment of D, U and S. Taking this into consideration; it becomes obvious that the satisfactory development of the plants in the growing trial according to their genotypic predisposition is a prerequisite for the proper assessment of the characteristics which describe the variety.

1.2 Conditions for conducting the examination

Document TGP/7 Development of Test Guidelines explains that “the test should be carried out under conditions ensuring satisfactory growth for the expression of the relevant characteristics of the variety and for the conduct of the examination”.

Specific and detailed practical guidance on certain aspects is given in the relevant Test Guidelines; however, general aspects that apply across all Test Guidelines are not covered there. This annex aims to address the aspects for a satisfactory growth of the varieties in a DUS test based on field trials under specific consideration of factors which are influenced by good (or bad) agricultural practices.

1.3 Factors which impact on the expression of characteristics

There are many factors which can affect the expression of the characteristics of a variety, e.g. pests, diseases, chemical treatment or the origin of the plant material examined like tissue culture or different rootstocks. The General Introduction states that, where the factor is not intended for DUS examination, it is important that its influence does not distort the DUS examination. Accordingly, the testing authority should ensure either that the varieties under test are all free of such factors or that all varieties included into the DUS test are subject to the same factor and that it has an equal effect on all varieties.

2. GAP and elements which impact on the true expression of characteristics

The following section will explain the guiding principles of good agricultural practices for the most important elements which play an important role on the satisfactory growth of varieties in terms of their impact on the true expression of the characteristic of the plant genotype. It is important to note that the below mentioned measures should be carried out in a way that their impact on the plants in the field trial is uniform and equal.

The examiner should keep records of all cultivation measures. These records should allow providing evidence in case an incident of growth conditions happens.

The below mentioned elements enumerate the most important considerations for DUS testing in a field trial:

2.1 Soil

As a principle, any field work has to be carried out under consideration of the conditions of the local weather and the location in question.

The natural soil fertility and biological activity shall be increased or maintained by proper rotation. The species involved into the rotation should be chosen in respect of the DUS trial which will follow. E. g. in order to avoid misinterpretation on uniformity, the pre-crop species should not be the same or similar to the next DUS trial.

The humus content should be increased by regular supply of organic substance or by reduction of field work such as tillering, if possible.

Measures to avoid erosion by wind or water are maintaining or planting hedges.

The area of the field which is chosen for a DUS trial should provide to a maximum for a uniform soil.

2.2 Water

Rainfall should be monitored, the amount of irrigation should be such that it provides for proper growth (if that facility exists).

The water used for irrigation should be analyzed on its quality in order to avoid damage to plants, dissipation should be avoided.

2.3 Fertilization

General principles according to the relevant applicable law ruling fertilization (if applicable) shall be applied.

Regular analyses of the composition of the soil with regard to the nutritional elements and to its humus content should guide the supply of fertilizers.

Supply from previous crops in the rotation should be considered.

2.4 Pest and disease management

General principles according to the relevant applicable law ruling pest management shall be applied.

Pests or diseases which allow a treatment by integrated pest management should be privileged to chemical treatment in the framework of sustainable development and to avoid side effects of pesticides.

2.5 Qualification of the personnel

The personnel who performs the agricultural work should have at least basic knowledge of the principles of DUS testing.

2.6 Technical equipment

Sowing or planting equipment should be cleaned in a way that technical mixture of varieties is avoided.

A regular maintenance of machines should be respected.

It shall be cleaned regularly so that infections are avoided.

2.7 Wildlife protection

When protecting trials against damage by wildlife (e.g. nets for birds or fences for hares or deer), this protection should hamper the animals to destroy the trial, it should not kill them.

Good agricultural practices should be a guiding principle for outdoor grown DUS trials in order to provide for satisfactory growth and to minimize any undesirable effect on the true expression of the genotype of a variety.

[Annex V follows]

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

	Comments of the TWPs in 2011	
General	The TWC received the presentation by Mr. Adrian Roberts (United Kingdom) on “An Adjustment to the COYD Method When Varieties are Grouped Within the DUS Trial” (see document TWC/29/25). The TWC agreed that the text should be included in TGP/8 Part II Section 3	TWC

New Section after Section COYU Statistical Methods for Very Small Sample Sizes (Drafter: Mr. Gerie van der Heijden (Netherlands))

Notes

13. The TC at its forty-fourth session held in Geneva from April 7 to 9, 2008, agreed to invite the Technical Working Parties to consider including statistical methods for very small sample sizes, subject to suitable methods which are in use by members of the Union being provided.

[DRAFT TEXT FOLLOWS]

3.5 STATISTICAL METHODS FOR VERY SMALL SAMPLE SIZES

	Comments of the TWPs in 2011	
General	The TWC agreed that it would be useful to extend the draft from a perspective of presenting possible solutions to the different situations presented, although this might be difficult from a statistical point of view. Input from other Technical Working Parties was welcomed.	TWC
	The TWV and the TWF agreed that it was important to emphasize that, “if data are to be statistically analyzed, then the assumptions behind the theory on which the statistical methods are based must be met - at least approximately” (see document TGP/8/1: Part I: 2. VALIDATION OF DATA AND ASSUMPTIONS, Section 2.3 “Assumptions for statistical analysis and the validation of these assumptions”). The TWV and the TWF agreed that the wording should be amended for consistency with the wording in document TGP/8/1: Part I: 1. DUS TRIAL DESIGN: “1.5.3.3.6.2.6 The test statistic is based on a sample of plants, trialled in a sample of growing conditions. Thus if the process were to be repeated at a different time, a different value of the test statistic would be obtained.	TWV TWF

	Because of this inherent variability, there is a chance that a different conclusion is arrived at compared to the conclusion which would be reached if the trial could be repeated indefinitely. Such “statistical errors” can occur in two ways, let us first consider distinctness conclusions: “- The conclusions based on the test statistic, i.e. from the DUS trial, is that two varieties are distinct, when they would not be distinct if the trial could be repeated indefinitely. This is known as a Type I error and its risk is denoted by α. […]”.	
	The TWO agreed that realistic examples should be included in the document, based on actual cases. If no such cases could be provided, the section should be deleted. The TWO noted that the United Kingdom would try to provide an example.	TWO

One of the main problems when applying a statistical test on small trials is that we do not have enough data available to limit the risk of making a wrong decision to an acceptable level. Every statistical test has a probability/risk of making wrong decisions: there is a Type I error, i.e. the risk of declaring two varieties different where in reality they are not significantly different, and a Type II error: declaring two distinct varieties not significantly different.

	<u>Comments of the TWPs in 2011</u>	
1.	The TWA, the TWV and the TWF considered Annex V and recommended to amend, in the first paragraph, “two varieties different” as “two varieties distinct” as follows: “One of the main problems when applying a statistical test on small trials is that we do not have enough data available to limit the risk of making a wrong decision to an acceptable level. Every statistical test has a probability/risk of making wrong decisions: there is a Type I error, i.e. the risk of declaring ~~two varieties different~~ <u>two varieties distinct</u> where in reality they are not significantly different, and a Type II error: declaring two distinct varieties not significantly different.”	TWA TWV TWF

In general we control the Type I error by fixing the significance level (α). However, especially with small trials, a low risk of Type I (low α) considerably increases the Type II error, or alternatively stated, such a test has a considerable lack of discriminating power. Another problem with small sample sizes is that we do not have enough data to test our assumptions.

From a statistical point of view it is possible to statistically compare the mean of a candidate variety after a single measurement on a single plant in a single year with a set of reference varieties, if at least several reference varieties are being measured in the same year as well as in one or more other years. For this, one could use any statistical package capable of analysing unbalanced two-way designs with the factors years and varieties. This analysis can be seen as an extension of the long-term LSD but is not standard UPOV practice. The test is based on the usual assumptions, which can however not be tested with such a small dataset. If we are willing to accept assumptions like normality, homogeneity of variance and additivity, e.g. from previous knowledge, the test is in principal valid, although lack of power is still a problem.

In general, small sample size may refer to different aspects of the variety trial:
(a) limited number of plants/measurements in a plot
(b) limited number of replications
(c) limited number of varieties
(d) limited number of years
or any combination of these aspects.

Ad (a). For any experiment, sound experimental design principles should be kept in mind at all times. With regard to the number of plants per plot, it is bad practise to use so few plants in a plot that measured plants are considerably influenced by their neighbours. A plant of a small variety next to a plant of a tall variety may lead to both plants having a more extreme expression than under the condition of neighbouring plants of similar height. This interaction effect hampers unbiased comparisons. To overcome this neighbouring effect, one often uses border plants. Alternatively one can group varieties in different height classes such that these effects are minimised within the groups. Also refer to TGP8 part 1, section 1.6.3.7 for further details.

Ad (b). The number of replications in a trial is often at least 2. Strictly speaking, for the COYD or long-term LSD we only use the variety means of the year for the analysis, so from a theoretical point of view a single replication per variety per year is sufficient. Of course having no replications within a year may lead to a significant increase of the uncertainty of the estimate of the variety mean and it limits the testing of assumptions for the analysis.

Ad (c). With regard to the number of varieties in the test, from a theoretical point as few as three or four varieties are sufficient if two or three years of data are used. However, in most cases, experience has taught us that such small experiments with just a few degrees of freedom are not really useful, as the discriminating power of the test is too low. A low power may be less of a problem, if we have just a few varieties and large and consistent differences between them.

Ad (d). Theoretically spoken, it is possible to make a decision based on a single year's observation of a candidate variety, when reference varieties are also observed and data from the reference varieties over several years are available. Several assumptions need to be made and these assumptions can not be tested. An important assumption is that the candidate variety to be tested does not exhibit a strong interaction from year to year with close reference varieties for the characteristic under study. However, the most important drawback is that the power of the test is very limited, i.e. the chance that a truly significant difference between a pair of varieties will indeed be declared significant in the analysis is very small. In that case, the conclusion would be that the two varieties are not sufficiently different to obtain a significant result given the small sample size. If this information is sufficient for rejection of the candidate variety is an open question, but probably not.

Historical data can be used to gain insight in the lack of power of the experiment, i.e. the risk of accidently rejecting a distinguishable variety. One can also use these data to get an impression of the best way to improve the experimental design.

The power of the test can be increased in several ways. If a reference variety is not tested in the same years as the candidate variety, the standard error of this difference is rather large. By putting the varieties in the same trial in the next year, the standard error for this difference can be reduced considerably.

Another way to increase the power of the test is by increasing the number of degrees of freedom for the residual term. This can be done by using more data from previous years, which is exactly what is done in the long-term LSD.

Note that small trials are troublesome for distinctness testing, but even more so for uniformity testing. The COYU requires a considerable number of plants per plot for a reasonable estimate of the standard deviation.
Another problem when we use small and unbalanced designs is that some variety differences are tested with greater power than others. The comparison of candidate varieties with reference varieties which are less frequent (or even absent) in the years of testing of the candidates will have a much larger standard error of difference. This might lead to rejecting a candidate which can not be declared sufficiently distinct, but which is due to bad luck since it is close to a reference not in the collection of reference varieties on the field. The procedure is in itself statistically valid and sound, but might be unwanted from a fair policy point of view.

	<u>Comments of the TWPs in 2011</u>	
	The TWA recommended to redraft the last paragraph of the document in such a way that a variety cannot be rejected on the basis that a similar variety is not available in the field in the reference collection.	TWA

[Annex VI follows]

ANNEX VI

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

New Section 11 Examining DUS in bulk samples (Drafter: Mr. Kristian Kristensen (Denmark))

Notes

The TC at its forty-fourth session held in Geneva from April 7 to 9, 2008, requested that for each statistical method an explanation of the requirements for its application and the situations where it would be appropriate to apply the method be included.

[DRAFT TEXT FOLLOWS]

11. EXAMINING DUS IN BULK SAMPLES

	Comments of the TWPs in 2011	
General	The TWA considered Annex VI and noted the new subsections 11.1 and 11.2 for "Examining DUS in bulk samples".	TWA
	The TWC agreed that the paper could be included in TGP/8 but that the content of sections 11.3 "Distinctness" and 11.4 "Uniformity" should be excluded from the main text and presented in an appendix.	TWC

11.1 Introduction

11.1.1 The term "bulk sampling" is here used for the process of merging some or all individual plants before recording the characteristics. Bulking is usually only applied where the measurement of the character is very expensive or very difficult to obtain for each individual plant. Some examples are: the content of potassium that is used for distinctness purpose for sugar beet varieties may be based on bulked samples because it would be very expensive to prepare samples and analyse the content of potassium for each individual plant. Likewise the weight of seeds that is used in field peas and field beans are often done after bulking the seeds from several plants.

	Comments of the TWPs in 2011	
11.1.1	The TWV and the TWO agreed that the example of sugar beet should be replaced by a crop for which there are UPOV Test Guidelines.	TWV TWO
	The TWF agreed that the example of sugar beet should be replaced by a crop for which there are UPOV Test Guidelines and to include an example of the fruit crops if it is available.	TWF

11.1.2 There are different degrees of bulking ranging from: 1) merging of pairs of plants, 2) merging 3 or 4 up to all plants within a plot and 3) merging all plants for each the variety. The degree of bulking may play an important role for the efficiency of the tests and may exclude even exclude some tests.

11.2 Consequences of bulking for DUS examination

The consequences of bulking will be more serious when testing for uniformity than when testing for distinctness.

11.2.1 Testing for Uniformity

11.2.1.1 If the test for uniformity is based on the number of off-types any bulking may completely mask the off-types as now only the mean the characteristic over the bulked plants can be evaluated.

