PD CEN/TR 16797-2:2015

BSI Standards Publication

Construction products: Assessment of release of dangerous substances — Guidance on the statistical assessment of declared values

Part 2: Technical and statistical background

bsi.

...making excellence a habit.™

National foreword

This Published Document is the UK implementation of CEN/TR 16797-2:2015.

The UK participation in its preparation was entrusted to Technical Committee B/557, Construction products - Assessment of dangerous substances.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

ISBN 978 0 580 85718 8

ICS 03.120.20; 91.100.01

Compliance with a British Standard cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 30 September 2015.

Amendments issued since publication

Date	Text affected

TECHNICAL REPORT

RAPPORT TECHNIQUE

TECHNISCHER BERICHT

CEN/TR 16797-2

August 2015

ICS 03.120.20; 91.100.01

English Version

Construction products: Assessment of release of dangerous substances - Guidance on the statistical assessment of declared values - Part 2: Technical and statistical background

Produits de construction - Evaluation de l'émission de substances dangereuses - Guide pour l'évaluation de la performance et la vérification de sa constance - Partie 2 : Données techniques et statistiques

This Technical Report was approved by CEN on 16 January 2015. It has been drawn up by the Technical Committee CEN/TC 351.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Former Yugoslav Republic of Macedonia, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.

EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

CEN-CENELEC Management Centre: Avenue Marnix 17, B-1000 Brussels

Contents

Page

European foreword

This document (CEN/TR 16797-2:2015) has been prepared by Technical Committee CEN/TC 351 “Construction products: Assessment of release of dangerous substances”, the secretariat of which is held by NEN.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN [and/or CENELEC] shall not be held responsible for identifying any or all such patent rights.

CEN/TR 16797, *Construction products: Assessment of release of dangerous substances — Guidance on the statistical assessment of declared values*, comprises the following two parts:

— *Part 1: Principles and rules of application*;

— *Part 2: Technical and statistical background* [the present document].

0 Introduction

0.1 General

CEN/TR 16797-1 [1] provides a brief introduction to declaring performance for the potential release, emission and/or content of dangerous substances from or in construction products and gives the principles which underpin the assessment and verification of constancy of performance (AVCP) of the product in respect of such declarations. The main rules of application are introduced, all of which satisfy the given principles.

CEN/TR 16797-2 provides a more detailed background and technical explanation together with examples and the statistical justification for the rules of application. The definitions and abbreviations listed in Clause 2 also apply to CEN/TR 16797-1. Annex D contains a model clause and the rules of application as normative text that may be copied into or cited by product standards. A recommended solution is to copy the model clause into the product standard and specify the 'rule of application given in CEN/TR 16797-2:2015, Annex D' to be used.

This Technical Report was developed on the basis of experience with the control of release into soil and water. As it is an assessment of data against a declared value regardless of the source of the data, it is the technical view of CEN/TC 351 that these procedures are also valid for the assessment of emission from construction products into indoor air and assessment of gamma radiation from construction products.

It is suggested that all product technical committees follow the principles set out in this CEN/TR and it is hoped that all regulators will accept that these principles achieve their objectives with respect to an acceptable AVCP procedure. The rules of application are examples of the ways in which these principles may be applied. There is no obligation on a product technical committee to adopt these rules of application and they are free to determine their own rules of application. The given rules of application may also be used as a benchmark for assessing alternative rules of application.

If product technical committees and producers could streamline their approaches in a way that could be accepted by all regulators, it might support a common understanding on the European market and it might encourage regulators to harmonize their existing different approaches and requirements on reliability and meaning of performance declarations in legislation and enforcement.

0.2 Background

CEN/TC 351 in its Resolution 162 (Milan 2011) has allocated to its TG 7 the task of drafting a technical report with the following content:

'This technical report provides guidance to Technical Committees on the evaluation of conformity of test results with regard to dangerous substances from or in construction products. Where further-testing is required, methods for assessing whether the product conforms to the information provided on the potential release of dangerous substances is described and justified. This technical report also describes the conditions and assumptions under which the proposals are given.'

In undertaking this task, TG 7 reviewed existing systems that are accepted by regulators. In practice these were the Dutch Soil Quality Decree system, a statistically based system where the rate of testing is a function of the closeness of the measured values to the regulatory limit value, and the German technical approval system. Technical Approvals normally apply to products not covered by CEN standards, but in Germany they are also being applied to the characteristic of release of dangerous substances, as this characteristic is not yet covered by CEN procedures. The German Technical Approval defines a rate of routine testing for release (not a constant between different products) and the product needs to achieve a release not greater than the defined value. The simplicity of this pass/fail criterion has merit, but there is no possibility of reducing the rate of testing (or incentive to the producer to achieve the lowest possible values of potential release) and when used

in a CEN system, the declaration of non-conformity over a period of months would have disproportionate consequences.

After reviewing the options, CEN/TC 351 is proposing a single statistically based system that includes the concept of a producer being able to reach a point where no-further-testing is required and a system that is applicable to a wide range of construction products, allowing a uniform assessment of release or emission which is not dependent of the type of construction product.

The objective of this CEN/TR is to provide guidance to product technical committees and information to regulators on the statistical assessment of declared values/classes for release, emission or content of dangerous substances.

0.3 Assessment of construction products

CEN/TR 15858 [2] describes three procedures for classifying the potential release of RDS: without-testing, without-further-testing and further-testing. Without-testing (WT) is a generic procedure where a comprehensive dossier of information containing information and previous test data on release from a product is prepared and submitted to the European Commission. The Commission appoints an expert group that includes regulators and product experts to review the dossier of information. If approved by the expert group and endorsed by the Commission, the product standard may include generic procedures for declaring classes of release based on composition, without the need for type testing and further-testing by the producer. Document DS 129 [4], however, describes this procedure as WFT on the basis that a dossier will not be approved without test data.

In this technical report, only the term WFT will be used, following the way it is used in DS 129.

Where a WFT procedure is not provided in the product standard or not selected by the producer, the producer undertakes type testing (TT) to determine the initial rate of further testing. Further-testing usually follows type testing and this is where every batch (called ‘batch testing’) or more usually a random batch selected from a defined period of production is tested (called ‘random testing’) and this test result is used with previous test data to show the consistency of production and to determine the frequency of testing. Further-testing by the producer (FT) and the assessment as to whether the declared value is validated using these test data is fully described in this CEN/TR.

If the FT (and rarely the TT) shows the product is consistently safe and the distance between the measured values and declared values is large, FT may end provided certain other requirements are met (essentially to ensure the composition of the product or the process is not changed significantly) and with the involvement of an independent third party. The criteria for no-further-testing (NFT) are defined in this technical report. This technical report recommends that any NFT decision is approved and audited by an independent third party.

NOTE CEN/TR 15858 [2] described this end to FT as ‘without-further-testing’, but this term is used in DS 129 [4] and in this technical report to mean something else; see text above.

All methods of verifying the declared value/class have equal status and validity. Figure 1 shows the outline procedure for these options.

The conditions under which the statistical assessment of declared values/classes described in this CEN/TR applies are:

— the product is covered by a harmonized European product standard requiring information on the potential release, emission or content of RDS or the contract of supply requires the provision of information on the potential release of RDS;

— there is one or more RDS that require FT to confirm the technical class or value declared by the producer;

— the FPC system has specifically addressed the control of RDS.

This CEN/TR provides guidance on the statistical assessment of declared values where type testing followed by further testing is required with respect to RDS for:

— potential release into soil, groundwater and surface water;

— potential emission into indoor air;

— content;

— potential gamma radiation or emission (exhalation) of radon gas into indoor air

to support:

— declarations of performance made under CE-marking;

— for showing conformity of RDS restricted at the Community level;

— for showing conformity of RDS restricted by provisions valid in the place of use;

— for showing conformity to limits for RDS specified in the contract of supply.

In all four cases the procedure for the statistical assessment of the declared value is the same.

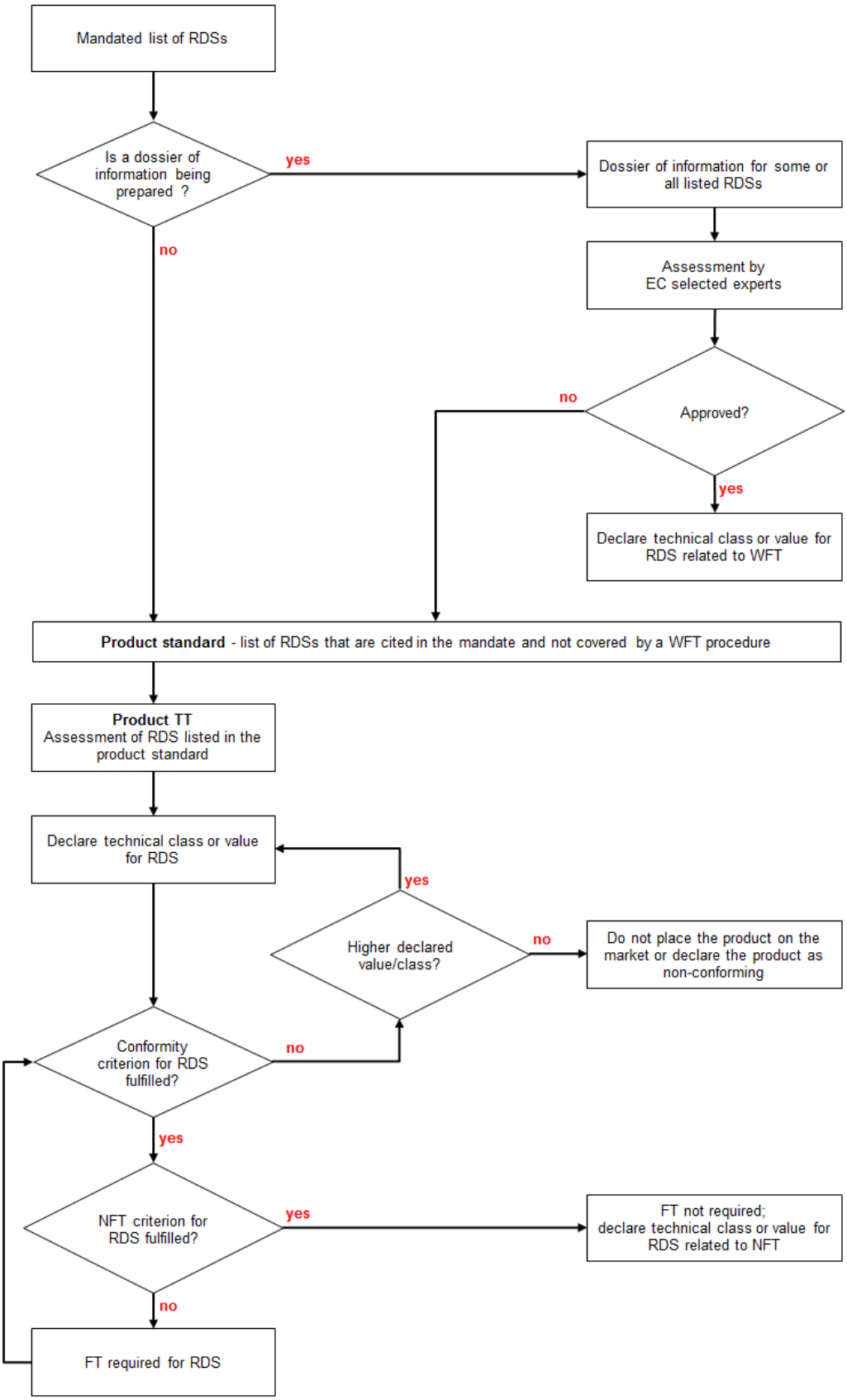


Figure 1 — Possibilities of WFT, FT and NFT for construction products

0.4 Reading guide to CEN/TR 16797-2

0.4.1 General introduction

A general introduction to the statistical assessment of declared values for the release, emission or content of RDS from or in construction products is given in Clauses 0, 1, 2 and 3. Terms and definitions are given in Clause 2 and Clause 3 puts CEN/TR 16797-2 into its context and identifies the tasks that need to be undertaken by CEN Technical Committees.

0.4.2 Rules of application

The rules of application are explained in Clauses 4 to 7. In Clause 4 the quality level of performance of the rules of application are defined. The methods and basic rules for the statistical assessment of declared values are described in Clause 5 followed by a technical description of the rules of application for assessment by variables (Clause 6) and assessment by attributes (Clause 7), all of which satisfy the given principles.

0.4.3 Statistical background

The statistical background of the rules of application is described in Clause 8. Clause 9 contains additional sampling requirements not addressed in CEN/TR 16220 [3], while Clause 10 gives guidance on the use of alternative, indirect test methods.

0.4.4 Annexes

Annex A contains examples of the rules of application described in this CEN/TR. Background information on the statistical distribution of RDS are given in Annex B. Annex C contains a checklist for technical aspects related to RDS that need to be addressed in product standards. In Annex D examples of model text are given on how to include normative requirements for statistical assessment of declared values. Annex E gives an extensive table with values for some of the statistical parameters used in this CEN/TR. Annex F gives values for the factor of a specific tool called the gamma rule.

1 Scope

This Technical Report provides guidance on the statistical assessment of declared values with respect to the release, emission and/or content of dangerous substances. This report provides statistically-based criteria for type-testing (TT), further-testing (FT) and where a product has been shown to be consistent with measured values for the release, emission or content that are significantly below the declared values, the point where no-further-testing (NFT) is permitted.

A series of fundamental principles are defined in CEN/TR 16797-1 and two statistical approaches are defined. The first approach is to use assessment by variables and this approach requires the data to be normally or log-normally distributed. This approach is recommended as the default option. The alternative approach based on assessment by attributes is appropriate for data sets that are not normally or log-normally distributed. The downside to this form of assessment is that more test data are needed for the same level of reliability. CEN/TR 16797-1 introduces these assessment procedures and CEN/TR 16797-2 provides more detail and the statistical proof that they satisfy the principles defined in CEN/TR 16797-1. With both of these approaches the minimum frequency of testing is a function of the distance between the mean value and declared value and the variability of the data set, i.e. the sample standard deviation.

To reduce the costs of testing, production plants producing a similar product may share data, e.g. be grouping the product into clusters for statistical assessment of declared values. Rules for the use of clusters are given in this document.

This document also contains rules for identifying outliers within a data set and guidance on using tests other than the reference method for FT.

A list of tasks for product technical committees is given in this document as is a model clause for including in product standards and rules of applications that may be cited in the product standard or copied into product standards.

2 Terms, definitions, abbreviations and symbols

2.1 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

NOTE The following definitions are used in this document for the purposes of using and explaining the system of assessment. These definitions align with those given in EN 16687 [6], ISO 3534-1 [10] and ISO 3534-2 [11].

2.1.1
assessment by attributes
assessment by noting the presence (or absence) of one or more particular characteristic(s) in each of the items in the group under consideration, and counting how many items do, or do not, possess the characteristic(s)

[SOURCE: [11] (adapted)]

Note 1 to entry: With respect to the assessment of the release, emission and/or content of RDS ‘presence’ or ‘possessing the characteristic’ means that the test result is greater than the declared value and ‘absence’ or ‘not possessing the characteristic’ means that the test results is equal to or lower than the declared value.

2.1.2
assessment by variables
assessment by measuring the magnitude(s) of the characteristic(s) of an item

[SOURCE: [11] (adapted)]

Note 1 to entry: With respect to the assessment of the release, emission or content of RDS 'magnitude of the characteristic' means the test result.

2.1.3
batch
amount of *construction product* at which *conformity* is established

Note 1 to entry: Within this document the size of a batch is undefined. The batch size should be defined in the product standard.

2.1.4
batch testing
procedure where every *batch* is tested prior to it being placed on the market

2.1.5
cluster
group of *production units* that are considered as one with respect to type testing and routine control of RDS

2.1.6
coefficient of variation
standard deviation divided by the *mean*

[SOURCE: [10]]

2.1.7
conformity
fulfilment of a requirement

2.1.8
construction product
product which is produced for incorporation in a permanent manner in construction works and placed as such on the market

[SOURCE [6]]

2.1.9
consumer's risk
probability of accepting a *batch* that has a release greater than the *declared value*

[SOURCE [11] (adapted)]

2.1.10
continuous random variable
random variable having a continuous distribution

[SOURCE [10]]

2.1.11
declared value
level, expressed as a numerical value, declared by the producer, having a very low probability of being exceeded in the production

Note 1 to entry: Where the term 'declared value' is used in this document, it may be interchanged with the terms 'regulatory class limit' or 'technical class limit'.

2.1.12
discrete random variable
random variable having a discrete distribution

[SOURCE [10]]

Note 1 to entry: The value of a discrete random variable is not expressed as a number but as a non-arithmetical expression, e.g. yes/no, pass/fail.

2.1.13
further-testing (FT)
assessment procedure where the AVCP requires routine testing by the manufacturer to verify that the *declared value* is being achieved

[SOURCE: [2] (adapted)]

Note 1 to entry: Further-testing is applied when the type assessment shows there is a risk that the declared value may be exceeded.

2.1.14
mean
sum of the test values divided by the number of test values

[SOURCE: [10] (adapted)]

Note 1 to entry: For a series of *n* random test results, i.e. {x_1, x_2, …., x_n}, the ***sample mean*** $\overline{x}$ is:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

2.1.15
no-further-testing (NFT)
deemed-to-conform assessment procedure for determining the *declared value* based on type assessment, type testing and routine testing where at some point it is shown that there is a very low risk that the *declared value* may be exceeded

2.1.16
non-conformity
non-fulfilment of a requirement

[SOURCE: [11]]

2.1.17
operation characteristic curve (OC-curve)
curve showing the relationship between the probability of acceptance of a *batch* and the percentage of the *production* that has a release greater than the *declared value* and the incoming quality level for a given assessment scheme

[SOURCE: [11] (adapted)]

2.1.18
outlier
member of a set of values which is inconsistent with the other members of that set

[SOURCE: [10]]

2.1.19
process
set of interrelated or interacting activities which transform inputs into outputs

[SOURCE: [11]]

2.1.20
producer's risk
probability of rejecting a *batch* that has a release equal to or lower than the *declared value*

[SOURCE: [11] (adapted)]

2.1.21
product
result of a *process*

[SOURCE: [11]]

2.1.22
production
entire (statistically assumed to be infinite) amount of *construction product* that has been and will be produced, that is divided into batches of equal size for testing purposes and that may be defined as the amount represented by:

— one type of construction product as defined in the product standard from one *production unit*; or

— a group of construction products with similar properties from one *production unit*; or

— one type of construction product as defined in the product standard that is grouped into a *cluster* for AVCP; or

— a group of construction products with similar properties from a *cluster*

Note 1 to entry: In the case of RDS, the declaration of performance (covered by CE-marking) is based on the properties of the production and not on the properties of single *batches* except when testing every batch.

2.1.23
production unit
location where a construction product is manufactured

2.1.24
random testing
procedure where a limited number of randomly selected *batches* is tested

2.1.25
sample
representative portion of material selected from a *batch* and which reflects the average properties of that *batch* as much as possible

[SOURCE: [6] (adapted)]

2.1.26
standard deviation
non-negative square root of the *variance*

[SOURCE: [10] (adapted)]

Note 1 to entry: This is a measure of the spread of results around their *mean*.

2.1.27
variance
sum of squared deviations of the test values from their *mean* divided by the number of test values minus one

[SOURCE: [10] (adapted)]

Note 1 to entry: For a series of *n* random test results, i.e. {x_1, x_2, …., x_n}, with sample mean $\overline{x}$ the ***sample variance*** s^2 is:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \quad (2)$$

2.1.28
without-further-testing (WFT)
deemed-to-conform assessment procedure for determining the declared value based on assessment of the generic product alone and specified in the relevant product standard

[SOURCE: [2] (adapted)]

Note 1 to entry: Assessment by the WFT procedure means that the release, emission or content of the relevant RDS in the construction product is determined on a generic level; an individual producer does not need to determine the release, emission or content of the relevant RDS in the construction product by type testing followed by further-testing. Type assessment is required to show that the product conforms to the requirements for using a declaration of performance based on WFT.

2.2 Abbreviations and symbols

For the purposes of this document, the following abbreviations and symbols apply:

AVCP	Assessment and verification of constancy of performance
CB	certification body
CEN	Comité Européen de Normalisation (European Committee for Standardization)
EEA	European Economic Area
FPC	factory production control
FT	further-testing
NB	notified body
NFT	no-further-testing
NPD	no performance determined
P{event}	probability of the event
RDS	regulated dangerous substance(s)
TC	technical committee
TG	task group
TR	technical report (CEN deliverable)
TT	type testing (determination of the prototype)
WFT	without-further-testing

For the purposes of this document, the following symbols apply:

a	slope
b	intercept
c_s	coefficient of variation of due to the sampling error
c_t	coefficient of variation due to the testing error
c_v	coefficient of variation of the test value
c_w	coefficient of variation due to the spatial and/or temporal variation
G_p	critical value according to Grubbs' test
H_0	null hypothesis
i	index number
k_0	critical value for k_n multiplication factor of the non-central t-distribution
L_D	declared value
m	number of increments per sample
n	number of test values
n_a	acceptable number of test values that exceed the declared value
n_e	number of test values that exceed the declared value
p	percentage of the production exceeding the declared value
PF	precision factor
s	sample standard deviation
x	test value
x'	equivalent value for the reference test
$\overline{x}$	sample mean
x_1	the last (most recent) test value
x_i	test value
x_{max}	largest test value
x_{min}	smallest test value
z_i	upper 100i percentage point of the standard normal distribution
α	uncertainty
β	fraction of production that exceeds the declared value
Γ	gamma factor
μ	true value of the mean of a property
μ_0	true production mean
σ	true value of the standard deviation

σ_s standard deviation due the sampling error

σ_t standard deviation due to the testing error (see test standard)

σ_v standard deviation of the test value

σ_w standard deviation of the spatial and/or temporal variation

3 Context

In order to put this CEN/TR into its context, the following is a brief review of the overall procedure for declaring performance from the viewpoint of a producer. Much more detail is provided in CEN/TR 15858 [2] and CEN/TR 16496 [5]. For the producer, the starting point is the product standard and/or the contract to supply. The actions before this point are for the product technical committee preparing the product standard and those preparing a generic dossier of information (see Figure 1).

The European product standard that has a revised or amended mandate for the product being produced should give a list of dangerous substances. Except where the dangerous substance is limited at the European level, these lists will not contain limit values as these values, if any, are set at the national level. Producers identify their market (national members of the EEA) to determine if it includes one or more regulated markets, i.e. ones with defined limits on the release, emission or content of dangerous substances, and identify from the list in the product standard those dangerous substances that are applicable in their regulated markets. In addition to the dangerous substances listed in the product standards, there may be the need to conform to other regulations valid in the place of use that set limits or ban certain additional dangerous substances. The applicable RDS may be some, all or more than those listed in the product standard. Producers are expected to know what regulations (legal requirements) are placed on the products they are putting on the market.

Producers may use the database CP-DS (http://ec.europa.eu/growth/tools-databases/cp-ds/) for information on national and European legislation on dangerous substances and construction products, and they may make use of the contact points for the CPR in each Member State.

NOTE The rules of application described in this CEN/TR are applicable to all RDS, including those not listed in product standards.

For those substances listed in the product standard, the producer is required to provide information to the market in the form of a technical class or declared value (where permitted in the product standard) under CE-marking. Essential characteristics (in particular for the release, emission or content of dangerous substances) shall be declared according to the cited harmonized standard or as NPD to legally place products in the market (EFTA). The regulations of some Member States may require the declaration of some essential characteristics for specific intended uses.

Some essential characteristics may be declared through the WFT procedure. In this case the declared value/class is fixed and does not require testing. For essential characteristics not covered by the WFT procedure, manufacturers shall follow the defined AVCP and declare the value/class or NPD.

Where a WFT procedure is applicable and used, the producer needs to document in the FPC system that the product conforms to the requirements for using the WFT procedure. After TT and periods of routine testing, the producer may have the opportunity to see if the test data satisfy the requirements for declaring performance with no-further-testing (the NFT-procedure). The criteria for such assessments are given in this CEN/TR, but they only have validity if these criteria are transposed into the relevant product standard or cited in a normative way as being applicable. The procedures that are relevant to the NFT-procedure are given in 5.4.

4 Quality level of performance with respect to the release, emission or content of RDS

With respect to the release, emission or content of RDS it is highly recommended to apply the same quality level of performance to all construction products. In this way users of construction products and regulators are able to compare performances of different (competitive) construction products without doubt about the risk of getting a non-conforming product, e.g. a product that has a higher release, emission or content of one or more RDS than declared by the producer.

For the release, emission and content of RDS the following two quality levels of performance are proposed:

The rules of application verify with a confidence of 90 % that the 50th percentile of the production is less than or equal to the declared value/class.

Assessment of NFT verifies with a confidence of 99 % that the 90th percentile of the production is less than or equal to the declared value/class.

The assessment system described in this CEN Technical Report and the according consumer's risk and producer's risk are based on this quality level of performance when the scale of declaration is a batch as defined in the product standard. The quality level of performance for the statistical assessment of declared values is based on the Dutch Soil Quality Decree which has proven to give sufficient protection for the environment with respect to the release of RDS from construction products.

5 Methods for statistical assessment of declared values for RDS

5.1 Goal of the assessment

The goal of assessing the potential release, emission or content of dangerous substances is to allow a producer to define a declared value or technical class for a product in such a way that testing is reduced to a minimum while users and regulators are still able to conclude with sufficient trust whether the product fulfils the (national) requirements for release, emission or content of dangerous substances.

To find a balance between testing effort and trust a risk based approach is proposed, meaning that the risk of exceeding the declared value is acceptable within the framework of (national) regulations.

Based on the Dutch experiences, a risk of exceeding a limit value of not more than about 20 % seems acceptable.

The Dutch system of assessment of the release of dangerous substances is based on a statistical approach of testing randomly selected batches. From an evaluation of the Dutch Soil Quality Decree in 2011 [14], it was concluded by the Ministry of Infrastructure and the Environment that the requirements of the decree provide sufficient protection for the environment with respect to the use of construction products. The statistical rules used to assess the release of dangerous substances give a maximum risk of exceeding a limit value of about 21 %.

It should be noted that the statistical rules to assess the release of dangerous substances are just one aspect within the framework of the Dutch Soil Quality Decree. A whole set of additional rules is part of the Dutch system to control the environmental quality of a construction product and its uses.

5.2 Assessment of the production

In its most simple form, assessment of the properties of a production is done by testing each and every batch. The batch conforms when the measured values of all tested properties fulfil the criteria specified. Testing each batch may lead to a lot of unnecessary testing and high cost for the producers

and their customers, because the risk of exceeding a declared value is not taken into account. If this risk is dealt with in a proper way, it is possible to limit the amount of testing without significant risk. Moreover, when the risk of exceeding the declared value defines the test frequency, the system of testing will be cost effective.

The method of assessment of the release, emission or content of dangerous substances in this CEN/TR incorporates a risk based testing system for use within the quality control system of the producer for routine testing.

If the risk is (very) low, assessment is based on testing randomly selected batches (**random testing**). The test frequency decreases gradually when the risk of exceeding the declared value becomes smaller. In the case of random testing the declared value is not a hard limit, because the risk of exceeding this value is low and it is only used to estimate the risk of exceeding this value. Testing randomly selected batches implies that not all batches are tested and no batches are rejected; however, the assessment of conformity applies to all batches, including the batches which are produced in between those tested. As not all batches are tested, there is always a (very) low risk that a batch which exceeds the declared value is part of the accepted production and placed on the market (this is true for all product characteristics, not just RDS). But as long as the product's conformity is established by testing randomly selected batches, the product may be put on the market before conformity is established, because there is no necessity to wait for the result as the risk of exceeding the declared value is (very) low (the reference leaching tests take 1 month to 2½ months, depending on the type of test).

On the other hand every batch is tested when the risk of exceeding the declared value is relatively high (**batch testing**). Under this condition this value is a hard limit. Exceeding the declared value means that the batch is rejected, consequently only conforming batches are placed on the market. This requires that the product is tested and conformity is established before it is placed on the market. Nevertheless, there are a few products that need to be placed on the market before the assessment of conformity is completed, e.g. screeds. Experience has shown that these products are unlikely to require batch testing.

The advantage of batch testing is that there is no uncertainty about conformity of the batch (except for the uncertainty due to the test error), although testing will be more expensive and time consuming. Batch testing allows a producer to prevent batches that do not exceed the declared value being rejected and batches that exceed this value being placed on the market. On the other hand it also highly motivates a producer to improve the quality of his product in order to reduce the testing effort and to be able to switch to random testing.

The described method for the assessment of conformity has the following characteristics:

— risk based approach: a maximum risk that a CE-marked product which has been placed on the market exceeds a declared value for the release, emission or content of a dangerous substance;

— where the risk of exceeding the declared value is high every batch is tested and the declared value is a hard limit;

— conformity is assessed on the scale of the batch size;

— the statistical assessment of declared values when based on assessment by variables is based on the running mean of at least five test results which include all valid test values that are collected during FT; for assessment by variables the safety margin to limit the risk of a batch which exceeds the declared value being placed on the market is obtained by taking into account the variation between batches due to the variation in product quality. In the case of assessment by attributes this is achieved by limiting the acceptable number of test values that exceed the declared value;

— conformity is assessed separately for each RDS. The combined risk for all RDS of a construction product is not considered and consequently any RDS can determine the test frequency and ultimately the non-acceptance of a batch;

— the distribution of test values is assumed to be either lognormal (assessment by variables) or unknown (assessment by attributes). Other distributions are not considered;

— variable test frequency based on the risk of exceeding the declared value;

— arbitrarily, the assessment method is based on a true maximum risk of about 10 % (in practice the true maximum risk will be lower than the theoretical maximum risk as producers will be motivated to aim at the lowest test frequency).

If a construction product is characterized by more than one relevant RDS, the theoretical maximum risk of exceeding one of the declared values is equal to $1 – 0{,}90^n$, where n is the number of relevant RDS and assuming a maximum risk of 10 % of exceeding the declared value for one relevant RDS (e.g. in the case of 3 relevant RDS the theoretical maximum risk that a declared value is exceeded equals 27 % instead of 10 % for one relevant RDS).

In practice the risk is lower than this theoretical maximum risk due to the facts that:

— the release of one of the RDS may be dependent on the release of another RDS, which means that only one of the two controls the risk of exceeding a declared value. For example, in some slags the leaching of Ba and SO_4 is controlled by the solubility of $BaSO_4$: the presence of oxidized sulphur (sulphates) prevents leaching of Ba;

— the system of assessing the release, emission or content of RDS is controlled by the RDS that gives the highest risk. In practice with most construction products one relevant RDS controls conformity.

Therefore it is hardly possible to estimate the true risk, unless the true risk for each of the RDS of a construction product is known which requires a lot of testing. For this reason, the assessment system is focused on assessing individual RDS only instead of the entire set of RDS of a construction product.

Although two distributions are considered, the use of assessment by variables is preferred for practical reasons; however, the use of assessment by variables requires a normally distributed dataset. As explained in 5.5.1 for dangerous substances a logarithmic transformation is used to (more or less) normalize data sets. If the data set already has a normal distribution by itself, an ln-transformation is not needed. Also an ln-transformation is not always the optimal choice and more appropriate statistical techniques, like the use of the Box-Cox transformation, may give a better result.

Because it is considered too complex for users to allow different normalization options in a product standard, ln-transformation is the best available and relatively easy to use option. Also, one of the goals of standardization is limiting options. Therefore the method of assessment in this CEN/TR focuses on ln-transformation with respect to assessment by variables. If it is shown by a user that no transformation is needed or that another type of transformation gives a better approximation of normality for a certain product or substance, it is recommended to allow other statistical techniques to improve normality of data sets. In all cases the formulas for assessment by variables presented in this CEN/TR may be used.

5.3 Assessment of clusters

The concept of clusters, i.e. group/groups of manufacturers sharing logistics of testing, test results and information within a national or international construction products approval scheme, may be an efficient and cost effective way of testing construction products that have a relatively low risk of exceeding declared values. Although this CEN/TR gives technical rules for the assessment of construction products only, some legal and commercial aspects are mentioned as they are believed to

be crucial for the functioning of clusters. One of these aspects is that clusters need to be managed by an independent organization, the so-called cluster organization, if the cluster comprises more than one producer.

Clusters may be organized in different ways.

1) A producer who owns several plants might decide to create an in-company cluster for some or all of his plants. In this case a third party cluster organization is not needed.

2) A trade association might create a cluster that is available to its members; one of the privileges of membership may be a reduced cost for AVCP with respect to dangerous substances.

3) A commercial organization makes cluster facilities available for any producer who wishes to join and their product can meet the technical criteria for membership.

In Cases 2) and 3) the involvement of an independent cluster organization is required. In a normal situation, where a producer tests his product independently, test data remain the property of the producer and any third party, e.g. certification body, is not permitted to disclose these data to another party or use these data for their own commercial gain. The cluster system however, requires a different relationship between the parties where a third party, the cluster organization, is free to use all the data to show conformity of the cluster and to exclude members if their results are outside the normal variation and/or lead to an increase in the frequency of testing. Such sharing of commercially sensitive information between different commercial organizations should meet the rules for free competition and would need the explicit approval of the Competition Authorities. Therefore it is suggested that the cluster organization should be a totally independent legal entity that has no connections to any of its members if the cluster comprises more than one producer.

