



BSI Standards Publication

Postal Services - Quality of Services - Measurement of the transit time of end-to-end services for single piece priority mail and first class mail

National foreword

This British Standard is the UK implementation of EN 13850:2012. It supersedes BS EN 13850:2002 which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee SVS/4, Postal services.

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.

© The British Standards Institution 2013.
Published by BSI Standards Limited 2013

ISBN 978 0 580 62469 8

ICS 03.240

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

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 January 2013.

Amendments issued since publication

Date	Text affected
------	---------------

EUROPEAN STANDARD
NORME EUROPÉENNE
EUROPÄISCHE NORM

EN 13850

December 2012

ICS 03.240

Supersedes EN 13850:2002+A1:2007

English Version

**Postal Services - Quality of Services - Measurement of the
transit time of end-to-end services for single piece priority mail
and first class mail**

Services postaux - Qualité de service - Mesure du délai
d'acheminement des services de bout en bout pour le
courrier prioritaire égrené et de première classe

Postalische Dienstleistungen - Dienstqualität - Messung der
Durchlaufzeit von Einzelbriefsendungen mit Vorrang und
Einzelbriefsendungen erster Klasse von Ende zu Ende

This European Standard was approved by CEN on 27 October 2012.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

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

Management Centre: Avenue Marnix 17, B-1000 Brussels

© 2012 CEN All rights of exploitation in any form and by any means reserved
worldwide for CEN national Members.

Ref. No. EN 13850:2012: E

Contents

Foreword.....	6
0 Introduction	7
0.1 General	7
0.2 Regulatory background	7
1 Scope.....	8
2 Normative references.....	8
3 Terms and definitions	9
4 Symbols and abbreviations	15
5 Transit time as a Quality-of-Service indicator	16
5.1 General	16
5.2 Transit time calculation	17
5.2.1 Measurement unit	17
5.2.2 Continuity of measurement	17
5.2.3 Calculation of the transit time	17
5.3 Service performance indicators	18
6 Methodology	18
6.1 Representative sample design	18
6.2 Minimum Sample Size (MSS)	19
6.2.1 Domestic measurement systems	19
6.2.2 Cross-border measurement systems	19
6.3 Determination of the design basis	20
6.3.1 General	20
6.3.2 Estimation of real mail flows	20
6.3.3 Design basis	21
6.4 Discriminant Mail Characteristics (DMC)	21
6.4.1 General	21
6.4.2 Determination of the discriminant mail characteristics	21
6.4.3 Geographical stratification	22
6.5 Geographical distribution of the panel	23
6.5.1 General	23
6.5.2 Small panels up to 90 panellists	23
6.5.3 Bigger panels over 90 panellists	24
6.6 Integrity of the measurement	25
6.7 Unbiased sample design	26
7 Report.....	26
7.1 Measurement results	26
7.2 Estimators	27
7.2.1 Accuracy	27
7.2.2 Panel turnover in relation to accuracy	27
7.3 Weighting of the results.....	28
7.3.1 Reasons for implementing a weighting system.....	28
7.3.2 Weighting caps	28
7.3.3 Design changes due to annual mail characteristic and postal flow changes	29
7.4 Content and timing	29
8 Quality control and auditing	30
9 The annexes	31
Annex A (normative) Accuracy calculation	32
A.1 Scope.....	32

A.1.1	General	32
A.1.2	Two stage sampling approach.	32
A.1.3	Covariance / Stratification / Accuracy calculation	32
A.1.4	The design factor	32
A.2	Symbols.	33
A.3	Variance calculation for one stratum	33
A.3.1	General calculation method	33
A.3.2	Relation-to-total variation	34
A.3.3	Intra-relation variation	34
A.4	Variance calculation for a stratified sample	35
A.4.1	Variance of a weighted sample design.	35
A.4.2	Final weight of the individual item.	35
A.4.3	Weighting basis.	36
A.4.4	Combination of weighting and covariance	36
A.5	Calculation of the confidence interval.	37
A.5.1	General	37
A.5.2	Normal approximation	37
A.5.3	Agresti-Coull approximation	39
A.5.4	Inverse Beta approximation	40
Annex B	(normative) Transit Time Calculation Rule	41
B.1	Working week transit time calculation rule / domestic and cross-border mail	41
B.2	Calculation rules	42
B.2.1	Rule 1: Collection Monday-Friday / Delivery Monday-Friday	42
B.2.2	Rule 2: Collection Monday-Friday / Delivery Tuesday-Saturday	43
B.2.3	Rule 3: Collection Monday-Friday / Delivery Monday-Saturday	44
B.2.4	Rule 4: Collection Monday-Saturday / Delivery Monday-Friday	45
B.2.5	Rule 5: Collection Sunday-Friday / Delivery Monday-Friday	46
B.2.6	Rule 6: Collection Monday-Saturday / Delivery Monday-Saturday	47
B.2.7	Rule 7: Collection Sunday-Friday / Delivery Monday-Saturday	48
Annex C	(normative) Quality control and auditing	49
C.1	Quality Control	49
C.1.1	Statistical design	49
C.1.2	Test item production	49
C.1.3	Provision of test items to the sender panellists	49
C.1.4	Sending test items	50
C.1.5	Receiving test items	50
C.1.6	Data collection	50
C.1.7	Data analysis and reporting	50
C.1.8	Archiving	51
C.1.9	Quality control and Information Technology (IT)	51
C.2	Auditing – general remarks	51
C.3	Audit of the design basis	52
C.3.1	General	52
C.3.2	Methodological audit	52
C.3.3	Results	52
C.4	Audit of the Quality-of-Service measurement system	53
C.4.1	Panel audit	53
C.4.2	Stability of the parameters	53
C.4.3	Instructions given to the panellists	53
C.4.4	General Audit of the system	53
Annex D	(normative) Relaxation related to flows with small real mail volumes	54
D.1	General	54
D.1.1	Scope	54
D.1.2	Measurement period	54
D.1.3	Minimum Sample Size (MSS)	54
D.2	Domestic mail flows	55
D.3	Cross-border mail flows	56
Annex E	(informative) Purpose of postal Quality of Service standards	58
E.1	General	58

E.2	Benefits of QoS standards.....	58
E.3	Use of the survey results for quality improvement	59
E.3.1	Detailed analysis	59
E.3.2	Other / broader concepts.....	59
Annex F	(informative) Considerations before implementing EN 13850	60
F.1	Limitations of EN 13850.....	60
F.2	Responsibilities.	60
F.2.1	General	60
F.2.2	Regulatory authority	61
F.2.3	Postal operator.	61
F.2.4	Independent Performance monitoring organisation	62
F.2.5	Auditor.....	62
F.3	Design of the measurement system.	63
F.3.1	Design parameters	63
F.3.2	Field of study.	64
F.3.3	Geographical coverage.....	65
F.3.4	Design requirements due to national peculiarities.....	66
F.4	Small mail volumes.....	66
F.4.1	General	66
F.4.2	Domestic.	66
F.4.3	Cross border	66
F.5	Measurement organisation.....	67
F.5.1	Role of the contractor	67
F.5.2	Independence.	67
F.5.3	Tender process	67
Annex G	(informative) Design basis.....	68
G.1	Discriminant characteristics.....	68
G.1.1	Representative sample design	68
G.1.2	Studies for the evaluation of possible candidates	69
G.1.3	Connection between Design Basis and Sample Design	71
G.2	Design basis.	72
G.2.1	Real mail studies for domestic mail	72
G.2.2	Real mail studies for cross border mail	75
G.2.3	Alternative design bases	75
G.3	Frequency of update.....	76
Annex H	(informative) Implementing EN 13850	77
H.1	Stages of the survey	77
H.1.1	Preparation.	77
H.1.2	Set-up	77
H.1.3	Pilot (testing phase).....	77
H.1.4	Faster implementation	78
H.1.5	Measurement period	78
H.2	Panellists	79
H.2.1	Representativeness	79
H.2.2	Risk of panellist identification	79
H.2.3	Induction and delivery	80
H.2.4	Panel turnover.	82
H.3	Validation and transit time calculation.	82
H.3.1	Data validation	82
H.3.2	Service standard	84
H.3.3	Transit-time calculation rule.....	85
H.3.4	Loss.....	86
H.3.5	Force majeure	86
H.4	Weighting	88
H.4.1	Weighting and stratification	88
H.4.2	Weighting caps.	93
H.5	Reporting of results	95
H.5.1	Reporting.	95
H.5.2	Archiving	97