11.2.1.2 For many continuous variables uniformity are tested using the COYU method which is based on logarithm of the standard deviation of individual plants within each plot. For this method the effect of moderate bulking is mainly caused by a decrease in the number of degrees of freedom and thereby larger uncertainty on the logarithm of the standard deviations. Moderate bulking (bulking pairs of plants) will in most cases decrease the power of tests. Further bulking, up to having only two bulked samples per plot will further decrease the power of the tests which means that the degree of non-uniformity must be much higher for it to be detected – about 3-4 times higher if 30 plants from each of two blocks were bulked into 2 groups of 15 plants for each of the two blocks before the recording was made. These calculations assume that equal amount of material were bulked from each plant. If that is not done the effect of bulking is expected to be larger.

11.2.1.3 In general, if all plants in a plot are bulked such that only a single sample is available for each plot, it becomes in general impossible to calculate the within plot variability and in such cases no tests for uniformity can be performed. In rare cases, where non-uniformity may be judged from values that can only be found in mixtures, non-uniformity may be detected even where a single bulk sample for each plot is used. For example, in the characteristic “erucic acid” in oil seed rape, values between 2% and 45% can only arise because of a lack of uniformity. However this only applies in certain special cases and even here the non-uniformity may only show up under certain circumstances.

11.2.1.4 Bulking across plots have the consequences that that part of the between plot (and block) variation will be included in the estimate standard deviation between bulks. If this variation is relatively large then this will tend to mask any differences in uniformity between varieties. In addition some noise may also be added because the ratio of material form the different plots may vary from bulk to bulk. Finally the assumptions for the present recommend method, COYU, may not be fulfilled in such cases. Therefore it is recommended only to bulk within plots.

11.2.2 Testing for distinctness

The effect of bulking will usually decrease the power of the distinctness much less than for the uniformity test – and may in some cases result in an ignorable small decrease in power. The reason for this is that both the COYD method and the 2×1% method both are based on means (per year and variety for COYD method and per year, block and variety for the 2×1% method). Therefore, the only loss of precision here is the increase in variability caused by fewer measurements. The uncertainty caused by the measurement is usually much smaller than the uncertainty caused by other sources such as plant, soil and climate. If the uncertainty caused by the measurement is very small (relatively to other sources of variation) it is thus expected that the decrease in power will be ignorable as long as there are at least one bulked sample per year and variety for the COYD method and one bulked sample per year, block and variety for the 2×1% method. Also here it is assumed that equal amount of material were bulked from each plant. If that is not the case the effect of bulking may not be as small as described here.

11.3 Distinctness

11.3.1 In the COYD method for examining distinctness the basic values to be used in the analyses are the annual variety means. As bulk sampling also gives at least one value for each variety per year, it will usually still be possible to use the COYD method for distinctness purposes for any degree of bulking, as long as at least one value is recorded for each variety in each year and that the bulk samples are representative for the variety. However, some problems may be foreseen: the assumption of data being normal distributed may be better fulfilled when the mean of many individual measurements are analyzed instead of the mean of fewer measurements or, in the extreme, just a single measurement.

11.3.2 The efficiency of the test of distinctness may be expected to be lower when based on bulked samples than when it is based on the mean of all individual plants in a year. The loss will be from almost zero upwards, depending on the importance of the different sources of variations. The variation which is relevant for the efficiency of variety comparisons is formulated in the following model:

$$\sigma^2_{total} = \sigma^2_{vy} + \sigma^2_p + \sigma^2_i + \sigma^2_m$$

where

σ^2_{total} is the total variance of a characteristic used for comparing varieties. The total variance is regarded as being composed of four sources of variation:

1: σ^2_{vy} the variance component due to the year in which the variety is measured

2: σ^2_p the variance component due to the plot in which the measurement was taken

3: σ^2_i the variance component due to the plant on which the measurement was taken

4: σ^2_m the variance component due to the inaccuracy in the measurement process

11.3.3 In cases where the data are not bulked the variance of the difference between two variety means, σ^2_{diff}, becomes:

$$\sigma^2_{diff} = 2\left\{\frac{\sigma^2_{vy}}{a} + \frac{\sigma^2_p}{ab} + \frac{\sigma^2_i}{abc} + \frac{\sigma^2_m}{abc}\right\}$$

where

a is the number of years used in the COYD method *b* is the number of replicates in each trial

c is the number of plants in each plot

11.3.4 Assuming that each bulk sample has been composed in such a way that it represents an equal amount of material from all the individual plants which have been bulked into that sample, the variance between two varieties based on *k* bulked samples (each of *l* plants) becomes:

$$\sigma_{diff}^2 = 2\left\{\frac{\sigma_{vy}^2}{a} + \frac{\sigma_p^2}{ab} + \frac{\sigma_i^2}{abkl} + \frac{\sigma_m^2}{abk}\right\}$$

where

k is the number of bulk samples

l is the number of plants in each bulk sample

11.3.5 Thus if all plants in each plot are divided in *k* groups of *l* plants each and an average measurement is taken for each of the k groups, then only the last term in the expression for σ_{diff}^2 has increased (as *kl* is equal to *c*). For many characteristics it is found that the variance caused by the measurements process is small and hence the bulking of samples will only have a minor effect on the conclusions reached by the COYD method. Only if the variance caused by the measurement process is relatively large can bulking have a substantial effect on the distinctness tests using COYD.

Example 1

Variances for comparing varieties were estimated (by the use of estimated variance components) for different degrees of bulking. The calculations were based on the weight of 100 seeds of 145 pea varieties grown in Denmark during 1999 and 2000. In this example, the contribution to the variance caused by the measurement process was relatively very small, which means that bulking will have a low influence on the test for distinctness. In a 3 year test with 30 plants in each of 2 blocks, the variance of a difference between two varieties was estimated to be 2.133 and 2.135, for no bulking and a single bulk sample per plot, respectively.

For other variables the variance component due to the measurement process may be relatively more important. However, it is likely that in most practical cases this variance component will be relatively small.

11.3.6 In some cases each bulk sample is not drawn from a specific set of plants (say, plant 1 to 5 in bulk sample 1, plant 6 to 10 in bulk sample 2 etc.), but bulk samples are formed from mixed samples of all plants in a plot. This means that different bulk samples may contain material from the same plants. It must be expected that similar results apply here, although, in this situation, the effect of bulking may have an increased effect because there is no guarantee that all plants will be equally represented in the bulk samples.

11.4 Uniformity

11.4.1 Bulking within plot

11.4.1.1 In COYU the test is based on the standard deviation of the individual plant observations (within plots) as a measurement of uniformity. The log of the standard deviations plus one are analyzed in an over-years analysis; i.e. the values $Z_{vy} = \log(s_{vy} + 1)$ are used in the analyses. The variance on these Z_{vy} values can be regarded as arising from two sources, a component that depends on the variety-by-year interaction and a component that depends on the number of degree of freedom used for estimating the standard deviation, s_{vy} (the fewer degrees of freedom the more variable the standard deviation will be). This can be written (note that the same symbols as used in the distinctness section will be used here with different meaning):

$$Var(Z_{vy}) = \sigma_{vy}^2 + \sigma_f^2$$

where this variance can be regarded as being composed of two sources of variation:

1: σ_{vy}^2 the variance component due to the year in which the variety is measured

2: σ_f^2 the variance component due to the number of degrees of freedom used in estimating s_{vy}

σ_f^2 is approximately $\frac{1}{2v}\left(\frac{\sigma}{\sigma+1}\right)^2$ when the recorded variable is normally distributed and the standard deviations do not vary too much. This last expression reduces to 0.5/v when $\sigma >> 1$. Here σ is the mean value of the s_{vy} values and v is the number of degrees of freedom used in the estimation of s_{vy}.

11.4.1.2 The variance caused by the year in which the variety is measured may be assumed to be independent of whether the samples are bulked or not, whereas the variance caused by the number of degrees of freedom will be increased when bulked samples are used because a lower number of degrees of freedom is available.

11.4.1.3 The variance of a difference between a Z_{vy} for a candidate variety and the mean of the reference varieties' Z_{vy} values may be written:

$$\sigma_{diff}^2 = (\sigma_{vy}^2 + \sigma_f^2)\left(\frac{1}{a} + \frac{1}{ar}\right)$$

where

a is the number of years used in the test

r is the number of reference varieties

Example 2

The effect of bulking in the test for uniformity, an estimate was made using the same data as for Example 1 I Part II, section 11.2.5 [cross ref.]. For a test using 50 reference varieties in 3 years with 30 plants per variety in each of 2 plots per trial the variance for comparing the Z_{vy} value for a candidate variety and the mean of the reference varieties' Z_{vy} will be 0.0004 if no bulking is done. This can be compared to 0.0041, 0.0016 and 0.0007 when 2, 4 and 10 bulk samples per plot were used. Thus, in this example, the *effect of bulking has a great influence on the test for uniformity. The variance increased, approximately by a factor of 10 when changing from individual plant records to just 2 bulk samples per plot. This means that the degree of non-uniformity must be much higher for it to be detected when 2 bulk samples are used instead of individual plant records.*

11.4.2 Bulking across plots

Bulking across plots means that part of the between plot (and block) variation will be included in the estimated standard deviation between bulked samples. If this variation is relatively large it will tend to mask any differences in uniformity between varieties. In addition some noise may also be added because the ratio of material from the different plots may vary from bulk to bulk. Finally the assumptions for the present recommended method, COYU, may not be fulfilled in such cases. Therefore it is recommended to bulk only within plots.

11.4.3 Taking just one bulk sample per plot

In general, if all plants in a plot are bulked such that only a single sample is available for each plot, it becomes impossible to calculate the within plot variability and in such cases no tests for uniformity can be performed. In rare cases, where non-uniformity may be judged from values that can only be found in mixtures, non-uniformity may be detected even where a single bulk sample for each plot is used. For example, in the characteristic "erucic acid" in oil seed rape, values between 2% and 45% can only arise because of a lack of uniformity. However this only applies in certain special cases and even here the non-uniformity may only show up under certain circumstance

[Annex VII follows]

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

New Section 12 - Examining characteristics using image analysis (Drafter: Mr. Gerie van der Heijden (Netherlands))

Notes

1. With regard to new proposals concerning the content of document TGP/8, the TWA, at its thirty-seventh session, held in Nelspruit, South Africa, from July 14 to 18, 2008, proposed to remove Section III: "Examination of characteristics using image analysis" from TGP/12 and to include that section in document TGP/8, on the basis that it did not concern characteristics, but methods of examining characteristics. The TWC, at its twenty-sixth session, agreed with that proposal. The TC-EDC, at its meeting on January 8, 2009, noted that the section on the examination of characteristics using image analysis would require further substantial development and would not be finalized in time for the initial adoption of document TGP/8 (document TGP/8/1) (see document TC/45/5 paragraph 25).

2. The TWC, at its twenty-sixth session, agreed as follows:

(a) for existing characteristics: to explain the need to compare the results of the characteristics examined by the old method and by image analysis. The TWC noted that it might, in some cases, lead to a modification of the existing characteristic, in which case it would be necessary for the Test Guidelines to provide a clear definition of the characteristic, including an outline of the algorithm which defined the characteristic;

(b) for new characteristics: to provide guidance on the need to meet the requirements for a characteristic to be used for DUS, as set out in the General Introduction, and the need to check for independence from other characteristics, in the same way as for other characteristics. In response to an observation from an expert from China, the TWC agreed that the guidance to be developed in document TGP/8 on image analysis should provide guidance on how to consider calibration of images, particularly images containing more than one object, to account for the differing distances of the objects from the camera.

3. The TWC at its twenty-seventh session agreed that existing text be moved to Part I and Mr. Gerie van der Heijden (Netherlands) and Mr. Nik Hulse (Australia) to provide additional information for Part II.

[DRAFT TEXT FOLLOWS]

12. EXAMINATION OF CHARACTERISTICS USING IMAGE ANALYSIS[1]

	Comments of the TWPs in 2011	
General	The TWC received presentations on image analysis by Mr. Adrian Roberts (United Kingdom) (document TWC/29/19), by Mr. Sami Markkanen (Finland) (document TWC/29/21), by Mr. David Hampel (Czech Republic) (document TWC/29/27) and by Mr. Gerie van der Heijden (Netherlands) (document TWC/29/29). The TWC agreed to propose the development of a questionnaire concerning software and hardware used for image analysis and invited UPOV members to make presentations on image analysis at the thirtieth session of TWC session, in 2012.	TWC

12.1 Introduction

Characteristics which may be examined by image analysis should also be able to be examined by visual observation and/or manual measurement, as appropriate. Explanations for observing such characteristics, including where appropriate explanations in Test Guidelines, should ensure that the characteristic is explained in terms which would enable the characteristic to be understood and examined by all DUS experts.

	Comments of the TWPs in 2011	
12.1	The TWV, the TWO and the TWF agreed that Section 12.1 should be reworded to explain that image analysis would be an alternative method for observing a characteristic, rather than a principal method for observing a characteristic.	TWV TWO TWF

12.2 Combined characteristics

12.2.1 The General Introduction (document TG/1/3, Chapter 4, Section 4) states that:

"4.6.3 Combined Characteristics

"4.6.3.1 A combined characteristic is a simple combination of a small number of characteristics. Provided the combination is biologically meaningful, characteristics that are assessed separately may subsequently be combined, for example the ratio of length to width, to produce such a combined characteristic. Combined characteristics must be examined for distinctness, uniformity and stability to the same extent as other characteristics. In some cases, these combined

[1] TWA and TWC agreed to move Section III "Examination of characteristics using image analysis" from TGP/12 to TGP/8.

