UPOV

International Union for the Protection of New Varieties of Plants

E

Technical Working Party on Automation and Computer Programs **TWC/37/5**

Thirty-Seventh Session
Hangzhou, China, October 14 to 16, 2019

Original: English
Date: September 30, 2019

RISKS ASSOCIATED WITH ASSESSMENT OF UNIFORMITY BY OFF-TYPES ON THE BASIS OF MORE THAN ONE GROWING CYCLE

Document prepared by experts from the United Kingdom

INTRODUCTION

1. The report of the Technical Working Party on Automation and Computer Programs (TWC) at its thirty-sixth session (see document TWC/36/15 "Report") concluded:

> "54. The TWC considered document TWC/36/7 'Risks associated with assessment of uniformity by off-types on the basis of more than one growing cycle' and received a presentation by experts from Germany and the United Kingdom, a copy of which would be provided as document TWC/36/7 Add..
>
> "55. The TWC agreed to invite the experts from Germany and the United Kingdom to develop examples demonstrating the risks and consequences for decisions on uniformity to be presented at its next session.
>
> "56. The TWC noted the importance of considering the risks associated with assessment of uniformity by off-types on the basis of more than one growing cycle and agreed that it would not be practical to develop tables with the allowed number of off-types for such cases. The TWC noted that, in future, software might be needed to calculate such risks."

2. A calculator based in Excel is presented. This enables the calculation of both appropriate thresholds in terms maximum allowable numbers of off-types as well as statistical risks. It works for both single and two-cycle systems for off-type assessment. At this stage it does not work for sequential approaches based on sub-samples within a single test or trial.

3. The text below and the software assumes that there is only one submission of plant material, even when uniformity is assessed over two cycles.

THE IMPORTANCE OF CONSIDERING RISKS WHEN ASSESSING UNIFORMITY BY OFF-TYPES

4 Population standards for off-types are given in Test Guidelines, defining the maximum allowable proportions of off-types.

5 The population standard is the maximum level of off-types in all individuals of variety. This is a hypothetical concept since we cannot assess all individuals. Instead, we examine a sample of individuals, with the number depending on circumstances and indicated in the Test Guidelines.

6. As we only look at a relatively small number of individuals, the proportion of off-types we see may not reflect perfectly the proportion in the variety.

7. For example, if the proportion of off-types in the variety is 5% and the sample size is 500 plants, we might find the following numbers of off-types on different occasions (random simulation): 29, 19, 21, 27, 30, 29, 32, 28, 21 or 22.

8. Some of these numbers are more than 5% of 500 plants, others less. If we were to set the maximum number of off-types to 25 (5% of 500), approximately half the time we would expect a variety with 5% off-types to fail this uniformity criterion.

9. We should allow for this natural sampling variability in the way that we assess varieties for off-types. We can do this by understanding the risks associated with the decisions.

TYPE OF RISKS TO CONSIDER

10. From a statistical viewpoint, we should consider two types of risks in particular:

Type I error: The conclusion based on the test statistic, i.e. from the DUS trial, is that a variety is not uniform, when it would be uniform if the trial could be repeated indefinitely.

Type II error: The conclusion based on the test statistic, i.e. from the DUS trial, is that a variety is uniform, when it would be non-uniform if the trial could be repeated indefinitely.

11. We can alter the chances of these two types of errors by the way we assess off-types. For a single cycle assessment, we can do this by changing the maximum number of off-types that is acceptable or by changing the sample size. The probabilities of these types of errors can be calculated using mathematical formulae.

12. In general, the maximum number of off-types is set to ensure that chance of type I errors is sufficiently low. In most cases, we aim to have the chance of type I errors being less than 5%. As the chance of type I errors is 1 minus the acceptance probability, this equates to an acceptance probability of 95%.

13. Given the sample size, the population standard and the acceptance probability, for a single cycle assessment we can identify the maximum number of off-types. These can be found in tables in document TGP/8/3 “Trial Design and Techniques used in the Examination of Distinctness, Uniformity and Stability” Part II Section 8. For example, the maximum number of off-types with a sample size of 500, a population standard of 5% and an acceptance probability of 95% is 33. Note that 33 off-types in 500 plants is 6.6%, a proportion that is larger than the population standard. So the approach gives a margin of safety for the candidate variety according to the sample size.