In theory Situations 2) and 3) could lead to an unlimited number of companies joining a cluster. In practice however, the number of companies will be limited, as the release properties of the products in one cluster need to be more or less the same in order to keep the risk of exceeding the declared value in control. Only low-risk construction products, where the release, emission or content is significantly below the declared value, may be grouped in one cluster.

NOTE An example is the cluster of Dutch bituminous mixtures (Asphalt Cluster). This cluster comprises most of the asphalt mixtures that are being placed on the market by the Dutch producers. Despite the cluster being quite large (about 45 production units), cluster control has been proven to be sufficiently reliable for checking the release of RDS.

According to the CEN Rules a standard should give requirements, methods for verification of these requirements and AVCP without assigning these tasks to a specific organization. Cluster management, however, should always include third party involvement (except for situation 1). Addressing this involvement collides with the CEN Rules. For this reason the technical rules for clusters in CEN standards are informative or could simply cite this CEN/TR (see D.4).

5.4 Assessment of no-further-testing

In this document no-further-testing (NFT) is considered an option for a specific product or cluster following TT or when sufficient test results have become available during FT for one or more RDS. This NFT is an assessment based on all available test values and based on the same principles as described in 5.2; however, the risk based approach of the method described in 5.2 is not appropriate for the assessment of NFT.

The described method of assessment for NFT has the following characteristics:

— risk based approach where the assessment is based on all available test values;

— NFT is assessed for each RDS separately;

— the distribution of test values is assumed to be either (log)normal or unknown. Other distributions are not considered.

As the background of historic data will probably be unknown, the batch size represented by the test results will probably be unknown as well. In case of assessment of NFT this uncertainty is included in the criteria.

If it is decided that the NFT option should not be a producer's decision, a third party would need to be involved. This leaves two options:

1) A low level decision taken by a third party. For practical reasons, a notified body or certification body seems most appropriate to take this decision. For this reason the system of AVCP should be 1+, allowing the NB or CB to take and test samples independently (see Figure 2).

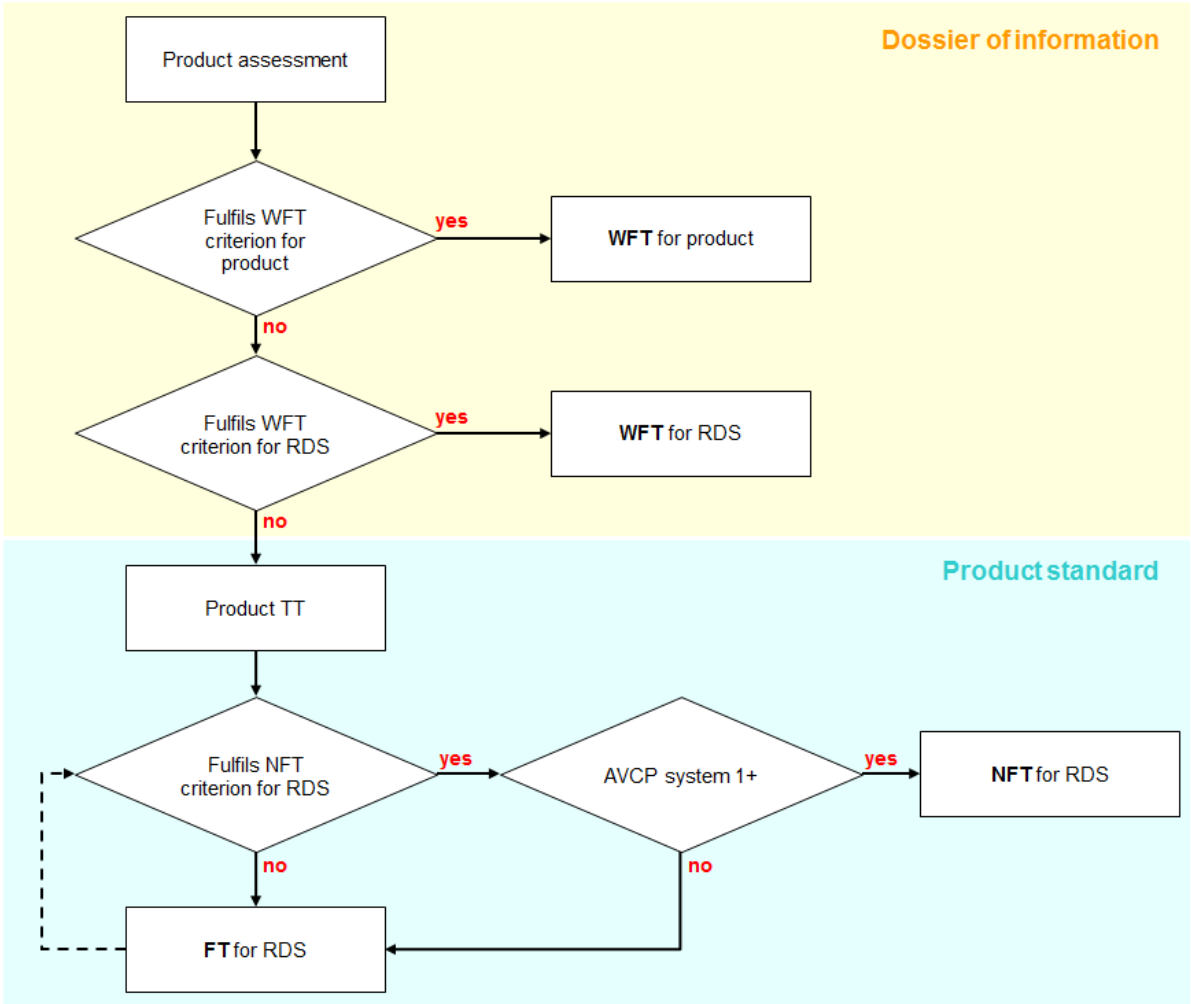


Figure 2 — WFT-FT-NFT procedures for construction products option 1: third party control

2) An update of the WFT dossier of information, in which case the decision is a high level generic one taken by the EC (see Figure 3). A disadvantage of this procedure is that it is not possible for a single producer to apply directly for a generic WFT procedure.

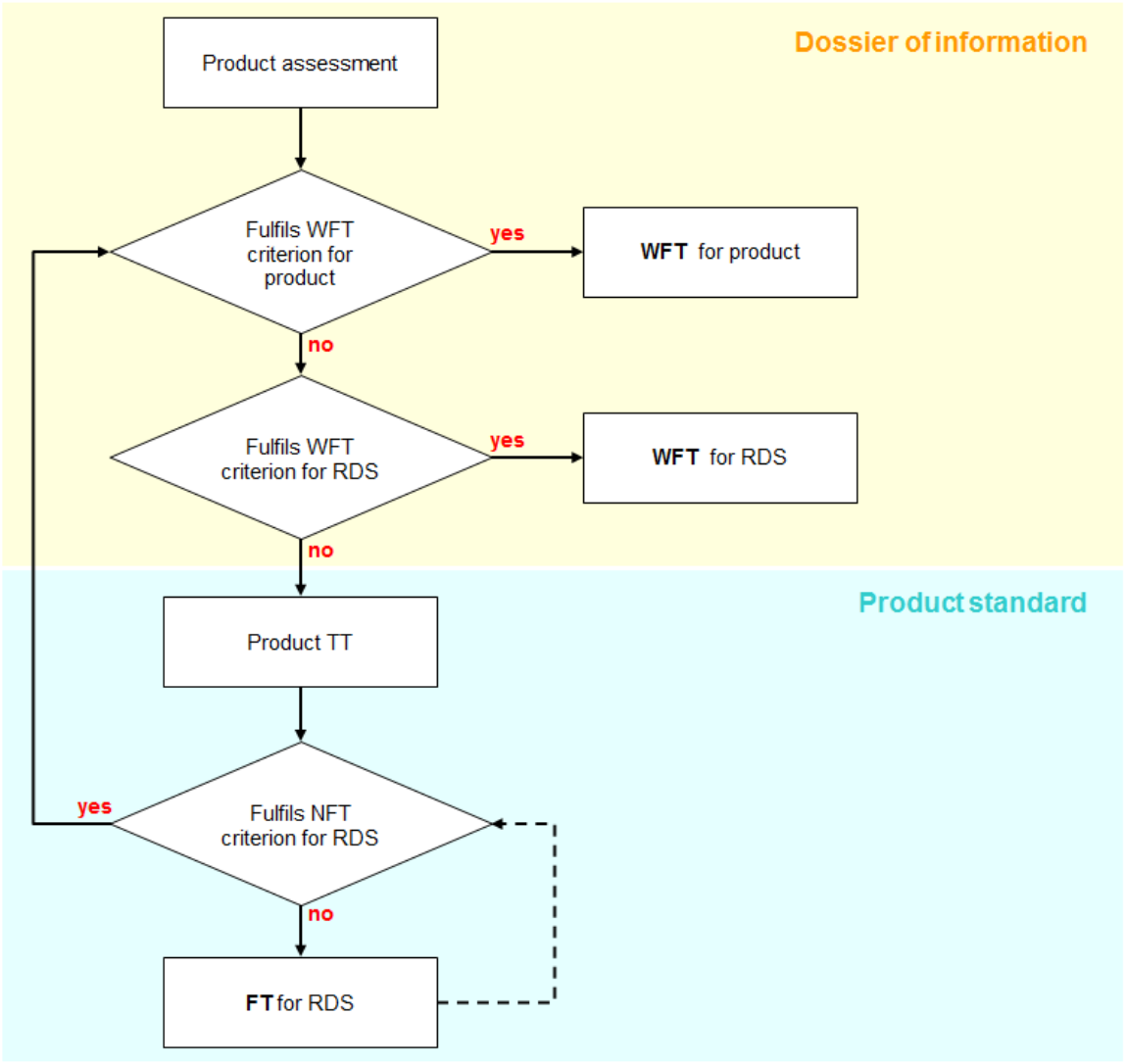


Figure 3 — WFT-FT-NFT procedures for construction products option 2: update dossier

Conditions for a NFT status of a dangerous substance are:

— a well-defined product with respect to composition, raw materials, (technical) properties relevant to the release, emission or content of dangerous substances, etc.,

— the product's composition, properties etc. remains within a defined range,

— a continual surveillance of the product's properties that are linked to the NFT decision.

In principle, the level of the system of AVCP will remain the same as in the original product mandate. Member States, however, may ask the European Commission to introduce another level of AVCP via a Commission Decision (legal act). As a system 1+ is most desirable with respect to dangerous substances, for some standards such a request is likely to be made.

In practice it will not be useful to apply NFT to all RDS for all construction products. In the case of some products the release, emission or content of certain RDS depends on the quality of the production process or the use of a specific raw material.

NOTE An example is the leaching of sulfate from recycling aggregate. Sulfate is related to the presence of gypsum and aerated concrete. The quality of the process to remove gypsum and aerated concrete from the construction waste at the construction site and the quality of the production process of the recycling aggregate (how does the producer deal with gypsum or aerated concrete contaminated construction waste) determine the quality of the final product.

In these special cases where the quality of the process determines the quality of the product, NFT is not an appropriate option. On the other hand, NFT is a useful option for intrinsic micro contaminations that cannot be controlled, but from which it is known that the release, emission or content of these RDS is significantly below the declared value.

5.5 Methods of assessment

5.5.1 Introduction

The statistical assessment of declared values is based on two statistical methods which depend on the distribution of the test values: assessment by variables or assessment by attributes. Assessment by variables assumes a lognormal distribution; assessment by attributes assumes an unknown distribution. Annex B gives background information on the statistical distribution of RDS. Where (log)normality cannot be established assessment by variables will probably lead to false conclusions and unjust rejection of products. In this case assessment by attributes may be more appropriate, although more test results will be needed.

As a starting point it is assumed that for all situations a lognormal distribution may be applied. If a normal distribution or other normalized distribution is more appropriate, the producer may choose to apply the normal or normalized distribution instead. This does not affect the quality level of performance or rules of application.

In the case of **outdoor soil and groundwater** (leaching) there are many test results available based on the experiences with the Dutch Soil Quality Decree where for many RDS the declared value is substantially higher than the detection limit of the leaching tests (depending on the type of construction product). For this reason assessment by variables is the most appropriate method even if the test values are close to or lower than the detection limit.

In the case of **indoor air** (emission) the results of the robustness testing seem to give a similar picture.

In the case of **content**, the declared value may be substantially higher than the detection limit of the test method or close to or even lower than the detection limit (according to the experiences with the Dutch Soil Quality Decree). Depending on the type of dangerous substance assessment by variables (e.g. heavy metals) or assessment by attributes (e.g. asbestos) may be used.

Nevertheless, for all situations it is possible to use assessment by attributes.

5.5.2 Assessment by variables

Because most test values for the release, emission or content of dangerous substances are relatively close to zero, it is assumed that test values have a skewed distribution and taking the log of the test values gives an approximately normal distribution (for brevity this is described as a lognormal distribution). For most dangerous substances this appears to be a correct assumption, although the distribution tends to change to normal when the lowest test values for a specific product and dangerous substance are considerably greater than zero.

As the level of release, emission or content is not known in advance and the effect on the risk of exceeding a declared value by assuming a lognormal distribution instead of taking into account the true distribution is minor or on the safe side, it was decided to ignore this in order not to complicate the method of assessment.

In case of a lognormal distribution the statistical principle is based on the so-called ***assessment by variables***. In this case the statistics for normal distributions apply to ln-transformed test values. This implies that the assessment is based on the sample mean and sample standard deviation of ln-transformed test values. Because the true mean and true standard deviation of a set of test values may change in time (e.g. due to minor changes in the production process, switching between different constituents, seasonal influences or oxidation of certain compounds), it is not possible to assume that the samples are taken from a constant production. Also, it will not be likely that the true percentage of the production exceeding the declared value is known. Therefore, criteria for assessment of conformity need to be based on statistics that include these unknown variations of the production process and unknown percentage of production exceeding the declared value.

For these reasons statistics based on the non-central t-distribution are used for assessment by variables. When defining conformity criteria based on assessment by variables, the number of samples and the probability of accepting a non-conforming product for a given percentage of production exceeding the declared value are chosen. The multiplication factor to fulfil these conditions is calculated or taken from a table.

5.5.3 Assessment by attributes

In the case where the assumption of a lognormal (normal or other normalized) distribution (see NOTE 2 in 5.2) is not appropriate, e.g. when the test values are close to or lower than the detection limit or when test values can be substantially affected by the production process, the statistics that apply to an unknown distribution may be used. For this ***assessment by attributes*** the statistical assessment of declared values is based on the number of test values that exceed the declared value within a defined number of test values. The number of test values that exceed the declared value cannot be greater than the number of test results that are permitted to exceed the declared value.

The advantage of using attributes is that it is not necessary to know the distribution of the test values and that it may be used for continuous random variables as well as discrete random variables. On the other hand, a larger number of test values than for assessment by variables are needed to obtain the same level of confidence in the declaration.

Statistics for assessment by attributes are based on the binomial distribution. When defining conformity criteria based on assessment by attributes, the number of samples and the probability of accepting a non-conforming product for a given percentage of production exceeding the declared value are chosen. The number of test values that are permitted to exceed the declared value to fulfil these conditions is calculated or taken from a table.

5.6 Description of the assessment procedure

5.6.1 Type testing

An early end to type testing is only possible if the mean value is low in relation to the declared value.

Type testing is the start of the assessment procedure and comprises testing every batch. Where using the method of variables, type testing comprises at least two, but nor more than ten, test results. Where using the method of attributes, type testing comprises at least four, but not more than twelve, test results. On the basis of the TT results and a declared value it is decided whether the production is conforming and the product should be placed on the market, see Figure 4. Also the frequency of FT testing is established: if the production conforms to the declared value in accordance with the random testing criteria given in 6.2.2 and 7.1.2, FT is based on random testing. If not, FT is based on batch testing.

If the criterion for random testing is not fulfilled the producer may change the declared value to some higher value (but such product may not satisfy the criteria for certain markets), or the producer may

decide that FT is based on ***batch testing***, in which case each batch is tested and every batch that exceeds the declared value is rejected.

In the case of ***random testing*** not all batches are tested, but a smaller number of randomly selected batches and no batches are rejected (thus, batches which exceed the declared value are accepted). Additionally the results may be checked for NFT if at least 5 test results are available.

As the introduction of testing RDS in product standards will affect ongoing production, it is suggested to carry on production in the case of an existing production and take decisions after finishing TT. If TT concerns a new product, TT should preferably be finished before the product is placed on the market. If that is not feasible, during TT conformity of each batch tested should be established before placing the batch on the market.

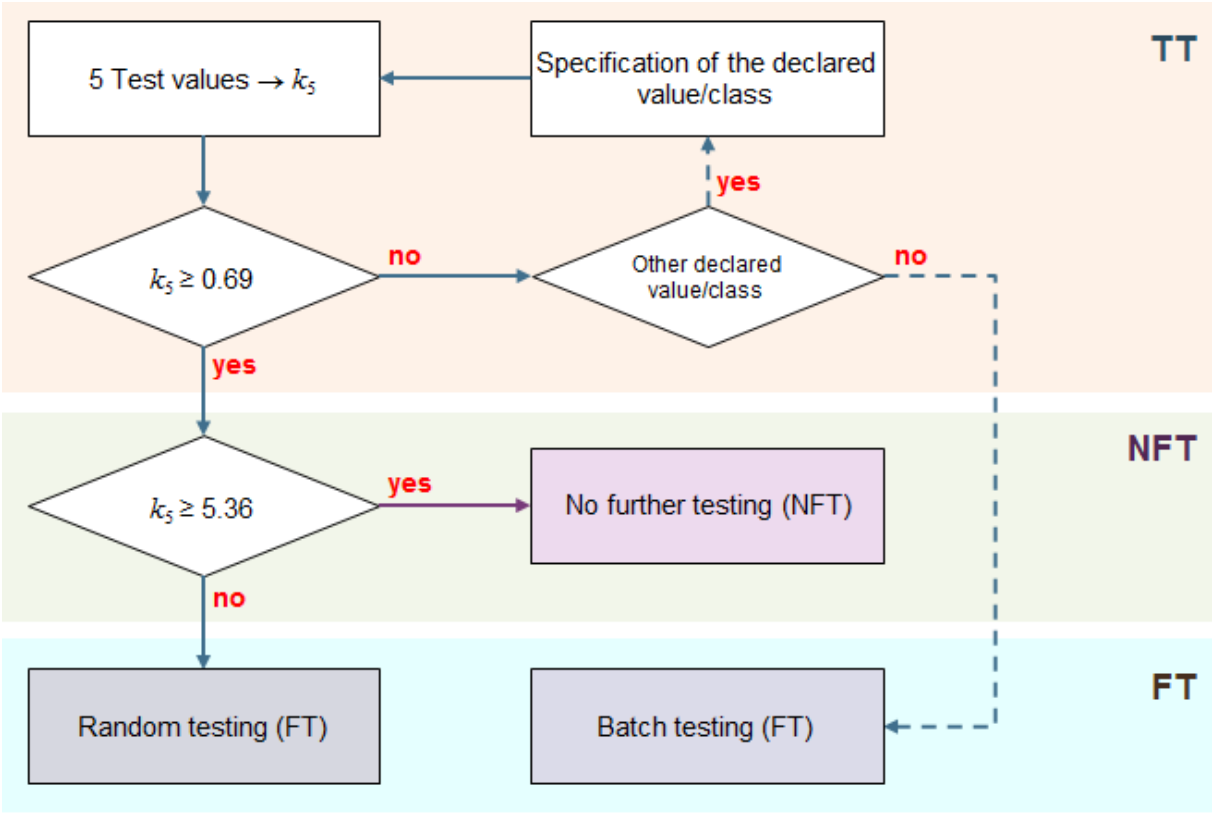


Figure 4 — TT procedure, example of assessment by variables based on 5 test values. In the case of NFT more requirements need to be satisfied, see 5.4

5.6.2 Further-testing

Type testing is followed by further-testing at a rate that is a function of the closeness of the measured values to the declared value. When the risk of exceeding the declared value becomes higher (the closer the measured values are to the declared value or the more variable the results) the more frequent is the testing. When the risk becomes too high every batch is tested.

When in FT the conformity criterion is combined with risk based criteria for the determination of the test frequency for testing randomly selected batches. This means that the test effort for a RDS increases and more batches are tested when the risk of exceeding the declared value becomes higher. When the risk becomes too high and conformity cannot be ensured, every batch is tested. For both methods (assessment by variables and assessment by attributes) two alternative criteria are given. One criterion is for a relatively small number of test values and one for a larger number of test values.

When FT is established, new test values are assessed in conjunction with previous test values (random testing) or as single values (batch testing). After every new test value the frequency of testing is determined and it is decided whether random testing or batch testing is to be applied or whether it is allowed to switch from batch testing to random testing, see Figure 5.

NOTE The criteria in Figure 4 and Figure 5 are based on assessment by variables (see 6.1.1 and 6.1.2). For assessment by attributes (see 7.1.1 and 7.1.2) equivalent criteria are applied.

As long as random testing continues all batches, including the tested batch, may be placed on the market before the test results are available. The batch that causes a change from random testing to batch testing is not subjected to the batch testing criterion. If the test value of this batch exceeds the declared value, a practical dilemma occurs, because random testing allows the product to be placed on the market before the test result is available so that rejection is not possible unless this batch has not been placed on the market. Where batch testing is required batches cannot be placed on the market before the test result is available.

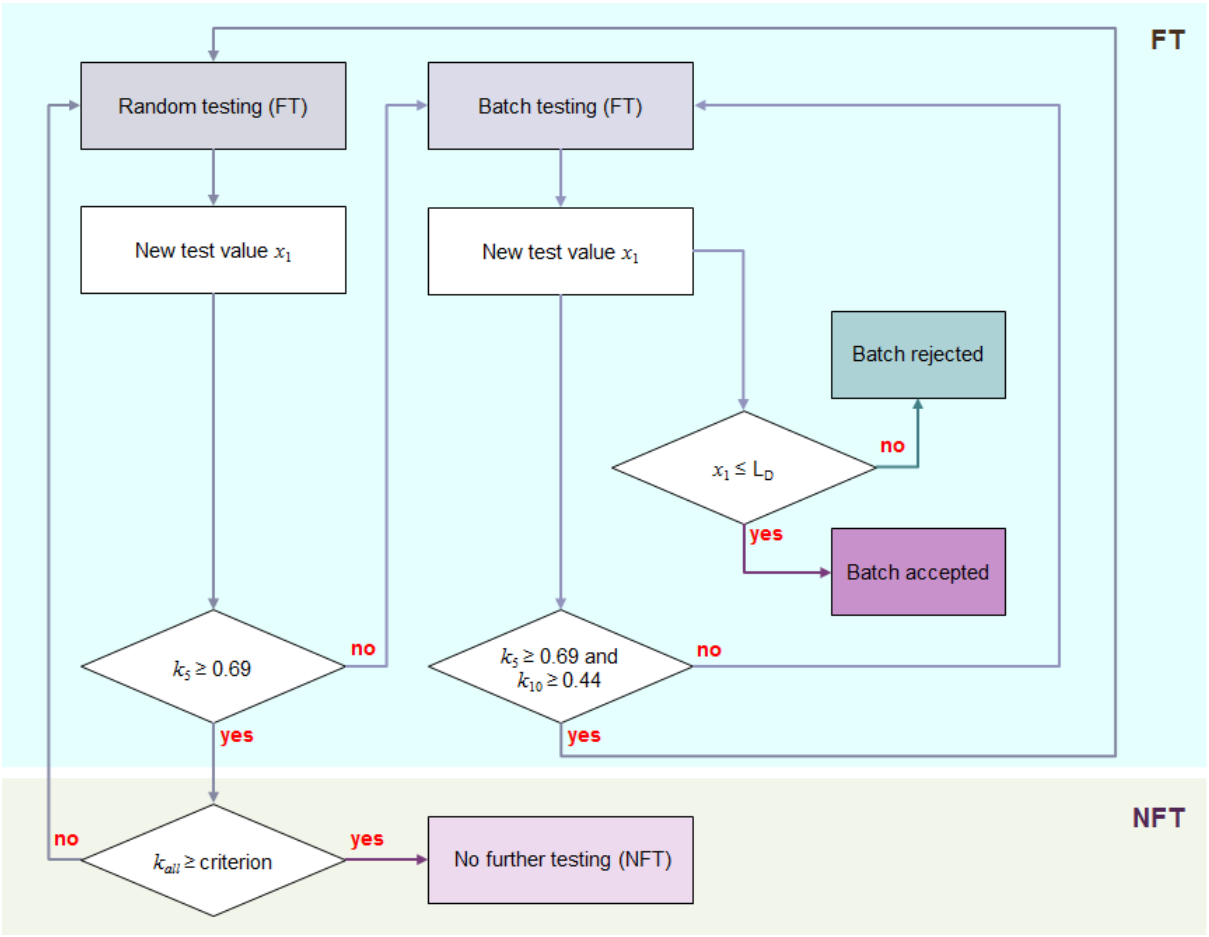


Figure 5 — FT procedure based on 5 test values and assessment by variables. In the case of NFT more requirements need to be satisfied, see 5.4. In the case of batch testing a return to random testing is delayed, see 6.2.2

5.6.3 Clusters

The assessment criteria described in this CEN/TR are applicable to both the production of a single production unit and to the production of clusters. In the case of a cluster additional requirements may be defined by the organization running the cluster with respect to the properties of the product(s) or quality system.

Cluster TT requires a thorough and complete description of the products in the cluster:

— definition of the product, including percentage or content ranges of constituents, composition, specific properties and other relevant aspects of the product;

— full description of all (potential) constituents, constituent parts, etc., including origin and relevant properties.

TT could then be focused on the 'extreme' combinations, so that TT covers the entire range of different combinations and all constituents.

The cluster should be run by an organization that is independent from any of the participating production units to avoid a conflict of interest.

The cluster organization may for instance be a notified body in the case of a group of producers or an independent quality control unit in the case of a group of production locations of one producer.

The cluster organization and production units should have a contract. This contract should include rights and obligations of the cluster organization and production unit, including on site quality control by the cluster organization. The technical requirements in the contract should be equal for all participating production units.

The cluster organization should establish and maintain a quality control manual setting out the procedures by which the requirements for quality control are satisfied. This manual should include:

— a description of the tasks, responsibilities and competences of the cluster organization and participating production units, including on site quality control by the cluster organization;

— a procedure for quality control by the participants, including sampling;

— a procedure for the coordination of all control activities by the cluster organization;

— a procedure for sample selection and testing by the cluster organization;

— a procedure for sample selection and testing by the cluster organization;

— a procedure and requirements for admittance to the cluster;

— a procedure for expelling from the cluster;

— a procedure for internal communication and protection of the anonymity of the participants;

— registration of the participating production units;

— maintenance of the product's definition.

If the test data of a RDS do not fulfil the requirements for applying the cluster system, all participating production units undertake separate testing of this RDS (or all RDS if this is considered more practical). In order to be sure that proper measures have been taken after the cluster system has been rejected and FT changed from random testing of the cluster to random (or batch) testing of all individual production units, each production unit is required to test at least one batch before it is allowed to return to random testing of the cluster. If the cluster consists of less than 5 production units, at least 5 batches need to be tested in total.

The sequence of testing all individual production units is repeated as long as the cluster system is rejected (see Figure 6).

When the risk of exceeding the declared value is too high for a specific RDS the cluster system cannot be applied for this RDS and all members need to test it based on the rules for single production units. As the risk level of switching from cluster testing to individual testing is lower than for switching from random testing to batch testing for a single production unit, cluster FT is based on a three step approach:

1) random testing for a cluster;

2) random testing for a single production unit;

3) batch testing for a single production unit.

If a new production unit is to be admitted to an existing cluster the following should be taken into account:

— the composition and other properties of the product should lie within the defined range of the cluster product(s);

— at least one sample should be tested to allow an assessment by variables or by attributes by adding the test value(s) to the cluster's data base.

Besides the technical criteria defined in 6.2 and 7.2, additional criteria might be considered, e.g. the requirement that the test frequency for the cluster product is not changed due to admitting a new production unit and/or that the test value lies within the range of the data base.

The organization that runs the cluster may define criteria for discarding a product from the cluster in order to protect the cluster from being disrupted, e.g. by defining when a production unit produces outlying results, or when a production unit changes its raw materials. If a product is discarded from the cluster all test data from this specific product should be deleted from the cluster data base.

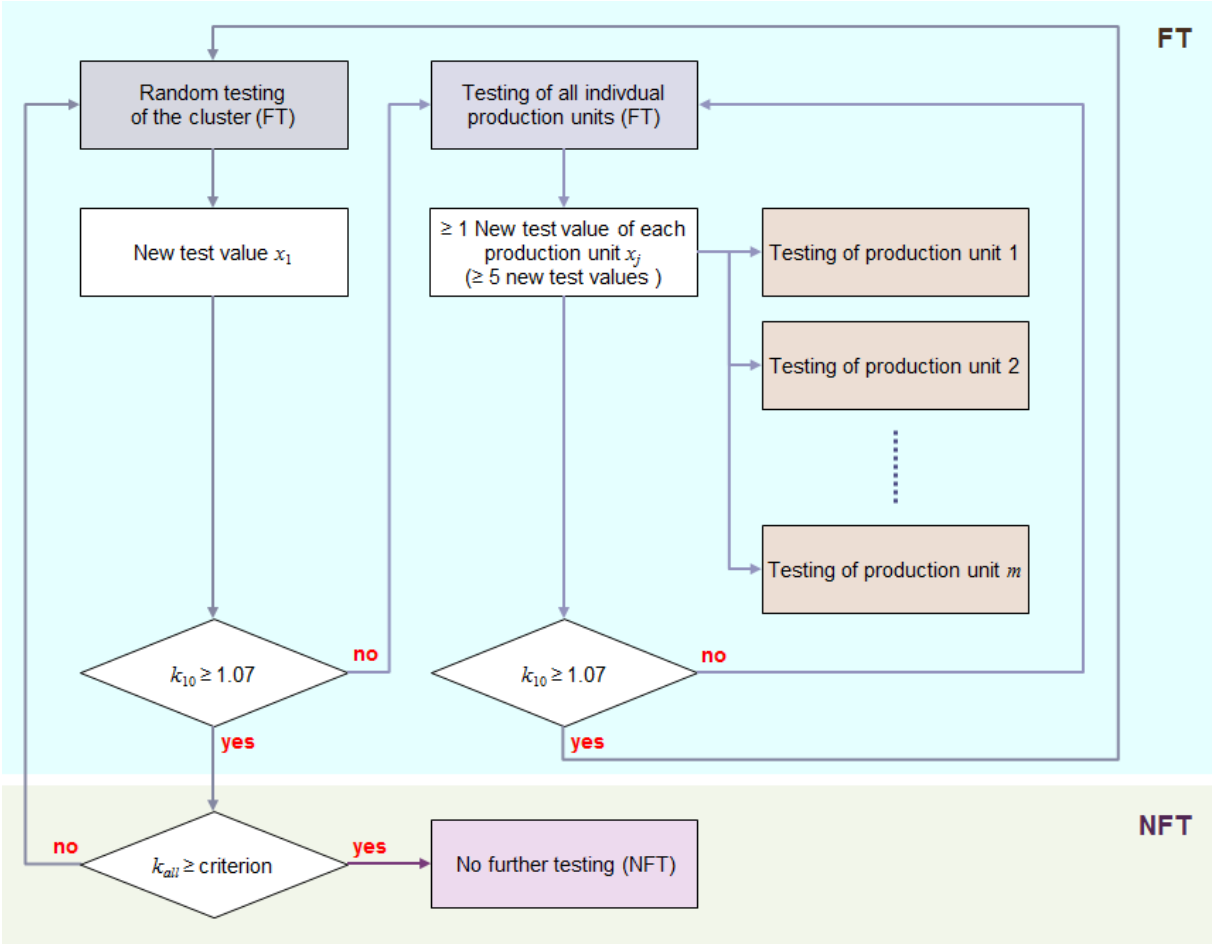


Figure 6 — FT cluster system based on 10 test values and assessment by variables

5.6.4 No-further-testing

In this document NFT is considered as an option within TT and FT when sufficient data are collected and there is sufficient proof that release values will stay within a certain range.

Because NFT is an irreversible decision, the criteria for NFT cannot be based on the same starting points as TT and FT. NFT assessment requires a more stringent approach, because in principle no new test data will be generated. Therefore the NFT criterion is based on a more stringent statistical basis and it is based on all available valid test results. Also additional organizational rules may be needed to apply NFT (see 5.4).

As with FT, the choice between assessment by variables and assessment by attributes depends on the assumption made about the distribution of the test values and the level of the declared value with respect to level of the detection limit of the test method. For NFT the considerations for the different situations are the same as for FT.

NOTE WFT in relation to product dossiers is not dealt with in this document.

5.7 Banned substances

In most products produced from primary materials banned substances should not be a problem with respect to the assessment of their presence. They are simply not allowed. In some products, however, e.g. recycled products, it is not always possible to entirely prevent or remove a banned substance from the input materials. For instance demolition waste may contain some asbestos, even when

asbestos is removed prior to demolition as much as possible. For these situations it is not realistic to set a declared value or regulatory limit value of zero or a value equal to the detection limit, but the declared value should be based on what is left after applying the best available techniques to remove a banned substance from waste.