> characteristics are examined by means of techniques, such as Image Analysis. In these cases, the methods for appropriate examination of DUS are specified in document TGP/12, 'Special Characteristics'."

12.2.2 Thus, the General Introduction clarifies that the use of image analysis is one possible method for examining characteristics which fulfil the basic requirements for use in DUS testing (see document TG/1/3, Chapter 4.2), which includes the need for the uniformity and stability of such characteristics to be examined. With regard to combined characteristics, the General Introduction also explains that such characteristics should be biologically meaningful.

12.3 Guidance on the use of image analysis

[to be developed by the Technical Working Party on Automation and Computer Programs (TWC)]

	Comments of the TWPs in 2011	
12.3	The TWA considered Annex VII and noted that the TWC would develop subsection 12.3 "Guidance on the use of image analysis".	TWA

The TWC, at its Twenty-sixth Session, agreed as follows:

(a) for existing characteristics: to explain the need to compare the results of the characteristics examined by the old method and by image analysis. The TWC noted that it might, in some cases, lead to a modification of the existing characteristic, in which case it would be necessary for the Test Guidelines to provide a clear definition of the characteristic, including an outline of the algorithm which defined the characteristic;

(b) for new characteristics: to provide guidance on the need to meet the requirements for a characteristic to be used for DUS, as set out in the General Introduction, and the need to check for independence from other characteristics, in the same way as for other characteristics

In response to an observation from an expert from China, the TWC agreed that the guidance to be developed in document TGP/8 on image analysis should provide guidance on how to consider calibration of images, particularly images containing more than one object, to account for the differing distances of the objects from the camera.]

The TWC also agreed that Mr. Gerie van der Heijden (Netherlands) should prepare a draft text for Section III, Subsection 3, taking into account the comments made above.]

[the TWA, at its thirty seventh session, agreed that for existing characteristics: to explain the need to compare the results of the characteristics examined by old method and by image analysis; for new characteristics: to provide guidance on the need to meet the requirements for a characteristic to be used for DUS, as set out in the General Introduction, and the need to check for independence from other characteristics]

	<u>Comments of the TWPs in 2011</u>	
New section	The TWC agreed that a new section should be prepared on the basis of the discussion on documents TWC/29/19, TWC/29/21, TWC/29/27 and TWC/29/29. Drafters: experts from Netherlands (first drafter), Czech Republic, Finland and the United Kingdom.	TWC

[Annex VIII follows]

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

New Section 13 - Methods for data processing for the assessment of distinctness and for producing variety descriptions: (Drafters: experts from France, Germany, Japan, Kenya and the United Kingdom)

Notes

1. The TWC at its twenty-sixth session agreed that the information provided in documents TWC/26/15 and TWC/26/23, presented by Mr. Vincent Gensollen (France) and Mr. Uwe Meyer (Germany), respectively, and an oral presentation by Ms. Mariko Ishino (Japan) included in document TWC/26/15 Add. provided valuable guidance on data processing for the assessment of distinctness and for producing variety descriptions and noted that UPOV did not have guidance on that matter in the TGP documents. It agreed that a new section should be created in document TGP/8/1, Part I as "Data processing for the assessment of distinctness and for producing variety descriptions for producing variety descriptions" and that the methods used by France, Germany and Japan should be included in a new section in document TGP/8/1, Part II as "Methods for data processing for the assessment of distinctness and for producing variety descriptions.

2. The TWC at its twenty-seventh session agreed that experts from Finland, France, Germany, Italy, Japan, Kenya and United Kingdom to provide a short description of the principles underlying the detailed methods provided in Part II.

3. Mrs. Sally Watson (United Kingdom) to provide an example for Section 13.1

[DRAFT TEXT FOLLOWS]

13. METHODS FOR DATA PROCESSING FOR THE ASSESSMENT OF DISTINCTNESS AND FOR PRODUCING VARIETY DESCRIPTIONS

[The TWC agreed that the information provided in documents TWC/26/15 and TWC/26/23, presented by Mr. Vincent Gensollen (France) and Mr. Uwe Meyer (Germany), respectively, and an oral presentation by Ms. Mariko Ishino (Japan) included in document TWC/26/15 Add. provided valuable guidance on data processing for the assessment of distinctness and for producing variety descriptions and noted that UPOV did not have guidance on that matter in the TGP documents. It agreed that a new section should be created in document TGP/8/1, Part I as "Data processing for the assessment of distinctness and for producing variety descriptions for producing variety descriptions" and that the methods used by France, Germany and Japan should be included in a new section in document TGP/8/1, Part II as "Methods for data processing for the assessment of distinctness and for producing variety descriptions. [..]The TWC agreed that Finland, France, Germany, Japan, Kenya and the United Kingdom should prepare information on their methods for inclusion in the next draft of document TGP/8]

	Comments of the TWPs in 2011	
General	The TWA considered Annex VIII and agreed that further guidance should be developed based on the information provided at the UPOV DUS Seminar, held in Geneva, in March 2010, and the examples provided in Annex VIII. The TWA noted that, for the time being, two examples have been provided.	TWA
	The TWC agreed that Mrs. Sally Watson should update the information on the species presented in the method from the United Kingdom and that it should be included in TGP/8. It is also agreed that the method provided by Japan should be included in TGP/8. received a presentation from Mr. Vincent Gensollen (France) on a method used in France for producing variety descriptions for herbage crops. The TWC agreed that the method presented by Mr. Gensollen should be included in document TGP/8.	TWC
	The TWV and the TWO considered Annex VIII in conjunction with Annex III.	TWV TWO
	The TWF noted that some other examples from Republic of Korea and other members presented at the Seminar on DUS testing should be added.	TWF

France

Example Illustrating how Variety Descriptions are Developed in Herbage Crops

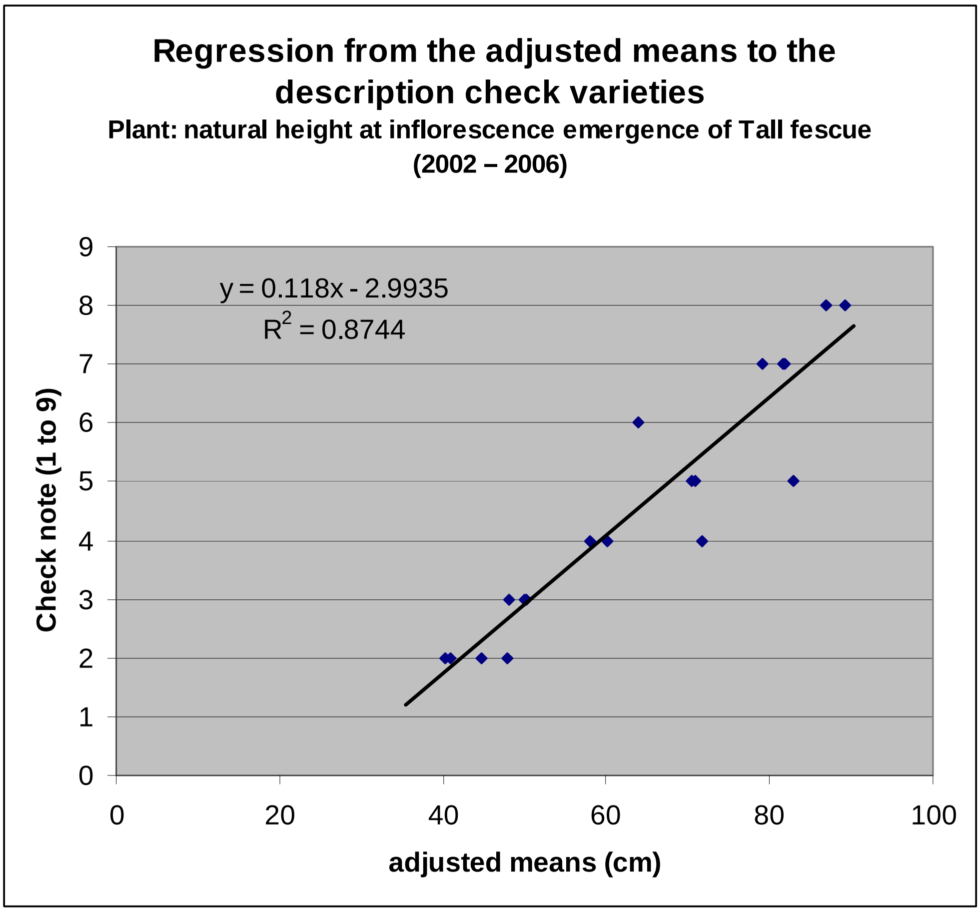
UPOV Test Guidelines Meadow Fescue, Tall Fescue, characteristic no. 10 “plant: natural height at inflorescence emergence” for tall fescue varieties

1. The data of this characteristic come from measurements on Single plant (MS) in spaced plant trials (A). In that case, the Combined Over Years Distinctness (COYD) analysis provides adjusted means of the reference varieties and the candidate varieties.

2. For the purpose of the description, we transform the adjusted means to notes. We use a linear regression from the adjusted means to “description check varieties”. The description check varieties are already well described example varieties (i.e. example varieties of the UPOV guide line or national example varieties).

3. The graph below shows the regression from the adjusted means to the description note. In this case 4 varieties had been described with the note 2, 2 varieties with note 3.

FIG. 1: LINEAR REGRESSION FROM THE ADJUSTED MEAN TO THE DESCRIPTION CHECK VARIETY



Regression square (R^2) = 0.8744.
The regression is valid if $R^2 > 0.6$.

Predicted note = 0.118 x adjusted mean - 2.9935.

From the equation above, we can compute the description note.

TAB 3: ADJUSTED MEAN AND DESCRIPTION NOTE FOR THE CHARACTERISTIC NATURAL HEIGHT AT INFLORESCENCE EMERGENCE OF TALL FESCUE VARIETIES.

Variety name	Adjusted mean (cm)	Check description note	Predicted note	Description note
C1	35.50	.	1.19423	1
BONAPARTE	44.71	2	2.28068	2
ELDORADO	47.90	2	2.65699	3
C2	48.15	.	2.68648	3
MONTSERRAT	48.15	3	2.68648	3
MURRAY	50.29	3	2.93893	3
C3	52.78	.	3.23266	3
TOMAHAWK	54.80	.	3.47095	3
BORNEO	58.11	4	3.86141	4
C4	58.94	.	3.95932	4
BARDAVINCI	60.28	.	4.11739	4
VILLAGEOISE	62.07	.	4.32855	4
C5	62.13	.	4.33563	4
DANIELLE	63.97	6	4.55268	5
DIVYNA	64.54	.	4.61992	5
C6	69.54	.	5.20975	5
GARDIAN	70.55	5	5.32889	5
EMERAUDE	70.91	5	5.37136	5
CENTURION	71.81	4	5.47753	5
SZARVASI 56	73.18	.	5.63914	6
BARCEL	79.41	.	6.37406	6
DULCIA	81.63	7	6.63594	7
LUNIBELLE	81.85	7	6.66190	7
C7	86.57	.	7.21869	7
BARIANE	87.02	8	7.27177	7
C8	87.44	.	7.32132	7
APRILIA	89.28	8	7.53837	8
C9	89.65	.	7.58202	8
FLEXY	90.31	.	7.65988	8

This example illustrates a simple way to obtain coherent notes with computations that can be performed without the need of a statistical package.

Japan

The Method to Adjust the Table of Assessment for Quantitative Characteristics

Japan
National Center for Seeds and Seedlings (NCSS)

TABLE OF CONTENTS

1. Introduction

1.1 This provides an explanation of the Japanese methods to adjust the table of assessment for quantitative characteristics in characteristics table of TG.

1.2 The method is based on the premise as below.

a) This method is mainly used for ornamental plants and vegetable crops.

b) Basically, DUS growing trial for ornamental plants and vegetable crops is assessed in two independent growing cycles. When we decide it is satisfactory for the assessment of DUS, further growing trial will not be done. This document explains the adjusting method of the quantitative characteristics from the result of DUS growing trial of one growing cycle.

c) The term “the table of assessment” means the table to evaluate the notes from the data of quantitative characteristics.

2. Method with the Fundamental Table of Assessment (FAT)

2.1 [Background]

2.1.1 For the assessment of note in most quantitative characteristics, the relative assessment based on the data of the example variety in one time seems to be general method. Especially when we start DUS growing trial about new species, we use this method. But, we seek more effective method to reduce the yearly variation for concerned species which we have examined for many years.

.

2.1.2 The method with FAT is used for this purpose. We make FAT as the adjustable base only for the species that had examined in sufficient number of DUS growing trials. FAT is adjusted every year to correct yearly variations of data.

2.2 [What is FAT?]

2.2.1 FAT is the table of assessment that made from the enough experimental data about the species. In the concrete, one of the experimental data is “Proposition by experts”. It is the table that is based on the expert’s experience and knowledge, and the table covers the full ranges of variations that the species or variety groupings show under the normal growth. The other of the experience is “Accumulated statistical data.” It is the data accumulated about several example varieties in sufficient number of DUS growing trials. We try to accumulate the data from sufficient number of growing trials. But it needs long time to accumulate the data in one site for many times. Before we get enough data to make FAT, we set the notes based on example variety’s data from one growing trial and our experiences. If we estimate the data accumulated in certain place for one species are enough stable, we make FAT based on the data. FAT is available only for species that had examined for sufficient experience of DUS growing trial about several example varieties.

