BS EN 13850:2012

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

bsi.

...making excellence a habit.™

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 It supersedes BS EN 13850:2002 which is withdrawn.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.

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

Ref. No. EN 13850:2012: E

Contents

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 service[1].

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 $\pm\varepsilon$, 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 toAnnex 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 inducting 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 ± 17,5 % * 20 % = ± 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/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 $j_1 \cdots j_k$ of characteristics $1 \cdots k$
W_{j_c}	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
$\mathrm{Exp}[\hat{p}]$	Expected value of $\hat{p}$. $\mathrm{Exp}[\hat{p}] = p$ if $\hat{p}$ is unbiased.
$\widehat{\mathrm{Var}}\,[\hat{p}]$	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:

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

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:

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

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:

$$\widehat{\mathrm{Var}}_{\mathrm{IRV-OLR}}[x_{ab}] := \frac{1}{4}\left[\frac{1}{n_a^2} * \widehat{\mathrm{Var}}_{\mathrm{IRV}}[\hat{p}_a] + \frac{1}{n_b^2} * \widehat{\mathrm{Var}}_{\mathrm{IRV}}[\hat{p}_b]\right]$$
$$= \frac{1}{4n^2}\left[\frac{x_a}{n_a(n_a-1)}\left(1-\frac{x_a}{n_a}\right) + \frac{x_b}{n_b(n_b-1)}\left(1-\frac{x_b}{n_b}\right)\right]$$

A.4 Variance calculation for a stratified sample

A.4.1 Variance of a weighted sample design

The sample is stratified according to the set of discriminant mail characteristics. According to 7.3.1, the stratum-results have to be weighted to get an unbiased result. The weighted estimator $\hat{p}_{weighted}$ is defined as:

$$\hat{p}_{weighted} = \sum_s \frac{N_s}{N} * \hat{p}_s$$

where $\hat{p}_s$ is the estimate for the on-time probability in stratum s and N_s is the amount of real mail in stratum s. The variance of the *Stratified Random Sample (StrRS)* shall be calculated as:

$$\widehat{\mathrm{Var}}_{\mathrm{StrRS}}[\hat{p}_{weighted}] := \sum_s \left(\frac{N_s}{N}\right)^2 \widehat{\mathrm{Var}}_{\mathrm{SRS}}[\hat{p}_s] = \sum_s \left(\frac{N_s}{N}\right)^2 \frac{\hat{p}_s(1-\hat{p}_s)}{n_s - 1} \tag{A.2}$$

where n_s shall be greater than one for all strata s.

All strata with no or only one valid test letter may be omitted from the calculation and stated as such in the report.

Thus, in order to get weighted results, the mean value $\hat{p}_{weighted}$ is obtained by weighting with the *Real mail Strata Weights (RSW)* $\frac{N_s}{N}$, while the variance is obtained by weighting with the squared RSW $\left(\frac{N_s}{N}\right)^2$.

A.4.2 Final weight of the individual item

If k characteristics prove to be discriminant, each characteristic having $i_c\ (c = 1, \ldots, k)$ modes, the stratification $s = 1, \ldots, S$ has $S := i_1 * i_2 * \ldots * i_k$ strata.

Without weighting, the sample is drawn as a biased SRS yielding n valid mail items. Applying the stratification one can display the mean value $\hat{p}_{biased}$ as a weighted sum of strata means:

$$\hat{p}_{biased} := \frac{x}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i := \frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n_s} x_{s_i} = \sum_{s=1}^{S}\frac{n_s}{n}\left(\frac{1}{n_s}\sum_{i=1}^{n_s} x_{s_i}\right) = \sum_{s=1}^{S}\left(\frac{n_s}{n}\hat{p}_s\right)$$

The *weighted* estimator $\hat{p}_{weighted}$ can be expected to yield an unbiased estimate of the on-time probability p:

$$\mathrm{Exp}[\hat{p}_{weighted}] = \sum_{s=1}^{S}\underbrace{\frac{N_s}{N}\frac{n}{n_s}}_{=:w_s}\mathrm{Exp}\left[\frac{n_s}{n}\hat{p}_s\right] = \sum_{s=1}^{S}\frac{N_s}{N}\mathrm{Exp}[\hat{p}_s] = \sum_{s=1}^{S}\frac{N_s}{N}p_s = p$$

Using the *Corrective Strata Weights* w_s, the variance of the StrRS can also be calculated as:

$$\widehat{\mathrm{Var}}_{\mathrm{StrRS}}[\hat{p}_{weighted}] = \sum_s w_s^2\, \widehat{\mathrm{Var}}_{\mathrm{SRS}}\left[\frac{n_s}{n}\hat{p}_s\right] = \frac{1}{n^2}\sum_s w_s^2\,\frac{n_s\, x_s}{n_s - 1}\left(1-\frac{x_s}{n_s}\right)$$

Applying the corrective weighting procedure, each valid test letter x_i may be assigned its *Individual Final Weight (IFW)* w_i:

$$w_i := w_{s,i} := \frac{w_s}{n} = \frac{N_s}{N * n_s} = \frac{N_s}{N} * \frac{n}{n_s} * \frac{1}{n} \quad \text{with} \quad w_i = \frac{1}{n} \quad \text{if} \quad \frac{n_s}{n} = \frac{N_s}{N}$$

(A.3)

The IFW shall comply with the capping system as described in 7.3.2.

A.4.3 Weighting basis

The target proportions of the modes of all discriminant characteristics are usually measured through real mail studies. In the context of this standard, these proportions are called *Real mail Mode Weights (RMW)* regardless of the chosen design basis. The distribution of modes in the test mail sample shall be corrected according to the RMW.

In real mail studies the RMW are measured, but usually not the real mail strata weights. They partially have to be defined to enable the calculation of the measurement results. The total set of all defined strata weights $\frac{N_s}{N}$, *the weighting basis*, leads to a set of *Calculated Mode Weights (CMW)* that shall be equal the RMW for all modes.

The final definition of the weighting basis is necessary before calculating the report. It should be done in a way that avoids extreme IFW (see A.4.2). The method of optimising the weighting basis is no part of this standard.

The CMW are defined as follows: For each mode j_c of each characteristic c, there is the number N_{j_c} $(j_c = 1, \ldots, i_c)$ of all real mail items of mode j_c. The CMW of each mode j_c can be defined as sum of the weighting basis in the following way:

$$CMW_{\mathbf{j_c}} := \frac{N_{\mathbf{j_c}}}{N} := \frac{1}{N} \sum_{j_1=1}^{i_1} \sum_{j_2=1}^{i_2} \cdots \sum_{j_{c-1}=1}^{i_{c-1}} \sum_{j_{c+1}=1}^{i_{c+1}} \cdots \sum_{j_k=1}^{i_k} N_{j_1 j_2 \cdots j_{c-1} \mathbf{j_c} j_{c+1} \cdots j_k} \tag{A.4}$$

where each stratum is defined over its modes $j_1 \cdots j_k$.

A.4.4 Combination of weighting and covariance

The variance of a Simple Random Sample (SRS) is calculated as:

$$\widehat{\text{Var}}_{\text{SRS}}[\hat{p}] = \frac{\hat{p}(1-\hat{p})}{n-1} = \frac{1}{n^2} \frac{n}{n-1} x \left(1 - \frac{x}{n}\right) \tag{A.5}$$

The variance of a Stratified End-to-End measurement system (StrEtE) shall then be calculated as:

$$\widehat{\text{Var}}_{\text{StrEtE}}[\hat{p}_{weighted}] := \frac{\widehat{\text{Var}}_{\text{EtE}}[\hat{p}]\, \widehat{\text{Var}}_{\text{StrRS}}[\hat{p}_{weighted}]}{\widehat{\text{Var}}_{\text{SRS}}[\hat{p}]} \tag{A.6}$$

To further simplify the handling of the effects of covariance and weighting this standard works with the concept of design factors (A.1.4).

The design factors for the unstratified End-to-End measurement system and for the stratified random sample are defined as [see Formulae (A.1), (A.2) and (A.5)]:

$$\text{df}_{\text{EtE}} := \frac{\widehat{\text{Var}}_{\text{EtE}}[\hat{p}]}{\widehat{\text{Var}}_{\text{SRS}}[\hat{p}]} \quad \text{and} \quad \text{df}_{\text{StrRS}} := \frac{\widehat{\text{Var}}_{\text{StrRS}}[\hat{p}_{weighted}] * (n-1)}{\hat{p}_{weighted} * (1 - \hat{p}_{weighted})} \tag{A.7}$$

The design factor for a Stratified End-to-End measurement system shall be calculated as [see Formulae (A.5 to A.7)]:

$$\mathrm{df}_{\mathrm{StrEtE}} := \mathrm{df}_{\mathrm{EtE}} * \mathrm{df}_{\mathrm{StrRS}} = \frac{\widehat{\mathrm{Var}}_{\mathrm{EtE}}[\hat{p}] * \widehat{\mathrm{Var}}_{\mathrm{StrRS}}[\hat{p}_{weighted}] * (n-1)}{\widehat{\mathrm{Var}}_{\mathrm{SRS}}[\hat{p}] * \hat{p}_{weighted} * (1-\hat{p}_{weighted})} \tag{A.8}$$

NOTE The calculations in Formulae (A.6) and (A.8) are approximations, which work well as long as the weighting caps in 7.3.2 are respected. With extreme weighting systems that do not completely respect the weighting caps the approximations tend to be conservative, i.e. the approximated design factor tends to be larger than the true design factor.

A.5 Calculation of the confidence interval

A.5.1 General

Using the design factor derived in Formula (A.8) one now can calculate the accuracy of the on-time performance estimator $\hat{p}$. The accuracy is expressed as the width of the confidence interval around $\hat{p}$. A maximum length 2ϵ of this width is defined in 7.2.1 based on a confidence level of $(1-\alpha) = (100\% - 5\%) = 95\%$.

Any confidence interval is based on a probability distribution. Depending on the chosen probability distribution the interval is symmetrical $[\hat{p}-\epsilon\,;\hat{p}+\epsilon]$ or not $[\hat{p}-\epsilon_{lower}\,;\hat{p}+\epsilon_{upper}]$, $\epsilon_{lower} > \epsilon_{upper}$, if $\hat{p} > 50\%$.

NOTE If the confidence interval is not symmetrical the maximum length 2ϵ is defined as: $2\epsilon := \epsilon_{lower} + \epsilon_{upper}$

The correct probability distribution for modelling the on-time performance in a simple random sample would be the Binomial distribution. Confidence intervals for this distribution are not straightforward to calculate. For the purpose of this standard appropriate approximation of the Binomial distribution will be recommended.

A.5.2 Normal approximation

A.5.2.1 The Normal confidence interval

In most cases, the Normal distribution will be an appropriate approximation of the Binomial distribution. The confidence interval and the accuracy for $\hat{p}_{weighted}$ are defined as:

$$\left[\hat{p}_{weighted} \pm 1.95996 * \sqrt{\frac{\hat{p}_{weighted} * (1-\hat{p}_{weighted})}{n-1} * \mathrm{df}_{\mathrm{StrEtE}}}\right] \tag{A.9}$$

$$\text{with} \quad \mathbf{2}\epsilon = 3.91992 * \sqrt{\widehat{\mathrm{Var}}_{\mathrm{StrEtE}}[\hat{p}_{weighted}]}$$

This simple Normal confidence interval is symmetrical and easy to use. The minimum sample size can be calculated using ϵ as defined in 7.2.1:

$$n_{minSS} := \min n^* \quad \text{with} \quad n^* \geq \mathrm{df}_{\mathrm{StrEtE}} * \left[3.84145 * \frac{\hat{p}_{weighted}(1-\hat{p}_{weighted})}{\epsilon^2}\right] + 1 \tag{A.10}$$

In practice the design factor may be applied directly on the minimum sample size of a SRS to determine the minimum sample size of the design. It has to be kept in mind that not only the sample size has to grow by the design factor but also the panel size if the design-factor itself is not intended to change.

A.5.2.2 Applicability of the Normal confidence interval

The Normal approximation will be sufficient as long as the on-time performance does not approach 100 %. With a growing performance level, one denotes a growing bias of the approximation.

Confidence intervals may be computed using the asymptotic normality of a maximum likelihood estimator as long as sample sizes are very large and the estimator of the probability is not anywhere near 0 or 1. In fact the discreteness of the binomial distribution and performance levels over 85 % often make the normal approximation work poorly even with moderate sample sizes. The result is a confidence interval that is often “liberal”. This means when 95 % is stated as the confidence level, the true confidence level is often lower.

The maximum tolerable bias with respect to using the Normal distribution confidence interval shall not be greater than 4 % of the non-performance quota.

This restriction will be fulfilled when a minimum number of non-performance items are recorded. This minimum depends on the performance level. The following table shall be used to decide if the number of recorded non-performance items allows the use of the Normal distribution:

Table A.1 — Minimum number of non-performance items for the use of the Normal distribution

P Performance	*NP* Non-Performance	Minimum *NP*-Items	Maximum Bias
99,0 %	1,0 %	**52**	0,04 %
98,0 %	2,0 %	**51**	0,08 %
97,0 %	3,0 %	**50**	0,12 %
96,0 %	4,0 %	**49**	0,16 %
95,0 %	5,0 %	**48**	**0,20 %**
94,0 %	6,0 %	**47**	0,24 %
93,0 %	7,0 %	**46**	0,28 %
92,0 %	8,0 %	**45**	0,32 %
91,0 %	9,0 %	**44**	0,36 %
90,0 %	10,0 %	**43**	**0,40 %**
87,5 %	12,5 %	**42**	**0,50 %**
85,0 %	15,0 %	**40**	**0,60 %**
82,5 %	17,5 %	**37**	**0,70 %**
80,0 %	20,0 %	**35**	**0,80 %**
77,5 %	22,5 %	**32**	**0,90 %**
75,0 %	25,0 %	**30**	**1,00 %**

EXAMPLE 1 If you have a performance of 90 % and at least 43 Non-Performance Items, you may use the Normal distribution.

EXAMPLE 2 If you have only 42 Non-Performance Items, you have to switch to the confidence intervals inA.5.3 or A.5.4.

EXAMPLE 3 If your performance falls below 90 % with 42 NP-Items, the use of the Normal distribution is allowed again.

EXAMPLE 4 If you have a performance below 82,5 % and 36 NP-Items (not tabled), the use of the Normal distribution is allowed, too.

NOTE The applicability of the Normal distribution can be reviewed if the sample size is known and if the performance can be estimated. Example: 3 500 valid mail items with a performance of 85 % include 3 500*0,85=525 non-performance items.

In case the normal distribution cannot be used, other confidence interval calculation methods exist that produce asymmetrical confidence intervals with a negligible bias. These calculation methods are detailed inA.5.3-A.5.4.

A.5.3 Agresti-Coull approximation

The first improved calculation method is called Agresti-Coull interval[3]. It is also based on the Normal distribution and can be used for all sample sizes with at least $n \geq 40$ items. For a SRS the Agresti-Coull Interval is calculated as:

$$\left[\hat{p}_{ac} \ \pm \ 1.95996 * \sqrt{\frac{\hat{p}_{ac}(1-\hat{p}_{ac})}{n+3.84145}}\right] \quad \text{with} \quad \hat{p}_{ac} := \frac{x+1.92072}{n+3.84145}$$

To incorporate the design factor in this calculation method one works with the effective sample size (ESS). The effective sample size of a design is the sample size of a SRS with the same variance as the design, i.e. the sample size of the design divided by the design factor.

For a stratified End-to-End measurement system the Agresti-Coull interval for $\hat{p}_{weighted}$ shall be calculated as:

$$\left[\hat{p}_{acESS} \ \pm \ 1.95996 * \sqrt{\frac{\hat{p}_{acESS}\,(1-\hat{p}_{acESS})}{n_{ESS}+3.84145}}\right] \quad \text{with} \quad \hat{p}_{acESS} := \frac{x_{ESS}+1.92072}{n_{ESS}+3.84145} \qquad \text{(A.11)}$$

$$n_{ESS} := \max n^* \ \text{ with } \ n^* \leq \frac{n}{\mathrm{df}_{\mathrm{StrEtE}}} \quad \text{and} \quad x_{ESS} := \max x^* \ \text{ with } \ x^* \leq \ \hat{p}_{weighted} * n_{ESS}$$

$$\text{providing} \quad \mathbf{2}\epsilon = 3.91992 * \sqrt{\frac{\hat{p}_{acESS}\,(1-\hat{p}_{acESS})}{n_{ESS}+3.84145}}$$

$$\text{and} \quad \epsilon_{lower} := (\hat{p}_{weighted} - \hat{p}_{acESS}) + 1.95996 * \sqrt{\frac{\hat{p}_{acESS}\,(1-\hat{p}_{acESS})}{n_{ESS}+3.84145}}$$

$$\epsilon_{upper} := (\hat{p}_{acESS} - \hat{p}_{weighted}) + 1.95996 * \sqrt{\frac{\hat{p}_{acESS}\,(1-\hat{p}_{acESS})}{n_{ESS}+3.84145}}$$

This adapted Normal distribution confidence interval is asymmetrical and can be used without sophisticated statistical software.

3 See *Agresti, A. and Coull, B. A.* (1998). “Approximate is better than ‘exact’ for interval estimation of binomial proportions”. The American Statistician 52(2), pp. 119-126.

A.5.4 Inverse Beta approximation

The second improved calculation method uses the *Inverse Beta Function [BetaInv].* It is based on the Beta distribution, which is the continuous form of the Binomial distribution. The Inverse Beta Function is implemented in most software packages for data analysis. The rule directly calculates the confidence interval without calculating the variance first. It also uses the effective sample size.

For a SRS the Inverse Beta interval can be calculated as[4]:

$$[\ \mathrm{BetaInv}\,[0.025\,;\,x\,;\,n-x+1]\ \ ;\ \ \mathrm{BetaInv}\,[0.975\,;\,x+1\,;\,n-x]\,]$$

For a stratified End-to-End measurement system the Inverse Beta interval for $\hat{p}_{weighted}$ shall be calculated as:

$$\Big[\ \mathrm{BetaInv}\,[0.025\,;\,x_{ESS}\,;\,n_{ESS}-x_{ESS}+1]\ \ ;\ \ \mathrm{BetaInv}\,[0.975\,;\,x_{ESS}+1\,;\,n_{ESS}-x_{ESS}]\,\Big] \qquad \text{(A.12)}$$

$$n_{ESS} := \max n^* \ \text{ with } \ n^* \le \frac{n}{\mathrm{df}_{\mathrm{StrEtE}}} \quad \text{and} \quad x_{ESS} := \max x^* \ \text{ with } \ x^* \le \hat{p}_{weighted} * n_{ESS}$$

$$\mathbf{2}\epsilon = \mathrm{BetaInv}\,[0.975\,;\,x_{ESS}+1\,;\,n_{ESS}-x_{ESS}] - \mathrm{BetaInv}\,[0.025\,;\,x_{ESS}\,;\,n_{ESS}-x_{ESS}+1]$$

$$\text{and} \quad \epsilon_{lower} := \hat{p}_{weighted} - \mathrm{BetaInv}\,[0.025\,;\,x_{ESS}\,;\,n_{ESS}-x_{ESS}+1]$$

$$\epsilon_{upper} := \mathrm{BetaInv}\,[0.975\,;\,x_{ESS}+1\,;\,n_{ESS}-x_{ESS}] - \hat{p}_{weighted}$$

This Beta distribution confidence interval is asymmetrical and easy to use. It requires special software, which provides the Inverse Beta Function.

[4] See *Blyth, Colin R.* (1986). “Approximate Binomial Confidence Limits”. Journal of the American Statistical Association Vol. 81, No. 395, Theory and Methods, pp. 843-850.

Annex B
(normative)

Transit Time Calculation Rule

B.1 Working week transit time calculation rule / domestic and cross-border mail

The results for domestic and cross-border mail shall be based on a working-week calculation rule that excludes all weekly non-working days and all national public holidays in the country of delivery after the date of posting.

NOTE For some fields of study the working week is a 5-day working week beginning with *Day1 = Monday* and ending with *Day 5 = Friday*. In this case, *Saturday* and *Sunday* are the weekly non-working days.

Published Regional holidays may be treated as national public holidays and subtracted in the calculation of the end-to-end transit time, too.

The working week may be defined according to its days of operation.

If for cross border mail the posting is done in the country of origin either on:

— a weekly non-working day or national public holiday or,

— one day before weekly non-working day(s) or national public holiday(s),

these non-working day(s) or national public holiday(s) have to be excluded.

The following tables offer a number of examples of methods for calculating the transit time, depending on the actual service offered at weekends and on whether delivery takes place five or six days of the week. The chosen calculation rule shall reflect the existing logistic processes in the field of study. It shall be stated in the report.

Additional cases to those presented in this annex may exist, which require the examples to be properly adapted.

Any calculation rule other than those seven rules presented in Annex B shall be defined before applying the standard. This definition shall be done in accordance with the regulatory authority.

The chosen calculation rule shall be applied consistently on all data from the measurement period.

B.2 Calculation rules

B.2.1 Rule 1: Collection Monday-Friday / Delivery Monday-Friday

Table B.1 shows a calculation rule applied for mail posted at least on each collection day of the week. Collection and delivery are from day one up to day five.

EXAMPLE Day one up to day five may be Monday to Friday, day six and day seven Saturday and Sunday. Collection is then from Monday to Friday; delivery is also from Monday to Friday.

Table B.1 — Collection Monday-Friday / Delivery Monday-Friday

Working Week 1					Non-Working		Working Week 2				
Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 1	Day 2	Day 3	Day 4	Day 5
J	J + 1	J + 2	J + 3	J + 4			J + 5	J + 6	J + 7	J + 8	J + 9
	J	J + 1	J + 2	J + 3			J + 4	J + 5	J + 6	J + 7	J + 8
		J	J + 1	J + 2			J + 3	J + 4	J + 5	J + 6	J + 7
			J	J + 1			J + 2	J + 3	J + 4	J + 5	J + 6
				J			J + 1	J + 2	J + 3	J + 4	J + 5
					J		J + 1	J + 1	J + 2	J + 3	J + 4
						J	J + 1	J + 1	J + 2	J + 3	J + 4

NOTE 1 J = Day of induction.

NOTE 2 Mail posted on non-working days are treated as mail posted on the next working day.

B.2.2 Rule 2: Collection Monday-Friday / Delivery Tuesday-Saturday

Table B.2 shows a calculation rule applied for mail posted at least on each collection day of the week. Collection is from day one up to day five; delivery is from day two up to day six (five days, too).

EXAMPLE Day one up to day five may be Monday to Friday, day six and day seven Saturday and Sunday. Collection is then from Monday to Friday, delivery is from Tuesday to Saturday.

