

DIN ISO 13528

ICS 03.120.30

**Statistische Verfahren für Eignungsprüfungen durch Ringversuche
(ISO 13528:2005);
Text Deutsch und Englisch**

Statistical methods for use in proficiency testing by interlaboratory comparisons
(ISO 13528:2005);
Text in German and English

Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons
interlaboratoires (ISO 13528:2005);
Texte en allemand et anglais

Gesamtumfang 107 Seiten

Normenausschuss Qualitätsmanagement, Statistik und Zertifizierungsgrundlagen (NQSZ) im DIN
Normenausschuss Wasserwesen (NAW) im DIN

Inhalt

Seite

Nationales Vorwort	5
0 Einleitung.....	6
0.1 Ziele von Eignungsprüfungen	6
0.2 ISO/IEC Guide 43.....	7
1 Anwendungsbereich.....	10
2 Normative Verweisungen	10
3 Begriffe	11
4 Statistische Leitlinien für die Planung und die Auswertung von Eignungsprüfungen	12
4.1 Eingriffs- und Warnsignale	12
4.2 Leitlinien für die Ermittlung der Grenzen der Unsicherheit des zugewiesenen Werts.....	14
4.3 Leitlinien für die Festlegung der Anzahl von Wiederhol- messungen	15
4.4 Homogenität und Stabilität von Proben.....	16
4.5 Konventionsmessverfahren	17
4.6 Angabe von Daten	17
4.7 Gültigkeitsdauer der Ergebnisse der Eignungsprüfung	18
5 Bestimmung des zugewiesenen Werts und der zugehörigen Standardunsicherheit.....	18
5.1 Wahl des Verfahrens zur Ermittlung des zugewiesenen Werts	18
5.2 Bekannte Werte aus der Probenherstellung	19
5.3 Zertifizierte Referenzwerte	20
5.4 Referenzwerte	21
5.5 Konsenswerte von Expertenlaboratorien.....	25
5.6 Konsenswert der Teilnehmer	26
5.7 Vergleich des zugewiesenen Werts	33
5.8 Fehlende Werte	34
6 Bestimmung der Standardabweichung für die Eignungsbeurteilung	35
6.1 Auswahl des Verfahrens	35
6.2 Vorgabewert	35
6.3 Werte aus Erkenntnissen.....	36
6.4 Wert aus einem allgemeinen Modell.....	38
6.5 Wert aus den Ergebnissen eines Versuchs zur Präzision	39

Contents

Page

0 Introduction.....	6
0.1 The aims of proficiency testing	6
0.2 ISO/IEC Guide 43	7
1 Scope.....	10
2 Normative references.....	10
3 Terms and definitions	11
4 Statistical guidelines for the design and interpretation of proficiency tests	12
4.1 Action and warning signals.....	12
4.2 Guidelines for limiting the uncertainty of the assigned value	14
4.3 Guidelines for choosing the number of replicate measurements	15
4.4 Homogeneity and stability of samples	16
4.5 Operationally defined measurement methods	17
4.6 Reporting of data.....	17
4.7 Period of validity of the results of proficiency tests	18
5 Determination of the assigned value and its standard uncertainty	18
5.1 Choice of method of determining the assigned value.....	18
5.2 Formulation.....	19
5.3 Certified reference values	20
5.4 Reference values	21
5.5 Consensus values from expert laboratories	25
5.6 Consensus value from participants	26
5.7 Comparison of the assigned value	33
5.8 Missing values	34
6 Determining the standard deviation for proficiency assessment.....	35
6.1 Choice of method	35
6.2 Prescribed value.....	35
6.3 By perception.....	36
6.4 From a general model	38
6.5 From the results of a precision experiment	39

	Seite		Page
6.6	Wert aus Daten, die während einer Runde eines Schemas für Eignungsprüfungen ermittelt wurden 40	6.6	From data obtained in a round of a proficiency testing scheme 40
6.7	Vergleich von Werten für die Präzision aus einer Eignungsprüfung mit etablierten Werten 40	6.7	Comparison of precision values derived from a proficiency test with established values 40
7	Berechnung von Leistungskenngrößen 41	7	Calculation of performance statistics.....41
7.1	Schätzwerte für die systematische Abweichungskomponente eines Laboratoriums 41	7.1	Estimates of laboratory bias 41
7.2	Prozentuale Differenzen..... 46	7.2	Percentage differences 46
7.3	Rangordnungen und prozentuale Rangordnungen 49	7.3	Ranks and percentage ranks 49
7.4	z -Werte 51	7.4	z -scores 51
7.5	E_n -Zahlen 54	7.5	E_n numbers 54
7.6	z' -Werte 54	7.6	z' -scores 54
7.7	Zeta-Werte (ζ) 56	7.7	Zeta-scores (ζ) 56
7.8	E_z -Wert 58	7.8	E_z score 58
7.9	Beispiel der Auswertung von Daten mit Angabe von Unsicherheiten..... 59	7.9	An example of the analysis of data when uncertainties are reported.....59
7.10	Kombinierter Leistungswert..... 65	7.10	Combined performance scores 65
8	Grafische Verfahren zur Kombination von Leistungswerten für mehrere Messgrößen in einer Runde einer Eignungsprüfung..... 67	8	Graphical methods for combining performance scores for several measurands from one round of a proficiency test 67
8.1	Anwendung 67	8.1	Application.....67
8.2	Histogramme von Leistungswerten 67	8.2	Histograms of performance scores.....67
8.3	Balken-Diagramme standardisierter systematischer Abweichungskomponenten von Laboratorien..... 69	8.3	Bar-plots of standardized laboratory biases 69
8.4	Balkendiagramme standardisierter Wiederholmessungen 71	8.4	Bar-plots of standardized repeatability measurements 71
8.5	Youden-Diagramm..... 71	8.5	Youden Plot 71
8.6	Diagramme von Wiederholstandardabweichungen 79	8.6	Plots of repeatability standard deviations 79
8.7	Geteilte Proben 83	8.7	Split samples 83
9	Grafische Verfahren zur Kombination von Leistungswerten über mehrere Runden eines Schemas für Eignungsprüfungen 88	9	Graphical methods for combining performance scores over several rounds of a proficiency testing scheme.....88
9.1	Anwendungen 88	9.1	Applications.....88
9.2	Shewhart-Regelkarte für z -Werte 89	9.2	Shewhart control chart for z -scores 89
9.3	Cusum-Karte für z -Werte..... 92	9.3	Cusum control chart for z -scores 92
9.4	Auftragungen standardisierter systematischer Abweichungskomponenten von Laboratorien gegen Laboratorienmittelwerte 94	9.4	Plots of standardized laboratory biases against laboratory averages.....94
9.5	Punktwolke..... 95	9.5	Dot plot.....95

	Seite		Page
Anhang A (normativ) Formelzeichen	97	Annex A (normative) Symbols.....	97
Anhang B (normativ) Prüfung von Proben auf Homogenität und Stabilität	98	Annex B (normative) Homogeneity and stability checks of samples.....	98
Anhang C (normativ) Robuste Auswertung	104	Annex C (normative) Robust analysis	104
Literaturhinweise	107	Bibliography.....	107

Nationales Vorwort

Dieses Dokument enthält die deutsche Übersetzung der Internationalen Norm ISO 13528:2005-09, die vom Technischen Komitee ISO/TC 69, *Applications of statistical methods*, Unterkomitee SC 6, *Measurement methods and results*, erarbeitet wurde, dessen Sekretariat vom JISC/JSA (Japan) gehalten wird. Der zuständige nationale Normenausschuss ist der NA 147, *Qualitätsmanagement, Statistik und Zertifizierungsgrundlagen* (NQSZ), Arbeitsausschuss *Angewandte Statistik*. Deutscher Mitträger ist der Normenausschuss NA 119, *Wasserwesen* (NAW), Arbeitsausschuss *Wasseruntersuchung*.

Im Bereich der Wasseruntersuchungen werden z -Wert, z' -Wert, Zeta-Wert und E_z -Wert als z -score, z' -score, Zeta-score beziehungsweise E_z -Score bezeichnet.

Es wird auf die Möglichkeit hingewiesen, dass für Teile dieses Dokuments Patentrechte bestehen. Das DIN übernimmt keine Verantwortung für die Kennzeichnung solcher Rechte.

Für die in diesem Dokument zitierten Internationalen Normen wird im Folgenden auf die entsprechenden Deutschen Normen hingewiesen:

ISO 5725-1	siehe DIN ISO 5725-1
ISO/IEC Guide 43-1:1997	siehe DIN V 55394-1:2000-09

Nationaler Anhang NA (informativ)

Literaturhinweise

DIN ISO 5725-1, *Genauigkeit (Richtigkeit und Präzision) von Messverfahren und Messergebnissen — Teil 1: Allgemeine Grundlagen und Begriffe*

DIN V 55394-1:2000-09, *Eignungsprüfung durch Vergleiche zwischen Laboratorien — Teil 1: Entwicklung und Durchführung von Programmen für die Eignungsprüfung*

Statistische Verfahren für Eignungsprüfungen durch Ringversuche

0 Einleitung

0.1 Ziele von Eignungsprüfungen

Eignungsprüfungen durch Ringversuche werden angewendet, um die Leistungsfähigkeit einzelner Laboratorien für bestimmte Prüfungen oder Messungen zu bestimmen und die fortgesetzte Leistungsfähigkeit von Laboratorien zu überwachen. Die Einleitung zum ISO/IEC Guide 43-1:1997 enthält eine umfassende Darstellung der Zwecke der Eignungsprüfung. In der Sprache der Statistik kann die Leistungsfähigkeit eines Laboratoriums durch drei Eigenschaften beschrieben werden: systematische Abweichungskomponente des Laboratoriums, Stabilität und Wiederholpräzision. Die systematische Abweichungskomponente des Laboratoriums und die Wiederholpräzision sind in ISO 3534-1, ISO 3534-2 und ISO 5725-1 definiert. Ein Maß für die Stabilität der Ergebnisse eines Laboratoriums ist die Präzision unter Zwischenbedingungen, wie in ISO 5725-3 definiert.

Die systematische Abweichungskomponente eines Laboratoriums kann anhand von Prüfungen mit Referenzmaterialien mit Hilfe des in ISO 5725-4 beschriebenen Verfahrens beurteilt werden, wenn solche Referenzmaterialien verfügbar sind. Andernfalls liefert die Eignungsprüfung durch Ringversuche ein allgemein verfügbares Mittel, um Angaben über die systematische Abweichungskomponente eines Laboratoriums zu erhalten, und die Verwendung von Daten aus Eignungsprüfungen, um Schätzwerte für die systematische Abweichungskomponente eines Laboratoriums zu erhalten, ist ein wichtiger Aspekt der Auswertung solcher Daten. Jedoch beeinflussen Stabilität und Wiederholpräzision durch Eignungsprüfungen gewonnene Daten. Deshalb ist es möglich, dass ein Laboratorium während einer Runde der Eignungsprüfung Daten erhält, die eine systematische Abweichungskomponente aufzeigen, die tatsächlich durch ungenügende Stabilität oder ungenügende Wiederholpräzision verursacht wird. Es ist deshalb wichtig, dass diese Aspekte der Leistungsfähigkeit von Laboratorien regelmäßig beurteilt werden.

Stabilität kann durch das wiederholte Prüfen aufbewahrter Proben oder durch regelmäßige Messungen an einem Referenzmaterial oder einem laboratoriumseigenen Referenzmaterial (einem Vorrat an Material, der von einem Laboratorium zur Verwendung als privates Referenzmaterial angelegt

Statistical methods for use in proficiency testing by interlaboratory comparisons

0 Introduction

0.1 The aims of proficiency testing

Proficiency testing by interlaboratory comparisons is used to determine the performance of individual laboratories for specific tests or measurements, and to monitor the continuing performance of laboratories. The Introduction to ISO/IEC Guide 43-1:1997 should be consulted for a full exposition of the purposes of proficiency testing. In statistical language, the performance of laboratories can be described by three properties: laboratory bias, stability and repeatability. Laboratory bias and repeatability are defined in ISO 3534-1, ISO 3534-2 and ISO 5725-1. The stability of a laboratory's results is measured by intermediate precision as defined in ISO 5725-3.

Laboratory bias may be assessed by tests on reference materials, when these are available, using the procedure described in ISO 5725-4. Otherwise, proficiency testing by interlaboratory comparisons provides a generally available means of obtaining information about laboratory bias, and the use of data from proficiency tests to obtain estimates of laboratory bias is an important aspect of the analysis of such data. However, stability and repeatability will affect data obtained in proficiency tests, so that it is possible for a laboratory to obtain data in a round of a proficiency test which indicate bias that is actually caused by poor stability or poor repeatability. It is therefore important that these aspects of laboratory performance are assessed regularly.

Stability may be assessed by re-testing of retained samples, or by making regular measurements on a reference material or an in-house reference material (a stock of material established by a laboratory to use as private reference material). These techniques are described in ISO 5725-3. Stability may also be assessed by plotting estimates of laboratory bias derived from proficiency tests in control charts. This can provide information about laboratory performance that is not apparent from the examination of the results of individual rounds of proficiency testing schemes, and is another important aspect of the analysis of such data.

Data suitable for assessing repeatability may be generated by tests carried out in the normal course of the work of a laboratory, or by extra tests carried out within a laboratory specifically to assess

wird) beurteilt werden. Diese Techniken werden in ISO 5725-3 beschrieben. Stabilität kann auch beurteilt werden, indem Schätzwerte der systematischen Abweichungskomponente eines Laboratoriums, die aus Eignungsprüfungen hergeleitet wurden, in Qualitätsregelkarten grafisch dargestellt werden. Dies kann Informationen über die Leistungsfähigkeit des Laboratoriums liefern, die bei der Überprüfung der Ergebnisse einzelner Runden von Schemata für Eignungsprüfungen nicht erkennbar werden, und stellt einen weiteren wichtigen Aspekt der Auswertung solcher Daten dar.

Daten, die zur Beurteilung der Wiederholpräzision geeignet sind, dürfen durch im üblichen Arbeitsverlauf eines Laboratoriums durchgeführte Prüfungen oder anhand von zusätzlichen Prüfungen erzeugt werden, die eigens zum Zweck der Beurteilung der Wiederholpräzision innerhalb eines Laboratoriums durchgeführt werden. Folglich ist die Beurteilung der Wiederholpräzision nicht unbedingt ein wichtiger Aspekt der Eignungsprüfung, obwohl es wichtig ist, dass Laboratorien ihre Wiederholpräzision auf irgendeine Weise überwachen. Die Wiederholpräzision kann durch die grafische Auftragung von Spannweiten von Doppelmessungen in einer Qualitätsregelkarte beurteilt werden, wie in ISO 5725-6 beschrieben.

Das Ablaufdiagramm (Bild 1) veranschaulicht, wie die in dieser Internationalen Norm beschriebenen Techniken anzuwenden sind.

0.2 ISO/IEC Guide 43

ISO/IEC Guide 43-1 beschreibt verschiedene Arten von Schemata für Eignungsprüfungen und gibt Hinweise hinsichtlich der Organisation und Erstellung von Schemata für Eignungsprüfungen. ISO/IEC Guide 43-2 gibt Hinweise zur Auswahl und Verwendung von Schemata für Eignungsprüfungen durch Akkreditierungsstellen für Laboratorien. Detaillierte Angaben hierzu sollten diesen Dokumenten entnommen werden. (Die Information wird hier nicht wiederholt.) ISO/IEC Guide 43-1 enthält einen Anhang, der kurz die statistischen Verfahren beschreibt, die bei Schemata für Eignungsprüfungen verwendet werden.

Diese Internationale Norm ergänzt ISO/IEC Guide 43, indem sie detaillierte Hinweise zur Anwendung statistischer Verfahren bei Eignungsprüfungen liefert, die im ISO/IEC Guide 43 fehlen. ISO 13528 basiert weitgehend auf einem harmonisierten Protokoll für die Eignungsprüfung von analytischen Laboratorien [1], aber ist zur Anwendung bei allen Messverfahren bestimmt.

repeatability. Consequently, the assessment of repeatability is not necessarily an important aspect of proficiency testing, although it is important that laboratories monitor their repeatability in some way. Repeatability may be assessed by plotting ranges of duplicate measurements on a control chart as described in ISO 5725-6.

The flowchart (Figure 1) illustrates how the techniques described in this International Standard are to be applied.

0.2 ISO/IEC Guide 43

ISO/IEC Guide 43-1 describes different types of proficiency testing schemes and gives guidance on the organization and design of proficiency testing schemes. ISO/IEC Guide 43-2 gives guidance on the selection and use of proficiency testing schemes by laboratory accreditation bodies. Those documents should be consulted for detailed information in those areas (the information is not duplicated here). ISO/IEC Guide 43-1 contains an annex that briefly describes the statistical methods that are used in proficiency testing schemes.

This International Standard is complementary to ISO/IEC Guide 43, providing detailed guidance that is lacking in that document on the use of statistical methods in proficiency testing. ISO 13528 is to a large extent based on a harmonized protocol for the proficiency testing of analytical laboratories [1], but is intended for use with all measurement methods.

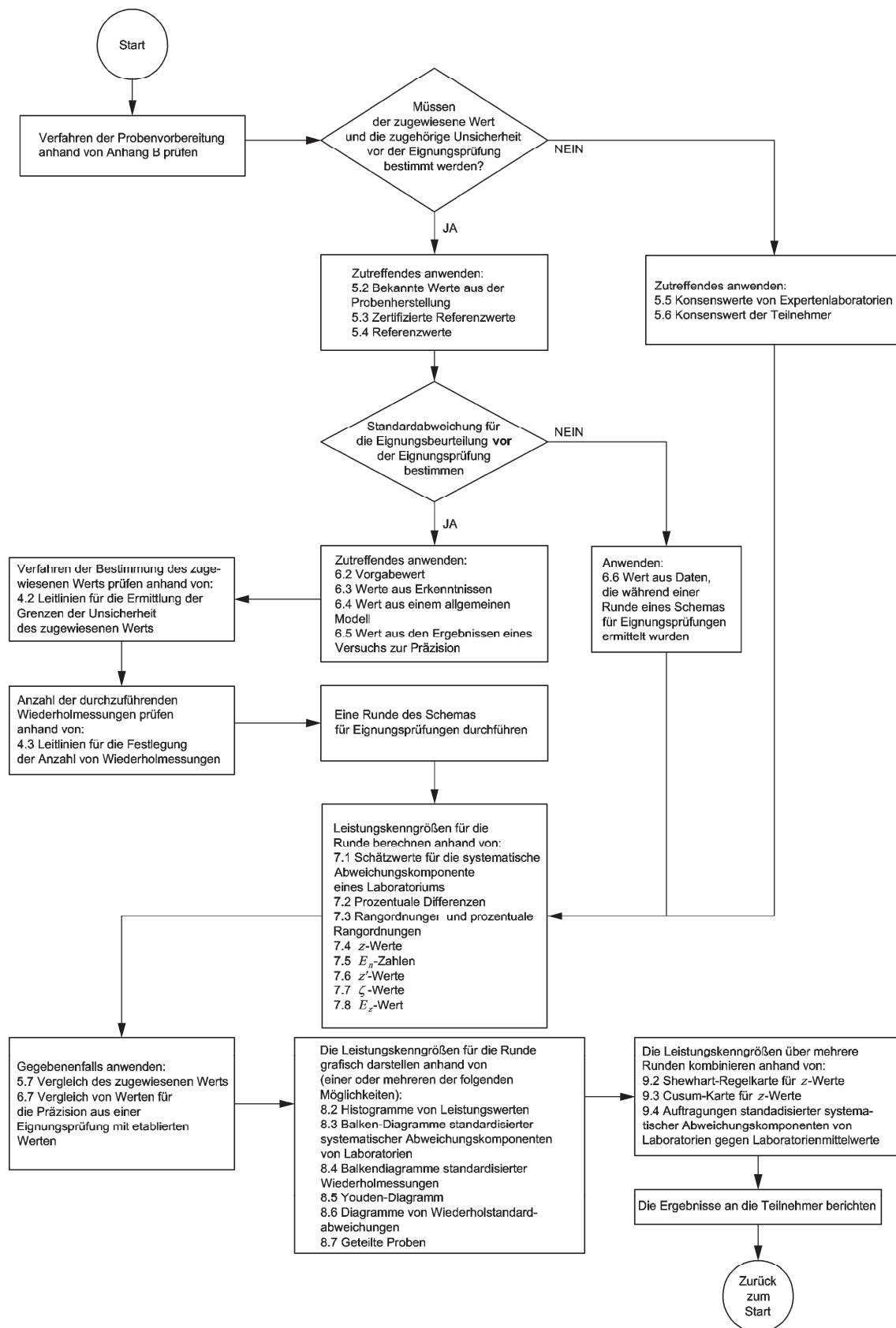


Bild 1 — Ablaufdiagramm der Aktivitäten, die die Verwendung von statistischen Verfahren erfordern, wenn ein Schema für Eignungsprüfungen ausgeführt wird

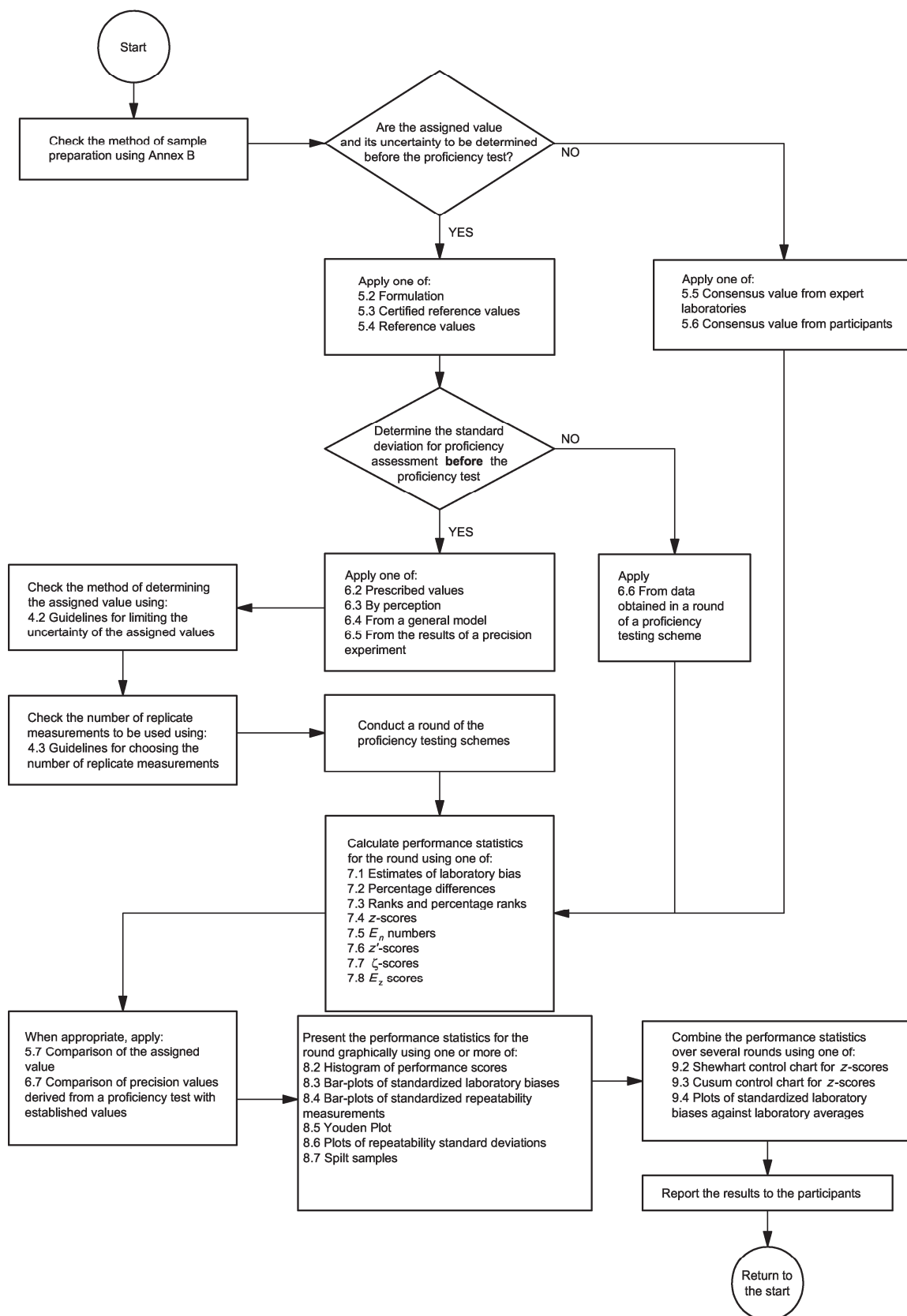


Figure 1 — Flowchart showing the activities requiring the use of statistical methods when operating a proficiency testing scheme

1 Anwendungsbereich

Diese Internationale Norm ergänzt ISO Guide 43 (alle Teile), indem sie detailliert zuverlässige statistische Verfahren beschreibt, die von Ausrichtern von Schemata für Eignungsprüfungen zur Auswertung der erhaltenen Daten anzuwenden sind, und indem sie Empfehlungen zur praktischen Anwendung dieser Verfahren durch Teilnehmer an solchen Prüfungen beziehungsweise durch Akkreditierungsstellen gibt.

Diese Internationale Norm kann angewendet werden, um nachzuweisen, dass die von Laboratorien erhaltenen Messergebnisse keine unannehmbare systematische Abweichungskomponente aufweisen.

Sie ist auf quantitative, nicht jedoch auf qualitative Daten anwendbar.

2 Normative Verweisungen

Die folgenden in Bezug genommenen Dokumente sind für die Anwendung dieses Dokuments unentbehrlich. Bei datierten Verweisungen gilt die jeweils in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die jeweils letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich Änderungen).

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*

ISO 3534-2:—¹⁾ N¹⁾, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO/IEC Guide 43-1:1997, *Proficiency testing by interlaboratory comparisons — Part 1: Development and operation of proficiency testing schemes*

1 Scope

This International Standard complements ISO Guide 43 (all parts) by providing detailed descriptions of sound statistical methods for organizers to use to analyse the data obtained from proficiency testing schemes, and by giving recommendations on their use in practice by participants in such schemes and by accreditation bodies.

This International Standard can be applied to demonstrate that the measurement results obtained by laboratories do not exhibit evidence of an unacceptable level of bias.

It is applicable to quantitative data but not to qualitative data.

2 Normative references

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

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*

ISO 3534-2:—¹⁾, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO/IEC Guide 43-1:1997, *Proficiency testing by interlaboratory comparisons — Part 1: Development and operation of proficiency testing schemes*

1) In Vorbereitung./To be published.

N1) Nationale Fußnote: 2006 erschienen.

3 Begriffe

Für die Anwendung dieses Dokuments gelten die Begriffe nach ISO 3534-1, ISO 3534-2 und ISO 5725-1 und die folgenden Begriffe.

3.1

Ringversuch

Organisation, Durchführung und Auswertung von Prüfungen oder Messungen an denselben oder ähnlichen Prüfgegenständen durch zwei oder mehr Laboratorien entsprechend vorherbestimmten Bedingungen

ANMERKUNG Angepasst nach ISO/IEC Guide 43-1.

3.2

Eignungsprüfung

Bestimmung der Leistungsfähigkeit eines Laboratoriums durch Ringversuche

3.3

zugewiesener Wert

Wert, welcher einer bestimmten Größe zugeordnet wurde und welcher, mitunter durch Absprache, als mit einer für einen gegebenen Zweck geeigneten Unsicherheit behaftet akzeptiert wird

3.4

Standardabweichung für die Eignungsbeurteilung

Maß für die Streuung, das bei der Beurteilung der Eignung auf der Grundlage der verfügbaren Information verwendet wird

3.5

z-Wert

standardisiertes Maß für die systematische Abweichungskomponente eines Laboratoriums, berechnet unter Verwendung des zugewiesenen Werts und der Standardabweichung für die Eignungsbeurteilung

3.6

Koordinator

Organisation oder Person, die die Verantwortung für die Koordination aller Tätigkeiten trägt, die im Rahmen eines Schemas für Eignungsprüfungen anfallen

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2, ISO 5725-1 and the following apply.

3.1

interlaboratory comparison

organization, performance and evaluation of tests or measurements on the same or similar test items by two or more laboratories in accordance with predetermined conditions

NOTE Adapted from ISO/IEC Guide 43-1.

3.2

proficiency testing

determination of laboratory testing performance by means of interlaboratory comparisons

3.3

assigned value

value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose

3.4

standard deviation for proficiency assessment

measure of dispersion used in the assessment of proficiency, based on the available information

3.5

z-score

standardized measure of laboratory bias, calculated using the assigned value and the standard deviation for proficiency assessment

3.6

coordinator

organization (or person) with responsibility for coordinating all of the activities involved in the operation of a proficiency testing scheme

4 Statistische Leitlinien für die Planung und die Auswertung von Eignungsprüfungen

(siehe ISO/IEC Guide 43-1:1997, 5.4.2)

4.1 Eingriffs- und Warnsignale

4.1.1 Diese Internationale Norm beschreibt einige einfache numerische oder grafische Kriterien, die auf die im Rahmen einer Eignungsprüfung erhaltenen Daten angewendet werden sollten, um festzustellen, ob Eingriffs- oder Warnsignale vorliegen. Selbst in einem gut geführten Laboratorium mit erfahrener Personal können mitunter anomale Ergebnisse ermittelt werden. Ebenso kann ein standardisiertes Messverfahren, auch wenn es in einem Präzisionsversuch validiert wurde, Fehler enthalten, die erst nach mehreren Runden eines Schemas für Eignungsprüfungen auffallen. Das Eignungsprüfungssystem selbst kann Fehler enthalten. Aus diesen Gründen sollen die hier gegebenen Kriterien nicht verwendet werden, um Laboratorien als ungeeignet für die Durchführung des betrachteten Messverfahrens zu beurteilen. Wenn die Eignungsprüfung angewendet wird, um Laboratorien als ungeeignet zu beurteilen, so müssen zwingend entsprechende Kriterien für diesen Zweck entwickelt werden.

4.1.2 Die hier angegebenen Kriterien sind so ausgelegt, dass sie Eingriffssignale geben, wenn die Standardabweichung für die Eignungsbeurteilung auf der (nach einem der in 6.4 bis 6.6 beschriebenen Verfahren) beobachteten Leistungsfähigkeit basiert und die Ergebnisse so aus dem Rahmen fallen, dass eine genauere Prüfung und Korrekturmaßnahmen angemessen erscheinen.

4.1.3 Der Koordinator sollte die wichtigsten Quellen der Variabilität verstehen, die bei Daten der Eignungsprüfung für die betreffende Messung zu erwarten sind. Erster Schritt jeder Auswertung sollte die Untersuchung der Verteilung von Ergebnissen daraufhin sein, ob Hinweise für unvorhergesehene Quellen von Variabilität vorliegen. Zum Beispiel könnte eine bimodale Verteilung ein Hinweis auf eine gemischte Grundgesamtheit von Ergebnissen sein, verursacht durch unterschiedliche Verfahren, kontaminierte Proben oder schlecht formulierte Anleitungen. In dieser Situation sollten solche Bedenken ausgeräumt werden, bevor die Auswertung oder Beurteilung fortgesetzt wird. Akkreditierungsstellen müssen Leitlinien für die Reaktion auf unannehmbare Leistungen bei Eignungsprüfungen haben. Maßnahmen als Folge solcher Fehlleistungen werden anhand dieser Leitlinien oder durch die Qualitätssicherungs-

4 Statistical guidelines for the design and interpretation of proficiency tests

(see ISO/IEC Guide 43-1:1997, 5.4.2.)

4.1 Action and warning signals

4.1.1 This International Standard describes some simple numerical or graphical criteria that should be applied to the data obtained in a proficiency test to see if they give rise to action or warning signals. Even in a well-run laboratory, with experienced staff, anomalous results may sometimes be obtained. Also, it is possible that a standardized measurement method, even though it has been validated by a precision experiment, may contain faults that become apparent only after several rounds of a proficiency testing scheme. The proficiency scheme itself may contain faults. For these reasons, the criteria given here shall not be used to condemn laboratories, as being unfit to perform the measurement method under examination. If proficiency testing is used to condemn laboratories, then it shall be necessary to devise appropriate criteria for that purpose.

4.1.2 The criteria given here are designed so that, when the standard deviation for proficiency assessment is based on observed performance (using one of the methods described in 6.4 to 6.6), the criteria give action signals when results are so exceptional as to merit investigation and corrective action.

4.1.3 The coordinator should have an understanding of the major sources of variability that can be anticipated in proficiency test data for the measurement in question. The first step in any analysis should be to examine the distribution of results for evidence of unanticipated sources of variability. For example, a bimodal distribution might be evidence of a mixed population of results caused by different methods, contaminated samples or poorly worded instructions. In this situation, the concern should be resolved before proceeding with analysis or evaluation. Accrediting bodies shall have policies for response to unacceptable performance in proficiency testing. Follow-up actions are determined by that policy or by the laboratory's quality procedures. However, there are generally recommended actions when a laboratory produces an unacceptable result in a proficiency test. Guidance for actions by labora-

verfahren des Laboratoriums bestimmt. Es gibt jedoch allgemein empfohlene Maßnahmen für den Fall, dass ein Laboratorium bei einer Eignungsprüfung ein unannehmbares Ergebnis produziert. Leitlinien für Maßnahmen von Laboratorien als Reaktion auf erfolglose Durchführung einer Eignungsprüfung werden in 4.1.4 gegeben.

4.1.4 Bei Plänen, bei denen die Standardabweichung für die Eignungsbeurteilung auf der beobachteten Leistungsfähigkeit basiert, muss das Laboratorium entscheiden, welche Nachforschungen und Korrekturmaßnahmen angemessen sind, wenn ein Ergebnis ein Eingriffssignal liefert; nötigenfalls ist diese Entscheidung in Abstimmung mit dem Koordinator oder einer Akkreditierungsstelle zu treffen. Spricht kein triftiger Grund dagegen, muss das Laboratorium seine Verfahrensweisen prüfen und eine oder mehrere Korrekturmaßnahmen ermitteln, die nach dem Ermessen des Laboratoriumspersonals wahrscheinlich das erneute Auftreten solcher Ergebnisse verhindern. Das Laboratorium kann sich hinsichtlich möglicher Ursachen für sein Problem durch den Koordinator beraten lassen oder den Koordinator bitten, weitere Fachleute zu Rate zu ziehen. Das Laboratorium muss an weiteren Runden des Schemas für Eignungsprüfungen teilnehmen, um die Wirksamkeit der Korrekturmaßnahmen zu beurteilen. Als geeignete Korrekturmaßnahme kommt eine der folgenden in Frage:

- a) Überprüfung, ob das Personal das Messverfahren versteht und befolgt;
- b) Überprüfung, ob alle Einzelheiten des Messablaufs korrekt sind;
- c) Überprüfung der Kalibrierung der Geräte und der Zusammensetzung von Reagenzien;
- d) Ersatz von fehlerverdächtigen Geräten oder Reagenzien;
- e) vergleichende Prüfung des Personals, der Geräte und/oder Reagenzien mit einem anderen Laboratorium.

Die Verwendung der Ergebnisse von Eignungsprüfungen durch Akkreditierungsstellen für Laboratorien ist in ISO/IEC Guide 43-2:1997, Abschnitt 6, beschrieben.

tories in response to unsuccessful performance on a proficiency test is given in 4.1.4.

4.1.4 In schemes where the standard deviation for proficiency assessment is based on observed performance, when a result gives an action signal, the laboratory shall decide what investigations and corrective actions are appropriate, in consultation with the coordinator or an accreditation body if necessary. Unless there is a valid reason not to do so, the laboratory shall examine its procedures and identify one or more corrective actions that, according to staff in the laboratory, are likely to prevent the recurrence of such results. The laboratory may ask the coordinator for advice on possible causes of its problem, or may ask the coordinator to consult other experts. The laboratory shall take part in further rounds of the proficiency testing scheme to assess the effectiveness of the corrective actions. Appropriate corrective actions may be one of the following:

- a) checking that staff understand and follow the measurement procedure;
- b) checking that all details of the measurement procedure are correct;
- c) checking the calibration of equipment and the composition of reagents;
- d) replacing suspect equipment or reagents;
- e) comparative tests of staff, equipment and/or reagents with another laboratory.

The use of the results of proficiency tests by laboratory accreditation bodies is described in ISO/IEC Guide 43-2:1997, Clause 6.

4.2 Leitlinien für die Ermittlung der Grenzen der Unsicherheit des zugewiesenen Werts

Der zugewiesene Wert X hat eine Standardunsicherheit u_X . Diese Unsicherheit hängt vom Verfahren ab, das verwendet wird, um sie zu bestimmen. Wenn sie durch Prüfungen in mehreren Laboratorien bestimmt wird, hängt diese Unsicherheit auch von der Anzahl von Laboratorien und möglicherweise von weiteren Faktoren ab. Verfahren zur Berechnung der Standardunsicherheit des zugewiesenen Werts werden in Abschnitt 5 angegeben.

Die Standardabweichung für die Eignungsprüfung, $\hat{\sigma}$, wird verwendet, um die Größe von Schätzwerten der in einer Eignungsprüfung gefundenen systematischen Abweichungskomponente eines Laboratoriums zu beurteilen. Verfahren zur Bestimmung der Standardabweichung für die Eignungsprüfung werden in Abschnitt 6 angegeben und Kriterien, die sie mit Schätzwerten der systematischen Abweichungskomponente eines Laboratoriums vergleichen, werden in Abschnitt 7 gegeben.

Wenn die Standardunsicherheit u_X des zugewiesenen Werts zu groß im Vergleich zur Standardabweichung für die Eignungsprüfung, $\hat{\sigma}$, ist, dann besteht ein Risiko, dass einige Laboratorien Eingriffs- oder Warnsignale erhalten, dies jedoch nicht wegen einer Ursache innerhalb des Laboratoriums, sondern auf Grund von Ungenauigkeiten bei der Bestimmung des zugewiesenen Werts. Daher muss die Standardunsicherheit des zugewiesenen Werts bestimmt und den an Schemata für Eignungsprüfungen teilnehmenden Laboratorien mitgeteilt werden (siehe ISO/IEC Guide 43-1:1997, A.1.4 und A.1.6).

Wenn

$$u_X \leq 0,3\hat{\sigma} \quad (1)$$

dann ist die Unsicherheit des zugewiesenen Werts vernachlässigbar und muss nicht in die Auswertung der Ergebnisse der Eignungsprüfung einbezogen werden.

Wenn diesen Leitlinien nicht entsprochen wird, dann muss der Koordinator Folgendes berücksichtigen:

- a) Es ist ein Verfahren zur Bestimmung des zugewiesenen Werts zu suchen, dessen Unsicherheit der oben genannten Leitlinie entspricht.

4.2 Guidelines for limiting the uncertainty of the assigned value

The assigned value X has a standard uncertainty u_X that depends on the method that is used to derive it, and also, when it is derived from tests in several laboratories, on the number of laboratories and, perhaps, on other factors. Methods for calculating the standard uncertainty of the assigned value are given in Clause 5.

The standard deviation for proficiency testing $\hat{\sigma}$ is used to assess the size of estimates of laboratory bias found in a proficiency test. Methods for obtaining the standard deviation for proficiency testing are given in Clause 6 and criteria that compare it with estimates of laboratory bias are given in Clause 7.

If the standard uncertainty u_X of the assigned value is too large in comparison with the standard deviation for proficiency testing $\hat{\sigma}$, then there is a risk that some laboratories will receive action and warning signals because of inaccuracy in the determination of the assigned value, not because of any cause within the laboratories. For this reason, the standard uncertainty of the assigned value shall be established and shall be reported to laboratories participating in proficiency testing schemes (see ISO/IEC Guide 43-1:1997, A.1.4 and A.1.6).

If

$$u_X \leq 0,3\hat{\sigma} \quad (1)$$

then the uncertainty of the assigned value is negligible and need not be included in the interpretation of the results of the proficiency test.

If these guidelines are not met, then the coordinator shall consider the following.

- a) Look for a method for determining the assigned value such that its uncertainty meets the above guideline.

- b) Es ist bei der Auswertung der Ergebnisse der Eignungsprüfung die Unsicherheit des zugewiesenen Werts zu verwenden (siehe 7.5 bezüglich der E_n -Zahlen oder 7.6 bezüglich des z' -Werts).
- c) Die Teilnehmer der Eignungsprüfung sind zu informieren, dass die Unsicherheit des zugewiesenen Werts nicht vernachlässigbar ist.

BEISPIEL Angenommen, der zugewiesene Wert X wird als Mittelwert $\bar{x}$ der Prüfungsergebnisse von 11 Laboratorien bestimmt und die Standardabweichung für die Eignungsprüfung als die Standardabweichung s derselben 11 Ergebnisse, also $\hat{\sigma} = s$. In erster Näherung kann die Standardunsicherheit des zugewiesenen Werts in dieser Situation zu $u_X = s/\sqrt{11} = 0,3 s$ abgeschätzt werden, so dass die Anforderung scheinbar erfüllt ist. Jedoch kann die Anforderung in dieser Situation mit weniger als 11 Laboratorien nicht erfüllt werden. Weiter wird die Unsicherheit des zugewiesenen Werts größer als $s/\sqrt{11}$ sein, wenn die Proben inhomogen oder instabil sind oder wenn es einen Faktor gibt, der eine gemeinsame systematische Abweichungskomponente in den Ergebnissen des Laboratoriums verursacht (zum Beispiel wenn alle dasselbe Referenznormal verwenden).

4.3 Leitlinien für die Festlegung der Anzahl von Wiederholmessungen

Wiederholstreuung trägt zur Streuung in den systematischen Abweichungskomponenten von Laboratorien bei einer Eignungsprüfung bei. Wenn die Wiederholstreuung zu groß im Vergleich zur Standardabweichung für die Eignungsprüfung ist, dann besteht ein Risiko, dass die Wiederholstreuung zu irregulär streuenden Werten (Ausreißer) in den Ergebnissen der Eignungsprüfung führt. In dieser Situation besteht die Möglichkeit, dass ein Laboratorium bei einer Runde eine große systematische Abweichungskomponente aufweist, in der nächsten Runde jedoch nicht, und die Ursache dafür wird nur schwierig zu ermitteln sein.

Wenn es als wünschenswert betrachtet wird, den Einfluss der Wiederholstreuung einzugrenzen, ist daher die Anzahl n der Wiederholmessungen, die von jedem Laboratorium bei einer Eignungsprüfung durchzuführen sind, so zu wählen, dass

$$\sigma_r/\sqrt{n} \leq 0,3\hat{\sigma} \quad (2)$$

Dabei ist σ_r die Wiederholstandardabweichung, die bei einem vorangegangenen Ringversuch ermittelt wurde.

- b) Use the uncertainty of the assigned value in the interpretation of the results of the proficiency test (see 7.5 on E_n numbers or 7.6 on the z' -score).
- c) Inform the participants in the proficiency test that the uncertainty of the assigned value is not negligible.

EXAMPLE Suppose that the assigned value X is determined as the average $\bar{x}$ of the results of tests in 11 laboratories, and that the standard deviation for proficiency testing is determined as the standard deviation s of these same 11 results, so $\hat{\sigma} = s$. As a first approximation, the standard uncertainty of the assigned value in this situation may be estimated by $u_X = s/\sqrt{11} = 0,3 s$, so that the requirement appears to be met. However, the requirement cannot be met in this situation with fewer than 11 laboratories. Further, the uncertainty of the assigned value will be larger than $s/\sqrt{11}$ if the samples suffer from non-homogeneity or instability, or if there is a factor that causes a common bias in the results of the laboratories (e.g. if they all use the same reference standard).

4.3 Guidelines for choosing the number of replicate measurements

Repeatability variation contributes to the variation between the laboratory biases in a proficiency test. If the repeatability variation is too large in comparison with the standard deviation for proficiency testing, then there is a risk that repeatability variation will cause the results of the proficiency test to be erratic. In this situation, a laboratory could have large bias in one round, but not the next, and they will have difficulty identifying the cause.

For this reason, when it is considered desirable to limit the influence of repeatability variation, the number of replicate measurements n made by each laboratory in a proficiency test shall be chosen so that:

$$\sigma_r/\sqrt{n} \leq 0,3\hat{\sigma} \quad (2)$$

where σ_r is the repeatability standard deviation that has been established in a previous interlaboratory experiment.

Der Faktor von 0,3 ist dadurch gerechtfertigt, dass die Wiederholstandardabweichung nicht mehr als etwa 10 % zur Standardabweichung für die Eignungsprüfung beiträgt, wenn dieses Kriterium erfüllt ist.

Weiterhin müssen alle Laboratorien dieselbe Anzahl von Wiederholmessungen durchführen. (Die später in dieser Internationalen Norm angegebenen Auswertungsverfahren basieren auf der Annahme, dass diese Anforderung erfüllt ist.) Wenn die durch Ungleichung (2) definierte Anforderung nicht erfüllt wird, muss die Anzahl der Wiederholmessungen erhöht werden, oder die Ergebnisse der Eignungsprüfung müssen mit Vorsicht interpretiert werden.

Dieser Ansatz geht davon aus, dass Laboratorien im Allgemeinen ähnliche Wiederholpräzisionen aufweisen. Es können sich Fälle ergeben, bei denen diese Annahme nicht erfüllt ist. Um die Verfahren, die in dieser Internationalen Norm beschrieben werden, anzuwenden, kann in solchen Fällen wie folgt vorgegangen werden. Der Koordinator sollte die Anzahl von Wiederholmessungen n mit Hilfe eines typischen Werts für die Wiederholstandardabweichung festsetzen. Dann sollte jedes Laboratorium überprüfen, ob es Ungleichung (2) mit seiner eigenen Wiederholstandardabweichung erfüllt. Wenn dies nicht der Fall ist, dann sollte es sein Messverfahren so modifizieren, dass es ein Prüfergebnis als Mittelwert einer Anzahl von Bestimmungen erhält, die so gewählt wurde, dass Ungleichung (2) erfüllt ist.

4.4 Homogenität und Stabilität von Proben (siehe ISO/IEC Guide 43-1:1997, 5.6.2 und 5.6.3)

Anhang B gibt Verfahren an, die geeignet sind zu überprüfen, ob die bei einer Eignungsprüfung zu verwendenden Proben hinreichend homogen und stabil sind.

Wenn ein Verfahren zur Probenvorbereitung angewendet wird, so dass das Homogenitätskriterium in Anhang B nicht erfüllt wird, müssen die Teilnehmer mehrere Proben prüfen. Alternativ muss die Standardabweichung für die Eignungsprüfung einen Zuschlag für die Heterogenität der Proben beinhalten, wie in Anhang B beschrieben.

The justification for the factor of 0,3 is that when this criterion is met, the repeatability standard deviation contributes no more than about 10 % of the standard deviation for proficiency testing.

Further, all laboratories shall carry out the same number of replicate measurements. (The methods of analysis given later in this International Standard assume that this requirement is met.) If the requirement inequality of (2) is not met, then the number of replicate measurements shall be increased, or the results of the proficiency test shall be interpreted with caution.

This approach assumes that laboratories have generally similar repeatability. Cases can arise where this is not so. In such cases, for the methods described in this International Standard to be applied, the following device may be used. The coordinator should fix the number of replicate measurements n , using a typical value for the repeatability standard deviation. Then, each laboratory should check that it satisfies Inequality (2) with its own repeatability standard deviation. If it does not, then it should modify its measurement procedure so that it obtains a test result as the average of some number of determinations chosen so that Inequality (2) is satisfied.

4.4 Homogeneity and stability of samples (see ISO/IEC Guide 43-1:1997, 5.6.2 and 5.6.3)

Methods are given in Annex B for checking that the samples to be used in a proficiency test are adequately homogeneous and stable.

When a method of sample preparation is used such that the homogeneity criterion in Annex B is not met, then replicate samples shall be tested by the participants, or the standard deviation for proficiency testing shall include an allowance for the heterogeneity of the samples, as described in Annex B.

4.5 Konventionsmessverfahren

Bei einem Konventionsmessverfahren ist das Messergebnis durch das Messverfahren bestimmt. Zum Beispiel kann die Größenverteilung eines partikulären Materials mit Hilfe von Sieben mit runden oder quadratischen Löchern bestimmt werden. Es mag keinen Grund geben, warum eine bestimmte Art von Sieb zu bevorzugen ist, wenn aber die Art des Siebs nicht vorgegeben ist, können Laboratorien, die verschiedene Arten von Sieben verwenden, unterschiedliche Ergebnisse erhalten. Wenn ein Teilnehmer ein anderes Verfahren anwendet als dasjenige, welches angewendet wird, um den zugewiesenen Wert zu bestimmen, dann können seine Ergebnisse selbst dann eine systematische Abweichungskomponente aufweisen, wenn kein Fehler in der Durchführung vorliegt.

Wird den Teilnehmern Wahlfreiheit zwischen Konventionsverfahren eingeräumt, kann es sein, dass kein gültiger Konsens unter ihnen ersichtlich ist. Dieses Problem kann auf zweierlei Weise gelöst werden:

- a) Wenn die Teilnehmer routinemäßig ein standardisiertes Verfahren anwenden, wird dieses zur Ermittlung des zugewiesenen Werts angewendet, und die Teilnehmer werden angewiesen, dieses Verfahren in der Eignungsprüfung anzuwenden;
- b) für jedes angewendete Verfahren wird ein eigener zugewiesener Wert bestimmt.

Eine ähnliche Situation ergibt sich, wenn die Messgröße vorgegeben wird, aber nicht das Verfahren, und es muss dieselbe Wahl getroffen werden.

4.6 Angabe von Daten

(siehe ISO/IEC Guide 43-1:1997, 6.2.3)

Für die Zwecke der in Eignungsprüfungen erforderlichen Rechnungen wird empfohlen, dass einzelne Messergebnisse nicht um mehr als $\sigma_r/2$ gerundet werden sollten.

Teilnehmer müssen angehalten werden, ihre tatsächlichen Messergebnisse anzugeben. Messergebnisse dürfen nicht verkürzt als einseitig beschränkte Intervalle angegeben werden (das heißt Ergebnisse dürfen nicht in der Form „< 0,1“ oder „unterhalb der Nachweisgrenze“ wiedergegeben werden). Gleichermaßen gilt: Wenn ein negatives Ergebnis beobachtet wird, muss der tatsächliche negative Wert angegeben werden (das heißt Ergebnisse dürfen auch dann

4.5 Operationally defined measurement methods

With an operationally defined measurement method, the measurement result is defined by the measurement procedure. For example, the size distribution of a particulate material may be determined using either square-holed or round-holed sieves. There may be no good reason why one type of sieve should be preferred, but unless the type of sieve is specified, laboratories that use different types of sieves may obtain differing results. If a participant uses a different method from that used to establish the assigned value, then their results may show a bias when no fault in execution is present.

If participants are free to choose between operationally defined methods, no valid consensus may be evident amongst them. Two recourses are available to overcome this problem:

- a) when a standardized method is in routine use by the participants, it is used to establish the assigned value, and participants are instructed to use it in the proficiency test;
- b) a separate value of the assigned value is produced for each method used.

A similar situation arises when the measurand is specified, but not the procedure, and the same choice has to be made.

4.6 Reporting of data

(see ISO/IEC Guide 43-1:1997, 6.2.3)

For the purposes of the calculations required in proficiency testing, it is recommended that individual measurement results should not be rounded by more than $\sigma_r/2$.

Participants shall be asked to report the actual values of their measurement results. Measurement results shall not be truncated (i.e. results shall not be reported in the form „< 0,1“ or “less than the detection limit”). Likewise, when a negative result is observed, the actual negative value shall be reported (i.e. results shall not be reported as zero even when logically the measurement result cannot be negative). Participants shall be informed that if they report truncated results on a sample, or zeros

nicht als null wiedergegeben werden, wenn das Messergebnis logischerweise nicht negativ sein kann). Die Teilnehmer müssen darüber in Kenntnis gesetzt werden, dass alle Daten für eine Probe von der Auswertung ausgeschlossen werden, wenn für diese Probe verkürzte Ergebnisse oder der Wert Null an Stelle von negativen Ergebnissen angegeben werden. Falls notwendig, darf auf dem zur Angabe der Ergebnisse verwendeten Formular die Möglichkeit vorgesehen werden, dass der Teilnehmer ankreuzt, dass ein Ergebnis unterhalb der Nachweisgrenze liegt.

4.7 Gültigkeitsdauer der Ergebnisse der Eignungsprüfung

Die Gültigkeitsdauer des von einem Laboratorium in einer einzelnen Runde eines Schemas für Eignungsprüfungen ermittelten Ergebnisses ist auf die Zeit beschränkt, während der das Laboratorium die Prüfung durchführte. Wenn also ein Laboratorium bei einer einzelnen Runde ein zufrieden stellendes Ergebnis erzielt, darf dieses Ergebnis nicht als Beleg für die Behauptung verwendet werden, das Laboratorium hätte bei anderen Gelegenheiten zuverlässige Daten gewonnen.

Ein Laboratorium, das ein Qualitätssystem anwendet und über eine große Zahl von Runden eines Schemas für Eignungsprüfungen zufrieden stellende Ergebnisse dokumentieren kann, ist berechtigt, diese Ergebnisse als Beleg dafür zu verwenden, dass es in der Lage ist, konsistent zuverlässige Daten zu erhalten.

5 Bestimmung des zugewiesenen Werts und der zugehörigen Standardunsicherheit

5.1 Wahl des Verfahrens zur Ermittlung des zugewiesenen Werts

Fünf Arten, den zugewiesenen Wert X zu bestimmen, werden in 5.2 bis 5.6. beschrieben. Die Wahl zwischen diesen Verfahren liegt in der Verantwortung des Koordinators in Abstimmung mit technischen Fachleuten, wie in ISO/IEC Guide 43-1 beschrieben. Die in 5.5 und 5.6 beschriebenen Verfahren sind wahrscheinlich nicht anwendbar, wenn die Anzahl der an einem Plan teilnehmenden Laboratorien klein ist. Die Verfahren zur Berechnung der Standardunsicherheit u_X des zugewiesenen Werts, die in diesem Abschnitt angegeben werden, werden üblicherweise für diejenigen Anwendungen geeignet sein, für die sie in dieser Internationalen Norm angewendet werden. Andere

when results are negative, then all the data for that sample will be excluded from the analysis. If necessary, the form used to report results may contain a box to allow a participant to indicate that a result is below the detection limit.

4.7 Period of validity of the results of proficiency tests

The period of validity of the result obtained by a laboratory in a single round of a proficiency testing scheme is limited to the time that the laboratory performed the test. Thus, if a laboratory achieves a satisfactory result in a single round, the result shall not be used to support a claim that the laboratory obtained reliable data on any other occasion.

A laboratory that operates a quality system and achieves a history of satisfactory results in many rounds of a proficiency testing scheme shall be entitled to use the results as evidence that it is able to obtain consistently reliable data.

5 Determination of the assigned value and its standard uncertainty

5.1 Choice of method of determining the assigned value

Five ways of determining the assigned value X are described in 5.2 to 5.6. The choice between these methods shall be the responsibility of the coordinator following consultation with technical experts as described in ISO/IEC Guide 43-1. The methods described in 5.5 and 5.6 are unlikely to be applicable when the number of laboratories participating in the scheme is small. The methods of calculating the standard uncertainty, u_X , of the assigned value given in this clause will usually be adequate for the applications for which they are used in this International Standard. Alternative methods may be used provided that they have a sound statistical basis and that the method used is described in the documented plan for the scheme.

Verfahren mit gesicherter statistischer Grundlage dürfen angewendet werden; das angewendete Verfahren ist dann im dokumentierten Plan zu beschreiben.

Die Bestimmung des zugewiesenen Werts liegt in der Verantwortung des Koordinators. Der zugewiesene Wert darf den Teilnehmern erst offenbart werden, wenn sie ihre Ergebnisse beim Koordinator gemeldet haben. Der Koordinator muss einen Bericht erstellen, aus dem Einzelheiten der Bestimmung des zugewiesenen Werts, die Identitäten der Laboratorien, die an seiner Bestimmung beteiligt waren, und Angaben zur Rückführbarkeit und Messunsicherheit des zugewiesenen Werts hervorgehen.

Der *Leitfaden zur Angabe der Unsicherheit beim Messen* gibt Hinweise zur Beurteilung von Messunsicherheiten.

Diese Internationale Norm empfiehlt die Verwendung von robusten statistischen Verfahren, wenn diese als die am besten geeigneten Verfahren angesehen werden (zum Beispiel wie in 5.5 und 5.6). Alternativ dürfen Verfahren zur Erkennung und Beseitigung von Ausreißern zur Anwendung gebracht werden, vorausgesetzt, dass die statistische Grundlage dieser Verfahren abgesichert ist und dass das angewendete Verfahren dokumentiert wird. Hinweise für die Anwendung von Prüfungen auf Ausreißer werden in ISO 5725-2 angegeben.

5.2 Bekannte Werte aus der Probenherstellung

[siehe ISO/IEC Guide 43-1:1997, A.1.1, Aufzählung a)]

5.2.1 Allgemeines

Das Prüfmaterial kann durch Mischen von Bestandteilen in festgelegten Anteilen oder durch das Hinzufügen eines festgelegten Anteils eines Bestandteils zu einem Ausgangsstoff hergestellt werden. In diesem Fall wird der zugewiesene Wert X durch Berechnung aus den verwendeten Massen hergeleitet.

Dieser Ansatz ist besonders wertvoll, wenn auf diese Weise Einzelproben hergestellt werden können und die Anteile der Bestandteile oder des Zusatzstoffs zu bestimmen sind: es besteht dann keine Notwendigkeit, einen Probenansatz herzustellen und sicherzustellen, dass dieser homogen ist. Wenn sich jedoch dadurch Proben ergeben, bei denen der Zusatzstoff weniger fest gebunden ist als in üblichen Materialien oder bei denen er in einer unterschiedlichen Form vorliegt, so kann die Anwendung eines anderen Ansatzes sinnvoller sein.

The determination of the assigned value shall be the responsibility of the coordinator. The assigned value shall not be disclosed to the participants until they have reported their results to the coordinator. The coordinator shall prepare a report giving details of how the assigned value was obtained, the identities of laboratories involved in its determination, and statements of the traceability and measurement uncertainty of the assigned value.

The *Guide to the expression of uncertainty in measurement* gives guidance on the evaluation of measurement uncertainties.

This International Standard recommends the use of robust statistical methods when it is considered that they are the most appropriate methods to use (for example, as in 5.5 and 5.6). Alternatively, procedures that involve the detection and removal of outliers may be used provided that they have a sound statistical basis and the method that is used is reported. Guidance on the use of tests for outliers is given in ISO 5725-2.

5.2 Formulation

[see ISO/IEC Guide 43-1:1997, A.1.1, item a)]

5.3.1 General

The test material may be prepared by mixing constituents in specified proportions, or by adding a specified proportion of a constituent to a base material. In this case, the assigned value X is derived by calculation from the masses used.

The approach is especially valuable when individual samples may be prepared in this way, and it is the proportion of the constituents or of the addition that is to be determined: there is then no need to prepare a bulk quantity and ensure that it is homogeneous. However, when formulation gives samples in which the addition is more loosely bonded than in typical materials, or in a different form, it may be preferable to use another approach.

5.2.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn der zugewiesene Wert aus der Zusammensetzung des Prüfmaterials berechnet wird, wird die Standardunsicherheit durch Kombination der Unsicherheiten nach dem im *Leitfaden zur Angabe der Unsicherheit beim Messen* beschriebenen Ansatz abgeschätzt. Zum Beispiel werden die Unsicherheiten in chemischen Analysen normalerweise jene sein, die mit gravimetrischen und volumetrischen Messungen verbunden sind.

Die Einschränkung dieses Verfahrens (bei der chemischen Analyse) besteht darin, dass Sorgfalt nötig ist, um sicherzustellen, dass

- a) der Ausgangsstoff effektiv keinen Anteil des hinzugefügten Bestandteils enthält oder dass der Anteil des hinzugefügten Bestandteils im Ausgangsstoff genau bekannt ist;
- b) die Bestandteile homogen vermischt werden (soweit erforderlich);
- c) alle Fehlerquellen ermittelt werden (zum Beispiel wird nicht immer erkannt, dass Glas Quecksilberverbindungen absorbiert, so dass die Konzentration einer wässrigen Lösung einer Quecksilberverbindung durch den Behälter verändert werden kann);
- d) keine Wechselwirkung zwischen den Bestandteilen und der Probenmatrix stattfindet.

5.2.3 Beispiel: Bestimmung des Zementgehalts von Festbeton

In diesem Fall können Betonproben durch das Abwiegen von Mengen der Bestandteile (Zement, Zuschläge und Wasser) und Mischen der Komponenten erstellt werden, um jede einzelne Betonprobe zu erzeugen. Der Ansatz ist zufriedenstellend, weil die Genauigkeit, mit der Proben erstellt werden können, weit höher ist als bei dem analytischen Verfahren zur Bestimmung des Zementgehalts.

5.3 Zertifizierte Referenzwerte

[siehe ISO/IEC Guide 43-1:1997, A.1.1 Aufzählung b)]

5.3.1 Allgemeines

Wenn das bei einer Eignungsprüfung verwendete Material ein zertifiziertes Referenzmaterial (en: certified reference material, CRM) ist, wird sein zertifizierter Referenzwert als der zugewiesene Wert X verwendet.

5.2.2 Standard uncertainty u_X of the assigned value

When the assigned value is calculated from the formulation of the test material, the standard uncertainty is estimated by combination of uncertainties using the approach described in the *Guide to the expression of uncertainty in measurement*. For example, in chemical analyses the uncertainties will usually be those associated with gravimetric and volumetric measurements.

The limitation of this method (in chemical analysis) is that care is needed to ensure that:

- a) the base material is effectively free from the added constituent, or that the proportion of the added constituent in the base material is accurately known;
- b) the constituents are mixed together homogeneously (where this is required);
- c) all sources of error are identified (e.g. it is not always realized that glass absorbs mercury compounds, so that the concentration of an aqueous solution of a mercury compound can be altered by its container);
- d) there is no interaction between the constituents and the matrix.

5.2.3 Example: Determination of the cement content of hardened concrete

In this case, concrete specimens may be prepared by weighing out quantities of the constituents (cement, aggregates and water) and mixing them together to form each concrete sample. The approach is satisfactory because the accuracy with the specimens can be prepared is far superior to that of the analytical method used to determine the cement content.

5.3 Certified reference values

[see ISO/IEC Guide 43-1:1997, A.1.1 item b)]

5.3.1 General

When the material used in a proficiency test is a certified reference material (CRM), its certified reference value is used as the assigned value X .

5.3.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn ein zertifiziertes Referenzmaterial als das Prüfmaterial verwendet wird, wird die Standardunsicherheit des zugewiesenen Werts aus den Angaben zur Unsicherheit hergeleitet, die dem Zertifikat zu entnehmen sind.

Die Einschränkung dieses Ansatzes besteht darin, dass die Bereitstellung einer Probe des zertifizierten Referenzmaterials für jeden Teilnehmer einer Eignungsprüfung aufwändig sein kann.

5.3.3 Beispiel: Los-Angeles-Wert von Zuschlagstoffen

Der „Los-Angeles-Wert“ ist ein Maß für die mechanische Festigkeit von Zuschlagstoffen, die im Straßenbau verwendet werden, und Ergebnisse der Prüfung werden in „LA-Einheiten“ ausgedrückt. Anlässlich eines Ringversuchs, der das Ziel verfolgte, ein Referenzmaterial zu zertifizieren, wurde eine große Anzahl von Proben eines bestimmten Zuschlagstoffs vorbereitet, und einige dieser Proben wurden bei einem Ringversuch verwendet, an dem 28 Laboratorien beteiligt waren. Dieser Ringversuch gestattete die Bestimmung eines zugewiesenen Werts von $X_{\text{CRM}} = 21,62$ LA-Einheiten mit einer Standardunsicherheit von $u_{X,\text{CRM}} = 0,26$ LA-Einheiten. Die verbleibenden Proben dieses Zuschlagstoffs konnten bei Eignungsprüfungen verwendet werden.

5.4 Referenzwerte

[siehe ISO/IEC Guide 43-1:1997, A.1.1 Aufzählung c)]

5.4.1 Allgemeines

Bei diesem Ansatz werden zunächst Proben des Prüfmaterials, das als Referenzmaterial (en: reference material, RM) vorgesehen ist, zur Verteilung an die Teilnehmer vorbereitet. Eine Anzahl zufällig ausgewählter Proben wird dann zusammen mit zertifizierten Referenzmaterialien in einem Laboratorium, das ein geeignetes Messverfahren anwendet, und unter Wiederholbedingungen (wie definiert in ISO 3534-2) geprüft. Der zugewiesene Wert X_{RM} des Prüfmaterials wird dann von einer Kalibrierung gegen die zertifizierten Referenzwerte der CRM abgeleitet.

5.3.2 Standard uncertainty u_X of the assigned value

When a certified reference material is used as the test material, the standard uncertainty of the assigned value is derived from the information on uncertainty provided on the certificate.

The limitation of this approach is that it can be expensive to provide every participant in a proficiency test with a sample of a certified reference material.

5.3.3 Example: Los Angeles value of aggregates

The “Los Angeles value” is a measure of the mechanical strength of aggregates that are used for road construction, and results of the test are measured in “LA units”. In an exercise to certify a reference material, a large number of samples of a particular aggregate were prepared, and some of these samples were used in an interlaboratory experiment involving 28 laboratories, allowing an assigned value of $X_{\text{CRM}} = 21,62$ LA units to be established with a standard uncertainty of $u_{X,\text{CRM}} = 0,26$ LA units. The remaining samples of this aggregate could be used in proficiency tests.

5.4 Reference values

[see ISO/IEC Guide 43-1:1997, A.1.1 item c)]

5.4.1 General

In this approach, samples of the test material that is to be the reference material (RM) are prepared first, ready for distribution to the participants. A number of the samples are then selected at random and tested along with certified reference materials, in one laboratory, using a suitable measurement method, and under repeatability conditions (as defined in ISO 3534-2). The assigned value X_{RM} of the test material is then derived from a calibration against the certified reference values of the CRMs.

5.4.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn der zugewiesene Wert eines Prüfmaterials von den Ergebnissen einer Reihe von Prüfungen an diesem Material und am CRM abgeleitet wird, wird die Standardunsicherheit des zugewiesenen Werts aus den Prüfergebnissen und den Unsicherheiten der zertifizierten Referenzwerte des CRMs abgeleitet. Wenn das Prüfmaterial und das CRM (hinsichtlich Matrix, Zusammensetzung und Größenordnung der Ergebnisse) nicht ähnlich sind, dann ist die sich daraus ergebende Unsicherheit ebenfalls einzubeziehen.

Dieses Verfahren gestattet es, den zugewiesenen Wert in einer Weise zu bestimmen, die auf die zertifizierten Werte der CRM rückführbar ist. Seine Standardunsicherheit kann berechnet werden. Das Verfahren vermeidet auch die Kosten für die Verteilung des CRMs an alle Teilnehmer. Dies sind gute Gründe dafür, es anderen Verfahren vorzuziehen. Jedoch geht das Verfahren davon aus, dass es keine Wechselwirkungen zwischen den verwendeten Materialien und den Prüfbedingungen gibt.

Das Beispiel in 5.4.3 veranschaulicht die Berechnung der erforderlichen Unsicherheit für den einfachen Fall, in welchem der zugewiesene Wert eines Prüfmaterials durch direkten Vergleich mit einem einzelnen CRM ermittelt wird.

5.4.3 Beispiel: Los-Angeles-Wert von Zuschlagstoffen

Das im Beispiel in 5.3 beschriebene CRM kann verwendet werden, um den zugewiesenen Wert für ein RM zu bestimmen, wobei es sich um einen anderen, ähnlichen Zuschlagstoff handelt. Diese Bestimmung erfordert die Durchführung einer Reihe von Prüfungen an Proben der beiden Zuschlagstoffe in einem Laboratorium unter Verwendung desselben Messverfahrens und unter Wiederholbedingungen. Es sei

X_{CRM} der zugewiesene Wert für das CRM

X_{RM} der zugewiesene Wert für das RM

D_i die Differenz (RM – CRM) zwischen den gemittelten Ergebnissen für das RM und das CRM bei den Proben mit der Zählnummer i

$\bar{D}$ der Mittelwert der Differenzen D_i

5.4.2 Standard uncertainty u_X of the assigned value

When the assigned value of a test material is derived from the results of a series of tests on that material and on CRM, the standard uncertainty of the assigned value is derived from the test results, and the uncertainties of the certified reference values of the CRM. If the test material and the CRM are not similar (in matrix, composition and level of results), then the uncertainty arising from this is also to be included.

This method allows the assigned value to be established in a manner that is traceable to the certified values of the CRMs, with a standard uncertainty that can be calculated, and avoids the cost of distributing the CRM to all the participants. These are good reasons for preferring it to other methods. However, the method assumes that there are no interactions between the materials used and the test conditions.

The example in 5.4.3 illustrates how the required uncertainty may be calculated in the simple case when the assigned value of a test material is established by direct comparison with a single CRM.

5.4.3 Example: Los Angeles value of aggregates

The CRM described in the example in 5.3 may be used to determine the assigned value for an RM which is another, similar, aggregate. This determination requires a series of tests to be carried out, in one laboratory, on samples of the two aggregates, using the same measurement method, and under repeatability conditions. Let

X_{CRM} is the assigned value for the CRM

X_{RM} is the assigned value for the RM

D_i is the difference (RM – CRM) between the average results for the RM and the CRM on the i^{th} samples

$\bar{D}$ is the average of the differences D_i

Dann gilt

$$X_{\text{RM}} = X_{\text{CRM}} + \bar{D} \quad (3)$$

Die Standardunsicherheit des zugewiesenen Werts des RM kann wie folgt berechnet werden:

$$u_{X;\text{RM}} = \sqrt{u_{X;\text{CRM}}^2 + u_D^2} \quad (4)$$

Tabelle 1 gibt ein Beispiel für Daten, die sich bei einer solchen Reihe von Prüfungen ergeben könnten, und zeigt, wie die Standardunsicherheit u_D der Differenz berechnet wird.

Mit diesen Ergebnissen gilt

$$X_{\text{RM}} = 21,62 + 1,73 = 23,35 \text{ LA-Einheiten} \quad (5)$$

und

$$u_{X;\text{RM}} = \sqrt{0,26^2 + 0,24^2} = 0,35 \text{ LA-Einheiten} \quad (6)$$

Dabei ist 0,26 die Standardunsicherheit des zugewiesenen Werts für das (im Beispiel in 5.3 gegebene) CRM, und 0,24 ist die Standardunsicherheit von $\bar{D}$.

Then

$$X_{\text{RM}} = X_{\text{CRM}} + \bar{D} \quad (3)$$

The standard uncertainty of the assigned value of the RM may be calculated as:

$$u_{X;\text{RM}} = \sqrt{u_{X;\text{CRM}}^2 + u_D^2} \quad (4)$$

Table 1 gives an example of data that might be obtained in such a series of tests, and shows how the standard uncertainty u_D of the difference is calculated.

With these results,

$$X_{\text{RM}} = 21,62 + 1,73 = 23,35 \text{ LA units} \quad (5)$$

and

$$u_{X;\text{RM}} = \sqrt{0,26^2 + 0,24^2} = 0,35 \text{ LA units} \quad (6)$$

where 0,26 is the standard uncertainty of the assigned value of the CRM (given in the example in 5.3), and 0,24 is the standard uncertainty of $\bar{D}$.

Table 1 — Berechnung der mittleren Differenz zwischen einem CRM und einem RM und der Standardunsicherheit dieser Differenz

Table 1 — Calculation of the average difference between a CRM and a RM, and of the standard uncertainty of this difference

Probe/ Sample	RM		CRM		Differenz der Mittelwerte/ Difference in average values RM – CRM LA-Einheiten/ LA units
	Prüfung 1/ Test 1 LA-Einheiten/ LA units	Prüfung 2/ Test 2 LA-Einheiten/ LA units	Prüfung 1/ Test 1 LA-Einheiten/ LA units	Prüfung 2/ Test 2 LA-Einheiten/ LA units	
1	20,5	20,5	19,0	18,0	2,00
2	21,1	20,7	19,8	19,9	1,05
3	21,5	21,5	21,0	21,0	0,50
4	22,3	21,7	21,0	20,8	1,10
5	22,7	22,3	20,5	21,0	1,75
6	23,6	22,4	20,3	20,3	2,70
7	20,9	21,2	21,5	21,8	–0,60
8	21,4	21,5	21,9	21,7	–0,35
9	23,5	23,5	21,0	21,0	2,50
10	22,3	22,9	22,0	21,3	0,95
11	23,5	24,1	20,8	20,6	3,10
12	22,5	23,5	21,0	22,0	1,50
13	22,5	23,5	21,0	21,0	2,00
14	23,4	22,7	22,0	22,0	1,05
15	24,0	24,2	22,1	21,5	2,30
16	24,5	24,4	22,3	22,5	2,05
17	24,8	24,7	22,0	21,9	2,80
18	24,7	25,1	21,9	21,9	3,00
19	24,9	24,4	22,4	22,6	2,15
20	27,2	27,0	24,5	23,7	3,00
Mittlere Differenz, $\bar{D}$ / Average difference, $\bar{D}$					1,73
Standardabweichung/Standard deviation					1,07
Standardunsicherheit von $\bar{D}$ / Standard uncertainty of $\bar{D}$ (Standardabweichung/ $\sqrt{20}$) / (Standard deviation/ $\sqrt{20}$)					0,24
ANMERKUNG Bei den Daten handelt es sich um Messungen der mechanischen Festigkeit des Zuschlagstoffs nach der Los-Angeles-(LA-)Prüfung.					
NOTE The data are measurements of the mechanical strength of aggregate, obtained from the Los Angeles (LA) test.					

5.5 Konsenswerte von Expertenlaboratorien

[siehe ISO/IEC Guide 43-1:1997, A.1.1 Aufzählung d)]

5.5.1 Allgemeines

Wie beim Ansatz unter Verwendung von *Referenzwerten* (5.4) werden zunächst Proben des Prüfmaterials zur Verteilung an die Teilnehmer vorbereitet. Einige dieser Proben werden dann von einer Gruppe von Expertenlaboratorien zufällig ausgewählt und analysiert. Alternativ kann die Gruppe von Expertenlaboratorien Teilnehmer einer Runde eines Schemas für Eignungsprüfungen sein, wenn der zugewiesene Wert und seine Unsicherheit nach Beendigung der Runde abgeleitet werden. Der zugewiesene Wert X wird berechnet als robuster Mittelwert der Ergebnisse (berechnet nach Algorithmus A im Anhang C), die von der Gruppe von Expertenlaboratorien angegeben wurden.

Anstelle von Algorithmus A dürfen andere Rechenverfahren angewendet werden, vorausgesetzt, dass sie eine verlässliche statistische Grundlage haben und das verwendete Verfahren im Bericht beschrieben wird.

5.5.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn jedes einzelne von p erfahrenen Laboratorien eine Messung x_i am Prüfmaterial zusammen mit einem Schätzwert u_i der Standardunsicherheit der Messung angibt und der zugewiesene Wert X als robuster Mittelwert nach Algorithmus A berechnet wird, wird die Standardunsicherheit des zugewiesenen Werts X wie folgt abgeschätzt:

$$u_X = \frac{1,25}{p} \times \sqrt{\sum_{i=1}^p u_i^2} \quad (7)$$

Geben die erfahrenen Laboratorien keine Standardunsicherheiten an oder sind die Unsicherheiten nicht unabhängig validiert (zum Beispiel durch eine Akkreditierungsstelle für Laboratorien), ist die Standardunsicherheit des zugewiesenen Werts wie in 5.6 beschrieben abzuschätzen.

ANMERKUNG Der Faktor 1,25 bildet das Verhältnis der Standardabweichung des Medians zur Standardabweichung des arithmetischen Mittelwerts für große Probenumfänge ($p > 10$) bei einer Normalverteilung ab. Bei normalverteilten Daten ist die Standardabweichung eines nach dem Algorithmus im Anhang C berechneten robusten Mittelwerts unbekannt; sie wird jedoch zwischen

5.5 Consensus values from expert laboratories

[see ISO/IEC Guide 43-1:1997, A.1.1 item d)]

5.5.1 General

As with the *reference values* approach (5.4), samples of the test material are prepared first, ready for distribution to the participants. Some of these samples are then selected at random and analysed by a group of expert laboratories. Alternatively, the group of expert laboratories may be participants in a round of a proficiency testing scheme, when the assigned value and its uncertainty will be derived after the round is completed. The assigned value X is calculated as the robust average of the results reported by the group of expert laboratories, calculated using Algorithm A in Annex C.

Other calculation methods may be used in place of Algorithm A, provided that they have a sound statistical basis and the report describes the method that is used.

5.5.2 Standard uncertainty u_X of the assigned value

When each of p expert laboratories reports a measurement x_i on the test material together with an estimate u_i of the standard uncertainty of the measurement, and the assigned value X is calculated as a robust average using Algorithm A, the standard uncertainty of the assigned value X is estimated as:

$$u_X = \frac{1,25}{p} \times \sqrt{\sum_{i=1}^p u_i^2} \quad (7)$$

When the expert laboratories do not report standard uncertainties, or when the uncertainties are not validated independently (e.g. by a laboratory accreditation body), the standard uncertainty of the assigned value shall be estimated as described in 5.6.

NOTE The factor 1,25 represents the ratio of the standard deviation of the median to the standard deviation of the arithmetic mean, for large samples ($p > 10$) from a normal distribution. For normally distributed data, the standard deviation of a robust average calculated using the algorithm in Annex C is not known, but will fall somewhere between the standard

der Standardabweichung des arithmetischen Mittelwerts und der Standardabweichung des Medians liegen, so dass die Gleichung einen konservativen Schätzwert für die Standardunsicherheit u_X ergibt. Für $p < 10$ ist der geeignete Faktor kleiner als 1,25, so dass die Gleichung doppelt konservativ ist.

Die Einschränkungen dieses Ansatzes bestehen darin, dass die Ergebnisse der Gruppe von Expertenlaboratorien eine unbekannte systematische Abweichungskomponente enthalten können und dass die angegebenen Unsicherheiten unzuverlässig sein können.

5.5.3 Beispiel: Petrografische Analyse von Zuschlagstoffen

Dieser Ansatz kann angewendet werden, wenn Proben der Zuschlagstoffe verteilt werden sollen und die Teilnehmer die Aufgabe haben, die petrografische Zusammensetzung der Proben zu bestimmen. Die Klassifizierung von Zuschlagstoffen erfordert Sachkenntnis und Erfahrung und es gibt keine Referenzmaterialien. Daher kann in diesem Fall der Konsens einer kleinen Gruppe von Fachleuten der am besten geeignete Weg zur Feststellung von zugewiesenen Werten sein.

5.6 Konsenswert der Teilnehmer [siehe ISO/IEC Guide 43-1:1997, A.1.1 Aufzählung e)]

5.6.1 Allgemeines

Bei diesem Ansatz ist der zugewiesene Wert X für das bei einer Runde eines Schemas für Eignungsprüfungen verwendete Prüfmaterial der robuste Mittelwert der Ergebnisse, die von allen Teilnehmern an dieser Runde angegeben werden; er wird nach Algorithmus A im Anhang C berechnet.

Andere Berechnungsverfahren als Algorithmus A dürfen verwendet werden, vorausgesetzt, dass sie auf einer gesicherten statistischen Grundlage beruhen und dass der Bericht das zur Anwendung gebrachte Verfahren beschreibt. So kann zum Beispiel die Berechnung in C.1 ^{N2)} angehalten werden, wenn der Median ermittelt wird, und an Stelle von C.2 darf der Median des Betrags der Abweichung verwendet werden.

Dieser Ansatz kann sich bei einem Konventionsmessverfahren als besonders nützlich erweisen, vorausgesetzt, dieses Verfahren ist standardisiert.

deviation of the arithmetic mean and the standard deviation of the median, so the formula gives a conservative estimate of the standard uncertainty u_X . For $p < 10$, the appropriate factor is less than 1,25, so the formula is then doubly conservative.

The limitations of this approach are that there may be an unknown bias in the results of the group of expert laboratories, and the claimed uncertainties may not be reliable.

5.5.3 Example: Petrographic analysis of aggregates

This approach may be used when samples of aggregates are to be distributed, and the participants are to determine the petrographic composition of the samples. Classification of aggregates requires skill and experience, and there are no reference materials available, so in this case the consensus of a small group of experts may well be the best way to establish assigned values.

5.6 Consensus value from participants [see ISO/IEC Guide 43-1:1997, A.1.1 item e)]

5.6.1 General

With this approach, the assigned value X for the test material used in a round of a proficiency testing scheme is the robust average of the results reported by all the participants in the round, calculated using Algorithm A in Annex C.

Other calculation methods may be used in place of Algorithm A, provided that they have a sound statistical basis and the report describes the method that is used. For example, the calculation in C.1 may be stopped at Equation (53) when the median is obtained, and the median absolute deviation may be used in place of C.2.

This approach may be particularly useful with an operationally defined measurement method, provided that the method is standardized.

N2) Nationale Fußnote: Die Erwähnung von „Equation (53)“ ist ein redaktioneller Fehler im Englischen und wird im deutschen Text fortgelassen.

5.6.2 Standardunsicherheit u_X des zugewiesenen Werts

Wenn der zugewiesene Wert als robuster Mittelwert nach Algorithmus A hergeleitet wird, wird die Standardunsicherheit des zugewiesenen Werts X wie folgt abgeschätzt:

$$u_X = 1,25 \times s^* / \sqrt{p} \quad (8)$$

Dabei ist s^* die robuste Standardabweichung der Ergebnisse, berechnet nach Algorithmus A im Anhang C. (Hier gilt als „Ergebnis“ für einen Teilnehmer der Mittelwert aller seiner Ergebnisse mit dem Prüfmateri al.)

Die Einschränkungen dieses Ansatzes bestehen darin, dass

- a) möglicherweise kein tatsächlicher Konsens unter den Teilnehmern besteht;
- b) der Konsens auf Grund von allgemein üblichen Verfahrensfehlern eine systematische Abweichungskomponente enthalten kann, die sich nicht in der Standardunsicherheit des zugewiesenen Werts, die wie oben beschrieben berechnet wird, widerspiegelt.

Keine dieser Bedingungen kommt bei der Ermittlung von Spurenbestandteilen selten vor.

5.6.3 Beispiel: Antikörper-Konzentrationen

Tabelle 2 listet Daten einer Runde einer Eignungsprüfung auf, bei der Konzentrationen von drei allergenspezifischen IgE-Antikörpern (Immunglobuline E) bestimmt wurden. Bild 2 stellt dieselben Daten in Histogrammform dar.

Zur Anwendung von Algorithmus A werden die Daten zunächst in aufsteigender Reihenfolge sortiert und die anfänglichen robusten Schätzwerte für den Mittelwert und die Standardabweichung werden berechnet. Nachfolgend wird das iterative Verfahren des Algorithmus angewendet. Tabelle 3 stellt diese Berechnungen für die Ergebnisse für den allergenspezifischen IgE-Antikörper d1 aus Tabelle 2 dar.

Die im Rahmen der Anwendung von Algorithmus A erforderlichen Berechnungen können wie folgt in einer Tabellenkalkulation durchgeführt werden.

5.6.2 Standard uncertainty u_X of the assigned value

When the assigned value is derived as a robust average calculated using Algorithm A, the standard uncertainty of the assigned value X is estimated as:

$$u_X = 1,25 \times s^* / \sqrt{p} \quad (8)$$

where s^* is the robust standard deviation of the results calculated using Algorithm A in Annex C. (Here a “result” for a participant is the average of all their measurements on the test material.)

The limitations of this approach are that:

- a) there may be no real consensus amongst the participants;
- b) the consensus may be biased by the general use of faulty methodology and this bias will not be reflected in the standard uncertainty of the assigned value calculated as described above.

Neither of these conditions is rare in the determination of trace constituents.

5.6.3 Example: Antibody concentrations

Table 2 gives data from a round of a proficiency test in which concentrations of three allergen-specific IgE (immunoglobulin E) antibodies were determined. Figure 2 displays these same data in histograms.

To apply Algorithm A, the data are first sorted into ascending order, initial robust estimates of the average and standard deviation are calculated, and then the iterative method of the algorithm is applied. Table 3 gives these calculations for the results for the allergen specific IgE antibody d1 from Table 2.

The calculations required by Algorithm A may be carried out in a spreadsheet as follows.

- a) **Schritt 1:** Die Daten werden in aufsteigender Reihenfolge in eine Spalte eingetragen, wie in Tabelle 3 für Iterationsschritt 0 ausgewiesen. Ihr Mittelwert und ihre Standardabweichung werden berechnet (10,91 und 3,13 in Tabelle 3). Anfängliche Werte für den robusten Mittelwert und die robuste Standardabweichung (10,85 und 3,53 in Tabelle 3) werden nach den Gleichungen in C.1 berechnet.
- b) **Schritt 2:** Die Daten werden in die nächste Spalte kopiert, wie in Tabelle 3 für Iterationsschritt 1 ausgewiesen. Die anfänglichen Werte für den robusten Mittelwert und die robuste Standardabweichung werden unter Verwendung der in C.1 angegebenen Gleichungen verwendet, um die Grenzwerte (5,56 und 16,15 in Tabelle 3) zu berechnen. Werte außerhalb der Grenzwerte werden durch die Grenzwerte ersetzt (2,18 wird durch 5,56 ersetzt, 16,30 durch 16,15). Der neue Mittelwert und die neue Standardabweichung der veränderten Daten (11,03 und 2,81 in Tabelle 3) werden berechnet. Nach den Gleichungen im Anhang C ist der robuste Mittelwert mit diesem Mittelwert (11,03) identisch und die robuste Standardabweichung (3,19) errechnet sich durch Multiplikation der Standardabweichung mit 1,134.
- c) **Schritt 3:** Bei einer Tabellenkalkulation entfällt jetzt die Notwendigkeit, weitere Datenreihen zu erzeugen. Stattdessen wird die Berechnung der Grenzwerte oben in der zweiten Spalte so verändert, dass diese den robusten Mittelwert und die robuste Standardabweichung vom unteren Ende derselben Spalte verwendet. Dies ergibt die Grenzwerte (6,24 und 15,82), die in Tabelle 3 unter dem Iterationsschritt 2 ausgewiesen werden. Die Berechnung kann dann vollendet werden, indem Daten außerhalb der Grenzwerte weiterhin durch die Grenzwerte ersetzt werden, bis die Iterationen konvergieren. Beim Ersetzen von Daten aktualisiert die Tabellenkalkulation automatisch die Mittelwerte und Standardabweichungen sowie Grenzwerte, jedoch werden die Änderungen dieser Werte zunehmend kleiner, bis sie schließlich insignifikant werden.

Die robusten Mittelwerte und Standardabweichungen für die beiden anderen allergenspezifischen IgE-Antikörper werden in ähnlicher Weise berechnet.

Interessanterweise wird das Ergebnis des Laboratoriums P für Antikörper d1 nach dem Grubbs-Test nicht als Ausreißer oder Nachzügler klassifiziert. Werden also bei diesem Beispiel Prüfungen auf

- a) **Step 1:** Enter the data in a column, in ascending order, as shown in Table 3 for Iteration 0. Calculate their average and standard deviation (10,91 and 3,13 in Table 3). Calculate initial values for the robust average and robust standard deviation (10,85 and 3,53 in Table 3) using the formulae given in C.1.
- b) **Step 2:** Copy the data into the next column, as shown in Table 3 for Iteration 1. Use the initial values for the robust average and robust standard deviation to calculate the cut-off values (5,56 and 16,15 in Table 3) using the formulae given in C.1. Replace data outside the cut-off values by the cut-off values (2,18 is replaced by 5,56 and 16,30 is replaced by 16,15). Calculate the new average and standard deviation of the altered data (11,03 and 2,81 in Table 3). According to the formulae given in Annex C, the robust average is the same as this average (11,03) and the robust standard deviation (3,19) is obtained by multiplying the standard deviation by 1,134.
- c) **Step 3:** With a spreadsheet, there is now no need to create further columns of data. Instead, change the calculation of the cut-off values at the top of the second column of data so that they use the robust average and robust standard deviation from the bottom of the same column. This will give the cut-off values (6,24 and 15,82) shown in Table 3 under Iteration 2. The calculation can then be completed by continuing to replace data outside the cut-off values by the cut-off values until the iterations converge. When data are replaced, the spreadsheet will automatically update the averages and standard deviations and cut-off values, but the changes in these values will become progressively smaller until they are no longer significant.

The robust averages and standard deviations for the other two allergen-specific IgE antibodies are calculated similarly.

It is interesting to note that the result for Laboratory P for d1 is not classed as an outlier or a straggler by Grubbs' test. Hence if one relies on the use of outlier tests as described in ISO 5725-2 in this example, the average and standard deviation

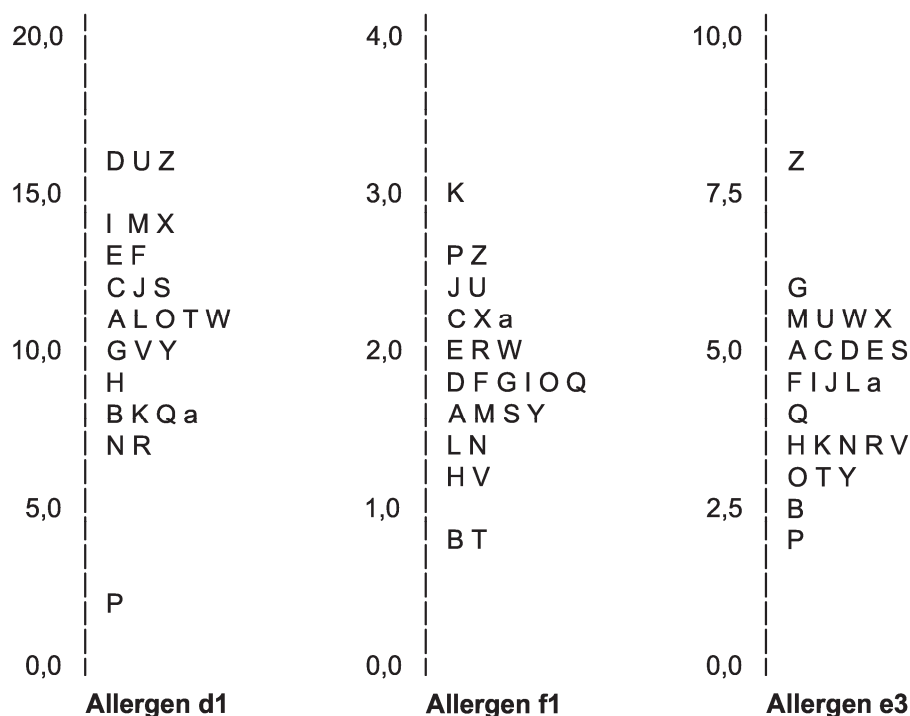
Ausreißer nach ISO 5725-2 verwendet, so werden Mittelwert und Standardabweichung aus allen Daten berechnet und es ergeben sich die Werte, die in Tabelle 3 für *Iterationsschritt 0* ausgewiesen sind. Beim robusten Verfahren haben der niedrige Wert des Laboratoriums P und die hohen Werte der Laboratorien D, U und Z keinen Einfluss auf die Werte der robusten Schätzwerte. Die Balkendiagramme (siehe Bild 9 in 8.3) würden die Werte einiger Laboratorien als einer weiteren Untersuchung würdig kennzeichnen. So zeigt zum Beispiel Bild 9, dass Laboratorium Z für alle drei Allergenspiegel hohe positive z -Werte aufweist.

are calculated from all the data, giving the values shown for *Iteration 0* in Table 3. With the robust method, the low result for Laboratory P and the high results for Laboratories D, U and Z have no influence on the values of the robust estimates. The bar-plots (see Figure 9 in 8.3) would identify the results of some laboratories as being worth investigation. For example, in Figure 9, Laboratory Z can be seen to give large positive z -scores at all three levels.

**Tabelle 2 — Konzentrationen dreier allergenspezifischer IgE-Antikörper (d1, f1 und e3) —
Daten wie von $p = 27$ Laboratorien angegeben**

**Table 2 — Concentrations of three allergen specific IgE antibodies (d1, f1 and e3) —
Data as reported by $p = 27$ laboratories**

Laboratorium/Laboratory	Konzentrationen/Concentrations		
	d1 kU/l	f1 kU/l	e3 kU/l
A	11,30	1,69	5,02
B	8,29	0,74	2,52
C	11,90	2,23	5,15
D	15,60	1,76	5,15
E	13,40	1,91	4,84
F	12,50	1,71	4,54
G	10,40	1,88	5,94
H	9,38	1,14	3,50
I	14,20	1,74	4,48
J	12,10	2,39	4,75
K	8,10	3,10	3,70
L	10,80	1,39	4,70
M	13,80	1,52	5,59
N	7,00	1,50	3,40
O	10,85	1,80	2,80
P	2,18	2,52	1,88
Q	8,39	1,83	3,80
R	6,95	1,92	3,52
S	11,80	1,58	4,86
T	10,90	0,80	2,80
U	16,30	2,39	5,60
V	9,71	1,21	3,33
W	10,50	1,93	5,35
X	13,60	2,23	5,53
Y	10,10	1,63	3,18
Z	16,07	2,69	8,22
a	8,47	2,16	4,64
p	27	27	27
Robuster Mittelwert x^* / Robust average x^*	11,03	1,83	4,35
Robuste Standardabweichung s^* / Robust standard deviation s^*	3,04	0,50	1,25
ANMERKUNG Bei den Daten handelt es sich um Anzahlen von Einheiten (en: Units, U) in eintausend (k) je Liter (l) Probe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.			
NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.			



ANMERKUNG 1 Bei den Daten handelt es sich um Anzahlen von Einheiten (en: Units, U) in eintausend (k) je Liter (l) Probe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.

NOTE 1 The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

ANMERKUNG 2 Die in Tabelle 2 angegebenen Zahlenwerte sind diejenigen, die ermittelt werden, wenn die Berechnung manuell bis auf zwei Dezimalen durchgeführt wird.

NOTE 2 The numerical values given in Table 2 are those that will be obtained when the calculation is carried out manually working to two decimal places.

Bild 2 — Konzentrationen dreier allergenspezifischer IgE-Antikörper — Histogramme für angegebene Daten (Daten aus Tabelle 2)

Figure 2 — Concentrations of three allergen specific IgE antibodies — Histograms of data as reported (data from Table 2)

**Tabelle 3 — Konzentrationen dreier allergenspezifischer IgE-Antikörper —
Berechnung des robusten Mittelwerts und der robusten Standardabweichung für Antikörper d1**

**Table 3 — Concentrations of three allergen specific IgE antibodies —
Calculation of the robust average and standard deviation for antibody d1**

Iterationsschritt/ Iteration	0	1	2	3	4	5
$\delta = 1,5 s^*$	—	5,30	4,79	4,62	4,58	4,56
$x^* - \delta$	—	5,56	6,24	6,41	6,45	6,47
$x^* + \delta$	—	16,15	15,82	15,65	15,61	15,59
P	2,18	5,56	6,24	6,41	6,45	6,47
R	6,95	6,95	6,95	6,95	6,95	6,95
N	7,00	7,00	7,00	7,00	7,00	7,00
K	8,10	8,10	8,10	8,10	8,10	8,10
B	8,29	8,29	8,29	8,29	8,29	8,29
Q	8,39	8,39	8,39	8,39	8,39	8,39
a	8,47	8,47	8,47	8,47	8,47	8,47
H	9,38	9,38	9,38	9,38	9,38	9,38
V	9,71	9,71	9,71	9,71	9,71	9,71
Y	10,10	10,10	10,10	10,10	10,10	10,10
G	10,40	10,40	10,40	10,40	10,40	10,40
W	10,50	10,50	10,50	10,50	10,50	10,50
L	10,80	10,80	10,80	10,80	10,80	10,80
O	10,85	10,85	10,85	10,85	10,85	10,85
T	10,90	10,90	10,90	10,90	10,90	10,90
A	11,30	11,30	11,30	11,30	11,30	11,30
S	11,80	11,80	11,80	11,80	11,80	11,80
C	11,90	11,90	11,90	11,90	11,90	11,90
J	12,10	12,10	12,10	12,10	12,10	12,10
F	12,50	12,50	12,50	12,50	12,50	12,50
E	13,40	13,40	13,40	13,40	13,40	13,40
X	13,60	13,60	13,60	13,60	13,60	13,60
M	13,80	13,80	13,80	13,80	13,80	13,80
I	14,20	14,20	14,20	14,20	14,20	14,20
D	15,60	15,60	15,60	15,60	15,60	15,59
Z	16,07	16,07	15,82	15,65	15,61	15,59
U	16,30	16,15	15,82	15,65	15,61	15,59
Mittelwert/ Average	10,91	11,03	11,03	11,03	11,03	11,03
Standardabweichung/ Standard deviation	3,13	2,81	2,72	2,69	2,68	2,68
Neuer Wert x^* / New x^*	10,85	11,03	11,03	11,03	11,03	11,03
Neuer Wert s^* / New s^*	3,53	3,19	3,08	3,05	3,04	3,04

ANMERKUNG Bei den Daten handelt es sich um Anzahlen von Einheiten (en: Units, U) in eintausend (k) je Liter (l) Probe; eine Einheit ist definiert als die Konzentration eines internationalen Referenzmaterials.

NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

5.7 Vergleich des zugewiesenen Werts

Werden die in 5.2 bis 5.4 beschriebenen Verfahren zur Bestimmung des zugewiesenen Werts X verwendet, so ist nach jeder Runde eines Schemas für Eignungsprüfungen der robuste Mittelwert x^* , hergeleitet aus den Ergebnissen der Runde, mit dem zugewiesenen Wert zu vergleichen. Wenn die in 5.5 und 5.6 beschriebenen Verfahren zur Bestimmung des zugewiesenen Werts verwendet werden, muss der Wert, soweit möglich, mit einem Referenzwert verglichen werden, der durch ein sachkundiges Laboratorium ermittelt wurde. Die Standardunsicherheit der Differenz x^*-X ist wie folgt abzuschätzen:

$$\sqrt{\frac{(1,25s^*)^2}{p} + u_X^2}$$

Dabei ist

- s^* die robuste Standardabweichung;
- p die Anzahl der Laboratorien.

Beträgt die Differenz mehr als das Doppelte ihrer Unsicherheit, ist der Grund hierfür festzustellen. Mögliche Gründe sind

- eine systematische Abweichungskomponente im Messverfahren,
- eine gemeinsame systematische Abweichungskomponente in den Ergebnissen der Laboratorien,
- die Nichtbeachtung der Einschränkungen des Verfahrens bei Anwendung des Verfahrens zur Probenherstellung nach 5.2,
- eine systematische Abweichungskomponente in den Ergebnissen der „Expertenlaboratorien“, wenn der Ansatz des „Konsenswerts von Expertenlaboratorien“ zu Grunde gelegt wird, und
- Verfahren mit systematischer Abweichungskomponente bei den Teilnehmern oder mehrere Laboratorien mit systematischer Abweichungskomponente, wenn der robuste Konsensmittelwert als zugewiesener Wert verwendet wird.

5.7 Comparison of the assigned value

When the methods described in 5.2 to 5.4 are used to establish the assigned value X , after each round of a proficiency testing scheme, the robust average x^* derived from the results of the round shall be compared with the assigned value. When the methods described in 5.5 and 5.6 are used to establish the assigned value, the value shall, where possible, be compared with a reference value obtained by a competent laboratory. The standard uncertainty of the difference x^*-X shall be estimated as:

$$\sqrt{\frac{(1,25s^*)^2}{p} + u_X^2}$$

where

- s^* is the robust standard deviation;
- p is the number of laboratories.

If the difference is more than twice its uncertainty, the reason shall be sought. Possible reasons are

- bias in the measurement method,
- a common bias in the results of the laboratories,
- failure to appreciate the limitations of the method when using the formulation method described in 5.2 approach,
- bias in the results of the “expert laboratories” when using the “consensus value from expert laboratories” approach, and
- biased participant method(s) or several biased laboratories when the robust consensus mean is used as the assigned value.