NOTE E.g. in the Netherlands the regulatory limit value for the weighted content of asbestos in recycled aggregates equals 100 mg/kg, which is a calculated value equal to the serpentine asbestos content plus 10 times the amphibole asbestos content, while the detection limit equals 1 mg/kg.

For these specific cases it is possible to use the assessment criteria for dangerous substances as described in this report.

6 Assessment by variables

6.1 Single production units

6.1.1 Type testing

For assessment by variables TT consists of testing at least two batches. It is however recommended to base TT on at least five batches. Each batch needs to conform if it is being placed on the market. TT may end if the value of k_n is greater than or equal to the critical value k_0 according to Table 1. For 6 to 9 test values, k_5 is used to determine if TT may be ended.

Table 1 — Criteria for ending TT for a single production unit – assessment by variables

Number of test values *n*	Conformity criterion for the batch	Criterion for ending TT
1	$x_1 \leq L_D$	not allowed
2	$x_2 \leq L_D$	$k_2 \geq 2{,}18$
3	$x_3 \leq L_D$	$k_3 \geq 1{,}09$
4	$x_4 \leq L_D$	$k_4 \geq 0{,}82$
5	$x_5 \leq L_D$	$k_5 \geq 0{,}69$
6	$x_6 \leq L_D$	$k_5 \geq 0{,}69$
7	$x_7 \leq L_D$	$k_5 \geq 0{,}69$
8	$x_8 \leq L_D$	$k_5 \geq 0{,}69$
9	$x_9 \leq L_D$	$k_5 \geq 0{,}69$
10	$x_{10} \leq L_D$	$k_{10} \geq 0{,}44$

The coefficient k_n is calculated from:

$$k_n = \frac{\ln\left(L_D\right) - \overline{x}}{s} \tag{3}$$

where

n = number of test values;

x_i = last test value;

$\overline{x}$ = running mean of the last consecutive n ln-transformed test values;

s = running standard deviation of the last consecutive n ln-transformed test values;

L_D = declared value.

NOTE In the case that a normal or normalized distribution is more appropriate, the ln-transformation of test values and declared value is not applied.

TT ends when there are ten test values, but this does not necessarily mean the end of batch testing. This depends on the value of $k_{10.}$

6.1.2 Further-testing

Further-testing is based on the last five or ten consecutive test values. In general FT will be based on ten values, but for certain conditions (e.g. changes after taking a corrective measure) five may be more appropriate. The production conforms when there is random testing but for the continuation of random testing:

$$k_5 \geq 0,69 \text{ or } k_{10} \geq 0,44 \quad (4)$$

where

k_5 = value of k_n of the last consecutive 5 ln-transformed test values;

k_{10} = value of k_n of the last consecutive 10 ln-transformed test values.

The frequency of random testing should be in accordance with Table 2. If the last consecutive 5 or more test values are lower than the detection limit, the test frequency equals 1 batch per 3 years.

If the criterion for random testing is not satisfied ($k_5 < 0,69$ or $k_{10} < 0,44$) the production should be batch tested and the conformity criterion applied to each batch is:

$$x_1 \leq L_D \quad (5)$$

where

x_1 = last (most recent) test value;

L_D = declared value.

When batch testing no account is taken of the uncertainty of measurement; when random testing, the uncertainty of measurement is reflected in the standard deviation.

Table 2 — Test frequency for a single production unit – assessment by variables

$k_n = \frac{ln(UL) - \bar{x}}{s}$	Number of test values [a]	Test frequency[b]	Period of production
$k_2 \geq 24{,}58$ $k_3 \geq 9{,}65$ $k_4 \geq 7{,}13$ $k_5 \geq 6{,}11$ $k_{10} \geq 4{,}63$	2 3 4 5 10	1 batch per 3 years	3 years of production
$18{,}50 \leq k_2 < 24{,}58$ $7{,}34 \leq k_3 < 9{,}65$ $5{,}44 \leq k_4 < 7{,}13$ $4{,}67 \leq k_5 < 6{,}11$ $3{,}53 \leq k_{10} < 4{,}63$	2 3 4 5 10	1 batch per year	1 year of production
$10{,}25 \leq k_2 < 18{,}50$ $4{,}26 \leq k_3 < 7{,}34$ $3{,}19 \leq k_4 < 5{,}44$ $2{,}74 \leq k_5 < 4{,}67$ $2{,}07 \leq k_{10} < 3{,}53$	2 3 4 5 10	1:10 batches (≥5 batches per 3 years)	10 consecutive batches
$4{,}88 \leq k_2 < 10{,}25$ $2{,}23 \leq k_3 < 4{,}26$ $1{,}69 \leq k_4 < 3{,}19$ $1{,}46 \leq k_5 < 2{,}74$ $1{,}07 \leq k_{10} < 2{,}07$	2 3 4 5 10	1:4 batches (≥10 batches per 3 years)	4 consecutive batches
$2{,}18 \leq k_2 < 4{,}88$ $1{,}09 \leq k_3 < 2{,}23$ $0{,}82 \leq k_4 < 1{,}69$ $0{,}69 \leq k_5 < 1{,}46$ $0{,}44 \leq k_{10} < 1{,}07$	2 3 4 5 10	1:2 batches (≥5 batches per year)	2 consecutive batches

[a] The criteria for 2, 3 and 4 test values can only be used to bridge TT and FT as long as there are less than 5 test values available.

[b] For random testing the batch size should be not more than 1/10 of a year's production for continually produced products. For batch testing a smaller batch size could be used to reduce the impact of rejection.

Where random testing has led back into batch testing, it is necessary to be sure that the product quality has stabilized before returning to random testing, so additional requirements are applied. Once batch testing is required at least 5 new consecutive batches are tested and the criteria $k_5 \geq 0{,}69$ and $k_{10} \geq 0{,}44$ satisfied before it is permitted to return to random testing (see Figure 5). Satisfying the criterion for k5 shows that the latest production is stable enough, while satisfying the criterion for k_{10} shows consistency with the previous production.

An example of TT and FT assessment by variables of a single production unit is shown in Annex A (EXAMPLE 1).

6.1.3 Rule of application for products where the test values are significantly below the declared value (Gamma rule)

In the case of continual very low test values and a very low risk of exceeding the declared value the following criteria may be applied in addition to the previously described rules of application based on variables.

If the last consecutive 5 or last consecutive 10 test values are smaller than $\Gamma \times L_D$ a test frequency according to Table 3 to may be applied, where:

Γ = gamma factor;

L_D = declared value.

Because the use of the gamma rule assumes a lognormal distribution of the test values, it can only be used as an alternative for assessment by variables.

Table 3 — Test frequency – gamma factor

Number of test values	Gamma factor Γ	Test frequency
5 10	0,31 0,41	1 batch per year
5 10	0,19 0,26	1 batch per 3 years

6.1.4 Rule of application for products where a small number of test results may be expected

If it is expected that only few test data will be available for TT and FT (due to large variation in products and little production volumes), the use of the gamma rule as describes in 6.1.3 may provide an alternative to assessment by variables as long there are less than 5 test values.

The use of the gamma rule assumes that the coefficient of variation of the test values is known. Annex F shows gamma factors for different coefficients of variation. When there are 5 or more test values available for FT, assessment by variables is considered more appropriate and more reliable because knowledge of the coefficient of variation is not required.

When applying the gamma rule the declared value is multiplied by the so-called gamma factor and TT may be ended if:

$$\sum_{i=1}^{n} \{x_i > \Gamma \cdot L_D\} = 0 \tag{6}$$

where

x_i = i^{th} test value;

n = number of test results;

Γ = gamma factor according to Table 4 or Annex F;

L_D = declared value.

Table 4 — Criteria for ending TT – Gamma rule for a coefficient of variation of 0,65

Number of test values n	Conformity criterion for the batch	Criterion for ending TT
1	$x_1 \leq L_D$	not allowed
2	$x_2 \leq L_D$	$\sum_{i=1}^{2}(x_i > 0,38 \times L_D) = 0$
3	$x_3 \leq L_D$	$\sum_{i=1}^{3}(x_i > 0,64 \times L_D) = 0$
4	$x_4 \leq L_D$	$\sum_{i=1}^{4}(x_i > 0,82 \times L_D) = 0$

For leaching data a coefficient of variation of 0,65 may be applied to leaching data according to Dutch experiences.

The FT test frequency for random testing should be in accordance with Table 5. For batch testing the rules of 6.1.2 of apply.

6.2 Cluster of production units

6.2.1 Type testing

The total number of tests required for cluster TT depends on the product's description, with a minimum of 5 test results in total, but each production unit should contribute at least one test result. The critical value for k_n depends on the number of test results and should be selected from Table 6. Cluster TT is ended if the critical value for ending cluster TT is met ($k_n \geq k_0$). At this point cluster TT switches to cluster FT.

If the critical value for ending cluster TT cannot be met ($k_n < k_0$) it is not permitted to apply the cluster system to the set of production units and product as tested for the RDS under consideration. In this case it may be decided to change the cluster product by leaving out extreme products, leave out specific production units, select a higher declared value or even decide not to apply the cluster system and assess each production unit separately according to 6.1.1.

Table 5 — Test frequency for random testing – Gamma rule for a coefficient of variation of 0,65

Number of test values n	Gamma factor Γ	Test frequency [a]
2 3 4	0,12 0,15 0,18	$\sum_{i=1}^{n}(x_i > \Gamma \cdot L_D) = 0$ 1 batch per 3 years
2 3 4	0,19 0,24 0,28	$\sum_{i=1}^{n}(x_i > \Gamma \cdot L_D) = 0$ 1 batch per year

Number of test values n	Gamma factor Γ	Test frequency [a]
2 3 4	0,35 0,44 0,51	$\sum_{i=1}^{n}(x_i > \Gamma \cdot L_D) = 0$ 1:10 batches (≥5 batches per 3 years)
2 3 4	0,51 0,65 0,76	$\sum_{i=1}^{n}(x_i > \Gamma \cdot L_D) = 0$ 1:4 batches (≥10 batches per 3 years)
2 3 4	0,64 0,82 0,96	$\sum_{i=1}^{n}(x_i > \Gamma \cdot L_D) = 0$ 1:2 batches (≥5 batches per year)
[a] For random testing the batch size should be not more than 1/10 of a year's production for continually produced products. For batch testing a smaller batch size could be used to reduce the impact of rejection.		

Table 6 — Criteria for ending TT for a cluster – assessment by variables

Number of test values	Criterion for ending cluster TT
1 to 4	not allowed
5	k_5 ≥ 1,46
6	k_6 ≥ 1,32
7	k_7 ≥ 1,23
8	k_8 ≥ 1,16
9	k_9 ≥ 1,11
10	k_{10} ≥ 1,07
11	k_{11} ≥ 1,03
12	k_{12} ≥ 1,00
13	k_{13} ≥ 0,98
14	k_{14} ≥ 0,96
15	k_{15} ≥ 0,94
16	k_{16} ≥ 0,92
17	k_{17} ≥ 0,91
18	k_{18} ≥ 0,90

Number of test values	Criterion for ending cluster TT
19	$k_{19} \geq 0{,}88$
20	$k_{20} \geq 0{,}87$
> 20	See Annex E.

6.2.2 Further-testing

The frequency of random testing for a cluster is based on the last 5, 10 or 20 consecutive test results and should be in accordance with Table 7. If the criterion for cluster testing is not satisfied ($k_5 < 1{,}46$, or $k_{10} < 1{,}07$, or $k_{20} < 0{,}87$) the cluster system cannot be applied and each production unit should be test individually according to 6.1.2.

Table 7 — Test frequency for a cluster – assessment by variables

$k_n = \frac{\ln(UL) - \overline{x}}{s}$	Number of test values	Test frequency [a]	Period of production
$k_5 \geq 6{,}11$ $k_{10} \geq 4{,}63$ $k_{20} \geq 4{,}01$	5 10 20	10 batches per 3 years, unless there are less than 20 production units, in which case the test frequency equals 1 batch per 3 years per 2 production units	at least 6 years of production per production unit
$4{,}67 \leq k_5 < 6{,}11$ $3{,}53 \leq k_{10} < 4{,}63$ $3{,}05 \leq k_{20} < 4{,}01$	5 10 20	10 batches per year, unless there are less than 20 production units, in which case the test frequency equals 1 batch per year per 2 production units	at least 2 years of production per production unit
$2{,}74 \leq k_5 < 4{,}67$ $2{,}07 \leq k_{10} < 3{,}53$ $1{,}77 \leq k_{20} < 3{,}05$	5 10 20	1:10 batches (≥5 batches per 3 years)	10 consecutive batches
$1{,}46 \leq k_5 < 2{,}74$ $1{,}07 \leq k_{10} < 2{,}07$ $0{,}87 \leq k_{20} < 1{,}77$	5 10 20	1:4 batches (≥10 batches per 3 years)	4 consecutive batches
[a] The batch size should be not more than 1/10 of a year's production of a production unit for continually produced products.			

If the last consecutive 5 or more test values are lower than the detection limit, the test frequency equals 1 batch per 3 years.

Where random testing for a cluster has led back into testing single production units, it is necessary to be sure that the product quality has stabilized before returning to random testing for the cluster, so additional requirements are applied. Once testing single production units is required at least one new test value for each production unit should be produced and the criteria $k_5 \geq 1{,}46$ and $k_{10} \geq 1{,}07$ ($n = 5$ or 10) or $k_{10} \geq 1{,}46$ and $k_{20} \geq 0{,}87$ ($n = 20$) satisfied before it is allowed to return to random testing for the cluster (see Figure 6). Satisfying the criterion for k_5 (or k_{10} in case of a larger cluster) shows that

the latest production is stable enough, while satisfying the criterion for k_{10} (or k_{20} in case of a larger cluster) shows consistency with the previous production.

An example of TT and FT assessment by variables of a cluster is shown in Annex A (EXAMPLE 3).

6.2.3 Rule of application for cluster products where the test values are significantly below the declared value (Gamma rule)

In the case of continual very low test values and a very low risk of exceeding the declared value the following criteria may be applied in addition to the previously described rules of application for clusters based on variables.

If the last consecutive 5, last consecutive 10 or last consecutive 20 test values are smaller than $\Gamma \times L_D$ a test frequency according to Table 8 may be applied, where:

Γ = gamma factor;

L_D = declared value.

Table 8 — Test frequency – gamma factor

Number of test values	Gamma factor Γ	Test frequency
5 10 20	0,31 0,41 0,52	1 batch per year
5 10 20	0,19 0,26 0,33	1 batch per 3 years

6.3 No-further-testing

For a dangerous substance and a single production unit NFT is permitted if at least 5 test values are available and the value of k_n, which is based on all available test results, is equal to or greater than the critical value k_0 according to Table 9.

Table 9 — Criteria for NFT – assessment by variables

Number of test values	Criterion for NFT
1 – 4	not allowed
5	$k_5 \geq 5{,}36$
6	$k_6 \geq 4{,}41$
7	$k_7 \geq 3{,}86$
8	$k_8 \geq 3{,}50$
9	$k_9 \geq 3{,}24$
10	$k_{10} \geq 3{,}05$

Number of test values	Criterion for NFT
11	$k_{11} \geq 2{,}90$
12	$k_{12} \geq 2{,}78$
13	$k_{13} \geq 2{,}68$
14	$k_{14} \geq 2{,}59$
etc.	see Annex E
∞	$k_{\infty} \geq 1{,}28$

For clusters every production unit should have contributed at least one test value to the data set used to establish NFT, while the entire range of products that fall within the definition of the cluster product should be covered.

6.4 Handling values lower than the detection limit

Measured values lower than the detection limit are allocated 0,7 × the detection limit value.

6.5 Identifying outliers

In order to identify a suspicious smallest or largest test value as an outlier, the following procedure, called Grubbs' test, may be helpful [13].

The largest test value of n consecutive test values should be considered an outlier when

$$\frac{\ln\left(x_{max}\right)-\overline{x}}{s} > G_{\mathrm{P}} \tag{7}$$

The smallest test value of n consecutive test values should be considered an outlier when

$$\frac{\overline{x}-\ln\left(x_{\mathrm{min}}\right)}{s} > G_{\mathrm{P}} \tag{8}$$

where

x_{max} = largest test value of *n* test values,

x_{min} = smallest test value of *n* test values,

$\overline{x}$ = mean of *n* ln-transformed test values,

s = standard deviation *n* ln-transformed test values,

G_{P} = 1 % critical value according to Grubbs' test (see Table 10 and Annex E).

Table 10 — Critical value GP for testing outliers

Number of test values	Critical value G_P
10	2,482
20	3,001
30	3,236
40	3,381

When a test value has been identified as an outlier, the cause should be investigated and documented. Outliers that are proven to be due to sampling or testing errors may be rejected. If a valid outlier identifies a batch that is significantly above the declared value, the producer should consider not placing this batch on the market (if practical) and take urgent action to identify the cause of the outlier and take corrective action. If such a batch is not placed on the market the test result may be removed from the database.

6.6 Choosing a declared value

Contrary to set class limit values a producer may benefit from the declaration of maximum release values, because the producer should find a balance between the value of the declared value, testing effort and product image or declare a value that leads to NFT. A low declared value will probably result in a high test frequency or even batch testing, while a high declared value will give a low test frequency. On the other hand a high declared value may not compare well against competitive construction products having lower declared values.

To estimate a suitable declared value Formula (3) is reversed:

$$L_D = \exp\left(\overline{x} + k_0 \times s\right) \tag{9}$$

where

L_D = declared value,

$\overline{x}$ = mean of *n* ln-transformed test values,

s = standard deviation of *n* ln-transformed test values,

k_0 = critical value, which depends on *n* and the risk of exceeding the declared value,

NOTE See Annex E for the risk of exceeding the declared value.

n = number of test values.

Annex E shows values of k_0 for a risk of exceeding the declared value of (0,1 to 50) %. It is recommended to use a risk of not more than 10 %, because of the high probability of batch testing when a relatively high risk (30 % or more) is used.

7 Assessment by attributes

7.1 Single production units

7.1.1 Type testing

For assessment by attributes TT consists of testing at least four batches. It is however recommended to base TT on at least seven batches. Each batch needs to conform if it is being placed on the market. TT may end if n_e (the number of test values that exceed the declared value) is not greater than the acceptable number of test results that are permitted to exceed the declared value (n_a) according to Table 11.

Table 11 — Criteria for ending TT for a single production unit – Assessment by attributes

Number of test values n	Conformity criterion for the batch	Criterion for ending TT
1	$x_1 \leq L_D$	not allowed
2	$x_2 \leq L_D$	not allowed
3	$x_3 \leq L_D$	not allowed
4	$x_4 \leq L_D$	$n_e = 0$
5	$x_5 \leq L_D$	$n_e = 0$
6	$x_6 \leq L_D$	$n_e = 0$
7	$x_7 \leq L_D$	$n_e \leq 1$
8	$x_8 \leq L_D$	$n_e \leq 1$
9	$x_9 \leq L_D$	$n_e \leq 1$
10	$x_{10} \leq L_D$	$n_e \leq 1$
11	$x_{11} \leq L_D$	$n_e \leq 1$
12	$x_{12} \leq L_D$	$n_e \leq 3$

A test result is considered non-conforming if

$$x > L_D \tag{10}$$

where

x = test value;

L_D = declared value.

TT based on assessment by attributes ends when there are twelve test values, but this does not necessarily mean the end of batch testing. This depends on the number of test values that exceed the declared value.

7.1.2 Further-testing

Further-testing is based on the last seven or twelve consecutive test values. In general FT will be based on twelve values, but for certain conditions (e.g. changes after taking a corrective measure) seven may be more appropriate. The production conforms when there is random testing but for the continuation of random testing:

$$n_e \leq 1 \ (n = 7) \text{ or } n_e \leq 3 \ (n = 12) \tag{11}$$

where

n_e = number of test results that exceed the declared value within the last consecutive n test values.

The frequency of testing should be in accordance with Table 12.

Table 12 — Test frequency for a single production unit – Assessment by attributes

n_e	Number of test values n	Test frequency [a]	Period of production
$n_e = 0$ $n_e \leq 1$	230 388	1 batch per year	1 year of production
$n_e = 0$ $n_e \leq 1$	22 38	1:10 batches (≥5 batches per 3 years)	10 consecutive batches
$n_e \leq 1$ $n_e \leq 3$	12 21	1:4 batches (≥10 batches per 3 years)	4 consecutive batches
$n_e = 0$ $n_e \leq 1$ $n_e \leq 3$	4 to 6 [b] 7 12	1:2 batches (≥5 batches per year)	2 consecutive batches

[a] For random testing the batch size should be not more than 1/10 of a year's production for continually produced products. For batch testing a smaller batch size could be used to reduce the impact of rejection.

[b] Only applicable where there are less than 7 test values available (at the start of FT following TT).

If the last consecutive 5 or more test values are lower than the detection limit, the test frequency equals 1 batch per 3 years.

If the criterion for random testing is not satisfied $n_e > 1$ ($n = 7$) or $n_e > 3$ ($n = 12$), the production should be batch tested and the conformity criterion applied to each batch is:

$$x_1 \leq L_D \tag{12}$$

where

x_1 last (most recent) test value;

L_D declared value.

When determining whether a batch exceeds the declared value no account is taken of the uncertainty of measurement. When random testing, this uncertainty is reflected in the acceptable number of test results that exceed the declared value.

In alignment with assessment by variables, where random testing has led back into batch testing, it is necessary to be sure that the product quality has stabilized before returning to random testing, so additional requirements are applied. At least 5 new consecutive batches need to be tested and the criterion $n_e \leq 1$ for the consecutive last 7 test results and $n_e \leq 3$ for the consecutive last 12 test results satisfied before it is permitted to return to random testing: satisfying the criterion for the consecutive last 7 test results shows that the latest production is stable enough, while satisfying the criterion for the consecutive last 12 test results shows consistency with the previous production.

An example of TT and FT assessment by attributes is shown in Annex A (EXAMPLE 2).

7.2 Cluster of production units

7.2.1 Type testing

The total number of tests required for cluster TT depends on the product's description, with a minimum of 7 test results in total, but each production unit should be required to contribute at least one test result. The acceptable number of test results that exceed the declared value depends on the number of test results and should be selected from Table 13.

Cluster TT is ended is if the actual number of test results that exceed the declared value (n_e) is equal to or smaller than the acceptable number of test results that are permitted to exceed the declared value (n_a). At this point cluster TT switches to cluster FT.

If the acceptable number of test results that are permitted to exceed the declared value for ending cluster TT cannot be met ($n_e > n_a$) it is not permitted to apply the cluster system to the set of production units and product as tested for the RDS under consideration. In this case it may be decided to change the cluster product by leaving out extreme products, leave out specific production units, select a higher declared value or even decide not to apply the cluster system and assess each production unit separately according to 7.1.1.

Table 13 — Criteria for ending TT for a cluster – assessment by attributes

Number of test values	Criterion for ending cluster TT
1 to 6	not allowed
7 to 11	$n_e = 0$
12 to 15	$n_e \leq 1$
16 to 20	$n_e \leq 2$
21 to 24	$n_e \leq 3$
25 to 28	$n_e \leq 4$
29 to 32	$n_e \leq 5$
33 to 36	$n_e \leq 6$
37 to 40	$n_e \leq 7$

Number of test values	Criterion for ending cluster TT
41 to 44	$n_e \leq 8$
45 to 48	$n_e \leq 9$
49 to 52	$n_e \leq 10$
53 to 56	$n_e \leq 11$
57 to 59	$n_e \leq 12$
60 to 63	$n_e \leq 13$
64 to 67	$n_e \leq 14$
68 to 71	$n_e \leq 15$

7.2.2 Further-testing

The frequency of random testing for a cluster is based on the last n consecutive test results as given in Table 14. If the criterion for cluster testing is not satisfied the cluster system cannot be applied and each production unit should be test individually according to 7.1.2.

Table 14 — Test frequency for a cluster – assessment by attributes

n_e	Number of test values n	Test frequency [a]	Period of production
$n_e = 0$ $n_e \leq 1$ $n_e \leq 2$	230 388 531	10 batches per year, unless there are less than 20 production units, in which case the test frequency equals 1 batch per year per 2 production units	at least 2 years of production per production unit
$n_e = 0$ $n_e \leq 1$ $n_e \leq 2$	22 38 52	1:10 batches (≥5 batches per 3 years)	10 consecutive batches
$n_e = 0$ $n_e \leq 1$ $n_e \leq 3$ $n_e \leq 5$	7 [b] 12 21 29	1:4 batches (≥10 batches per 3 years)	4 consecutive batches

[a] The batch size should be not more than 1/10 of a year's production of a production unit for continually produced products.
[b] Only applicable when less than 12 test results are available.

If the last consecutive 5 or more test values are lower than the detection limit, the test frequency equals 1 batch per 3 years.

Where random testing for a cluster has led back into testing single production units, it is necessary to be sure that the product quality has stabilized before returning to random testing for the cluster, so additional requirements are applied. Once testing single production units is required at least one new

test value for each production unit should be produced and $n_e \leq 1$ ($n = 12$) or $n_e \leq 3$ ($n = 21$) or $n_e \leq 5$ ($n = 29$), depending on the available number of test values, before it is allowed to return to random testing for the cluster.

An example of TT and FT assessment by variables of a cluster is shown in Annex A (EXAMPLE 4).

7.3 No-further-testing

For a dangerous substance NFT is allowed if at least 44 test values are available and the number of test results that exceed the declared value n_e, which should be based on all available test results, is equal to or smaller than the acceptable number of test results that are permitted to exceed the declared value n_a according to Table 15.

For clusters every production unit should have contributed at least one test value to the data set used to establish NFT, while the entire range of products that fall within the definition of the cluster product should be covered.

Table 15 —Criteria for NFT – assessment by attributes

Number of test values	Criterion for NFT
1 to 43	not allowed
44 to 63	$n_e \leq 0$
64 to 80	$n_e \leq 1$
81 to 96	$n_e \leq 2$
97 to 112	$n_e \leq 3$
113 to 126	$n_e \leq 4$
127 to 141	$n_e \leq 5$
142 to 155	$n_e \leq 6$
156 to 169	$n_e \leq 7$
170 to 182	$n_e \leq 8$
183 to 196	$n_e \leq 9$
197 to 209	$n_e \leq 10$

7.4 Handling values lower than the detection limit

Measured values lower than the detection limit are allocated 0,7 × the detection limit value.

7.5 Identifying outliers

There is no appropriate statistical test for outliers when the distribution is unknown. In this situation the process data should be checked to determine if there is a valid reason for excluding this result. In this

situation Grubbs' test as described in 6.5 cannot be used to determine statistical outliers as Grubbs' test requires a data set with a more or less (log)normal distribution.

7.6 Choosing a declared value

Contrary to set class limit values a producer may benefit from the declaration of maximum release values, because the producer should find a balance between the value of the declared value, testing effort and product image or declare a value that leads to NFT, see 6.6.

In the case of assessment by attributes it is recommended to choose a declared value equal to or greater than the highest test value, depending on the number of available test values.

8 Statistical principles of the rules of application

8.1 Introduction

A batch represents the amount of construction product that is representative for the intended use of that product. If the average release, emission or content of a RDS from that batch is not greater than the declared value, the batch when considered as a whole poses no significant risk to the environment. Therefore the process of sampling a batch and testing samples should provide a test result that approximates to the true value of a RDS related property as close as possible.

The proposed system of statistical assessment of declared values is based on the Dutch Soil Quality Decree. The statistical assessment of declared values is based on the conformity of a part of the production that comprises one or more batches of approximately equal size. The number of batches in this production part is a function of the frequency of testing; the size of the production part varies but the batch size is fixed. Batches are sampled by taking at least one sample which is tested for the substance(s) of interest. Sampling should be such the sample is representative for the batch, meaning that the sample should produce a test value close to the true average value of that batch. Conformity of a part of the production is established by comparing the mean value of the tested batch(es) with the declared value. Because testing random batches introduces risks of accepting batches which exceed this value, the assessment should include a safety margin to limit the risk that the potential release from a batch placed on the market exceeds the declared value. This risk is set at a maximum of about 10 % (see 5.2).

NOTE 1 Some aspects of the Dutch Soil Quality Decree, however, do not match the basic characteristics introduced in 5.2. The Dutch system is aimed at the 'input' of FPC and is designed to establish whether the mean value of the entire production fulfils the regulatory limit value (to establish with a probability of 90 % that the 50th percentile of the production does not exceed the regulatory limit value). The proposed assessment system is aimed at the 'output' of FPC and is designed to establish that the risk of a batch which exceeds the declared value being placed on the market is not greater than about 10 %.

NOTE 2 Samples are assumed to reflect the properties of the batch. Therefore, it is not required to define further requirements for the properties of samples (size, number of increments, etc.) within the scope of this document, as the percentage > L_D is independent of the number of sample increments and other properties of the samples.

When assessing whether the production is suitable for random testing, test values from rejected batches obtained during batch testing are not removed from the data set unless it is obvious and explainable that a rejected batch is not according to the normal production (see 6.5 and 7.5). This differs from testing technical properties where failing batches can be related to different production circumstances in most cases.

8.2 Assessment of a production part

NOTE In this section the statistical principle is mainly explained on the basis of assessment by variables. For assessment by attributes the principles are similar. The only difference is the method of handling the test results.

The risk of placing a batch which exceeds the declared value on the market is related to the running mean and running standard deviation within a 'running' part of the production (see Figure 7).

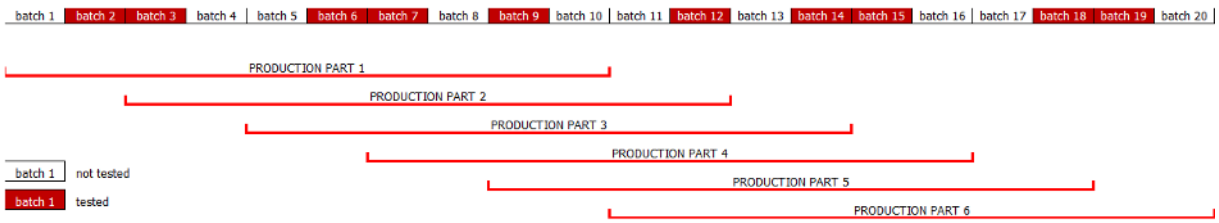


Figure 7 — Testing consecutive production parts in case of randomly testing every 1 of 2 batches

In case of assessment by variables, the criterion for random testing equals:

$$\overline{x} + k \cdot s \leq L_{\mathrm{D}} \tag{13}$$

where

$\overline{x}$ = running mean of n test values,

s = running standard deviation of n test values,

L_{D} = declared value,

k = multiplication factor based on the non-central t-distribution. This factor depends on the number of batches and the true percentage of production exceeding the declared value,

n = number of batches.

In order to limit the consumer's risk, e.g. the risk of a batch which exceeds the declared value being placed on the market, the mean of a production part is established with a probability of at least 90 % (uncertainty less than 10 %). In other words: **a part of the production of which the 50th percentile equals the declared value has a probability of acceptance of not more than 10 %**. On the other hand this criterion should not lead to rejection of a conforming production part, e.g. a production part of which the 90th percentile equals the declared value. Therefore **a part of the production of which the 90th percentile equals the declared value has a probability of acceptance of at least 90 %**. This means that the assessment is based on an OC-curve that falls within the light (grey) shaded area of Figure 8. In the case of assessment by variables the OC-curve defined by n = 5 and k = 0,69 fulfils the minimum requirements of both criteria. Parts of the production with a true mean greater than the declared value have a probability of acceptance less than 10 %.

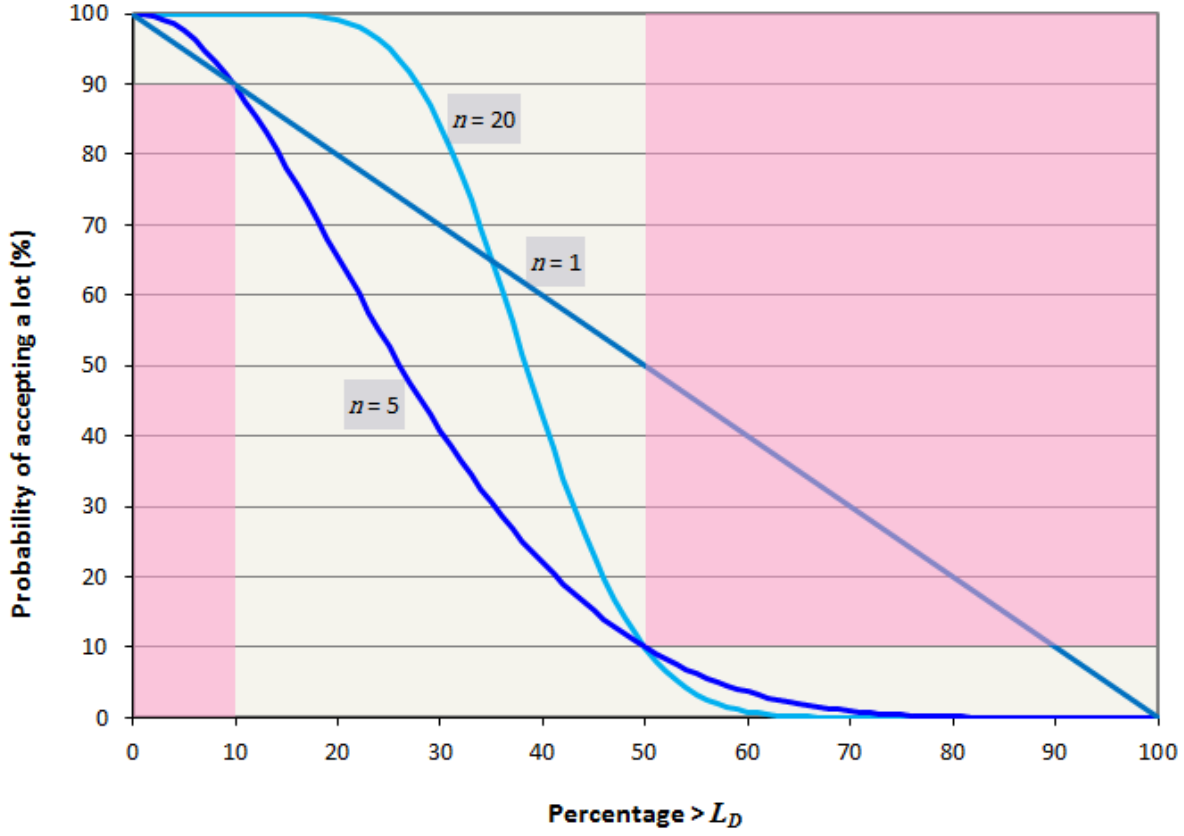


Figure 8 — Criteria for a conforming part of the production – OC-curves for 1, 5 and 20 samples based on assessment by variables

The percentage of production exceeding the declared value is defined at a given batch size. The variation in test values reflects the variation of the measured property between batches, the variation within a batch and the error of the test method.