H.6	Audit	98
H.6.1	General	98
H.6.2	Position of the auditor	98
H.6.3	Audit report	98
H.6.4	Selection of the auditor	99
H.6.5	Frequency of audit	99
H.7	Implementation timetables	99
Annex I	(informative) Application of the accuracy calculation.	103
I.1	Limitations of the accuracy calculation methods provided	103
I.1.1	Participants with high mail loads	103
I.1.2	Disproportional models beyond the capping system.	103
I.2	Recommendations for the application of the rules.	104
I.2.1	Unstratified end-to-end sample	104
I.2.2	Stratified simple random sample.	104
I.2.3	Approximation of the Binomial distribution.	105
I.2.4	Accuracy.	105
I.2.5	Accuracy application	105
I.3	The sample size.	107
I.4	General Example for a national yearly result.	107
I.4.1	Introduction	107
I.4.2	Design factor for an unstratified end-to-end sample.	109
I.4.3	Design factor for a stratified random sample	110
I.4.4	Accuracy calculation.	111
I.5	Simplified scenarios.	113
I.5.1	General	113
I.5.2	Transit time results up to 96 %.	113
I.5.3	Fully proportional sample	113
I.5.4	Single induction / delivery point.	114
I.5.5	Induction / delivery point with only one letter.	114
Annex J	(informative) Changes to the 2007 version of EN 13850	115
J.1	Methodology.	115
J.1.1	Accuracy and Minimum Sample Size (MSS)	115
J.1.2	MSS for flows with small real mail volumes	115
J.2	Transit-time calculation rule.	116
J.3	Accuracy calculation method.	116
J.3.1	Improved applicability.	116
J.3.2	Reduced bias in calculation.	116

Foreword

This document (EN 13850:2012) has been prepared by Technical Committee CEN/TC 331 "Postal Services", the secretariat of which is held by NEN.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by June 2013, and conflicting national standards shall be withdrawn at the latest by June 2013.

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.

This document supersedes EN 13850:2002+A1:2007.

According to the CEN/CENELEC Internal Regulations, the national standards organisations of the following countries are bound to implement this European Standard: 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 the United Kingdom.

0 Introduction

0.1 General

The European Commission emphasises the need to have common rules for the development of community postal services and the improvement of Quality-of-Service (QoS). The Commission has identified requirements for postal QoS-Measurement systems that include:

- Independent end-to-end measurement capabilities;
- A focus on national and cross-border distribution service performance;
- A single, uniform and reliable system for monitoring distribution service performance within the Union.

The Commission has acknowledged that the different postal traditions and cultures in Europe would not allow for the establishment of one common unified European measurement system and that national systems should have sufficient freedom to reflect national needs and peculiarities. On the other hand, they should fulfil a defined set of minimum requirements to satisfy the information interests of the Commission, the regulatory authority, postal customers and postal operators themselves. Any regulatory authority is free to adapt to national circumstances where the standard gives room to do so.

The objective of the measurement is to estimate the end-to-end transit time QoS given to the customer domestically in each European country and cross-border between the European countries. This European Standard refers to a number of principles and minimum requirements to be applied for the measurement of the end-to-end transit time service level.

0.2 Regulatory background

The regulatory basis of EN 13850 is laid out in the 97/67/EC, as amended by Directive 2002/39/EC and Directive 2008/6/EC.

Main guidance is given in Chapter 6 Quality of Service. Article 16 states: "Member States shall ensure that quality-of-service standards are set and published in relation to Universal Service in order to guarantee a postal service of good quality".

Furthermore, EN 13850 is mandatory for measuring the performance levels of single piece priority or first class mail which falls under the universal service¹.

For intra-community cross-border mail of the fastest standard category a minimum QoS level is laid down in the Directive 97/67/EC. At least 85 % of all letters shall have an end-to-end transit time of J+3 and less and at least 97 % of all letters shall have an end-to-end transit-time of J+5 and less.²

The mandate for this revised version of EN 13850:2002+A1:2007 is the Third mandate for Postal Services – M428:2008 which states that EN 13850 shall "take into account the local / regional / national specificities as well as the experience since its implementation, with the aim of having a more generic method in order to satisfy regulatory needs".

¹ See also: "Letter to all Members of the Postal Directive Committee, 21.03.2005, Brussels, Markt/E4/JR/DS/HM D(2005) – 2346" (N676, CEN/TC331)

² See also: "Postal Directive 97/67/EC: Article 18.1 and Annex" and "Postal Directive 2008/6/EC: Article 18.1 and Annex 2, Article 1"

1 Scope

This European Standard specifies methods for measuring the end-to-end transit time of domestic and cross-border Single Piece Priority Mail (SPPM), collected, processed and delivered by postal service operators. It considers methods using representative end-to-end samples for all types of single piece priority mail services for addressed mail with defined transit-time service levels offered to the customer. This standard is applicable to the measurement of End-to-End priority mail services.

The standardised QoS-measurement method provides a uniform way for measuring the end-to-end transit time of postal items. Using a standardised measurement method will assure that the measurement will be done in an objective and equal way for all operators in accordance with the requirements of the Directive 97/67/EC and its amendments.

It is not the purpose of this standard to measure the postal operators' overall performance in a way that provides direct comparison of postal service providers.

This European Standard relates to the measurement of the SPPM services given to household and business customers that post mail at street letterboxes, over the counter at post offices or have pick-ups at their offices. To cover flows with smaller mail volumes this European Standard includes flexibility areas for adapted implementation. For technical reasons this European Standard may not be suitable for the measurement of very small volumes of mail.

The end-to-end service measured may be provided by one operator or by a group of operators working either together in the same distribution chain or parallel in different distribution chains. This European Standard is not applicable for the measurement of end-to-end transit times in fields of study with more than one induction operator (Multi-Operator Environments), which require different methodologies. The method for end-to-end measurement specified in this European Standard is also not designed to provide results for the measurement of parts of the distribution chain.

This European Standard is not applicable for the measurement of end-to-end transit times of bulk mailers' services and hybrid mail, which require different measurement systems and methodologies (see, for example, EN 14534 Measurement of the transit time of end-to-end services of bulk mail).

This European Standard includes specifications for the quality control and auditing of the measurement system.

This European Standard does not specify:

- the minimum acceptable level of accuracy that will be required by the national regulatory authority;
- the target(s) that the regulatory authority might set;
- how the regulatory authority should determine whether the target(s) have been met.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

EN ISO 9001, *Quality management systems — Requirements (ISO 9001)*

EN ISO 9004, *Managing for the sustained success of an organization — A quality management approach (ISO 9004)*

EN ISO 19011:2011, *Guidelines for auditing management systems (ISO 19011:2011)*

ISO 3534-1:2006, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-2:2006, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 3534-3:1999, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*

ISO 10005, *Quality management systems — Guidelines for quality plans*

ISO 10007, *Quality management systems — Guidelines for configuration management*

ICC/ESOMAR, *International Code of Marketing and Social Research Practice (1995 revision)*

3 Terms and definitions

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

3.1

accuracy

closeness of agreement between a test result and the accepted reference value

Note 1 to entry: The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.

[SOURCE: ISO 3534:2006]

Note 2 to entry: In this standard the accuracy is expressed as $\pm\epsilon$, where 2ϵ is the length of the confidence interval at the confidence level 95 % for the parameter being estimated, namely the probability of attaining the transit time target.

3.2

aggregation

compounding of primary data into an aggregate for the purpose of expressing them in a summary form

3.3

audit

systematic and independent examination to determine whether activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve objectives

Note 1 to entry: The organisation carrying out the audit is called the *auditor*.

Note 2 to entry: A (full) audit may be carried out as an *initial audit* of a new or substantially changed system or as an initial audit by a new auditor. It may also be carried out as a *re-audit* of the same system by the same auditor in the next audit cycle.

Note 3 to entry: If an audit results in objections, then the auditor may require corrective actions until a defined deadline. A final check of these corrective actions is called *corrective audit*.

3.4

average (arithmetic mean)

sum of values divided by the number of values

[SOURCE: ISO 3534:2006]

3.5

bring service

mail collection or mail delivery service, specifically contracted by the customer

3.6

business panellist

panellist with an address other than a household address such as a company or an organisation

3.7

characteristic

property, which helps to identify or differentiate between items of a given population

Note 1 to entry: The characteristics may be either quantitative - by variables, or qualitative - by attributes.

[SOURCE: ISO 3534:2006]

Note 2 to entry: In this standard the population is SPPM items and the characteristics are related to type of senders, type of receivers, times and types of induction or delivery, physical aspects of test items, franking, etc.

3.8

city

geographically defined area according to national classification systems

3.9

clearance

operation of collecting postal items by a postal service provider

3.10

conformity

fulfilment of specified requirements

3.11

corrective action

action taken to eliminate the causes of an existing non-conformity, defect or other undesirable situation in order to prevent recurrence

3.12

country

territory of a nation with its own government

3.13

cross-border mail

mail from or to another state or from or to a third country

3.14

customer

natural or legal person purchasing a postal service from a postal operator

3.15

date of delivery

date on which a postal item is delivered to the address or to the addressee

3.16

date of induction (J)

date on which a postal item is posted, provided posting takes place before the last collection of that day

Note 1 to entry: The term *date of induction* has the same meaning as the term *date of deposit* in the Directive 97/67/EC.