Table B.2: Collection Monday-Friday / Delivery Tuesday-Saturday

Working Week 1					Non-Working		Working Week 2				
Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 1	Day 2	Day 3	Day 4	Day 5
J	J + 1	J + 2	J + 3	J + 4	J + 5			J + 6	J + 7	J + 8	J + 9
	J	J + 1	J + 2	J + 3	J + 4			J + 5	J + 6	J + 7	J + 8
		J	J + 1	J + 2	J + 3			J + 4	J + 5	J + 6	J + 7
			J	J + 1	J + 2			J + 3	J + 4	J + 5	J + 6
				J	J + 1			J + 2	J + 3	J + 4	J + 5
					J			J + 1	J + 2	J + 3	J + 4
						J		J + 1	J + 2	J + 3	J + 4

NOTE 1 J = Day of induction.

NOTE 2 Mail posted on non-working days are treated as mail posted on the next working day.

B.2.3 Rule 3: Collection Monday-Friday / Delivery Monday-Saturday

Table B.3 shows a calculation rule applied for mail posted at least on each collection day of the week. Collection is from day one up to day five, delivery from day one up to day six.

EXAMPLE Day one up to day five may be Monday to Friday, day six Saturday and day seven Sunday. Collection is then from Monday to Friday, delivery is from Monday to Saturday.

NOTE 1 Rule 3 corresponds to the official transit-time calculation rule in EN 13850:2002, B.1 "*Five-day working-week calculation rule / domestic and cross-border mail*".

Table B.3 — Collection Monday-Friday / Delivery Monday-Saturday

Working Week 1					Non-Working		Working Week 2				
Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 1	Day 2	Day 3	Day 4	Day 5
J	J + 1	J + 2	J + 3	J + 4	J+5		J + 5	J + 6	J + 7	J + 8	J + 9
	J	J + 1	J + 2	J + 3	J+4		J + 4	J + 5	J + 6	J + 7	J + 8
		J	J + 1	J + 2	J+3		J + 3	J + 4	J + 5	J + 6	J + 7
			J	J + 1	J+2		J + 2	J + 3	J + 4	J + 5	J + 6
				J	J+1		J + 1	J + 2	J + 3	J + 4	J + 5
					J		J + 1	J + 1	J + 2	J + 3	J + 4
						J	J + 1	J + 1	J + 2	J + 3	J + 4

NOTE 2 J = Day of induction.

NOTE 3 Mail posted on non-working days are treated as mail posted on the next working day.

B.2.4 Rule 4: Collection Monday-Saturday / Delivery Monday-Friday

Table B.4 shows the calculation rule applied for mail collected on each day of the week excluding Sunday (six days) and posted at least on each collection day. Mail is delivered from Monday through Friday (five days).

Table B.4 — Collection Monday-Saturday / Delivery Monday-Friday

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
J	J + 1	J + 2	J + 3	J + 4			J + 5	J + 6	J + 7	J + 8
	J	J + 1	J + 2	J + 3			J + 4	J + 5	J + 6	J + 7
		J	J + 1	J + 2			J + 3	J + 4	J + 5	J + 6
			J	J + 1			J + 2	J + 3	J + 4	J + 5
				J			J + 1	J + 2	J + 3	J + 4
					J		J + 1	J + 2	J + 3	J + 4
						J	J + 1	J + 1	J + 2	J + 3

NOTE 1 J = Day of induction.

NOTE 2 Mail posted on non-working days are treated as mail posted on the next working day.

B.2.5 Rule 5: Collection Sunday-Friday / Delivery Monday-Friday

Table B.5 shows the calculation rule applied for mail collected on each day of the week excluding Saturday (six days) and posted at least on each collection day. Mail is delivered from Monday through Friday (five days).

Table B.5 — Collection Sunday-Friday / Delivery Monday-Friday

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
J	J + 1	J + 2	J + 3	J + 4			J + 5	J + 6	J + 7	J + 8
	J	J + 1	J + 2	J + 3			J + 4	J + 5	J + 6	J + 7
		J	J + 1	J + 2			J + 3	J + 4	J + 5	J + 6
			J	J + 1			J + 2	J + 3	J + 4	J + 5
				J			J + 1	J + 2	J + 3	J + 4
					J		J + 1	J + 2	J + 3	J + 4
						J	J + 1	J + 2	J + 3	J + 4

NOTE 1 J = Day of induction.

NOTE 2 Mail posted on non-working days are treated as mail posted on the next working day.

B.2.6 Rule 6: Collection Monday-Saturday / Delivery Monday-Saturday

Table B.6 shows the calculation rule applied for mail collected and delivered on each day of the week excluding Sunday (six days) and posted at least on each collection day.

Table B.6 — Collection Monday-Saturday / Delivery Monday-Saturday

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
J	J + 1	J + 2	J + 3	J + 4	J+5		J + 6	J + 7	J + 8	J + 9
	J	J + 1	J + 2	J + 3	J+4		J + 5	J + 6	J + 7	J + 8
		J	J + 1	J + 2	J+3		J + 4	J + 5	J + 6	J + 7
			J	J + 1	J+2		J + 3	J + 4	J + 5	J + 6
				J	J+1		J + 2	J + 3	J + 4	J + 5
					J		J + 1	J + 2	J + 3	J + 4
						J	J + 1	J + 1	J + 2	J + 3

NOTE 1 J = Day of induction.

NOTE 2 Mail posted on non-working days are treated as mail posted on the next working day.

B.2.7 Rule 7: Collection Sunday-Friday / Delivery Monday-Saturday

Table B.7 shows the calculation rule applied for mail collected on each day of the week excluding Saturday (six days) and posted at least on each collection day. Mail is delivered from Monday through Saturday (six days, too).

Table B.7 — Collection Sunday-Friday / Delivery Monday-Saturday

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
J	J + 1	J + 2	J + 3	J + 4	J+5		J + 6	J + 7	J + 8	J + 9
	J	J + 1	J + 2	J + 3	J+4		J + 5	J + 6	J + 7	J + 8
		J	J + 1	J + 2	J+3		J + 4	J + 5	J + 6	J + 7
			J	J + 1	J+2		J + 3	J + 4	J + 5	J + 6
				J	J+1		J + 2	J + 3	J + 4	J + 5
					J		J + 1	J + 2	J + 3	J + 4
						J	J + 1	J + 2	J + 3	J + 4

NOTE 1 J = Day of induction.

NOTE 2 Mail posted on non-working days are treated as mail posted on the next working day.

Annex C
(normative)

Quality control and auditing

C.1 Quality Control

C.1.1 Statistical design

Quality assurance & control activities shall include:

- documentation of the sample design and the design basis;
- development and maintenance of a master plan for update and systematic review of the design;
- control of the statistical design to ensure that it reflects the target proportions for both, the discriminant mail characteristics and geographical stratification;
- formalised reviews of the mail production and induction plans in terms of structure and volume in order to ensure the stability of discriminant parameters;
- control of panel requirements, including the number of panellists, structure and geographical spread of the panel and the maximum panel workload;
- reporting on the panel size, structure and geographical spread.

C.1.2 Test item production

Quality assurance & control activities shall include:

- development and maintenance of a specific production plan;
- adaptation of the production plan to changed target proportions;
- consistency of the plan production with the sample requirements;
- during every production batch selection and inspection of a predefined random sample of produced test items to ensure compliance with specifications. The result of these inspections shall be documented. The sample shall be as a minimum 3 % of the produced test items. The samples shall be selected based on specified Acceptable Quality Level's (AQL) according to EN ISO 19011;
- review the production procedures, quality control and evaluation criteria used during test item production.

C.1.3 Provision of test items to the sender panellists

Quality assurance & control activities shall include:

- development and maintenance of a provision schedule for providing the sender with test letters for the next mail allocation period;
- documentation that the test items have been provided according to schedule.

C.1.4 Sending test items

Quality assurance & control activities shall include:

- continuous review of each sender's performance individually and on an aggregated basis according to predefined rules;
- provision of a specific instruction manual to every sender detailing all tasks to be conducted;
- review of the training provided to the senders;
- review of the induction reporting documents;
- reporting of the result of the individual and aggregated controls.

C.1.5 Receiving test items

Quality assurance & control activities shall include:

- continuous review of each receivers performance individually and on an aggregated basis according to predefined rules;
- provision of an instruction manual to every receiver detailing the tasks to be undertaken;
- review of the training provided to receivers;
- review of the receipt reporting documents;
- reporting of the result of the individual and aggregated controls.

C.1.6 Data collection

Quality assurance & control activities shall include:

- control of reply rates from every participant;
- comparison of reply rates with production targets;
- continuous checks on the quality of the data capture to ensure that test item data have been entered correctly.

C.1.7 Data analysis and reporting

Quality assurance & control activities shall include:

- analysis of data for consistency on ad-hoc basis;
- analysis of non-conforming information to ensure that corrective actions are implemented;
- maintenance of a stringent data analysis procedure;
- control of the weights resulting from the weighting procedure for the discriminant mail characteristics.

C.1.8 Archiving

Quality assurance & control activities shall include:

- storage of the test item information under suitable conditions (physically and/or electronically) in an orderly and retrievable manner for an adequate time at least until the report and the auditing are completed.

C.1.9 Quality control and Information Technology (IT)

Quality assurance & control activities and IT system requirements shall follow the rules laid down in:

- EN ISO 9001;
- EN ISO 9004;
- ISO 10005;
- ISO 10007;

C.2 Auditing – general remarks

According to Clause 8 the measurement system shall be audited every three years.

Exceptions from this rule are years when:

- the supplier changes,
- changes to the methodology of the measurement occur.

In these cases, the audit has to be done in the year after the changes have occurred.

It may also be that the regulatory authority requires additional audits.

Audit shall be implemented by an independent body approved by the regulatory authority in order to check the design, set-up and maintenance of the QoS measurement system, including an evaluation if:

- the requirements of this European Standard are fulfilled;
- changes in the sample design (e.g.: design basis, geographical stratification, discriminant mail characteristics) that prove to be more efficient for a given operator, remain coherent with the requirements in this European Standard.

The audit shall verify that the body in charge of the measurement system is independent from the postal operators.

C.3 Audit of the design basis

C.3.1 General

As a minimum, the audit of the design basis shall cover the:

- the type and structure of the design basis;
- the geographical stratification;
- the discriminant mail characteristics.

Detailed and complete information on the reasoning, data collection methodologies, accuracy and results shall be provided to the auditor.

It is acceptable that the audit of the design basis and the audit of the test mail measurement system are performed separately by different auditors.

EXAMPLE The standard design basis has to be audited. There is already a regular audit in place for existing real mail studies which have a scope beyond the requirements of EN 13850. The regular audit may then be extended to the audit-requirements of this standard to certify the appropriateness of the applied design basis. In a second step the test mail measurement system may be audited by the same or a different auditor independently of the RMS audit.

C.3.2 Methodological audit

- definition of the required structural information;
- chosen parameter values based on best practice and experience;
- external and internal data sources used;
- results from measurement systems (e.g. real mail studies) set up to provide structural information, including their sample design and measurement accuracy.

C.3.3 Results

- reasoning for the choice of and possible changes in parameter values
- reasoning for structural changes in historical series of results (market changes, changes in logistics etc).

C.4 Audit of the Quality-of-Service measurement system

C.4.1 Panel audit

As a minimum, the audit of the QoS measurement system shall check that:

- the panel represents all types of induction and all types of delivery;
- the panellists are randomly distributed within each stratum;
- the required number of panellists and the required proportion of postal areas are covered;
- the required panel turnover is respected whenever possible;
- the panellists' workload is appropriate according to the guidelines of 6.6 and that it is related to the postal density in the field of study.
- the panel performance is monitored. The monitoring is based – among other criteria – on historical series of QoS results per panellists.

The replacement of a panellist should not be influenced by the postal operator's on-time performance of that panellist.

C.4.2 Stability of the parameters

For the stratification, check that the stability of the mail allocation process is insured throughout the measurement period.

C.4.3 Instructions given to the panellists

Some requirements for this European Standard rely on the instructions given to the panellists, especially for time of posting. This shall be checked by direct and anonymous interviews of panellists.

C.4.4 General Audit of the system

All quality control procedures in C.1 shall be audited.

Annex D
(normative)

Relaxation related to flows with small real mail volumes

D.1 General

D.1.1 Scope

This annex gives information on how to categorise mail flows for measuring purposes and explains how the *Minimum Sample Size* (MSS) for small and medium sized real mail flows may be obtained by measuring under a consecutive number of years.

The following sections describe under which conditions and how relaxation is possible. Two fundamental aspects to be relaxed under certain conditions are:

— *Measurement period* (5.1): “The measuring system shall provide one annual figure for each relevant field of study”.

— *Minimum Sample Size:* Domestic (6.2.1) and Cross-Border (6.2.2)

D.1.2 Measurement period

Only for cross-border fields of study it is allowed to relax the requirement of one annual figure. When the real mail volume in a given cross-border field of study is smaller than a certain limit or small compared to other fields of study then the MSS may be reached after a longer measurement period than one year.

The results for cross-border flows reaching full accuracy after 2 or 3 years should not be reported until before these years have been completed. After that, the results should continue to be reported annually on a year-rolling basis.

EXAMPLE A certain small country-to-country flow is measured annually with 1/3 of the MSS. It will reach the MSS after three years of measurement:

End of year 1: No reporting of results.
End of year 2: No reporting of results.
End of year 3: Reporting of results based on accumulated period from year 1 to year 3.
End of year 4: Reporting of results based on accumulated period from year 2 to year 4.
End of year 5: Reporting of results based on accumulated period from year 3 to year 5.
And so on…

D.1.3 Minimum Sample Size (MSS)

The prescribed MSS calculation requirements in section 6.2 are based on the knowledge and experience of QoS monitoring in many countries over many years. Therefore, the aim should be to reach this MSS for any domestic or cross-border transit time measurement. Despite this, under certain conditions situations may exist, in which it is advisable to relax the MSS requirements. These conditions are:

— the real mail volume in a given field of study is smaller than a certain limit and

— the transit-time performance for a given field of study is below a certain limit.

Only if both of these conditions hold at the same time it is allowable to measure the transit time with a lower MSS. Often the second condition will only be fulfilled temporarily until there is improvement in the transit time results. In order to meet the above conditions 1 and 2, the models for domestic and cross-border measurement include a “bounded sample size” concept, which implies a sample size relaxation.

D.2 Domestic mail flows

In some of the smaller EU countries, the annual domestic mail volumes are very low. In the following approach countries are categorised into four size bands and the bounded sample size concept is applied:

— **Category 1: Large size mail volumes**

Definition: Total annual real mail volume above 500 million mail pieces

Measurement: Application of 6.2.1, i.e. the MSS requirement

— **Category 2: Medium size mail volumes**

— Definition: Total annual real mail volume between 200 million and 500 million mail pieces.

Measurement: Application of 6.2.1 for countries above J+n performance of 85 %, fixed sample size of 4 950 for countries below J+n performance of 85 % according to the bounded sample size concept.

— **Category 3: Small size mail volumes**

— Definition: Total annual real mail volume between 1,5 million and 200 million mail pieces.

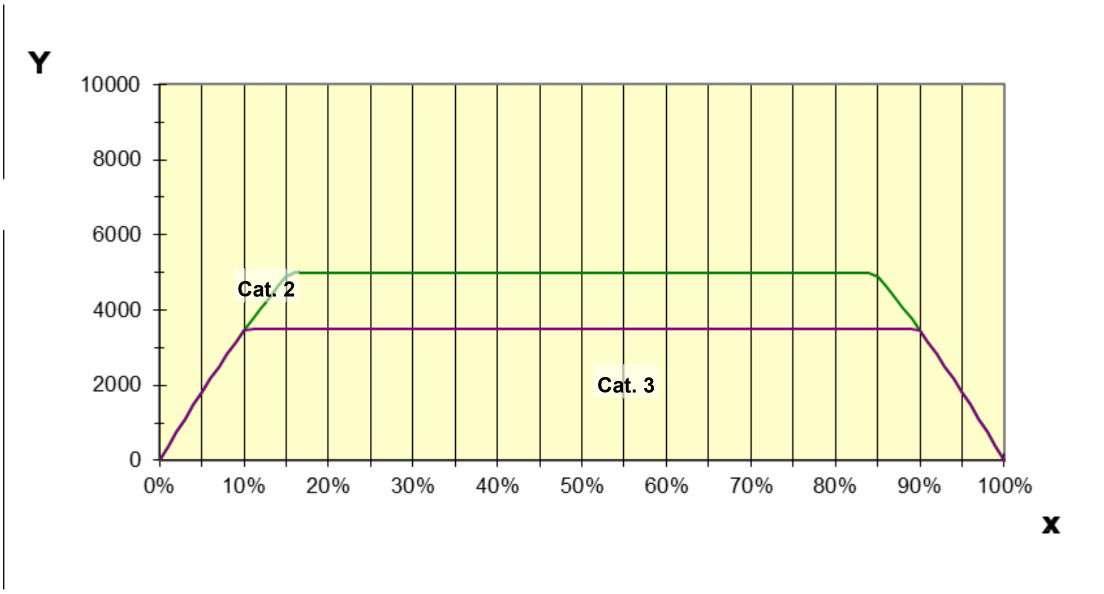
Measurement: Application of 6.2.1 for countries above J+n performance of 90 %, fixed sample size of 3 500 for countries below J+n performance of 90 % according to the bounded sample size concept.

— **Category 4: Very small size flows**

— Definition: All flows with volumes below 1,5 million mail pieces per year.

Measurement: For each of these domestic measurement systems the test mail would increase the real mail volume in the total field of study by more than 0,25 %. Since the measurement itself would add considerable amounts of mail to the logistic system, the practical value and neutral position of such a measurement is questionable. In category 4 cases it is recommended to broaden the field of study or alternatively to include further operators.

Minimum effective domestic sample sizes (categories 2 and 3)



Key

X performance J+n

Y domestic (± 1 %)

Figure D.1 — Collection Sunday-Friday / Delivery Monday-Saturday

The relation between performance level and minimum sample size is illustrated in Figure D.1. The upper bound for countries in category 2 is at 4.900. For countries in category 3, it is at 3.500. For countries with on-time performances above 85 % or 90 %, the minimal sample sizes are lower.

D.3 Cross-border mail flows

The MSS calculation requirement of 6.2.2 may not be reasonable for all cross border mail flows. An alternative way to reach this accuracy level is measuring over a number of consecutive years. It is recommended that in these circumstances the test period should be extended to 2 or 3 years. The accuracy requirements may be relaxed for flows with small volume and poor performance.

After all cross border mail flows have been sorted by real mail volume for each outbound country, each of them falls into one of the following four different categories:

— **Category 1: Large size flows**

Definition: At least 5 largest (real mail volume) outbound flows per country and
All other outbound flows with annual real mail volumes of more than 1,45 Mio items plus
At least representing 65 % of total outbound EU volume plus
At least the 3 largest inbound flows per country.

Measurement: Application of 6.2.2, i.e. the MSS requirement

— **Category 2: Medium size flows**

Definition: At least representing 10 largest outbound flows together with category 1 and
All other outbound flows with annual real mail volumes of more than 580.000 items plus
At least representing in total, with Cat.1 flows, 80 % of total outbound EU volume.

Measurement: Per 2-year period Application of 6.2.2 for flows above J+n performance of 80 %, fixed sample size of 290 for flows below J+n performance of 80 % according to the bounded sample size concept.

— **Category 3: Small size flows**

Definition All remaining flows up to 100 % of total EU volume with exception of category 4.

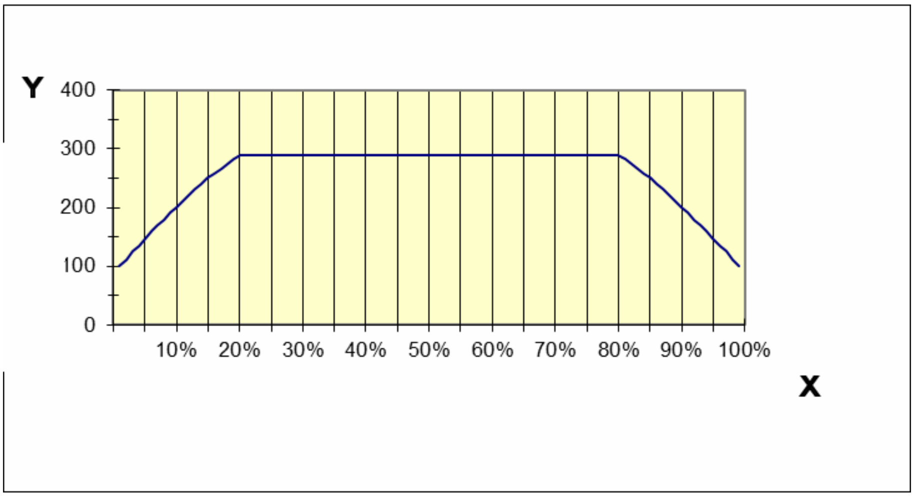
Measurement: Per 3-year period application of 6.2.2 for flows above J+n performance of 80 %, fixed sample size of 290 for flows below J+n performance of 80 % according to the bounded sample size concept.

— **Category 4: Very small size flows**

Definition All flows with volumes of below 11.500 mail pieces per year.

Measurement On each of these flows, the test mail would increase the real mail volume more than 2,5 %. Since the practical value and neutral position of such a measurement is rather questionable, these flows may be excluded from the measurement.

For all flows in categories 2 and 3, the MSS will only be reached after 2 or 3 years, respectively. The bounded sample size for flows in categories 2 and 3 is illustrated in Figure D.2.



Key

X performance J+n
Y cross-border (± 5%)

Figure D.2 — Minimum effective cross-border sample sizes (categories 2 & 3)

For flows with J+n performance levels from 80 % to 100 % there is no change to the original approach applied in category 1. For all other small and medium flows with J+n performance below 80 % (and above 20 %), the required effective sample size is limited by the upper bound at 290 items.

Annex E
(informative)

Purpose of postal Quality of Service standards

E.1 General

EN 13850 has been developed in order to:

— Define the minimum requirements that allow postal operators to comply with the requirements of the Directive 97/67/EC and its amendments.

— Provide a standardised method that shall be used throughout the European Union for measuring the transit time of end-to-end SPPM services.

For any minimum service level required by the European Commission or by national regulations for operators in the Universal Service, requirements can be made that the service has to be measured according to EN 13850 and its amendments.

To satisfy the minimum requirements, the various sections in the standard are divided in normative and informative sections. The normative sections have to be applied in all measurements to comply with the standard EN 13850. In the normative sections, the use of the word 'shall' refers to a specific normative requirement.

The informative sections are present to assist all stakeholders in the understanding, interpretation and practical use of the standard. In addition, they highlight areas where it might be useful to exceed the said standard.

E.2 Benefits of QoS standards

The main benefit of standardised QoS measuring methods is that they create a level playing field for all stakeholders, whilst complying with the Directive 97/67/EC and its amendments. More specifically, the following benefits can be achieved by applying EN 13850:

— Postal operators will have an absolute view on how measurements have to take place in order to satisfy the EU regulations. This means the measurements can be pre-specified and can therefore be more cost-efficient.