Figure 8 shows that random testing on the basis of only one batch per production part may lead to an unacceptable high risk of accepting a batch that exceed the declared value, because the 90 % probability will not be met if the percentage > L_D lies between 50 % and 90 %. Therefore testing conformity of a part of the production should be based on more than one batch, preferably at least five batches.

Figure 9 shows the probability of random testing and batch testing based on the OC-curve for n = 5 and k = 0,69. For instance for a product that is characterized by 40 % non-conforming batches the probability that random testing is concluded from the assessment is 22 % and the probability that batch testing is concluded from the assessment is 78 % when the assessment is based on the mean and standard deviation of 5 test values.

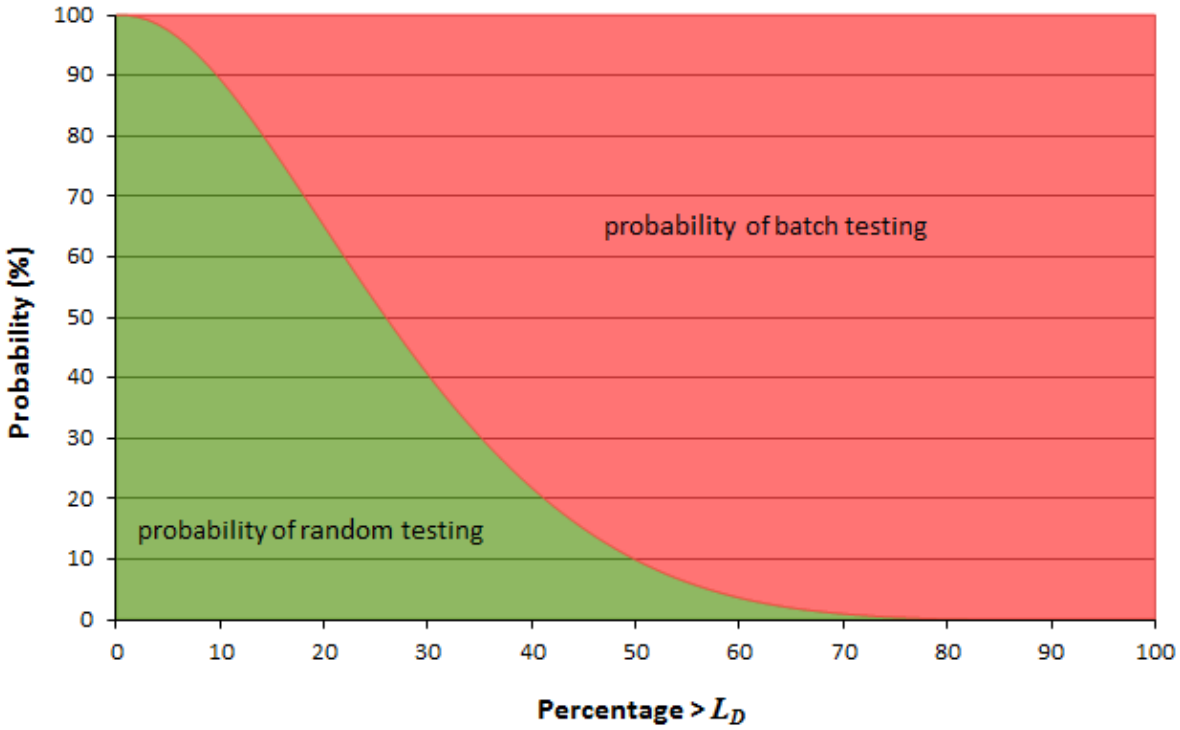
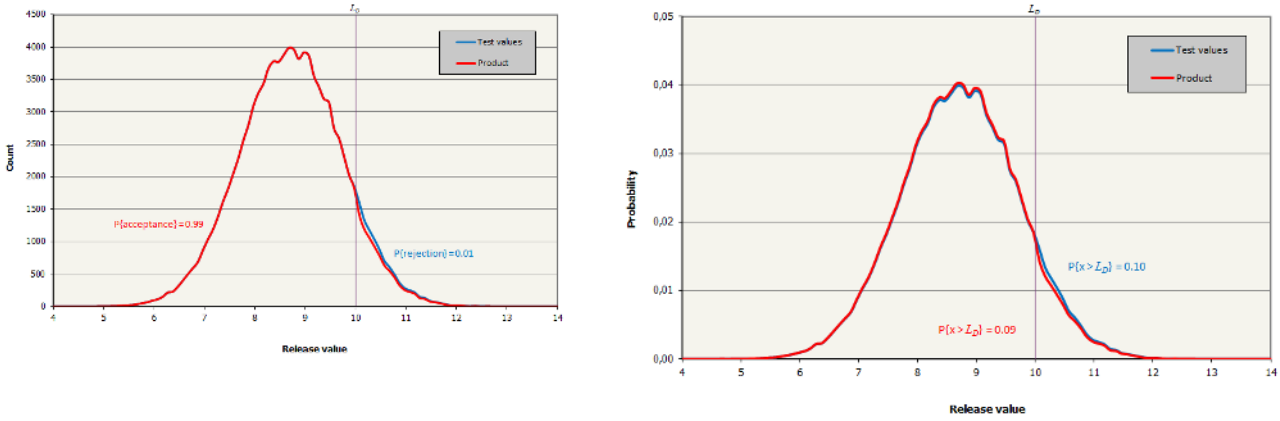


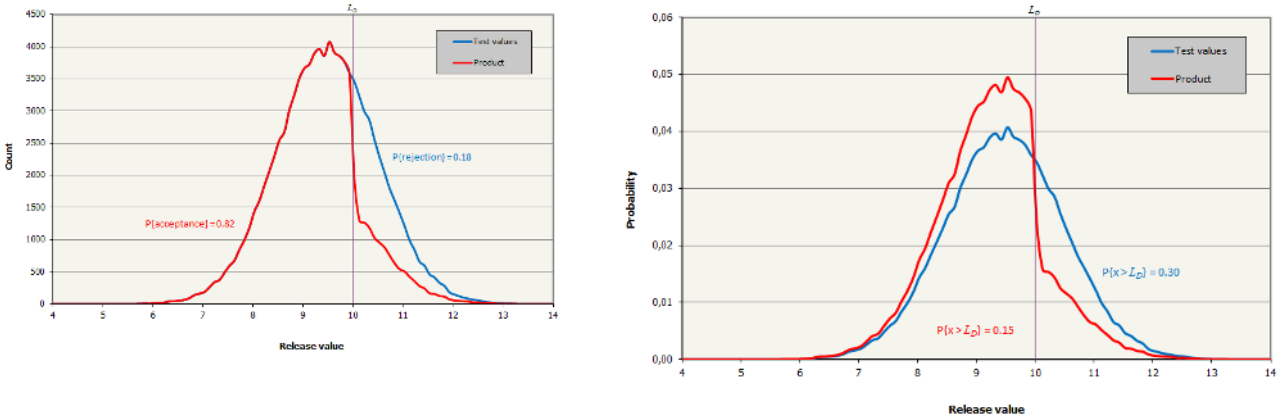
Figure 9 — Probability of random testing and batch testing – assessment by variables 5 batches per production part (n = 5, k = 0,69); P{random testing} + P{batch testing} = 1

In order to show the effects of random testing and batch testing on the product quality, Figure 10 shows the difference between the distribution of the entire production ('input quality', blue lines) and the distribution of the accepted batches after the assessment, excluding batches that exceed the declared value in case of batch testing ('output quality', red lines). The latter equals the quality of the product that is placed on the market.

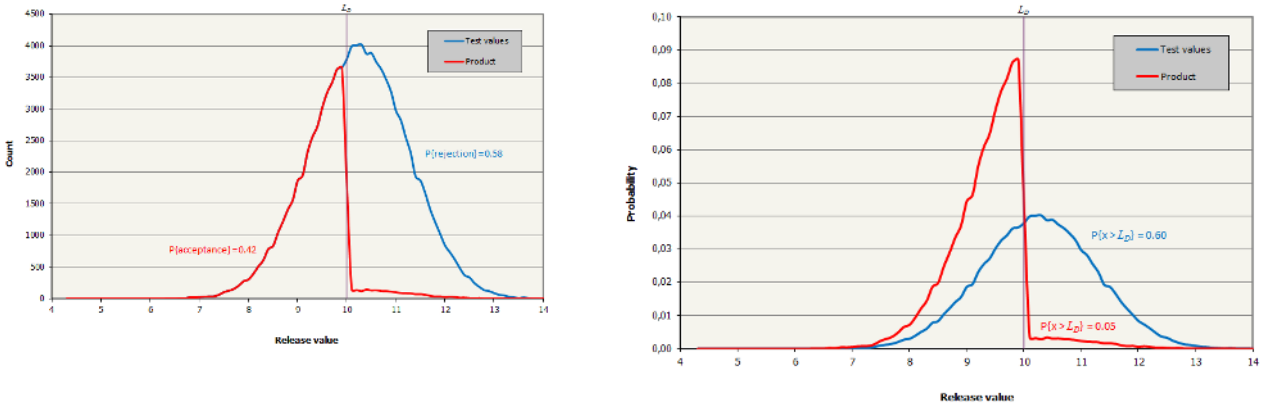
The two top graphs show the result for a normally distributed production with 10 % > L_D. Comparing the blue and red lines, there is hardly any difference between the distribution of the production and the distribution of the accepted batches, because the probability of rejecting a batch is rather low (1 %). In this case the product that is placed on the market is characterized by 9 % > L_D.



a) 10 % > L_D



b) 30 % > L_D



c) 60 % > L_D

Figure 10 — Change of the histogram (left) and distribution (right) of three different product qualities (from top to bottom a production with 10 %, 30 % and 60 % > L_D) due to combining random testing and batch testing – assessment by variables 5 batches per production part (n = 5, k = 0,69). The blue lines represent the quality of the entire production and the red lines represent the quality of the product that is placed on the market

The two middle graphs show the result for a normally distributed production with 30 % > L_D. Comparing the blue and red lines, there is a large difference between the distribution of the production and the distribution of the accepted batches, because the probability of rejecting a batch is substantial (18 %). The histogram in the left graph clearly shows the effect of the assessment method. A high proportion of the values greater than the declared value have been rejected and removed from the population of the 'output', causing the right tail of the original histogram being cut off (effect of batch testing). Some of the values greater than the declared value remain (effect of random testing).

As a result, the distribution also changes (right graph). Due to cutting off part of the high values, the distribution of the 'output' has a higher peak and shorter right tail than the normally distributed 'input'. Also, the mean and standard deviation of the 'output' are both lower than the mean and standard deviation of the 'input'. In this case the product that is placed on the market is characterized by 15 % > L_D.

The two bottom graphs show the result for a normally distributed production with an extreme 60 % > L_D and a probability of rejecting a batch equal to 58 %. In this case the product that is placed on the market is characterized by 5 % > L_D, while the distribution of the 'output' has a high skewness and almost no right tail.

8.3 Test error

Within the random testing procedure the test error (uncertainty of the test method) is included in the assessment; however, the influence of the test error on the results of the assessment (assuming no significant bias) becomes smaller as the number of test results increases (the test error is reduced by a factor equal to the square root of the number of test results). Therefore, the test error has only a small effect on the consumer's and producer's risk in the case of random testing (the test error is reduced by a factor equal to the square root of the number of test results).

Within the batch testing procedure the test error does affect the result of the assessment, because in practice only one test result will be available due to economic reasons. There are three possibilities to deal with this error:

1) no correction of the declared value;

2) a positive correction of the declared value;

3) a negative correction of the declared value.

Ignoring the test error will not affect the consumer's risk and producer's risk, because the test error may give a lower as well as a higher test value (the probability of both effects is the same). The average effect is zero.

Adding a correction value to the declared value increases the consumer's risk, because the threshold value to accept a batch is raised so that the probability to accept a batch which exceeds the declared value is higher. It does not affect the producer's risk, which remains zero.

Subtracting a correction value from the declared value decreases the consumer's risk and increases the producer's risk, because the threshold value to accept a batch is lowered so that the probability to accept a batch which exceeds the declared value becomes lower and the probability to reject a conforming batch becomes higher.

In one way or another, a positive and negative correction has an undesirable effect for the customer or producer. Therefore the test error is chosen to be ignored in the proposed assessment method.

NOTE The effect of the test error will become smaller if more than one sample is tested per batch. If the test error is relatively large with respect to the declared value and the probability of batch testing is substantial, it may

be attractive for a producer to test more than one sample per batch in the case of batch testing. Even in the case of random testing this might be attractive, because this will lower the probability of batch testing.

8.4 Assessment by variables

8.4.1 Type testing and further-testing

Assessment by variables is based on the assumption that test values have a normal or lognormal distribution. In this case the statistics developed for normal distributions are applicable and the assessment is based on the continuous random variable k_n (see Clause 6). The value of k_n is distributed according to a non-central t-distribution, which is a generalization of the Student's t-distribution using a non-centrality parameter. In the case of sampling the estimated true sample mean may deviate from the true mean of the production. The non-central t-distribution gives the probability that a t-test will correctly reject a false null hypothesis of mean μ when the production mean is actually μ_0 (H_0: $\mu = \mu_0$); that is, it gives the power of the t-test. The power increases as the difference $\mu_0 - \mu$ increases, and also as the sample size n increases [8]. Critical values for k_n depend on the number of test values and the probability that a certain percentile of the production exceeds the declared value L_D.

The probabilities of ending TT for the values of (n,k_0) in Table 1 are shown in Figure 11 and the probabilities of achieving random testing during FT are shown in Figure 12. According to Figure 11 the criterion for a part of the production of which the 90th percentile equals the declared value with a probability of at least 90 % is not met in the case of $n < 5$ and the producer's risk is greater than 10 % (43 % for $n = 2$).

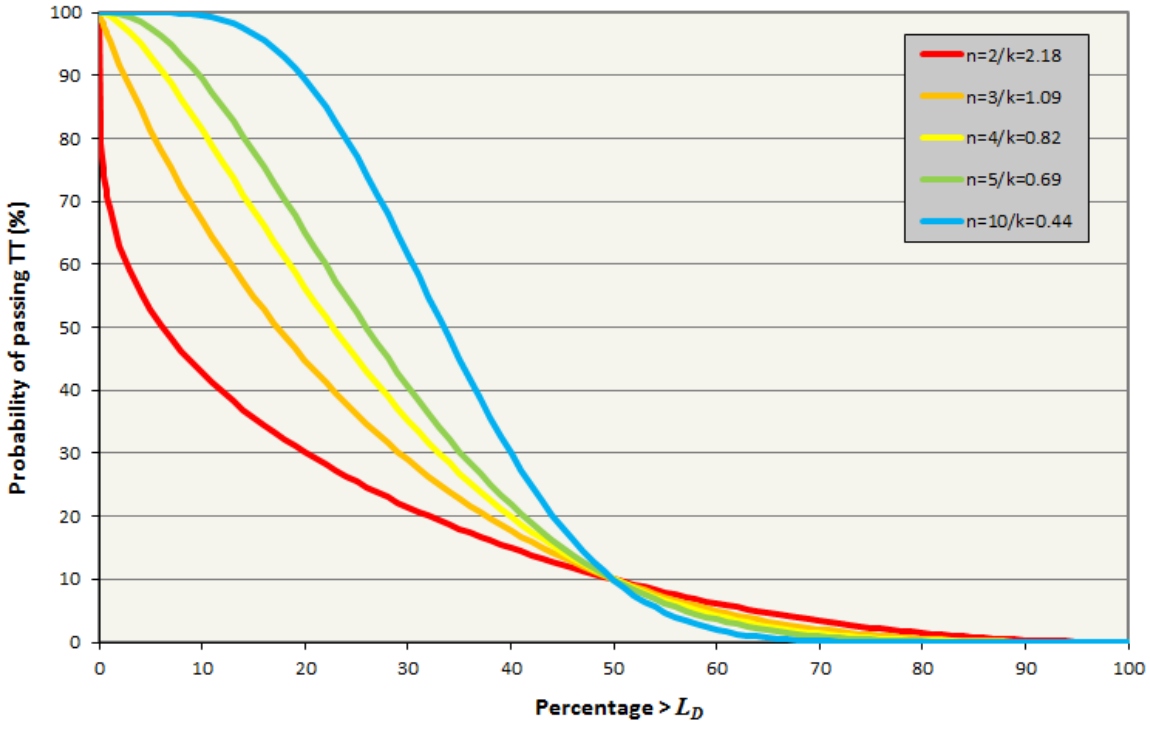


Figure 11 — Probability of ending TT – assessment by variables (OC-curves)

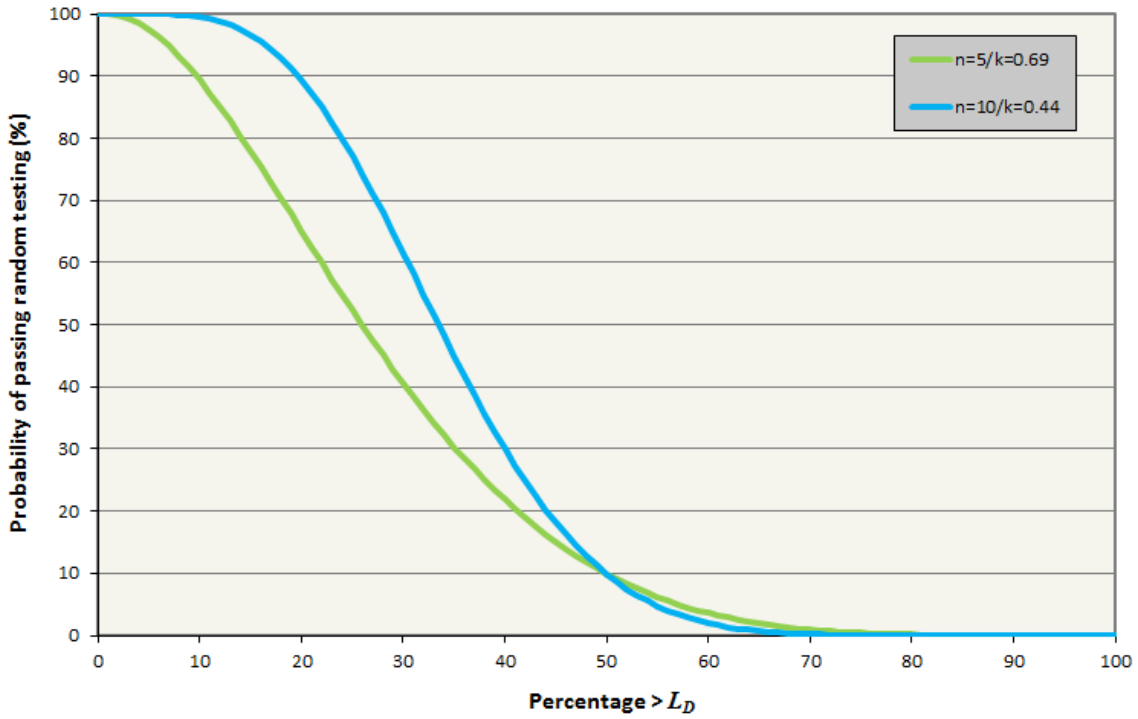


Figure 12 — Probability of achieving random testing for FT – assessment by variables (OC-curves)

8.4.2 Test frequency for further-testing

The proposed method of assessment includes a risk-based variable test frequency. The lower the risk of exceeding the declared value, the lower is the test frequency and vice versa. Only in the case of random testing is the test frequency variable. Within the conditions for random testing, the test frequency is linked to the percentage of the production that exceeds the declared value: to apply the test frequency linked to a percentage of the production exceeding the declared value of p % the (1-p)th percentile of the production is less than or equal to the declared value with a probability of at least 90 %. Within random testing the 70th, 90th, 99th and 99,9th percentiles are used to control the test frequency. Again fulfilment of this criterion is estimated from the value of k_n (assessment by variables according to Clause 6).

Critical values for k_n for different percentiles and a probability of 90 % are given in Table 16. In the general case test frequencies are based on 5, 10 or 20 test values and these percentiles, see Table 2 and Table 7. The probabilities of achieving the frequency criteria are shown in Figure 13.

Table 16 — Critical values for kn for different percentiles and a probability of 90 %

Percentage > L_D p	**Percentile**	**Critical value k_0**		
		n = 5	**n = 10**	**n = 20**
50 %	50th	0,69	0,44	not applicable
30 %	70th	1,46	1,07	0,87
10 %	90th	2,74	2,07	1,77
1 %	99th	4,67	3,53	3,05
0,1 %	99,9th	6,11	4,63	4,01

The criteria for batch testing are included in these tables and figures, because the system of batch testing and random testing as a whole constitutes a complete set of complementary criteria.

Figure 14 shows the probability of a testing frequency being applied for n = 5. Other values of n produce similar graphs. According to this, it appears that a certain product quality (defined by the percentage > L_D) is not bound to one test frequency. As a result the test frequency may vary even though the product quality remains unchanged. This is illustrated in another way in Figure 15, which shows the 80 % probability range of possible values of k_5 for different percentages of the production exceeding the declared value for n = 5. This figure shows that the 80 % range of possible k_5 values is broad enough to contain two or three different test frequencies.

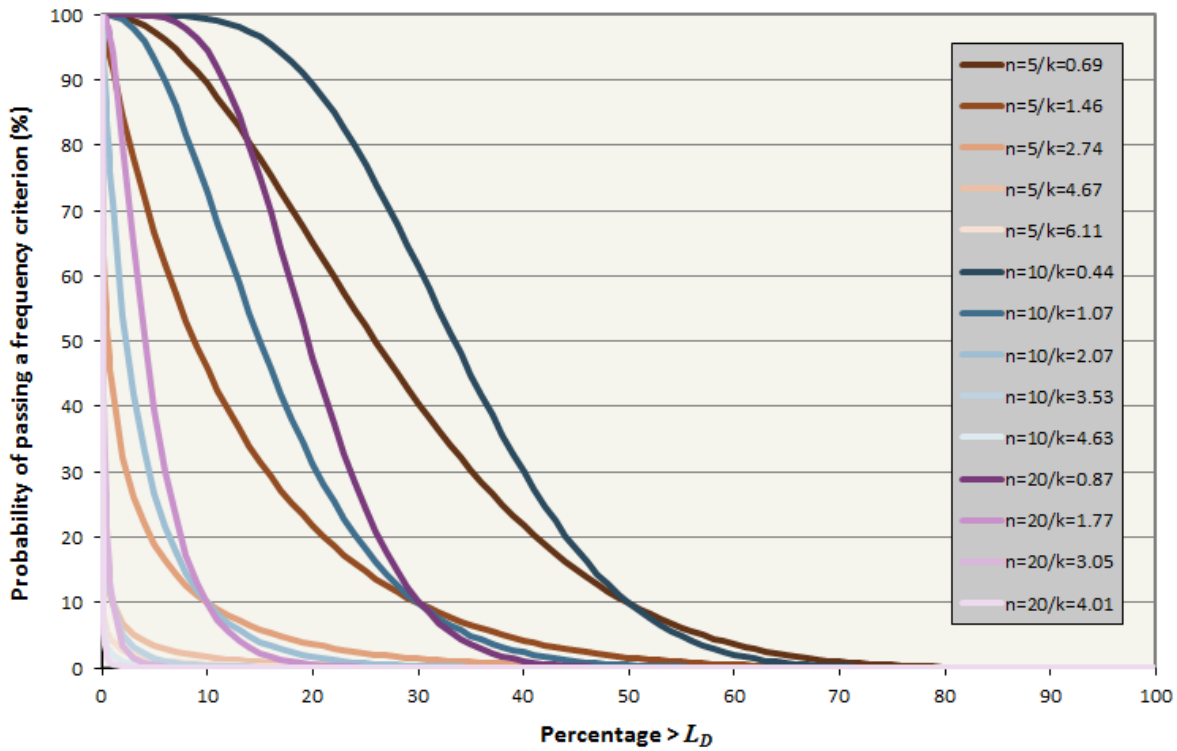


Figure 13 — Probability of achieving frequency criteria – assessment by variables (OC-curves)

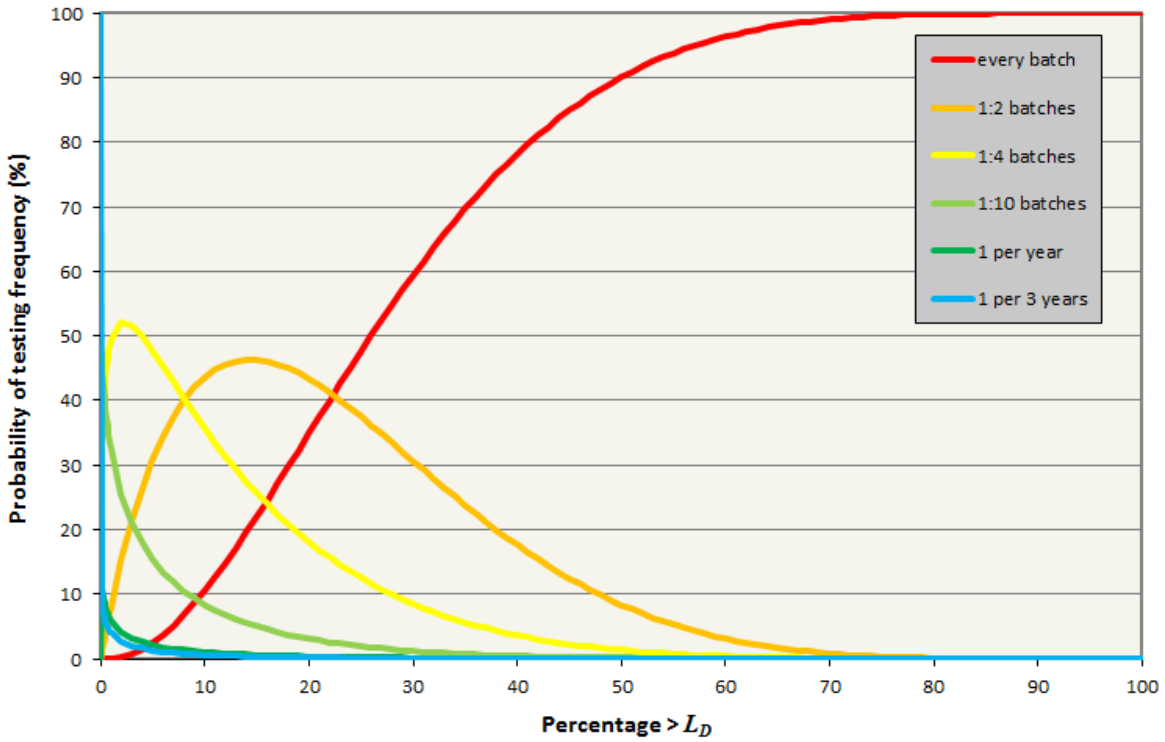


Figure 14 — Probability of achieving a specific testing frequency for n = 5 – assessment by variables (curves have been derived from the OC-curves of Figure 13)

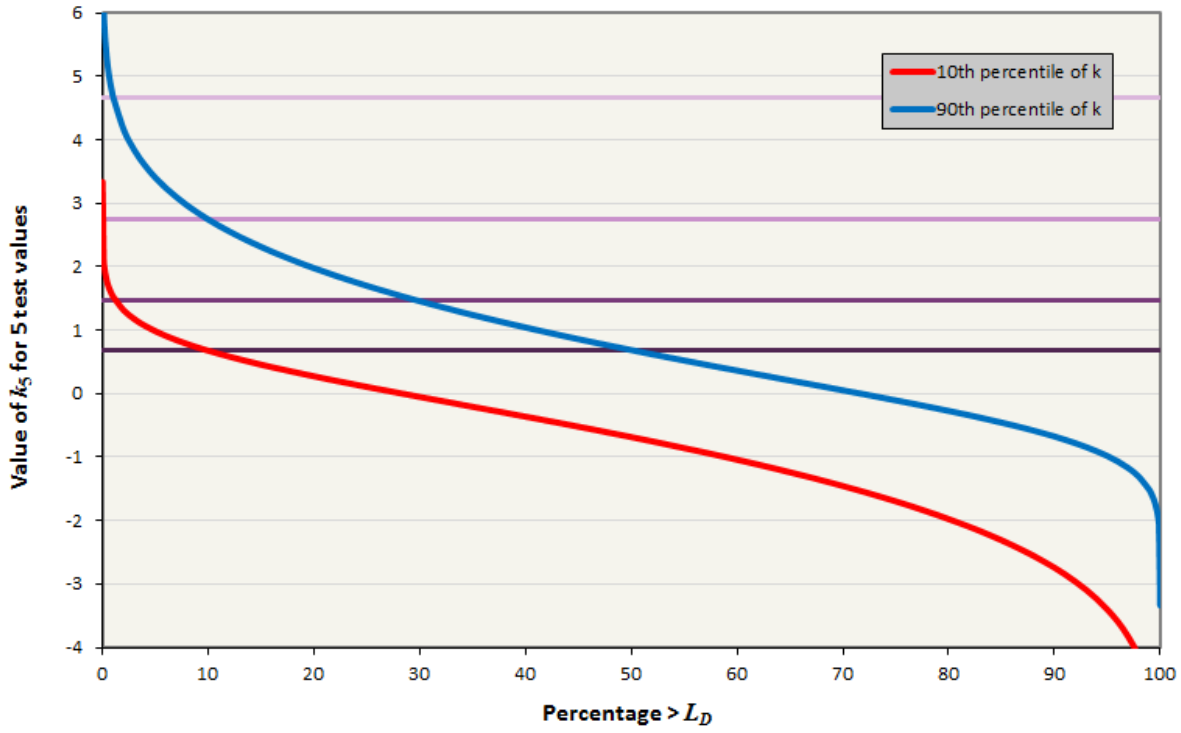


Figure 15 — Range of possible k_5 values for n = 5 — Assessment by variables

Despite the possibility of different test frequencies for the same product quality, the test frequencies in Table 2 and Table 7 are well balanced, because in practice a producer will always be stimulated to aim at the lowest possible test frequency, meaning that variations in the value of k_n due to random product variability will be lower in practice than suggested by these figures. Figure 14 and Figure 15 are based on total randomness which may not be realistic.

Furthermore it should not be forgotten that the consumer's risk is sufficiently low, because variable test frequencies are bound to random testing only and do not lead to an increase of the consumer's risk.

To give an impression of the test effort as a result of applying the system of random testing and batch testing as described in this section, Figure 16 shows the percentage of batches that are actually tested as a function of the percentage of production exceeding the declared value. This figure illustrates that in the case of a conforming production (a production of which the 90th percentile is smaller than or equal to the declared value) the test effort remains low and that the test effort rapidly increases when the percentage > L_D of the production increases. Therefore the system of random testing and batch testing encourages a producer to improve or optimize the quality of the product with respect to RDS.

Also, the batch size is important for the assessment system to function. In general, a smaller batch size requires a larger test effort, but has a relatively small impact if a batch is rejected, while a larger batch size requires a smaller test effort, but the impact and costs of rejecting a batch may be substantial. Also, the producer's choice of the CE-marked declared value is important as it determines the percentage > L_D. This means that there should be a balance between the declared product quality and batch size. This CEN/TR recommends a minimum batch of 1/10th of a year's production.

In order to find a proper balance between test effort and the impact of rejecting a batch, it may be appropriate to apply a relatively small batch size during batch testing and allow a larger batch size during random testing.