2.3 [Composition of FAT]

2.3.1 Table 1 shows the part of example FAT, the characteristic “length of leaf blade”. There are nine notes. In the note 5,
Range : 70-79 mm
Interval : 10 mm,
Median : 75 mm
Standard example variety of the note 5 : ‘EV-B’

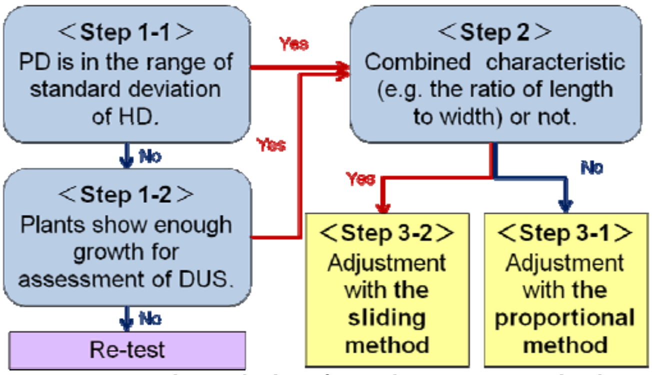
Table 1: Example FAT for the characteristic “ length of leaf blade”

Characteristics	Note	1	2	3	4	5	6	7	8	9
Length of leaf blade (mm)	Range	～ 39	40 ～ 49	50 ～ 59	60 ～ 69	70 ～ 79	80 ～ 89	90 ～ 99	100 ～ 109	110 ～
	Interval		10	10	10	10	10	10	10	
	Median		45	55	65	75	85	95	105.	
	Example variety			EV-A		EV-B				

2.4 [Practical adjusting methods for use of FAT]

2.4.1 【Overview of the methods】

2.4.1.1 There are two methods in adjustment of FAT. One is the proportional method, the other is the sliding method. PD indicates Present data, the data of the example variety measured in this time. HD indicates Historical data, the mean of the data of the example variety measured in sufficient times of DUS growing trial.



**PD: Present data = The data of Example Variety measured in this time*
HD: Historical data = Mean of the Data of Example Variety measured in sufficient number of DUS growing trial

Fig. 1: Flow chart of the practical adjusting method with FAT

2.4.1.2 Figure 1 shows the practical adjusting method.
Step 1-1: Check whether PD is in the range of standard deviation of HD
Step 1-2: Check whether plants show satisfactory growth for assessment of DUS
Step 2 : Check whether the characteristic is combined characteristic or not.
Step 3-1: Adjustment FAT with the proportional method
Step 3-2: Adjustment FAT with the sliding method

2.4.2 【Step 1-1: Check whether PD is in the range of standard deviation of HD】

2.4.2.1 We confirm the example variety's normal growth by checking step 1-1. If step 1-1 is not satisfied, we should check whether the growing trial can be done reasonably and properly or not.

2.4.2.2 The examples are as follows.
Characteristic "length of leaf blade"
HD: 74.0mm
Standard deviation: 5.01
Range of the standard deviation: 69.0-79.0mm

2.4.2.2.1 If PD is 70.3mm, PD is in the range of standard deviation of HD. → Go to step 2

2.4.2.2.2 If PD is 83.6mm, PD is out of the range of standard deviation of HD. → Go to step 1-2.

2.4.3 【Step 1-2:Check whether plants show satisfactory growth for assessment of DUS】

2.4.3.1 The purpose of step 1-2 is to check whether the growing trial can be done reasonably and properly or not.

2.4.3.2 If the example variety we expect to use for adjustment doesn't show satisfactory growth, we can use another example variety (which shows satisfactory growth and has enough experimental data) for adjustment of FAT. In this case, we estimate plants in this growing trial shows satisfactory growth for evaluation of DUS.→ Go to step 2

2.4.3.3 In the case other varieties also show unusual growth, we should try to make clear the reason with assistance of the plant species expert. After taking into account the distance from the range of standard deviation of HD and the advice of our expert and examiner, we estimate whether we can evaluate DUS in this growing trial.

We can evaluate DUS.→ Go to step 2
We can't evaluate DUS. → Re-test

2.4.4 【Step 2: Check whether the characteristic is combined characteristic or not】

2.4.4.1 The purpose of step 2 is to decide which method, the proportional method or the sliding method, is more suitable for the characteristic. In the proportional method, range and interval of notes are adjusted at once. In the sliding method, range is adjusted on the one hand and interval is not changed. It means that the proportional method is not suitable for the characteristics that need fixed interval. In the concrete, the combined characteristics are generally stable than other characteristics and they need fixed interval. In such case, the sliding method is applied.

2.4.4.2 Characteristic "length of leaf blade"
It is not the combined characteristic. → Go to step 3-1

2.4.4.3 Characteristic "Leaf: ratio length/width"
It is the combined characteristic. → Go to step 3-2

2.4.5 【Step 3-1: Adjustment FAT with the proportional method】

2.4.5.1 We calculate the proportion of the measured data in this time to the mean of the historical data about an example variety. FAT multiplied by the proportion gives the adjusted table of assessment in this time.

2.4.5.2 The examples are as follows.
Characteristic " length of leaf blade"
PD: 70.3mm
HD: 74.0mm
Proportion (PD/HD) =0.95

2.4.5.3 The upper line of Figure 2 is FAT expressed in a number line. FAT multiplied 0.95 gives the adjusted table of assessment of this time, the lower line.

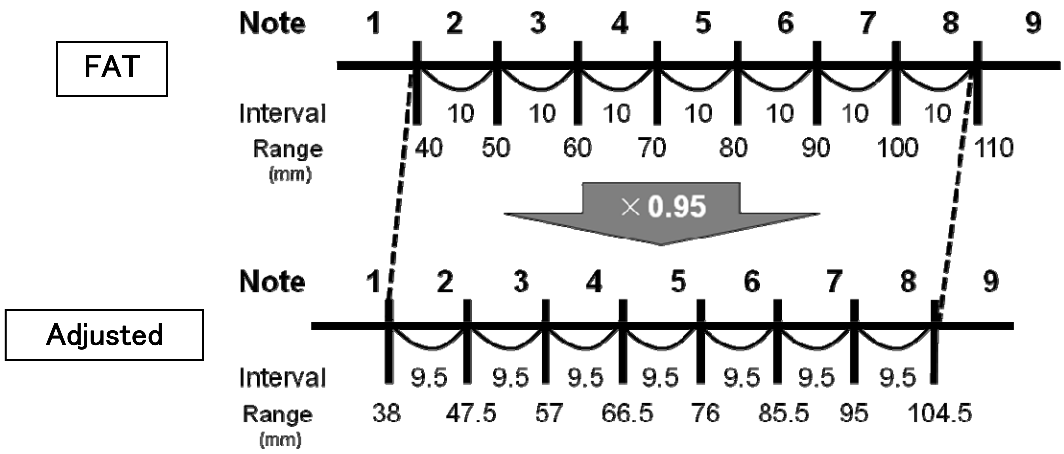


Fig.2: Adjustment FAT with the proportional method

2.4.5.4 We take the note 5 as an example,
The minimum of the range is 70. 70 multiplied by 0.95 make 66.5.
The maximum of the range is 80. 80 multiplied by 0.95 make 76.
The interval of the note 5 changes from 10 to 9.5.

2.4.6 【Step 3-2: Adjustment FAT with the sliding method】

2.4.6.1 We do subtraction the mean of the historical data from the measured data in this time about an example variety. FAT added to the difference is the adjusted table of assessment in this year.

2.4.6.2 The examples are as follows.
Characteristic "Leaf: ratio length/width"
PD of the example variety of the note 5 (EV) is 1.16.

2.4.6.3 The upper line of Figure 3 is FAT expressed in a number line. PD of EV, 1.16 is allocated in the note 4 in FAT. We should adjust FAT as the median of the note 5 becomes the same value to PD of EV, 1.16. FAT subtracted 0.19 gives the table of assessment of this time, the lower line.

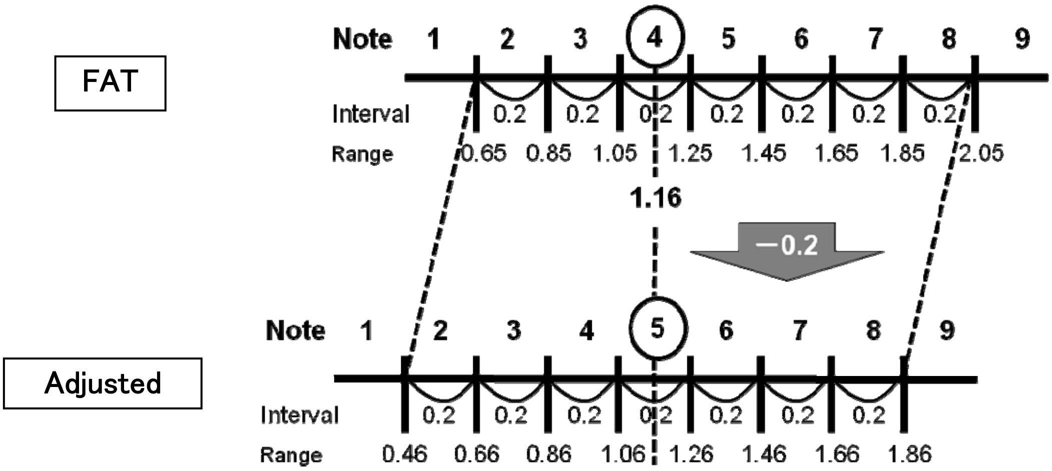


Fig.3: Adjustment FAT with the sliding method

2.4.6.4 We take "the note 5" as an example.
The minimum of the range 1.25－0.19 = 1.06.
The maximum of the range 1.45－0.19 = 1.26.
The interval is not adjusted.
The median of the note 5 = PD of EV, 1.16.

2.4.6.5 Generally, there are several example varieties in a characteristic. But we select one example variety from them for adjustment of FAT. We basically use the least variable example variety during many years' DUS growing trials about each characteristic.

2.5 [Difference between self-pollinated varieties and cross-pollinated varieties]

2.5.1 We use the same method to self-pollinated varieties and cross-pollinated varieties. But the adjustable range changes according to dispersion of HD of example variety. Because our methods are based on the data of example variety, the propagation type of example variety is automatically reflected in the adjustable range.

2.5.2 Table 2 shows the example data. In general, there is tendency that the dispersion of the self-pollinated varieties is lower than that of the cross-pollinated varieties. In this example, HD of two varieties is the same. But the dispersion of self- pollinated varieties example variety is lower than that of cross-pollinated varieties.

Table 2: Example data of self-pollinated example variety and cross-pollinated example variety

Trial number	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Historical Data(HD)	Dispersion	Standard deviation	Coefficient of variance
Self E.V.	80	84	81	83	86	88	83	80	87	88	84.0	9.78	3.13	11.64
Cross E.V.	75	84	74	83	87	96	84	75	88	94	84.0	59.11	7.69	70.37

*E.V.is example variety

2.5.3 Figure 4 shows the normal curve of two varieties of different propagating type. The curve of self-pollinated example variety is narrower than that of cross-pollinated example variety. As I said earlier, if the data of this year is in the range of standard deviation, we can adjust FAT. Therefore, the adjustable range of self-pollinated varieties becomes narrower than that of cross-pollinated ones automatically.

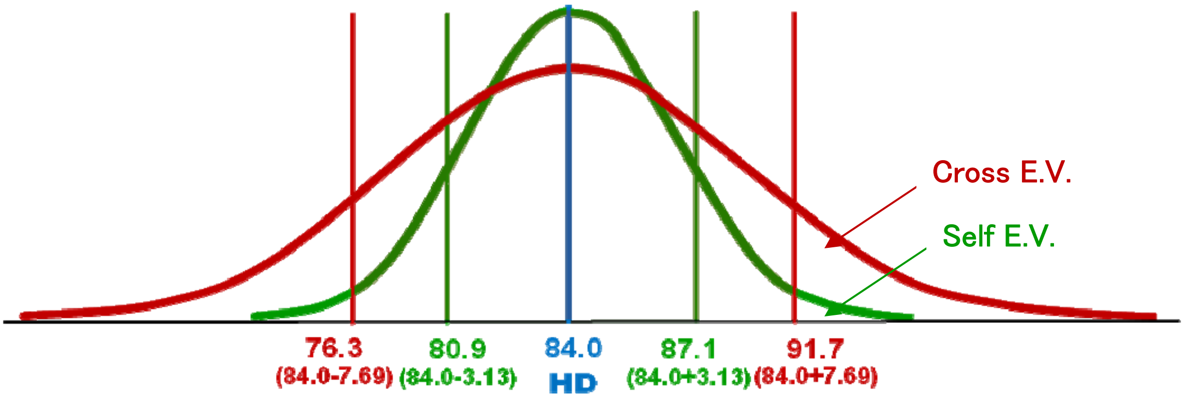


Fig.4: Normal curve of self-pollinated example variety (Self EV) and cross-pollinated example variety (Cross EV)

3. Conclusions

3.1 We have two methods to adjust FAT. One is the proportional method, and the other is the sliding method. In the proportional method, we calculate the proportion of the measured data in this time to the mean of the historical data (HD) about example variety. FAT multiplied by the proportion is the adjusted table of assessment in this time. The sliding method is applied to the characteristics that need fixed interval. We do subtraction the mean of the HD from the measured data in this time about example variety. We can get the adjusted table of assessment in this time by adding the difference to FAT.