— Regulatory authorities can refer to and require measurements to comply with this standard. For any minimum service levels required by the European Commission or by national regulation for operators in the Universal Service, requirements can be made that the service be measured according to EN 13850. This makes it easier for the regulatory authority to control and check the measurement processes in place in a country.

— Using standardised measuring methods 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. This also means that measurements across years within a country can be compared, which Postal Users can use to form a long-term opinion of the services they receive.

— Standards are an efficient way to set up a measurement system. They could also make it easier to find independent performance monitoring organisations familiar with the methods and to evaluate the work of performance monitoring organisations for measuring and auditing. The standardised measuring methods may also eliminate debates and uncertainties about results from varying measuring methods.

— The use of the standard will ensure that reliable and correct information is collected, which can be presented to regulatory authority and the public in an understandable way.

E.3 Use of the survey results for quality improvement

E.3.1 Detailed analysis

EN 13850 sets out minimum requirements in order to produce one overall figure using a panel based measurement system. It is recognised that operators or other bodies may wish to go beyond these minimum requirements in order to produce information to identify and correct specific areas of poor performance.

Although the one overall figure can be used for quality improvements, postal operators may want to include different objectives into the design of the survey. The most common design aspects to consider outside the single annual figure are:

— sub-results for operational units or for periods of the year,

— fast provision of localised data for direct quality adjustments.

The survey design can be expanded to collect this information. As with the main design, real mail studies should form the basis of the expanded design.

E.3.2 Other / broader concepts

E.3.2.1 General

EN 13850 provides minimum requirements and allows measurement systems to be enhanced. It should be noted that other methodologies are available to find problematic areas and feed the need for quality improvement and that the standard may not be the only method to use for this purpose.

E.3.2.2 Panellist methodologies

Although a panel of business and private individuals is described as the preferred method of measurement in the standard, in certain cases other methodologies might give better results.

E.3.2.3 Technical registrations

— EN 13850 allows the use of test items containing electronic or other advanced technology so that each test item can be monitored at pre-defined points or throughout its whole journey to allow for more detailed diagnostic analysis of the transit time.

EXAMPLE 1 Many postal operators use Radio Frequency Identification (RFID) transponders in each test letter and have set up receiver systems at crucial points in their logistic network to register the moments at which the test letters pass these logistic nodes.

If these are used, the organisation operating the measuring system shall take steps to ensure that the diagnostic system:

— does not introduce biases in the end-to-end transit-time results,

— cannot be identified at any stage of the distribution and

— does not interfere with automated distribution processes.

In order to remove any panellist-induced bias from the survey, technical registration of the actual sent and receipt dates can be considered. With the help of electronic devices, the actual times of posting and actual times of receipt at the panellist's address can be captured, therefore taking away any ambiguity around the panellist's ability to register the correct dates and times.

EXAMPLE 2 Several postal operators use Radio Frequency Identification (RFID) receiver systems in each panellist's letterbox to register the moment the test letters arrives at its designated delivery point.

Annex F
(informative)

Considerations before implementing EN 13850

F.1 Limitations of EN 13850

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. 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, including Directive 2008/6/EC of 20th February 2008.

EN 13850 covers mail services with a defined transit-time service level which is defined in terms of a J+n transit-time target. It does not include mail services with a defined delivery day or a defined delivery-time service-level like, for example, ‘Friday delivery’ or ‘delivery up to 9 am’.

Transit times resulting from this measurement do not measure the postal operators' overall performance in a way that provides direct comparison of postal service operators.

Since measurements can differ widely between countries, both in terms of accuracy achieved as well as additional rules agreed with the regulatory authority, straightforward comparisons of the results are not valid. In line with this, different fields of study cannot be compared either (i.e. domestic versus cross-border performance in a particular country).

If results from different study domains are put together in any form of compilation, the results should be accompanied by additional information about key characteristics of the respective study domains. By doing so, it should be secured that indicators are given why the results may differ and why the compilation cannot serve as a straightforward comparison.

In addition, the QoS measurement under this standard does not consider other aspects of service performance, which might be relevant to customers and users, such as the time of collection or delivery or the days on which deliveries are made. Other types of research like customer satisfaction research can handle different aspects of the service experience customers have.

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. It is also not applicable for measuring the End-to-End transit times of bulk mailers' services and hybrid mail, which require different measurement systems and methodologies. If an operator is accepting SPPM only under special conditions of posting, it may be more appropriate to use EN 14534 for bulk mail.

F.2 Responsibilities

F.2.1 General

The implementation of this standard should be carried out according to a framework designed and decided upon by the European Union as laid down in the Directive 97/67/EC and its amendments, including Directive 2008/6/EC and the Third mandate for Postal Services – M428:2008.

At the national level the main stakeholders involved in this implementation are the regulatory authority, the operator, the contractor (i.e. performance monitoring organisation) and the auditing party whose actions contribute to implement this standard correctly and to guarantee the accuracy and reliability of the QoS measurement.

This annex is a summary of the responsibilities of the three main stakeholders as laid down in the mandatory and informative parts of this standard. As such, it provides for each stakeholder an overview of his tasks. The responsibilities listed in this annex are either marked as mandatory [mand.] or informative [info.].

F.2.2 Regulatory authority

The regulatory authority has the main responsibility of ensuring the measurement of the quality of domestic and cross-border SPPM services within the Universal Service by:

- requiring the QoS monitoring of the Universal Service Provider and by
- assuring that the measurement systems (domestic, cross-border) of the operator are implemented in compliance with this standard.

In order to achieve these objectives, the regulatory authority has to rely on the auditor whose expertise is needed to fully ascertain the adequacy of the measurement design and of the implementation done by the independent performance monitoring organisation. Therefore, the approval of the auditor is one of the responsibilities of the regulatory authority [mand. (1), see C.2].

The regulatory authority has the responsibility to take into account the national peculiarities of the countries in the field of study [info., see F.3.4].

EXAMPLE Exclusion of P.O. boxes from the measurement system.

As this standard gives room to some interpretations in the survey design and the measurement itself, the regulatory authority may make decisions on some specific points, in particular the:

- removal of some measurement periods in case of "*force majeure*" events and after discussion with the operator [mand. (2), see 5.2.2];
- choice for potential discriminant characteristics outside the examples in 6.4.2 [mand. (3), see 6.4.2],
- number and choice of the postal areas for the stratification [mand. (4), see 6.4.3],
- audit cycle [mand. (5), see Clause 8].
- any calculation rule that differs from the suggestions in Annex B [mand. (6), see B.1];
- measurement of existing SPPM services in the context of the Universal Service [info., see E.3] and the
- frequency of the update of the design basis [info., see G.3],

If the regulatory authority chooses to make decisions on the points above, these decisions are subject to an audit by an external auditor, independent of the regulatory authority himself.

If the regulatory authority is commissioning the measurement system, he has to provide the design basis. The decision for a design basis for each discriminant mail characteristic has to be carried out in accordance with the reported operator [mand. (7)., see 6.3].

F.2.3 Postal operator

If the operator is commissioning the measurement system, the operator has to provide the design basis. The decision for a design basis for each discriminant mail characteristic has to be carried out in accordance with the responsible regulatory authority if the SPPM in the field of study lies within the Universal Service [mand. (1), see 6.3 and info, see G.2.3].

The postal operator usually has the responsibility of implementing the measurement process according to this standard and through the following actions (if applicable):

- the implementation of the real mail studies, which leads to the real mail flows and the weights of the strata. The methodology, the results and the accuracy of these studies shall be documented [mand. (2), see 6.3.2 and 7.4 and info., see G.2.1-G.2.2],

— the provision of logistic or management data to estimate the real mail flows and the weights of the strata. The data sources and the estimation of the real mail flows shall be documented [mand. (3), see 6.3.2 and 7.4],

— the provision of experts who are able to fix certain DMCs. The resulting design basis shall be documented [mand. (4), see 6.3.2 and 7.4],

— the choice of the independent performance monitoring organisation (contractor) charged with the actual measurement for example by a tender process [info. (5), see F.5.3],

— the choice of the auditor charged with auditing the design and deployment of the measurement system, for example by a tender process [info. (6), see H.6.3],

— the provision of adequate geographical postal grids for stratification and recruitment [mand. (7), see 6.4.3 and 6.5],

— the communication of the results to the regulatory authority on a yearly basis [info.].

F.2.4 Independent Performance monitoring organisation

The independent performance monitoring organisation has to ensure that the postal operators have not wrongly influenced the results of the measurement.

The independent performance monitoring organisation is required to,

— ensure that the design of the measurement survey is in compliance with this standard (design basis, geographical stratification, panel distribution, sampling, weighting, calculation, etc.) [mand. (1), see Def. 3.30]

— manage the panel according to the International Chamber of Commerce / ESOMAR International Code of Marketing and Social Research Practice [mand. (2), see 6.6],

— ensure that the panel of senders and receivers is independent of and remain unknown to all postal operators and the regulatory authority [mand. (3), see Def 3.30 and 6.6],

— ensure that individual induction and delivery points are kept unknown to all postal operators involved [mand. (4), see 6.6],

— produce test letters that are not identifiable within the postal network concerning appearance as well as mail volume [mand. (5), see 6.6],

— define and perform all the necessary validation checks and actions to ensure the quality and accuracy of the data collected for the QoS measurement [info., see H.3.1],

— perform the reporting including the weighting procedures and accuracy calculation for the measurement results [mand. (6), see Annex A],

— document all material that is necessary for the performance of the audit including a system of maximum workload caps for the panellists [mand. (7), see 6.6] and the physical archiving of test letters at the end of the contract [info., see H.5.2].

F.2.5 Auditor

The auditing body has the main responsibility of certifying the compliance of the measurement system with this standard [mand. (1), see 7.4] through an exhaustive check of all the critical points of the implemented system, in particular the:

— independence of the measurement system [mand. (2), see 6.1 and C.2],

— studies for the determination and revision of the discriminant mail characteristics [mand. (3), 6.4.2 and C.2],

- methodology, data collection and results of the processes leading to the determination of the design basis including the decision for the type of design basis for each DMC [mand. (4), see 6.4],
- test mail study's methodology, including a possible choice for a disproportional sample design [mand. (5), see 6.1 and 7.3.1],
- adequacy of the geographical stratification [mand. (6), see 6.4.3]
- weighting method and calculation rules which lead to the measurement figures,
- panel management including maximum workload caps and possible re-selection of panellists [mand. (7), see 6.2, 6.6 and C.4.1 and C.4.3],
- integrity of the measurement [mand. (8), see 6.6],
- cases of *force majeure*, of other external influences and the deduction of corresponding periods [mand. (9), see 5.2.2],
- implemented quality control procedures [mand. (10), see C.4.4].

The audit should be performed at least every three years in form of an initial audit and following re-audits [mand. (11), see 7]. The auditor should be fully independent from the audited parties and hence should not be chosen among the competitors of the independent performance monitoring organisation.

F.3 Design of the measurement system

F.3.1 Design parameters

The outlines of the design of an adequate SPPM measurement system are laid down in this standard. The standard furthermore shows a high degree of flexibility by allowing for a number of design parameters.

- **Domestic / Cross Border:** The system can be placed within a country, or it can focus on selected country-to-country mail flows;
- **Field of Study:** Any field of study within the context of postal Single Piece Priority Mail distribution can be selected (see F.3.2);
- **Measurement Period:** Usually being a one year period, based on the calendar year (see 7.4); Multi-year periods are possible in the context of the relaxations of Annex D;
- **Sample Size:** Any sample size can be chosen that respects the Minimum Sample Size requirements of 6.2;
- **Panel Size:** Any panel size can be chosen that respects the maximum panellist workload requirements of 6.2 and 6.6;
- **Performance Monitoring Organisation:** An independent measurement organisation has to be selected, which will bring in its methodology in panel management, data handling, data validation and reporting (see F.5).
- **Geographic Stratification:** A geographic subdivision of the total induction area and the total delivery area should be chosen, which ideally reflects the postal logistic structures in the field of study (see 6.4.3);
- **Discriminant Mail Characteristics (DMC):** Depending on the field of study, a set of mail characteristics will be determined with modes that show a significant difference in on-time performance and that have therefore to be controlled in the design (see 6.4);
- **Design Basis:** A choice has to be made on the structural information that provides a basis for the modelling of the geographic stratification and the DMC. This design basis is the structure for which the test mail

measurement system will be representative. The choice of the design basis will influence the reliability of the total measurement system (see 6.3);

— **Time of Reporting:** A date not later than three months following the end of the measurement period has to be fixed, on which the report will be delivered.

The choice of design parameters has to be documented in the report (see 7.4).

F.3.2 Field of study

F.3.2.1 General

The first step in designing the measurement system is to decide on the field of study.

The *field of study* for a SPPM measurement system that follows EN 13850 is a SPPM flow between defined induction and delivery areas. To define the field of study one needs to specify the induction operator, the induction areas, the delivery areas and the types of SPPM covered by the study.

NOTE The cross-border end-to-end distribution process normally includes two operators in the distribution chain from induction to delivery. The field of study will be defined via the induction operator only, since the operator offers the end-to-end service to the customer. It is the operator's obligation to maintain the end-to-end service at a certain QoS level.

The SPPM has to be characterised according to the transit-time service-level offered to the customer and has to be defined in terms of a J+n transit-time target.

EXAMPLE 1 In many countries priority mail is defined as a SPPM service with a domestic transit-time target of J+1.

The field of study may be an aggregation of mutually exclusive flows with individually defined transit-time targets.

EXAMPLE 2 In a domestic measurement system regional mail flows within urban areas may have a transit-time target of J+1, all other mail flows a transit-time target of J+2. The on-time performance of each test letter is evaluated in relation to the transit-time target of the respective mail flows.

F.3.2.2 Domestic services

For domestic services, the field of study defines the country or parts of the country in which items will be posted (induction area(s)), the operator included who will accept the mail (induction operator) and the country or parts of the country where the items will be delivered (delivery area(s)).

NOTE 1 In the standard domestic case the induction area and the delivery area are identical. The study concentrates on the QoS in this area (*area view*). The domestic sample size requirements of this standard are therefore *per area*.

NOTE 2 It is important to make sure that all parts of the country served by the operator are covered in the field of study.

F.3.2.3 Cross border services

For cross border services, the induction or the delivery areas may be defined as parts or the whole of a single country or as a group of countries.

NOTE In the standard cross border case the induction area and the delivery area lie in different countries. The study concentrates on the QoS on the link between the induction and the delivery area (*link view*). The cross border sample size requirements of this standard are therefore *per link*.

The field of study should cover all parts of the countries served by the included operators.

It may be useful to measure the overall service to / from a group of countries in order not to increase the real mail flow significantly by test mail in the case of an operator forwarding or receiving only small amounts of mail to / from several countries.

F.3.3 Geographical coverage

It has to be ensured that the panels of senders and receivers are properly distributed across the field of study. How a good geographical coverage can be achieved is laid down in 6.4.3.

The required geographical coverage in the field of study is primarily a question of panel recruitment, which goes beyond the requirements of the geographical stratification. 6.4.3 is in place to ensure that:

— Panel recruitment respects a representative variety of smaller geographic regions within the field of study,

— All relevant geographic regions have an equal and fair chance to be part of the measurement,

— Designers of the measurement systems are encouraged to decide for a more detailed recruitment grid that pays respect to finer structures in the logistic network as shown in 6.4.3, Table 3 and Table 4.

The system of geographical coverage may not be related to the geographical stratification. It forms no part of the weighting process.

With a representative placement of panellists throughout the field of study, a sound base is established for the representativeness of the sample of test letters itself. Therefore, the minimum geographical stratification of the test letter sample can be limited to three induction areas and three delivery areas within the field of study (see 6.4.3).

In case the client organisation commissioning the measurement system decides for a detailed logistic structure as a template for the geographical recruitment grid, not all areas have to be filled with panellists. The recruitment can be described as a three-step process:

— Choose a grid of postal areas,

— Sample a random subset of areas from the grid. The minimum size of this subset is given in 6.4.3, Table 3 and Table 4,

— Recruit the panellists in each area of the subset at a random location within the area.

The choice of certain areas should be, in principle, non-discriminant to the transit-time result. This secures that the choice of any subset of areas will lead to the same result as sampling in the total field of study.

The minimum sample size (see 6.2) and the maximal workload per panellist (see 6.2 and 6.6 and H.2.2) define the required number of panellists (panel size).

Table 3 directly assigns the minimum number of geographical areas to be covered to certain panel sizes. This direct assignment works for small panels up to 90 panellists as used primarily in cross-border transit-time measurements of small to medium mail flows as well as in small volume domestic transit-time measurement systems.

For measurement systems with more than 90 panellists, it is relevant how detailed the chosen grid of postal areas is. It can be a fine one, resulting in a low average of panellists per area, or it can be coarse, resulting in a high number of panellists per area. A fine grid imposes stronger restrictions on the recruitment process, but it might be closer to logistic realities like, for example, the network of distribution areas.

The size of the chosen grid and the panel size define the average number of panellists per area.

In panels of more than 90 panellists, Table 4 assigns to exemplary averages a minimum percentage of areas to be covered. They range from 0,25 panellists per area to three panellists per area. For all grids below 0,25 panellists per area, 25 % of all postal induction areas and 25 % of all postal delivery areas have to be covered. For all grids above three panellists per area, all areas have to be included in the recruitment.

EXAMPLE It may be concluded from a calculated minimum sample size that this sample can be handled with 100 panellists. One may further decide for an existing postal grid with 97 cells, resulting in an average of 1.03 panellists per area. In this case, according to Table 4, at least 55 % of all areas (97*0.55 = **54** areas) have to be covered. Another sample might be handled with 80 panellists. Table 3 determines that a minimum of **27** areas forms the recruitment grid here.

F.3.4 Design requirements due to national peculiarities

Those seeking to implement EN 13850 are advised to consult requirements and guidance from the regulatory authority and legislation that apply in their jurisdictions.

Definitions should be approved at national level, because definitions like urban, rural and types of postal areas differ from country to country depending on the number and the concentration of population and the size of the country.

EXAMPLE 1 A common definition of “urban” as “cities and communities with more than 50,000 inhabitants” and “rural” as everything else would not work throughout Europe. Using this definition, only four cities in Ireland would be classed as urban. Similarly, only 10 communities in Norway have population of more than 50,000, four of them close to Oslo and three others within 200 km of the capital. In the UK, the City of London (the business quarter of the capital) would be classed as a “rural” area because it has a very small population.

For these reasons, flexibility is given within the standard to use definitions, which consider national peculiarities. This especially applies to the set of possible discriminant characteristics.

EXAMPLE 2 Applicable types of induction vary from country to country. Some countries have in addition to the types listed in 6.4.2 further types of induction in place like, for example, “franchised service points” in Germany or “mobile postman” in Austria and the Czech Republic.

Design requirements that pay respect to national peculiarities are marked throughout the standard by phrases like “if applicable”.

F.4 Small mail volumes

F.4.1 General

In May 2004 and again in January 2007 the European Union was enlarged by a number of additional European countries. This made it necessary to adjust EN 13850 to suit the needs of the additional countries and to acknowledge the fact that the number of cross-border country-to-country flows increased considerably. Many of these additional countries or flows carry only small mail volumes.

F.4.2 Domestic

In principle, all countries have a domestic market with SPPM volumes large enough to apply EN 13850 without modification. To reach the full application, especially in the area of the real mail studies, a transitional period may however be required. It may be appropriate to run real mail studies and the first measurement year in parallel and afterwards to adjust the first yearly result by weighting (see 6.3.3.2). Countries with smaller SPPM volumes are granted relaxed MSS requirements by allowing smaller minimum test-letter samples when the estimated performance is below certain levels (see D.2).

EXAMPLE Country A has a domestic SPPM volume of 150 million mail pieces per year. Their J+n performance is 74 %. The required annual MSS is therefore relaxed from 8.125 test letters to a yearly sample size of 3.500 test letters.

F.4.3 Cross border

Similar to domestic measurements, cross border flows with smaller SPPM volumes are granted relaxed MSS requirements by allowing smaller minimum test letter samples when the estimated performance is below certain levels. Another type of relaxation is the extension of the measuring period (see also D.3).

This redesign of the requirements for the cross-border study is not limited to the new links with the new EU members, but applies on a similar basis to links between all participating countries.

EXAMPLE The flow from country B to country C is the seventh largest outbound link for country B. Its annual mail volume is 900.000 items. It is the 4th largest inbound flow for country C. This means that the link ”country B → country C” falls into category 2 of medium size flows. The J+3-performance on this flow is 68 %. The required MSS of the 2-year measurement result is therefore relaxed from 370 test letters to 290 test-letters. This sample size should ideally be divided evenly into two annual samples of 145 test letters.

F.5 Measurement organisation

F.5.1 Role of the contractor

A performance-monitoring organisation is responsible for undertaking the survey. Usually this performance-monitoring organisation will act as a contractor for a client organisation commissioning the measurement system, which may be, for example, the postal operator or a regulatory authority. The performance-monitoring organisation is a neutral body between the postal operator, the regulatory authority and the public.

F.5.2 Independence

The measurement is carried out by an independent performance-monitoring organisation. This is defined as a body in charge of the monitoring of the QoS which is "external to, and having no links of ownership or control with the postal operators thus monitored". The independence should particularly relate to all aspects through which the postal operators may possibly have an influence on the measurement results. These aspects are, for example:

- Addresses of panellists shall not be known to the postal operator
- Procedures for checking, validating and removing erroneous measurement results or panellists shall be transparent to the auditor and applied by the contractor.

The credibility of the measurement and its results can only be achieved through full independence of the performance-monitoring organisation.

F.5.3 Tender process

Factors normally taken into account in choosing an independent performance monitoring organisation as a contractor include the capability, experience and reputation of the organisation as well as any legal or regulatory requirements and, of course, cost. To cover all these points it is common to go through a formal tender process. For this process, the following will usually need to be taken into consideration:

- The tender process should be initiated early enough in order to leave sufficient time for the selection of a contractor and for the subsequent set-up and pilot phase conducted by the contractor (see time lines in H.7).
- It should follow all existing national and international legal requirements on a tender process dealing with a contract of a certain financial volume, including demands on publication and timing.
- The client organisation commissioning the measurement system should specify to all bidders the same broad design requirements but can let the bidders come up with their own detailed sample size and panel size suggestions. In this case as much information as possible should be given to the bidders, in particular:
  - A reference to this standard
  - The design basis and data relating to discriminant mail characteristics (if available),
  - The current (perhaps estimated) end-to-end transit-time performance,
  - Any design specifications going beyond what the standard requires (e.g.: an additional requirement of a measurement accuracy of ±1 %).

If no stable estimate of the current end-to-end transit-time performance is available, the published or required service level should be taken.

  - The client organisation commissioning the measurement system can then select from a variety of system designs from the different bidders.