Note 2 to entry: Last collection refers to the advertised last time for collection (not the actual time).

3.17

date of posting

date on which a postal item is posted (irrespective of whether it is posted before the advertised last collection of that day)

3.18

delivery point

physical location at which delivery of postal items by a postal operator takes place and where they leave the operator's responsibility

3.19

design basis

structure in the field of study for which the design of the measurement is representative. The design basis should be defined before the start of the measurement

Note 1 to entry: If a design basis other than measured real mail flows is selected, then statements regarding the representativity of the measurement shall be made in relation to the chosen design basis.

3.20

design factor

ratio of the variance of the estimator of the QoS indicator in the given sample design by the variance of the estimator in an elementary sample design of the same size. The design factor is always related to a given sample design and estimator

3.21

discriminant characteristic

characteristic affecting the outcome

Note 1 to entry: In this standard a characteristic is discriminant when the transit time significantly differs according to the different modes of the characteristic (see 6.4.2).

3.22

distribution

process from collecting mail at collecting points through sorting at the mail centre(s) to the delivery of mail items to the addressee

3.23

domestic mail

mail items sent and received within one country

3.24

effective sample size

total sample size divided by the design factor

3.25

end-to-end

routing from the access point to the network up to the point of delivery to the addressee

3.26

estimate

value of an estimator obtained as a result of an estimation

[SOURCE: ISO 3534:2006]

3.27

estimator

statistic used to estimate a population parameter

[SOURCE: ISO 3534:2006]

Note 1 to entry: In this European Standard, an estimator is a function of the observed values of test-item transit times allowing the estimation of the probability of attaining the transit time target.

3.28

field of study

total SPPM flow between defined postal areas

Note 1 to entry: Some examples for field of study could be:

- Domestic – one induction operator in one country
- Domestic – one induction operator in one part of a country
- Cross-border – one induction operator on a country-to-country link

- Cross-border – one induction operator to one delivery operator
- Cross-border – one induction operator to a group of delivery operators
- Cross-border – one induction operator to one country
- Cross-border – one country to one delivery operator
- Cross-border – one induction operator to a group of countries
- Cross-border – a group of countries to one delivery operator
- Cross-border – one country to one country

Note 2 to entry: Some mail flows between postal operators may not meet the technical requirements in this standard to qualify as fields of study, e.g. limited mail volumes (see Annex D).

3.29
geographical coverage

spread of postal services within a pre-defined geographical area

3.30
independent performance monitoring organisation

body charged with the monitoring of the QoS according to the methodology specified in this standard, which is external to, and having no links of ownership or control with the postal operator thus monitored

3.31
induction

deposition of mail into the postal network

3.32
inspection

activity such as measuring, examining, testing or gauging one or more characteristics of an entity and comparing the results with specified requirements in order to establish whether conformity is achieved for each characteristic

[SOURCE: ISO 3534:2006]

3.33
last collection time

advertised last time for collection or contracted latest time for collection

Note 1 to entry: This is often not equal to the actual collection time, because from the postal work-organisation point of view, the collection usually happens some time later than the advertised last collection time (e.g. the collection routing timetable can only be defined with some tolerance).

3.34
metered mail

mail franked by franking machines

3.35
office of exchange

place where a postal operator accepts cross-border mail from a postal operator of another country and prepares mail for the transfer to other countries

3.36
on-time performance

proportion of postal items within a given period of time with transit times meeting the specification

3.37
on-time probability

probability of the event that the transit time T of a mail item meets the transit time target t , i.e. does not exceed the specified number t of days: $P(T \leq t)$

3.38

one-Operator field of study

field of study with exactly one induction operator

Note 1 to entry: A one-operator field of study may be defined in a multi-operator postal environment.

Note 2 to entry: End-to-end postal operation in a one-operator field of study may include several postal operators for processing and / or delivery.

3.39

pick-up time

published time by which the postal operator commits to ensuring that the day's post is available in the P.O. Box for collection

3.40

panel turnover

permanent and active exchange of established panellists with new panellists

3.41

panel rotation

active change in the subset of established panellists, which are chosen for participation in a study, or in their tasks in the study from one period to the other

3.42

postal area

one of the parts into which a postal operator's whole territory is divided and which is suitable for characterising postal distribution peculiarities

3.43

postal catchment area

postal area served by a domestic sorting centre or by an office of exchange for cross border mail outbound or inbound. Catchment areas may be different for outbound and inbound mail or for different fields of study

3.44

postal item

item addressed in the final form in which it is to be carried by a postal service provider

3.45

postal performance indicator

expression used to characterise the performance of a postal operator

Note 1 to entry: In this European Standard, the performance indicator is derived from postal transit time statistics.

3.46

postal service

services involving the clearance, sorting, transport and delivery of postal items

3.47

priority item, first class item, A-class item

postal item sent with priority as defined nationally

3.48

private panellist

panellist with a household address

3.49

professional panellist

panellist who is paid to perform specific tasks, usually posting mail in various pre-defined posting points during a day and posting more frequently than private or business panellists

3.50

quality

totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs

3.51

quality assurance

all the planned and systematic activities implemented within the quality system and demonstrated as needed, to provide adequate confidence that an entity will fulfil requirements for quality

3.52

quality control

operational techniques and activities that are used to fulfil requirements for quality

[SOURCE: ISO 3534:2006]

3.53

quality evaluation

systematic examination of the extent to which an entity is capable of fulfilling specified requirements

3.54

real mail flow

number of postal items of a given type on a given link within the postal network

3.55

real mail study

studies on real mail flows or real mail characteristics involving sampling of real mail items

3.56

rural

geographical entity with less than a specified number of inhabitants

3.57

service standard

standard that specifies requirements to be fulfilled by a service, to establish its fitness for purpose

Note 1 to entry: For end-to-end transit time monitoring systems the service standard is the number of qualifying days within which items should be delivered. This service standard is also called 'transit time target'.

3.58

single piece mail

postal items posted and distributed via a postal service for which a 'single piece tariff' for individual postal items is set in the general terms and conditions of the postal service provider that receives the payment

Note 1 to entry: Single piece mail may be inducted using different modes of payment as long as the induction of individual postal items is not restricted.

Note 2 to entry: Excluded are all postal services that contain further requirements on the induction like, for example, the registration of items, minimum induction volumes, equal contents or a pre-sortation of the inducted mail.

3.59

sorting centre

place where the main sorting of mail is done

3.60

stamped mail

postal items paid for with postage stamps

3.61

stratification

division of a population into mutually exclusive and exhaustive subpopulations (called strata), which are thought to be more homogeneous with respect to the characteristics investigated than the total population

[SOURCE: ISO 3534:2006]

3.62

study domain

subpopulations for which separate results may be appropriate

Note 1 to entry: Study domains could be defined for example by geographic segmentation or by product

3.63

test item

postal item produced in the test measurement system for the purpose of measuring transit time Quality-of-Service. A test item should be assigned to a stratum in the field of study. It should be manufactured, inducted and delivered according to the discriminant mail characteristics defining the stratum

3.64

test period

period under which measurement has been carried out and for which the results are presented in a separate test report

3.65

time of delivery

time when a postal item is delivered at its delivery point

3.66

time of posting

time when a postal item is posted at its induction point

Note 1 to entry: The time of posting may be before or after the actual last time of collection.

3.67

transit time

number of days elapsed between date of induction and date of delivery of a mail item

Note 1 to entry: The transit time is calculated according to 5.2.3

3.68

urban

geographical entity with not less than a specified number of inhabitants

3.69

user

any natural or legal person benefitting from postal service provision as a sender or an addressee

3.70

weighting

compensation for the difference between the distributions of characteristics in the measurement system compared to the chosen design basis

4 Symbols and abbreviations

CMW	Calculated Mode Weights
df	Design factor
DMC	Discriminant Mail Characteristic
EC	European Commission
ESS	Effective Sample Size
EtE	End-to-End
IFW	Individual Final Weight
IRV	Intra Relation Variation
IT	Information Technology
J	Date of induction
MSS	Minimum Sample Size
OLR	One Letter Relations

QoS	Quality of Service
RMW	Real mail Mode-Weights
RMS	Real Mail Studies
RSW	Real mail Strata-Weights
RtT	Relation-to-Total Variation
SPPM	Single-Piece Priority Mail
SRS	Simple Random Sample
StrRS	Stratified Random Sample
StrEtE	Stratified End-to-End Sample
SWB	Standard Weighting Basis
USO	Universal Service Obligation
WB	Weighting Basis

5 Transit time as a Quality-of-Service indicator

5.1 General

The overall transit time QoS level is to be expressed as the percentage of mail distributed within J+n days end-to-end according to the Directive 97/67/EC and its amendments. It specifies a set of requirements for the design of a QoS measuring system for SPPM, involving the selection and distribution of test items sent and received by selected panellists. The sample design gives the specifications for the panellists and items to be representative of the chosen design basis. The design basis is the most appropriate structural information available to characterise all real mail distributed in the field of study.