14. If the test is set up this way, increasing the sample size tends to reduce the chance of type II errors. For example, with a sample size of 500, a population standard of 5% and an acceptance probability of 95%, the chance of a variety with 10% off-types having an acceptable number of off-types (type II error rate) is just 0.5%. However if the sample size is reduced to 50 plants so that the maximum acceptable number is now 5, then the type II error rate increases to 63%.

15. Guidance on sample sizes can be found in Test Guidelines. General guidance can be found in TGP/8/3 on statistical aspects of assessment of uniformity by off-types.

CONSIDERING RISKS WHEN UNIFORMITY BY OFF-TYPES IS ASSESSED ON BASIS OF MORE THAN ONE GROWING CYCLE

16. When examining off-types over two or more cycles, it is necessary to consider which risks are most important when setting thresholds in each cycle. This is particularly true for approaches 1 and 2 as set out in Annex I of document TWP/1/17 Rev.. There are two key choices:

a) To consider risks in each cycle separately;
b) To consider risks for the complete decision process over multiple cycles.

17. The advantage of following a) is that it is easier to work out the maximum number of off-types required for each cycle to achieve the required acceptance probability; tables in TGP/8/3 will give these. For b), tables have not yet been produced.

18. However, b) has the advantage that the selected acceptance probability will be achieved for the whole test, which will have the effect of reducing the chance of type II errors. The choice should lead to greater harmonisation in standards of off-types testing across member states. This would be irrespective of whether off-types are assessed in one cycle or more than one cycle, and is independent of approach. This is because risks are balanced appropriately for the overall decision on uniformity.

19. For example, let consider approach 1 with 50 plants in each cycle, a 1% population standard with acceptable probability of 95%. [In approach 1, the two cycles are assessed separately. If there is inconsistency in the verdicts for the two cycles, a third cycle is assessed]. If the risks are considered separately for each cycle then the maximum number of off-types in each cycle would be set at 2. If the overall risks are considered, the maximum number is 1. The overall chance of declaring a variety with 5% off-types as uniform is 56% when the maximum number is 2 and 19% when it is 1. So setting the number of off-types allowed based on the acceptance probability for a single cycle results in a much higher overall chance of type II errors.

20. Software has now been developed in Excel that enables examination of risks as provided in the Annex II to this document.

PROPOSAL

21. Whichever approach is used, when uniformity is assessed over two or more cycles, it is proposed that the thresholds for the acceptable number of off-types should be chosen to give the required overall acceptance probability (or equivalently, the required type I error over all cycles as this is 1 minus the acceptance probability). This will lead to greater harmonisation of standards, whatever approach is used.

SOFTWARE

22. Software has been developed in Excel. The Excel workbook uses no macros so should be easy to distribute with low security concerns.

23. The software can calculate the two types of statistical risk for either a one-cycle or a two-cycle situation. For this, it requires the acceptance probability, the population standard, the sample size (number of individuals examined) in each cycle, and the maximum allowable numbers of off-types (for two-cycles, this is per cycle and possibly for the combined sample depending on the approach).

24. The software can also suggest the thresholds in terms of maximum allowable numbers of off-types. This requires the acceptance probability, the population standard and the sample size in each cycle. In the case of two-cycles, the thresholds are calculated to achieve **either** the overall acceptance probability **or** the acceptance probability for individual cycles.

25. The software is meant to facilitate discussion on setting up of off-type tests, either using overall cycles or individual cycle acceptance probabilities. It should also be a useful tool for crop experts.

26. The authors would be happy for the software to be freely distributed. However, it would be prudent for it to be tested by an independent party.

27. A demonstration of the software will be presented at the thirty-seventh session of the TWC.

28. The TWC is invited to:

(a) note that a software was developed in Excel that enables examination of risks, as provided in Annex II to this document; and

(b) consider the proposal that the thresholds for the acceptable number of off-types should be chosen to give the required overall acceptance probability, when uniformity is assessed over two or more cycles, as set out in paragraph 21 of this document.