Annex G
(informative)

Design basis

G.1 Discriminant characteristics

G.1.1 Representative sample design

G.1.1.1 Representativeness in a postal end-to-end network

The design of the measurement system should secure that the test letters are allocated as a representative sample of all SPPM in the field of study.

The most basic way to achieve a representative sample would be to draw a simple random sample (SRS) of real mail letters and observe their transit time. Unfortunately, such a design would be impossible to implement with high measurement accuracy. To secure a high-quality measurement, the design relies on pre-fabricated test letters sent and received by a group of selected panellists.

This design approach requires that the test letters that are added to the existing real mail stream do not differ structurally from it. The relevant structures are all those mail characteristics, which have a significant influence on the transit-time result.

EN 13850 deals with end-to-end transit-time results. This implies that a letter is sent by a sender from an induction point to a delivery point where a receiver receives the letter. Possible characteristics therefore are structures related to the senders, the receivers, the postal logistic network and the test letters themselves. The standard deals with these structures in the following form:

— The postal logistic network can be described in geographic terms. The geographic structures should be chosen in a way that mirrors the postal network.

— A number of mail characteristics have a default mode for certain senders or receivers, for example ‘metered mail’ for certain business senders and ‘P.O. boxes’ for certain receivers. They are called panellist-based mail characteristics. These panellist based mail characteristics shall be reflected in the panel structure in order to be able to allocate representative shares of their modes in the sample. Example: Test letters can only be allocated for metered mail franking when an adequate number of receivers has rented a franking machine.

— All other characteristics can be attributed to the test letters themselves and are independent of the panellists in the mail allocation process like, for example, format and weight.

EN 13850 sets out the following candidates for discriminant characteristics that should be taken into account: Geographical area, Type of payment, Type of induction, Time of posting, Type of delivery, Formats, Weight steps and Addressing method. These mail characteristics and their modes listed in 6.4.2 have been included because they have frequently been found to be discriminant.

The characteristics and modes, which are discriminant, depend in practice on the detailed operation of the postal system; it is possible, that factors, which are important to some operators, may not be important to others.

G.1.1.2 Formats and weights

For most of the mail characteristics listed in 6.4.2 possible modes are described in detail. Format and weight are left to be characterised in the context of the field of study. Format and weight classes usually differ considerably between different countries and operators.

For both characteristics at least the two most common modes in the field of study are to be considered. The two characteristics might be combined. The definition of the modes should be orientated on postal products which are distributed by differing logistic processes or are likely to yield different results within a certain process.

In case that there are more than two common modes conceivable, all of them should be included. Modes with more than 20 % real mail share or modes with known distribution problems are likely candidates.

G.1.1.3 Type of induction and delivery

For ‘type of induction’ and ‘type of delivery’ the most common modes are listed in 6.4.2. Due to national peculiarities a greater number of possible modes exists, for example

— Induction at franchised postal outlets, at sorting centres or via ‘mobile postman’, who not only deliver but also collect mail,

— Combined delivery of letters and parcels (in contrast to letter-delivery only).

Therefore, the list of modes in 6.4.2 is non-exhaustive and has to be updated in relation to national and operator peculiarities.

G.1.1.4 Additional mail characteristics

When considering whether additional geographical factors, mail characteristics or modes of mail characteristics should be included in the measurement design, the following should be taken into account:

— The extent to which the factors affect the provided QoS,

— The ability to define and apply these factors consistently
(Preferably based on a nationally established definition),

— The ability to understand the definition of these factors,
(Technical terms like ‘machine readable’ may not be straightforward).

— The availability of relevant real mail information,

— The ability to audit the design basis for the additional mail characteristic.

The first step to define a representative sample is to evaluate, which of the possible mail characteristics can be regarded as *significant* for the end-to-end transit-time performance.

G.1.2 Studies for the evaluation of possible candidates

G.1.2.1 Type and extent of the evaluation

The evaluation of possible candidates for discriminant mail characteristics can be done in variety of ways:

— Practical in-house experience of the operator with his logistic processes has shown that certain processes are likely to deliver a less or better than average transit-time performance.

— External sources of information like, for example, the complaints management have singled out scenarios in which regularly poor performance has been reported by the customers.

— Based on experience, a pre-test either with real or test mail can be performed to verify the conclusions.

— If a test mail measurement system already exists, results from the measurement system can be taken for evaluation.

For the first-time evaluation usually the first three methods are used; for follow-up evaluations one can rely on existing transit-time measurement information.

All evaluations should be governed by the desire to detect as many discriminant mail characteristics as possible. Any new or improved discriminant mail characteristic increases the representativity of the test mail measurement and reduces the risk of overlooking a relevant bias to the results.

It is important to note that for the application of the fourth method all characteristics and modes that are to be inspected have to be part of test mail measurement system in order to be able to perform the evaluation. These include at least the characteristics and modes as listed in 6.4.2.

In the first evaluation at the set up of the measurement system, all characteristics and modes as listed in 6.4.2 have to be checked. Further promising candidates can be included in the evaluation.

In follow-up evaluations, all mail characteristics found to be discriminant in the previous evaluation have to be evaluated again. They can be dropped in the next evaluation only if

— They have not proven to be discriminant and

— They are not part of the modes listed in 6.4.2.

New characteristics and modes can always be added to the evaluation.

G.1.2.2 A quick-check of significance

The statistical check of significance of certain mail characteristics relies on common statistical methodology. The evaluation process should ideally be performed in a joint effort between the operator and the independent performance monitoring organisation.

For evaluations based of the results of the test mail measurement system, the following definition applies:

— A characteristic is discriminant, if at least two modes of the characteristic show transit-time results which differ significantly.

Any check of significance will depend for each pair of modes on the following four variables:

— The definition of the modes

— The sample-sizes of modes

— The performance levels of modes

— The difference between the performance levels of modes

The following table has been included as an initial quick-check for pair-wise significance for common transit-time values. It is based on a *t-test* with a level of significance of $1 - \alpha = 95\%$.

Table G.1 — Quick-Check for Significance

		— **Performance levels larger than:**				
		75 %	**80 %**	**85 %**	**90 %**	**95 %**
— **Sample size larger than**	**310**	9,0 %	8,1 %	7,1 %	5,7 %	3,9 %
	1.250	4,6 %	4,2 %	3,7 %	3,1 %	2,2 %
	5.000	2,4 %	2,2 %	1,9 %	1,6 %	1,1 %
	20.000	1,2 %	1,1 %	0,98 %	0,82 %	0,59 %
	80.000	0,60 %	0,55 %	0,49 %	0,41 %	0,30 %

The table is to be read in the following way:

- First you look up the minimum sample size of the two modes and take the row with the highest value that is still smaller than this sample size.
- Then you take the mode with the lower performance level and choose the column with the highest percentage that is still lower than this performance level.

The decision rule is:

- If the difference between the performance levels of both modes is greater than the tabled percentage, then the two modes are discriminant.
- If the difference is considerably smaller, then the two modes are likely not to be discriminant.
- If the difference is smaller, but not considerably so, a more detailed test has to be performed in order to be able to make the decision.

EXAMPLE 1 The total sample size is 5.000 letters. In the sample 2.000 are stamped, 2.000 are metered and 1.000 are postage paid. Possible comparisons are: *stamped vs. metered, stamped vs. postage paid* and *metered vs. postage paid.* For the first check we denote a performance of stamped mail of 87 % and of metered mail of 91 %. The difference in performance is 4 %. 2.000 letters are smaller than 5.000 but not **1.250**. 87 % is smaller than 90 % but not **85 %**. Therefore the critical value is **3,7 %** (2^{nd} row, 3^{rd} column). Since 4 % is greater than 3,7 %, the difference between stamped mail and metered mail is discriminant and therefore 'franking' is discriminant.

In some cases the definition of the modes is straightforward (e.g. 'franked' versus 'metered mail') in others one is free to define the modes (e.g. 'format' or 'weight'). Significance is highly sensitive to the definition of the modes.

In the first run of the check, the modes should be defined according to practical experience and to where high differences in transit-time performance can be expected.

After comparison of all initial modes of a discriminant characteristic it may turn out for characteristics with more than two modes, that, for example, only one of them is discriminant to all the others. Therefore the following procedure should be applied:

With characteristics with more than two modes two types of significant differences may occur.

- **Type I:** One or more of the modes differ significantly from one or more of the other modes,
- **Type II:** One or more of the modes differ significantly from the total group of all other modes.

If for a characteristic type I differences exist, but no type II differences, all modes should be used as defined. If both type, I and type II differences exist, all modes with no type II difference might be grouped together into one mode.

EXAMPLE 2 Postage paid letters differ significantly from stamped letters as well as from metered letters. Stamped and metered letters do not differ significantly from each other. In this case we have two type I differences. If the postage paid letters also differ significantly from the *common group* of stamped or metered letters, postage paid letters also form a type II mode. Stamped and metered letters might therefore be grouped into a common mode.

G.1.3 Connection between Design Basis and Sample Design

The sample has to be representative according to all modes that have proven to be discriminant for the end-to-end transit-time.

The sample has the form of a stratified sample. The structure (strata) of the stratification is defined by the discriminant mail characteristics and their modes. The transit time will be calculated as a weighted mean of the measured strata results.

The knowledge of the real mail SPPM flow proportions between different modes of all discriminant mail characteristics is required for the panel design, mail allocation process and weighting procedure.

To evaluate these real mail flows, preferably the so called real mail studies have to be established. Since the real mail studies measure values of all those structures that are used to build the design of the performance measurement system, the measured values of all those structures have to be accurate and reliable.

In practice, the test mail monitoring system measures a logistic reality exactly as defined by the discriminant mail characteristics. If the mail characteristics structure resulting from the design basis information deviates from reality, there is a probability that the test mail monitoring system defined in this standard will not be representative. As a measure of caution, the design basis used is to be documented for the audit.

— the requirements of this standard on the information sources of structural information are the provision of:

— rough estimates of the discriminant characteristics at the planning stage of the design,

— accurate and reliable estimates before the definition of the weighting basis and the calculation of the report.

The audit has to verify that the used design basis information is:

— gathered at least according to the actual definition of the discriminant mail characteristics,

— gathered with an adequate accuracy,

— likely to be (still) valid in the measurement period.

Special provisions have to be taken for characteristics that show significant short-term changes in the real mail volumes associated to their modes.

G.2 Design basis

G.2.1 Real mail studies for domestic mail

G.2.1.1 General

Real mail studies (RMS) shall consider all SPPM flows of a given field of study. They shall cover all discriminant mail characteristics for which real mail information is required. Most postal operators manage real mail studies on a continuous basis.

Real mail studies consist of systems where samples of real postal items are drawn when they go through the postal network. The sampled postal items are looked at and their characteristics are described in specific forms to collect the appropriate information (e.g.: day of induction, type of delivery, size and weight of the postal item, etc.).

The samples are designed in accordance with the stage of the postal process where they are drawn, either at the induction stage or at the sorting centres stage, the delivery offices stage, or at the office of exchange in case of RMS for cross-border mail.

Depending on the discriminant mail characteristics (DMC) that have to be measured and on the postal operators' mail process, several types of real mail studies can be implemented. The following table shows possibilities, of which data can be collected at which stage for exemplary DMC.

Table G.2 — Possible real mail studies for exemplary Discriminant Mail Characteristics

Characteristic	Possible Modes	How to collect relevant information?
Induction Area	Postal Area	At the Post offices and sorting centres.
Delivery Area	Postal Area	In every stretch of the process.
Type of Payment	Stamped, Metered, Postage Paid	In every stretch of the process.
Type of Induction	Mailbox, Post office, Pick-Up, Franchisees	Induction point (or in the incoming point of the sorting centre if there are different modes of transportation)
Type of Delivery	Door, P.O.-Box	Delivery office (or the sorting centre in case of different sorting processes)
Format	(to be defined)	In every stretch of the process.
Weight	(to be defined)	At the Post offices and sorting centres. (Use data from sorting machines)
Addressing	Typed, Hand Written	In every stretch of the process.
Weekday	Mo-Fr, Sa, So	Statistics on mail volumes per day.

Real mail studies have to be designed to provide global national data for each DMC. Considering the DMCs to be measured and the information obtained through the methodologies described above, the following picture appears:

Table G.3 — Observation points for real mail studies

		Postal operators' real mail studies	
		Observation at induction	**Observation at delivery**
Type of geographical area	Induction area	Yes	Yes if included in postmark
	Delivery area	Yes	Yes
Discriminant parameters	Day of posting	Yes	Yes
	Time of posting	Yes	No
	Type of induction	Yes	No
	Type of delivery	Partially	Yes
	Mail characteristics	Yes	Yes

The review of the real mail studies aims at reflecting changes in the postal environment. Reasons to review and perform new real mail studies may be the following events:

— General
  — Substantial increase / decrease of the economy in general;
  — Substantial increase / decrease of the number of (a specific type of) postal items.

— Operational
  — Introduction of new SPPM postal products;
  — Introduction of new logistics that affect the distribution of mail;
  — Change of tariff structure.

This list serves only as an example. There might be many more reasons based on national peculiarities and/or general effects that may prove the need for the recurrent performance of real mail studies.

G.2.1.2 Documentation

For all DMC the results of the real mail studies have to be documented in an orderly manner. Elements of the documentation that are required by this standard are:

- measured mail characteristic;
- measured type of mail (if not “single piece priority mail total”) mentioning all types of
  - single piece priority mail excluded,
  - non single piece priority mail (e.g. bulk priority mail) included;
- sampling stage in the distribution chain (induction / delivery);
- geography of measurement (e.g. included mail centres);
- measurement period;
- sample size and sample stratification;
- results;
- development of results over time (if available);
- accuracy.

The documentation should be laid down after each real mail measurement period.

G.2.1.3 Adequate representativeness

An adequate representativeness for the real mail studies is achieved if it can be avoided that a possible measurement error in the real-mail study results may have a significant effect on the on-time performance measurement. The effect is illustrated in the following example:

EXAMPLE A real mail measurement of mode A of a DMC has a positive measurement error of 5 %, of mode B a negative measurement error of -5 % (55 %:45 % measured instead of 50 %:50 % in reality). The real difference in performance between mode A and B is 8 % (mode A: 86 % performance, mode B: 94 %). The measurement error in real mail measurement results in a bias in test mail measurement: 55 %*86 % + 45 %*94 % = 89,6 % instead of 50 %*86 % + 50 %*94 % = 90,0 %. The measurement error in real mail measurement of ± 5 % translates into a bias of 0,4 % in on-time performance.

The required real mail study accuracy depends on the discriminative power of the DMC measured. The more the DMC discriminates, the more accurate the corresponding real mail study has to be.

It is advisable to base the test mail measurement on a real mail measurement with at least the sample size of the test mail study. If the sample size of the real mail study is considerably larger than the test mail sample, the real mail studies measurement error and its consequences on the on-time performance result can be neglected.

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 priority mail flows. In these cases, the real mail studies measurement-period should exclude all times of the year with extraordinary bulk mail activity in order to minimise the resulting bias.

G.2.2 Real mail studies for cross border mail

For cross-border mail, real mail information is required for both posting and delivery countries. In many cases, the operators in the two countries will collect and share information on real mail flows in their own countries. If shared information is unavailable, the operator commissioning the survey will have to collect information about the flows in both countries.

Often postal operators implement real mail studies at the offices of exchange, for inbound and for outbound mail. The methodology of these studies will be the same as that for domestic mail real mail studies, that is to say that for

- Inbound mail, which only requires to look at the address to gather the information, they will provide all the appropriate parameters for geographical stratification and discriminant parameters,
- Outbound mail, collecting induction area information will depend on whether the postmark provides this information or not as already said in Table G.3. If the postmark does not include the induction area, induction points have to be observed.

G.2.3 Alternative design bases

G.2.3.1 General

A decision for a non real-mail based design may lead to:

- different data requirements (and data collection costs);
- a shorter implementation phase;
- higher risk of bias (i.e. loss in representativity);
- additional transparency requirements in the report.

G.2.3.2 Alternative design bases: Proxies for existing real mail flows

Each postal operator usually collects and stores some kind of information on his logistic processes using process monitoring routines, accounting routines, quality control routines etc., which give him insights into the handled mail volumes. Although the information is gathered for reasons beyond the objectives of this standard, it may be utilised for developing a design basis.

It may require some effort to recompose the information into a stable design basis, but it is directly available without producing any extra costs. It may even be the case, that it is gathered more regularly than it would be possible in the framework of a real mail study and that it may therefore be less outdated.

The type of information and its source has to be stated in the report to enable the reader to evaluate its reliability.

G.2.3.3 Requirements for the reporting

All new types of design bases require a special type of reporting in order to guarantee a sufficient level of transparency for the reader:

- data from Real Mail Studies (Standard Design Basis) No reference required. Results are ‘representative for existing operations’;
- logistic / Management data Report the type of logistic or management data for each of the applied discriminant mail characteristics;

G.3 Frequency of update

Real mail studies and other forms of re-evaluation of the structural information laid down in the design basis have to be performed at a minimum once every third year. It is not always necessary for these studies to be undertaken continuously over the 12-month period. They should, however, cover different days of the week and periods of the year.

For smaller mail flows the collection of the required amount of structural information may take up to three years.

The reason why the evaluations shall be repeated is that the set of discriminant mail characteristics may have changed over time.

The set of discriminant mail characteristics should always be revised before the start of a real mail studies measurement period and at least once every third year.

There may be situations where the re-evaluation should be made more often or where real mail studies should even be implemented as a continuous measurement:

— Frequent changes are to be expected in the logistic processes,

— Rapid changes are to be expected in real mail shares of the modes,

— Triggers (for example the complaint management) point at changes in the QoS of certain modes.

The frequency of the re-evaluation should be decided upon in accordance with the regulatory authority to simplify the auditing process.

It is also recommended to consider the effect of any changes in the design basis on the transit-time performance. If a characteristic is not very discriminant, in practice, big changes in the real mail mode volumes may have little effect on the performance and a re-evaluation once in three years would be sufficiently frequent. However, if a characteristic is clearly discriminant then small changes in the real mail volumes per mode could have such a significant effect on the QoS that re-evaluations that are more frequent are needed.

The frequency of the re-evaluations may be different for different discriminant mail characteristics. Whatever frequency has been chosen, it is important that the design basis used for the calculation of the test mail measurement report is as up-to-date as possible to be representative for the test mail measurement period.

EXAMPLE If a continuous real mail study is in place, the real mail measurement period could by synchronised with the test mail measurement period with a time lag of one to three months. So for *a January to December* test mail measurement period, for example, a *November to October* real mail Study result could be chosen.

Annex H
(informative)

Implementing EN 13850

H.1 Stages of the survey

H.1.1 Preparation

H.1.1.1 General

Altogether, the preparation of the test mail survey can take 9 to 18 months. The preparation phases presented in H.1.1.2 to H.1.1.5 may overlap each other.

H.1.1.2 Test mail survey planning phase

As a starting point for the test mail survey, an outline specification for the survey characteristics and dimensions has to be prepared. This specification should identify, in particular, the geographical stratification that will be used and the discriminant parameters that will need to be considered. Beyond the requirements of EN 13850, the specification may include internal information requirements for quality control procedures of the postal operator. Based on the outline specification, a call for tender has to be written (3 to 6 months).

H.1.1.3 Real mail studies planning & set-up phase

For all discriminant mail characteristics based on the standard design, the requirements for the real mail studies, which will provide the information required for the test-mail sample plan, have to be compiled. The real mail studies have to be organised. The real mail studies will in general be performed by the postal operator himself. Real mail planning and set-up may be done in parallel to the test mail survey planning. (3 to 6 months)

H.1.1.4 Real mail studies pilot and final adjustment

The first measurement period of the real mail studies has to be conducted and adjusted. If alternative design bases with additional data requirements have to be calculated, the required data should be obtained now. Meanwhile it is possible to proceed with the tender process for the test mail survey, the selection of an independent survey operator and to begin contract negotiations before the results of the real mail studies are known (preliminary information regarding the discriminant characteristics can be used in the tender process instead). (3 to 9 months)

H.1.1.5 Adjustment of survey design and contract phase

Once full results of the real mail studies or any alternative approaches to determine the design basis are available, the principal design of the test mail survey can be finalised and contractual negotiations with the survey operator can be completed. During this time, required real mail studies will start to be performed on a regular basis, either continuously or in periodic waves at least every three years. (3 months)

H.1.2 Set-up

A period will be necessary for the set-up of the survey. During this time, the panel have to be recruited and trained. The necessary measurement logistics have to be put in place, which includes all relevant software development. (3 to 6 months)

H.1.3 Pilot (testing phase)

A period will be necessary for the testing of the panel and the survey logistics. The survey should run with its full functionality, although it does not have to start with the full sample size. This has to be achieved in time to be able to test all panellists sufficiently before the first measurement period. (6 to 9 months)

The timetable for preparation (H.1.1), set-up (H.1.2) and pilot phase (H.1.3) should be planned as to have them completed at the end of a calendar year.

From the findings of the pilot phase, final adjustments have to be made to the survey operations. After that, the contractor should be able to show that survey logistics and panellists are running within predefined parameters.

Conclusion: Where there are neither systems to measure end-to-end transit-time QoS nor existing real mail studies in place it could take 18 up to 30 months from the beginning of the planning period to the start of the measurement period. (See, for example, time line in Table H.12 in H.7).

H.1.4 Faster implementation

H.1.4.1 General

In a number of cases, faster implementation may be possible: Some sort of real mail collection system or source of logistic or management data is usually available to the postal operators. Many operators even have some external system already in place for measuring end-to-end transit-time performance. Existing real mail information and results from this measurement may be used to evaluate possible discriminant characteristics and to finalise the test-mail survey sample-plan, saving a considerable amount of implementation time.

H.1.4.2 Parallel run of test and real mail studies

If there is no existing real mail study in place it may be possible to speed up the process of finalising the end-to-end transit-time QoS survey design by using existing logistic or management information available to the postal operator instead of real mail study information. The real mail studies would run in parallel with the QoS survey, and the results of the QoS survey would be adjusted retrospectively using weighting at the end of the first year of operation. (See time line in Table H.13 in H.7. Time saved: up to 12 months to the start of the measurement period).

Depending on the accuracy and fit of the logistic or management information, weighting of a higher degree resulting in a violation of certain weighting caps might by necessary after the first measurement period (see 7.3.2). Consensus should be reached between all parties involved (operator, regulatory authority and independent performance monitoring organisation) if this is an adequate trade-off for the time saved.

H.1.4.3 Minor modifications

If external systems to measure QoS already exist with designs already based on real mail studies and if only minor modifications are required, it should be possible for the independent performance monitoring organisations to make the modifications required for compliance with EN 13850 with little delay since neither elaborate planning nor pilot testing phases are necessary. (See time line in Table H.14 in H.7. Planning and adaptation: Up to 6 months to start of measurement period)

H.1.4.4 Major modifications

If external systems to measure QoS already exist with designs already based on real mail studies, but major modifications are required before the surveys are compliant with EN 13850, there may be significant contractual issues with the independent survey operator. Time will have to be allowed for this, the bigger the survey the more time.
To secure stable survey operations a survey pilot phase is required after the survey adjustment phase. (Planning and adaptation: Up to 18 months to start of measurement period).