3.2 We use the same method to self-pollinated varieties and cross-pollinated varieties to assess the quantitative characteristics. The difference between self-pollinated varieties and cross-pollinated varieties is the allowable range of the value of PD to estimate whether we can adjust the FAT or not. The adjustable range changes according to dispersion of HD of an example variety. Generally, the adjustable range of self-pollinated varieties becomes narrower than that of cross-pollinated varieties because the dispersion of the former is narrower than that of latter. Because our methods are based on the enough experimental data of example variety, the dispersion of HD according to the propagation type of example variety is automatically reflected in the adjustable range.

United Kingdom

Handling Measured, Quantitative Characteristics for Vegetable and Herbage Crops Tested in the United Kingdom

1. This document provides an explanation of how measured, quantitative characteristics are handled and used to develop variety descriptions in the United Kingdom for vegetable and herbage crops.

2. In vegetable and herbage crops, which are mostly cross-pollinated except for pea which is self-pollinated, the trials are conducted according to the UPOV Test Guidelines.

3. For the measured, quantitative characteristics, as part of the determination of distinctness, COYD is applied on the original scale of the characteristics.

4. To develop variety descriptions, over-year variety means are calculated on the original scale of the characteristics. These over-year means are then converted to notes.

5. For each crop the over-year variety means of the varieties in trial are calculated from their yearly means in trials. For herbage crops the past 10 years are used, whereas for vegetable crops all years are included in which the reference collection varieties have been tested. As not all varieties are present in all years, a fitted constants analysis is used to adjust the over-year means for the different years varieties were present in. This is done using the DUSTNT module FITC in conjunction with the module FIND.

6. The over-year means are converted to notes using the DUSTNT module VDES. This permits two methods of division of the range of expression into states and notes as follows, where the number of states is as given in the UPOV Test Guideline:-

a) By use of delineating varieties to divide the range of expression into states.

b) By division of the range of expression of the over-year means for the reference collection varieties into equal-spaced states.

These methods are illustrated by an example in Figures 1 and 2 respectively.

7 For vegetable crops excluding potato method (b) is used to divide the range of expression into states and notes, and for herbage crops method (a) is used.

8 For herbage crops the DUSTNT module SAME is used to check whether there are varieties with the same variety description.

9 For herbage crops the DUSTNT module MOST, is used in conjunction with the modules SSQR and DIST to find most similar varieties based on multivariate distances.

Figure 1: Example illustrating how Variety Descriptions are developed in Herbage crops using delineating varieties in United Kingdom

Characteristic: UPOV No 20, Inflorescence: number of spikelets

10. The five states for this characteristic are defined by the following delineating reference varieties (shown in bold in the table below).

Reference variety	**Delineates**
R2	Upper limit of state 1
R5	Lower limit of state 3
R10	Upper limit of state 3
R14	Lower limit of state 5

11. To obtain notes for the candidate varieties (C1…C5) for this characteristic, the over-year variety means of the candidate and reference varieties are calculated from their yearly means in a fitted constants analysis. The yearly and over-year variety means, sorted by the latter, are shown below.

12. As the yearly means for candidates C1 and C2 are between those for varieties R2 and R5, they have note 2.
As the yearly mean for candidate C3 is between those for varieties R10 and R14, it has note 4.
As the yearly mean for candidate C4 is between those for varieties R5 and R10, it has note 3.
As the yearly mean for candidate C5 is less than that for variety R2, it has note 1.

Reference variety	Yearly means 1	2	3	4	5	6	7	8	9	10	Over-year mean	Note
R1	*	*	*	22.44	23.09	20.40	22.83	23.71	20.79	22.33	21.95	1
R2	*	*	*	23.36	22.88	21.65	21.39	24.23	19.49	23.27	**22.05**	1
R3	*	*	*	*	*	22.26	21.35	24.57	20.13	23.14	22.2	2
R4	19.77	22.05	22.17	25.33	21.84	20.57	22.57	23.55	21.80	23.55	22.32	2
R5	21.15	23.13	23.75	24.74	23.74	23.67	23.80	25.25	21.71	24.55	**23.55**	3
R6	*	*	*	*	24.64	23.00	23.76	25.02	22.16	24.25	23.62	3
R7	*	*	*	*	*	21.47	25.93	24.65	23.07	25.24	23.98	3
R8	*	*	25.00	24.92	24.97	23.51	24.55	26.03	22.31	25.88	24.34	3
R9	*	24.33	25.43	24.18	25.73	23.13	24.74	26.19	23.59	25.90	24.56	3
R10	*	*	*	*	*	22.22	24.82	26.28	25.14	25.56	**24.72**	3
R11	*	*	*	*	*	*	25.35	27.77	24.60	27.11	25.83	4
R12	25.13	27.58	28.57	27.01	27.98	25.42	28.52	27.88	27.30	27.27	27.27	4
R13	*	*	*	*	28.34	26.31	27.68	30.01	26.63	28.41	27.71	4
R14	26.77	27.49	28.65	28.90	29.33	28.19	28.22	29.76	27.91	28.00	**28.32**	5
R15	*	*	*	*	29.48	28.4	30.34	29.85	27.48	29.5	28.99	5
Candidate variety												
C1	*	*	*	*	*	*	*	22.93	22.65	23.36	22.57	2
C2	*	*	*	*	*	*	*	24.84	22.25	23.17	23.01	2
C3	*	*	*	*	*	*	*	26.97	24.73	27.39	25.95	4
C4	*	*	*	*	*	*	*	*	22.63	26.08	24.47	3
C5	*	*	*	*	*	*	*	*	20.98	22.12	21.67	1
Year means	22.30	24.17	24.99	25.27	25.12	23.36	24.75	25.93	23.37	25.31		

Figure 2: Example illustrating how Variety Descriptions are developed in Peas by division of the range of expression into equal-spaced states in United Kingdom

Characteristic: UPOV No 15, Stipule: length

13. To obtain notes for the candidate varieties (C1…C5) for this characteristic, the over-year variety means of the candidate and reference varieties are calculated from their yearly means in a fitted constants analysis. The yearly and over-year variety means, sorted by the latter, are shown below.

14. The five states for this characteristic are defined here by division of the range of expression of the over-year means for the reference collection varieties into equal-spaced states. The range of expression is 109 (= 139 - 30). So each state is of width 109/5 = 21.8, and the upper limits of states 3, 4, 5 and 6 are 51.8, 73.6, 95.4 and 117.2 respectively.

15. If the technical experts judge the range of variation to be large, the 3-7 scale may be expanded to a 1-9 scale.

16. As the yearly means for candidates C1 and C2 are less than 51.8, they have note 3.
As the yearly mean for candidate C3 is between 51.8 and 73.6, it has note 4.
As the yearly mean for candidate C4 is between 73.6 and 95.4, it has note 5.
As the yearly mean for candidate C5 is greater than 117.2, it has note 7.

Reference variety	Yearly means 1	2	3	4	5	6	7	8	9	Over-year mean	Note
R1	*	*	*	*	*	21	36	22	24	30.0	3
R2	*	*	*	29	39	29	39	25	28	35.4	3
R3	*	55	65	68	48	44	59	56	28	54.7	4
R4	72	61	73	45	59	52	68	56	53	59.9	4
R5	*	*	*	*	*	68	70	58	60	68.4	4
R7	*	*	77	61	73	72	80	64	61	72.2	4
R8	*	*	*	*	96	107	102	101	91	102.7	6
R9	121	120	113	78	117	102	109	105	79	104.7	6
R10	*	97	112	95	124	110	117	112	88	108.7	6
R11	*	*	*	122	121	128	105	102	85	117.7	7
R12	*	*	*	*	110	130	129	106	97	114.6	7
R13	*	*	*	*	*	132	133	130	112	131.2	7
R15	*	*	*	*	*	121	155	157	106	139.0	7
Candidate variety											
C1	*	*	*	*	*	*	55	32	27	43.3	3
C2	*	*	*	*	*	*	55	58	25	51.2	3
C3	*	*	*	*	*	*	*	46	44	55.7	4
C4	*	*	*	*	*	*	*	75	54	75.2	5
C5	*	*	*	*	*	*	*	124	102	123.5	7
Year means	97	84	91	75	84	81	88	79	65		

[Annex IX follows]

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

New Section - Guidance of data analysis for blind randomized trials (Drafters: France and Israel to provide examples)

Notes

Comments: proposed by the TC at its forty-fifth session

	Comments of the TWPs in 2011	
General	The TWA noted the information provided in Annex IX.	TWA
	The TWC considered that this was a section to be developed by crop experts.	TWC
	The TWV, The TWO and the TWF agreed that the experts from France should develop guidance on data analysis for blind randomized trials from their experience, including their use of blind randomized trials for disease resistance.	TWV TWO TWF

[Annex X follows]

ANNEX X

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

New Section - Statistical methods for visually observed characteristics (Drafter: Kristian Kristensen)

Notes

Comments: the TC at its forty-sixth session requested the TWC to investigate this subject for possible inclusion in the revision of document TGP/8.

	Comments of the TWPs in 2011	
General	The TWO and the TWV noted the proposals in Annex X.	TWO TWV

[DRAFT TEXT FOLLOWS]

Section 10 – Minimum number of comparable varieties for the Relative Variance Method

THE COMBINED OVER-YEARS METHOD FOR NOMINAL CHARACTERISTICS

	Comments of the TWPs in 2011	
Title	The TWA and the TWV considered Annex X and noted the new draft for the part concerning “The combined over-years method for binomial characteristics”. It recommended to modify the title of the three parts of “Section 10 – Minimum number of comparable varieties for the Relative Variance Method” as follows: THE COMBINED OVER-YEARS METHOD FOR NOMINAL-SCALED CHARACTERISTICS THE COMBINED OVER-YEARS METHOD FOR ORDINAL-SCALED CHARACTERISTICS THE COMBINED OVER-YEARS METHOD FOR BINOMIAL-SCALED CHARACTERISTICS	TWA TWV

Summary of requirements for application of the method

1. The method is appropriate to use for assessing distinctness of varieties where:

- The characteristic is nominal and recorded for individual plants (usually recorded visually)
- There are some differences between plants
- The observations are made over at least two years or growing cycles on a single location

- There should be at least 20 degrees of freedom for estimating the random variety-by-year interaction term.
- The expected number of plants for each combination of variety and note should be at least one – and for most of the combinations the number should be at least 5.

	Comments of the TWPs in 2011	
	The TWC agreed that it would be necessary to explore the consequences of the decisions for DUS examination, as the method is a test for differences in the distribution (both location and dispersion). Also, the consequences of excluding certain varieties from the test, as they did not have sufficient numbers in some cells, should be further investigated.	TWC

Summary

2. The method can be considered as an alternative to the χ^2-test for independence in a contingency table. The χ^2-test only takes the variation caused by random sampling into account and may thus be too liberal if additional sources of variation are present. The combined over-years method for nominal characteristics takes other sources of variation into account by including a random variety-by-year interaction term (as for the COYD method described in TGP/8/1 Part II: 3). The inclusion of the random effect is expected to decrease the number of distinct pairs of varieties compared to the χ^2-test for independence, but to better ensure that the decisions are consistent over coming years. The method is based on a generalisation of the traditional analyses of variance and regression methods for normally distributed data, which are called "generalized linear mixed models".

3. The combined over-years method for nominal characteristics involves

- Calculating the number of plants for each note for each variety in each of the two or three years of trials, which results in a 3-way table (see the example)
- Analyse the data using appropriate software
- Compare each candidate to the reference varieties and the other candidates at the appropriate level of significance to see which varieties the candidate is distinct from
- Check if the variety-by-year interaction term for distinct pairs is considerably larger than the average for all variety pairs

Technical description of the method

4. The method is based on a generalized linear mixed model using the generalized logit as link function assuming that the data are multinomial distributed (for more information on generalised linear mixed models see e.g. McCulloch and Searle, 2001 or Agresti, 2002). The model resembles the COYD method for normally distributed characteristics by including the year×variety interaction as a random effect. However, for each of the n-1 notes of a nominal characteristic there will be a random effect which is assumed to be normally distributed with a constant variance. The model can be written as:

$(Y_{1jk}, Y_{2jk}, Y_{3jk}, ..., Y_{njk})$ are multinomial distributed with parameters $(\pi_{1jk}, \pi_{2jk}, \pi_{3jk}, ..., \pi_{njk})$

$$\log\left(\frac{\pi_{ijk}}{\pi_{njk}}\right) = \mu_i + \beta_{ij} + \delta_{ik} + E_{ijk} \qquad \text{for } i = 1, 2, ..., n-1$$

where

Y_{ijk} is the number of plants for variety j in year k for note i

μ_i is the effect of note i $(i = 1, 2, 3, ..., n-1)$

β_{ij} is the effect of variety j for note i $(i = 1, 2, 3, ..., n-1,\ j = 1, 2, 3, ..., v)$

δ_{ik} is the effect of year k for note i $(i = 1, 2, 3, ..., n-1,\ k = 1, ..., y)$

E_{ijk} is the random effect of variety j in year k for note i $(i = 1, 2, 3, ..., n-1,\ j = 1, 2, 3, ..., v,$ $k = 1, ..., y)$

E_{ijk} is assumed to be normally distributed with mean zero and a constant variance for each of the $n-1$ levels of the note, i.e. $E_{ijk} \sim N(0, \sigma_i^2)$

n, v and y are the number of notes, varieties and years, respectively

5. In the formulation above it is assumed that the last note (number n) is taken as the reference note in the generalized logit. For improving the performance of the analyses it is recommended to ensure that the note used as the reference is the note that occurs most often (Agresti, 2002). The estimates of the parameters μ_i, δ_{ik}, and β_{ij} can be used to estimate the relative number of plants with a given note for each variety, and the differences between pairs of varieties can be quantified and tested by estimating the differences between β_{ij}-β_{il} for each of the n-1 notes. The overall test will be the result of a contrast for each of those notes using an F-test with n-1 degrees of freedom in the numerator, whereas the degrees of freedom in the denominator will depend on the actual pair of varieties and the size of the random variety-by-year interaction, but will usually be in the range between $(y\text{-}1)(v\text{-}1)$ and $(n\text{-}1)(y\text{-}1)(v\text{-}1)$. The relative number of plants for each note and variety may be calculated as follows:

First calculate: $\hat{P}_{ij} = \hat{\mu}_i + \hat{\beta}_{ij} + \frac{1}{y}\sum_{k}^{y} \hat{\delta}_{ik}$ for $i = 1, 2, ..., n-1$ and each variety, j

$$\text{Then calculate } \hat{\pi}_{ij} = \begin{cases} e^{\hat{P}_{ij}} / (1 + \sum_{l=1}^{n-1} e^{\hat{P}_{lj}}) & \text{for } i = 1, 2, ..., n-1 \\ 1 - \sum_{l=1}^{n-1} \hat{\pi}_{lj} & \text{for } i = n \end{cases}$$

where

$\hat{\pi}_{ij}$ is the estimated relative number of plants with note i for variety j

Other terms as defined above.