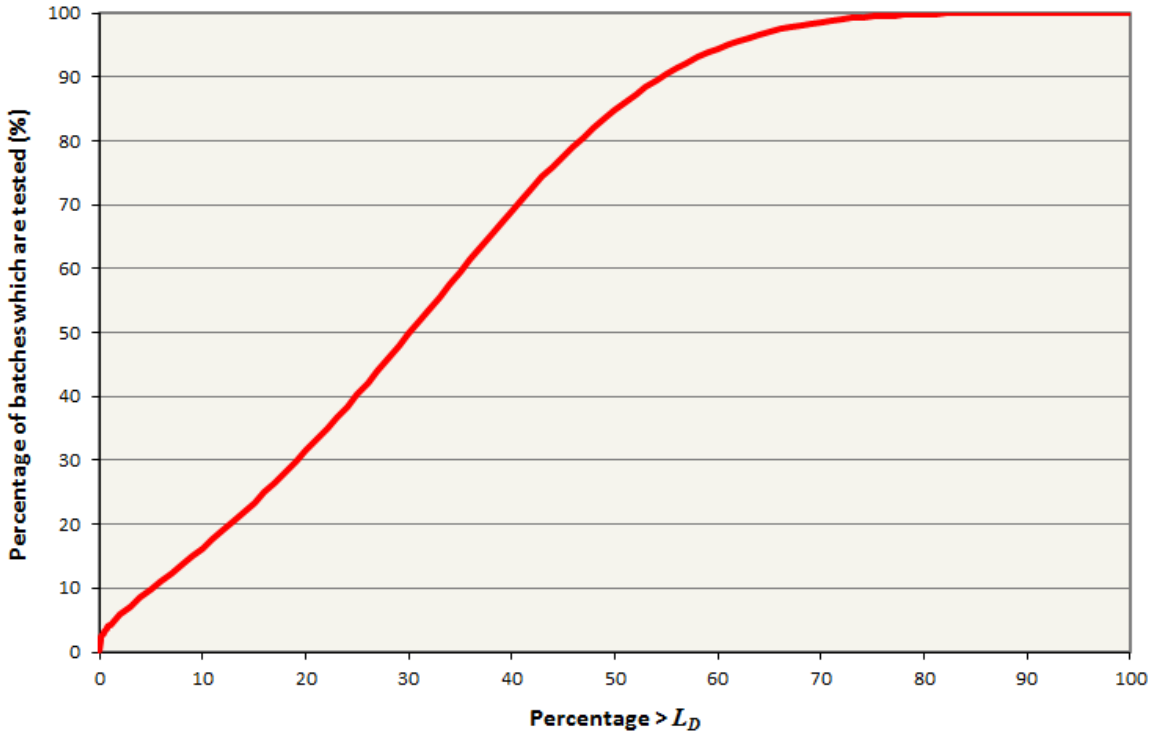


Figure 16 — Percentage of batches which are actually tested – assessment by variables (*n* = 5)

8.4.3 Gamma rule

Where the variability of test results is mainly the result of the error of the test method instead of the variability of the product quality, the test frequency will have little relation with the product quality. In order to prevent an unnecessary high test frequency, the gamma rule may be applied [9]. The principle of this rule is to check whether there are no test values that exceed the declared value reduced by a factor, the so-called gamma factor.

To establish with a probability of (1-*α*)·100 % that not more than *β*·100 % of the production exceeds the declared value, the gamma factor *Γ* equals:

$$\Gamma = \exp\left(\left(z_{1-\alpha^{1/n}} - z_{\beta}\right)\sigma\right) \tag{14}$$

where

α = uncertainty,

β = fraction of production that exceeds the declared value,

z_i = upper 100i percentage point of the standard normal distribution,

σ = $\sqrt{\ln\left(1+c_v^2\right)}$

where

c_v = coefficient of variation of the test results, which is a constant in case of a lognormal distribution,

n number of test results.

The gamma factor may only be estimated if the coefficient of variation is known. This coefficient is unknown for most construction products. In the case of leaching sufficient historic data are available and a value of 0,65 may be used, giving a value for *σ* of 0,59.

NOTE The value of 0,65 is taken from the Dutch Soil Quality Decree. This value is an average derived from leaching data of numerous construction materials.

Values for the gamma factor for different coefficients of variation are given in Annex F.

The effect of applying the gamma rule is that the average test frequency decreases, because the probability of testing the production once per year or one per 3 years becomes greater. It does not decrease the probability of batch testing.

8.4.4 No-further-testing

Examples of the OC-curves representing the probability of achieving NFT are shown in Figure 17 (see 6.3).

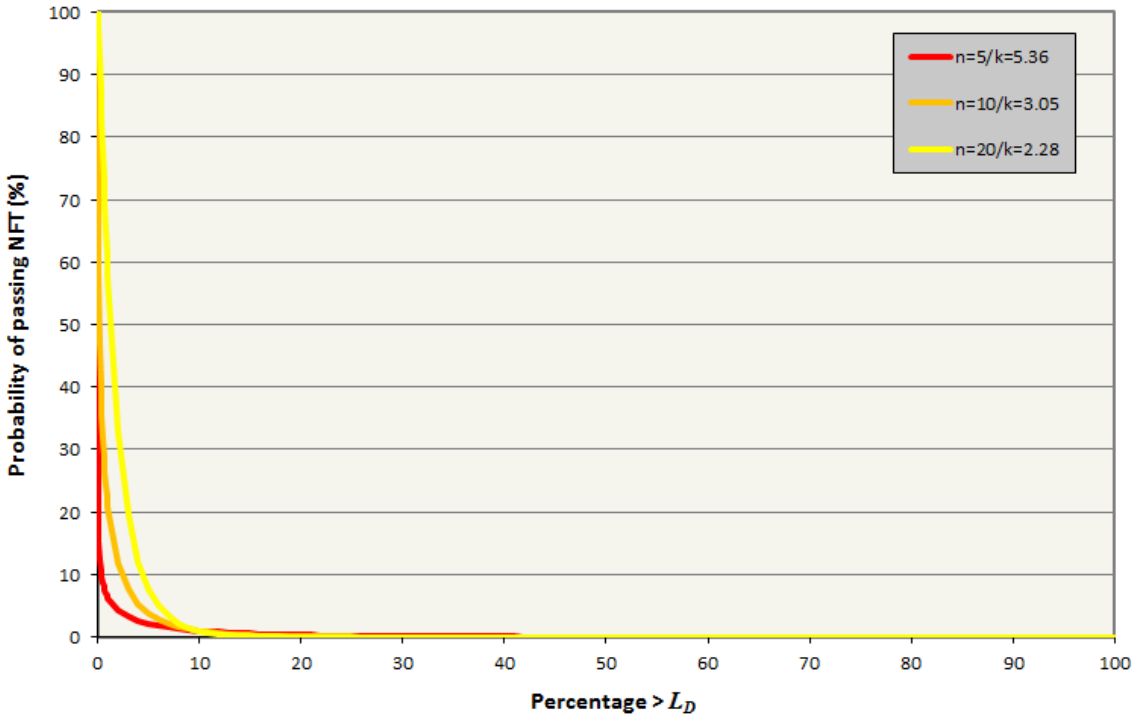


Figure 17 — Probability of passing NFT – assessment by variables (OC-curves)

8.5 Assessment by attributes

8.5.1 Type testing and further-testing

For assessment by attributes the same statistical assumptions apply as for assessment by variables, except for the distribution of the test values, which is assumed to be unknown. The assessment is based on the number of test values that exceed the declared value in the last n test values. Critical values for the number of test values that are permitted to exceed the declared value depend on n and the probability that a certain percentile of the production exceeds the declared value L_D and are based on the binominal distribution.

The probabilities of ending TT for the values of (n,n_a) in Table 11 are shown in Figure 18 and the probabilities of achieving random testing during FT are shown in Figure 19. According to Figure 18 the criterion for a part of the production of which the 90th percentile equals the declared value with a probability of at least 90 % is not met in the case of n < 9 and the producer's risk is greater than 10 % (64 % for n = 2).

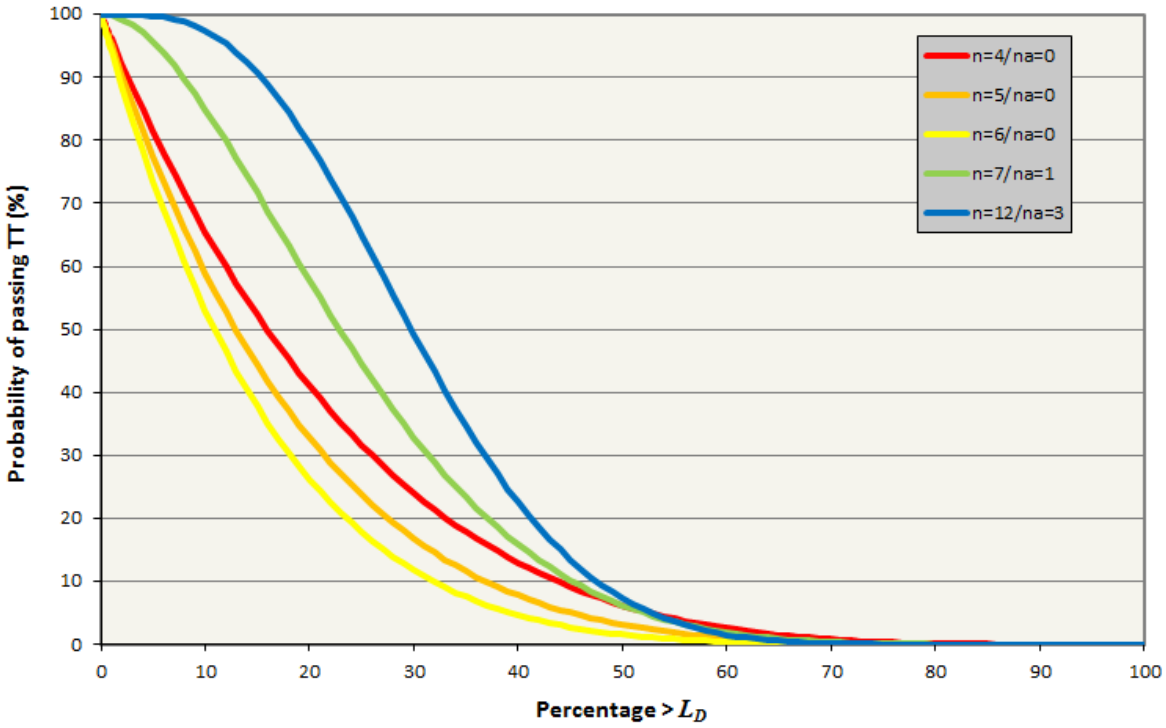


Figure 18 — Probability of ending TT – assessment by attributes (OC-curves)

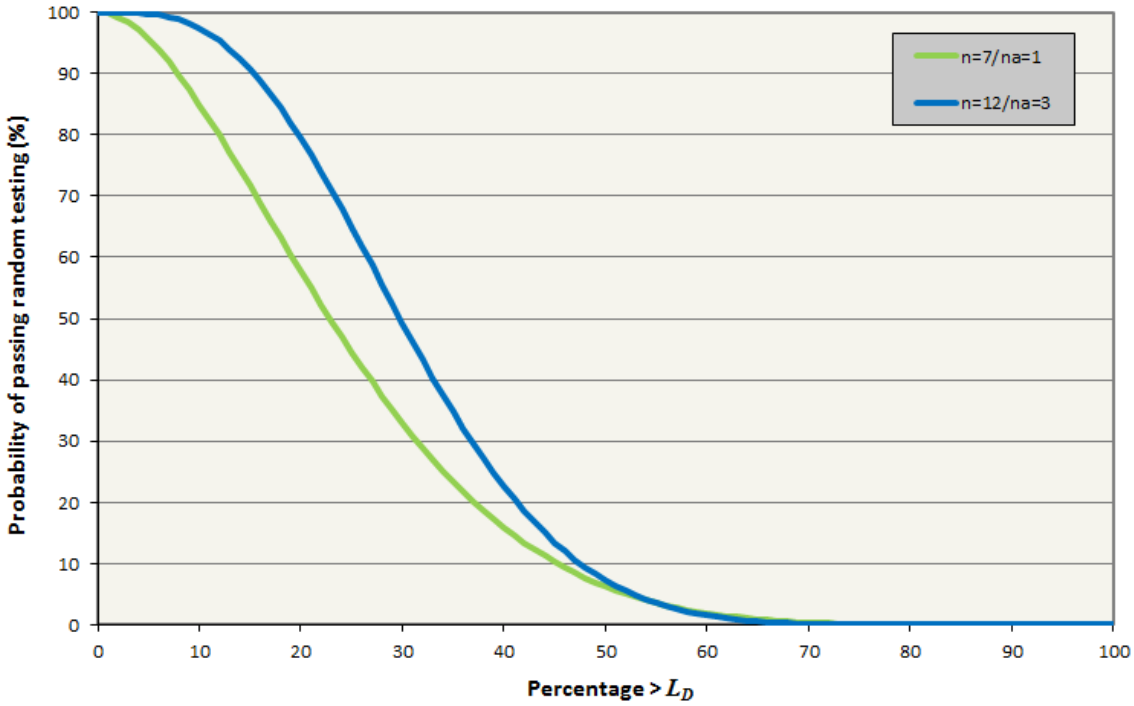


Figure 19 — Probability of achieving random testing for FT – assessment by attributes (OC-curves)

In case of assessment by attributes combinations of n and n_a for different percentiles and a probability of 90 % are given in Table 17. Two alternatives (I and II) are given depending on the number of test values. In practice it will be almost impossible to obtain a very low test frequency due the required large number of test values. The probabilities of achieving the frequency criteria are shown in Figure 20.

Table 17 — Values for n and na for different percentiles and a probability of 90 %

Percentile	n_a/n	
	I	II
50th	1/7	3/12
70th	1/12	3/21
90th	0/22	1/38
99th	0/230	1/388
99,9th	0/2302	1/3889

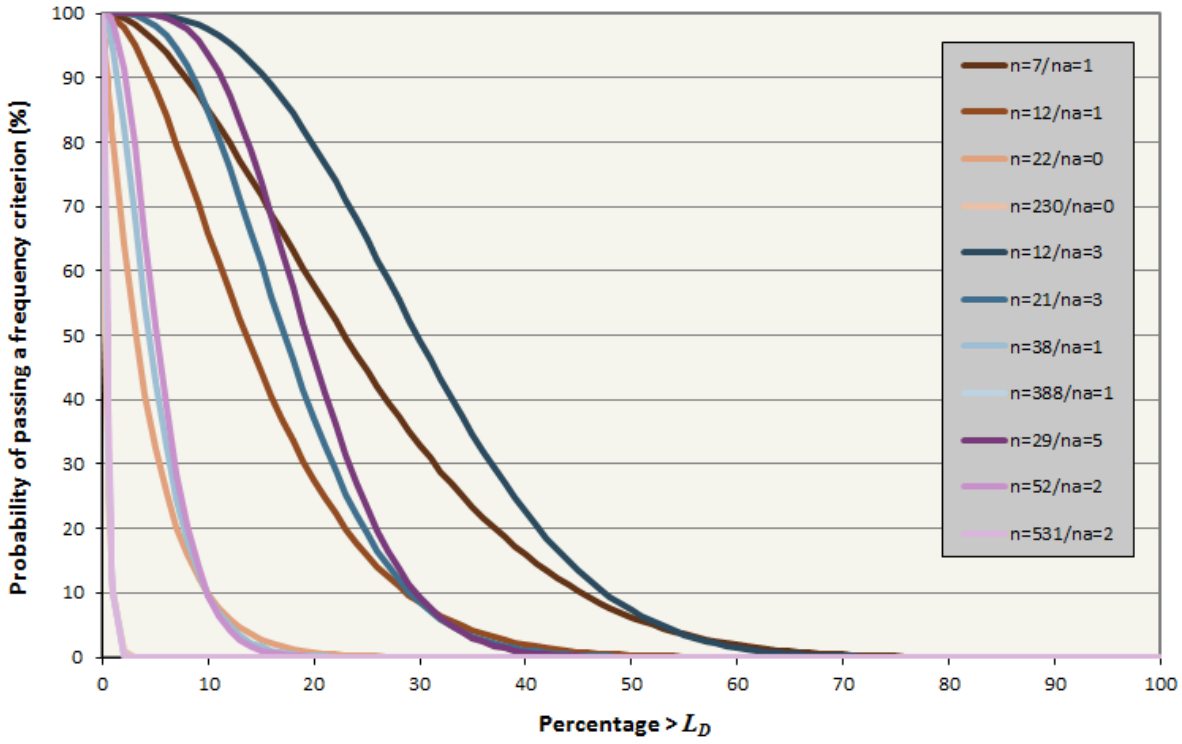


Figure 20 — Probability of passing frequency criteria – assessment by attributes (OC-curves)

8.5.2 No-further-testing

Examples of the OC-curves representing the probability of achieving NFT are shown in Figure 21 (see 7.3).

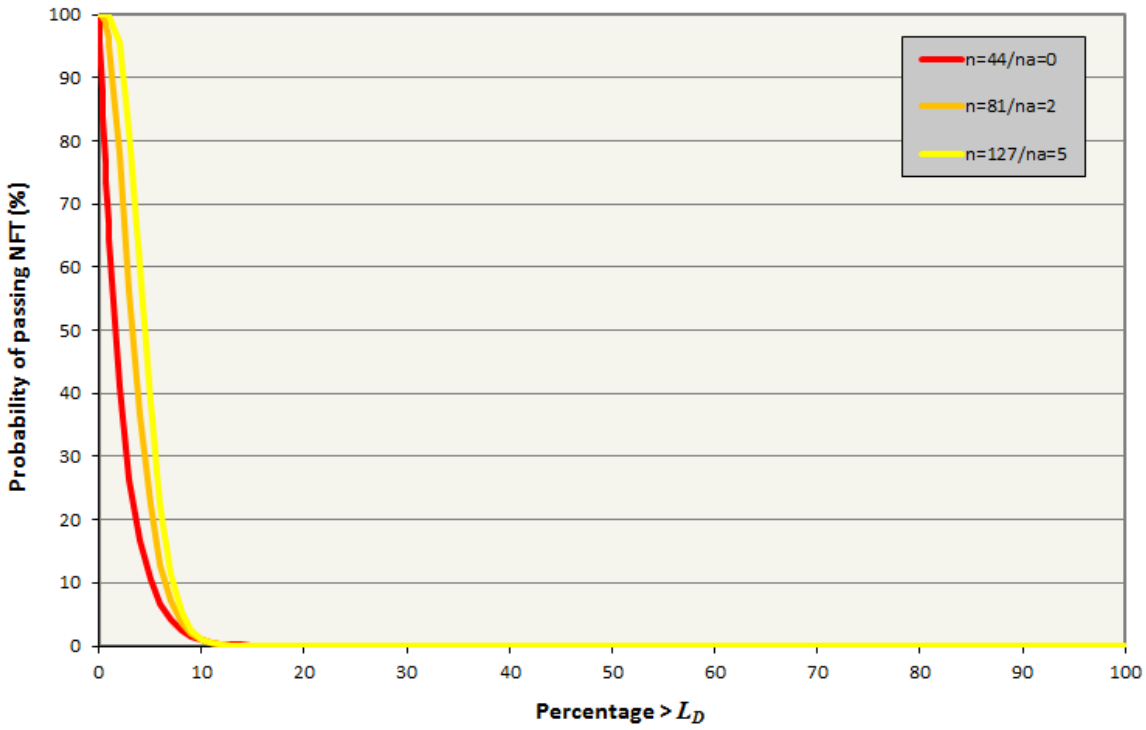


Figure 21 — Probability of achieving NFT – assessment by attributes

8.6 Consumer's and producer's risk

8.6.1 Acceptance and non-acceptance of batches that exceed the declared value

8.6.1.1 General

The probability of acceptance and non-acceptance of batches that exceed the declared value depends on the following parameters:

— OC-curve of the random testing/batch testing criterion (8.4.1 and 8.5.1);

— reduced testing due to a variable test frequency during random testing (6.1.2, 6.2.2, 7.1.2 and 7.2.2);

— criterion for returning from batch testing to random testing (6.1.2, 6.2.2, 7.1.2 and 7.2.2).

8.6.1.2 OC-curve

The OC-curve of the random testing/batch testing criterion determines the probability of random testing and therefore the probability of accepting batches that exceed the declared value:

$$P\{\text{accepting a batch} > L_D \mid \text{testing}\} = P\{\text{random testing}\} \times P\{\text{batch} > L_D\} \quad (15)$$

In case of random testing no batches are rejected (thus, batches which exceed the declared value are accepted); in case of batch testing every batch that exceeds the declared value is rejected (thus, only batches that do not exceed this value are accepted). Due to this distinction the producer's risk is reduced to zero (if measurement uncertainty is excluded as recommended in this CEN/TR), while the consumer's risk remains limited. The probabilities of accepting and rejecting batches with respect to the choice of the OC-curve are shown in Figure 22.

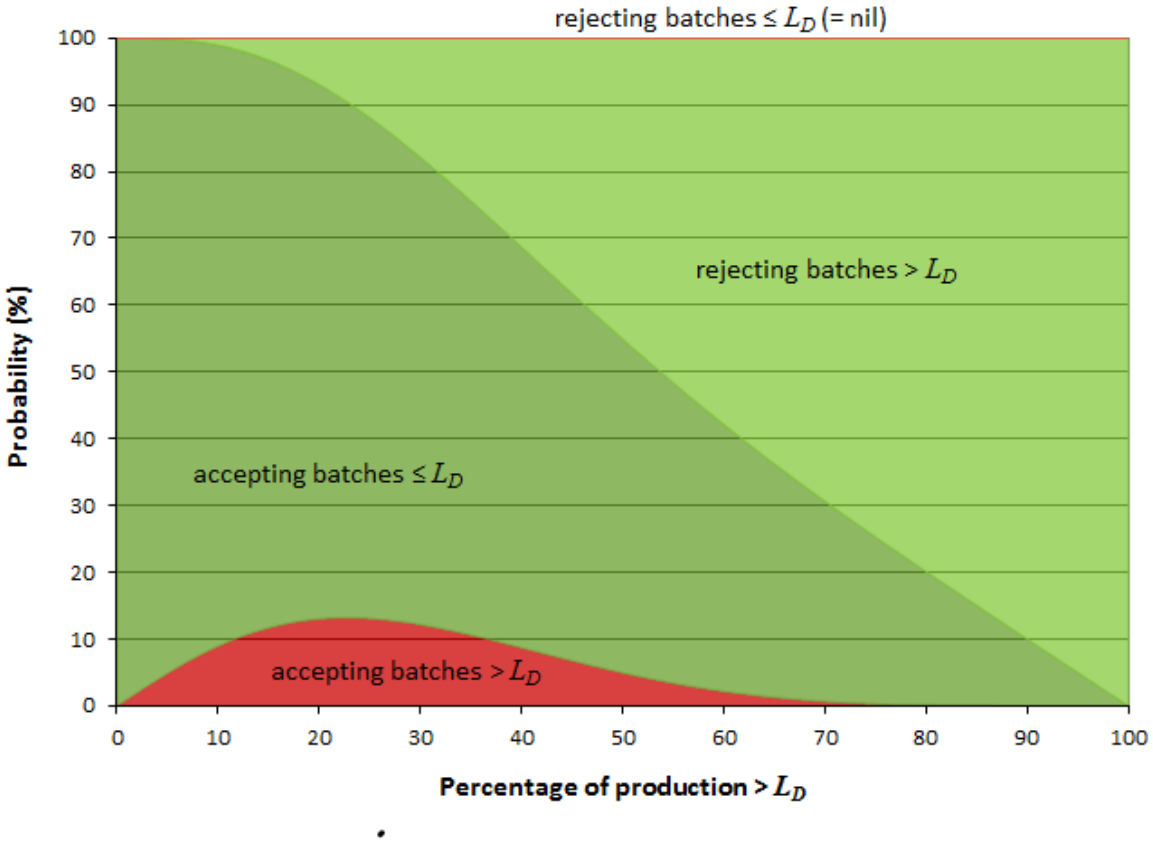


Figure 22 — Probability of accepting and rejecting batches – assessment by variables 5 batches per production part (*n* = 5,k = 0,69); P{accepting a batch > LD} + P{accepting a batch ≤ LD} + P{rejecting a batch > LD} + P{rejecting a batch ≤ LD} = 1

8.6.1.3 Reduced testing

The average test frequency determines the probability of testing a batch ($P\{\text{testing a batch}\}$). When a batch is tested the probability of accepting a batch that exceeds the declared value is given by Formula (15). When a batch is not tested (in the period between two successive batch tests) the probability of accepting a batch that exceeds the declared value equals:

$$P\{\text{accepting a batch} > L_D \mid \text{not testing}\} = P\{\text{batch} > L_D\} \tag{16}$$

The combined effect of the OC-curve and reduced testing (variable test frequency) on the probability of accepting a batch that exceeds the declared value equals:

$$P\{\text{accepting a batch} > L_D\} = P\{\text{testing a batch}\} \times P\{\text{random testing}\} \times P\{\text{batch} > L_D\} + \left(1 - P\{\text{testing a batch}\}\right) \times P\{\text{batch} > L_D\} \tag{17a}$$

or

$$P\{\text{accepting a batch} > L_D\} = P\{\text{batch} > L_D\} \times \left[1 - P\{\text{testing a batch}\} \times \left(1 - P\{\text{random testing}\}\right)\right] \tag{17b}$$

The probabilities of accepting and rejecting batches with respect to the choice of the OC-curve combined with reduced testing as described in 8.4.2 are shown in Figure 23. As may be expected, reduced testing increases the probability of accepting a batch that exceeds the declared value, thus theoretically increasing the consumer's risk.

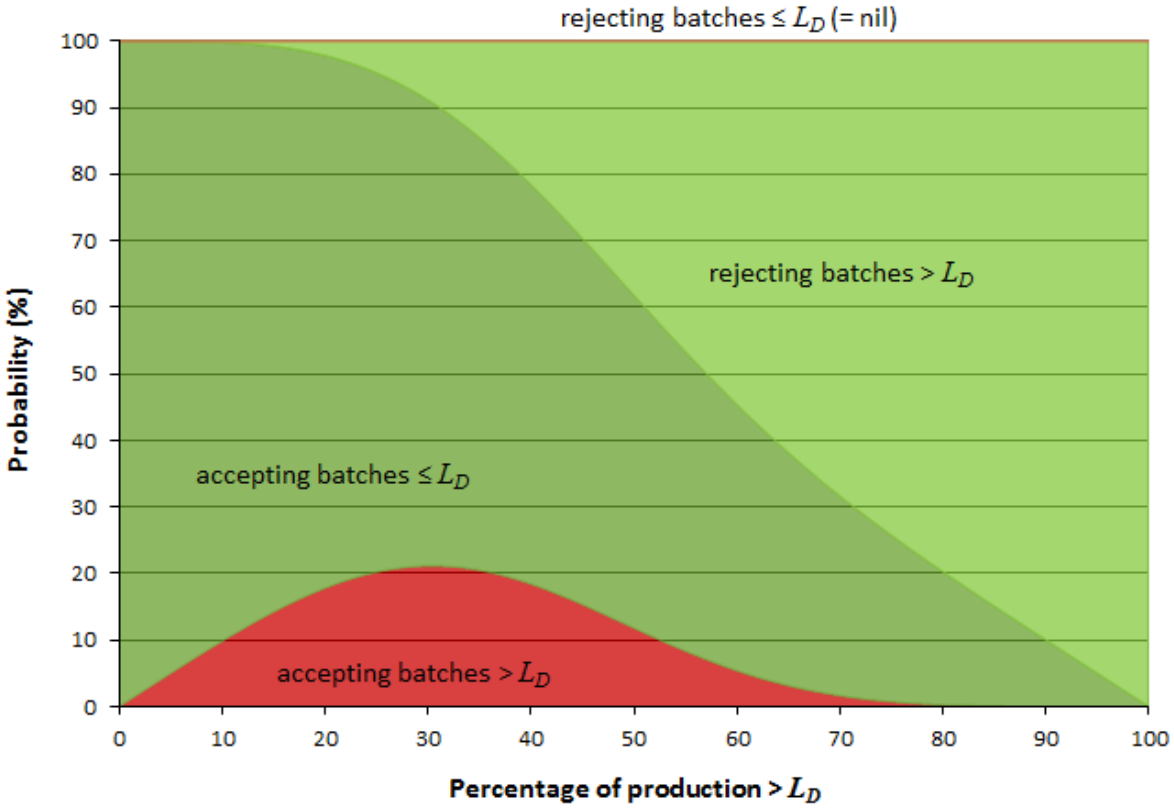


Figure 23 — Probability of accepting and rejecting batches with respect to the OC-curve combined with reduced testing – Assessment by variables 5 batches per production part

8.6.1.4 Criterion for returning from batch testing to random testing

Once batch testing is required at least 5 consecutive batches will have been tested before a return to random testing is permitted. The effect of this imposed delay is a reduction of the probability of accepting a batch that exceeds the declared value (due to an increase of $P\{\text{testing a batch}\}$); however, a theoretical approach of the effect is complicated and therefore the probabilities of accepting and rejecting batches have been determined by simulation.

Figure 24 shows these probabilities with respect to the OC-curve, reduced testing and imposed delay. The simulation shows that the imposed delay has a positive effect for products with (20 to 75) % > L_D. The maximum reduction of the probability of accepting a batch that exceeds the declared value is about 6 %. Outside this range this delay has no effect on the probability of accepting batches that exceed the declared value.

The effect of an imposed delay depends on the required number of batch tests. The consumer's risk decreases when the number of batches that need to be tested before a return to random testing is possible increases, simply due to the fact that the probability of testing a batch increases.

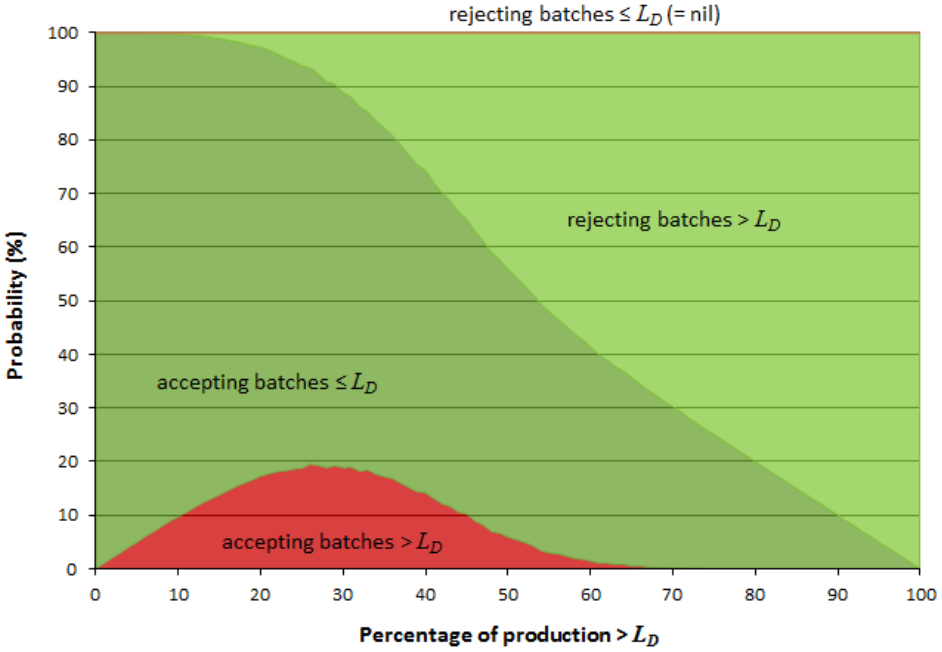


Figure 24 — Probability of accepting and rejecting batches with respect to the OC-curve, reduced testing and imposed delay – Assessment by variables 5 batches per production part

8.6.2 Estimation of the consumer's risk

The consumer's risk is related to the product as placed on the market and not to the entire production. Because only accepted batches are placed on the market, the consumer's risk is not equal to the probability of accepting a batch that exceeds the declared value. Instead, the consumer's risk equals:

$$\text{consumer's risk} = \frac{P\{\text{accepting a batch} > L_D\}}{P\{\text{accepting a batch} \leq L_D\} + P\{\text{accepting a batch} > L_D\}} \quad (18)$$

For the assessment system described in this CEN/TR the consumer's risk has been estimated using a simulation, because the effect of the imposed delay when returning from batch testing to random testing cannot be calculated easily from a theoretical approach.

The estimated consumer's risk is shown in Figure 25 (assessment by variables) and Figure 26 (assessment by attributes). Assessment by attributes gives a somewhat lower consumer's risk than assessment by variables. The consumer's risk is highest for productions that comprise 15 % to 50 % batches that exceed the declared value.