NOTE Any modification would be a major one if it includes a change in contractor.

H.1.5 Measurement period

The whole measurement system is based on a maximum cycle of three years.

The real mail studies, if not performed continuously, have to be performed at least every three years. The update of the standard design basis or any alternative design basis has to be performed at least every three years, too. (See 6.1). The measurement can have duration of up to 36 months (See Annex D)

There has to be a detailed initial audit of the survey after the first year. It shall be followed by independent re-audits at least every three years that. More detailed independent audits are necessary after measurement periods when the supplier changes or significant changes to the methodology of the measurement occur. The audit cycle has to be agreed with the regulatory authority.

Full survey operation will start at the beginning of a calendar year and is required to last for 12, 24 or 36 months in order to meet the requirements of EN 13850. (For the necessary measurement period, see alsoAnnex D)

It is advisable to start with initial data analysis after the first quarter of the measurement period in order to monitor

— the compliance of the survey with the design specifications and

— the quality of the panel performance.

The report will be issued not later than three months following the end of the measurement period.

H.2 Panellists

H.2.1 Representativeness

Each panellist usually represents a geographical area (urban / rural), type of induction (e.g. post box, post office, pick up), type of delivery, franking method and a certain pattern of last collection times (see 6.4.2). In total, the use of the panel enables a representative sample of the total mail flows in the field of study. The management of the panel shall ensure that this representativeness is maintained over time and that no bias occurs.

The fitting of individuals into the model is achieved by

— respecting their “normal” behaviour for panellist characteristics e.g. induction of metered mail for certain businesses,

— control of variable features such as mail characteristics and date of posting,

— exchange of sampling points/panellists over time.

As long as the characteristics to be represented by a panellist are not compromised, a panellist’s tasks may be executed by someone with different characteristics in the framework of this standard.

EXAMPLE 1 In areas where private panellists are hard to recruit (city centres etc.), but where the test mail volume is high, *professional induction panellists* might be considered for sending purposes. A professional induction panellist basically performs the tasks of a number of panellists and inducts various test letters across the geography within a given time frame.

EXAMPLE 2 In cases where certain business panellists are hard to recruit, private panellists might be able to perform the sending tasks instead of a business panellist. For instance, in some cases metered items to be posted into a mailbox might be posted by private panellists (test items pre-metered by the independent performance-monitoring organisation).

H.2.2 Risk of panellist identification

To keep the identity of panellists unknown to the operator and regulatory authority – both sender and receiver – the outward appearance of test mail shall be as inconspicuous as possible, and their volume has to be kept on an ordinary level (see 6.6).

The definition of the induction and delivery areas in the reported results should be unspecified enough to minimise the risk of panellist identification. The level to which geographical information is broken down for the operator should be broad enough that each sub-area contains panellists and that therefore no sub-areas can be identified that contain no panellists. These measures secure, that quality-improvement efforts are applied by the operator(s) on the whole field of study and not only on single employees or induction- / delivery-points.

As a consequence, for most countries not all digits/letters of the postcode should be revealed.

To define a system of appropriate maximum workload caps is the responsibility of the independent performance monitoring organisation and shall refer to the issues stated in 6.6. These issues either refer to

— noticeable mail characteristics or to

— a high test mail volume in comparison to the panellists' usual real mail volume.

Apart from the integrity of measurement, maximum and minimum workload caps play a vital role in securing a high level of panellist performance. The adequate selection of maximum workload caps is subject to the audit.

The mail supply packages / envelopes for the senders should not bear references to the independent performance-monitoring organisation by company logo or sender address or by incorporation into a metered franking. It is recommended to put on a camouflaged / neutral sender address for return letters, e.g. a post box, to avoid the opening of letters and detection of the panellist IDs on the test letters inside.

The outward appearance of mail supply could be varied by type of envelope, font type and mode of franking.

H.2.3 Induction and delivery

H.2.3.1 Induction and last collection

The organisation operating the test-mail measurement system shall ensure that panellists usually send test mail at the prescribed date and at the prescribed time (if time slots are part of the stratification).

— If time of posting is discriminant and the induction point has more than one collection, panellists have to be instructed on the importance of keeping the last collection time and the time intervals that are scheduled.

— If time of posting is not discriminant or there is only one collection, panellists are left to perform their induction on the scheduled day of posting according to their natural posting behaviour.

Provision has to be made by the independent performance-monitoring organisation for items posted after the last collection. Confirmation of time of posting relative to the advertised or contracted last time of posting is a usual requirement. 'Last collection' refers to the advertised last time for collection of the selected post box or to the contracted latest time for collection of mail within the field of study (not the actual time). If a postal item is inducted after the last collection then J should be adjusted to the next day of collection for that type of mail.

In practical terms, panellists may sometimes need to induct on a calendar day ahead of the date of induction to meet a specific requirement. For example, in some rural areas there may be an early last collection time, particularly at weekends.

EXAMPLE 1 In these circumstances panellists may be permitted to induct a planned Sunday deposit test item on Saturday evening provided it is inducted after the last Saturday collection and before the Sunday collection; the test item will be counted as having a Sunday date of induction.

In case time of posting is a discriminant mail characteristic, panellists have to post their letters in certain time intervals. For the individual panellists this is only relevant, if there is more than one daily collection at their induction point(s). If this is the case the panellists have to be instructed before which of the announced collections they have to post the test letter. The longest time interval is from midnight up to the first collection. No letter should be posted after the last collection.

Panellists should orientate themselves according to the announced collection times and not according to any other indication if the clearance has already taken place or not (e.g. a 'next clearance' indicator on the letterbox).

EXAMPLE 2 In a J+1 field of study a letterbox has three collections: 15:00, 18:00 and 21:00. The 21:00 clearance guarantees next day delivery for local mail only. Therefore 18:00 is the last J+1-collection. Posting instructions either prescribe to post before the first clearance (15:00) or to post between the first and the last J+1-clearance (15:00-18:00).

Delays before the letter reaches the postal operator e.g. by the use of an independent collection service, would bias the results. If a collection service is managed by a third party beyond the control of the panellist, this collection service / panellist should be excluded from the study (see 6.7).

If a collection service by the contracted postal operator observed in the measurement is used, the contracted time of pick up gives the start of the measurement of transit time.

In a large company, mail will often be handled by a mailroom. Delays between panellist and mailroom or within the mailroom are not allowed to enter the measurement. One way of doing this is to ensure that the test letters are posted by the mailroom supervisor.

H.2.3.2 Delivery and correct addressing

Test mail shall bear a complete, correct, unambiguous, and legible address of delivery according to the rules published by the responsible organisation. The applied rules should not differ significantly from the normal addressing behaviour of the customers.

Provision has to be made by the organisation operating the measuring system that addresses of the panellist database are properly put on the envelopes. This is usually achieved by manufacturing the complete test mail in a printing facility under supervision of the organisation operating the measuring system instead of having sender panellists doing this. Special care shall be taken in the case of handwritten addresses.

NOTE For handwritten addresses it is advisable to either print them as 'checked and scanned' handwritten addresses or in an unsuspicious font of 'handwriting' (*machine quasi-handwritten addresses*) instead of using real handwriting on the envelopes, which might be prone to writing mistakes.

Delays outside the responsibility of the postal service operator, e.g. by the use of an external delivery service, would bias the results (see 6.7). If a delivery service is managed by a third party beyond the control of the panellist, this delivery service / panellist should be excluded from the study (see 6.7).

Delays between panellist and mailroom or within the mailroom are not allowed to enter the measurement. One way of doing this is to ensure that the test letters are received by the mailroom supervisor.

In case of a six day delivery service by the operator it has to be recorded if the panellist is able to receive mail on the sixths day (usually the Saturday). Companies can be closed on Saturdays and therefore cannot always distinguish between mail that has been delivered on a Saturday and mail that has been delivered on a Monday. In this case these panellists have to be recorded as being not active on Saturdays which has to be respected in the validation and transit-time calculation.

H.2.3.3 P.O. boxes and pick-up times

Practical difficulties can occur in identifying the correct date of delivery for an item delivered to a P.O. Box (see 6.7). The date of delivery should be recorded on the day on which the item is placed into the box. If the box is not visited and cleared every day or if it is visited too early before all the mail has been sorted into the box the date of delivery may be recorded wrongly.

The extent of the difficulty can depend on the local operation, e.g. whether customers can access their P.O. Boxes before all the mail has been sorted, and on the ability of the independent performance monitoring organisation to recruit reliable panellists.

If there is no published pick-up time, this could make the accurate measurement of the day of delivery difficult.

EXAMPLE A panellist collects the mail at 10:00, which is usually available by then, but the postal administration aims to provide post by 10.30 without publishing a pick-up time. On this day, the mail is put in the box at 10.10, after the panellist has left. Is this a failure? The panellist would report it as delivered on the following day when he collects the next day's post.

In such cases, rules will need to be determined as to whether P.O. Boxes should be excluded entirely, or as to the time at which the postal operator should be given the benefit of the doubt.

It is also necessary to consider whether the results of the survey will be biased by not sampling P.O Boxes and this depends on what proportion of mail is delivered to P.O. Boxes and on whether P.O. Boxes can be expected to have different QoS from other types of delivery.

The argument for excluding P.O. Boxes is strong if practical difficulties are likely to arise and if the likelihood of bias is small.

H.2.4 Panel turnover

Panel turnover describes the overall exchange of panellists through the ending of the participation of one panellist and subsequent replacement by another. This replacement is required if a panellist

- Drops out / ends participation on his own accord,
- Is identified by the postal operator and has to leave the system,
- Has to be removed because of insufficient participation,
- Has to be replaced because of changes in the stratification requirements,
- Has to be removed when his participation time expires.

The objectives here (see 7.2.2) are to maintain as far as possible the independence of the measurement by both

- Avoiding the possibility of identification by over-use of induction and receiving points over an extended period and
- Allowing greater representation of posting and receiving points in the survey.

There is a trade-off between costs for panel size (panel recruitment and maintenance) and the costs for sample size (test letter production and sending). Ongoing variation in induction and receiving points also decreases the correlation effect and therefore the required minimum sample size. Another consideration is that experienced and reliable panellists are exchanged by new panellists who usually require training and more intense control of their reliability.

For mailbox induction points, this need not necessarily imply a change in the private panellists carrying out the posting of test items, provided they are able to rotate their use of mailboxes. In practice, however, most private posting panellists will be able to use only a limited range of mailboxes, and in order to meet the objectives above, they should be treated in the same way as private panellists representing receiving points.

For optimal correlation reduction and stability of results, the exchange of panellists should be ongoing throughout the measurement period. Therefore an appropriate part of all household panellists (including those posting test items) should be replaced every year. Due to the limited number of businesses and the difficulty and cost of their recruitment, business panellists are allowed to stay longer in the panel.

The result should be a complete replacement of the private induction and delivery points after four years, rather than relying solely on drop-out rates.

Normal drop-out rates might result in a ”natural” panel turnover of a scale that reaches the panel size after four years. Nevertheless, in certain sections of the panel this natural panel turnover may occur repeatedly throughout the four-year period, whereas in other sections there is no exchange at all. Therefore the natural panel turnover alone does not guarantee a complete replacement.

H.3 Validation and transit time calculation

H.3.1 Data validation

H.3.1.1 General

To produce reliable results it is essential to guarantee that only valid information is included in the database. Therefore, all recorded information from the panellists should undergo a rigid data validation process.

Data validation routines are set up to detect data inconsistencies that give significant reason either to correct the data set or to remove it. Each detection routine should be connected to a well-defined corrective action. The validation routines are usually performed in two steps, an item based validation and a panellist based validation.

Improved validation routines rely not only on the information provided by the panellists, but also on additional information connected to the process like, for example, the date of data capture or a defined maximum data capture delay.

H.3.1.2 Item-based validation

Item-based validation processes contain checks that can be applied to the data of any single letter. Item validation processes can be performed directly after new incoming information is recorded in the database. Item validation is a continuous process performed during the measurement period.

It evaluates the validity of the posting and delivery dates and checks them in comparison to other reference dates like the date of mail allocation, the postmark date (if available) and the date of data capture into the system.

Examples of item related inconsistencies questioning the date of induction or delivery are:

- no date information: Date information is not available either for the date of induction or the date of delivery;
- unsure date information: The information is questioned by the panellist himself ("I'm not sure, but…")
- invalid date information: Information is in no valid date format ("41.06.2008");
- date of induction and/or delivery is before the date of mail allocation;
- date of induction and/or delivery is after the date of data capture;
- date of delivery is before the date of induction;
- date of delivery is a Sunday or a national or regional public holiday on which no delivery is made;
- postmark date is equal to or after the date of delivery.

Other examples of item-related inconsistencies require that a closer look at the time of posting has to be taken:

- time of posting is after the last collection time (date of induction has to be set to the next working day);
- no last collection time (the last collection time for the point of induction is unknown);
- no time of posting;
- geographical postmark information (if available) differs from the corresponding geographical information known for the point of induction (for example differing induction mail centres).

Many item based validation routines can be integrated as checks into Internet based online data capture routines. Online entries by panellists can be checked in real time, making it possible to point out mistakes directly and – by doing so – giving the panellist an immediate possibility to reconsider his entry.

H.3.1.3 Panellist based validation

Panellist based validation is applied on all items posted or received by a panellist within a given measurement period. It is performed at a minimum after the end of the measurement period and before the calculation of the report to ensure that a complete picture of the reliability of the information provided by a panellist can be drawn.

Panellist based inconsistencies are not as clear-cut as item based inconsistencies. Checks on them relate to accumulations of inconsistencies of certain types, which as single incidents do not suffice to remove the corresponding items from the database. They are usually of the following form: "If more than x items show the inconsistency and/or at least y letters have been sent/received by the panellist and/or the panellist has previously shown questionable performance then…"

Examples for inconsistencies related to individual panellists are: More than x…

— …% of letters have a *postmark date* (if available) that is not equal to the date of induction;

— …letters have *geographical postmark information* (if available) that differs from the corresponding geographical information of the point of induction that the panellist should have used (for example differing induction mail centre);

— …letters with long transit-times (J+3 and more) are received on one day and no letters are received on the previous two working days.

Sometimes panellist based validation rules can be necessary but complex.

EXAMPLE 1 When an operational transit time calculation rule is applied in a country with Saturday delivery (see H.3.3.3), the following scenario should be evaluated: "More than *x* delayed test letters are received by a company on a Monday and it has received at least *y* letters during the measurement period. The company has stated to be able to receive letters on a Saturday but no test letters are received by the company on any Saturday during the measurement period."

It is important to note that the sequence of validation steps is not in all cases irrelevant.

EXAMPLE 2 For validation steps working with the induction date it is important to first establish the induction date adjusting the date of posting for letters posted after the last collection time.

Since one item can show multiple inconsistencies, it is equally important to consider whether:

— corrective action is sequentially executed immediately after each validation check or whether,

— one final bundle of corrective actions is determined after performing all validation checks.

EXAMPLE 3 One item shows two inconsistencies type A and type B. There are altogether 1+8 items with inconsistency A and 1+3 items with inconsistency B. The validation rules determine that an accumulation of at least 6 items with inconsistency A lead to a removal of these items. The same holds for a group of at least 4 items with inconsistency B. With final corrective actions, all 1+8+3=12 items will be removed. If corrective action A is performed *before* applying check B, only 3 items are left in group B, which would not qualify as a significant inconsistency any more.

Final corrective actions are preferable because they make better use of all available information although they are more complicated in implementation. In using sequential corrective actions, one has to consider the ranking of the validation steps, because the information basis is reduced by each step.

The definition of all necessary validation checks and actions is the responsibility of the performance-monitoring organisation.

All insights from the item validation and panellist validation processes should be used not only to correct the database. They should also become part of a continuous routine, monitoring individual performances of all panellists.

H.3.2 Service standard

On-time performance of the end-to-end transit time is measured against the transit time target, the so-called *service standard*, i.e. the number of working days within which the items should be delivered. EN 13850 does not specify what the service standard should be, so it may be specified by the regulatory authority or legislation or may be chosen and published by the operator. The European Commission has, for example, proposed a service standard of three days for cross-border priority mail within the Universal Service Obligation (USO).

The field of study may be an aggregation of mutually exclusive flows with individually defined service standards.

On-time performance results need to be compared against the same service standard.

H.3.3 Transit-time calculation rule

H.3.3.1 General

The requirements of Annex B define how to calculate the number of transit-time days from date of induction to date of delivery. These days shall be postal working days, i.e. ones during which the mail will be handled by the postal operators for collection, processing or delivery.

The rules take into account induction and delivery on specific dates and transit across specific dates.

H.3.3.2 Mandatory calculation rule

B.1 provides the mandatory calculation rule for all SPPM measurement systems that measure according to EN 13850. It is adequate to produce comparable End-to-End transit-time results for similar fields of study.

Transit-time calculation here is based on a five-day working week (minimum service), irrespective if the operator offers its service from Monday to Friday or Tuesday to Saturday as long as the measurement covers a working week with five consecutive working days (irrespective of holidays).

Letters for which the induction and the delivery day fall on the same date have a transit-time of J+0.

EXAMPLE 1 A letter is posted on Monday after the published last collection time. Therefore, the day of induction is one day after the day of posting. If the actual collection is late on that day and made after the time of posting then the letter might reach the addressee on the next day, being the day of induction.

EXAMPLE 2 A letter is posted on Saturday before the last collection time and received on Monday. If Saturdays and Sundays are subtracted, the next working day is Monday, which becomes the day of induction.

Non-working days and national holidays are excluded from the transit time calculation. These days may influence the offered service on the induction as well as on the delivery side of the mail distribution process. They are excluded, if applicable

— In the country of delivery, except on the day of posting, when the letter is still in the country of induction,

— In the country of origin, if posting is done on a nonworking day or a national holiday or one day before that day. This pays respect to different holiday collection patterns applied throughout Europe. It is assumed here that there is a significant probability that the letter has already left the country of origin two days after posting.

EXAMPLE 3 A letter is posted in Sweden on Monday and it arrives in Germany on Friday. That implies a transit time of 4 days. If there is a holiday in Sweden on Tuesday, no collection will be made on Monday. Therefore, the transit time will be reduced by one day to 3 days. If Tuesday is also a holiday in Germany, it will NOT be subtracted twice. If Wednesday is a holiday in Germany Wednesday will be subtracted, too, so that the final transit-time result will be 2 days.

EXAMPLE 4 A letter is posted in Sweden on Monday and it arrives in Germany on Friday. If Monday and Tuesday are holidays in Germany, only, Tuesday will be subtracted, since the letter will still be in Sweden on Monday (3 days transit-time).

EXAMPLE 5 A letter is posted in Germany on Monday and it arrives in Sweden on Friday. If there is a holiday in Germany on Tuesday, a collection will be made on Monday, but no regular processing will be done in the night from Mondayto Tuesday. Therefore, the transit time will be reduced by one day to 3 days.

EXAMPLE 6 A letter is posted in Sweden on Friday and it arrives in Germany two weeks later on Monday. This implies a transit time of 0+0+1+1+1+1+1+0+0+1= 6 days. If the letter is posted on Saturday, it is handled as posted on Monday (transit time 5 days). Adding to that, if the letter is already received on Saturday, it is nevertheless handled as received on Monday, resulting in no change in transit-time (5 days again).

For cross-border mail, only the calculation method of B.1 is allowed for the purpose of this standard.

H.3.3.3 Additional calculation rules

Since the minimum requirements of EN 13850 are based exclusively on the calculation rule of B.1, in addition any user of this standard is free to define his own operational calculation rule for additional reporting purposes.

The inclusion of extra working days in the operational calculation rules sometimes requires extra rules for specific scenarios. These extra rules should become part of the additional calculation rule.

EXAMPLE 1 If a valid Saturday delivery of an item is recorded, then Saturday is a valid day for transit time calculation regardless of whether the receiver is private or business or whether the business is able to identify post received on Saturdays. If the Item arrives on the next working day, then the Saturday should be subtracted if the receiver is a business that cannot identify Saturday items.

Published regional holidays may be subtracted, because the customer would not expect any service on a regional holiday. It is important however to ensure that customers have been informed and are aware that postal services will be affected by the holiday.

EXAMPLE 2 Due to the federal structure of Germany several holidays are valid for certain 'Bundesländer' only. For the users in these 'Bundesländer', the holiday is perceived as any other national holiday.

The calculation of transit-time results between domestic regions with different public holiday patterns should be handled like the calculation of transit-time results for cross-border flows.

H.3.4 Loss

EN 13850 does not measure lost mail. It has to produce a report within a reasonable timeframe after the end of the measurement period. Therefore, items not delivered within 30 days may be ignored. This is because it can be difficult to tell whether a missing test item has been lost within the postal system, lost outside the postal system, has never been posted or has been received but overlooked by the recipient. The following definitions apply:

— *Letter on-time:* Delivered within the service standard, for example domestically J+1,

— *Letter delayed:* Transit-time beyond the service standard up to J+30,

To be able to report all delayed letters, final reports can be calculated not sooner as all possible J+30 items had a reasonable chance to be reported back, processed and validated.

H.3.5 Force majeure

H.3.5.1 Best practice

Services and measurement within the postal sector depend on clear and unified principles. To ensure a common understanding and handling within operators and regulatory authorities when unforeseen situations occur it is useful to set up guidelines or principles that should be followed.

Force majeure is a common principle with different approaches depending on history, knowledge e.g. The definition of *force majeure* for the purpose of this standard is derived from the UNIDROIT PRINCIPLES OF INTERNATIONAL COMMERCIAL CONTRACTS (2004):

— Non-performance by a party is excused if that party proves that the non-performance was due to an impediment beyond its control and that it could not reasonably be expected to have taken the impediment into account at the time of the conclusion of the contract or to have avoided or overcome it or its consequences.

— When the impediment is only temporary, the excuse shall have effect for such period as is reasonable having regard to the effect of the impediment on the performance of the contract.

— The party who fails to perform shall give notice to the other party of the impediment and its effect on its ability to perform.

The impediment mentioned above may affect the ability of the operator to perform as well as the ability of the independent performance monitoring organisation to measure.

It is advisable to keep a dialogue on *force majeure* events with the regulatory authority. All cases should be resolved immediately after the event in order not to delay the report at the end of the measurement period. This is especially advisable when internal preliminary reports are calculated for example monthly or quarterly.

H.3.5.2 Domestic systems

The postal operator that claims *force majeure* should notify the auditor and the regulatory authority within an agreed period as soon as the operator knows the range of the incident and its consequences by reporting the:

- kind of incident claimed as a *force majeure* event,
- range of the influence on postal operations, and the public consequences,
- expected time frame and actions taken to limit the effects,
- range of the requested withdrawal period.