6. As a large year×variety interaction for a specific pair of varieties may cause that pair to be distinct, for instance if a very large difference occurs in one of the years but not in other years. To avoid that situation the year×variety interaction for each pair of varieties is compared to the average year×variety pair interaction using the quotient between the mean square for the interaction of the actual pair of varieties and the average interaction of all variety pairs. This quotient is here called F_3 and may be based on a joint contrast for the interaction of each of the n-1 notes. This will result in a quotient (F_3) which it is suggested is

tested approximately by assuming that the quotient is F-distributed with *(n*-1*)(y*-1*)* and *(n*-1*)(y*-1*)(v*-1*)* degrees of freedom. The F_3 may be calculated as:

$$F_{3jl} = \frac{1}{n-1}\sum_i T_{ijl} / \overline{T}_{i..}$$

where

F_{3jl} is the quotient based on the mean interaction, called the F_3, for variety pair j,l

T_{ijl} is the mean square for the j,l pair of varietes for note *i*.

$\overline{T}_{i..}$ is the average mean square of all pair of varietes for note *i*.

7. It may be valuable also to calculate a quotient that can be used to get a measure of how much each variety contributes to the interaction. Such a quotient, called F_4, may be based on a joint contrast for the interaction of each of the *n*-1 notes. This will result in a quotient (F_4) which it is suggested is tested approximately by assuming that the quotient is approximately F-distributed with (*n*-1)(*y*-1) and (*n*-1)(*y*-1)(*v*-1) degrees of freedom. The F_4-value may be calculated as:

$$F_{4j} = \frac{1}{(n-1)(y-1)}\sum_i \left(\sum_k \hat{E}_{ijk}^2 / MS_{Ei} \right)$$

where

F_{4j} is the quotient based on the mean interaction, called the F_4, for variety j

$MS_{Ei} = \frac{1}{(v-1)(y-1)}\sum_{j,k} \hat{E}_{ijk}^2$ is the mean square of the interaction terms for note *i*

8. More details on the method and comparison of the method with other methods can be found in Kristensen (2011?).

Example

9. For demonstration a subset of varieties from a DUS experiment with sugar beets was chosen. The notes for hypocotyl colour (Table 1) were analysed. Because some varieties had notes with zero plants in both years, there were difficulties in meeting the requirements mentioned above. Therefore, the varieties *M*, *N*, *O*, *Q*, *R*, *S* and *V* were excluded from the analyses shown here.

10. The estimated percent of plants in each note for each variety are shown in table 2.

11. Treating varieties *A* and *B* as candidates and the remaining varieties *C, D, …, U,* as reference varieties, the F-values and the P-values for testing the hypothesis of no difference between candidate and reference varieties were calculated. The F-values and the P-values are shown in Table 3. The F_3-values and their significances are also shown in Table 3.

12. Using the 1% level of significance as a decision rule for comparing the candidates with the reference varieties, we found that candidate *A* was distinct from 7 of the other varieties, while candidate *B* was distinct from 5 of the other varieties. The largest F_3-values were found for the variety pairs *B*-*K* and *A*-*K*. This seemed to be caused mainly by variety *K*, which had

many green and no red hypocotyls in year 1, but few green and many red hypocotyls in year 2.

Table 1. Number of individual with each note for hypocotyl colours for some varieties in sugar beets

Variety	Colour							
	1 Green		2 White		3-5 Red[1]		7 Orange	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
A	30	21	9	1	15	25	46	53
B	5	9	9	5	48	46	38	40
C	0	3	17	12	31	35	52	50
D	1	0	7	8	71	77	21	15
E	0	3	5	0	80	72	20	25
F	30	28	0	4	30	30	40	38
G	33	25	12	2	16	24	39	49
H	72	76	2	4	3	2	23	18
I	3	2	4	2	37	29	56	67
J	82	82	2	0	7	5	9	13
K	52	7	16	33	0	44	32	16
L	50	37	17	9	5	12	28	42
M	0	0	12	2	58	56	30	42
N	0	0	9	8	74	69	17	23
O	0	0	12	10	58	65	30	25
P	25	22	0	10	17	11	58	57
Q	0	0	0	10	65	64	35	26
R	0	0	0	0	75	55	25	45
S	0	0	6	1	53	61	41	38
T	83	92	5	1	3	1	9	6
U	54	30	12	13	3	4	31	53
V	0	0	6	18	71	63	23	19

1) Sum of three different reddish colours (pink, red and dark red)

Table 2. Estimated percent of plants for each note of each variety

Variety	Colour			
	1 Green	2 White	3-5 Red	7 Orange
A	25.8	3.9	19.8	50.5
B	7.0	6.8	47.2	39.1
C	1.5	14.3	33.0	51.1
D	0.5	7.5	74.2	17.8
E	1.5	1.8	74.7	22.0
F	29.1	1.7	30.1	39.2
G	29.5	5.6	20.1	44.8
H	74.1	2.9	2.5	20.5
I	2.5	2.9	33.0	61.6
J	82.2	0.9	6.0	11.0
K	27.7	29.3	14.0	29.0
L	44.0	12.7	8.0	35.2
P	23.9	3.4	14.1	58.7
Q	88.0	2.5	2.0	7.5
U	41.7	12.8	3.5	42.0

Table 3. Differences and F_3 values together with P-values for relevant pairs of varieties

Variety	Candidate A				Candidate B			
	F	$P_{dif.}$	F_3	P_{F3}	F	$P_{dif.}$	F_3	P_{F3}
A	-	-	-	-	2.34	0.1157	0.50	0.6855
B	2.34	0.1157	0.50	0.6855	-	-	-	-
C	5.70	0.0062	0.57	0.5829	2.06	0.1432	0.02	0.9826
D	6.29	0.0033	0.50	0.6485	2.05	0.1404	0.42	0.7800
E	5.40	0.0063	0.41	0.6601	1.35	0.2866	0.19	0.8542
F	0.52	0.6757	1.20	0.2671	3.20	0.0522	0.50	0.7097
G	0.16	0.9224	0.01	0.9976	2.79	0.0786	0.46	0.7701
H	6.91	0.0036	0.94	0.4998	14.33	<.0001	0.15	0.9024
I	5.44	0.0073	0.24	0.7018	2.27	0.1143	0.24	0.9500
J	10.36	0.0004	0.19	0.8365	17.65	<.0001	0.18	0.9506
K	2.19	0.1361	3.17	0.0405	4.54	0.0189	4.31	0.0071
L	2.02	0.1621	0.11	0.9719	6.55	0.0051	0.64	0.7790
P	0.21	0.8896	1.79	0.0934	2.67	0.0847	0.92	0.4270
T	13.62	<.0001	0.65	0.7695	21.42	<.0001	0.05	0.9946
U	2.34	0.1202	0.52	0.7387	7.38	0.0027	1.18	0.8181

The F_4 values for each variety in the analysis of the hypocotyl colours are shown in Figure 1. The largest F_4 value was found for variety *K*. The value seemed to be extremely large and an explanation for the unusual result should be sought.

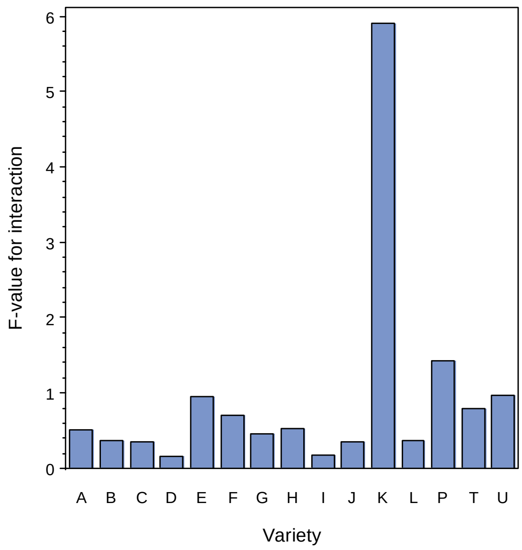


Figure 1: F_4-values for each variety's contribution to the interaction for nominal characteristic hypocotyl colour

Software

13. The procedure GLIMMIX of SAS (SAS Institute Inc., 2010) can be used to estimate the parameters of the generalised linear mixed model, and the data-step facilities (and/or the procedure IML) of the same package can be used for the remaining calculations. However, similar facilities may be found in other statistical packages,

THE COMBINED OVER-YEARS METHOD FOR ORDINAL CHARACTERISTICS

Summary of requirements for application of the method

14. The method is appropriate to use for assessing distinctness of varieties where:

- The characteristic is ordinal and recorded for individual plants (usually recorded visually)
- There are some differences between plants
- The observations are made over at least two years or growing cycles on a single location
- There should be at least 20 degrees of freedom for estimating the random variety-by-year interaction term.
- The distribution of the characteristic should be unimodal, i.e. notes with large number of plants should occur next to each other, zeros at one or both ends of the scale should not cause problems as long as most varieties have plants that fall in different notes
- The total number of plants for each variety should not be too low, at least 5 times the number of notes the variety covers

Summary

15. The method can be considered as an alternative to the χ^2-test for independence in a contingency table. The χ^2-test only takes the variation caused by random sampling into account and may thus be too liberal if additional sources of variation are present. Also the χ^2-test does not take the ordering of the notes into account. The combined over-years method for ordinal characteristics takes other sources of variation into account by including a random

variety-by-year interaction term (as for the COYD method described in TGP/8/1 Part II: 3). It takes the ordering of notes into account by using a cumulative function over the ordered notes. The inclusion of the random effect is expected to decrease the number of distinct pairs of varieties compared to the χ^2-test for independence, but to better ensure that the decisions are consistent over coming years. Taking the ordering of notes into account is expected to increase the power of the test and thus to increase the number of distinct pairs.

16. The method is based on a generalisation of the traditional analyses of variance and regression methods for normally distributed data, which are called "generalized linear mixed models".

17. The combined over-years method for nominal characteristics involves

- Calculating the number of plants for each note for each variety in each of the two or three years of trials, which results in a 3-way table (see the example)
- Analyse the data using appropriate software
- Compare each candidate to the reference varieties and the other candidates at the appropriate level of significance to see which varieties the candidate is distinct from
- Check if the variety-by-year interaction term for distinct pairs is considerably larger than the average for all variety pairs

<u>Technical description of the method</u>

18. The method is based on a generalized linear mixed model using the cumulative logit as link function assuming that the data are multinomial distributed (for more information on generalised linear mixed models see e.g. McCulloch and Searle, 2001 or Agresti, 2002). The model resembles the COYD method for normally distributed characteristics by including the

18. The method is based on a generalized linear mixed model using the cumulative logit as link function assuming that the data are multinomial distributed (for more information on generalised linear mixed models see e.g. McCulloch and Searle, 2001 or Agresti, 2002). The model resembles the COYD method for normally distributed characteristics by including the

$(Y_{1jk}, Y_{2jk}, Y_{3jk}, ..., Y_{njk})$ are multinomial distributed with parameters $(\pi_{1jk}, \pi_{2jk}, \pi_{3jk}, ..., \pi_{njk})$

$$\log\left(\frac{\sum_{l=1}^{i} \pi_{ljk}}{\sum_{l=i+1}^{n} \pi_{ljk}}\right) = \mu_i + \beta_j + \delta_k + E_{jk}$$

where

Y_{ijk} is the number of plants for variety j in year k for note i

μ_i is the effect of note i $(i = 1, 2, 3, ..., n-1)$

β_j is the effect of variety j $(j = 1, 2, 3, ..., v)$

δ_k is the effect of year k $(k = 1, ..., y)$

E_{jk} is the random effect of variety j in year k. E_{jk} is asumed to be normally distributed with zero mean and constant variance, i.e. $E_{jk} \sim N(0, \sigma^2)$

n, v and y are the number of notes, varieties and years, respectively

year×variety interaction as a random effect. The model can be written as:

19. The estimates of the parameters μ_i, δ_k and β_j can be used to estimate the relative number of plants with a given note for each variety, and the differences between the estimates of β_j-β_l can be used to quantify and test the difference between variety *j* and variety *l*. The average relative number of plants for each note and variety can be calculated by the formulas:

First calculate: $\hat{P}_{ij} = \hat{\mu}_i + \hat{\beta}_j + \frac{1}{y}\sum_k^y \hat{\delta}_k$ for $i = 1, 2, ..., n-1$ and each variety, j

$$\text{Then calculate } \hat{\pi}_{ij} = \begin{cases} e^{\hat{P}_{ij}}/(1+e^{\hat{P}_{ij}}) & \text{for } i = 1 \\ e^{\hat{P}_{ij}}/(1+e^{\hat{P}_{ij}}) - e^{\hat{P}_{i-1j}}/(1+e^{\hat{P}_{i-1j}}) & \text{for } i = 2, 3, ..., n-1 \\ 1 - e^{\hat{P}_{i-1j}}/(1+e^{\hat{P}_{i-1j}}) & \text{for } i = n \end{cases}$$

where

$\hat{\pi}_{ij}$ is the average relative number of plants with note i for variety j

20. As a large year×variety interaction for a specific pair of varieties may cause that pair to be distinct, for instance if a very large difference occurs in one of the years but not in other years. To avoid that situation the year×variety interaction for each pair of varieties is compared to the average year×variety pair interaction using the quotient between the mean square for the interaction of the actual pair of varieties and the average interaction of all variety pairs. This quotient is here called F_3. This will result in a quotient (F_3) which it is suggested is tested approximately by assuming that the quotient is F-distributed with *y*-1 and (*y*-1)(*v*-1) degrees of freedom. The F_3 value is calculated as:

$$T_{jl} = \frac{1}{y-1}\begin{cases} (\hat{E}_{j1} - \hat{E}_{j2} - \hat{E}_{l1} + \hat{E}_{l2})^2 & \text{for } y=2 \\ (\hat{E}_{j1} - \hat{E}_{j3} - \hat{E}_{l1} + \hat{E}_{l3})^2 + (\hat{E}_{j1} - 2\hat{E}_{j2} + \hat{E}_{j3} - \hat{E}_{l1} + 2\hat{E}_{l2} - \hat{E}_{l3})^2 & \text{for } y=3 \end{cases}$$

$$F_{3jl} = T_{jl}/\bar{T}_{..}$$

where

F_{3jl} is the quotient, called the F_3 value, for variety pair j, l

$\bar{T}_{..} = \frac{1}{v(v-1)/2}\sum_{j<l} T_{jl}^2$ is the average mean square for all pairs of varieties

21. It may be valuable also to calculate a quotient that can be used to get a measure of how much each variety contributes to the interaction. Such a quotient, called F_4, may be based on the interaction terms. This will result in a quotient (F_4) which it is suggested is tested approximately by assuming that the quotient is approximately F-distributed with (*y*-1) and (*y*-1)(*v*-1) degrees of freedom. The F_4-value is calculated as:

$$F_{4j} = \frac{1}{y-1}\sum_k \hat{E}_{jk}^2 / MS_E$$

where F_{4j} is the quotient, called the F_4 value, for variety j

22. More details on the method and comparison of the methods with other methods can be found in Kristensen (2011?).

<u>Example</u>

23. For demonstration a subset of varieties from a DUS experiment with winter wheat was chosen. The notes for anthocyanin coloration on coleoptiles (Table 4) were analysed.

24. The estimated percent of plants in each note for each variety are shown in table 5.

Table 4. Number of individual plants with each note for anthocyanin coloration on coleoptiles for some varieties in winter wheat

Variety	Note									
	1 absent or very weak		3 weak		5 medium		7 strong		9 very strong	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
A	98	86	1	3	0	0	0	0	0	0
B	4	14	14	65	178	20	0	0	0	0
C	6	0	32	6	56	83	0	4	0	0
D	1	4	5	13	75	82	17	1	1	0
E	84	62	106	19	3	0	0	0	0	0
F	96	100	4	0	0	0	0	0	0	0
G	96	100	4	0	0	0	0	0	0	0
H	77	84	23	16	0	0	0	0	0	0
I	8	4	15	16	55	69	4	1	0	0
J	95	93	3	0	2	0	0	0	0	0

Table 5. Estimated percent of plants for each note of each variety

Variety	Note				
	1 absent or very weak	3 weak	5 medium	7 strong	9 very strong
A	97.9	1.9	0.1	0.0	0.0
B	3.9	36.5	59.1	0.6	0.0
C	1.4	17.8	79.1	1.5	0.1
D	0.4	6.1	88.2	5.1	0.2
E	62.9	33.7	3.4	0.0	0.0
F	98.9	1.1	0.1	0.0	0.0
G	98.9	1.1	0.1	0.0	0.0
H	81.0	17.6	1.4	0.0	0.0
I	2.0	23.1	73.8	1.1	0.0
J	98.6	1.3	0.1	0.0	0.0

25. Treating varieties *A* and *B* as candidates and the remaining varieties *C, D, …, J* as reference varieties, the F-values and the P-values for testing the hypothesis of no difference between candidate and reference varieties were calculated. The F-values and the P-values are shown in Table 6. The F_3-values and their significances are also shown in Table 6.

26. For the data shown here candidate *A* could be separated from 4 of the other varieties when using a 1% level of significance while candidate *B* could be separated from 5 of the

other varieties. The F_3 values were not significantly larger than 1 for any of the tested variety pairs shown in table 3. The largest F_3 was found for the variety pair B-C and seemed to be caused by a stronger anthocyanin coloration of variety B than variety C in year 1 while in year 2 the anthocyanin coloration was stronger in variety C than in variety B. The second largest F_3 was found for the variety pair A-B and here the stronger anthocyanin coloration of variety B in 2007 seemed to be the cause.

Table 6. Differences and F_3 values together with P-values for relevant pairs of varieties

Variety	Candidate A				Candidate B			
	Difference	$P_{Difference}$	F_3	P_{F3}	Difference	$P_{Difference}$	F_3	P_{F3}
A	-	-	-	-	7.06	0.0009	2.47	0.1503
B	7.06	0.0009	2.47	0.1503	-	-	-	-
C	8.11	0.0004	0.38	0.5548	1.04	0.4648	4.78	0.0566
D	9.33	0.0001	1.42	0.2644	2.26	0.1327	0.15	0.7111
E	3.33	0.0471	0.67	0.4353	-3.73	0.0232	0.57	0.4691
F	-0.61	0.7152	1.56	0.2425	-7.68	0.0008	0.10	0.7551
G	-0.61	0.7152	1.56	0.2425	-7.68	0.0008	0.10	0.7551
H	2.41	0.1319	0.21	0.6612	-4.66	0.0079	1.25	0.2920
I	7.77	0.0005	0.03	0.8561	0.71	0.6176	1.92	0.1992
J	-0.40	0.8088	1.68	0.2273	-7.46	0.0009	0.08	0.7882

The F_4 values for each variety in the analysis of anthocyanin coloration on coleoptiles are shown in Figure 2. It is seen that only two varieties have a value larger than 1. The largest F_4 is found for variety *C*.

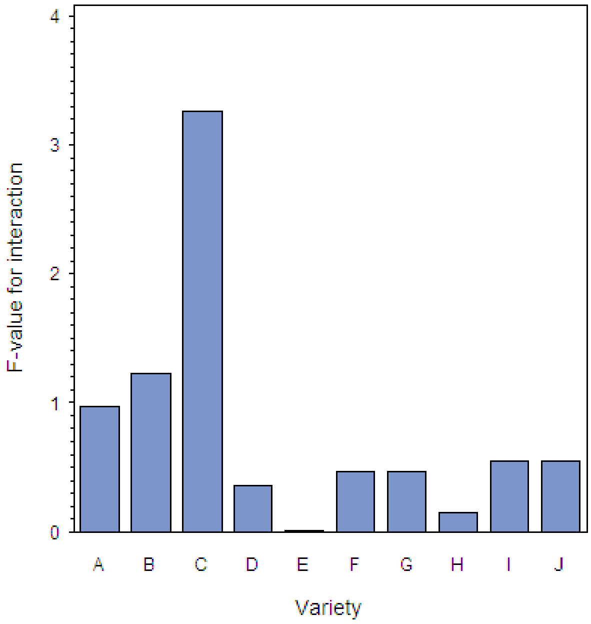


Figure 2 F_4-values for each variety's contribution to the interaction for ordinal characteristic anthocyanin coloration on coleoptiles

Software

27. The procedure GLIMMIX of SAS (SAS Institute Inc., 2010) can be used to estimate the parameters of the generalised linear mixed model, and the data-step facilities (and/or the procedure IML) of the same package can be used for the remaining calculations. However, similar facilities may be found in other statistical packages,

THE COMBINED OVER-YEARS METHOD FOR BINOMIAL CHARACTERISTICS

Summary of requirements for application of the method

28. The method is appropriate to use for assessing distinctness of varieties where:
- The characteristic is recorded for individual plants (usually recorded visually) using a scale with only 2 levels (such as present/absent or similar)
- There are some differences between plants
- The observations are made over at least two years or growing cycles on a single location
- There should be at least 20 degrees of freedom for estimating the random variety-by-year interaction term.
- The expected number of plants for each combination of variety and note should be at least one – and for most of the combinations the number should be at least 5.

Summary

29. The method can be considered as an alternative to the χ^2-test for independence in a contingency table. The χ^2-test only takes the variation caused by random sampling into account and may thus be too liberal if additional sources of variation are present. The combined over-years method for ordinal characteristics take other sources of variation into account by including a random variety-by-year interaction term (as for the COYD method described in TGP/8/1 Part II: 3). The inclusion of the random effect is expected to decrease the number of distinct pairs of varieties compared to the χ^2-test for independence, but to better ensure that the decisions are consistent over coming years.

30. The method is based on generalisation of the traditional analyses of variance and regression methods for normally distributed data, which are called “generalized linear mixed models”.

31. The combined over-years method for binomial characteristics involves
- Calculating the number of plants for each note for each variety in each of the two or three years of trials, which results in a 3-way table
- Analyse the data using appropriate software
- Compare each candidate to the reference varieties and the other candidates at the appropriate level of significance to see which varieties the candidate is distinct from
- Check if the variety-by-year interaction term for distinct pairs is considerably larger than the average for all variety pairs

Technical description of the method

32. The method is based on a generalized linear mixed model using the logit as link function assuming that the data are binomial distributed (for more information on generalised linear mixed models see e.g. McCulloch and Searle, 2001 or Agresti, 2002). However, the binomial distribution is a simplified case of the multinomial distribution and because of there are only two levels there will be no distinction between nominal and ordinal scale. The methods described in section xx.xx and xx.xx for "The Combined Over-Years Method for Nominal Characteristics" and "The Combined Over-Years Method for Ordinal Characteristics", respectively both reduce to the same method for binomial distributed data when only two possible notes are present. Thus for more details on the method, the reader should consult either of those two methods, and the method will not be described further here.

References and literature

Agresti, A., 2002, Categorical data analysis, 2nd edition. Wiley & Sons, Inc. 710 pp.

Kristensen, K. 2011? Analyses of visually accessed data from DUS trials using a combined over years analysis for testing distinctness. Submitted to Biuletyn Oceny Odmian (Cultivar Testing Bulletin).

McCulloch, C.E. and Searle, S.R., 2001, Generalized, Linear, and Mixed Models. John Wiley & Sons, Inc; New York. 325 pp.

SAS Institute Inc. 2010, SAS/STAT® 9.22 User's Guide. Cary, NC: SAS Institute Inc. 8460 pp. (online access: **http://support.sas.com/documentation/cdl/en/statug/63347/PDF/default/statug.pdf**, accessed 15th November 2010)

[Annex XI follows]

ANNEX XI

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

New Section - Guidance for the development of variety descriptions (Drafter to be agreed)

Notes

1. Comments: the TC at its forty-sixth session requested that, in the revision of document TGP/8, consideration should be given to guidance on the development of variety descriptions with information from:

(i) more than one growing cycle in one location, and
(ii) more than one location

2. In the establishment of guidance for the development of variety descriptions, the Technical Working Parties (TWPs) are invited to consider the discussions at the CAJ in respect to the status and use of the "official" variety description (see document CAJ/61/8, paragraphs 1, 2 and 6 and the examples provided in the annexes to this document)

3.

	Comments of the TWPs in 2011	
General	The TWA and the TWO noted the information provided in Annex XI.	TWA TWO
	The TWC considered that, for the time being, there was no information to develop guidance for the development of descriptions with information from more than one location.	TWC
	The TWV and the TWF agreed that the experts from the Netherlands should draft guidance on the development of variety descriptions with information from more than one growing cycle in one location and more than one location.	TWV TWF

[Annex XII follows]

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

Section 4 – 2x1 % Method - Minimum number of degrees of freedom for the 2x1% Method (Drafter: Sally Watson)

Notes

The TWC at its twenty-seventh session proposed to are commendation in the number of degrees of freedom for the 2x1% Method of at least 10, and preferably at least 20, degrees of freedom. The TC at its forty-sixth session agreed not to include the recommendation in document TGP/8/1 and that the proposal of the TWC be further discussed for a future revision of document TGP/8.

[DRAFT TEXT FOLLOWS]

4. 2x1% METHOD
5.