The system for measuring the distribution of the end-to-end transit time of SPPM items shall be robust and shall give a statistical measurement at a defined level of accuracy.

These requirements also consider fields of study that cover flows with small mail volumes. Provisions for these areas are detailed in Annex D, where the fields of study are categorised in up to four mail-flow classes. References to higher flexibility in some of these classes are given in several notes throughout this standard.

All measurements shall refer to the end-to-end transit time.

The sample design shall be representative of the chosen design basis. The design basis shall be selected according to its ability to characterise all real mail distributed in the field of study.

The measuring system shall provide one annual figure for the field of study.

NOTE 1 The figure presented annually may require a test period of 1, 2 or 3 years according to Annex D.

NOTE 2 For cross-border mail flows in categories 2 or 3 of Annex D it will take 2 or 3 years, respectively, to reach the required minimum sample size.

The measurement methodology shall be objective and shall be auditable.

The field of study shall be used consistently throughout the measurement period.

If the field of study is composed of several study domains, it is the QoS measurement for the overall field, which shall comply with the requirements of this European Standard.

The group of countries in a cross-border field of study should be used consistently throughout the measurement period.

5.2 Transit time calculation

5.2.1 Measurement unit

The transit time of a postal item shall be measured in units of days and expressed as J+n days. The day of induction J is the date of the next collection after posting.

5.2.2 Continuity of measurement

The measurement system shall be continuous. Posting shall cover all months and weeks of the year and at least all collection days of the week in accordance with the definition of the measurement unit and the transit-time calculation rule. All periods of the year shall be included as well as Christmas, Easter and summer holiday periods.

Non-functioning of the postal operator and days of strikes or industrial disputes shall not be discounted. However, in case of “*force majeure*” events, deduction of corresponding periods may be considered. Any deduction shall be indicated in the reporting and be subject to audit.

Any intended deduction shall be reported to the regulatory authority without delay. Agreement with the regulatory authority on all planned deductions due to *force majeure* is required prior to the calculation of the annual report.

For an event to qualify as *force majeure*, the incident shall fulfil the following minimum requirements. It shall;

- not be caused by the operators involved in the distribution and / or their subcontractors,
- be unforeseeable and,
- be unavoidable by them.

It shall;

- be a rare event,
- have a provable impact on several consecutive days of distribution.

Thus in case of, for example, natural disaster or terror attacks it should be allowed to consider the deduction of the corresponding period during which operation is affected in such a way that transit times cannot be guaranteed by “normal” postal operation.

Test items shall be posted on all seven days of the week. Published days without collection (i.e. non-collection days made known to the public in advance) may be excluded.

5.2.3 Calculation of the transit time

For the purpose of this European Standard, transit times for domestic and cross-border mail shall be calculated according to the working week calculation rule as presented in Annex B. Published regional holidays may be subtracted in the calculation of transit time.

The calculation of the transit time takes into account test items posted before the last collection time of the day for the type of mail in the field of study. The last collection time is taken as stated on the postal letterbox, published at the post office or otherwise announced by the postal operator. Test items posted after that time shall be considered as posted on the next collection day.

If a test item is posted after the last collection time, then the day of induction J should be adjusted to the next working day for this type of mail.

Restricted collections, e.g. later collection times for delivery within a local area, may be taken into account provided they have been advertised to customers.

It is necessary to ensure that panellists can identify the delivery of the item to the address, or the receipt of the item to a named individual.

If P.O. Boxes are to be included in the sample, panellists will need to be able to confirm that they collect the mail daily after the published pick-up time.

If there are items included in the survey which are not letterbox friendly or for some other reason cannot be delivered without the presence of the panellist, it is the first attempt to deliver that should count as the date of delivery if the date of the first attempt to deliver can be retrieved from a written notification.

Only valid test mail items shall be included in the calculations.

5.3 Service performance indicators

The following indicators shall be used in the presentation of the transit-time service level result. Only valid test mail items shall be included in the calculations. All postal items delivered up to J+30 shall be considered in the calculations. Postal items not delivered by J+30 may be excluded, because they may be deemed as lost or are not detectable in the system any more within the reporting period.

- *On-time performance*: The percentage of postal items delivered within the defined service standard. The result shall be presented as the percentage of postal items arriving within the transit time J+n, where n represents the number of qualifying days for the service standard. All reports shall state the level of on-time performance accuracy achieved in the measurement period.
- *Cumulative distribution of delivery days*: The cumulative percentage of mail delivered within a given period from J+1 to J+10 shall be reported.

6 Methodology

6.1 Representative sample design

The study shall be based on a defined methodology.

The methodology shall be based on test mail volumes representative of the chosen design basis. It shall define indicators that are representative of the transit time QoS provided in the measurement period.

The design basis shall be defined before the start of the measurement. If a design basis other than measured real mail flows is selected then statements regarding the representativity of the measurement shall be made in relation to the chosen design basis.

It is common practice to define the design basis as the total system of existing real mail flows in the field of study (standard design basis). Statements regarding the representativity of the measurement shall be interpreted in line with the standard design basis if no reference to an alternative design basis is made.

The measurement shall be carried out by an independent performance-monitoring organisation. The test mail method shall consist of a process in which panellists act as senders and/or receivers. Senders induct test items into the postal operator's mail network and register date and time of posting; receivers register the date of delivery.

Senders and receivers shall be spread over all the field of study based on the panel size and the corresponding minimum number of postal areas to be covered, in order to fulfil the specifications of minimum sample size, maximum panellist workload, stratification and geographical coverage. The sending and receiving process shall be organised in order to fulfil the specifications of the sample design. The test mail shall be manufactured in order to fulfil the specifications of the discriminant mail characteristics.

A representative sample design may be realised either by a strict proportionality of the test mail flows with respect to the design basis, or by an over-representation or an under-representation of some strata. The latter requires corrective weighting, which allows restoring the proportionality.

At least once in every three years or earlier when major changes occur the sample design and weighting system shall be adjusted.

6.2 Minimum Sample Size (MSS)

6.2.1 Domestic measurement systems

Without calculation, 9625 items shall be taken as the Minimum Sample Size (MSS) for a domestic measurement system. With this MSS reliable results can be achieved for all possible performance levels.

The MSS may be reduced, if a performance level $\hat{p}$ better than 50 % can be expected. The MSS then shall be determined using the following table:

Table 1 — Minimum Sample Sizes for selected performance levels (Domestic)

96,35 %	95 %	92,5 %	90 %	87,5 %	85 %	82,5 %	80 %	75 %	70 %	65 %	60 %	55 %	50 %
1.350	1.850	2.700	3.500	4.250	4.950	5.600	6.200	7.250	8.125	8.800	9.275	9.550	9.625

For any performance level $\hat{p}$ the MSS is the value tabled below the highest performance level still lower than $\hat{p}$.

EXAMPLE Let $\hat{p}$ be expected to be 91 %. Then the 90 % is the highest tabled performance still lower than 91 %. The MSS is then 3500.

This determination of the MSS can be applied for all performance levels from 50 % up to 96,35 % (see also A.5.2.2). For performance levels of 96,35 % and above at least 1.350 items have to be taken. For performance levels below 50 % at least 9.625 items have to be taken.

NOTE For fields of study with small mail flows these MSS-requirements are relaxed (see Annex D).

A maximum of 12 letters per week shall be allocated to any domestic sender except business senders with a mode of payment other than 'stamped'. A maximum of 24 letters per week shall be allocated to them.

A maximum of 12 letters per week shall be allocated to any domestic receiver. On average, no receiver shall get more than 6 letters per week during his time of participation in the measurement period.

The given weekly workloads are maximum values. In most cases the actual workloads should be well below the maximum.

6.2.2 Cross-border measurement systems

Without calculation 386 items shall be taken as the Minimum Sample Size (MSS) for a cross-border measurement system. With this MSS reliable results be achieved for all possible performance levels.

The MSS may be reduced, if a performance level $\hat{p}$ better than 50 % can be expected. The MSS then shall be determined using the following table:

Table 2 — Minimum Sample Sizes for selected performance levels (Cross-Border)

97,5 %	95 %	92,5 %	90 %	87,5 %	85 %	82,5 %	80 %	75 %	70 %	65 %	60 %	55 %	50 %
120	148	175	201	226	250	270	290	325	351	370	381	385	386

For any performance level $\hat{p}$ the MSS is the value tabled below the highest performance level still lower than $\hat{p}$.