An agreement should be achieved between the operator and the regulatory authority.

The postal operator shall provide the auditor with the following information in a written report:

- the kind of incident that has occurred and reasons for claiming it as an event of *force majeure*,
- documentation of actions taken to limit the consequences, and further, who and how various stakeholders are informed about the incident and the available alternatives.

H.3.5.3 Cross-border systems

The postal operator that claims *force majeure* should notify all other operators involved immediately and the auditor and the regulatory authority within an agreed period. At the time the operator knows the range of the incident and its consequences he reports to them the

- kind of incident claimed as a *force majeure* event,
- range of the influence on postal operations, and the public consequences,
- expected time frame and actions taken to limit the effects,
- range of the requested withdrawal period.

If an involved party claims a *force majeure* event, an agreement should be achieved between the operators and the regulatory authority involved.

The postal operator shall provide the auditor with the following information in a written report:

- the kind of incident that has occurred and reasons for claiming it as an event of *force majeure*,
- documentation of actions taken to limit the consequences, and further, who and how various stakeholders are informed about the incident and the available alternatives.

H.3.5.4 Examples of force majeure

- natural disasters; earthquake, flooding or other extreme weather conditions (which are unlikely in that region or country) causing damage to e.g. goods, infrastructure, people and making the postal operator unable to perform its obligations,
- war or terrorist activity causing physical damage to e.g. goods, infrastructure, people or creating a psychological distress that results in non-performance,
- general strike; an external strike outside the operators influence and where all major transportation systems are blocked on a nation-wide level.

The following events may not qualify as *force majeure*:

- strike within the operators influence,

- periods of the year or days with an unusually large volume of mail and / or parcels, independent of the induction point.

H.4 Weighting

H.4.1 Weighting and stratification

H.4.1.1 General

As the QoS measurement system has to provide results that reflect the design basis, usually being a system of real mail flows; the test mail flows for each stratum should then be proportional to the corresponding real mail flows. Strata with an extremely low real mail volume can be excluded from the measurement system (see section 7.3.2 on weighting caps).

Theoretically, that proportionality could be achieved through a Simple Random Sample (SRS). It would require extremely high target volumes in order to secure that all the relevant strata are part of the sample. On top of that, there would be no guarantee that these relevant strata would be sampled according to the corresponding proportions in the design basis.

An alternative to the SRS is to use a stratified sampling method; according to that method, each mode of each discriminant mail characteristic should be sampled independently. The design then defines the target volume corresponding to each particular mode.

In a fully proportional design, one would define a target proportion for the modes of each discriminant mail characteristic identical to the proportion of these modes in the design basis.

Each particular stratum is defined as a certain combination of modes of discriminant mail characteristics.

The results measured on different strata are aggregated as a weighted mean, using a weighting basis reflecting the design basis. Possible options for building up the weighting basis are described in section H.4.1.3.

The stratified sampling method is advantageous as:

- it allows a better control of the strata to be sampled; therefore it increases the representativeness of the total sample and reduces the sampling error (for instance, oversampling of irrelevant strata or undersampling of important ones),

- it allows calculating a weighted result that shows less variability than the simple average of a SRS.

EN 13850 makes it mandatory to implement a fully proportional design except for the geographical stratification, where non-proportional target volumes are allowed. When proportionality is not achieved, a weighting system should be applied in order to restore it. The following is a list of examples:

- a deviation from the real mail situation has been allowed (for instance, if the geographical stratification is based on a disproportional model),

- no real mail study was available at the time the QoS survey was implemented and the RM study had to be run in parallel with the survey; a retrospective weighting can be used to adjust the results at the end of the first year of operation,

- the target proportions defined in the sample design could not be achieved due to cases of non-response or invalid test items (corrective weighting).

Where portions of real mail may temporarily shift within the field of study, it might be advisable to implement a systematic weighting procedure in order to keep the measurement flexible. In such situations, a retroactive weighting could be applied, based on updated real-mail information obtained during the measured period.

The use of a fully proportional model prevents that extreme weights are attributed to very small subsamples and that the failure of a small number of test items has a major impact on the final measurement. It also improves the accuracy of the final weighted result.

H.4.1.2 Real mail distribution and Real Mail Weights (RMW)

Let us consider a set of discriminant mail characteristics $C_1, C_2, C_3, \ldots, C_M$ that have to be measured according to the standard. For a given mail characteristic C_m, one can define a:

- vector of modes $(C_m^1, C_m^2, C_m^3, \ldots, C_m^{im})$, each mode being a possible value taken by the mail characteristic.
- vector of *real mail Mode Weights* or *RMW* $(RMW_m^1, RMW_m^2, RMW_m^3, \ldots, RMW_m^{im})$, with RMW_m^1 being the real mail weight associated to the mode C_m^1. A real mail mode weight is the proportion of the total real mail flow that is associated to the real mail flow of letters which correspond to the mode. The set of RMWs is measured through real mail studies.

EXAMPLE Consider the format of a letter to be a mail characteristic that has been proven discriminant. The standard format may have a real mail flow share of 80 %, which is the RMW of the mode 'Standard' of the characteristic 'Format'.

H.4.1.3 Weighting Basis (WB) and Calculated Mode Weights (CMW)

Each stratum J is the result of a crossing between the mail characteristics required by the statistical design. The stratum can be defined by a vector of the involved mail characteristics' modes $(C_1^{J_1}, C_2^{J_2}, \ldots, C_M^{J_M})$.

To build up a weighting system, the stratification requires the definition of a set of strata weights, called the weighting basis (WB), i.e. a set of weights to be applied at the stratum level (RMW are only collected at the mode level!).

The Calculated Mode Weights (CMW) are calculated as marginal sums of the weighting basis for each mode of each discriminant characteristic (see Formula (A.4)).

The chosen weighting basis is valid if all CMW equal their corresponding RMW. The definition of the weighting basis is up to the performance monitoring organisation under the restriction that the RMW have to be respected.

The most straightforward definition of a weighting basis is the Standard Weighting Basis (SWB) in which the stratum weights are defined as the product of the corresponding mode weights. The standard stratum weight SWB_J for the stratum J will be calculated by multiplying the real mail proportions associated to all the modes that define the stratum.

$$SWB_J = \prod_{m=1}^{M} RMW_m^{J_m} \qquad \text{(H.1)}$$

The standard weighting basis can be taken to define a proportional sample design.

NOTE The definition of the SWB relies on the assumption that the discriminant characteristics are independent in the population. This might not always be the case. *Example:* Franked mail ('addressing method') will not be posted to the same degree via induction at the post office ('mode of induction') as metered mail will do.

The calculation of the standard weighting basis (SWB) might lead to strata weights which are quite small and difficult to implement into a stable sample design. It may also be the case that the assumption of the independence of the characteristics does not correspond to reality. In these cases the independent performance monitoring organisation is free to perform optimisation routines to modify the SWB into an alternative weighting basis that is closer to reality and secures a more stable final weighting routine. Nevertheless, the resulting CMW have to respect the system of RMW.

H.4.1.4 Individual Final Weight (IFW)

Similarly, the Individual Final Weight (IFW) of an item belonging to the stratum J is given by

$$IFW_{j\in J} = \frac{WB_J}{n_J} \tag{H.2}$$

where n_J is the number of valid items sampled in the stratum *J*.

H.4.1.5 Alternate formulation: Corrective factors

The IFW can also be calculated based on corrective factors. Corrective factors can be calculated at the mode level and at the stratum level; in both cases, they reflect the over- or underrepresentation of the subset in the sample as compared to its target weight, should the sample be fully proportional.

— For a given mode C_m^i of a given mail characteristic C_m (including geographical strata), the corrective factor would be calculated as follows:

$$Corr_m^i = \frac{CMW_m^i}{p_m^i} \tag{H.3}$$

with $p_m^i =$ proportion of the total sample corresponding to the mode C_m^i

For a stratum *J* defined by a vector $(C_1^{J_1}, C_2^{J_2}, \ldots, C_M^{J_M})$ of the mail characteristics' modes, the corrective factor would be:

$$Corr_J = \frac{WB_J}{p_J} \tag{H.4}$$

with $WB_J =$ weight corresponding to the stratum *J*

p_j = proportion of the total sample belonging to the stratum *J*

If we choose the standard weighting basis shown in Formula (H.1), we can rewrite Formula (H.4) as follows:

$$Corr_J = \frac{n}{n_J} * \prod_{m=1}^{M} RMW_m^{J_m} \tag{H.5}$$

with $RMW_m^{J_m} =$ RMW associated to the J_mth mode $C_m^{J_m}$ of the DMC C_m

n_J = number of valid items sampled in the stratum J

n = number of valid items in the total sample.

At the stratum level, the corrective factor can also be interpreted as the ratio between the weight of an item belonging to the stratum according to the actual sampling (that might be not fully-proportional) and the weight of the item if all items were given an equal weight (which would correspond to a fully-proportional sampling). Since the weight of any item would be equal to 1/n under the assumption of a fully-proportional model, one can write:

$$Corr_J = \frac{IFW_{j\in J}}{\frac{1}{n}} = IFW_{j\in J} * n \tag{H.6}$$

which establishes the link between the corrective factor of a given stratum and the IFW of an item belonging to that stratum.

H.4.1.6 Illustrative example

Let us consider a simplified example with three geographical strata (*Geo1, Geo2* and *Geo3*) and a discriminant characteristic MC with three modes (*MC1, MC2* and *MC2)*; the real mail proportions associated to the different modes are given in Table H.1 and Table H.2. Table H.3 gives the split of the sample (2000 valid items) between the nine resulting strata.

Table H.1 — RMW corresponding to the modes of the geographical characteristic

Geo1	30 %
Geo2	50 %
Geo3	20 %
Total	100 %

Table H.2 — RMW corresponding to the modes of the discriminant characteristic MC

MC1	10 %
MC2	35 %
MC3	55 %
Total	100 %

Table H.3 — Number of valid items per stratum

Sample	MC1	MC2	MC3	Total
Geo1	100	250	350	700
Geo2	120	310	450	880
Geo3	80	140	200	420
Total	300	700	1000	2000

The standard stratum weights are calculated by crossing proportions in Table H.1 and Table H.2 according to Formula (H.1); the resulting standard weighting basis is summed up in Table H.4:

Table H.4 — Standard Weighting Basis

SWB	MC1	MC2	MC3	Total (CMW)
Geo1	3,00 %	10,50 %	16,50 %	30,00 %
Geo2	5,00 %	17,50 %	27,50 %	50,00 %
Geo3	2,00 %	7,00 %	11,00 %	20,00 %
Total (CWM)	10,00 %	35,00 %	55,00 %	100,00 %

Table H.4a shows an alternative valid weighting basis that also respects the RMW (the CMW for both discriminant characteristics are equal to the corresponding RMW in both examples):

Table H.4a — Alternative Weighting Basis

AWB	MC1	MC2	MC3	Total (CMW)
Geo1	3,33 %	9,50 %	17,17%	30,00 %
Geo2	4,00 %	18,50 %	27,50%	50,00 %
Geo3	2,67 %	7,00 %	10,33%	20,00 %
Total (CMW)	10,00 %	35,00 %	55,00%	100,00 %

Table H.5 shows the individual final weights (IFW) of the standard weighting basis to be applied to the items in each stratum, given the sample volumes as shown in Table H.3:

Table H.5 — Individual Final Weights (IFW) for the standard weighting basis in each stratum

IFW SWB	MC1	MC2	MC3
Geo1	0,0300 %	0.0420 %	0,0471 %
Geo2	0,0417 %	0.0565 %	0,0611 %
Geo3	0,0250 %	0.0500 %	0,0550 %

Table H.5a shows the individual final weights (IFW), should the alternative weighting basis be used, as presented in Table H.4a:

Table H.5a — Individual Final Weights (IFW) for the alternative weighting basis in each stratum

IFW	MC1	MC2	MC3
Geo1	0,033 %	0,038 %	0,049 %
Geo2	0,033 %	0,060 %	0,061 %
Geo3	0,033 %	0,050 %	0,052 %

The calculation of corrective factors requires us to compare the real mail weights with the sampling proportions for each mode of each discriminant variable.

At the stratum-level, the sampling proportions can be derived from Table H.3 by dividing the volume sampled in each stratum by the total sample-size, all strata being put together.

Table H.6 — Sampling proportions per stratum

Sample Prop	MC1	MC2	MC3	Total
Geo1	5,00 %	12,50 %	17,50 %	35,00 %
Geo2	6,00 %	15,50 %	22,50 %	44,00 %
Geo3	4,00 %	7,00 %	10,00 %	21,00 %
Total	15,00 %	37,00 %	48,00 %	100,00%

— Corrective factors are calculated for each stratum by dividing the stratum weight by the sample proportion (Tables H.7 and H.7a).

Table H.7 — Corrective factors at the stratum level for the standard weighting basis

CF SWB	MC1	MC2	MC3	Total
Geo1	0,60	0,84	0,94	0,86
Geo2	0,83	1,13	1,22	1,14
Geo3	**0,50**	1,00	1,10	0,95
Total	0,67	1,00	1,10	

A corrective factor higher than 1 means, that the stratum is underrepresented. The correction compensates for that undersampling by attributing the stratum a weight higher than its actual weight in the sample. For example, the real mail proportion associated to the stratum defined by the *Geo2* region and the *MC3* mail characteristic is equal to 27,5 % while the sampling proportion is only 22,5 %; therefore a corrective factor of 1,22 is applied to that stratum.

A corrective factor lower than 1 means, that the stratum is overrepresented in the sample. In the case of a fully-proportional stratified sample, all the corrective factors would be equal to 1.

For some modes of MC1 the corrective factors are quite low. If not the standard, but the alternative weighting basis would have been chosen, the corrective factors for MC1 would be less extreme and more levelled.

Table H.7a — Corrective factors at the stratum level for the alternative weighting basis

CF AWB	MC1	MC2	MC3	Total
Geo1	0,67	0,76	0,98	0,86
Geo2	0,67	1,19	1,22	1,14
Geo3	0,67	1,00	1,03	0,95
Total	0,67	1,00	1,10	

H.4.2 Weighting caps

H.4.2.1 Necessity for weighting caps

The weighting process allows restoring proportionality in case a disproportional model has been applied. However, extreme deviations from the real-mail proportions should be avoided. For example, if extreme weights are attributed to very small subsamples, a single failure can have a major impact on the aggregated results and the reliability of the measurement becomes questionable.

This would have been the case for instance in the example presented in H.4.1.6 if the following sample had been selected (instead of the sample presented in Table H.3):

Table H.8 — Example of sample with extreme deviation from the real-mail distribution

Sample	MC1	MC2	MC3	Total
Geo1	100	279	321	700
Geo2	120	420	340	880
Geo3	80	***1***	339	420
Total	300	700	1000	2000

In this example, a single item has been allocated to the Geo3-MC2. Given that in the standard weighting basis a weight of 0.07 has been attributed to that stratum, a failure on that single item would affect the weighted result by 7 % (when that single item only represents 0,05 % of the total sample).

The corrective factors for the standard weighting basis in this example are:

Table H.9 — Corrective factors at the stratum level for the SWB in a case of major deviation

	MC1	MC2	MC3	Total
Geo1	0,60	0,75	1,03	0,86
Geo2	0,83	0,83	1,62	1,14
Geo3	0,50	***140,00***	0,65	0,95
Total	0,67	1,00	1,10	

To a lesser extent, a similar bias would be introduced each time the sampling design deviates too widely from the real-mail proportions.

Weighting caps have been introduced in order to keep that deviation within certain boundaries. They will be applied first at the mode level for each discriminant characteristics, then at the item level.

H.4.2.2 Caps applied at the mode level

That first level of caps aims at preventing the sample to deviate from the required design. Therefore, as the design is only fixed per characteristic (no requirement at the stratum level), each mail characteristic will be considered separately; only marginal proportions (CMW) are taken into account.

Concerned by the caps (i.e. considered as *capping-relevant*) are:

- For geographical characteristics:
  - All the modes
- For the other discriminant mail characteristics:
  - The two main modes
  - All remaining modes with a real-mail proportion greater or equal 17,5 %

Indeed, modes of discriminant mail characteristics with a less than 17,5 % share in real mail volume

- Do usually not have a significant influence on the total QoS if they are not explicitly known for showing a transit-time result that is differing considerably from the general mean value,
- Are not easy to keep on target in the implementation of the sample design.

Therefore as a relaxation these modes are left out of the capping system.

For each capping-relevant mode, the proportion of valid test-mail is not allowed to differ relatively by more than 20 % from the corresponding CMW. For each capping-relevant mode of each discriminant mail characteristic (*m*=1, 2,…, M) that constraint can be expressed as follows, using the same notation as in H.4.1.5:

$$0.80 * CMW_m^k \leq p_m^k \leq 1.20 * CMW_m^k$$

or, using the notion of corrective factors at the mode level, as defined in Formula (H.3):

$$0.8333\ldots \leq Corr_m^k \leq 1.25$$

Table H.10 and Table H.11 show the lower and upper bounds defined by the caps when applying the capping-system to the example developed above. For the geographical characteristic, all the modes are by definition capping-relevant. For the MC characteristic, only the *MC2* and *MC3* modes are capping-relevant (as *MC1* has according to Table H.2 a RMW of 10 % which is lower than 17,5 %).

Table H.10 — Lower and upper bounds for the marginal sampling proportion of the modes of the geographical strata

	CMW	Lower-cap	Upper-cap
Geo1	30 %	24 %	36 %
Geo2	50 %	40 %	60 %
Geo3	20 %	16 %	24 %

Table H.11 — Lower and upper bounds for the marginal proportion of the modes of the MC DMC

	CMW	Lower-cap	Upper-cap
MC1	10 %	Not applied	Not applied
MC2	35 %	28 %	42 %
MC3	55 %	44 %	66 %

The lower and upper bounds for the corrective factors are 0.8333 and 1.25 respectively in all applicable cases.

Both samples in Tables H.3 and H.8 comply with the caps for the corrective factors applied at the mode level (see Tables H.7 and H.9). Still, the corrective factors presented in Table H.9 show extreme deviations at the stratum level. This is the reason why a second level of caps has been fixed.

H.4.2.3 Caps at the item level

To prevent such deviations as presented in Table H.9 (i.e. compliance with the design in terms of RMW, but major deviation at the stratum level) a cap as been fixed at the item level as well.

IFW for each valid test item shall not be smaller than 25 % or higher than 400 % of the value of *1/sample size*. That constraint can be expressed as follows:

$$\frac{0.25}{n} \leq IFW_{i \in J} \leq \frac{4}{n}$$

Given the definition of the corrective factor in Formula (H.5), this constraint can be rewritten as:

$$0.25 \leq Corr_{J} \leq 4 \qquad \text{(H.7)}$$

In other words, the cap applied at the item level can be expressed as a constraint on the corrective factors at the stratum level: for any stratum, the corrective factor shall be between 0,25 and 4.

Based on that second criterion, the sample in Table H.8 would not comply with the standard (see Table H.9 above: the strata corrective factor related to the Geo3-MC2 stratum, being equal to 140, does not fit with the requirement of Formula (H.7)).

Formulae (H.5) and (H.7) can also be used in order to calculate a tolerance interval for each nJ: For each stratum J qualifying for the capping system, the following condition should be respected:

$$0.25 * n * WB_{J} \leq n_{J} \leq 4 * n * WB_{J} \qquad \text{(H.8)}$$

H.5 Reporting of results

H.5.1 Reporting

EN 13850 specifies technical requirements for the measurement including the preparation of reports on the results of the measurement. However, it does not state how and when results should be shared with customers and users as well as it does not state who will order the measurement to take place, and who will pay for the measurement; these subjects are outside the standard and will be determined by legal and regulatory requirements.

Reports shall be available within three months of the end of the measurement period.

NOTE 1 The measurement period can be twelve months for mail flows with large real mail volumes. For medium sized mail flows it can be 24 months. For small sized cross-border mail flows it can be 36 months (see D.3).

The period of the measurement should be stated. The minimum requirement in EN 13850 is to provide figures for the calendar year. It should include a description of any exclusion due to *force majeure*.

NOTE 2 In D.3 coverage of an extended measuring period and reporting of a longer period of time is described.

The measurement period may be shifted to another 12-month period, for example the fiscal year, if agreed with the regulatory authority.

For measurements that are, for example, extended over 36 months as explained in D.3 the reporting of results can be as follows:

— End of year 1 and 2: No reporting of results required,
End of year 3: Reporting of results based on accumulated period from year 1 to year 3,
End of year 4: Reporting of results based on accumulated period from year 2 to year 4,
End of year 5: Reporting of results based on accumulated period from year 3 to year 5,

The report should give the identity of the independent performance monitoring organisation.

The report should describe the discriminant mail characteristics measured in the survey. When interpreting the results from the measurement it is important to consider the national and other peculiarities that apply in each specific case. If operators do not cover all discriminant factors laid down in 6.4.2, this should be clear from the summary of the discriminant mail characteristics covered by the measurement.

Results are presented in terms of on-time performance and cumulative percentage of mail delivered within n days for n = 1 to 10. The total sample size and the accuracy of the results should be given in the report.

If QoS is different for different areas in the field of study, these areas may be treated as study domains. The on-time performance and cumulative percentage of items delivered within n days could then be reported for each study domain. These reports cannot replace the report for the overall field of study.

If results are presented for study domains, geographical regions or periods of the year, information should also be given on the accuracy of those results.

The report should describe any failures to meet the requirements of EN 13850. In the case of failures it should describe how these failures would be overcome in the future.

H.5.2 Archiving

Archiving is an important element of quality control.

All data captured from the panellists will normally undergo various validation checks and procedures which may lead to corrections of the original data or removal of data sets. To make it possible for an auditor to assess the validity of these validation checks, it is essential that the original data has been archived. The most important types of data are the posting and delivery dates as they were observed by the panellists.

Depending on how the panellists do record posting and delivery dates, archiving can be done in different ways (see also C.1.8):

— *Storage of physical test letter document:* The sender and/or the receiver panellists record the observed posting and delivery dates on paper. These forms are returned to the monitoring organisation and stored in an orderly and retrievable manner.

— *Storage of test letter document images:* The sender and/or the receiver panellists record the observed posting and delivery dates on paper. These forms are returned to the monitoring organisation where they are scanned and where the images are stored with their unique ID number in a secure location.

— *Storage of original data sets from panellists:* The sender and receiver panellists record the observed posting and delivery dates electronically (e.g. online on the Internet). The original data records are stored in a secure location.

Data validation is often supported by information taken from the test letter envelopes (e.g. postmark dates etc.). This information can be recorded by the panellists and archived in one of the above ways.

Alternatively, it can also be recorded by the monitoring organisation. In this case the test letter envelopes should be archived in one of the following ways:

— *Storage of physical test letter envelopes:* All test letter envelopes are returned to the monitoring organisation. They are labelled with their unique ID number and stored in an orderly and retrievable manner. If the test letters contained any documents filled out by senders and/or receivers then these documents should be stored together with the envelopes.