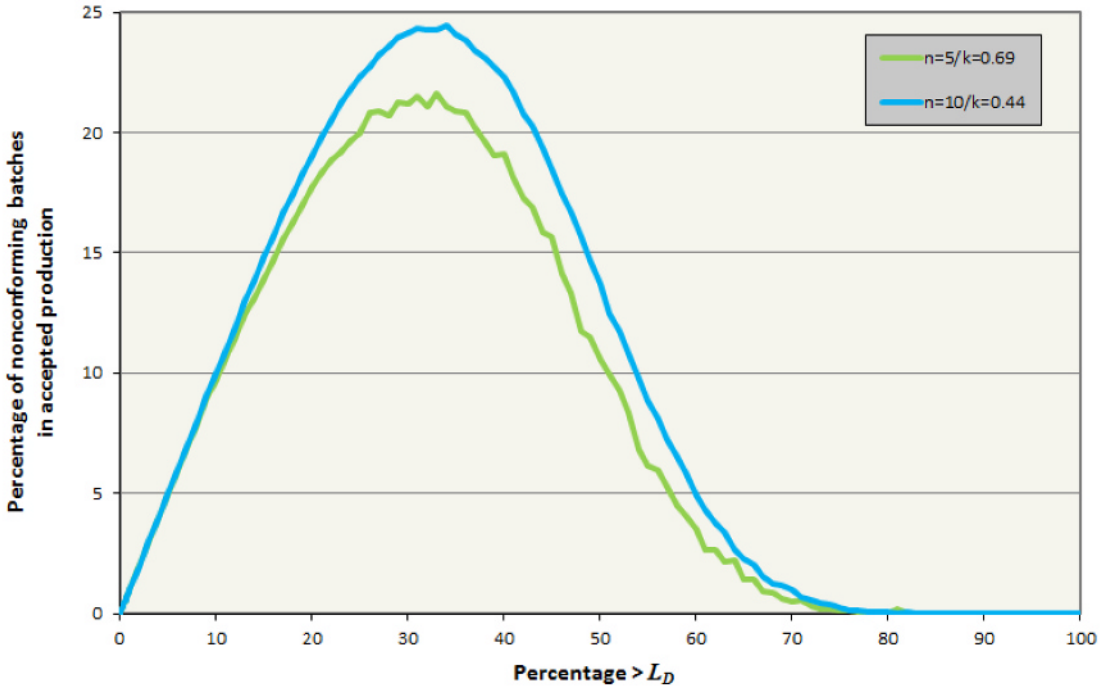


Figure 25 — Consumer's risk – assessment by variables

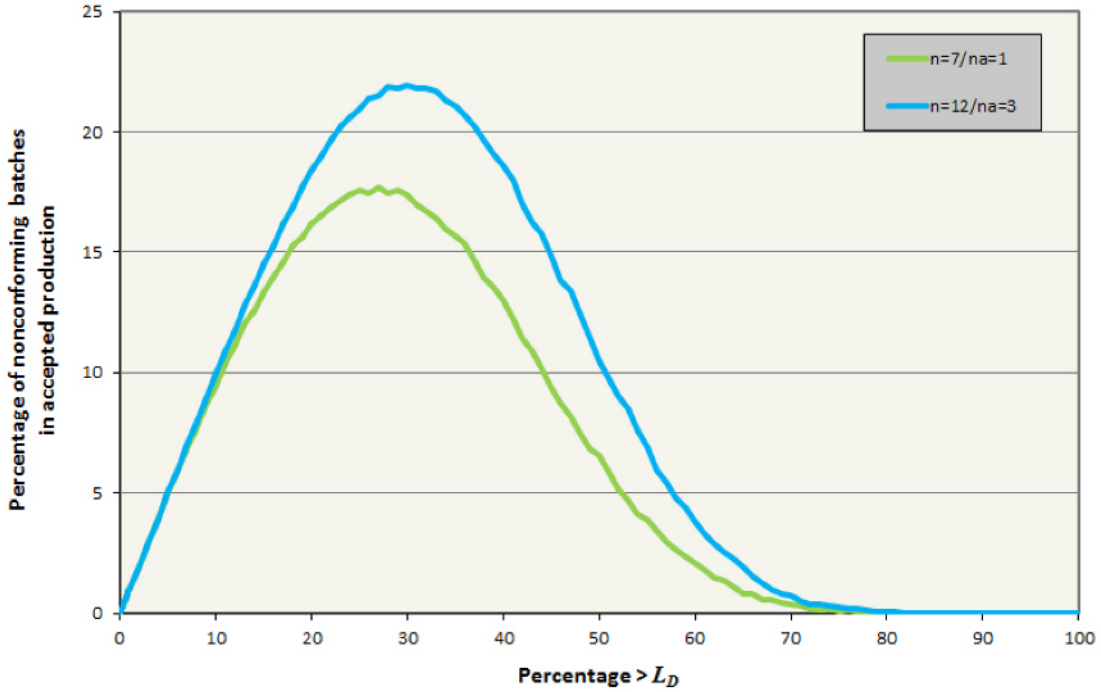


Figure 26 — Consumer's risk – assessment by attributes

8.6.3 Practical approach

In practice a producer will study the trend after each result and take action if possible to prevent batch testing if the average approaches an adverse change point, meaning that a test result will be probably influenced by an action based on a previous test result if the previous test result is critical. A producer

will always tend to reduce the testing effort, and therefore to improve the product and aim at the lowest percentage > L_D possible. Therefore production control creates a certain 'natural' dependency between consecutive batches.

In general this dependency causes the percentage > L_D to decrease in time, especially if FT tends to batch testing. Also, the batches which are produced in between two tests will have a relation with the ones that are tested due to the continuity of the production process and the aim to produce a constant product.

The estimated consumer's risk in Figure 25 and Figure 26, however, is based on a total random extraction of test values from a (log)normal distribution characterized by a chosen percentage > L_D and a constant μ and σ, assuming that there is no dependency between two consecutive test values and all intermediate values. This, of course, is not a realistic approach of how a producer deals with testing RDS (or any other property of a construction product) within FPC.

As the probability of batch testing becomes rather high for productions that comprise more than 20 % batches > L_D (see Figure 9) and the required testing effort (and costs) increases (see Figure 16), it seems likely that for construction products the percentage > L_D will be (much) less than 20 % and that this percentage decreases in time.

NOTE Taking this into consideration and the fact that the Dutch system, which is based on the criteria presented in this CEN/TR proves to work well in practice, the maximum risk of about 10 % which was the starting point of the assessment method (see 5.2) will probably not be exceeded in practice, even though the theoretical probabilities suggest otherwise.

9 Additional sampling requirements

9.1 General

Assessment of the release, emission or content of RDS takes place on the scale of batch sizes. This requires a test value that lies as close to the true average value of a batch as possible. As each test value will be the result of testing one or more samples taken from the batch, additional sample requirements for each construction product should be defined in order to obtain a representative test value. The technical report CEN/TR 16220 gives guidelines on these requirements.

This report only deals with requirements that affect the statistical representativeness of the test value:

— number of increments,

— number of samples.

The choice of these sampling parameters depends on the heterogeneity, type and properties of the construction material, the type of RDS, the error of sampling and testing and the costs of sampling and testing. A representative test value may be obtained from testing one sample that comprises a large number of increments as well as from testing several samples that comprise a smaller number of increments.

The latter option provides information on the variability within the batch (on the scale of the samples).

The variability of a test value, and consequently the uncertainty, is affected by:

— the spatial and/or temporal variability of the product within the batch (product heterogeneity),

— the error due to sampling and sample reduction,

— the error due to sample preparation and testing.

The variance of the test value is the sum of the variances of the different effects:

$$\sigma_v^2 = \sigma_w^2 + \sigma_s^2 + \sigma_t^2 \tag{19}$$

where

σ_v = standard deviation of the test value,

σ_w = standard deviation of the spatial and/or temporal variation,

σ_s = standard deviation due to the sampling error,

σ_t = standard deviation due to the testing error (see test standard).

The variability of the test value may be reduced by taking several increments and testing more than one sample in the following way:

$$\sigma_v^2 = \frac{\sigma_w^2}{n \times m} + \frac{\sigma_s^2 + \sigma_t^2}{n} \tag{20}$$

where

n = number of samples,

m = number of increments per sample.

The observed (spatial or temporal) variability depends on the size of the batch, where as a general rule the between batch variability will decrease when the batch size increases. The observed within batch variability (when more than one sample from a batch is analysed) will depend on the size of the increments and the number of increments combined in a sample. For most construction products only the standard deviation of the test value and the reproducibility of the test method are known. The standard deviation may be estimated from the test results (preferably from batch tests based on more than one sample per batch), while the latter is (or should be) given in the test standard. The standard deviation due to spatial variation, sampling, sample reduction and sample preparation are probably unknown. In order to estimate the standard deviation of the spatial variation, which is needed to estimate the number of increments and samples, a value for the standard deviation due to sampling, sample reduction and sample preparation needs to be chosen.

9.2 Probabilistic sampling

In practice it is almost impossible to get reliable information on sampling errors without many tests with at least two samples per batch. It will not be feasible to determine each error separately, but only to determine combined errors, e.g. the combined relative error due to product heterogeneity (spatial and/or temporal variability), sampling and testing. If the coefficient of variation (ratio of the standard deviation and mean) of this combined error is known, it is possible to estimate the number of increments and number of samples with the following formula:

$$n \times m \geq \left(\frac{z_{1-\frac{\alpha}{2}} \sqrt{1 + c_w^2}}{\ln\left(1 + PF\right)} \right)^2 \tag{21}$$

with

$$c_w^2 = c_v^2 - c_s^2 - c_t^2 \tag{22}$$

where

n = number of samples,

m = number of increments per sample,

$z_{\frac{1}{2}\alpha}$ = upper $\frac{1}{2}a$ percentage point of the standard normal distribution,

PF = precision factor, which is the ratio between the maximum deviation and median value,

c_w = coefficient of variation of one increment due to product heterogeneity,

c_v = coefficient of variation of the test value of a RDS,

c_s = coefficient of variation due to sampling, sample reduction and sample preparation,

c_t = coefficient of variation due to testing.

Based on Dutch experiences the value of the coefficient of variation due to sampling, sample reduction and sample preparation c_s may be assumed to equal 0,20. Another possibility is to assume that $c_s^2 + c_t^2$ is negligible, so that $c_w \approx c_v$. Although this assumption may be questionable, it is on the safe side, because it gives an overestimation of the required number of samples and increments.

The assumption of a lognormal distribution of the test values includes the assumption of a constant coefficient of variation.

Instead of using Formula (21) the number of samples and increments may be read from Figure 27 where the number of samples × increments is shown as a function of the coefficient of variation and the precision factor. In general a precision factor of 0,10 to 0,20 will give satisfying results. The precision factor is a chosen value based on the maximum variation that is assumed to be acceptable in relation to the goal of the sampling.

As the sampling parameters and coefficient of variation of the increments will probably have different values for the various RDS, the actual choice of the sampling parameters becomes rather complex. Therefore a more pragmatic approach is given as an alternative. Based on the Dutch experience the combinations in Table 18 may be used, assuming that the producer's aim is to provide a constant product quality. Only in specific situations or specific RDS the method based on Formula (21) should be used.

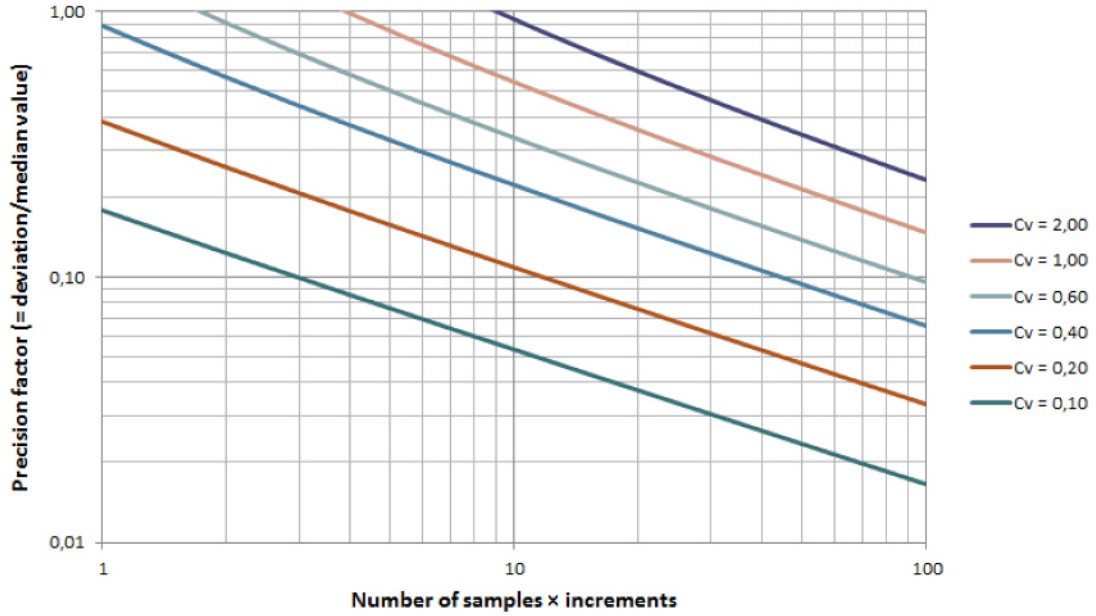


Figure 27 — Number of samples and increments at 90 % confidence (α = 10 %, $z_{1/2\alpha}$ = 1,645)

Table 18 — Number of increments and samples

Number of samples	Number of increments per sample
1	12 (monolithic construction products) 32 (granular construction products)
2	6
3	4

9.3 Judgemental sampling

Judgemental sampling requires detailed knowledge of the aspects of the release, emission or content of RDS from a construction product. In this case the number of increments and the number of samples will be circumstantial and not subjected to a statistical approach although the statistical principles of probabilistic sampling can be applied in the same way.

10 Indirect tests

10.1 General

Indirect (non-reference) tests are not permitted for TT. In certain situations, such as FT, indirect tests may be used to estimate the release, emission or content of RDS. The use of quick, indirect tests for FT purposes should only be permitted if the frequency of testing for a RDS is more than once per year and it can be proven that the indirect test gives a reliable result with respect to the result by the reference test.

An indirect test is reliable enough when:

— the use of the indirect test leads to the same conclusion with respect to accepting and rejecting batches as the use of the reference test method, or

— the indirect test produces a test value that is always equal to or greater than the test value produced with the reference test.

This requires that the property measured with the indirect test is in some way related to the release measured with one of the standard test methods.

For the continual validation of an indirect test, the producer should at least once a year test a sample simultaneously with the indirect test as well as with the reference test and verify the regression model.

In the case of a dispute, the reference test data has precedence over indirect test results. Enforcement testing should use the reference test method.

10.2 Correlation

In the case of a developed and documented correlation between the indirect test and reference test and it is justifiable to apply a linear regression model, it is permitted to express the value measured with the indirect test according to this regression model:

$$x' = ax + b \tag{23}$$

where

x' = equivalent value for the reference test,

x = indirect test value,

a = slope of the estimated regression line $y = ax + b$

where

x = indirect test value and

y = reference test value,

b = intercept of the estimated regression line $y = ax + b$

Correlation is established on the basis of at least 15 samples and every test value needs to be equal to at least 3 times the detection limit.

A justifiable correlation is satisfied when a linear correlation gives a (weighted) coefficient of determination $R^2 \geq 0,7$ and a slope ≥ 0,2.

From a statistical point of view, a linear relationship gives the most satisfying result. In the case of lognormal distributed test values, the use of ln-transformation of the test values may be required.

As the indirect test as well as the reference test have a testing error, it is not realistic to entirely allocate the difference between the test values to the indirect test. Therefore the use of a confidence interval of the linear regression is not preferred. It makes more sense to use the average regression formula (Formula (23)) instead.

If the slope is smaller than 0,2, the indirect test becomes too insensitive with respect to the reference test and the uncertainty may become too high.

If the indirect test is based on the same release mechanism as the reference test, it may be expected that any regression model between the indirect test and reference test will pass through the origin. If

the estimated intercept of the regression line (*b*) significantly deviates from zero on a linear scale, it may be possible that an external effect, not related to release, affected the measurements.

It is not permitted to extrapolate the linear regression model. Extrapolation is assumed to occur when the x value deviates more than 20 % from the x range of the regression model.

10.3 No correlation

If the correlation between the indirect test and reference test is not sufficient, the use of a linear regression model becomes unreliable and it is not allowed to correct test values. In this case only indirect tests are allowed that produce a real test value (not a detection limit) that is always equal to or greater than the test value produced with the reference test. Using an appropriate statistical test, e.g. Wilcoxon signed-rank test for paired data, it needs to be shown with a confidence of 90 % that the indirect test produces a test value equal to or greater than the reference test.

An indirect test does not necessarily need to be a release test, e.g. leaching test. Also indirect properties, like dry density or compressive strength, may be used to find a relation with the reference release test; however, to establish confidence in the indirect test some physical or chemical relationship to the RDS is essential.

Annex A

Examples of the rules of application

A.1 EXAMPLE 1: Assessment by variables for a single production unit

This example shows the evolution of the assessment based on assessment by variables during TT and FT. It includes the use of the criteria in 6.1.1 (TT) and 6.1.2 (FT). The example shows that TT may be ended when there are 2 test values.

TT

After 2 tests the value of k_2 equals 5,42 while the critical value for 2 test values equals 2,18, see Table 1. This means that TT may end and random testing commences at a frequency of 1 batch in 4, see Table 2.

FT

For the following production (Samples 3 to 11) batches are tested randomly (the example only records the batches that were tested and not the three batches that were not tested); however, the test value of Sample 11 exceeds the declared value. Because Sample 11 is taken from a randomly tested production, the batch from which the sample has been taken is not rejected. From a practical point of view, however, the producer might want to exclude this batch from being placed on the market and take the necessary measures if practical.

The test result of Sample 11 causes the value of k_5 to drop below the critical value (0,69). As soon as this test result is known, FT changes to batch testing. As long as all batches are being tested the declared value acts as a hard limit, meaning that batches that exceed this value are rejected (Samples 14, 15, 21 and 22).

A return to random testing is only permitted if all the following are satisfied:

1) the value of $k_5 \geq 0{,}69$;

2) at least 5 consecutive batches have been tested during the batch testing period;

3) the value of $k_{10} \geq 0{,}44$.

All three conditions are met for the first time for Sample 26. As soon as the test result of Sample 26 is known, FT switches to random testing at a frequency of 1 batch in 2, see Table 2. After this point no further disruption of the test results occurs and FT continues to be based on random testing.

In total one batch which exceeds the declared value has been accepted out of 5 within a total of 37 tested batches.

Based on the test frequency in total 84 batches have been produced, meaning that 47 batches have not been tested.

Production

μ	110
σ	80
Percentage > L_D	10 %
L_D	210
ln(L_D)	5,35

				n test values						Frequency of testing	10 test values					
	Sample	x_i	ln(x_i)	***n***	**Mean**	**St.dev.**	k_n	k_0	**Conclusion**			**Mean**	**St.dev.**	k_{10}	k_0	**No. of batches batch testing**	**Switch**
TT	1	77	4,34							batch accepted							
	2	57	4,04	2	4,19	0,21	5,42	2,18	TT may be ended	batch accepted	1 in 4						
	Start								random testing								
FT	3	11	2,40	3	3,59	1,05	1,67	1,09	random testing	batch accepted	1 in 2					0	
	4	83	4,42	4	3,80	0,95	1,63	0,82	random testing	batch accepted	1 in 2					0	
	5	66	4,19	5	3,88	0,84	1,75	0,69	random testing	batch accepted	1 in 4					0	
	6	120	4,79	5	3,97	0,92	1,50	0,69	random testing	batch accepted	1 in 4					0	
	7	170	5,14	5	4,19	1,06	1,09	0,69	random testing	batch accepted	1 in 2					0	
	8	99	4,60	5	4,63	0,36	2,00	0,69	random testing	batch accepted	1 in 4					0	
	9	110	4,70	5	4,68	0,34	1,95	0,69	random testing	batch accepted	1 in 4					0	
	10	120	4,79	5	4,80	0,20	2,69	0,69	random testing	batch accepted	1 in 4	4,34	0,75	1,34	0,44	0	
	11	380	5,94	5	5,03	0,55	0,58	0,69	batch testing	batch accepted	1 in 1	4,50	0,91	0,93	0,44	0	no
	12	130	4,87	5	4,98	0,55	0,67	0,69	batch testing	batch accepted	1 in 1	4,58	0,90	0,85	0,44	1	no
	13	69	4,23	5	4,91	0,63	0,70	0,69	batch testing	batch accepted	1 in 1	4,77	0,51	1,15	0,44	2	no
	14	290	5,67	5	5,10	0,70	0,36	0,69	batch testing	batch rejected	1 in 1	4,89	0,56	0,81	0,44	3	no

			n test values								10 test values					
Sample	x_i	ln(x_i)	***n***	**Mean**	**St.dev.**	k_n	k_0	**Conclusion**		**Frequency of testing**	**Mean**	**St.dev.**	k_{10}	k_0	**No. of batches batch testing**	**Switch**
15	500	6,21	5	5,39	0,82	−0,05	0,69	batch testing	batch rejected	1 in 1	5,09	0,64	0,40	0,44	4	no
16	110	4,70	5	5,14	0,79	0,26	0,69	batch testing	batch accepted	1 in 1	5,08	0,65	0,41	0,44	5	no
17	210	5,35	5	5,23	0,78	0,15	0,69	batch testing	batch accepted	1 in 1	5,11	0,65	0,37	0,44	6	no
18	130	4,87	5	5,36	0,61	−0,02	0,69	batch testing	batch accepted	1 in 1	5,13	0,63	0,34	0,44	7	no
19	68	4,22	5	5,07	0,76	0,37	0,69	batch testing	batch accepted	1 in 1	5,08	0,68	0,38	0,44	8	no
20	75	4,32	5	4,69	0,45	1,45	0,69	batch testing	batch accepted	1 in 1	5,04	0,72	0,43	0,44	9	no
21	340	5,83	5	4,92	0,68	0,63	0,69	batch testing	batch rejected	1 in 1	5,03	0,71	0,45	0,44	10	no
22	230	5,44	5	4,93	0,70	0,59	0,69	batch testing	batch rejected	1 in 1	5,08	0,72	0,37	0,44	11	no
23	150	5,01	5	4,96	0,70	0,55	0,69	batch testing	batch accepted	1 in 1	5,16	0,65	0,28	0,44	12	no
24	85	4,44	5	5,01	0,64	0,53	0,69	batch testing	batch accepted	1 in 1	5,04	0,66	0,47	0,44	13	no
25	22	3,09	5	4,76	1,07	0,55	0,69	batch testing	batch accepted	1 in 1	4,73	0,77	0,80	0,44	14	no
26	44	3,78	5	4,35	0,94	1,06	0,69	random testing	batch accepted	1 in 2	4,63	0,83	0,86	0,44	15	yes
27	95	4,55	5	4,18	0,75	1,56	0,69	random testing	batch accepted	1 in 4	4,56	0,79	1,00	0,44	0	
28	130	4,87	5	4,15	0,71	1,69	0,69	random testing	batch accepted	1 in 4	4,56	0,79	1,00	0,44	0	
29	210	5,35	5	4,33	0,90	1,14	0,69	random testing	batch accepted	1 in 2	4,67	0,82	0,83	0,44	0	
30	55	4,01	5	4,51	0,63	1,32	0,69	random testing	batch accepted	1 in 2	4,64	0,84	0,85	0,44	0	
31	99	4,60	5	4,67	0,49	1,38	0,69	random testing	batch accepted	1 in 2	4,51	0,73	1,15	0,44	0	
32	37	3,61	5	4,49	0,69	1,25	0,69	random testing	batch accepted	1 in 2	4,33	0,70	1,46	0,44	0	
33	170	5,14	5	4,54	0,73	1,10	0,69	random testing	batch accepted	1 in 2	4,34	0,71	1,41	0,44	0	
34	110	4,70	5	4,41	0,60	1,56	0,69	random testing	batch accepted	1 in 4	4,37	0,72	1,36	0,44	0	
35	39	3,66	5	4,34	0,67	1,49	0,69	random testing	batch accepted	1 in 4	4,43	0,62	1,48	0,44	0	
36	190	5,25	5	4,47	0,79	1,11	0,69	random testing	batch accepted	1 in 2	4,57	0,63	1,23	0,44	0	
37	89	4,49	5	4,65	0,63	1,11	0,69	random testing	batch accepted	1 in 2	4,57	0,63	1,24	0,44	0	

A.2 EXAMPLE 2: Assessment by attributes for a single production unit

This example shows the evolution of the assessment based on assessment by attributes during TT and FT and it is based on the same dataset as example 1. It includes the use of the criteria in 7.1.1 (TT) and 7.1.2 (FT). This example shows that TT may be ended when there are 4 test values.

TT

As the first four test values are lower than the declared value, TT is ended and random testing may be started at a frequency of 1 batch in 2, see Table 12.

FT

For the following production (Samples 3 to 14) batches are tested randomly.

The test value of Sample 11 exceeds the declared value. Because Sample 11 is taken from a randomly tested production, the batch from which the sample has been taken is not rejected; however, from a practical point of view the producer might want to exclude this batch from being placed on the market and take the necessary measures if practical.

The test result of Sample 14 causes the value of n_e to exceed the acceptable number of test results that exceed the declared value (1). As soon as this test result is known, FT changes to batch testing. As long as all batches are being tested the declared value acts as a hard limit, meaning that batches that exceed this value are rejected (Samples 15, 21 and 22).

A return to random testing is only permitted if all the following are satisfied:

1) the value of n_e for the consecutive last 7 test results ≤ 1;

2) at least 5 consecutive batches has been tested during the batch testing period;

3) the value of n_e for the consecutive last 12 test results ≤ 3.

All three conditions are met for the first time for Sample 28. As soon as the test result of Sample 28 is known, FT switches back to random testing at a frequency of 1 batch in 2, see Table 12. After this point no further disruption of the test results occur and FT continues to be based on random testing.

In total 2 batches which exceed the declared value have been accepted out of 5 within a total of 37 batches.

Based on the test frequency in total 76 batches have been produced, meaning that 39 batches have not been tested.

Production

μ	110
σ	80
Percentage > L_D	10 %
L_D	210

			n test values					Frequency of testing	12 test values			
	Sample	x_i	***n***	n_e	n_a	**Conclusion**			n_e	n_a	**No. of batches batch testing**	**Switch**
TT	1	77					batch accepted					
	2	57	2	0	0		batch accepted					
	3	11	3	0	0		batch accepted				0	
	4	83	4	0	0	TT may be ended	batch accepted	1 in 2			0	
	Start					random testing						
FT	5	66	4	0	0	random testing	batch accepted	1 in 2			0	
	6	120	4	0	0	random testing	batch accepted	1 in 2			0	
	7	170	7	0	1	random testing	batch accepted	1 in 2			0	
	8	99	7	0	1	random testing	batch accepted	1 in 2			0	
	9	110	7	0	1	random testing	batch accepted	1 in 2			0	
	10	120	7	0	1	random testing	batch accepted	1 in 2			0	
	11	380	7	1	1	random testing	batch accepted	1 in 2			0	
	12	130	7	1	1	random testing	batch accepted	1 in 4	1	3	0	
	13	69	7	1	1	random testing	batch accepted	1 in 4	1	3	0	
	14	290	7	2	1	batch testing	batch accepted	1 in 1	2	3	0	no
	15	500	7	3	1	batch testing	batch rejected	1 in 1	3	3	1	no

		n test values					Frequency of testing	12 test values			
Sample	x_i	***n***	n_e	n_a	**Conclusion**			n_e	n_a	**No. of batches batch testing**	**Switch**
16	110	7	3	1	batch testing	batch accepted	1 in 1	3	3	2	no
17	210	7	3	1	batch testing	batch accepted	1 in 1	3	3	3	no
18	130	7	2	1	batch testing	batch accepted	1 in 1	3	3	4	no
19	68	7	2	1	batch testing	batch accepted	1 in 1	3	3	5	no
20	75	7	2	1	batch testing	batch accepted	1 in 1	3	3	6	no
21	340	7	2	1	batch testing	batch rejected	1 in 1	4	3	7	no
22	230	7	2	1	batch testing	batch rejected	1 in 1	5	3	8	no
23	150	7	2	1	batch testing	batch accepted	1 in 1	4	3	9	no
24	85	7	2	1	batch testing	batch accepted	1 in 1	4	3	10	no
25	22	7	2	1	batch testing	batch accepted	1 in 1	4	3	11	no
26	44	7	2	1	batch testing	batch accepted	1 in 1	3	3	12	no
27	95	7	2	1	batch testing	batch accepted	1 in 1	2	3	13	no
28	130	7	1	1	random testing	batch accepted	1 in 2	2	3	14	yes
29	210	7	0	1	random testing	batch accepted	1 in 2	2	3	0	
30	55	7	0	1	random testing	batch accepted	1 in 2	2	3	0	
31	99	7	0	1	random testing	batch accepted	1 in 2	2	3	0	
32	37	7	0	1	random testing	batch accepted	1 in 2	2	3	0	
33	170	7	0	1	random testing	batch accepted	1 in 4	1	3	0	
34	110	7	0	1	random testing	batch accepted	1 in 4	0	3	0	
35	39	7	0	1	random testing	batch accepted	1 in 4	0	3	0	
36	190	7	0	1	random testing	batch accepted	1 in 4	0	3	0	
37	89	7	0	1	random testing	batch accepted	1 in 4	0	3	0	
38	25	7	0	1	random testing	batch accepted	1 in 4	0	3	0	

A.3 EXAMPLE 3: Assessment by variables for a cluster

This example shows the evolution of the assessment of a cluster product based on assessment by variables during TT and FT. It includes the use of the criteria in 6.2.1 (TT) and 6.2.2 (FT). The cluster comprises 6 production units and TT is based on 6 samples and for the start of TT one sample is taken from each production unit. The test data show that TT may be ended when there are 6 test values and that the cluster system may be applied to the intended cluster.

TT

The results of 6 tests give a value of k_6 of 1,59 while the critical value k_0 for 6 test values equals 1,32, see Table 6. This means that TT may end and random testing commences.

FT

The start of FT is based on 5 test values. For the calculation of k_5 the lowest of the 6 test values is ignored to stay on the safe side. Because k_5 (1,55) ≥ k_0 (1,46) FT starts off with random testing of the cluster production at a test frequency of 1 batch in 4, see Table 7. For the following production (Samples 3 to 13) batches are tested randomly. In time more test values become available allowing the cluster organization to do the assessment on the consecutive 10 last test results and following on the consecutive 20 last test results.

Test value of Sample 13 exceeds the declared value. Because Sample 13 is taken from a randomly tested production, the batch from which the sample has been taken is not rejected. From a practical point of view the producer (production unit 1) might want to exclude this batch from being placed on the market and take the necessary measures if practical.

The test result of Sample 13 causes the value of k_{10} to drop below the critical value (1,07). As soon as this test result is known, FT changes to testing the production of each production unit separately instead of testing the cluster production. From that moment on each production unit is being assessed as described in example 1. As shown in the example production unit 1 needs to switch to batch testing (k_3 < 1,09), while the other production units should switch to random testing.

NOTE 1 Instead it may be decided to ban production unit 1, which caused the disruption, from the cluster.

NOTE 2 Because of the complexity a more detailed description of the testing of each production unit is omitted in this example. This example only describes the functioning of the cluster system.

A return to cluster testing is only permitted if:

1) all production units have produced at least one test value; and

2) the value of k_{10} ≥ 1,07 or k_{20} ≥ 0,87.

After 6 test results (Samples 14 to 19) the condition of return is not met, so a second round of 6 test values needs to be produced. After this second round (Samples 20 to 25) both conditions are met. As soon as the test result of Sample 25 is known, FT switches to random testing of the cluster production at a test frequency of 1 batch in 4, see Table 7. After this point no further disruption of the test results occur and FT continues to be based on random testing of the cluster production.

In total one batch which exceeds the declared value has been accepted out of one within a total of 33 batches.

	Individual productions							Total production
μ	110	60	100	100	120	100		100
σ	70	50	80	60	20	30		68
							Percentage > L_D	10 %
L_D	180							
ln(L_D)	5,19							

		Production unit										*n* test values						
		1		2		3		4		5		6							
	Sample	$x_{1.i}$	ln($x_{1.i}$)	$x_{2.i}$	ln($x_{2.i}$)	$x_{3.i}$	ln($x_{3.i}$)	$x_{4.i}$	ln($x_{4.i}$)	$x_{5.i}$	ln($x_{5.i}$)	$x_{6.i}$	ln($x_{6.i}$)	***n***	**Mean**	**St.dev.**	k_n	k_0	**Conclusion**	**Switch**
TT	1	110	4,70																	
	2			75	4,32															
	3					100	4,61													
	4							40	3,69											
	5									120	4,79									
	6											130	4,87	6	4,49	0,44	1,59	1,32	cluster testing approved	
	Start													5	4,45	0,48	1,55	1,46	random testing cluster	
FT	7	53	3,97											5	4,38	0,52	1,54	1,46	random testing cluster	
	8	1,66	B	33	3,50									5	4,16	0,63	1,63	1,46	random testing cluster	
	9			2,22	R	54	3,99							5	4,22	0,59	1,65	1,46	random testing cluster	
	10					2,06	B	110	4,70					10	4,31	0,49	1,78	1,07	random testing cluster	
	11							1,40	B	130	4,87			10	4,33	0,51	1,69	1,07	random testing cluster	
	12									6,46	R	150	5,01	10	4,40	0,55	1,43	1,07	random testing cluster	
	13	270	5,60									2,51	R	10	4,50	0,67	1,03	1,07	testing prod. units	no
	14	0,54	B	37	3,61														testing prod. units	no

	Production unit												*n* test values						
	1		2		3		4		5		6								
Sample	$x_{1,i}$	$\ln(x_{1,i})$	$x_{2,i}$	$\ln(x_{2,i})$	$x_{3,i}$	$\ln(x_{3,i})$	$x_{4,i}$	$\ln(x_{4,i})$	$x_{5,i}$	$\ln(x_{5,i})$	$x_{6,i}$	$\ln(x_{6,i})$	***n***	**Mean**	**St.dev.**	k_n	k_0	**Conclusion**	**Switch**
15			3,11	R	48	3,87												testing prod. units	no
16					2,63	R	34	3,53										testing prod. units	no
17							1,92	R	130	4,87								testing prod. units	no
18									7,62	R	120	4,79						testing prod. units	no
19	48	3,87									2,69	R	10	4,47	0,70	1,03	1,07	testing prod. units	no
20	0,82	R	20	3,00														testing prod. units	no
21			2,91	R	32	3,47												testing prod. units	no
22					2,57	R	11	2,40										testing prod. units	no
23							1,71	R	110	4,70								testing prod. units	no
24									4,86	R	64	4,16						testing prod. units	no
25	50	3,91									1,29	R	20	4,13	0,78	1,35	0,87	random testing cluster	yes
26	1,05	R	23	3,14									20	4,05	0,79	1,45	0,87	random testing cluster	
27			3,26	R	90	4,50							20	4,07	0,80	1,40	0,87	random testing cluster	
28					2,36	R	53	3,97					20	4,10	0,79	1,39	0,87	random testing cluster	
29							1,84	R	110	4,70			20	4,13	0,80	1,33	0,87	random testing cluster	
30									4,89	R	82	4,41	20	4,12	0,79	1,36	0,87	random testing cluster	
31	87	4,47									1,55	R	20	4,10	0,78	1,41	0,87	random testing cluster	
32	1,13	R	11	2,40									20	3,97	0,83	1,47	0,87	random testing cluster	
33			4,31	R	50	3,91							20	3,88	0,74	1,78	0,87	random testing cluster	

The outlined cells give the value of k_n of each production unit and assessment method (R = random testing, B = batch testing).