EXAMPLE Let $\hat{p}$ be expected to be 91 %. Then the 90 % is the highest tabled performance still lower than 91 %. The MSS then is 201.

This determination of the MSS can be applied for all performance levels from 50 % up to 97,5 % (with at least three non-performing items). For performance levels of 97,5 % and above at least 120 items have to be taken. For performance levels below 50 % at least 385 items have to be taken.

NOTE For fields of study with small mail flows these MSS-requirements are relaxed (see Annex D).

A maximum of six letters per week shall be allocated to any cross-border sender, except business senders with a mode of payment other than 'stamped'. A maximum of 12 letters per week shall be allocated to them. A maximum of six letters per week shall be allocated to any cross-border receiver.

The given weekly workloads are maximum values. In most cases the actual workloads should be well below the maximum.

6.3 Determination of the design basis

6.3.1 General

The design basis may be determined in different ways. The standard design basis is based on real mail flows. The standard design basis will be determined via a system of real mail studies estimating these flows. The standard design basis shall be used at least after the first measurement period (see 6.3.3).

Before and during the first measurement period the use of alternative design bases may be necessary.

Different discriminant mail characteristics may have different design bases.

The design basis shall be provided by the stakeholder(s) commissioning the measurement.

The design may be based on estimated real mail volumes in the field of study. The methodology and reasoning leading to the determination of the design basis shall be audited by an independent auditor.

6.3.2 Estimation of real mail flows

6.3.2.1 Real mail studies

If available and reliable, real mail studies shall be used as the standard design basis.

Real mail studies shall be designed before or parallel to the set up of the test-mail measurement system in order to sample real mail flow information and to obtain the structural information needed for the determination of the sample design. The real mail measurement period may run before or parallel to the test mail measurement period.

The real mail studies shall:

- consider all single piece priority flows of a given field of study;
- collect statistics on single piece priority real mail flows and real mail characteristics according to the minimum information requirements of the statistical design;
- be implemented with a statistical accuracy coherent with the statistical objectives and the accuracy of the QoS monitoring system itself;
- include detailed information about the precision of each estimated quantity;
- be implemented either by the postal operators themselves or by an external body.

The real mail studies shall be performed at a minimum once every third year.

Existing real mail studies that may be used for the purpose of this standard may not be able to distinguish between single piece and bulk mail on all real mail flows. If a distinction is not possible, this shall be indicated in the report.

6.3.2.2 Logistic / management data

The distribution of real mail flows corresponding to certain mail characteristics may be estimated using existing logistic or management data that is available out of regular postal operation.

EXAMPLE The total mail flow on a certain link may be estimated via the number of trucks used to maintain operation.

The type of logistic or management data on which the design is based shall be stated in the report.

6.3.3 Design basis

6.3.3.1 First measurement period

Depending on the timing of the real mail studies and the resulting availability of real mail flow data at the planning stage of the first measurement period, each of the two types of design bases may be an appropriate choice:

- *1st choice*: Measured real mail flows (standard design basis, historical data) from older or recent real mail studies;
- *2nd choice*: Estimates via logistic or management data (alternative design basis, historical data);

It is possible that the type of design basis varies between the discriminant mail-characteristics, for example if real mail study data is only available for some of the discriminant mail characteristics but not for all. For each discriminant mail characteristic the most reliable design basis available shall be chosen.

6.3.3.2 Running system

In a running system after the first measurement period (running phase) always the most recent real-mail study data shall be used for a reliable approximation of the real mail flow situation in the planning and reporting stages of the test mail measurement.

The first calculation in the running phase that shall use the standard design basis is the calculation of the report for the first measurement period. For this report real mail studies results for all DMCs have to be available.

The estimation of the distribution of the following real mail characteristics as a minimum is mandatory:

- The discriminant characteristics according to 6.4.2;
- The geographical stratification according to 6.4.3.

To estimate these distributions it is necessary to measure the size of the mail flows

- For each mode of the discriminant characteristics (see 6.4.2);
- For the chosen grid of geographical induction and delivery areas.

6.4 Discriminant Mail Characteristics (DMC)

6.4.1 General

Outward appearance of test items as well as sending and receiving processes shall not deviate from real mail, and it shall not be possible for postal operators at any stage in their operations to identify them as test items. The test items shall have correctly written addresses.

6.4.2 Determination of the discriminant mail characteristics

The geographical network of possible induction and delivery points in the field of study is regarded as discriminant for the transit time of all letters distributed within the network. Discriminant geographical mail characteristics include the location of the induction point, the location of the delivery point or logistic differences related to geography.

Geographical stratification shall be based on postal areas and shall be the basis of the sample design. It shall be replaced by other geographical parameters only if it is demonstrated by auditable proof, that these are more discriminant. The proof would most likely come from the results of a QoS survey.

The list of all other discriminant mail characteristics shall be reviewed for each operator at a minimum before each review of the real mail studies or at least every three years. It shall be based on the results of appropriate studies

on the subject. These studies shall be audited in accordance with Annex C. Results from the SPPM monitoring system itself may be used as decision basis if they are based on the most recent 12 months data available.

The scope of the review is to check each element on a list of potential candidates for DMCs if they are discriminant or not. *Only those elements of the list that prove to be discriminant have to be included into the stratification.*

The following list of possible characteristics shall be evaluated as a minimum at least in the modes as listed below if applicable in the field of study:

1. Mail characteristic referring to the induction / delivery point:

- *type of geographical area* by (i) urban or (ii) rural;
- *type of payment* by (i) stamped, (ii) metered or (iii) postage paid;
- *type of induction* by (i) mail street box, (ii) post office, (iii) collection from senders premises; (iv) induction in sorting centres
- *time of posting* by at least 2 intervals between collections in case of more than one collection per day;
- *type of delivery* by (i) street address (ii) P.O. box or (iii) delivery to receivers' premises.

2. Mail characteristics referring to the test letter itself:

- *formats* by at least two modes;
- *weight steps* by at least two modes;
- *addressing method* by (i) hand written and (ii) typed;
- *induction weekday* by (i) Monday (ii) Tuesday (iii) Wednesday (iv) Thursday (v) Friday (vi) Saturday (vii) Sunday.

It may turn out that some characteristics are only discriminant in certain modes.

EXAMPLE Type of payment may only be discriminant as 'stamped' versus 'other than stamped'.

For each discriminant characteristic at minimum the two main modes shall be represented.

The revision shall be made with auditable proof. The revised set of discriminant mail characteristics shall determine the stratification of the design. Newly set discriminant mail characteristics shall be included into the design in the next measurement period.

6.4.3 Geographical stratification

- The relation between the induction point and the delivery point characterises each letter geographically. Both points together represent a certain induction area, a certain delivery area, a certain distance and the complexity of the geographical postal logistics between them.
- The stratification by geographical mail characteristics shall mirror possible logistic differences in mail distribution between different induction/delivery areas and distribution distances. The geographical stratification is therefore based on the postal logistic structure in the field of study.
- For induction as well as delivery, the field of study shall be divided into at least three mutually exclusive postal areas. The postal areas shall be derived from the logistic structure of the postal operators. They shall provide complete coverage of the geography in the field of study.

NOTE For any cross-border mail flow at least three mutually exclusive postal areas are required in the country of origin and in the country of delivery, resulting in at least $3+3 = 6$ postal areas for each cross-border mail flow.

- The choice and number of postal areas shall be agreed with the regulatory authority.

- The grid of all possible relations between the chosen postal areas shall form the geographical stratification. With at least three induction areas and at least three delivery areas, there shall be **at least 3 strata** (relations). Postcode/zip code areas may be used as a tool to define postal areas.
- The postal areas used for different fields of study may differ from each other.

For cross-border mail flows between adjacent countries, areas within each country (usually near the border), which show higher cross-border trade exchange as well as higher private mail flows, should be taken into consideration (provision for the requirements of trans-national European regions).

For cross-border mail flows in categories 2 or 3 of Annex D, the representative stratification need not be achieved annually. It should be tried to come as close as possible to optimal representation annually in order to be able to reach a representative stratification for each full multi-year result.

For cross-border mail flows in categories 2 or 3 of Annex D the annual sample sizes that compose a multi-year result shall be spread evenly over the test period.

For cross-border mail flows in categories 2 or 3 of Annex D and a multiple-country field of study the minimum stratification shall apply on each of the participating countries inbound and/or outbound.

If the survey design is changed during the test period then checks shall be made that the results of the survey before and after the design change shall be combined in a way, which is representative of the design basis in the test period.

6.5 Geographical distribution of the panel

6.5.1 General

The geographical distribution of the panel shall be done according to random sampling. The whole of the geographical area defined in the field of study shall be eligible.

The panels of senders and receivers shall be dispersed over a geographical recruitment grid based on postal zones served by the operators. The dispersion of the panel shall guarantee that all postal areas defining the geographical stratification (see 6.4.3) are covered. The geographical recruitment grid however may be different from the geographical stratification.