— *Electronical storage of test letter envelopes:* All test letter envelopes and inside documents, if available, are returned to the monitoring organisation. They are scanned and the images are stored with their unique ID number in a secure location.

— *Electronical storage of test letter envelope-information:* All test letter envelopes and inside documents, if available, are returned to the monitoring organisation. All relevant information is coded and stored with an ID number unique to the test letter in a secure location.

In cases where the contract with the independent performance monitoring organisation ends before the audit of the measurement period, physical archiving of test-letter documents and test-letter envelopes should at least be done by the contractor for the last month of the contracted measurement period until the end of the audit.

H.6 Audit

H.6.1 General

This section refers to the audit of the measurement system including the design basis.

The aim of the audit is to verify the compliance of the design basis and the test mail measurement system and its processes to the standard requirements. The audit report indicates whether the measurement as performed by the independent performance monitoring organisation is compliant to the standard and it reports the level of compliance.

The field of investigation for the audit includes the independence of the measurement system, the test mail study's methodology, the adequacy of the geographical stratification, the weighting method and calculation rules which lead to the measurement figures, the panel management, the integrity of the measurement and the implemented quality control procedures.

H.6.2 Position of the auditor

It is the responsibility of the postal operator / regulatory authority to indicate the requirements for the measurement system. It is also the responsibility of the postal operator and/or regulatory authority to ensure that the requirements of the measurement system are in line with this standard. The responsibility of the independent performance monitoring organisation is to implement and run the measurement system as requested by the postal operator and/or regulatory authority.

The auditor has a central position in between all parties. The auditor would not only investigate the compliance of the requirements of the measurement system as set out by the postal operator / regulatory authority, but would also investigate the compliance of the performance monitoring organisation's processes and activities to the requirements for the measurement system (and thus indirectly to this standard).

To make sure that an audit can be performed correctly, all parties concerned (independent performance monitoring organisation and postal operator and/or regulatory authority) should provide all documentation and information required by the auditor. Appropriate documentation should be part of the measurement system right from the start of the measurement period.

The documentation for the audit is provided in commercial confidence and all parties would need to formally respect this. In some instances, a review of documentation only would not be sufficient for an auditor to fully assess the compliance to the standard. For the panel audit the auditor should also have the possibility e.g. to listen to live panellist interviews and/or contact panellists for the purpose of checking training procedures and instructions. The means chosen for the auditing should however be fully in line with the aim of the audit. All means of auditing chosen (in this case the interview of panellists) have to be in line with and comply with national legislation.

H.6.3 Audit report

The auditor finalises the audit by providing an audit report. This report contains the outcome of the audit and should be handed over to the postal operator, the regulatory authority and the independent performance monitoring organisation.

The audit report describes in detail all areas of non-compliance to the requirements of this standard and possible remedial action. The auditor and all parties concerned should work out a schedule to solve the shortcomings and prepare for a corrective audit.

The measurement system in total is considered compliant to the standard when the auditor decides that the normative requirements from the standard are met. Only the requirements applicable to the specific type of measurement (e.g. field of study) are to be considered during the audit.

H.6.4 Selection of the auditor

The auditor should not be chosen among the competitors of the independent performance monitoring organisation.

For measurements under the responsibility of a single regulatory authority the auditor should be chosen in agreement with the regulatory authority.

In case of cross-border measurements, the organisation responsible for the measurement should select the auditor, in agreement with the postal operators and approved by the regulatory authority.

H.6.5 Frequency of audit

In case of multi-years contracts, the audit should take place within the first 18 months of running.

In case of compliance:

— A re- audit should take place within 3 years after the previous audit if no major changes occur and if not requested otherwise by the regulatory authority.

— This re-audit should focus on aspects and procedures that have changed since the previous audit. It should also focus on areas where an update is to be expected.

In case of non-compliance:

— A corrective audit should take place within 6 months.

— This corrective audit needs to cover the areas of non-compliance.

In case of a one year contract for the measurement, the initial audit should start in the first six months of the measurement period and finish after the end of the measurement period. In case of compliance for a one-year contract, no additional audit is necessary.

NOTE If a one year contract is prolonged by a follow-up contract within the first contract period, the series of contracts can be handled in the same way as a multi-year contract.

H.7 Implementation timetables

In principle, a number of steps are required in order to implement QoS measurement in accordance with EN 13850. They follow the stages of the survey as laid down in H.1.

In this annex the time required for each step is assessed and the total time needed to implement the standard, from the first planning session to the report of the first measurement period, is described for three exemplary cases:

— complete system implementation (Small to medium sized real mail and test mail domestic measurement systems),

— complete system implementation with a parallel run of the first real mail and test mail measurement periods (some sort of End-to-End test mail system is already in place but no real mail measurement system)

— system implementation of EN 13850 (with minor modifications to existing real mail and test mail measurement systems),

Each of the three cases is presented in form of a time line to illustrate the implementation plan (see Figures H.12-H.14).

Table H.12 — Time line for complete system implementation of small to medium sized real mail and test mail domestic measurement systems

ID	**Task name for Case I**	**Duration**	**Year 1**		**Year 2**				**Year 3**				**Year 4**				**Year 5**
	Complete Set-up	[weeks]	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr 1
	Total: Case I	**130**		**P**		**C**			**[-------**	**--------**	**--------**	**-------]**	**R**				
1	Test mail survey Planning phase	13		P													
2	Real mail studies Planning & Set-up	13															
3	Real mail studies Pilot phase	13			[------]												
4	Final adjust RM studies operations	13															
5	Selecting Survey operator + Contract	26				C											
6	Final adjustment of Survey design	13															
7	Running real mail studies (cont.)	13				[------]											
8	Survey Set-up phase	13															
9	Survey Pilot phase (test phase)	26					[-------	-------]									
10	Final adjust of survey operations	26															
11	Measurement period [---]	52							[-------	--------	--------	-------]	[-------	--------	--------	-------]	[--------
12	Data analysis & Reporting	52											R				R

The whole time line spans from the beginning of the planning phase over the contact with the monitoring organisation to the report of the first measurement period. (see also H.1 Stages of the survey and here especially H.1.1-H.1.3 and H.1.5). This time line describes a timing of 10 quarters in total, assuming that all phases are running without major difficulties. This timeline is the minimum timeline that has to be taken into account when setting up a completely new system.

Table H.13 — Time line for system adjustment without existing real mail measurement with a parallel run of the first real mail and test mail measurement periods

ID	Task name for Case II	Duration	Year 1		Year 2				Year 3				Year 4				Year 5
	QoS-Measurement in place w/o real mail	[weeks]	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr 1
	Total: Case II	**130**		**P**	**C**				**[-------**	**--------**	**--------**	**-------]**	**R**				
1	Test mail survey Planning phase																
2	Real mail studies Planning & Set-up	26		P													
3	Adjustment of Survey design	26															
4	Agree survey Contract adjustments	26			C												
5	Real mail studies Pilot phase	39			[-------	--------	-------]										
6	Survey Adjustment phase	26															
7	Final adjust RM studies operations	13															
8	Survey Pilot phase (test phase)	26					[-------	-------]									
9	Final adjust of survey operations	26															
10	Running real mail studies (cont.)	52						[-------	--------	--------	-------]	[-------	--------	--------	-------]	[-------	---------
11	Measurement period [---]	52	--------	-------]	[-------	--------	--------	-------]	[-------	--------	--------	-------]	[-------	--------	--------	-------]	[--------
12	Data analysis & Reporting	52											R				R

The whole time line spans from the beginning of the planning phase over the contact with the monitoring organisation to the report of the first measurement period. (see also H.1 Stages of the survey and here especially H.1.4.2). This second time line describes the situation of time line 1 shortened to 10 quarters by modifying an existing test mail measurement system and using a parallel run of the first real mail and test mail measurement periods designed according to this standard. During the first measurement period the test mail measurement runs on alternative real mail estimates.

Table H.14 — Time line for system implementation of EN 13850 with minor modifications to existing real mail and test mail measurement systems

ID	**Task name for Case III**	**Duration**	**Year 1**		**Year 2**				**Year 3**				**Year 4**				**Year 5**
	Minor Modifications	[weeks]	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr1	Qtr2	Qtr3	Qtr4	Qtr 1
	Total: Case III	**91**	**A C**						**R**								
1	Test mail survey Planning phase																
2	Real mail studies Planning & Set-up																
3	Adjustment of Survey design	13	A														
4	Agree survey Contract adjustments	13	C														
5	Adjustment RM studies operations	26															
6	Survey Adjustment phase	26															
7	Real mail studies Pilot phase																
8	Survey Pilot phase (test phase)																
9	Final adjust of survey operations																
10	Running real mail studies (cont.)	52	-------]	[-------	--------	--------	-------]	[-------	--------	--------	-------]	[-------	--------	--------	-------]	[-------	---------
11	Measurement period [---]	52	--------	-------]	[-------	--------	--------	-------]	[-------	--------	--------	-------]	[-------	--------	--------	-------]	[--------
12	Data analysis & Reporting	52							R				R				R

The whole time line spans from the beginning of the adjustment phase over the review of contact with the monitoring organisation to the first conforming report. (see also H.1 Stages of the survey and here especially H.1.4.3). This third time line describes the situation where real mail and test mail measurement systems are already in place and only minor modifications have to be conducted together with an existing contractor to make the measurement system conform to the standard.

Annex I
(informative)

Application of the accuracy calculation

I.1 Limitations of the accuracy calculation methods provided

I.1.1 Participants with high mail loads

The accuracy calculation methods provided were developed for the measurement of Single Piece Priority Mail (SPPM). This measurement includes inductions and deliveries that comprise a limited number of test items per induction and delivery.

For fields of study where the size of inductions increases, for example for the measurement of Bulk Mail, or where the size of deliveries increases, for example for the measurement of Response Services, the calculation methods are not applicable without modifications.

Studies on a number of existing bulk mail measurement systems have shown, that for them temporal correlations have to be taken into consideration as well, for example correlations between test items that are inducted on the same weekday. The size of the design factors that have to be expected in non single-piece systems will increase considerably.

The measurement of non single-piece systems will be dealt with in adapted versions of this standard, for example EN 14534 Measurement for the transit time of end-to-end services for bulk mail.

I.1.2 Disproportional models beyond the capping system

The accuracy calculation methods provided were developed for a proportional sample. The standard in general allows for a disproportional approach for the discriminant mail characteristic “geography” only [see 7.3.1.1].

To keep the biasing effects due to cases of non-response or invalid test items low, a system of weighting caps is introduced that demands the maintenance of the proportionality of the model [see 7.3.2].

The calculation methods work well as long as the weighting caps in 7.3.2 are respected. With extreme weighting systems that do not completely respect the weighting caps the approximations tend to be conservative, i.e. the approximated design factor tends to be larger than the true design factor. Disproportional models have therefore a higher risk of overestimating the design factor thus producing a higher bias in the calculation of the variability of the measurement results.

I.2 Recommendations for the application of the rules

I.2.1 Unstratified end-to-end sample

The calculation of the accuracy of the unstratified end-to-end sample gives insights into the loss of accuracy due to correlation effects between test items. These correlations might exist between items sent at the same induction point, received at the same delivery point or distributed via the same induction-delivery point relation.

The loss of accuracy is measured in form of a design factor (DF). A measurement system with a DF of 4.0 loses the double amount of accuracy due to correlation in comparison with a measurement system with a design factor of 2.0. A system with a DF of 1.0 shows no correlation effects at all.

The calculation of the accuracy of the unstratified end-to-end sample can also be used to identify the magnitude of the three sources of correlation [see terms [R], [I] and [D] in A.3.2]. These insights should be used to improve the design by allowing for more induction or delivery points.

Since the accuracy of an unstratified end-to-end sample depends on the amount of existing correlation in the measurement period, its size cannot be planned completely before the start of the measurement. Any optimisation of the accuracy level is an ongoing design calibration process.

I.2.2 Stratified simple random sample

Weighting is a common statistical method for correcting surveys when the distribution of valid samples obtained does not correspond to the fully proportional design.

EXAMPLE A survey of 10 000 items was planned, with 17 % to be posted on a Monday, 18 % on Tuesday, 20 % on Wednesday, 20 % on Thursday, 20 % on Friday and 5 % on Saturday (these proportions define the weighting basis (WB) to be applied). Actually 9.374 letters were analysed with the shortfall varying from day to day.

To ensure that the results represent each day of the week in the right proportion an Individual Final Weight (IFW) is calculated for each item, depending on the day of the week. This IFW is obtained by dividing the weighting basis corresponding to that day by the number of valid items on that day. For example, for an item posted on a Monday the IFW would be equal to 17 %/1612 = 0,010546 %.

Table I.1 — Example of a stratified sample

	Design volume	Valid volume	On-Time	Proportion OT	Item Weight
Monday	1 700	1 612	1 485	92,12 %	0 010546 %
Tuesday	1 800	1 750	1 653	94,46 %	0 010286 %
Wednesday	2 000	1 880	1 722	91,60 %	0 010638 %
Thursday	2 000	1 940	1 799	92,73 %	0 010309 %
Friday	2 000	1 742	1 450	83,24 %	0 011481 %
Saturday	500	450	381	84,67 %	0 011111 %
Total	10 000	9 374	8 490	90,57 %	

The total weighted performance is calculated by attributing to each item the IFW corresponding to the day it was posted and by dividing

— The total of the IFW corresponding to successful items by

— The total of the IFWs for the whole sample. Hence:

$$Weighted - OTP = \frac{1485 * 0.00010546 + 1653 * 0.00010286 + \dots}{1612 * 0.00010546 + 1750 * 0.00010286 + \dots} = 94.41\%$$

The weighted performance can also be calculated by weighting the performances measured on each weekday with the corresponding WB. Using that method:

$$Weighted-OTP = 17\%*92.12\%+18\%*94.46\%+20\%*91.60\%+\ldots = 94.41\%$$

Both methods are equivalent.

Restoring the proportionality of the sample means that the contribution, which each part (stratum) of the sample makes to the total result, gets exactly the importance related to the design basis.

In any weighting process it might happen that the weights get very small or very big. Very small IFWs might reduce the contribution of the concerned items to a degree where the contribution does not justify the cost of sampling. Very big IFWs might increase the contribution of single items to a degree, where one delayed item may have a considerable impact on the total result. Both scenarios are not desirable. Therefore, caps are imposed on the weighting system to keep the size of the IFWs on an appropriate level (see 7.3.2).

If the sampling is done on a continuous basis, it should be adapted for cases of extreme weighting already during the measurement period.

I.2.3 Approximation of the Binomial distribution

Generally an approximation method available should be chosen that delivers the smallest possible bias. But, of course, questions of applicability, ease of handling and acceptance by the reader are important factors as well.

From the bias point-of view the **Inverse Beta approximation** (A.5.4) is to be preferred. It shows the smallest bias, although it is also the most conservative method, i.e. resulting in the widest confidence interval of the three methods described. It utilises the Inverse Beta function for which software with statistical applications is required.

Similar to the Inverse Beta is the **Agresti-Coull approximation** (A.5.3). The bias is slightly higher; is slightly less conservative. It can be re-calculated with a pocket-calculator by the reader of the report and is therefore transparent in its application, but the calculation has several steps.

Both approximations result in an unsymmetrical confidence interval, usually unfamiliar to the reader.

The **Normal approximation** (A.5.2) is well known and applied. It is easy to calculate in one step (see A.5.2.1). It is symmetrical and has a commonly known structure (x.x % ± y.y %). Unfortunately scenarios exist in which the bias gets out of hand and where it should not be used any more (see A.5.2.2)

I.2.4 Accuracy

The accuracy of a measurement is described by the width of the confidence interval of the estimate regarding the QoS-indicator of interest. *The smaller the width, the higher the accuracy.* The confidence interval covers the true QoS-indicator with a probability of 95 %. The estimate $\hat{p}$ for the QoS indicator p is the most probable position within the interval.

For a symmetrical confidence interval its width can be described in the form (±x %). For an unsymmetrical interval the width is (-x1 %;+x2 %).

It is possible to define lower bounds for the accuracy in form of a tolerable maximal width of the confidence interval. It can have the form $\pm\epsilon$ or 2ϵ with an epsilon of choice.

The minimum sample size MSS required by this standard (see Table 1 and Table 2, 6.2) is set up to tolerate an epsilon of 1 % (domestic ϵ) and 5 % (cross-border ϵ) in cases where the design factor can be assumed to be 1. The calculation of the MSS tables is based on the Agresti-Coull approximation (see A.5.3).

Design factors beyond 1 will lead to higher epsilon values, approximately to a degree of $\epsilon * \sqrt{df_{StrEtE}}$.

I.2.5 Accuracy application

It is often of interest, how the estimate of the QoS-indicator, i.e. the measurement result, compares to a given value, for example a quality target. In the implementation of such a comparison, the accuracy plays a key role. Based on the accuracy, i.e. on the confidence interval around the QoS-estimate, it can be determined, which scenario applies:

— Scenario I: With a probability of at least 95 % the measurement result lies below a given QoS-value / target.

— Scenario II: With a probability of at least 95 % the measurement result lies above a given QoS-value / target.

— Scenario III: No decision on the relative position of measurement result and given value / target can be made with at least 95 % security.

The following examples provide an illustration how scenarios and accuracy are related:

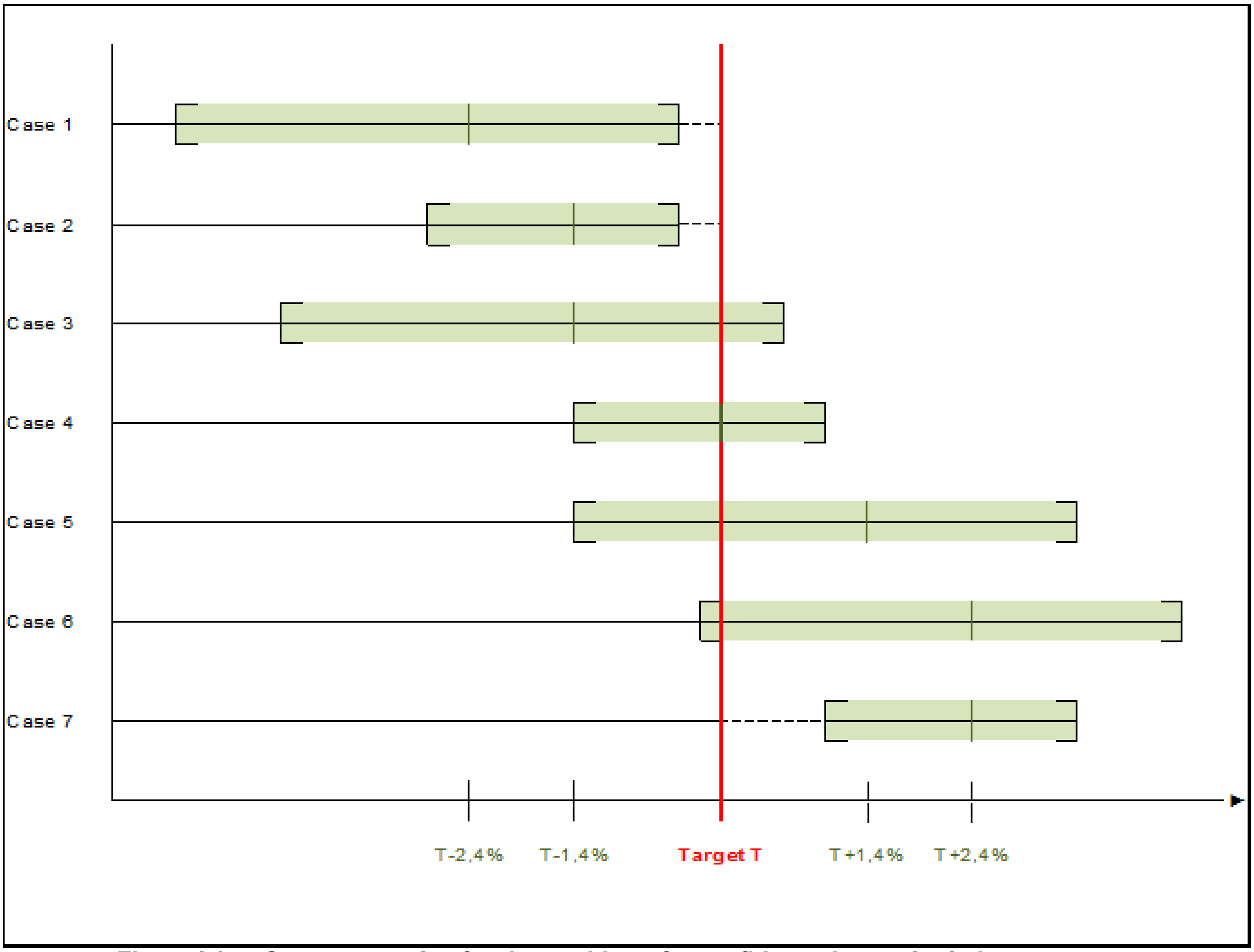


Figure I.1 — Seven examples for the position of a confidence interval relative to a target

Cases one and two fall in scenario I ("below the given value"). The measurement result in case two is nearer to a given value / target, but it keeps distinguishable because of the higher accuracy.

Case seven falls in scenario II ("above the given value").

Cases three to six fall in scenario III. Here the accuracy is too low to differentiate between given value / target and measurement result, since both results are close to each other. In case four the measurement result falls on the given value / target, showing the highest probability that result and given value / target are actually the same.

A QoS-target usually requires an indication whether the measurement result supports the statement that the target has not been fulfilled. Based on scenario I one can state with 95 % statistical security:

An end-to-end transit time target has been missed if the target value falls above the 95 % confidence interval.

I.3 The sample size

A key figure for the set-up of the test mail measurement system is the required Minimum Sample Size (MSS), which is fixed in 6.2. Beyond the requirements of this standard (see 6.2) it may be desirable to recalculate the MSS into an adapted MSS_{adapt} in order to secure certain measurement accuracy. A calculation method for the MSS_{adapt} is provided in Formula (A.10) for samples, where the *Normal confidence interval* can be applied (see also A.5.2.2):

$$MSS_{adapt} := \min n^* \quad \text{with} \quad n^* \geq \mathrm{df}_{\mathrm{StrEtE}} * \left[3.84145 * \frac{\hat{p}_{weighted}(1-\hat{p}_{weighted})}{\epsilon^2}\right] + 1$$

As can be seen, the MSS_{adapt} depends on three parameters, the

— Design factor of the measurement design df_{StrEtE},

— Weighted estimate $\hat{p}_{weighted}$ of the true on-time probability p and

— Required accuracy ϵ.