A.4 EXAMPLE 4: Assessment by attributes for a cluster

This example gives the evolution of the assessment of a cluster product based on assessment by attributes during TT and FT. It includes the use of the criteria in 7.2.1 (TT) and 7.2.2 (FT). This example shows that TT may be ended when there are 7 test values and that the cluster system may be applied to the intended cluster.

TT

The cluster comprises 6 production units and for the start of TT two samples were taken from production unit 1 and one from each of the others. In this example TT may end after having tested 7 batches as none of the batches exceeded the declared value (the value of $n_e \leq 1$, see Table 13).

FT

None of the 7 test values exceeded the declared value ($n_e = 0$) and so the cluster moved to FT and random testing at a frequency of 1 batch in 4, see Table 14. For the following production (Samples 8 to 19) batches are tested randomly. In time more test values become available allowing the cluster organization to do the assessment on the consecutive 12 last test results and following on the consecutive 21 last test results.

The test value of Samples 15 and 19 exceed the declared value. Because those samples were taken from a randomly tested production, the batches from which the samples have been taken are not rejected. From a practical point of view the producer (production unit 1) might want to exclude these batches from being placed on the market and take the necessary measures if practical.

The test result of Sample 19 causes the number of test results that exceed the declared value within the consecutive 12 last test values to exceed the acceptable number of test results that are permitted to exceed the declared value (1). As soon as this test result is known, FT changes to testing the production of each production unit separately instead of testing the cluster production. From that moment on each production unit is being assessed as described in Example 2. As shown in the example production units 1 and 3 need to switch to batch testing ($n_e > 0$), while the other production units should switch to random testing.

Instead it may be decided to ban the production units that caused the disruption from the cluster.

NOTE Because of the complexity a more detailed description of the testing of each production unit is omitted in this example. This example only describes the functioning of the cluster system.

A return to cluster testing is only permitted if:

1) all production units have produced at least one test value; and

2) the value of n_e for the consecutive last 12 test results ≤ 1 or the value of n_e for the consecutive last 21 test results ≤ 3 or the value of n_e for the consecutive last 29 test results ≤ 5.

After 6 test results (Samples 20 to 25) the conditions of return are met and as soon as the test result of Sample 25 is known, FT switches to random testing of the cluster production.

The production of production unit 3 does not seem in control yet, for the next test result of this unit exceeds the declared value again (Sample 27). As a result FT again changes to testing the production of each production unit separately as soon as this exceeding test result is known. Because this sample was taken from a randomly tested production, the batch from which the sample was taken is not rejected. From a practical point of view the producer (production unit 3) might want to exclude this batch from being placed on the market and take the necessary measures if practical.

Again each production unit is being assessed as described in example 2: production units 1 and 3 need to switch again to batch testing (n_e > 0), while the other production units should switch to random testing. So, after a second round of 6 test results (Samples 28 to 33) the condition of return is met again and as soon as the test result of Sample 33 is known, FT switches to random testing of the cluster production. After this point no further disruption of the test results occur and FT continues to be based on random testing of the cluster production.

In total 3 batches that exceed the declared value have been accepted out of 4 within a total of 35 batches.

	Individual productions						Total production
μ	110	60	100	100	120	100	100
σ	70	50	80	60	20	30	68
						Percentage > L_D	10 %
L_D	180						

		Production unit						*n* test values					
		1	**2**	**3**	**4**	**5**	**6**					**No. batches batch testing**	**Switch**
	Sample	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	$x_{4,i}$	$x_{5,i}$	$x_{6,i}$	***n***	n_e	n_a	**Conclusion**		
TT	1	94											
	2		92										
	3			120									
	4				82								
	5					140							
	6						86						
	7	130						7	0	1	cluster testing approved		
	Start							7	0	0	random testing cluster		
FT	8		26					7	0	0	random testing cluster	0	
	9		1	35				7	0	0	random testing cluster	0	
	10		B	1	89			7	0	0	random testing cluster	0	
	11			B	1	120		7	0	0	random testing cluster	0	
	12				B	2	130	12	0	1	random testing cluster	0	
	13	73				B	2	12	0	1	random testing cluster	0	
	14	0	110				B	12	0	1	random testing cluster	0	

	Production unit						*n* test values					
	1	**2**	**3**	**4**	**5**	**6**					**No. batches batch testing**	
Sample	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	$x_{4,i}$	$x_{5,i}$	$x_{6,i}$	***n***	n_e	n_a	**Conclusion**		**Switch**
15	R	0	330				12	1	1	random testing cluster	0	
16		R	1	99			12	1	1	random testing cluster	0	
17			B	0	140		12	1	1	random testing cluster	0	
18				R	0	110	12	1	1	random testing cluster	0	
19	370				R	0	12	2	1	testing production units	0	
20	1	27				R	12	2	1	testing production units	1	no
21	B	0	290				21	3	3	testing production units	2	no
22		R	2	41			21	3	3	testing production units	3	no
23			B	0	130		21	3	3	testing production units	4	no
24				R	0	85	21	3	3	testing production units	5	no
25	39				R	0	21	3	3	random testing cluster	6	yes
26	1	30				R	21	3	3	random testing cluster	0	
27	B	0	200				21	4	3	testing production units	0	
28		R	3	44			21	4	3	testing production units	1	no
29			B	0	130		29	4	5	testing production units	2	no
30				R	0	110	29	4	5	testing production units	3	no
31	100				R	0	29	4	5	testing production units	4	no
32	1	23				R	29	4	5	testing production units	5	no
33	B	0	100				29	4	5	random testing cluster	6	yes
34		R	3	19			29	4	5	random testing cluster	0	
35			B	0	130		29	4	5	random testing cluster	0	

The outlined cells give the value of n_e of each production unit and assessment method (R = random testing, B = batch testing).

A.5 EXAMPLE 5: No-further-testing (NFT) – assessment by variables

This example gives the evolution of the assessment of NFT based on assessment by variables during TT and FT. It includes the use of the criteria in 6.1.1 (TT), 6.1.2 (FT) and 6.3 (NFT). This example shows that TT may be ended when there are 2 test values.

TT

This example describes the assessment of NFT based on assessment by variables. TT is based on 2 samples. In this example TT is ended when there are two test results (the value of k_2 equals 2,63 while the critical value k_0 for 2 test values equals 2,18, see Table 1).

FT

FT starts off with random testing at a frequency of 1 batch in 2, see Annex E). During production (Samples 3 to 29) batches are tested randomly. The test value of Sample 11 exceeds the declared value, but because Sample 11 is taken from a randomly tested production, the batch from which the sample has been taken is not rejected; however, from a practical point of view the producer might want to exclude this batch from being placed on the market and take the necessary measures if practical.

At the start there are insufficient data to pass the NFT criterion. As FT continues more test results become available. In time sufficient data are collected to show that the risk of exceeding the declared value is less than 10 % (the true probability of exceeding the declared value of the data set of the example is 5 %). After 29 samples NFT may be applied.

Production

μ	110
σ	80
Percentage > LD	5 %
LD	260
ln(LD)	5,56

		n test values								all test values					
	Sample	x_i	ln(x_i)	***n***	**Mean**	**St.dev.**	k_n	k_0	**Conclusion**		***n***	**Mean**	**St.dev.**	k_{all}	k_0	**FT/NFT**
TT	1	150	5,01							batch accepted						
	2	100	4,61	2	4,81	0,29	2,63	2,18	TT may be ended	batch accepted						
	Start								random testing							
FT	3	88	4,48	3	4,70	0,28	3,10	1,09	random testing	batch accepted						
	4	47	3,85	4	4,49	0,48	2,24	0,82	random testing	batch accepted						
	5	150	5,01	5	4,59	0,48	2,03	0,69	random testing	batch accepted	5	4,59	0,48	2,03	5,36	FT
	6	57	4,04	5	4,40	0,46	2,52	0,69	random testing	batch accepted	6	4,50	0,48	2,20	4,41	FT
	7	110	4,70	5	4,42	0,47	2,42	0,69	random testing	batch accepted	7	4,53	0,45	2,31	3,86	FT
	8	61	4,11	5	4,34	0,49	2,49	0,69	random testing	batch accepted	8	4,48	0,44	2,47	3,50	FT
	9	120	4,79	5	4,53	0,43	2,40	0,69	random testing	batch accepted	9	4,51	0,42	2,48	3,24	FT
	10	210	5,35	5	4,60	0,54	1,79	0,69	random testing	batch accepted	10	4,59	0,48	2,02	3,05	FT
	11	280	5,63	5	4,92	0,59	1,08	0,69	random testing	batch accepted	11	4,69	0,55	1,58	2,90	FT
	12	100	4,61	5	4,90	0,60	1,10	0,69	random testing	batch accepted	12	4,68	0,53	1,67	2,78	FT
	13	28	3,33	5	4,74	0,89	0,92	0,69	random testing	batch accepted	13	4,58	0,63	1,56	2,68	FT
	14	33	3,50	5	4,48	1,05	1,03	0,69	random testing	batch accepted	14	4,50	0,67	1,58	2,59	FT
	15	98	4,58	5	4,33	0,94	1,31	0,69	random testing	batch accepted	15	4,51	0,65	1,63	2,52	FT

	n test values									all test values					
Sample	x_i	ln(x_i)	***n***	**Mean**	**St.dev.**	k_n	k_0	**Conclusion**		***n***	**Mean**	**St.dev.**	k_{all}	k_0	**FT/NFT**
16	110	4,70	5	4,14	0,67	2,11	0,69	random testing	batch accepted	16	4,52	0,63	1,67	2,46	FT
17	83	4,42	5	4,11	0,64	2,26	0,69	random testing	batch accepted	17	4,51	0,61	1,73	2,41	FT
18	46	3,83	5	4,21	0,52	2,61	0,69	random testing	batch accepted	18	4,47	0,61	1,78	2,36	FT
19	100	4,61	5	4,43	0,35	3,24	0,69	random testing	batch accepted	19	4,48	0,59	1,82	2,31	FT
20	190	5,25	5	4,56	0,51	1,95	0,69	random testing	batch accepted	20	4,52	0,60	1,73	2,28	FT
21	100	4,61	5	4,54	0,51	2,01	0,69	random testing	batch accepted	21	4,52	0,59	1,77	2,24	FT
22	110	4,70	5	4,60	0,51	1,90	0,69	random testing	batch accepted	22	4,53	0,57	1,79	2,21	FT
23	75	4,32	5	4,70	0,34	2,54	0,69	random testing	batch accepted	23	4,52	0,56	1,84	2,18	FT
24	80	4,38	5	4,65	0,37	2,47	0,69	random testing	batch accepted	24	4,52	0,55	1,89	2,15	FT
25	130	4,87	5	4,57	0,23	4,35	0,69	random testing	batch accepted	25	4,53	0,54	1,89	2,13	FT
26	70	4,25	5	4,50	0,27	3,96	0,69	random testing	batch accepted	26	4,52	0,54	1,94	2,11	FT
27	66	4,19	5	4,40	0,27	4,29	0,69	random testing	batch accepted	27	4,51	0,53	1,99	2,09	FT
28	64	4,16	5	4,37	0,29	4,09	0,69	random testing	batch accepted	28	4,50	0,52	2,03	2,07	FT
29	87	4,47	5	4,39	0,29	3,99	0,69	random testing	batch accepted	29	4,49	0,51	2,07	2,05	**NFT**

Annex B

Distribution of test values

B.1 General

The normal statistical techniques to test conformity assume that the data are normally distributed or that normality can be obtained by a suitable mathematical transformation. If the (transformed) data do not follow a normal distribution, these statistical techniques cannot be used, or using them may lead to false conclusions, increasing either the consumer's risk or producer's risk. Therefore reliable or safe assumptions need to be made with respect to the distribution of the test results.

Because release, emission or content values of dangerous substances cannot be negative the so-called Box-Cox transformation is a useful technique for trying to normalize a dataset with unknown distribution; however, this method is too complex to apply in a product standard. Therefore the easier solution is to use the logarithmic transformation (ln-transformation). This assumes that the true distribution is approximated by a lognormal distribution and this is often a reasonable estimation.

B.2 Leaching

For leaching, the distribution of test values is assumed to be lognormal. This means that ln-transformed test values have a normal distribution. In practice it is not always possible to prove a lognormal distribution as is shown in Table B.1. This table shows a summary of the evaluation of FPC leaching data of 16 different construction products[1)] and gives the number of data sets with a certain distribution.

Normality may be tested by for instance the extended W-test according to Shapiro and Wilk (1963) and Royston (1982) [7].

Table B.1 shows that a lognormal distribution is found for about half the data sets. Nevertheless, for some data sets a normal distribution is found. It appears that the lognormal distribution tends to change to normal if the coefficient of variation is relatively small ($s << \bar{x}$ however, in these cases the distance between the test values and the detection limit was large.

In the case of an unknown distribution more than 90 % of these distributions are characterized by a positive skewness and positive kurtosis (fat right tail and a higher peak compared to a normal distribution). Compared to a lognormal distribution the tails of the unknown distribution may be short or fat and the peak may be steep or flat. In general unknown distributions resemble a lognormal distribution which means that a lognormal distribution is a sufficiently reliable assumption for testing dangerous substances.

Shorter tails and a higher peak indicate that the product is more constant than assumed, while fat tails and a flat peak show more variation. Possible reasons for fat tails (relatively larger variation) are:

— the mean value and/or standard deviation of the production change in time, for instance due to changes in the production process or properties of the constituents;

— the product's leaching properties change while the product is kept in stock (e.g. due to oxidation or carbonation);

— chemical interaction or reaction between constituents;

1) The test results contain both up-flow percolation tests and tank tests. For each construction product 4 to 7 substances have been tested and for each substance 9 to 97 test values (average 47) were available. The distribution differs per substance and per product.

— mixing of different productions (for example as a result of clustering similar products);

— an uncontrollable or out of control production.

Table B.1 — Distribution of leaching data of 16 different construction products

RDS	lognormal distribution	normal distribution	unknown distribution
Ba	6	3	7
Cr	3	0	0
Mo	3	0	3
Se	3	0	0
V	6	0	5
Br	2	0	3
Cl	4	3	9
F	7	0	3
SO_4	11	1	4
Total	45	7	34

In case of a shorter tails the assumption of a lognormal distribution gives an overestimation of the risk of exceeding the declared value, which means that the consumer's risk is lower. In case of fat tails the consumer's risk is higher. But the effect of a deviation from a lognormal distribution seems limited in practice and the effect on the consumer's and producer's risk is small.

B.3 Release into air

No data available.

B.4 Content

For content, the distribution of test values is assumed to be lognormal. Table B.2 shows a summary of the evaluation of FPC leaching data of 7 different construction products[2)] and gives the number of data sets with a certain distribution. This table shows that the assumption of a lognormal distribution is sufficiently founded. The same remarks made about the distribution of leaching data apply to content.

Substances banned from construction products should not be present in any test result; however, for practical reasons a very low value may be taken as effectively meaning zero content. With such low values an unknown distribution should be assumed.

2) For each construction product 1 to 8 substances have been tested and for each substance 13 to 57 test values (average 42) were available. The distribution differs per substance and per product.

Table B.2 — Distribution of content data of 7 different construction products

RDS	lognormal distribution	normal distribution	unknown distribution
As	0	0	1
Cd	1	0	0
Cr	1	0	1
Cu	1	0	1
Ni	1	0	0
Pb	0	0	1
Se	1	1	0
Zn	2	0	0
Mineral oil	3	0	2
PAH (sum)	1	0	2
Total	11	1	8

Annex C

Checklist for Technical Committees

In order for these several options to become available in product standards, the relevant CEN Technical Committee will be required to:

a) Respond to the amendment to the mandate either accepting the list of mandated dangerous substances for further-testing or decide if some of the list of dangerous substances do not need further-testing because the release, emission or content is much lower than the declared value. If so, prepare or have prepared by another body a dossier of information to justify a WFT procedure for these RDS and submit it to the European Commission for approval.

b) Where regulatory and/or technical classes have not been established by the European Commission, establish technical classes or permit declared values for those dangerous substances that are linked to satisfying particular regulations and including an 'NPD class' for use, where appropriate. Decide whether the information under CE-marking may include declared values.

c) Revise or amend the relevant product standards to include a technical provision covering:

   1) the agreed list of dangerous substances for inclusion in the standard;

   2) where relevant, the dangerous substances for which performance may be declared on the basis of an approved WFT procedure together with the conditions for the WFT procedure to apply;

   3) for each dangerous substance, whether the technical class or declared value is to be based on release, emission or content and the reference test methods;

   4) a requirement that the FPC system specifically addresses the control of dangerous substances;

   5) the procedure for the statistical assessment of declared values with respect to dangerous substances;

   6) the procedure for NFT, see 5.4;

   7) requirements for CE-marking with respect to dangerous substances.

Further, with respect to RDS the following subjects need to be implemented in the product standard:

— definition of the product;

— definition of a batch (≤1/10th of a year's production for continually produced products);

— number of samples per batch (1);

— number of sample increments;

— procedures for TT, FT and NFT (see D.2 and D.3 for model text);

— assessment by variables and/or attributes;

— number of batches for the running mean (variables) or n_a (attributes);

— test frequency;

— return conditions from batch testing to random testing;

— use of a cluster system (see D.4 for model text);

— hierarchy of test methods, if not already covered for other properties;

— handling test values < detection limit;

— use of the gamma rule and establish the coefficient of variation;

— handling outliers (see D.5 for model text);

— use of indirect tests for FT purposes (see D.6 model text).

Annex D

Model clauses for product standards

D.1 Introduction

This annex contains a model clause that may be freely copied into product standards. The CEN Product Committee should select the options and complete the product-specific missing information. The subsequent clauses of Annex D are rules of application. These may also be copied into product standards or cited in product standards.

This model clause is covering for dangerous substances only and it should be considered in conjunction with the model standard and clauses provided by CEN.

The rules of application are written as normative requirements to support easy implementation in product standards; while the use of a particular rule of application is not normative (unless it is made so by the product standard). It is emphasized that the definition of the product and batch size are crucial for the reliability of the assessment, but remain undefined in this annex because they depend on the type of product. The batch size should be not more than 1/10 of a year's production.

D.2 Model clause for product standards

[*The CEN/TC should clearly define or describe the product and include a definition of the term 'batch'.*]

D.2.1 Statistical assessment of declared values for dangerous substances

Table D.1 lists the dangerous substances cited in notified regulations for [*product*], the declaration of performance and the reference test method. The producer shall declare a value for all the listed dangerous substances or use the no-performance-determined option if applicable.

Table D.1 — Substances listed in notified regulations for [product]

Substance	Declared value	Reference test method
Substance 1 [*List from mandate*]	[*Select from:* *Technical class (defined by the product committee)* *Declared value (defined by the producer)*]	.....
Substance 2	.....	.....
.....	.....	.....

The reference test method shall be used for type testing. For further-testing, indirect (non-reference) test methods may be used provided they satisfy the criteria given in CEN/TR 16797-2:2015, D.6.

The rules of application shall verify with a confidence of 90 % that the 50th percentile of the production is less than or equal to the declared value.

NOTE 1 Guidance and background to the AVCP with respect to dangerous substances is given in CEN/TR 16797–2.

For continually produced products, the scale of declaration shall be not more than [*scale, which should not be more than one tenth of the production over one year*]. The scale of declaration shall be defined in the factory

production control manual, or if the product is not continually produced, the maximum batch size shall be defined.

The minimum number of samples per batch shall be [*one or some other number*] comprising at least [*number*] increments. Sampling shall be in accordance with [*give reference or specify*].

For single production units, either the rule of application given in CEN/TR 16797-2:2015, [*D.3.1 or D.3.2*] (assessment by variables) or D.3.3 (assessment by attributes) shall be used for the statistical assessment of the declared value of each substance that require further-testing. The rule of application shall be defined in the factory production control manual.

NOTE 2 Guidance on these rules of application is given in CEN/TR 16797–1.

[*Add clause when appropriate*] For clusters of production units, either the rule of application given in D.4.3 (assessment by variables) or D.4.4 (assessment by attributes) shall be used for the statistical assessment of the declared value of each substance that require further-testing. The rule of application shall be defined in the factory production control manual.

NOTE 3 Guidance on these rules of application is given in CEN/TR 16797–1.

D.2.2 No-further-testing

Assessment of no-further-testing (NFT) verifies with a confidence of 99 % that the 90th percentile of the production is less than or equal to the declared value when the scale of declaration is a batch as defined in this standard.

For added confidence in a NFT decision, such a decision and the subsequent product controls should be audited by the appropriate accredited certification body.

D.3 Rules of application for single production units

D.3.1 Rule of application using assessment by variables

D.3.1.1 General

Assessment by variables assumes a lognormal distribution of the data and is based on the parameter k_n. The formula for k_n is:

$$k_n = \frac{\ln(L_D) - \overline{x}}{s}$$

where

n = number of test values;

$\ln$ = natural logarithm of the value;

$\overline{x}$ = running mean of the last consecutive *n* ln-transformed test values;

s = running standard deviation of the last consecutive *n* ln-transformed test values;

L_D = declared value.

If the data are proven to be normally distributed instead of lognormally distributed, k_n may be based on non-transformed values in which case the formula for k_n is:

$$k_n = \frac{L_D - \overline{x}}{s}$$

where

n = number of test values;

$\overline{x}$ = running mean of the last consecutive n test values;

s = running standard deviation of the last consecutive n test values;

L_D = declared value.

The value of k_n shall be calculated whenever a new test value is available. Values lower than the detection limit shall be allocated 0,7 × the detection limit value.

If a test result is suspected to be erroneous due to sampling or testing errors, CEN/TR 16797-2:2015, D.5 may be used to establish whether or not the test value is an outlier.

D.3.1.2 Type testing

Type testing (TT) shall comprise testing at least two consecutive batches and not more than 10 consecutive batches. Type testing shall continue until the criterion for k_n in Table D.2 is satisfied.

Table D.2 — Assessment by variables: Conformity criteria for TT

Number of test results	Conformity criterion for the batch	Criterion for ending TT
1	x_1 ≤ declared value	none, as the second batch shall be tested
2	x_2 ≤ declared value	k_2 ≥ 2,18, transfer to Table D.3, Column I
3	x_3 ≤ declared value	k_3 ≥ 1,09, transfer to Table D.3, Column II
4	x_4 ≤ declared value	k_4 ≥ 0,82, transfer to Table D.3, Column III
5	x_5 ≤ declared value	k_5 ≥ 0,69, transfer to Table D.3, Column IV
6 to 9	x_i ≤ declared value	k_5 ≥ 0,69, transfer to Table D.3, Column IV
10 [a)]	x_{10} ≤ declared value	k_5 ≥ 0,69, transfer to Table D.3, Column IV or [b] k_{10} ≥ 0,44, transfer to Table D.3, Column V

[a] If not already ended, TT ends when there are 10 test results. Testing every batch shall continue if it is required by the criterion given in Table D.3.

[b] The producer may opt to use the running mean of the last 5 or the last 10 consecutive tests results for FT. The number of consecutive tests shall be defined in the factory production control manual.

D.3.1.3 Further-testing

Further-testing (FT) is split into batch testing where every batch is tested and random testing where at least one batch within the number of successive batches defined in Table D.3 is tested.

During batch testing, any batch that is greater than the declared value shall be declared as non-conforming.

During random testing the minimum frequency of testing shall be at least as specified in Table D.3.

Table D.3 — Assessment by variables: Minimum test frequency for FT

Number of test results					Minimum test frequency [b c d e]
2	3	4	Last consecutive 5 [a]	Last consecutive 10 [a]	
I	II	III	IV	V	VI
$k_2 \geq 24{,}58$	$k_3 \geq 9{,}65$	$k_4 \geq 7{,}13$	$k_5 \geq 6{,}11$	$k_{10} \geq 4{,}63$	1 batch per 3 year
$18{,}50 \leq k_2 < 24{,}58$	$7{,}34 \leq k_3 < 9{,}65$	$5{,}44 \leq k_4 < 7{,}13$	$4{,}67 \leq k_5 < 6{,}11$	$3{,}53 \leq k_{10} < 4{,}63$	1 batch per year
$10{,}25 \leq k_2 < 18{,}50$	$4{,}26 \leq k_3 < 7{,}34$	$3{,}19 \leq k_4 < 5{,}44$	$2{,}74 \leq k_5 < 4{,}67$	$2{,}07 \leq k_{10} < 3{,}53$	1:10 batches (≥5 batches per 3 years)
$4{,}88 \leq k_2 < 10{,}25$	$2{,}23 \leq k_3 < 4{,}26$	$1{,}69 \leq k_4 < 3{,}19$	$1{,}46 \leq k_5 < 2{,}74$	$1{,}07 \leq k_{10} < 2{,}07$	1:4 batches (≥10 batches per 3 years)
$2{,}18 \leq k_2 < 4{,}88$	$1{,}09 \leq k_3 < 2{,}23$	$0{,}82 \leq k_4 < 1{,}69$	$0{,}69 \leq k_5 < 1{,}46$	$0{,}44 \leq k_{10} < 1{,}07$	1:2 batches (≥5 batches per year)
$k_2 < 2{,}18$	$k_3 < 1{,}09$	$k_4 < 0{,}82$	$k_5 < 0{,}69$	$k_{10} < 0{,}44$	test every batch [f g]

[a] It is the producer's choice to base the assessment on the running mean of the last 5 or 10 consecutive test results. The number of consecutive tests shall be defined in the factory production control manual.

[b] Where every batch is not being tested, the batch for testing shall be selected at random from within the period of production given in the last column.

[c] If the last consecutive 5 test values are lower than the limit of detection of the test method, the minimum test frequency is 1 batch per 3 years.

[d] If the last five consecutive test values are lower than 0,31 × declared value, the minimum test frequency is 1 batch per year; if they are lower than 0,19 × declared value, the minimum test frequency is 1 batch per three years.

[e] If the last 10 consecutive test values are lower than 0,41 × declared value, the minimum test frequency is 1 batch per year; if they are lower than 0,26 × declared value, the minimum test frequency is 1 batch per three years.

[f] Any batch above the declared value is classified as non-conforming and not placed on the market.

[g] Where batch testing is required during continuous production, at least five new batches shall be tested and the results give a $k_5 \geq 0{,}69$ and a $k_{10} \geq 0{,}44$ before it is permitted to change from testing every batch to random testing.

D.3.1.4 No-further-testing

For closely defined production that is consistently proven to be safe with respect to the release, emission or content of dangerous substances, no-further-testing (NFT) is required to validate the declaration of performance if the production satisfies the criteria given in Table D.4.

For added confidence in a NFT decision, such a decision and the subsequent product controls should be audited by the appropriate accredited certification body.

If the producer wishes to use NFT for all of some of the substances the producer shall:

— define the product with respect to composition, raw materials and technical and other properties relevant to the release, emission of content of the substances subjected to NFT;

— continually monitor these items and establish that the product lies within the defined range.

Assessment of NFT shall be based on all available valid test results.

Table D.4 — Assessment by variables: Criteria for NFT

Number of test results	Criterion for NFT as critical values where k_n ≥ the relevant values below
5	k_5 ≥ 5,36
6	k_6 ≥ 4,41
7	k_7 ≥ 3,86
8	k_8 ≥ 3,50
9	k_9 ≥ 3,24
10	k_{10} ≥ 3,05
11	k_{11} ≥ 2,90
12	k_{12} ≥ 2,78
13	k_{13} ≥ 2,68
14	k_{14} ≥ 2,59
> 14	[a]
∞	k_∞ ≥ 1,28
[a] Use values given in CEN/TR 16797-2:2015, Annex E.	

D.3.2 Rule of application using assessment by variables and the gamma rule

D.3.2.1 General

Assessment by variables assumes a lognormal distribution of the data and is based on the parameter k_n. The formula for k_n is:

$$k_n = \frac{\ln(L_D) - \overline{x}}{s}$$

where

n = number of test values;

ln = natural logarithm of the value;

$\overline{x}$ = running mean of the last consecutive *n* ln-transformed test values;

s = running standard deviation of the last consecutive *n* ln-transformed test values;

L_D = declared value.

If the data are proven to be normally distributed instead of lognormally distributed, k_n may be based on non-transformed values in which case the formula for k_n is:

$$k_n = \frac{L_D - \overline{x}}{s}$$

where

n = number of test values;

$\overline{x}$ = running mean of the last consecutive n test values;

s = running standard deviation of the last consecutive n test values;

L_D = declared value.

The value of k_n shall be calculated whenever a new test value is available. Values lower than the detection limit shall be allocated 0,7 × the detection limit value.

The use of the gamma rule assumes a lognormal distribution and a known coefficient of variation for the test data. If the coefficient of variation is not known statistical assessment of declared values shall be assessed according to D.3.1 or D.3.3.

For leaching data or content a coefficient of variation of 0,65 may be assumed. Gamma factors for other values of the coefficient of variation are given in CEN/TR 16797-2:2015, Annex F.

If a test result is suspected to be erroneous due to sampling or testing errors, CEN/TR 16797-2:2015, D.5 may be used to establish whether or not the test value is an outlier.

D.3.2.2 Type testing

Type testing (TT) is based on the use of the gamma rule and shall comprise testing at least two consecutive batches and not more than four consecutive batches. Type testing shall continue until the criterion for batches in Table D.5 is satisfied.

Table D.5 — Assessment by variables: Conformity criteria for TT for assessment by the gamma rule for data that have a coefficient of variation of 0,65

Number of test results	Conformity criterion for the batch	Criterion for ending TT
1	x_1 ≤ declared value	none, as the second batch shall be tested
2	x_2 ≤ declared value	all x_i ≤ 0,64 × declared value (i = 1 to 2), transfer to Table D.6
3	x_3 ≤ declared value	all x_i ≤ 0,82 × declared value (i = 1 to 3), transfer to Table D.6
4	x_4 ≤ declared value	all x_i ≤ 0,96 × declared value (i = 1 to 4), transfer to Table D.6

If after four consecutive batch tests the criterion in Table D.5 is not satisfied, TT may be continued according to D.3.1.2.

D.3.2.3 Further-testing

Further-testing (FT) is split into batch testing where every batch is tested and random testing where at least one batch within the number of successive batches defined in Table D.6 (2, 3 or 4 test results) or Table D.3 (5 or more test results) is tested.

During batch testing, any batch that is greater than the declared value or technical class limit shall be declared as non-conforming.

During random testing the minimum frequency of testing shall be at least that specified in Table D.6 (2, 3 or 4 test results) or Table D.3 (5 or more test results).

Table D.6 — Assessment by variables: Minimum test frequency for FT for testing by the gamma rule for data that have a coefficient of variation of 0,65

Number of test results	Every test result	Minimum test frequency
2	all x_i ≤ 0,12 × declared value (i = 1 to 2)	1 batch per 3 years
3	all x_i ≤ 0,15 × declared value (i = 1 to 3)	
4	all x_i ≤ 0,18 × declared value (i = 1 to 4)	
2	all x_i ≤ 0,19 × declared value (i = 1 to 2)	1 batch per year
3	all x_i ≤ 0,24 × declared value (i = 1 to 3)	
4	all x_i ≤ 0,28 × declared value (i = 1 to 4)	
2	all x_i ≤ 0,35 × declared value (i = 1 to 2)	1:10 batches (≥5 batches per 3 years)
3	all x_i ≤ 0,44 × declared value (i = 1 to 3)	
4	all x_i ≤ 0,51 × declared value (i = 1 to 4)	
2	all x_i ≤ 0,51 × declared value (i = 1 to 2)	1:4 batches (≥10 batches per 3 years)
3	all x_i ≤ 0,56 × declared value (i = 1 to 3)	
4	all x_i ≤ 0,76 × declared value (i = 1 to 4)	
2	all x_i ≤ 0,64 × declared value (i = 1 to 2)	1:2 batches (≥5 batches per year)
3	all x_i ≤ 0,82 × declared value (i = 1 to 3)	
4	all x_i ≤ 0,96 × declared value (i = 1 to 4)	
2	one x_i > 0,64 × declared value (i = 1 to 2)	test every batch
3	one x_i > 0,82 × declared value (i = 1 to 3)	
4	one x_i > 0,96 × declared value (i = 1 to 4)	

D.3.2.4 No-further-testing

For closely defined production that is consistently proven to be safe with respect to the release, emission or content of dangerous substances, no-further-testing (NFT) is required to validate the declaration of performance if the production satisfies the criteria given in Table D.4.