EXAMPLE The stratification may be based on the catchment areas of the delivery mail centres. A delivery office is usually linked to a delivery mail centre which in turn serves several delivery offices. The catchment areas of the postal delivery offices may serve as a (finer) geographical recruitment grid. In this example the panellists shall be dispersed over a high number of delivery offices (see below) covering all delivery mail centres included in the stratification.

6.5.2 Small panels up to 90 panellists

The minimal number of postal zones in the grid to be covered by the panel will be linked to the size of the panel; the rule shall be applied separately on the induction and on the delivery side.

Table 3 — Minimal number of postal zones to be covered in panels up to 90 panellists

Number of panellists	Minimal number of postal zones to be covered
10	4
20	7
30	10
40	14
50	17
60	20
70	24
80	27
90	30

EXAMPLE 1 Suppose that in a cross-border study 30 panellists are required for induction and 40 for delivery. Then a minimum of 10 postal zones should be covered in the sending country and a minimum of 14 postal zones should be covered in the receiving country.

EXAMPLE 2 Suppose that 43 panellists are required. By interpolation, the minimum number of zones to be covered is 15.

6.5.3 Bigger panels over 90 panellists

For studies with more than 90 panellists, the panel shall cover at least 30 postal zones spread over the whole field of study. In this case, the coverage additionally depends on the relation between panel size and number of zones available. The rule shall be applied separately on the induction and on the delivery side.

If on average three or more panellists could be recruited in each geographical zone, all postal zones in the field of study shall be included into the measurement.

Table 4 — Minimal number of postal zones to be covered in panels over 90 panellists

Average Number of Panellists per Area	Minimum Percentage of Areas to be Covered
0,25	25 %
0,36	30 %
0,51	37 %
0,72	45 %
1,03	55 %
1,47	67 %
2,10	82 %
3,00	100 %

EXAMPLE 1 Suppose that in a domestic study, 119 panellists are required and these are to be distributed over 81 areas by random sampling. Then there are on average 1.47 panellists per area. The minimum number of areas to be covered is 67 % of 81 equals 54.3 areas, which should be rounded up to 55 areas. Thus, the panellists should be distributed so that 55 up to 81 areas are covered.

EXAMPLE 2 Suppose that 120 panellists are required and that these are to be distributed over 44 areas by random sampling. Then there are on average 2.73 panellists per area. 2.73 lie between 2.1 and 3.0 in the table; by interpolation, the minimum percentage of areas to be covered is 95,0 %. The minimum number of areas to be covered is 94,6 % of 44 = 42 areas. Thus, the panellists should be distributed so that at maximum two areas are omitted.

6.6 Integrity of the measurement

For the integrity of the measurement, it is essential to secure that no improper influence is exercised on the measurement. Therefore, it has to be guaranteed that:

- Persons who might be able to influence the measurement results are not recruited

EXAMPLE Persons connected to the measurement (e.g. staff of postal operators or the regulatory authority).

- The panels of senders and receivers shall not be known to the staff of the postal operators.

The panel of senders and receivers shall therefore be independent of postal operators and shall be managed according to the ICC/ESOMAR “International Code of Marketing and Social Research Practice” including its attachments. This requires, among other things, that the identity of participants in research or measurement shall not be revealed to third parties outside the independent performance monitoring organisation and the auditing body. This applies both to individuals and to organisations. Together with the identity of the panellists, the exact location of induction and delivery points of the active panel shall remain unknown to the postal operator.

This requirement shall ensure that a postal operator cannot influence either the behaviour or reporting of the panellists, or the QoS provided specifically to them, thus introducing bias. If their identity is kept secret, there can be no opportunity for influencing behaviour.

To keep the identity of panellists concealed a series of measures has to be taken:

- Single-item results which may be used for quality improvement procedures by the operator shall omit any information on the exact location of the induction and delivery points.

It shall not be possible to identify any sender or receiver.

It shall not be possible to identify individual employees processing a single item.

It shall not be possible to identify delivery areas in the field of study, which are not covered by the measurement due to panel allocation.

- The test letters shall have an outward appearance that shows no obvious differences from the letters usually sent by the sender if the sender inducts at a post office or by a pick-up service.
- The weekly workload for any sender who inducts at a post-office shall, at most, double the panellist's usual amount of inducted real mail letters per post office or by a pick-up service.
- The weekly workload for any household receiver shall be kept at an inconspicuous level. The maximum weekly workload that shall be allocated to a household receiver shall be twelve letters.

In case the panellist takes part in more than one test mail measurement system, this workload cap shall apply to the total amount of test mail received from all test mail measurement systems.

The workload should in most cases be well below twelve letters.

- Furthermore, for every receiver, the independent performance monitoring organisation has to consider:
 - the weekly amount of international mail that may plausibly be received without raising suspicion,
 - the weekly amount of larger than standard mail formats that may pass as plausible, too,
 - whether other test letter formats exist which might increase the risk of identification,
 - whether high or low individual real mail volumes are received during the measurement,
 - the level of contact to the delivery postman.

Based on these factors and the requirements of 6.2, a system of workload caps that guarantees with auditable proof that the risk of receiver identification is kept low in the panel shall be developed and respected.

6.7 Unbiased sample design

The design of the measurement system shall respect the discriminant mail characteristics as detailed in 6.4.2, 6.4.3 and the weighting procedure in 7.3.

It shall neutralise biases that may occur with

- induction or deliveries at P.O. boxes;
- mail room effects;
- deliveries on non-working days in companies (for instance Saturdays);
- bring services and other special collection or delivery services.

P.O. boxes may be excluded from the measurement system, for example if an inclusion would jeopardise the validation of the delivery of the test items.

If changes in the design basis occur during the measurement period, for example due to changing results of continuous real mail studies, the sample design should allow for a correction of the growing bias as fast as possible to keep final weighting on a low level. This could be achieved by a continuous adaptation of the mail allocation plan in the measurement period.

If there are published multiple collections per day at certain induction points and the time of posting has proved to be discriminant, it shall be prescribed to panellists using these induction points between which collections they have to make the posting. In this case, the distribution of the actual time of posting has to be determined with respect to the published collection times. The prescribed spread of the test mail between the published collections has to be in proportion to the chosen design basis (see 6.3).

EXAMPLE The result of the real mail studies may be that, for induction points with two collections, on average 70 % of the postings are made before the 1st and 30 % between the 1st and the last collection.

If more than two published collections per day exist, it may turn out that the significance of the time of posting can only be proved for letters inducted before the *last* and before the *second to last* collection (two intervals). All other collections may turn out to be insignificant to the transit-time result. In this case, only these two intervals should be used for stratification.

If only one published collection per day exists or the time of posting has not proved to be discriminant, the time of posting shall not be influenced by the independent organisation running the monitoring system. The natural posting behaviour of the panel shall be respected. The panel shall record the time of posting of each item.

7 Report

7.1 Measurement results

The domestic measurement system shall provide one annual result for the overall transit time QoS-level according to the design basis in the field of study. The cross-border measurement system shall provide one annual result for the overall transit time QoS-level according to the design basis in the field of study.

For cross-border mail flows in categories 2 or 3 of Annex D, the results shall be cumulative over 2 or 3 years during the test period.

The measured transit time QoS-level shall be an estimate of the probability that the transit time of a randomly selected item has attained the end-to-end transit time target (service standard).

7.2 Estimators

7.2.1 Accuracy

Probabilities for attaining the transit time target shall be calculated based on the sampled test letters and an estimator. According to the two different types of measurement results, two different estimators are necessary. Let:

- $\hat{p}_1$ be the estimator for p_1 the true probability for attaining the transit time target for domestic mail;
- $\hat{p}_2$ be the estimator for p_2 the true probability for attaining the transit time target for cross-border mail.

For the calculated estimators $\hat{p}_i$, $i = 1, 2$ the accuracy per measurement period is given by the maximum length $2\epsilon_i$ of the confidence interval for $\hat{p}_i$ for a given confidence level $(1 - \alpha) = 95\%$.

The accuracy of the transit time QoS-level shall be assessed by calculating the variance of the estimator $\hat{p}_i$ and the corresponding design factor. This calculation shall take into account for each test letter as a minimum:

- the induction / delivery point and date,
- all discriminant mail characteristics,
- an indication if the transit time target has been attained or not,

in accordance with Annex A.

7.2.2 Panel turnover in relation to accuracy

On-time QoS indicators of test items sent from the same panellist, that is test items sent from a small geographic area, may be correlated with each other reducing the accuracy of the measurement. This might be the case even if the test items are not sent within a short period.

A turnover of the panel is one way to increase the calculated accuracy by increasing the number of induction and/or delivery points during the measurement period. This advantage of exchanging induction or delivery points by exchange of panellists should be balanced against the importance of experienced and reliable panellists.