	Comments of the TWPs in 2011	
General	The TWA, the TWO and the TWF noted the information provided in Annex XII.	TWA TWO TWF
	The TWC agreed that the explanation proposed in Annex XII should be included in document TGP/8.	TWC

4.1 Requirements for application of method

4.1.1 The 2x1% Criterion is an appropriate method for assessing the distinctness of varieties where:

– the characteristic is quantitative;

– there are some differences between plants (or plots) of a variety;

– observations are made on a plant (or plot) basis over two or more years;

– there are at least 10, and preferably at least 20, degrees of freedom for the residual mean square used to estimate the standard error in the t-test in each year;

– To have replicated plots

	Comments of the TWPs in 2011	
4.1.1	The TWV noted that at least 10 degrees of freedom were required for the residual mean square used to estimate the standard error in the t-test in each year. The TWV proposed that further clarification was needed with regard to the significance of the wording "preferably at least 20 degrees of freedom".	TWV

4.2 The 2x1% Criterion (Method)

4.2.1 For two varieties to be distinct using the 2x1% criterion, the varieties need to be significantly different in the same direction at the 1% level in at least two out of three years in one or more measured characteristics. The tests in each year are based on Student's two-tailed t--test of the differences between variety means with standard errors estimated using the residual mean square from the analysis of the variety x replicate plot means.

4.2.2 With respect to the 2x1% criterion, compared to COYD, it is important to note that:

- Information is lost because the criterion is based on the accumulated decisions arising from the results of t-tests made in each of the test years. Thus, a difference which is not quite significant at the 1% level contributes no more to the separation of a variety pair than a zero difference or a difference in the opposite direction. For example, three differences in the same direction, one of which is significant at the 1% level and the others at the 5% level would not be regarded as distinct
- Some characteristics are more consistent over years than others in their expression of differences between varieties. However, beyond requiring differences to be in the same direction in order to count towards distinctness, the 2x1% criterion takes no account of consistency in the size of the differences from year to year.
- It is recommended that there should be at least 10, and preferably at least 20, degrees of freedom for the residual mean square used to estimate the standard error in the t-test in each year. This is in order to ensure that the residual mean square is based on sufficient data to be a reliable estimate of the varieties-by-replicates variation used in the standard error in the t-test. Assuming replicates are arranged in blocks, 20 degrees of freedom corresponds to 11 varieties in three replicates, or 5 varieties in six replicates, whereas, ten degrees of freedom corresponds to 6 varieties in three replicates, or 3 varieties in six replicates.

The fewer the degrees of freedom for the residual mean square below 20, the greater the loss in precision in the estimate of the standard error in the t-test. This is compensated for by the critical t-value used in the t-test being larger, which results in a reduction in the power of the test: meaning that there is a reduced chance of declaring varieties as being distinct.

[Annex XIII follows]

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

Section 9 - The Combined-Over-Years Uniformity Criterion (COYU) - Minimum number of degrees of freedom for COYU (Drafter: Sally Watson)

Notes

The TWC at its twenty-seventh session proposed to change the recommendation in the minimum number of degrees of freedom for COYU to "there should be at least 10, and preferably at least 20, degrees of freedom for the varieties-by-years mean square in the COYD analysis of variance, or if there are not, then Long-Term COYD can be used". The TC at its forty-sixth session agreed to maintain the previous recommendation of 20 degrees of freedom and to consider the proposal of the TWC for a future revision of document TGP/8.

	Comments of the TWPs in 2011	
General	The TWA, the TWO and the TWF noted the information provided in Annex XIII.	TWA TWO TWF
	The TWC agreed that the explanation proposed in Annex XIII should be included in document TGP/8.	TWC

[PROPOSED REVISED TEXT]

3.1 Summary of requirements for application of method

COYD is an appropriate method for assessing the distinctness of varieties where:

– the characteristic is quantitative;

– there are some differences between plants (or plots) of a variety;

– observations are made on a plant (or plot) basis over at least two years or growing cycles, and these should be carried out at a single location;

– there should be at least 10, and preferably at least 20 degrees of freedom for the varieties-by-years mean square in the COYD analysis of variance, or if there are not, then Long-Term COYD can be used (see 3.6.2 below);

[…]

3.5 Use of COYD

3.5.1 COYD is an appropriate method for assessing the distinctness of varieties where:

– the characteristic is quantitative;

– there are some differences between plants (or plots) of a variety;

– observations are made on a plant (or plot) basis over two or more years;

– There should be at least 10, and preferably at least 20 degrees of freedom for the varieties-by-years mean square in the COYD analysis of variance, or if there are not, then Long-Term COYD can be used (see 3.6.2 below);

The reason for this recommendation is to ensure that the varieties-by-years mean square is based on sufficient data to be a reliable estimate of the varieties-by-years variation for the LSD. Twenty degrees of freedom corresponds to 11 varieties common in three years of trials, or 21 varieties common in two years, whereas, ten degrees of freedom corresponds to 6 varieties common in three years of trials, or 11 varieties common in two years. Trials with fewer varieties in common over years are considered to have small numbers of varieties in trial. The fewer the degrees of freedom for the residual mean square below 20, the greater the loss in precision in the estimate of the varieties-by-years variation used in the LSD. This is compensated for by the critical t-value, t_p, used in the LSD being larger, which results in a reduction in the power of the test: meaning that there is a reduced chance of declaring varieties as being distinct.

[…]

~~3.6.2.2~~1. In trials with small numbers of varieties the variety-by-year tables of means can be expanded to include means for earlier years, and if necessary, other established varieties. As not all varieties are present in all years, the resulting tables of variety-by-year means are not balanced. Consequently, each table is analyzed by the least squares method of fitted constants (FITCON) or by REML, which produces an alternative varieties-by-years mean square as a long-term estimate of variety-by-years variation. This estimate has more degrees of freedom as it is based on more years and varieties.

[…]

3.7 Implementing COYD

COYD is an appropriate method for assessing the distinctness of varieties where:

– the characteristic is quantitative;

– there are some differences between plants (or plots) of a variety;

– observations are made on a plant (or plot) basis over two or more years;

– There should be at least 10, and preferably at least 20 degrees of freedom for the varieties-by-years mean square in the COYD analysis of variance, or if there are not, then Long-Term COYD can be used (see 3.6.2 above) ;

The COYD method can be applied using TVRP module of the DUST package for the statistical analysis of DUS data, which is available from Dr. Sally Watson (Email: *info@afbini.gov.uk*) or from *http://www.afbini.gov.uk/dustnt.htm.* Sample outputs are given in Part II section 3.10.

	Comments of the TWPs in 2011	
3.7	The TWV agreed that it would be necessary to provide data in support of the proposal to reduce the minimum degrees of freedom for the varieties-by-years mean square in the COYD analysis of variance from 20 to 10. The TWV agreed that the following wording in Section 3.1 "Summary of requirements for application of method" should be amended because it meant that Long-Term COYD could be used with less than 10 degrees of freedom: "- there should be at least 10, and preferably at least 20, degrees of freedom for the varieties-by-years mean square in the COYD analysis of variance, or if there are not, then Long-Term COYD can be used (see 3.6.2 below);"	TWV

[…]

[Annex XIV follows]

TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION

Section 10 – Minimum number of comparable varieties for the Relative Variance Method (Drafter: Nik Hulse (Australia)).

Notes

The TC at its forty-sixth session agreed that a recommendation on the minimum number of comparable varieties to be included in the trial in the Relative Variance Method be included in a revision of document TGP/8.

[DRAFT TEXT FOLLOWS]

MINIMUM NUMBER OF COMPARABLE VARIETIES FOR THE RELATIVE VARIANCE METHOD

	Comments of the TWPs in 2011	
General	The TWA considered Annex XIV. The expert from Germany noted that according to TGP/10, comparable varieties should be considered for assessing uniformity, and according to TGP/9, similar varieties for assessing distinctness. There was no agreement on this proposal from the expert from Australia. The TWA recommended that the TWC provide guidance on the adequate sample size of comparable varieties to be used in order to correctly assess uniformity.	TWA
	The TWC received a presentation from Mr. Nik Hulse (Australia). The TWC conditionally agreed with the proposal made by Australia. Doubts were expressed regarding some assumptions of the method and further investigation will be done by Australia with respect to these assumptions and the F value used in the calculations.	TWC
	The TWV, the TWO and the TWF noted the comments made by the TWA and TWC concerning the minimum number of comparable varieties for the Relative Variance Method.	TWV TWO TWF

Note: Uniformity assessment on the basis of the Relative Variance method is set out in Chapter 10 of TGP/8/1. The first two paragraphs of 10.1 should be numbered 10.1.1 and 10.1.2 respectively and it is proposed that the following text is inserted;

10.1.3 Chapter 5 of the document "Examining Uniformity", TGP/10/1 explains that where it is not possible to visualize off-types then a comparison is made to comparable varieties as follows;

> "5.1 The General Introduction, Chapter 6.4, explains that, in cases where there is a high levell of variation in the expressions of characteristics for the plants within a variety, it is not possible to visualize which plants should be considered as off-types and the off-type approach for the assessment of uniformity is not appropriate. It clarifies that in such cases, uniformity can be assessed by considering the overall levell of variation, observed across all the individual plants, to determine whether it is similar to comparable varieties. In this approach, relative tolerance limits for the levell of variation are set by comparison with comparable varieties, or types, already known ("standard deviations approach"). The standard deviations approach means that a candidate variety should not be significantly less uniform than the comparable varieties."

10.1.4 In many situations relatively large scale trials are conducted with a large number of comparable varieties. In these cases an approach such as COYU may be considered appropriate. However, in trials where the number of available comparable varieties is typically low the Relative Variance method may be used.

10.1.5 For example, Chapter 7 of TGP/8/1 describes the Match approach and the varieties included in the trial as follows;

> "7.2.3 The Match method typically involves relatively small scale trials where the number of varieties in the trials is limited to the candidate varieties and the most similar varieties of common knowledge."

10.1.6 Comparable varieties can be considered to be those that are similar in their relevant characteristics to the candidate variety and are sufficiently uniform. Consequently, the number of comparable varieties used for examining uniformity is determined by the number of similar varieties included in the trial for the purpose of examining distinctness.

10.1.7 Other varieties may be included in the trial for reasons other than that they are the most similar varieties to the candidate. For example, check or example varieties may be included to verify the expression of particular characteristics. The DUS examiner can exclude these as comparable varieties in the examination of uniformity.

[Annex XV follows]

Title of document		2011						2012						2013					
		TC-EDC	TC/47	CAJ/63	TWPs	CAJ/64	C/45	TC-EDC	TC/48	CAJ/65	TWPs	CAJ/66	C/46	TC-EDC	TC/49	CAJ/67	TWPs	CAJ/68	C/47
TGP/8 PART I: DUS TRIAL DESIGN AND DATA ANALYSIS																			
Annex I	New Section 2 - Data to be recorded *(Drafter: Mr. Uwe Meyer (Germany))*				x						x								
Annex II	New Section 3 - Control of variation due to different observers *(Drafter: Mr. Gerie van der Heijden (Netherlands))*				x						x								
Annex III	New Section 6 – Data processing for the assessment of distinctness and for producing variety descriptions *(Drafters: experts from Finland, France, Germany, Japan, Kenya and the United Kingdom)*				x						x								
Annex IV	New Section – Information of good agronomic practices for DUS field trials (Drafters: Mrs. Anne Weitz (European Union) and Argentina and France to contribute[1])				x						x			Sections that will go for adoption by the Council in 2013	Sections that will go for adoption by the Council in 2013	Sections that will go for adoption by the Council in 2013	Sections that will go for adoption by the Council in 2013	Sections that will go for adoption by the Council in 2013	Adoption of TGP/8/2
TGP/8 PART II: TECHNIQUES USED IN DUS EXAMINATION																			
Annex V	Section 1: The GAIA Methodology				x						x								
	New Section after Section COYU Statistical Methods for Very Small Sample Sizes *(Drafter: Mr. Gerie van der Heijden (Netherlands))*				x						x								
Annex XII	Section 4 – 2x1 % Method - Minimum number of degrees of freedom for the 2x1% Method (Drafter: Mrs. Sally Watson (United Kingdom)[2])				x						x								
	Section 5: Pearson's Chi Square Test Applied to Contingency Tables				x						x								
Annex XIII	Section 9 - The Combined-Over-Years Uniformity Criterion (COYU) - Minimum number of degrees of freedom for COYU (Drafter: Mrs. Sally Watson (United Kingdom)[3])				x						x								
Annex XIV	Section 10 – Minimum number of comparable varieties for the Relative Variance Method *(Drafter: Nik Hulse (Australia))*				x						x								
Anex VI	New Section 11 Examining DUS in bulk samples *(Drafter: Mr. Kristian Kristensen (Denmark))*				x						x								
Annex VII	New Section 12 - Examining characteristics using image analysis *(Drafter: Mr. Gerie van der Heijden (Netherlands))*				x						x								
Annex VIII	New Section 13 - Methods for data processing for the assessment of distinctness and for producing variety descriptions: (Drafters: Finland, France, Germany, Japan, Kenya and the United Kingdom)				x						x								
Annex IX	New Section - Guidance of data analysis for blind randomized trials *(Drafter: France[4] and Israel[5] to provide examples)*				x						x								
Annex X	New Section - Statistical methods for visually observed characteristics *(Drafter: Denmark, France and the United Kingdon[2])*				x						x								
Annex XI	New Section - Guidance for the development of variety descriptions *(Drafter to be agreed)*				x						x								

[1] Offer made at the 39th TWA Session
[2] Agreed by the TWC at its 28th Session
[3] Agreed by the TWC at its 28th Session
[4] Agreed at the TWZ at its 39th Session
[5] Agreed at the TWV at its 44th Session

[Endnotes follow]

ENDNOTES

---

[a] TWC: to be edited
[b] TWC: to be edited
[c] Rewording proposed by TWC
[d] Rewording proposed b TWC

[End of Endnotes and of document]