For added confidence in a NFT decision, such a decision and the subsequent product controls should be audited by the appropriate accredited certification body.

If the producer wishes to use NFT for all of some of the substances the producer shall:

— define the product with respect to composition, raw materials and technical and other properties relevant to the release, emission of content of the substances subjected to NFT;

— continually monitor these items and establish that the product lies within the defined range.

Assessment of NFT shall be based on all available valid test results.

D.3.3 Rule of application using assessment by attributes

D.3.3.1 General

Assessment by attributes is based on the number of test values that exceed the declared value (n_e) for the last consecutive n test results.

Whenever a new test value is available the value of n_e shall be determined. Values lower than the detection limit shall be allocated 0,7 × the detection limit value.

D.3.3.2 Type testing

Type testing (TT) shall comprise testing at least four consecutive batches and not more than 12 consecutive batches. Type testing shall continue until the criterion for n_e in Table D.7 is satisfied.

Table D.7 — Assessment by attributes: Conformity criteria for TT

Number of test results	Conformity criterion for the batch	Criterion for ending TT
1 to 3	x_i ≤ declared value	none, as the fourth batch shall be tested
4 to 6	x_i ≤ declared value	n_e = 0 for the last 4 test results, transfer to Table D.8, Column I
7	x_7 ≤ declared value	n_e ≤ 1 for the last 7 test results, transfer to Table D.8, Column II
8	x_8 ≤ declared value	n_e ≤ 1 for the last 7 test results, transfer to Table D.8, Column II
9	x_9 ≤ declared value	n_e ≤ 1 for the last 7 test results, transfer to Table D.8, Column II
10	x_{10} ≤ declared value	n_e ≤ 1 for the last 7 test results, transfer to Table D.8, Column II
11	x_{11} ≤ declared value	n_e ≤ 1 for the last 7 test results, transfer to Table D.8, Column II
12 [a]	x_{12} ≤ declared value	n_e ≤ 1 for the last 7 test results or [b] n_e ≤ 3 for the last 12 test results, transfer to Table D.8, Column III

[a] If not already ended, TT ends when there are 12 test results. Testing every batch shall continue if it is required by the criterion given in Table D.8.

[b] The producer may opt to use the last 7 or the last 12 consecutive tests results to start FT. The number of consecutive tests shall be defined in the factory production control manual.

D.3.3.3 Further-testing

Further-testing (FT) is split into batch testing where every batch is tested and random testing where at least one batch within the number of successive batches defined in Table D.8 is tested.

During batch testing, any batch that is greater than the declared value shall be declared as non-conforming.

During random testing the minimum frequency of testing shall be at least as specified in Table D.3, Column III.

Table D.8 — Assessment by attributes: Minimum frequency of testing for FT

Number of the last n consecutive test results in the assessment and the number of results above the declared value value n_e						**Minimum test frequency** [a b]
I		**II**		**III**		
n	n_e	n	n_e	n	n_e	
—	—	—	—	38 22	≤ 1 0	1:10 batches [c] (≥5 batches per 3 years)
—	—	—	—	21 12	≤ 3 ≤ 1	1:4 batches [c] (≥10 batches per 3 years)
4	0	7	≤ 1	12 7	≤ 3 ≤ 1	1:2 batches [c] (≥5 batches per year)
4	> 1	7	> 1	12	> 3	test every batch [d e]

[a] If the last consecutive 5 test values are lower than the limit of detection of the test method, the minimum test frequency is 1 batch per 3 years.

[b] Where every batch is not being tested, the batch for testing shall be selected at random from within the period of production given in the last column.

[c] It is the producer's choice to base the assessment on the last 7/12/22 or 12/21/38 consecutive test results. The number of consecutive tests shall be defined in the factory production control manual.

[d] Any batch above the declared value is classified as non-conforming and not placed on the market.

[e] Before it is permitted to change from testing every batch at least five new batches shall be tested and the test results having not more than one result above the declared value in the last 7 consecutive test results and not more than three results above the declared value in the last 12 consecutive test results.

D.3.3.4 No-further-testing

For closely defined production that is consistently proven to be safe with respect to the release, emission or content of dangerous substances, no-further-testing (NFT) is required to validate the declaration of performance if the production satisfies the criteria given in Table D.9.

For added confidence in a NFT decision, such a decision and the subsequent product controls should be audited by the appropriate accredited certification body.

If the producer wishes to use NFT for all of some of the substances the producer shall:

— define the product with respect to composition, raw materials and technical and other properties relevant to the release, emission of content of the substances subjected to NFT;

— continually monitor these items and establish that the product lies within the defined range.

Assessment of NFT shall be based on all available valid test results.

Table D.9 — Assessment by attributes: Criteria for NFT

Number of test results	Criterion for NFT: not more than the following number of test results above the declared value
44 — 63	0
64 — 80	1
81 — 96	2
97 — 112	3
113 — 126	4
127 — 141	5
142 — 155	6
156 — 169	7
170 — 182	8
183 — 196	9
197 — 209	10

D.4 Rules of application for clusters of production units

D.4.1 General

The rules of application for clusters of production units (D.4) are complementary to the rules of application for single production units (D.3).

D.4.2 Management of a cluster of production units

The cluster shall be run by an organization that is independent from any of the participating production units to avoid a conflict of interest if the production units are from more than one commercial enterprise.

The cluster organization may for instance be an accredited notified body in the case of a group of producers or an independent quality control unit in the case of a group of production locations of one producer.

If one single producer is involved, the involvement of an independent cluster organization is not required. In this situation the producer may manage the cluster.

The cluster organization and production unit are required to have a contract. This contract shall include rights and obligations of the cluster organization and production unit, including on site quality control by the cluster organization. The technical requirements in the contract shall be equal for all participating production units.

The use of a cluster system requires a complete description of the cluster product. The cluster organization shall define the cluster product with respect to composition, raw materials and/or constituent parts and technical and other properties relevant to the release, emission or content of RDS.

The cluster organization shall establish and maintain a quality control manual setting out the procedures by which the requirements for quality control are satisfied. This manual shall include:

— a description of the tasks, responsibilities and competences of the cluster organization and participating production units, including on site quality control by the cluster organization;

— a procedure for quality control by the participating production units, including sampling;

— a procedure for the coordination of all control activities by the cluster organization;

— a procedure for sample selection and testing by the cluster organization;

— a procedure and requirements for admittance to the cluster;

— a procedure for expelling from the cluster;

— a procedure for internal communication and protection of the anonymity of the participating production units;

— registration of the participating production units;

— maintenance of the product's definition.

D.4.3 Rule of application using cluster assessment by variables

D.4.3.1 Type testing

Type testing (TT) shall comprise testing at least one batch from each production unit with a minimum of five batches for the entire cluster. Type testing shall continue until the criterion for k_n in Table D.10 is satisfied and it is permitted to apply the cluster assessment.

If the criterion for k_n in Table D.10 is not satisfied, it is not permitted to apply cluster assessment to the set of production units and product as tested for the substance under consideration. In this case each production unit shall be assessed according to D.3.1.

D.4.3.2 Further-testing

Further-testing (FT) is split into random cluster testing, where at least one batch within the number of successive batches of the cluster as defined in Table D.11 is tested, and testing each production unit separately according to D.3.1.

During random cluster testing the minimum frequency of testing shall be at least that specified in Table D.11.

If the test data of a substance do not fulfil the requirements for applying cluster assessment all production units shall undertake separate testing of this substance according to D.3.1 until it is permitted to change to random cluster testing (see Footnote [i] in Table D.11).

D.4.3.3 No-further-testing

For closely defined production that is consistently proven to be safe with respect to the release, emission or content of dangerous substances, no-further-testing (NFT) is required to validate the declaration of performance if the production satisfies the criteria given in Table D.4.

For added confidence in a NFT decision, such a decision and the subsequent product controls should be audited by the appropriate accredited certification body.

If the cluster wishes to use NFT for all of some of the substances the cluster shall:

— define the product with respect to composition, raw materials and technical and other properties relevant to the release, emission of content of the substances subjected to NFT;

— continually monitor these items and establish that the product lies within the defined range.

Assessment of NFT shall be based on all available valid test results.

Table D.10 — Cluster assessment by variables: Conformity criteria for TT

Number of test results	Criterion for ending TT and applying cluster assessment
1 to 4	none, as the fifth batch shall be tested
5	$k_5 \geq 1{,}46$
6	$k_6 \geq 1{,}32$
7	$k_7 \geq 1{,}23$
8	$k_8 \geq 1{,}16$
9	$k_9 \geq 1{,}11$
10	$k_{10} \geq 1{,}07$
11	$k_{11} \geq 1{,}03$
12	$k_{12} \geq 1{,}00$
13	$k_{13} \geq 0{,}98$
14	$k_{14} \geq 0{,}96$
15	$k_{15} \geq 0{,}94$
16	$k_{16} \geq 0{,}92$
17	$k_{17} \geq 0{,}91$
18	$k_{18} \geq 0{,}90$
19	$k_{19} \geq 0{,}88$
20	$k_{20} \geq 0{,}87$
> 20	[a]

[a] Use values given in CEN/TR 16797–2:2015, Annex E.

Table D.11 — Cluster assessment by variables: Minimum test frequency for FT

Number of test results n			Minimum test frequency [b c d e f]
Last consecutive 5 [a]	Last consecutive 10 [a]	Last consecutive 20 [a]	
$k_5 \geq 6{,}11$	$k_{10} \geq 4{,}63$	$k_{20} \geq 4{,}01$	[g]
$4{,}67 \leq k_5 < 6{,}11$	$3{,}53 \leq k_{10} < 4{,}63$	$3{,}05 \leq k_{20} < 4{,}01$	[h]
$2{,}74 \leq k_5 < 4{,}67$	$2{,}07 \leq k_{10} < 3{,}53$	$1{,}77 \leq k_{20} < 3{,}05$	1:10 batches (≥5 batches per 3 years)
$1{,}46 \leq k_5 < 2{,}74$	$1{,}07 \leq k_{10} < 2{,}07$	$0{,}87 \leq k_{20} < 1{,}77$	1:4 batches (≥10 batches per 3 years)
$k_5 < 1{,}46$	$k_{10} < 1{,}07$	$k_{20} < 0{,}87$	test single production units [i]

[a] It is the cluster's choice to base the assessment on the running mean of the last 5, 10 or 20 consecutive test results. The number of consecutive tests shall be defined in the factory production control manual.

[b] The batch for testing shall be selected at random from within the period of production given in the last column.

[c] If the last consecutive 5 test values are lower than the limit of detection of the test method, the minimum test frequency is according to Footnote [g] of this table.

[d] If the last five consecutive test values are lower than 0,31 × declared value, the minimum test frequency is according to Footnote [h] of this table; if they are lower than 0,19 × declared value, the minimum test frequency is according to Footnote [g] of this table.

[e] If the last 10 consecutive test values are lower than 0,41 × declared value, the minimum test frequency is according to Footnote [h] of this table; if they are lower than 0,26 × declared value, the minimum test frequency is according to Footnote [g] of this table.

[f] If the last 20 consecutive test values are lower than 0,52 × declared value, the minimum test frequency is according to Footnote [h] of this table; if they are lower than 0,33 × declared value, the minimum test frequency is according to Footnote [g] of this table.

[g] Where there are less 20 production units: 1 batch per year per 2 production units. Where there are 20 or more production units: 10 batches per year for the entire cluster

[h] Where there are less 20 production units: 1 batch per 3 years per 2 production units. Where there are 20 or more production units: 10 batches per 3 years for the entire cluster

[i] Where testing single production units is required at least one new test value for each production unit shall be produced with a minimum of five and the criteria $k_5 \geq 1{,}46$ and $k_{10} \geq 1{,}07$ (n = 5 or 10) or $k_{10} \geq 1{,}46$ and $k_{20} \geq 0{,}87$ (n = 20) satisfied before it is permitted to change to random cluster testing. The sequence of testing all individual production units is repeated as long as the cluster system is rejected.

D.4.4 Rule of application using cluster assessment by attributes

D.4.4.1 Type testing

Type testing (TT) shall comprise testing at least one batch from each production unit with a minimum of seven batches for the entire cluster. Type testing shall continue until the criterion for n_e in Table D.12 is satisfied and it is permitted to apply the cluster assessment.

If the criterion for n_e in Table D.12 is not satisfied, it is not permitted to apply cluster assessment to the set of production units and product as tested for the substance under consideration. In this case each production unit shall be assessed according to D.3.3.

Table D.12 — Cluster assessment by attributes: Conformity criteria for TT

Number of test results	Criterion for ending TT and applying cluster assessment
1 to 6	none, as the seventh batch shall be tested
7 to 11	n_e = 0 for the last 7 test results, transfer to Table D.13, Column I
12 to 15	n_e ≤ 1 for the last 7 test results, transfer to Table D.13, Column II
16 to 20	n_e ≤ 2 for the last 7 test results, transfer to Table D.13, Column II
21 to 24	n_e ≤ 3 for the last 7 test results, transfer to Table D.13, Column III
25 to 28	n_e ≤ 4 for the last 7 test results, transfer to Table D.13, Column III
29 to 32	n_e ≤ 5 for the last 7 test results, transfer to Table D.13, Column IV
33 to 36	n_e ≤ 6 for the last 7 test results, transfer to Table D.13, Column IV
37 to 40	n_e ≤ 7 for the last 7 test results, transfer to Table D.13, Column IV
41 to 44	n_e ≤ 8 for the last 7 test results, transfer to Table D.13, Column IV
45 to 48	n_e ≤ 9 for the last 7 test results, transfer to Table D.13, Column IV
49 to 52	n_e ≤ 10 for the last 7 test results, transfer to Table D.13, Column IV
53 to 56	n_e ≤ 11 for the last 7 test results, transfer to Table D.13, Column V
57 to 59	n_e ≤ 12 for the last 7 test results, transfer to Table D.13, Column V
60 to 63	n_e ≤ 13 for the last 7 test results, transfer to Table D.13, Column V

Number of test results	Criterion for ending TT and applying cluster assessment
64 to 67	n_e ≤ 14 for the last 7 test results, transfer to Table D.13 Column V

D.4.4.2 Further-testing

Further-testing (FT) is split into random cluster testing, where at least one batch within the number of successive batches of the cluster as defined in Table D.13 is tested, and testing each production unit separately according to D.3.3.

During random cluster testing the minimum frequency of testing shall be at least that specified in Table D.13, Column V.

If the test data of a substance do not fulfil the requirements for applying cluster assessment all production units shall undertake separate testing of this substance according to D.3.3 until it is permitted to change to random cluster testing (see Footnote [d] in Table D.13).

Table D.13 — Assessment by attributes: Minimum frequency of testing for FT

Number of the last n consecutive test results in the assessment and the number of results above the declared value value n_e										Minimum test frequency [a] [b]
I		II		III		IV		V		
n	n_e	n	n_e	n	n_e	n	n_e	n	n_e	
—	—	—	—	22	0	38 22	≤ 1 0	52 38 22	≤ 2 ≤ 1 0	1:10 batches [c] (≥5 batches per 3 years)
—	—	—	—	21 12	≤ 3 ≤ 1	29 21 12	≤ 5 ≤ 3 ≤ 1	29 21 12	≤ 5 ≤ 3 ≤ 1	1:4 batches [c] (≥10 batches per 3 years)
7	> 1	12	> 1	12	> 1	12	> 1	12	> 1	test single production units [d]

[a] The batch for testing shall be selected at random from within the period of production given in the last column.

[b] If the last consecutive 5 test values are lower than the limit of detection of the test method, the minimum test frequency is 1 batch per 3 years per 2 production units where there are less 20 production units and 10 batches per 3 years for the entire cluster where there are 20 or more production units.

[c] It is the producer's choice to base the assessment on the last 12/22, 21/38 or 29/52 consecutive test results. The number of consecutive tests shall be defined in the factory production control manual.

[d] Where testing single production units is required at least one new test value for each production unit shall be produced with a minimum of five and the criteria $n_e \leq 1$ (n = 12), $n_e \leq 3$ (n = 21) or $n_e \leq 5$ (n = 29) satisfied before it is permitted to change to random cluster testing. The sequence of testing all individual production units is repeated as long as the cluster system is rejected.

D.4.4.3 No-further-testing

For closely defined production that is consistently proven to be safe with respect to the release, emission or content of dangerous substances, no-further-testing (NFT) is required to validate the declaration of performance if the production satisfies the criteria given in Table D.9.

For added confidence in a NFT decision, such a decision and the subsequent product controls should be audited by the appropriate accredited certification body.

If the cluster wishes to use NFT for all of some of the substances the cluster shall:

— define the product with respect to composition, raw materials and technical and other properties relevant to the release, emission of content of the substances subjected to NFT;

— continually monitor these items and establish that the product lies within the defined range.

Assessment of NFT shall be based on all available valid test results.

D.5 Identifying outliers

Outliers that are proven to be due to sampling or testing errors may be rejected. In other cases the outlier shall be included except where the cause is determined and excluded from further production.

The largest test value of n consecutive test values should be considered an outlier when

$$\frac{\ln\left(x_{max}\right)-\overline{x}}{s} > G_{\mathrm{P}}$$

The smallest test value of n consecutive test values should be considered an outlier when

$$\frac{\overline{x}-\ln\left(x_{min}\right)}{s} > G_{\mathrm{P}}$$

where

x_{max} = largest test value of n test values,

x_{min} = smallest test value of n test values,

$\overline{x}$ = mean of *n* ln-transformed test values,

s = standard deviation n ln-transformed test values,

G_{P} = critical value according to Table D.14.

NOTE For other values of n, see CEN/TR 16797–2:2015, Annex E.

Table D.14 — Critical value G_P for testing outliers

Number of test values	Critical value G_P
10	2,482
20	3,001
30	3,236
40	3,381
50	3,482

When a test value has been identified as outlier, the cause shall be investigated and documented. If the outlier is due the production process a corrective action may be needed.

D.6 Use of indirect tests

D.6.1 General

The use of indirect (non-reference) tests for FT purposes is permitted if the test frequency for a RDS is more than once a year and it is proven and documented that the indirect test gives a reliable result with respect to the result of the reference test. An indirect test is considered reliable enough when:

— the use of the indirect test leads to the same conclusion with respect to accepting and rejecting batches as the use of the reference test method, or

— the indirect test produces a test value that is always equal to or greater than the test value produced with the reference test.

For the continual validation of an indirect test, the producer shall at least once a year test a sample simultaneously with the indirect test as well as with the reference test and verify the regression model.

Indirect tests are not permitted for TT or in the case of a dispute.

D.6.2 Correlation

In the case of a developed and documented correlation between the indirect test and reference test and it is justifiable to apply a linear regression model, it is permitted to express the value measured with the indirect test according to this regression model:

$$x' = ax + b$$

where

x' = equivalent value for the reference test,

x = indirect test value,

a = slope of the estimated regression line $y = ax + b$

where

x = indirect test value and

y = reference test value,

b = intercept of the estimated regression line $y = ax + b$.

Correlation shall be established on the basis of at least 15 samples, while every test value shall be equal to at least 3 times the detection limit.

A justifiable correlation is satisfied when a linear regression model is applied and the (weighted) coefficient of determination $R^2 \geq 0{,}7$ and the slope ≥ 0,2. It is not permitted to extrapolate the linear regression model. Extrapolation is assumed to occur when the x value deviates more than 20 % from the x range of the regression model.

NOTE If the indirect test is based on the same release mechanism as the reference test, it may be expected that any regression model between the indirect test and reference test will pass through the origin. If the estimated intercept of the regression line (b) significantly deviates from zero on a linear scale, it may be possible that an external effect, not related to release, affected the measurements.

D.6.3 No correlation

If the correlation between the indirect test and reference test is not sufficient, the use a linear regression model becomes unreliable and it is not allowed to correct test values. In this case the use of indirect tests is only possible if the test produces a real test value (not a detection limit) that is always equal to or greater than the test value produced with the reference test. Using an appropriate statistical test, e.g. Wilcoxon signed-rank test for paired data, it shall be shown with a confidence of 90 % that the indirect test produces a test value equal to or greater than the reference test.

An indirect test does not necessarily need to be a release test, e.g. leaching test. Also indirect properties, like dry density or compressive strength, may be used to find a relation with the reference release test.

Annex E

Critical values for assessment by variables

Number of test values	Critical value k_0						G_P
	Risk of exceeding the declared value					NFT	
	50 %	30 %	10 %	1 %	0,1 %		
2	2,18	4,88	10,25	18,50	24,58	-	-
3	1,09	2,23	4,26	7,34	9,65	-	1,155
4	0,82	1,69	3,19	5,44	7,13	-	1,496
5	0,69	1,46	2,74	4,67	6,11	5,36	1,764
6	0,60	1,32	2,49	4,24	5,56	4,41	1,973
7	0,54	1,23	2,33	3,97	5,20	3,86	2,139
8	0,50	1,16	2,22	3,78	4,95	3,50	2,274
9	0,47	1,11	2,13	3,64	4,77	3,24	2,387
10	0,44	1,07	2,07	3,53	4,63	3,05	2,482
11	0,41	1,03	2,01	3,44	4,51	2,90	2,564
12	0,39	1,00	1,97	3,37	4,42	2,78	2,636
13	0,38	0,98	1,93	3,31	4,34	2,68	2,699
14	0,36	0,96	1,90	3,26	4,27	2,59	2,755
15	0,35	0,94	1,87	3,21	4,22	2,52	2,806
16	0,34	0,92	1,84	3,17	4,16	2,46	2,852
17	0,32	0,91	1,82	3,14	4,12	2,41	2,894
18	0,31	0,90	1,80	3,11	4,08	2,36	2,932
19	0,31	0,88	1,78	3,08	4,04	2,31	2,968
20	0,30	0,87	1,77	3,05	4,01	2,28	3,001
21	0,29	0,86	1,75	3,03	3,98	2,24	3,031
22	0,28	0,85	1,74	3,01	3,95	2,21	3,060

Number of test values	Critical value k_0						G_P
	Risk of exceeding the declared value					NFT	
	50 %	30 %	10 %	1 %	0,1 %		
23	0,28	0,85	1,72	2,99	3,93	2,18	3,087
24	0,27	0,84	1,71	2,97	3,90	2,15	3,112
25	0,26	0,83	1,70	2,95	3,88	2,13	3,135
26	0,26	0,82	1,69	2,94	3,86	2,11	3,158
27	0,25	0,82	1,68	2,92	3,84	2,09	3,179
28	0,25	0,81	1,67	2,91	3,83	2,07	3,199
29	0,24	0,81	1,66	2,90	3,81	2,05	3,218
30	0,24	0,80	1,66	2,88	3,79	2,03	3,236
31	0,24	0,80	1,65	2,87	3,78	2,01	3,253
32	0,23	0,79	1,64	2,86	3,77	2,00	3,270
33	0,23	0,79	1,64	2,85	3,75	1,98	3,286
34	0,22	0,78	1,63	2,84	3,74	1,97	3,301
35	0,22	0,78	1,62	2,83	3,73	1,96	3,316
36	0,22	0,77	1,62	2,82	3,72	1,95	3,330
37	0,21	0,77	1,61	2,82	3,71	1,93	3,343
38	0,21	0,77	1,61	2,81	3,70	1,92	3,356
39	0,21	0,76	1,60	2,80	3,69	1,91	3,369
40	0,21	0,76	1,60	2,79	3,68	1,90	3,381
42	0,20	0,75	1,59	2,78	3,66	1,88	3,404
44	0,20	0,75	1,58	2,77	3,65	1,87	3,425
46	0,19	0,74	1,57	2,76	3,63	1,85	3,445
48	0,19	0,74	1,57	2,74	3,62	1,83	3,464
50	0,18	0,73	1,56	2,73	3,60	1,82	3,482
52	0,18	0,73	1,55	2,73	3,59	1,81	3,499

Number of test values	Critical value k_0						G_P
	Risk of exceeding the declared value					NFT	
	50 %	30 %	10 %	1 %	0,1 %		
54	0,18	0,72	1,55	2,72	3,58	1,80	3,516
56	0,17	0,72	1,54	2,71	3,57	1,78	3,531
58	0,17	0,72	1,54	2,70	3,56	1,77	3,546
60	0,17	0,71	1,53	2,69	3,55	1,76	3,560
62	0,16	0,71	1,53	2,69	3,54	1,75	3,573
64	0,16	0,71	1,52	2,68	3,54	1,75	3,586
66	0,16	0,70	1,52	2,67	3,53	1,74	3,598
68	0,16	0,70	1,51	2,67	3,52	1,73	3,610
70	0,15	0,70	1,51	2,66	3,51	1,72	3,621
72	0,15	0,70	1,51	2,66	3,51	1,71	3,633
74	0,15	0,69	1,50	2,65	3,50	1,71	3,643
76	0,15	0,69	1,50	2,65	3,49	1,70	3,653
78	0,15	0,69	1,50	2,64	3,49	1,69	3,664
80	0,14	0,69	1,49	2,64	3,48	1,69	3,673
85	0,14	0,68	1,49	2,63	3,47	1,67	3,695
90	0,14	0,68	1,48	2,62	3,46	1,66	3,716
95	0,13	0,67	1,48	2,61	3,45	1,65	3,735
100	0,13	0,67	1,47	2,60	3,44	1,64	3,754
105	0,13	0,66	1,47	2,59	3,43	1,63	3,771
110	0,12	0,66	1,46	2,59	3,42	1,62	3,787
115	0,12	0,66	1,46	2,58	3,41	1,61	3,803
120	0,12	0,65	1,45	2,57	3,40	1,60	3,817
125	0,12	0,65	1,45	2,57	3,39	1,60	3,830
130	0,11	0,65	1,44	2,56	3,39	1,59	3,843

Number of test values	Critical value k_0						G_P
	Risk of exceeding the declared value					NFT	
	50 %	30 %	10 %	1 %	0,1 %		
135	0,11	0,65	1,44	2,56	3,38	1,58	3,856
140	0,11	0,64	1,44	2,55	3,38	1,58	3,867
145	0,11	0,64	1,44	2,55	3,37	1,57	3,879
150	0,11	0,64	1,43	2,55	3,37	1,57	3,889
160	0,10	0,64	1,43	2,54	3,32	1,56	3,910
170	0,10	0,63	1,42	2,53	3,31	1,55	3,927
180	0,10	0,63	1,42	2,53	3,30	1,54	3,946
190	0,09	0,63	1,41	2,52	3,30	1,53	3,963
200	0,09	0,62	1,41	2,51	3,29	1,52	3,978
210	0,09	0,62	1,41	2,51	3,29	1,52	3,991
220	0,09	0,62	1,40	2,50	3,28	1,51	4,007
230	0,08	0,62	1,40	2,50	3,28	1,51	4,017
240	0,08	0,62	1,40	2,50	3,27	1,50	4,032
250	0,08	0,61	1,40	2,49	3,27	1,50	4,042

Annex F

Gamma factor

Coefficient of variation	Number of test values n	Percentile				
		50th $\alpha = 0{,}050$ $\beta = 0{,}500$	70th $\alpha = 0{,}075$ $\beta = 0{,}300$	90th $\alpha = 0{,}100$ $\beta = 0{,}100$	99th $\alpha = 0{,}100$ $\beta = 0{,}010$	99,9th $\alpha = 0{,}100$ $\beta = 0{,}001$
0,35	1	0,57	0,51	0,42	0,29	0,23
	2	0,77	0,68	0,55	0,39	0,30
	3	0,89	0,78	0,63	0,44	0,34
	4	0,98	0,85	0,68	0,48	0,37
	5	1,00	0,91	0,72	0,51	0,39
	10	n.a.	n.a.	n.a.	0,60	0,46
	20	n.a.	n.a.	n.a.	0,69	0,53
0,50	1	0,46	0,40	0,30	0,18	0,13
	2	0,70	0,59	0,44	0,27	0,19
	3	0,85	0,71	0,52	0,32	0,22
	4	0,97	0,80	0,59	0,36	0,25
	5	1,00	0,88	0,64	0,39	0,27
	10	n.a.	n.a.	n.a.	0,49	0,34
	20	n.a.	n.a.	n.a.	0,60	0,42
0,65	1	0,38	0,31	0,22	0,12	0,07
	2	0,64	0,51	0,35	0,19	0,12
	3	0,82	0,65	0,44	0,24	0,15
	4	0,96	0,76	0,51	0,28	0,18
	5	1,00	0,85	0,57	0,31	0,19
	10	n.a.	n.a.	n.a.	0,41	0,26
	20	n.a.	n.a.	n.a.	0,52	0,33

Coefficient of variation	Number of test values n	Percentile				
		50th α = 0,050 β = 0,500	**70th α = 0,075 β = 0,300**	**90th α = 0,100 β = 0,100**	**99th α = 0,100 β = 0,010**	**99,9th α = 0,100 β = 0,001**
0,80	1	0,31	0,25	0,16	0,08	0,05
	2	0,59	0,45	0,29	0,14	0,08
	3	0,79	0,60	0,38	0,18	0,11
	4	0,95	0,72	0,45	0,22	0,13
	5	1,00	0,82	0,51	0,25	0,14
	10	n.a.	n.a.	n.a.	0,35	0,20
	20	n.a.	n.a.	n.a.	0,46	0,27
1,00	1	0,25	0,19	0,12	0,05	0,03
	2	0,53	0,39	0,23	0,10	0,05
	3	0,76	0,55	0,32	0,13	0,07
	4	0,94	0,68	0,39	0,16	0,09
	5	1,00	0,79	0,45	0,19	0,10
	10	n.a.	n.a.	n.a.	0,29	0,15
	20	n.a.	n.a.	n.a.	0,40	0,21
1,20	1	0,21	0,16	0,09	0,03	0,02
	2	0,49	0,35	0,19	0,07	0,03
	3	0,73	0,51	0,27	0,10	0,05
	4	0,94	0,64	0,35	0,13	0,06
	5	1,00	0,77	0,41	0,15	0,07
	10	n.a.	n.a.	n.a.	0,24	0,12
	20	n.a.	n.a.	n.a.	0,36	0,17
1,50	1	0,17	0,12	0,06	0,02	0,01
	2	0,44	0,29	0,15	0,05	0,02
	3	0,69	0,46	0,23	0,07	0,03

Coefficient of variation	**Number of test values** n	**Percentile**				
		50th $\alpha = 0{,}050$ $\beta = 0{,}500$	**70th** $\alpha = 0{,}075$ $\beta = 0{,}300$	**90th** $\alpha = 0{,}100$ $\beta = 0{,}100$	**99th** $\alpha = 0{,}100$ $\beta = 0{,}010$	**99,9th** $\alpha = 0{,}100$ $\beta = 0{,}001$
	4	0,93	0,60	0,29	0,09	0,04
	5	1,00	0,74	0,36	0,12	0,05
	10	n.a.	n.a.	n.a.	0,20	0,09
	20	n.a.	n.a.	n.a.	0,31	0,13

n.a. = not applicable

Bibliography

[1] CEN/TR 16797-1:2015, *Construction products: Assessment of release of dangerous substances — Guidance on the statistical assessment of declared values — Part 1: Principles and rules of application*

[2] CEN/TR 15858, *Construction products — Assessment of the release of regulated dangerous substances from construction products based on the WT, WFT/FT procedures*

[3] CEN/TR 16220, *Construction products —Assessment of release of dangerous substances — Complement to sampling*

[4] EC Expert Group on Dangerous substances, DS 129rev5, 2010, *The possibilities of reducing testing (Without Further Testing) against Essential Requirement no 3 (ER3)* (also circulated as CEN/TC 351 N 0286)

[5] CEN/TR 16496:2013, *Construction Products — Assessment of release of dangerous substances — Use of harmonised horizontal assessment methods*

[6] EN 16687, *Construction products - Assessment of release of dangerous substances - Terminology*

[7] ROYSTON P. Remark AS R94: *A remark on algorithm AS 181: The W-test for normality. Appl. Stat.* 1995, **44** (4) pp. 547–551

[8] LENTH R.V. Algorithm AS 243: *Cumulative distribution function of the non-central t-distribution. Appl. Stat.* 1989, **38** (1) pp. 185–189

[9] Redig, F., *Enkele berekeningen in verband met nieuwe regelgeving*, Faculteit Wiskunde en Informatica TU Eindhoven, May 2002 (in Dutch)

[10] ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

[11] ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

[12] ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

[13] ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*

[14] Agentschap NL, *Evaluatie Besluit bodemkwaliteit, Versie 5 definitief*, April 2011 (in Dutch)

This page deliberately left blank

British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards -based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at bsigroup.com/standards or contacting our Customer Services team or Knowledge Centre.

Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at bsigroup.com/shop, where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to bsigroup.com/subscriptions.

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member**.

PLUS is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit bsigroup.com/shop.

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email bsmusales@bsigroup.com.

BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK

Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

Copyright

Useful Contacts:

Customer Services
Tel: +44 845 086 9001
Email (orders): orders@bsigroup.com
Email (enquiries): cservices@bsigroup.com

Subscriptions
Tel: +44 845 086 9001
Email: subscriptions@bsigroup.com

Knowledge Centre
Tel: +44 20 8996 7004
Email: knowledgecentre@bsigroup.com

Copyright & Licensing
Tel: +44 20 8996 7070
Email: copyright@bsigroup.com

bsi.

...making excellence a habit.™