Any induction point used by a private sender and any receiving household address shall not be used for more than four consecutive years.

For an optimal correlation reduction the exchange of induction and delivery points should be performed continuously throughout the years.

The induction and / or delivery point of a panellist may be reselected after being inactive for a minimum of three months, if it can be shown by auditable proof that no alternative induction or delivery point was available to be selected with reasonable effort in the measurement period (see C.4.1).

EXAMPLE If a field of study includes metered mail, but the number of postal customers inducing metered mail is limited, the choice for possible new panellist that are able to meter mail may be restricted to a small group only.

A panellist may pause and be reselected *at any time* if his induction and delivery points change with the reselection (Panel rotation).

The exact induction and delivery points of the active panel shall remain unknown to the postal operator. If panellists become known, they shall be excluded and not be reselected.

7.3 Weighting of the results

7.3.1 Reasons for implementing a weighting system

7.3.1.1 Weighting according to the sample design

For all discriminant mail characteristics – except geography – the sample design shall be based on a proportional model where the test mail flows for each stratum are proportional to the chosen design basis, usually being a system of real mail flows.

The geographic stratification may be based on a disproportional model. In a disproportional model, an aggregated result can only be achieved as a sum of weighted strata results.

For reasons beyond this standard, it might be desirable to set up a disproportional model for a number of discriminant mail characteristics apart from geography. This shall only be done if the chosen weighting system is transparent in the report and if it is able to restore proportionality with individual final weights (see 7.3.2.3) with a maximal magnitude of $\frac{1}{MSS}$ to any single item.

7.3.1.2 Weighting due to non-response and invalid test items

The sample design defines the target proportions of the modes of all discriminant mail characteristics. In the test mail sample these target proportions will usually not be achieved due to cases of non-response or invalid test items. Corrective weighting of the strata shall therefore always be implemented in the calculation to restore proportionality.

The discriminant mail characteristic modes shall be weighted in reference to the chosen design basis, usually being the correct real mail proportions. The strata weights shall be determined in accordance with the design basis.

7.3.2 Weighting caps

7.3.2.1 General

- A system of controls is necessary to avoid overall results that are based on an extreme weighting of some of the measured strata results. This standard uses two types of controls, one at the level of each discriminant characteristic and one at the stratum level.

7.3.2.2 Weighting caps for each discriminant characteristic

- The proportions of valid mail per mode shall not differ relatively by more than *20 % of the target proportions* set for the modes of each discriminant characteristics at the beginning of the measurement period.

EXAMPLE (characteristic: *Weekday*, mode: *Monday*): A sample requirement that 17,5 % of test items shall be posted on Monday shall be fulfilled with a margin of $\pm 17,5 \% \cdot 20 \% = \pm 3,5 \%$ on an annual basis.

NOTE 1 Due to small sample sizes for countries in categories 2 and 3 of Annex D, for domestic measurement the above constraints are relaxed to a maximum relative deviation of 30 %.

NOTE 2 For cross-border mail flows in categories 2 or 3 of Annex D, the above constraint need not be fulfilled annually but only per full multi-year result.

The weighting caps as defined above shall apply only to the following capping-relevant modes:

- all modes of the geographical characteristics;
- the two main modes of each discriminant characteristic, e.g. the modes with the highest real mail proportions (the highest Real mail Mode Weights RMW) and
- all other modes with a target proportion not smaller than 17,5 %.

7.3.2.3 Weighting caps for each individual item

For the aggregation of results, each test mail item is assigned its Individual Final Weight (IFW). The IFW depends on both, the proportion of valid mail and the target proportion of the stratum to which the test mail item belongs. It has the form of a corrective factor for the stratum, divided by the overall sample size.

The IFW for each valid test item shall not be smaller than 25 % or higher than 400 % of the value $1/\text{sample size}$.

EXAMPLE With a sample size of 100 test letters each weight shall be within the interval [0,25 % ; 4,00 %].

NOTE 1 Due to small sample sizes for countries in categories 2 or 3 of Annex D, for domestic measurement the above constraints are relaxed to an interval of 20 % up to 500 %.

NOTE 2 For cross-border mail flows in categories 2 or 3 of Annex D, the above constraint need not be fulfilled annually but only per full multi-year result.

The weighting caps as defined above shall only apply to strata completely characterised by main modes, which are relevant for capping themselves. These capping-relevant modes are defined in 7.3.2.2.

7.3.3 Design changes due to annual mail characteristic and postal flow changes

The sample design in reference to the geographical model and discriminant mail characteristics shall be reviewed at least every third year based on up-to-date information.

In case of a running system with stable mail flows information from:

- the most recent twelve months period available, if the real mail studies are conducted continuously, or
- the last real mail studies, if the real mail studies are conducted periodically,

shall be used at the time when the weighting basis for the calculation of the report is defined.

In case of the first measurement period or a running system with shifting mail flows, information from real mail studies conducted continuously during the measurement period (parallel run of real and test mail studies) shall be used at the time when the weighting basis for the calculation of the report is defined.

7.4 Content and timing

Reports on the transit time QoS for cross-border and domestic mail shall be provided at least once a year, based on the calendar year. The reports shall be issued not later than three months following the end of the test period. All reports shall include at least the following information:

- name and address of (i) the independent performance monitoring organisation and (ii) the monitored postal operators;
- field of study including (i) geography and (ii) mail types (categories of SPPM measured);
- period of the test mail measurement, stating whether based on items posted or delivered within the period;
- national as well as regional public holidays excluded from transit-time calculation as well as approved incidents of *force majeure* during the measurement period;
- type and details of the chosen design basis according to 6.3, including period(s) of the real mail measurement(s) (if applicable) on which the stratification and the sampling plan are based;
- information if all real mail measurements are based on SPPM only;
- total sample size including possible application of Annex D;
- methodology used, by reference to this standard including the chosen stratification;

- consistency of the
 - field of study, especially regarding geography;
 - methodology, especially regarding the design basis and the chosen stratification;
- results according to 5.3 including
 - the calculation method(s) of the transit time according to Annex B;
 - the weighting system according to 7.3.1.1;
- Assessment of the degree to which the weighting caps are respected.

The weighting caps are ideally fully respected. If not all caps are respected, the report shall show to what proportion the caps are respected and how strong possible violations of individual caps are.

- Accuracy of the results;

If a major change in the field of study or methodology has to take place within the measurement period, additional results shall be calculated for the periods before and after the structural change. The accuracy for these additional results shall be reported.

- failures, if any, to meet standard requirements;
- a reference for the *source* where the following information on the system is documented and may be obtained:
 - specific information on the sample design and the design basis including public data;
 - statistical assumptions and detailed calculation methods;
 - *audit certification*, this audit certification shall precise whether the results of the different checks prescribed in Annex C are in accordance with the requirements of this European Standard. Failures shall be reported in details.

If the audit is not concluded at the provision of the report, the report shall at least state: (i) the commissioned auditor, (ii) the status of the audit and (iii) the foreseen date of audit certification.

8 Quality control and auditing

Quality control procedures shall be applied through all phases and activities of the measurement system according to Annex C.

Validation of the test mail shall be applied through all phases and activities of the measurement system.

There shall be a detailed initial audit of the measurement system after the first measurement period. It shall be followed by independent re-audits at least every three years that keep track of any changes in the performance of the measurement system. More detailed independent audits are necessary after measurement periods when:

- the supplier changes,
- significant changes to the methodology of the measurement occur (for example any change in the set of mail characteristic that have been proved discriminant).

The audit cycle shall be agreed with the regulatory authority.

NOTE The regulatory authority might require a shorter audit cycle.

The auditor or auditors shall be fully independent from the tasks and parties, which are the topic of their audit(s). The auditor(s) must also have the proven skills and capacity to perform such an audit.

The audit shall be done according to this standard (see also Annex C).

9 The annexes

The following annexes will provide guidance for the implementation of SPPM end-to-end measurement systems that comply with this standard. Annex A to Annex D are mandatory; Annex E to Annex J are informative. Annex E to Annex H follow a process view on the implementation and maintenance processes from the planning stage up to the audit.

Annex E puts this standard in the context of other postal Quality of Service standards including details on the benefits of its application and its connection to quality improvement efforts.

Annex F deals with the planning stage of the measurement. It sums up all necessary considerations before implementing EN 13850. It starts with a summary of the responsibilities of all organisations involved in the measurement. Details are given on the design of the measurement system including scenarios with small mail volumes.

— Relaxations related to flows with small real mail volumes are set down in Annex D.

Annex F concludes with recommendations for selecting an independent performance monitoring organisation (contractor) based on the chosen design approach.