EXAMPLE Let $\hat{p}$ be expected to be 89 % and the design factor to be 1. An accuracy of 1 % is required. Then the MSSadapt is:

$$MSS_{adapt} := 3762 \geq 3761.780 = 1 * \left[3.84145 * \frac{89\% \, (1 - 89\%)}{0.0001}\right] + 1$$

The adapted minimum sample size is directly proportional to the design factor. The design factor mainly depends on the number of induction and delivery points used, the more are taken in, the lower the design factor gets. This fact creates a trade-off situation between panel size and sample size: The bigger the panel, the more induction and delivery points can be used, the lower the design factor gets and with it the sample size.

Usually the design factor is not known prior to the implementation of the measurement system and may vary significantly between measurement systems. A sample of European domestic transit-time measurement systems already in place showed on *average design factors between 5 and 6,* which could be used as an initial starting point.

The true on-time probability is also not known. Here either a robust guess from experience could be used or – to be on the safe side – 50 %. After the first measurement period the sample size calculations for the next periods can be based on the results for p and df from the first period.

The MSS does in general not depend on the choice of the discriminant mail characteristics, although they may have an influence on the transit-time result and the design factor.

I.4 General Example for a national yearly result

I.4.1 Introduction

— Let us assume a panel of four senders S1 - S4 and four receivers R1 - R4. 80 letters are sent from four induction points to four delivery points. Each sender is sending 20 test letters. Each receiver receives 20 test letters.

NOTE This is the same example as in the previous version of this standard.

We assume that all test letters are sent on three induction days (Monday, Tuesday and Wednesday). The measurement is based on a five-day working week and the official calculation rule. It leads to transit times from J+1 to J+3, resulting in three delivery days (Tuesday to Friday and Monday to Tuesday).

Table I.2 — Mail-flow matrix from panellist S1-S4 to panellist R1+R2

Se/Re		Panellist R1						Panellist R2						Sum
	day	Tu	We	Th	Fr	Mo	Tu	Tu	We	Th	Fr	Mo	Tu	
Panel-list S1	Mo	2	**1**					-	**1**					4
	Tu		2						1					3
	We			2	-	-	**1**			1				4
Panel-list S2	Mo	1	**1**					1	**1**					4
	Tu		1	-	**1**				-	**1**	**1**			4
	We			1	**1**					1	**1**			4
Panel-list S3	Mo	1						1						2
	Tu								2					2
	We			1						1				2
Panel-list S4	Mo	1						2	**1**					4
	Tu		1	-	**1**				2					4
	We			1						1	-	**1**		3
Sum		5	6	5	3	0	1	4	8	5	2	1	0	40

The mail-flows from all induction points to the first two delivery points can be found in Table I.2, to the last two delivery points in Table I.3. On-time J+1-relations are marked as light-grey areas. Delayed test letter are marked in bold.

Table I.3 — Mail-flow matrix from panellist S1-S4 to panellist R3+R4

Se/Re		Panellist R3						Panellist R4						Sum
	day	Tu	We	Th	Fr	Mo	Tu	Tu	We	Th	Fr	Mo	Tu	
Panel-list S1	Mo	2						1						3
	Tu		1	-	**1**				1					3
	We			2						-	-	-	**1**	3
Panel-list S2	Mo	1						1	**1**					3
	Tu								1	**1**				2
	We			1						1	**1**			3
Panel-list S3	Mo	2						2						4
	Tu		2	**1**					2	-	**1**			6
	We			2						2				4
Panel-list S4	Mo	2						1						3
	Tu		1						1	**1**				3
	We			2						1				3
Sum		7	4	8	1	0	0	5	6	6	2	0	1	40

Let us further assume that there is a geographical stratification with two induction regions (I1-I2) and two delivery regions (D1-D2).

— Senders S1 and S2 are in region I1,

— Senders S3 and S4 are in region I2.

— Receivers R1 and R2 are in region D1,

— Receivers R3 and R4 are in region D2.

The resulting four strata are:

- Stratum 1: I1-D1
- Stratum 2: I1-D2
- Stratum 3: I2-D1
- Stratum 4: I2-D2

For the accuracy calculation according to Annex A we need the input parameters listed in Table I.4.

I.4.2 Design factor for an unstratified end-to-end sample

The calculation of the design-factor for an unstratified end-to-end sample follows the procedure denoted in A.3.

The variance of the example calculated as an unstratified End-to-End sample is, according to Formula (A.1), the sum of the Relation-to-Total variation and the Inter-Relation variation.

The **Relation-to-Total variation** is the variation between the mean results of different induction-delivery point links. In this example it is $\frac{1}{80^3} * 34.317$. It is composed of three parts:

- **[I]** Contribution of the induction point: $\frac{1}{80^2} * 19.567$
- **[D]** Contribution of the delivery point: $\frac{1}{80^2} * -0.433$
- **[R]** Contribution of the relation itself: $\frac{1}{80^2} * 15.183$

Table I.4 — Input parameters for the variance calculation

Group				I.4.2.1 Sample Size				On-Time			
Sender	**Receiver**	**Stratum**	**Relation**	n_a	n_b	n_s	n_{ab}	x_a	x_b	x_s	x_{ab}
S1			S1-R1	20			8	13			6
S2			S1-R2	20			3	12			2
S3			S1-R3	20			6	14			5
S4			S1-R4	20			3	20			2
	R1		S2-R1		20		6		15		3
	R2		S2-R2		20		6		10		2
	R3		S2-R3		20		2		18		2
	R4		S2-R4		20		6		16		3
		I1-D1	S3-R1			23	2			14	2
		I1-D2	S3-R2			17	4			13	4
		I2-D1	S3-R3			17	7			18	6
		I2-D2	S3-R4			23	7			14	6
			S4-R1				4				3
			S4-R2				7				5
			S4-R3				5				5
			S4-R4				4				3

The results for [I], [D] and [R] can be interpreted in the following way:

The highest variability is between the mean results of different induction points and of different induction-delivery point relations (19.576 and 15.183). Variability between delivery points is small (-0.433).

This information can be used to decide how to decrease the variability most efficiently in future samples.

In the above example one could increase the number of induction points to reduce the existing *induction point correlation* and therefore to minimise the design factor (resulting in an increased panel size of the sender panel).

With the new senders one could furthermore improve the test letter allocation plan which defines the letter allocation to individual induction-delivery point links. The improvement could be done by decreasing the existing clustering of test letters on the same *induction-delivery point relations* (resulting in an increased spread of test letters over the existing relations).

Apart of [I], [D] and [R] a fourth and last source for variation remains:

— **[IRV]** The Intra-Relation variation $\frac{1}{80^2} * 15.581$

The IRV is the sum of the variability within each induction-delivery point link (‘Intra-Relation’).

The measured *performance level* is $\hat{p} = 73.750\%$.

According to Formula (A.5), the variance of a simple random sample with the same performance p would be 0.245%. The design-factor for the End-to-End unstratified sample then is:

$$
\begin{aligned}
df_{EtE} &= \frac{1}{80^2} * (19.567 - 0.433 + 15.183 + 15.581) * \frac{1}{0.245\%} \\
&= \frac{49.898}{6400 * 0.00245} = 3.1815
\end{aligned}
$$

This design needs about 3 times more test-letters to achieve the same accuracy a simple random sample (SRS) which uses each sender and receiver only once. The SRS would require 20 times more panellists.

I.4.3 Design factor for a stratified random sample

Let us assume that in reality 60 % of the mail is inducted in region I1 and 40 % in I2. The same holds for the delivery areas. 60 % of the mail is delivered in region D1 and 40 % in D2. This leads to the standard weighting basis:

Table I.5 — Standard weighting basis

N_s/N	D1	D2	Total
I1	36 %	24 %	60 %
I2	24 %	16 %	40 %
Total	60 %	40 %	100 %

The standard weighting basis may be simplified into a weighting basis where three of four strata have the same weight. Both, the standard and the simplified weighting basis fulfil the 60 % / 40 % marginal distributions.

Table I.6 — Simplified weighting basis

N_s/N	D1	D2	Total
I1	40 %	20 %	60 %
I2	20 %	20 %	40 %
Total	60 %	40 %	100 %

According to A.4.2 the matrix of corrective factors for the simplified weighting basis can then be calculated as:

Table I.7 — Corrective factors

	D1	D2	Marginal
I1	139 %	94 %	83 %
I2	94 %	70 %	125 %
Marginal	83 %	125 %	

The capping system requires that the marginal corrective factors for the modes of the discriminant mail characteristics lie between 83 % and 125 % and that the corrective factors for the strata lie between 25 % and 400 %. Both restrictions are fulfilled in this example.

The variance of the stratified random sample is calculated according to A.4.2. The results are shown in Table I.8. The weighted on-time performance level is: $\hat{p}_{weighted} = 70.588\%$

Table I.8 — Variance of the stratified sample * 80²

	D1	D2	Total
I1	11,438	3,322	14,760
I2	2,325	1,320	3,645
Total	13,763	4,642	18,405

According to A.4.1 the variance of a simple random sample with the same weighted performance would be:

$$\frac{\hat{p}_{weighted} * (1-\hat{p}_{weighted})}{(n-1)} = \frac{70.588\% * 29.412\%}{79} = 0.263\%$$

Following from this, the design-factor for the stratified sample is:

$$\begin{aligned} df_{StrRS} &= \frac{1}{80^2} * (11.438 + 3.322 + 2.325 + 1.320) * \frac{1}{0.263\%} \\ &= \frac{18.405}{6400 * 0.00263} = 1.0943 \end{aligned}$$

A good stratification usually has a design factor near 1. If the significance of the discriminant mail characteristics is high, the design factor of the stratified sample may even be smaller than 1.

I.4.4 Accuracy calculation

I.4.4.1 General

The design factor for the stratified End-to-End sample is according to Formula (A.8):

$$df_{StrEtE} = df_{EtE} * df_{StrRS} = 3.1815 * 1.0943 = 3.4815$$

If the sample size were calculated under the assumption of a simple random sample (SRS), the actual design would require *about* 3.4815 *times more test letters* to achieve the same accuracy.

This simple logic only works if the design factor can be kept stable when switching from the smaller to the enlarged design.

The design factor depends on the number of test letters per induction and delivery point and furthermore on the spread of test letters over the existing induction-delivery point relations. In this example, the number of induction points and the spread are important sources of variation . To stabilise the design factor, the enlarged design therefore not only need more items, it needs more induction points, too.

If z times more induction points are used and the associated new induction panellists could also be used for receiving letters, the number of existing induction-delivery point relations would grow approximately by the factor z^2.

Resulting from this fact, the *stabilising panel enlargement factor* will lie in the range of $\sqrt{df_{StrEtE}} = 1.8659$ (same amount of letters per induction-delivery point relation) up to 3.4815 (same amount of letters per induction point).

I.4.4.2 Normal confidence interval

With this design factor the Normal confidence interval can be calculated according to Formula (A.9):

$$\left[70.588\% \pm 1.95996 * \sqrt{0.263\% * 3.4815}\right] = [70.588\% \pm 18.748\%] = [51.841\% ; 89.336\%]$$

This leads to an accuracy of $2\epsilon = 37.495\%$.

According to Formula (A.10), the minimum sample size associated to the weighted on-time performance would be $n_{minSS} = 3087$ if a minimum accuracy of, for example, $\epsilon = \pm 3\%$ is required and if the normal confidence interval is used for calculation. This also requires, as mentioned above, that the level of correlation is not rising i.e. that the design factor is kept stable.

To achieve this accuracy, a minimum panel size in the range of $\sqrt{\frac{n_{minSS}}{n}} = \sqrt{\frac{3087}{80}} * 4 = 24,85$ up to $\frac{3087}{80} * 4 = 154,35$ senders and receivers would be necessary.

For the use of the Normal confidence interval with a performance of $p_{weighted} = 70.588\%$ at least 30 non-performance items would be necessary according to Table A.1 (A.5.2.2). In our example we only have 21, therefore the use of the Normal confidence interval is not recommended.

I.4.4.3 Alternative confidence intervals

The first alternative confidence interval is the *Agresti-Coull interval*. For the calculation one first has to determine the effective sample size (see Formula (A.11)). In the example we have:

$$\frac{80}{3.4815} = 22,979; \quad n_{ESS} = 22 < 22.979 < 23; \quad 22{*}0,70588 = 15.53; \quad x_{ESS} = 15 < 15.53 < 16;$$

$$\hat{p}_{acESS} = \frac{15 + 1.92072}{22 + 3.84145} = 65.479\%$$

This means that the existing sample with 80 letters has the same accuracy that a simple random sample with 22 letters would have. The midpoint of the new Agresti-Coull interval is $\hat{p}_{acESS} = 65.479\%$

With these parameters the Agresti-Coull interval can be calculated around $\hat{p}_{weighted} = 70.588\%$ (not centred):

$$\left[65.479\% \pm 1.95996 * \sqrt{\frac{65.479\% * (1 - 65.479\%)}{22 + 3.84145}},\right] = [65.479\% \pm 18.331\%] = [47.148\% ; 83.810\%]$$

This leads to an accuracy of $2\epsilon = 36.662\%$.

For the use of the Agresti-Coull confidence interval a minimum of $n_{ESS} \geq 40$ would be necessary. In our example we only have 22 letters, therefore the use of the Agresti-Coull confidence interval is also not recommended.

The last of the alternative confidence intervals is the *Inverse Beta approximation*. The approximation can be directly calculated using the effective sample size:

$$[BetaInv[0.025; 15; 8] \; ; \; BetaInv[0.975; 16; 7]] = [45.128\% ; 86.135\%]$$

This leads to an accuracy of $2\epsilon = 41.008\%$.

With such a small sample size one sees that only the Inverse Beta approximation leads to a confidence interval that is conservative enough:

Table I.9 — Comparison of confidence intervals

	Lower Limit	Performance	Upper Limit	Accuracy
Normal	51,84 %	70,59 %	89,34 %	37,50 %
Agresti-Coull	47,15 %	70,59 %	83,81 %	36,66 %
Inverse Beta	45,13 %	70,59 %	86,14 %	41,01 %

The table shows that the lower limit is decreasing when the fit of the interval improves.

I.5 Simplified scenarios

I.5.1 General

The accuracy calculation method of Annex A is a complex tool that has a wide field of application in all sorts of end-to-end scenarios. Not all of these scenarios require the full range of complexity offered in Annex A.

I.5.2 Transit time results up to 96 %

For domestic measurement systems with performance levels up to 96 % the Normal confidence interval can be used. In this case, usually at least 50 delayed items are registered by the measurement system.

The Normal confidence interval is convenient since it is:

- Well known and applied,
- Easy to calculate (even by hand, see A.5.2.1) and
- Symmetrical (±X.X %).

The alternative confidence intervals of A.5.3 and A.5.4 are provided for extreme results beyond 96 % and for sub-results of domestic and cross-border measurements with small sample sizes.

I.5.3 Fully proportional sample

In case of a fully proportional sample the weighting is greatly simplified. The individual final weight is: $\frac{1}{sample\ size}$ for all items. The design factor of a proportional end-to-end sample is simply the variance of an end-to-end sample divided by the variance of a Simple Random Sample SRS (see Formula (A.7)).

To achieve a fully proportional sample one requires:

- A fully proportional design,
- A valid sample that does not differ from the proportions fixed in the design.

To approximate a fully proportional sample in a proportional design a flexible mail allocation process has to be installed to continuously adjust the existing valid sample.

I.5.4 Single induction / delivery point

The accuracy calculation method is not restricted to results from SPPM measurement systems only. It can be applied to many forms of End-to-End measurement systems.

For example, in bulk mail scenarios the induction volume of a single induction panellist might be big enough to make a sub-result reasonable for the service provided to this panellist. In this case one has only to deal with a single induction point. If each delivery point receives at least two letters Formula (A.1) simplifies to:

$$\widehat{\mathrm{Var}}_{\mathrm{EtE}}[\hat{p}] = \frac{1}{n^2}\left[\sum_b n_b \frac{x_b}{n_b - 1}\left(1 - \frac{x_b}{n_b}\right)\right]$$

In case each delivery point receives only one letter, we have a simple random sample (SRS).

If only some of the delivery points receive exactly one letter and some more than one letter one has to introduce a special solution for the group of letters on one-letter relations (OLR). Formula (A.1) then becomes:

$$\widehat{\mathrm{Var}}_{\mathrm{EtE}}[\hat{p}] = \frac{1}{n^2}\left[n_{OLR}\frac{x_{OLR}}{n_{OLR} - 1}\left(1 - \frac{x_{OLR}}{n_{OLR}}\right) + \sum_{b,\, n_b > 1} n_b \frac{x_b}{n_b - 1}\left(1 - \frac{x_b}{n_b}\right)\right]$$

The same reasoning is applicable for fields of study with only one delivery point, for example in case of response service measurements. Here Formula (A.1) becomes:

$$\widehat{\mathrm{Var}}_{\mathrm{EtE}}[\hat{p}] = \frac{1}{n^2}\left[n_{OLR}\frac{x_{OLR}}{n_{OLR} - 1}\left(1 - \frac{x_{OLR}}{n_{OLR}}\right) + \sum_{a,\, n_a > 1} n_a \frac{x_a}{n_a - 1}\left(1 - \frac{x_a}{n_a}\right)\right]$$

I.5.5 Induction / delivery point with only one letter

In case of an accuracy calculation for outbound sub-results, for example long-distance letters from a special induction area, it is possible that certain delivery points receive only one test letter. This has consequences for the calculation of the intra-relation variation for one-letter cases (last sum of Formula (A.1)). One has two possibilities here:

— To go beyond the sub sample and use the delivery-point items from the total sample. (If $n_b = 1$ in the sub-sample it might be greater than one in the total sample). In this case the formula can be kept unchanged.

— If one is restricted to the sub-sample one has to adapt the last sum of Formula (A.1) as follows:

$$\sum_a \sum_b {}_{n_{ab}=1,\, n_b>1} \left\{\frac{1}{4n_a}\frac{x_a}{(n_a-1)}\left(1 - \frac{x_a}{n_a}\right) + \frac{1}{4n_b}\frac{x_b}{(n_b-1)}\left(1 - \frac{x_b}{n_b}\right)\right\}$$
$$+ \sum_a {}_{n_{ab}=1,\, n_b=1} \left\{\frac{1}{n_a}\frac{x_a}{(n_a-1)}\left(1 - \frac{x_a}{n_a}\right)\right\}$$

In case of an accuracy calculation for inbound sub-results, for example long-distance letters to a special induction area, the same reasoning is applicable.

Annex J
(informative)

Changes to the 2007 version of EN 13850

J.1 Methodology

J.1.1 Accuracy and Minimum Sample Size (MSS)

In the 2002 version of EN 13850 the requirements for the Minimum Sample Size (MSS) were given in terms of accuracy requirements for domestic and cross-border measurement systems. Basis of this layout was an accuracy calculation method which linked any accuracy requirements directly to a corresponding MSS.

The improved accuracy calculation method of Annex A introduced a number of consequences which make it difficult to hold up this link in any case. These consequences are:

— It is not possible to be sure if a MSS calculated before the start of the measurement will precisely satisfy the accuracy requirements after the end of the measurement period. This is due to the fact that one cannot be sure of the actual amount of correlation between test letter results in the measurement period. Securing the accuracy requirements would require an over-sizing of sample and panel or an ongoing MSS calibration process without guarantee of ever fulfilling the accuracy requirements.

— The new MSS calculation rule is only applicable on moderate on-time service levels. For extreme results based on small test mail flows or on high on-time service levels no known MSS calculation rule exists.

— The reduced bias in accuracy calculation (see J.3.2) results in a MSS which would be considerably higher than the level accepted for years in existing transit-time measurement systems.

The link was therefore abandoned. The MSS requirements are now given directly as tabled values without any need for calculation. They are on the same level as in the previous version of this standard. Existing measurement systems which fulfil the former MSS requirements will also fulfil the MSS requirements of6.2.

The improved accuracy calculation is on the other hand still part of the reporting.

J.1.2 MSS for flows with small real mail volumes

EN 13850:2002+A1:2007, i.e. the consolidated version of EN 13850:2002 and EN 13850/prA1:2005, took into consideration real mail volumes for which the corresponding test mail volumes should not go beyond a certain level. The reasons are:

— EN 13850 should provide an acceptable entry level for users of this standard with a smaller mail base;

— The test mail volumes should not raise the size of existing mail flows significantly with possible consequences for the on-time performance;

— The costs of measurement should be in proportion to the importance of the mail flow.

The main body of EN 13850/prA1:2005 was implemented as a new Annex F in EN 13850:2002+A1:2007. This annex has now become Annex D. Accuracy requirements were transformed into minimum sample size requirements according to the line of argument in J.1.1. The three existing mail-flow categories were redefined.

For domestic mail flows, the boundaries between the categories moved from

— [“below 50 million” (very small), “50-100 million” (small), “100 and above million” (large and medium)] mail pieces per year in 2004 to

— [“1,5-200 million” (small), “200-500 million” (medium), “500 and above” million (large)] now.

— A new Category 4 was introduced for very small mail volumes below 1.5 million mail pieces per year.

— For cross-border mail flows only small changes were made

— 1.45 million instead of 1.5 million boundary in Category 1;

— J+1 performance boundary of 80 % instead of 75 % in Categories 2 and 3, see D.3).

— A new Category 4 was introduced for very small mail flows below 11.500 mail pieces.

J.2 Transit-time calculation rule

The transit-time calculation rule of Annex B has been amended for special cases of national or regional holidays in the cross-border case. Holidays are now respected in the country of induction as well as in the country of delivery.

In cases of national or regional holidays the new transit-time calculation rule might lead to different results than the former calculation rule.

J.3 Accuracy calculation method

J.3.1 Improved applicability

Annex A of this standard includes a number of areas of improved applicability in comparison to the previous version of EN 13850.

The accuracy calculation is applicable on stratified samples now. For measurement systems with a weighting system that is in line with 7.3, Annex A provides an unbiased accuracy calculation method, even for samples that are not fully proportional. This closes a gap, since 7.3 explicitly requires a weighting system based on a stratified sample.

Since the first instalment of EN 13850 transit-time performance has improved considerably throughout Europe. The high transit-time performance results that are common now in many European countries may generate a considerable bias if the common Normal confidence interval is used in smaller samples. Adding to that, the interest in QoS sub-results for regional quality improvement has grown, too, and with it the requirement for an accuracy calculation for small sub-samples.

Annex A covers these cases with the introduction of two alternative types of confidence intervals, the Agresti-Coull interval and the Inverse Beta interval. Both intervals take care of small samples and/or high transit-time performance results. A.5.2.2 assists with a decision rule when the alternative confidence intervals should be used.

J.3.2 Reduced bias in calculation

Annex A of this standard removes a bias in calculation that was reported by users of the previous version of EN 13850.

Extensive tests with measurement data from different operators throughout Europe has shown, that the old accuracy calculation method leads to accuracy results that were not conservative enough. The accuracy of large European transit-time measurement systems tended to be approximately four times overestimated.

This bias in the accuracy calculation has been corrected.

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.

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 Group Headquarters

389 Chiswick High Road London W4 4AL UK

bsi.

...making excellence a habit.™