Adrian Roberts, BioSS, United Kingdom
Sally Watson, AFBI, United Kingdom

[Annex I follows]

Assessing uniformity by off-types
Calculator for number of off-types and risks

Adrian Roberts
BioSS
UK

Sally Watson
AFBI
UK

Background

- Development of guidance for uniformity by off-types based on two growing cycles
- Document TWC/36/7 looked at risks
- Clear that software would help clarify the merits of different approaches and also be useful for experts

Software

Here we present new software
- Based in Excel (no macros)

Calculates:
- Statistical risks
- Maximum number of off-types allowed

It works for:
- One cycle
- Two cycles
- But not sub-samples (sequential sampling)

One-cycle test

- Count the number of off-types in a sample of individuals
- If the number exceeds a threshold, variety is not uniform
- Threshold depends on:
  - Number of individuals (TGs)
  - Acceptance probability (later, TGP/8 and TGs)
  - Population standard ((later, TGP/8 and TGs)

One-cycle test

- Count the number of off-types in a sample of individuals
- If the number exceeds a threshold, variety is not uniform
- Threshold depends on:
  - Number of individuals (TGs) e.g. 250
  - Acceptance probability (later, TGP/8 and TGs) e.g. 95%
  - Population standard ((later, TGP/8 and TGs) e.g. 1%

= 5

Two-cycle test

- Basic scheme is two growing cycles
  - Counts of off-types in each cycle
  - Number of individuals same in each cycle (more or less)
  - Three approaches
- Two of the approaches
  - Look at each cycle separately
  - differ in how they deal with conflicting results between cycles
- Third approach
  - Combines the information from both cycles
  - Then similar to one-cycle

Approach 1

Looks at each cycle separately

Compare number of off-types in a cycle to the cycle threshold

Third growing cycle in the case of inconsistent results

Approach 2

Looks at each cycle separately

Compare number of off-types in a cycle to the cycle threshold

Combining the results of two growing cycles in the case of inconsistent results

Approach 3

Combining the results of two growing cycles

Example

Population Standard = 1%						
Acceptance Probability ≥ 95%						
Sample Size in each of growing cycles 1 and 2 = 50						
Maximum number of Off-Types = 2						
Sample Size in growing cycles 1 and 2 combined = 100						
Maximum number of Off-Types = 3						
		Growing cycle		Decision		
		First	Second	Approach 1	Approach 2	Approach 3
	Number of Off-Types	1	1	uniform	uniform	uniform ←consistent
		2	2	uniform	uniform	non-uniform ←inconsistent
		0	3	third growing cycle	uniform	uniform
		1	3	third growing cycle	non-uniform	non-uniform
		0	10*	third growing cycle†	non-uniform†	non-uniform†
		10**	0	third growing cycle*†	non-uniform*†	non-uniform*†

† Care is needed when considering results that were very different in each of the growing cycles, such as when a type of off-type was observed at a high level in one growing cycle and was absent in another growing cycle.
* A variety may be rejected after a single growing cycle if the number of off-types found is sufficiently high.

Statistical risks

Type I error: declare variety non-uniform when population is uniform

Type II error: declare variety uniform when population is non-uniform

Terminology

Population standard: the required maximum proportion of off-types in a variety

Terminology

Population standard: the required maximum proportion of off-types in a variety

Acceptance probability: probability of accepting a variety with the population standard of off-types

Terminology

Population standard: the required maximum proportion of off-types in a variety

Acceptance probability: probability of accepting a variety with the population standard of off-types

- The acceptance probability depends on the way that off-types are assessed
- Given the sample size, the threshold is set so that the acceptance probability is exceeded.

Statistical risk

Tests are set up to achieve a set type I error

- Type I error = 1 – acceptance probability
- 5% in example

Different test can then be compared through the type II errors

- Type II errors are calculated at different levels of off-types in population
- e.g. 2 , 5 and 10 times the population standard

Overall vs stagewise errors

Can set type I error for each stage or growing cycle or for the overall test

We claim that it is better to use overall error → better reflects true risks for applicant & testing authority

Note – not important for single cycle test

Example

pop std=1%, accept prob=95%, no per cycle =50

Approaches 1 & 2 have type I error set to 5% per cycle **not overall**

Approach	Type I error	Type II error			Max off-types	
		2%	5%	10%	Per cycle	Combined
1	0.06%	98.3%	56.1%	3.47%	2	n/a
2	1.27%	89.3%	32.6%	1.39%	2	3
3	1.84%	85.9%	25.8%	0.78%	n/a	3

Approach 3 has the lowest type II errors

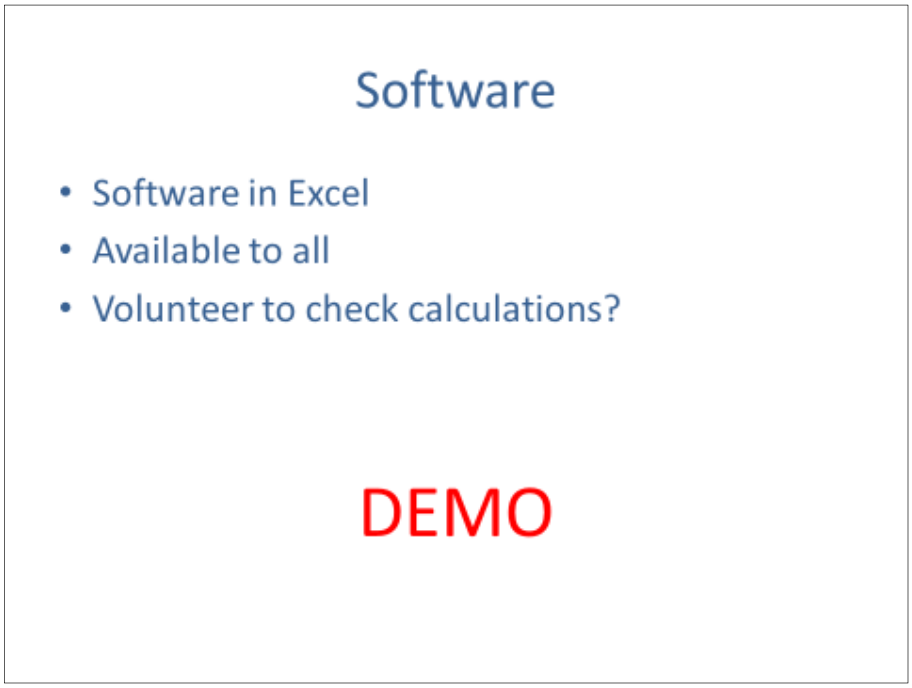
Example revisited

pop std=1%, accept prob=95%, no per cycle =50

Approaches 1 & 2 now have type I error set to 5% **overall**

Approach	Type I error	Type II error			Max off-types	
		2%	5%	10%	Per cycle	Combined
1	2.26%	82.7%,	19.1%	0.33%	1	n/a
2	1.84%	85.9%	25.8%	0.79%	1	3
3	1.84%	85.9%	25.8%	0.78%	n/a	3

Approach 1 has the lowest type II errors; approach 2 and 3 are not far behind



[Annex II follows]

ANNEX II

Introduction

This workbook is intended to help in the setting up of tests for assessing off-types, either in one cycle or over two cycles.

Given the sample size in terms of numbers of plants, the population standard and the acceptance probability, the workbook will calculate the maximum allowable number of off-types (per cycle). The population standard and acceptance probability are defined and explained in TGP/8, and are set for each crop in the Test Guidelines.

The workbook also calculates statistical risks associated with the chosen test, in terms of type I and type II errors -see below.

How to use the calculators

For single cycle tests, there are two calculator sheets:

"Single cycle - Optimise". Here you can modify the blue cells according to the test that you intend to employ (excluding the maximum number of off-types). The sheet calculates the number of off-types threshold that achieves the given acceptance probability. The sheet also gives the actual achieved acceptance probability, since it is not usually possible to exactly achieve the required probability.

"Single cycle - Risks". Here you can modify the blue cells according to the test that you intend to employ (including the maximum number of off-types). The sheet calculates the type I and II error probabilities. For this sheet, you must define the maximum number of off-types you wish to apply. This can be derived using TGP/8 or the following sheet. The chance of type II errors is for a specified multiple of the population standard (see below and TGP/8 for more details, typical values are 2, 5 and 10). So for example, in the case of a multiple of 5 and a population standard of 1%, the calculation is the chance that a variety with 5% off-types passes the assessment.

For two cycle tests, there are two calculator sheets:

"Two cycles - Optimise". Here you can modify the blue cells according to the test that you intend to employ (excluding the maximum number of off-types). The sheet calculates the number of off-types threshold that achieves the given acceptance probability. The sheet also gives the actual achieved acceptance probability, since it is not usually possible to exactly achieve the required probability. Two methods are used for calculating the off-type thresholds: firstly, based on the overall statistical risks - this is recommended by the authors; secondly, based on individual cycle risks - this carries a larger type II error risk and is not applicable to approach 3. Note that calculations for approach 2 currently only work for a maximum number of offtypes of 200.

"Two cycles- Risks". Here you can modify the blue cells according to the test that you intend to employ (including the maximum number of off-types in each cycle and in the combined sample if relevant). The sheet calculates the type I and II error probabilities. For this sheet, you must define the maximum number of off-types you wish to apply, for each cycle and for the combined sample (relevant for approaches 2 and 3). This can be derived using TGP/8 or the following sheet. The chance of type II errors is for a specified multiple of the population standard (see below and TGP/8 for more details, typical values are 2, 5 and 10). So for example, in the case of a multiple of 5 and a population standard of 1%, the calculation is the chance that a variety with 5% off-types passes the assessment. The sheet also gives risks for individual cycles. Note that calculations for approach 2 currently only work for a maximum number of offtypes of 200.

Functionality to add in future

Sequential tests as described in the draft text section 4.8 for TGP/8.

Type I and type II errors

The type I error is the chance that off-type assessment based on a sample indicates that the variety is non-uniform when the proportion of off-types in the variety (as opposed to the sample) is less than or equal to the population standard.

The type II error is the chance of the off-type assessment indicates that the variety is uniform when the proportion of off-types in the variety (as opposed to the sample) is greater than the population standard. We usually calculate the chance in a specific circumstance, such as when the proportion of off-types is 5 times the population standard.

Key assumptions

a) Samples are from a single submission of seed/plants.
b) The submission is representative of the variety.
c) That enviromental effects do not affect the off-type assessment.

About

Authors: Adrian Roberts (BioSS) & Sally Watson (AFBI)

Version: 0.4

Date: 22 July 2019

Password for unprotecting sheets: offtypes
Note there are hidden sheets with calculations. We recommend that these are kept hidden unless you wish to review the underlying calculations. There are no macros.

This sheet allows you to find the optimum number of off-types for single-cycle assessment

Acceptance probability =	95%
Population standard =	1%
Sample size (n) =	250

Blue cells can be changed

max no. of off-types	Achieved overall acceptance probablity
5	95.88%

⇧
required results

This sheet gives the type I and type II risks/errors for single cycle assessment, ***given the maximum numbers of off-types***

Population standard (p)=	1%	**Blue cells can be changed**
Sample size (n) =	50	
Multiple of population standard that Type II errors are evaluated at =	5	

		Achieved ACCEPTANCE PROBABILITY	**TYPE I ERROR**	**TYPE II ERROR**	
Single-cycle probabilities		Probability accept variety with Population Standard of off-types, ie with 1% of off-types	Probability reject variety with Population Standard of off-types, ie with 1% of off-types	Probability accept variety with 5 x Population Standard of off-types, ie with 5% off-types	
Max no. of off-types	5	100.00%	0.00%	96.22%	high risk test

This sheet allows you to find the optimum number of off-types for a two-cycle assessment with each approach

Acceptance probability = 95%
Population standard = 1%
Per cycle sample size (n) = 50

Blue cells can be changed

Based on optimising to overall acceptance probability (**recommended**)

	maximum number of off-types per cycle	maximum number of off-types for combined sample	Achieved overall acceptance probablity
Approach 1	**1**	n/a	97.74%
Approach 2	**1**	**3**	98.16%
Approach 3	n/a	**3**	98.16%

⇧
required results

Based on optimising to acceptance probability for each cycle

	maximum number of off-types per cycle	maximum number of off-types for combined sample	Achieved overall acceptance probablity
	2	n/a	99.94%
	2	3	98.73%

⇧
required results

Two cycle approaches
See future revision of TGP/10 for full details. Note results from growing cycles using different lots of plant material should not be combined.

Approach 1:
A variety is considered uniform if it is within the uniformity standard in both of the two growing cycles.
A variety is considered non-uniform if it fails to meet the uniformity standard in both of the two growing cycles.
If at the end of the two growing cycles the variety is within the uniformity standard in one growing cycle but is not within the uniformity standard in the other growing cycle, then uniformity is assessed in a third growing cycle. If in the third growing cycle the variety is within the uniformity standard, the variety is considered uniform. If in the third growing cycle the variety fails to meet the uniformity standard, the variety is considered non-uniform.

Approach 2:
A variety is considered uniform if it is within the uniformity standard in both of the two growing cycles.
A variety is considered non-uniform if it fails to meet the uniformity standard in both of the two growing cycles.
If at the end of the two growing cycles the variety is within the uniformity standard in one growing cycle but is not within the uniformity standard in the other growing cycle, then the numbers of off-types are combined over the two cycles and compared to threshold corresponding to the combined sample size.

Approach 3:
A variety is considered uniform if the total number of off-types at the end of the two growing cycles does not exceed the number of allowed off-types for the combined sample.

A variety is considered non-uniform if the total number of off-types at the end of the two growing cycles exceeds the number of allowed off-types for the combined sample.

This sheet gives the type I and type II risks/errors for two-cycle assessment with approaches 1, 2 and 3, ***given the maximum numbers of off-types***

Population standard (p)=	1%
Per cycle sample size (n) =	50
Multiple of population standard that Type II errors are evaluated at =	5

Blue cells can be changed

			Achieved ACCEPTANCE PROBABILITY: Probability accept variety with Population Standard of off-types, ie with 1% of off-types	TYPE I ERROR: Probability reject variety with Population Standard of off-types, ie with 1% of off-types	TYPE II ERROR: Probability accept variety with 5 x Population Standard of off-types, ie with 5% off-types	
Over-cycles probabilities						
Approach 1	Per cycle max no. of off-types	2	99.94%	0.06%	56.07%	high risk test
Approach 2	Per cycle max no. of off-types	2	98.73%	1.27%	32.60%	moderate risk
	Combined sample max no. of off- types	3				
Approach 3	Combined sample max no. of off- types	3	98.16%	1.84%	25.78%	low risk test
Single-cycle probabilities						
Approach 1	Per cycle max no. of off-types	2	98.62%	1.38%	54.05%	
Approach 2	Per cycle max no. of off-types	2	98.62%	1.38%	54.05%	

Two cycle approaches
See future revision of TGP/10 for full details. Note results from growing cycles using different lots of plant material should not be combined.

Approach 1:
A variety is considered uniform if it is within the uniformity standard in both of the two growing cycles.
A variety is considered non-uniform if it fails to meet the uniformity standard in both of the two growing cycles.
If at the end of the two growing cycles the variety is within the uniformity standard in one growing cycle but is not within the uniformity standard in the other growing cycle, then uniformity is assessed in a third growing cycle. If in the third growing cycle the variety is within the uniformity standard, the variety is considered uniform. If in the third growing cycle the variety fails to meet the uniformity standard, the variety is considered non-uniform.

Approach 2:
A variety is considered uniform if it is within the uniformity standard in both of the two growing cycles.
A variety is considered non-uniform if it fails to meet the uniformity standard in both of the two growing cycles.
If at the end of the two growing cycles the variety is within the uniformity standard in one growing cycle but is not within the uniformity standard in the other growing cycle, then the numbers of off-types are combined over the two cycles and compared to threshold corresponding to the combined sample size.

Approach 3:
A variety is considered uniform if the total number of off-types at the end of the two growing cycles does not exceed the number of allowed off-types for the combined sample.

A variety is considered non-uniform if the total number of off-types at the end of the two growing cycles exceeds the number of allowed off-types for the combined sample.

A variety may be rejected after a single growing cycle, if the number of off-types exceeds the number of allowed off-types for the combined sample (over two cycles).

[End of Annex II and of document]