For the exact definition of the sample design, decisions have to be made on the structure for which the sample should provide representative results. This structure is the design basis. Annex G gives guidance on how to determine the set of discriminant mail characteristics for the field of study. It shows the choices one has to decide for an appropriate and reliable design basis including real mail studies results as well as alternative design bases for each discriminant mail characteristic. Annex G gives further insights into the connection between design basis and test mail sample design. It provides recommendations on how often the design basis should be updated.

After the sample design is fixed and the planning stage is completed, the test mail survey can go into its implementation phase. Important guidance for the implementation of the test mail measurement system is given in Annex H. It starts with a summary of the stages of the survey. Since the survey is panel based, details on all panellist-related issues are given, including the recruitment, postal characteristics and turnover of the panel. Data recorded by the panellists has to be validated. A list of possible validation issues is provided. Only valid data can be used for transit-time calculation.

— The transit-time calculation rules are defined in Annex B.

Since the sample is usually stratified, rules for the calculation of a weighted transit-time result are given.

— The calculation rule for the transit-time result and its accuracy can be found in Annex A. Further statistical background to these rules is laid down in Annex I.

All results are laid down in the report, which is given further attention in Annex H. After the report, the measurement system has to be audited as specified.

— All requirements related to quality control and auditing can be found in Annex C.

Annex H concludes by providing charts with possible implementation timetables and an overview of possible future developments.

Finally, Annex J gives an overview over the major changes in relation to the 2007 version of EN 13850.

Annex A (normative)

Accuracy calculation

A.1 Scope

A.1.1 General

The main scope of this annex is to provide an *accuracy calculation method* for the probability p for attaining the end-to-end transit time target.

The annex covers all mathematical detail necessary for an independent performance monitoring organisation to implement a comprehensive system for accuracy calculation.

The annex provides the general methodology for accuracy calculation as well as variants for special applications. Examples and explanations of the underlying methodology can be found in Annex I.

A.1.2 Two stage sampling approach

The sampling process is based on a two stage sampling approach.

In stage one all panellists are sampled from all postal users in the field of study according to the stratification. In stage two, the test letters are allocated to:

- the panellists, grouped by their discriminant user characteristics and to
- all discriminant mail characteristics that do not refer to the panellists themselves.

Each test letter is characterised by:

- the induction point a and the sender using it including his sender characteristics,
- the delivery point b and the receiver using it including his receiver characteristics and
- all other discriminant mail characteristics (see also 6.4.2).

A.1.3 Covariance / Stratification / Accuracy calculation

The calculation for the on-time probability is governed by three design principles:

- the incorporation of induction point / delivery point effects on the measured variation (co-variance),
- the incorporation of the stratification via a weighting system,
- the use of the appropriate probability distribution for the accuracy calculation.

A.3 presents a variance calculation method that deals with the covariances. This method will be imbedded into a weighted system in A.4. A.5 defines the final accuracy calculation method using the variances and including the optimal choice for the underlying probability distribution.

A.1.4 The design factor

In case of a simple random sample (SRS), it is straightforward to calculate the minimum sample size based on the expected or required accuracy of the design.

The incorporation of the design principles of A.1.3 leads to a more complex sample design that in most cases also has a higher degree of variation.

The *design factor* (df) is a measure for the added variation. It is defined as the ratio of the variance of the on-time estimator $\hat{p}$ in the given sample design by the variance of the on-time estimator $\hat{p}$ in a SRS of the same size. The design factor is always related to a given sample design and probability estimator.

EXAMPLE A design factor of the size 2 implies that the given design varies twice as much as a SRS.

The accuracy of any end-to-end sample design is not only improved by an increased number of items, but also by an increased number of senders and/or receivers. It is advisable to strive for a low number of test letters per induction / delivery point in order to reduce correlation effects.

A.2 Symbols

The following list is an overview of all mathematical symbols used in the annex together with a definition of each symbol:

N	Amount of real mail
n	Number of test letters
x_i	equals 1 if the test letter i is on time, 0 otherwise
x	$\sum_i x_i$ Number of test letters on-time
$\hat{p}$	x/n Estimate for the on-time probability
a	Index for a sender with at least 2 valid letters, for example in n_a
b	Index for a receiver with at least 2 valid letters, for example in x_b
ab	Index for the relation between sender a and receiver b
s	Index for a stratum with modes $j_1 \dots j_k$ of characteristics $1 \dots k$
W_{jc}	Real mail mode weight for mode j of characteristic c
W_s	Real mail stratum weight for stratum s
w_s	Corrective stratum weight for stratum s
$\text{Exp}[\hat{p}]$	Expected value of $\hat{p}$. $\text{Exp}[\hat{p}] = p$ if $\hat{p}$ is unbiased.
$\widehat{\text{Var}}[\hat{p}]$	Sample variance of $\hat{p}$
df	Design factor
ϵ	Half of the required accuracy as defined in clause 7.2.1

A.3 Variance calculation for one stratum

A.3.1 General calculation method

The following general variance calculation method shall be applied to all unstratified samples.

The variance of the on-time probability estimator $\hat{p}$ in the End-to-End (EtE) measurement system shall be calculated as:

$$\begin{aligned} \widehat{\text{Var}}_{\text{EtE}}[\hat{p}] = & \frac{1}{n^2} \left[\sum_a \left(x_a - \frac{n_a}{n} x \right)^2 + \sum_b \left(x_b - \frac{n_b}{n} x \right)^2 - \sum_a \sum_b \left(x_{ab} - \frac{n_{ab}}{n} x \right)^2 \right. \\ & + \sum_a \sum_{b, n_{ab} > 1} n_{ab} \frac{x_{ab}}{n_{ab} - 1} \left(1 - \frac{x_{ab}}{n_{ab}} \right) \\ & \left. + \sum_a \sum_{b, n_{ab} = 1} \left\{ \frac{1}{4n_a} \frac{x_a}{(n_a - 1)} \left(1 - \frac{x_a}{n_a} \right) + \frac{1}{4n_b} \frac{x_b}{(n_b - 1)} \left(1 - \frac{x_b}{n_b} \right) \right\} \right] \quad (\text{A.1}) \end{aligned}$$

It consists of two elements,

- the *Relation-to-Total Variation* accounting for the differences between the results of different induction / delivery points and
- the *Intra-Relation Variation* accounting for the differences between the test letters on one relation.

A.3.2 Relation-to-total variation

The relation-to-total variation is to be calculated in two terms,

One with an induction point of view
$$+ \frac{1}{n^2} \sum_a \left(x_a - \frac{n_a}{n} x \right)^2$$

The other with a delivery point of view
$$+ \frac{1}{n^2} \sum_b \left(x_b - \frac{n_b}{n} x \right)^2$$

Both terms have to be corrected by the term
$$- \frac{1}{n^2} \sum_a \sum_b \left(x_{ab} - \frac{n_{ab}}{n} x \right)^2$$

In other words, the *Relation-to-Total Variation* (RtT) can be split in three meaningful terms:

[R] The difference between the *relations'* mean-values and the general mean value

$$\sum_a \sum_b \left(\frac{n_{ab}}{n} \right)^2 (\hat{p}_{ab} - \hat{p})^2$$

[I] The corrected difference between the *induction points'* mean values and the general mean value

$$\sum_a \left(\frac{n_a}{n} \right)^2 (\hat{p}_a - \hat{p})^2 - \sum_a \sum_b \left(\frac{n_{ab}}{n} \right)^2 (\hat{p}_{ab} - \hat{p})^2$$

[D] The corrected difference between the *delivery points'* mean values and the general mean value

$$\sum_b \left(\frac{n_b}{n} \right)^2 (\hat{p}_b - \hat{p})^2 - \sum_a \sum_b \left(\frac{n_{ab}}{n} \right)^2 (\hat{p}_{ab} - \hat{p})^2$$

The partitioning of RtT into *RtT-I*, *RtT-R* and *RtT-D* gives valuable insights into the main contributors of the Relation-to-Total variation and therefore gives answers to the question what causes the RtT variation in the first place, the induction point *I*, the delivery point *D* or certain mail-flow relations *R*.

A.3.3 Intra-relation variation

The intra-relation variation (IRV) is calculated for each relation with more than one valid test letter in one term:

$$\widehat{\text{Var}}_{\text{IRV}} [\hat{p}_{ab}] := \left(\frac{n_{ab}}{n} \right)^2 \widehat{\text{Var}}_{\text{SRS}} [\hat{p}_{ab}] = \frac{1}{n^2} \frac{n_{ab}^2}{n_{ab} - 1} \hat{p}_{ab} (1 - \hat{p}_{ab}) = \frac{1}{n^2} \frac{n_{ab} x_{ab}}{n_{ab} - 1} \left(1 - \frac{x_{ab}}{n_{ab}} \right)$$

For induction point / delivery point relations with *only one valid test letter* the Intra-Relation variation cannot be calculated as defined above. For these one-letter relations the following approximation of the Intra-Relation variation shall be used: