

## Glossar

### Streuemaße

Als Streumaße werden hauptsächlich die Varianz  $\hat{\sigma}^2$  und die Standardabweichung  $\hat{\sigma}$  verwendet. Für die Standardabweichung hat sich bisher keine einheitliche Abkürzung durchsetzen können. Außer  $\hat{\sigma}$  sind z.B. s und sdv gebräuchlich. Die „Überdachung“ eines Maßes zeigt in allen Fällen eindeutig an, dass es sich um eine aus einer Stichprobe geschätzte Größe handelt. Die Standardabweichung der Grundgesamtheit  $\sigma$  kann so leicht abgegrenzt und mit der Standardabweichung der Stichprobe  $\hat{\sigma}$  verglichen werden.

Die prozentuale relative Standardabweichung RSD (oder besser RSD%) ist sehr gut geeignet, Streuungen verschiedener Verfahren und Messungen zu vergleichen. Sie wird oft auch Variationskoeffizient genannt, dadurch erklärt sich der englische Ausdruck für diese Größe: CV, coefficient of variation.

$$\hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

$$RSD\% = \frac{\hat{\sigma}}{\bar{x}} \cdot 100\%$$

Daneben wird der Standardfehler  $\hat{\sigma}(\bar{x})$  verwendet. Er sagt nichts über die Streuung der Einzelwerte aus, sondern über die Streuung des Lagemaßes. Weil durch  $\sqrt{n}$  geteilt wird, ist der Standardfehler kleiner als die Standardabweichung. Standardabweichung und Standardfehler sind leicht ineinander umrechenbar. Aus Gründen der Vergleichbarkeit muss immer eindeutig dargestellt werden, welches der beiden Streumaße verwendet wird. Es ist sinnvoll, den Standardfehler zu verwenden, wenn durch Mehrfachmessung und Mittelwertbildung die zufällige Streuung einer Analysenmethode herabgesetzt wird (vgl. Gl. 5, Abschnitt 3.1.).

$$\hat{\sigma}(\bar{x}) = \frac{\hat{\sigma}}{\sqrt{n}}$$

### Präzisionsmaße [21, 22]

Die **Präzision** einer Analysenmethode ist die Übereinstimmung zwischen Analyseergebnissen an gleichen Proben. Hohe Präzision ist gegeben, wenn die Analyseergebnisse wenig streuen.

Die Streuung der Daten kann als Standardabweichung  $\hat{\sigma}$  oder prozentuale relative Standardabweichung RSD% angegeben werden. Zur Angabe der Präzision gehören immer die zugrundeliegenden Bedingungen, mit denen die Analyseergebnisse erhalten wurden (s.u.).

Ein Vergleich der verschiedenen Präzisionsebenen gibt Hinweise, welche Einflussfaktoren den größten Einfluss auf die Analysenvariabilität haben.

Die **Systempräzision** (injection repeatability, system precision) wird durch wiederholte Messung derselben Probe(nlösung) bestimmt. Sie wird im wesentlichen vom Messinstrument beeinflusst.

Unter Wiederholbedingungen wird die **Wiederhol-Standardabweichung** (repeatability,  $\hat{\sigma}_r$ ,  $RSD_r(\%)$ ) ermittelt. Wiederholbedingungen liegen vor, wenn das gesamte Analysenverfahren, inklusive Probenvorbereitung, in einem Labor vom gleichen Operator mit demselben Messinstrument auf identisches, homogenes Testmaterial angewandt wird und die Messungen innerhalb eines kurzen Zeitraums nacheinander durchgeführt werden (innerhalb einer Serie, mit denselben Referenzstandards).

Die **Mehrtages-Standardabweichung** (Intralabor-Standardabweichung, Labor-Vergleichspräzision, entspricht der engl. day-to-day-precision o. intermediate precision) gibt die Streuung an, wenn die gleiche, vollständige Methode an verschiedenen Tagen, möglicherweise durch unterschiedliche Analytiker und mit unterschiedlicher Ausrüstung, aber innerhalb des gleichen Labors, durchgeführt wird. Dieses Präzisionsmaß berücksichtigt (bei Kalibrierungen) die Streuung des Standards sowie zusätzlich zur zufälligen Streuung Einflüsse durch zeitliche Veränderungen (z.B. Temperaturschwankungen).

Die **Ringversuchs-Standardabweichung** (Interlabor-Standardabweichung, Vergleichspräzision, reproducibility) wird mit identischer Methode und identischem Testmaterial in unterschiedlichen Labors ermittelt. Bei diesem Präzisionsmaß werden zusätzlich zur zufälligen Streuung der Analysenergebnisse besonders auch systematische Veränderungen durch äußere Einflußgrößen (z.B. Temperatur, Raumfeuchte, Reagentienqualität, Ausbildungsstand des Laborpersonals ...) berücksichtigt.

Über Monate und Jahre nähert sich die Mehrtages-Standardabweichung immer stärker der Vergleichs-Standardabweichung an, da auch innerhalb eines Labors über längere Zeiträume Personal und Instrumentierung wechseln. Wir verwenden daher im Verlauf dieses Artikels den Begriff **Vergleichsstandardabweichung** ( $\hat{\sigma}_R$ ,  $RSD_R(\%)$ ) als übergeordneten Begriff.

Es ist nicht einfach, verlässliche Werte für die Größe der Vergleichsstandardabweichung zu erhalten. Meist wird sich zunächst an der experimentell leicht zugänglichen Mehrtages-Standardabweichung als Abschätzung orientiert. Realistische Werte für die Vergleichsstandardabweichung sind jedoch in der Regel höher als die Mehrtages-Standardabweichung [6]. Es ist aber zur Zeit noch nicht möglich, die zu erwartende Zunahme von der Mehrtages-Standardabweichung zur Vergleichsstandardabweichung quantitativ anzugeben.

## **„Analytische Unsicherheit und Rationale Spezifikationsfindung – Schwerpunkt: Gehaltsbestimmungen“**

Hermann Wätzig<sup>1)</sup> und Joachim Ermer<sup>2)</sup>

Fachgruppe Arzneimittelkontrolle/Pharmazeutische Analytik der DPhG

<sup>1)</sup>Institut für Pharmazeutische Chemie, TU Braunschweig, D-38106 Braunschweig

<sup>2)</sup>Global Analytical Development, Aventis Pharma AG, D-65926 Frankfurt am Main

**Wie können Spezifikationen sinnvoll nach dem Stand der Wissenschaft festgelegt werden, damit Qualitäts- und Sicherheitsrisiken vermieden und gleichzeitig eine kosteneffektive Qualitätskontrolle von Arzneimitteln erreicht werden kann? Mit dieser Frage beschäftigte sich der Workshop der Fachgruppe Arzneimittelkontrolle/Pharmazeutische Analytik der DPhG am 31.01.02 in Frankfurt/Main.**

Der Zusammenhang zwischen Spezifikationsgrenzen und analytischer Variabilität wird für die Bereiche Gehaltsbestimmung der Reinsubstanz, Gehaltsbestimmung aus der Formulierung und Reinheitskontrolle diskutiert. Bei der Festlegung von Spezifikationen müssen – natürlich unter Voraussetzung der Arzneimittelsicherheit - jeweils die produktionsbedingte und die analytisch bedingte Variabilität berücksichtigt werden.

Es wurde vorgeschlagen, bei Gehaltsbestimmungen in Wirkstoffmonografien in Zukunft eine Mindestanzahl von 3 Bestimmungen durchzuführen und eine maximal zulässige Standardabweichung vorzugeben. Zielstandardabweichungen (TSD) für Klassen von Methoden wurden in Ringversuchen festgelegt. Neue Spezifikationsgrenzen werden aus der maximal zulässigen Summe der Verunreinigungen und der zu erwartenden analytischen Variabilität berechnet.

Bei Formulierungen werden überwiegend die Werte 95 - 105% vom deklarierten Gehalt als Spezifikationsgrenzen festgelegt. In den meisten Fällen ist dies vernünftig, aber auch in einigen Fällen unbefriedigend. Höhere analytische Variabilität durch Matrixeffekte oder niedrige Wirkstoffkonzentration kann zum Teil nicht mehr mit vertretbarem Aufwand kompensiert werden, dies erfordert individuell angepasste Spezifikationen. Die Horwitz-Gleichung ist ein sehr vielversprechendes Konzept zur Abschätzung dieser Variabilität, besonders bei geringeren Konzentrationen. Aus dieser Gleichung kann der Erwartungsbereich der zu einem Wirkstoffgehalt gehörenden Präzision abgeleitet werden. Aus diesem Erwartungsbereich kann eine akzeptable Obergrenze für die experimentell ermittelte Standardabweichung abgeleitet werden, die zur Festlegung von weniger engen, problemangepassten Spezifikationsgrenzen für niedrige Wirkstoffkonzentrationen verwendet wurde. Ein erster pragmatischer Vorschlag zur Festlegung von Spezifikationen für Verunreinigungen von Wirkstoffen wurde zusätzlich vorgestellt.

### **1. Einleitung**

Die Qualitätsprüfung von Arzneistoffen und Arzneimitteln muß rational konzipiert sein. Dies setzt zunächst eine gezielte Ermittlung potentieller Risiken voraus. Im Anschluss wird die erforderliche Qualität, z.B. die maximal zulässige Menge von Verunreinigungen, festgelegt. Qualitätserfordernisse sind von Substanz zu Substanz unterschiedlich und daher nicht verallgemeinerbar [1].

Auf dieser Grundlage müssen bei der Festlegung von Spezifikationen die produktionsbedingte und die analytisch bedingte Variabilität berücksichtigt werden. Die Variabilität durch die

Produktion ist bei neuen Arzneimitteln schwer zu erfassen. Die analytische Variabilität lässt sich verhältnismäßig leichter bestimmen, aber auch hier ist der notwendige Aufwand teilweise beträchtlich.

Der Zusammenhang zwischen Spezifikationsgrenzen und analytischer Variabilität für die Bereiche Gehaltsbestimmung von Wirkstoffen, Gehaltsbestimmung aus der Formulierung und quantitative Bestimmung von verwandten Substanzen müssen getrennt diskutiert werden, denn die Konzentrationsbereiche, die zu erwartenden Matrixeffekte und auch die eingesetzten analytischen Techniken sind unterschiedlich. Nach eingehender Diskussion möchte die Fachgruppe Arzneimittelkontrolle/Pharmazeutische Analytik der DPhG ein Positionspapier zu diesem Themenkomplex vorlegen.

## 2. Spezifikationen für Wirkstoffe

### 2.1. Hintergrund

In den allgemeinen Vorschriften zur Gehaltsbestimmung von Wirkstoffen (drug substances) im Europäischen Arzneibuch ist die notwendige Anzahl von Wiederholbestimmungen nicht angegeben. Es wäre theoretisch ausreichend, bei Gehaltsbestimmungen nur eine Einzelbestimmung durchzuführen. Dieser einzelne Messwert konnte aber zufällig ausserhalb des in der Monographie angegebenen Bereichs liegen, obwohl die untersuchte Substanz die Spezifikationen bei genauerer Untersuchung erfüllte; ebenso konnte es passieren, dass ein Ergebnis durch die Messwertstreuung zufällig noch im zulässigen Bereich lag, obwohl eine eingehende Untersuchung eine Abweichung gezeigt hätte. Außerdem war unklar, wie man sich verhalten sollte, wenn der eine ermittelte Wert knapp außerhalb der gesetzten Spezifikationen lag. Ein Meßfehler war hier wahrscheinlicher als eine qualitätsbedingte Abweichung, also wurde die Untersuchung wiederholt; aber wie oft? Wenn sie nur einmal wiederholt wurde und das Ergebnis dann spezifikationskonform war: welchem Ergebnis sollte man nun trauen, dem ersten oder dem zweiten? Sollen also weitere Messungen erfolgen? Wenn ja, wieviele?

Selbst bei minderwertiger Qualität (z.B. 92% Gehalt) kann sich auch z.B. bei einer von 97 bis 102% gesetzten Spezifikationen zufällig ein Meßwert von 97.1% ergeben. Wenn lang genug gemessen wird, wird dieses Zufallsergebnis irgendwann produziert, vielleicht schon beim 1. oder 2. Mal, vielleicht auch erst beim 26. Wert oder später. Es sind Fälle von Missbrauch aufgetreten, bei denen sehr oft gemessen wurde und alle Messwerte ignoriert wurden, die auf nichtspezifikationskonforme Qualität hindeuteten. Bei diesem „in die Spezifikation hineintesten“ handelt es sich um eine unzulässige Manipulation (vgl. dazu [2]).

Eine neue Regelung soll hier Klarheit schaffen. Miller schlägt vor, dass in einer zukünftigen Ausgabe des Europäischen Arzneibuchs mindestens 3 unabhängige Bestimmungen mit individueller Probenaufarbeitung verlangt werden sollen [3, 4]. Von diesen werden Mittelwert und Standardabweichung bestimmt. Dann wird überprüft, ob die ermittelte Standardabweichung im Rahmen des zulässigen liegt (Tabelle 1). Je größer die Zahl der Bestimmungen ist, desto höhere Werte für die Standardabweichung sind zulässig, denn höhere Datenzahlen ermöglichen eine bessere Schätzung der Standardabweichung und erfordern somit einen geringeren Sicherheitsabstand (s.a. 3.2.).

Liegt die Standardabweichung unterhalb des tabellierten Wertes wird überprüft, ob der Mittelwert der Bestimmungen innerhalb der in der jeweiligen Monographie genannten Grenzen liegt. Auch hier wirkt sich eine hohe Datenzahl positiv aus, da hierdurch die zufällige Streuung des Mittelwertes verringert ist und es daher unwahrscheinlicher wird, dass

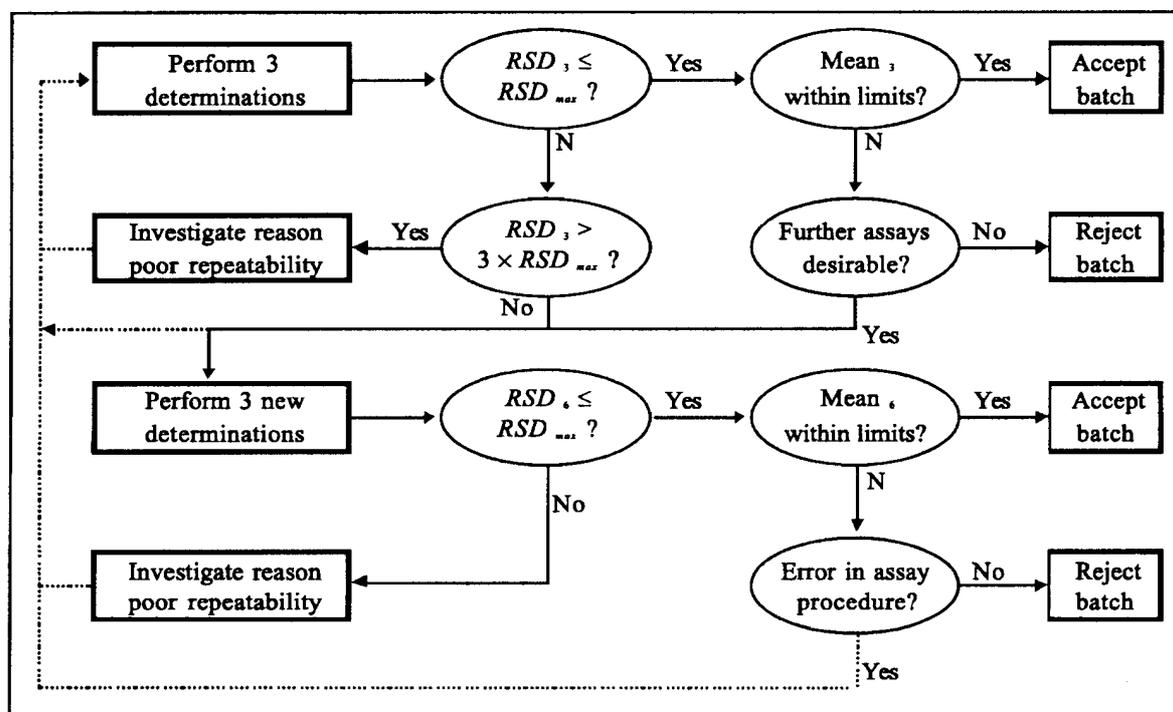
der Mittelwert durch zufällig sehr hohe oder niedrige Messwerte ausserhalb der Spezifikationen liegt.

Tabelle 1. Maximal zulässige relative Wiederhol-Standardabweichung in Abhängigkeit von der Anzahl der durchgeführten Bestimmungen; A ist hier der obere in der Monographie angegebene Grenzwert – 100%, wobei angenommen wird, dass diese obere Grenze die Reproduzierbarkeit der analytischen Methode widerspiegelt [3].

A (%)	(2)	3	4	5	6
1.0	0.11	0.29	0.42	0.52	0.60
1.5	0.17	0.44	0.63	0.78	0.90
2.0	0.22	0.59	0.84	1.04	1.20
2.5	0.28	0.73	1.05	1.29	1.50
3.0	0.33	0.88	1.26	1.55	1.80

Liegt die Streuung über den in Tabelle 1 angegebenen Werten, können zunächst weitere Bestimmungen bis zu einer maximalen Gesamtzahl von  $n=6$  durchgeführt werden. Wenn die Grenzen nun eingehalten werden, ist die Untersuchung abgeschlossen. Liegt die Streuung weiterhin zu hoch, muß die Ursache durch sachkundiges Personal festgestellt werden. Im Anschluss wird die komplette Untersuchung wiederholt, wobei eine andere Person die Analyse durchführt.

Liegt der Mittelwert ausserhalb der Spezifikation und die Messwertstreuung ist für eine Beurteilung geeignet, dann wird dieses Ergebnis zunächst durch eine andere Person bestätigt. Danach wird das nichtspezifikationskonforme Ergebnis abschließend dokumentiert (Diagramm).



Akzeptanz- und Ablehnungskriterien für Gehaltsbestimmungen von Wirkstoffen (Vorschlag aus [3], Fig. 1)

## 2.2. Berechnung von Spezifikationsgrenzen

Was kann man tun, wenn für einen neuen Wirkstoff eine Monographie erstellt oder er analog einer bestehenden beurteilt werden soll? Am besten man orientiert sich an den Spielregeln, nach denen Spezifikationsgrenzen für neue Monographien festgelegt werden [4]. Die folgenden Ausführungen basieren auf einer Anzahl von drei Wiederholbestimmungen.

Die untere und die obere Spezifikationsgrenze (engl. lower bzw. upper specification limit; LSL und USL) können einfach nach Gleichung 1a und 1b festgelegt werden [5]:

$$(1a) \quad LSL = 100\% - \%erwartete\ Verunreinigung - 3 \cdot TSD$$

$$(1b) \quad USL = 100\% + 3 \cdot TSD$$

Die Zielstandardabweichung (engl.: target standard deviation; TSD) entspricht der geometrisch gemittelten (gepoolten) Wiederholpräzisions-Standardabweichung. Werte für TSDs sind in umfangreichen Ringversuchen bestimmt worden [4]. Ein Auszug ist in Tabelle 2 wiedergegeben.

Tabelle 2. Zielstandardabweichungen (target standard deviations; TSDs), die in Ringversuchen ermittelt wurden [4]. TSDs sind geometrisch gemittelte Wiederhol-Standardabweichungen.

Methoden	Substanzbeispiele	TSD
<b>Titrationen</b>		
- im wässrigen Milieu mit Alkali, Farbindikator	Salicylsäure	0.2 %
- im wässrigen Milieu mit Alkali, potentiometrische Indikation	Salicylsäure	0.3 %
- im nicht-wässrigen Milieu mit Alkali, Farbindikator	Ephedrinhydrochlorid, racemisch	0.4 %
- im nicht-wässrigen Milieu mit Alkali, potentiometrische Indikation	Ephedrinhydrochlorid, racemisch	0.4 %
<b>UV-Spektrometrie</b>	Prednisolonacetat, u.a. <sup>1)</sup>	0.6 % - 1.3 %
<b>Flüssigchromatographie (HPLC)</b>	Cloxacillin-Natrium	0.6 % <sup>2)</sup>

1) Cinnarizin, Dienestrol, Albendazol und Methylprednisolon-hemisuccinate

2) ev. teilweise höher, vgl. UV-Spektrometrie

Beispiel (modifiziert nach [3]): Gehaltsbestimmung von Liothyronin-Natrium mittels HPLC

Gesamtmenge erlaubter substanzverwandter Verunreinigungen: 5%

Erlaubte Gesamtmenge Chloride (als NaCl berechnet): 2%

TSD(HPLC) = 0.6%;  $3 \cdot 0.6\% = 1.8\%$ , aufgerundet: 2%

Unteres Limit:  $100\% - 5\% - 2\% - 2\% = 91\%$

Oberes Limit:  $100\% + 2\% = 102\%$

Auf Grund der umfangreichen experimentellen Absicherung können TSD's als zuverlässige Abschätzungen der wahren Wiederholstandardabweichung  $\sigma_r$  betrachtet werden. Wenn die Zielstandardabweichungen für eine Klasse von Methoden ähnlich sind, würde dies eine schnelle Ermittlung von sinnvollen Spezifikationsgrenzen ermöglichen. Es ist jedoch noch zu früh, die Zahlenwerte für alle Klassen von Methoden endgültig festzulegen. Es ist zum Beispiel nicht einleuchtend, warum für die UV-Spektrometrie eine größere TSD angegeben werden soll als für die Kombination LC-UV, obwohl hier durch Injektion und Peakintegration zusätzliche Fehler hinzukommen. Ein Wert von 0.6% als relative Wiederholstandardabweichung erscheint nur unter günstigen Bedingungen erreichbar, da schon im Rahmen eines Systemtauglichkeits-Tests (system suitability tests, SST) zum Teil deutlich größere Werte ermittelt werden [6, 7]. In Tabelle 3 sind Systempräzisionen für die LC und für weitere Techniken aus SSTs zusammengestellt worden [6]. In einer ähnlichen Studie ist in einem anderen Fall eine durchschnittliche Systempräzisionen von 0.8% ermittelt worden, in 5 von 18 Versuchsserien aber auch nahe oder über 1% ([7], vgl. 3.2.1.). Unzureichende Integrationsalgorithmen können auch heute noch bei ungünstigen Verläufen der Basislinie zu erheblich größeren Streuungen führen [8, 9].

Wir möchten daher eine Datenerhebung zu dieser Thematik durchführen und bitten um Zusendung von Informationen über die verschiedenen Präzisionsebenen von Analysemethoden (System-, Wiederhol-, Mehrtages- und Ringversuchspräzision bzw. Vergleichspräzision).

Tabelle 3: System Performance, SST; entspr. Systempräzision, für verschiedene analytische Techniken (modifiziert aus [6]).

Technik	durchschn. RSD%	zweithöchste und höchste gefundene RSD%	n Messungen zur jew. Berechnung von RSD%	N Mess-Serien
HPLC, automatisiert	0.3	0.5, 0.6	5	22
GC, Direktinj.	0.7	0.9, 1.0	6	10
GC, headspace	1.1	1.8, 2.3	6	10
CE	0.7	1.1, 1.2	6	16
HPTLC <sup>1)</sup>	1.4 - 1.9	2.9, 2.9	8 - 16	20

1) Unterschiede bei Auswertung über Peakhöhen oder -flächen und durch Anzahl der ausgewerteten Bahnen pro TLC-Platte

Die Spezifikationsgrenzen können auch mit Hilfe der Mittelwerts-Ringversuchsstandardabweichung  $\hat{\sigma}_{R\bar{x}}$  (2) festgelegt werden, wenn diese für eine bestimmte Methode bekannt ist (3a, 3b) [4, 10]:

$$(2) \quad \hat{\sigma}_{R\bar{x}} = \sqrt{\frac{\sum (\bar{x}_i - \bar{x})^2}{N-1}}$$

$$(3a) \quad LSL = 100\% - \%erwartete\ Verunreinigung - 1.69 \cdot \hat{\sigma}_{R\bar{x}}$$

$$(3b) \quad USL = 100\% + 1.69 \cdot \hat{\sigma}_{R\bar{x}}$$

Hier ist  $\bar{x}_i$  der Mittelwert im jeweiligen Ringversuch,  $\bar{\bar{x}}$  ist der Mittelwert aller Mittelwerte und N ist die Anzahl von Labors, die am Ringversuch teilgenommen haben.

Die so erhaltenen Spezifikationsgrenzen sind besser der konkreten Fragestellung angepasst, ähneln aber den nach Gl. 1 ermittelten Werten. Der unterscheidende Faktor von ca. 2 gibt empirisch die Variabilitäts-erhöhung von der Wiederhol- zur Vergleichspräzision an (d.h. den Beitrag des Standards ( $\sqrt{2}$ ) sowie zusätzliche Einflüsse).

Wenn für eine Klasse von Methoden noch keine TSD existiert, oder wenn bei einer Methode ungewöhnlich hohe Streuungen auftreten, die eine Neuklassifizierung nahelegen, dann können entsprechende TSDs, wie in [4] detailliert beschrieben, in Ringversuchen ermittelt werden. Wegen des hohen Aufwands für solche Ringversuche ist es empfehlenswert, Kooperationspartner zu finden, die TSDs für Methoden aus der gleichen Klasse untersuchen.

### 3. Spezifikationen für Formulierungen

#### 3.1. Rationale der 95 – 105%-Grenzen

Für Gehaltsbestimmungen aus Formulierungen bei der Beantragung von europäischen Zulassungen gilt die Grundregel, dass der ermittelte Gehalt ohne angemessene Begründung den Bereich zwischen 95 und 105% des deklarierten Gehalts nicht überschreiten darf [11]. Werden diese Grenzen eingehalten, können sie bei der Zulassung eines Arzneimittels angegeben werden und müssen nicht extra begründet werden.

Es besteht Konsens darüber, dass dies in der Regel eine vernünftige Vorgehensweise ist und dass diese Grenzen in den meisten Fällen eingehalten werden können. Ebenso besteht Übereinstimmung, dass es begründete Einzelfälle gibt, in denen Arzneimittel mit weiter gesteckten Grenzen zugelassen werden müssen. Es besteht aber Diskussionsbedarf, ob sich allgemeine Regeln finden lassen, in welchen Fällen größere Bereiche sinnvoll und akzeptabel sind.

Der Spezifikationsbereich 95-105% kann mit einer (wahren) Gesamtstreuung von etwa 1.5 - 2% für einen einzelnen Wert eingehalten werden. Zur Abschätzung wird die t-Verteilung verwendet, da Aussagen aus der Normalverteilung nur für sehr große Datenzahlen möglich sind. Für n=6 Messungen findet man einen t-Faktor von 2.57 für eine Irrtumswahrscheinlichkeit von 0.05 und n-1 = 5 Freiheitsgrade;  $2.57 \cdot 2\% = 5.14\%$ .

Die Gesamtstreuung setzt sich aus der analytisch bedingten und den in der Produktion verursachten Abweichungen zusammen. Es gilt die Additivität der Varianzen, also der quadrierten Standardabweichungen (4):

$$(4) \quad \hat{\sigma}_{ges}^2 = \hat{\sigma}_{ana}^2 + \hat{\sigma}_{prod}^2$$

Die Varianz eines Produktionsprozesses ist jedoch sehr schwer zu bestimmen, denn zur Bestimmung muss immer ein analytisches Verfahren eingesetzt werden, man kann also die produktionsverursachte und analytisch bedingte Streuung nur zusammen bestimmen. Durch ein geeignetes verschachteltes Versuchsdesign (engl. nested design) kann man die zugehörigen Varianzanteile zwar berechnen: man vermisst dazu verschiedene Chargen mehrfach und vergleicht die Intra- und Interchargenvarianz. Für ein solches Design ist jedoch eine sehr hohe Datenzahl notwendig, es ist also aufwändig.

Es gibt auch keinen ausreichend großen Datenbestand, aus dem auf zukünftige, vergleichbare Fälle geschlossen werden könnte. Die beste Information über die produktionsbedingte Streuung liefern sicher Prozessregelkarten (Kontrollkarten). Diese Information kann aber erst bei laufender Produktion gewonnen werden und steht vor der Zulassung noch nicht zur Verfügung.

In der Praxis behilft man sich, indem man der produktionsbedingten maximalen Abweichung einen festen Wert zuweist, wobei dieser Wert den schlimmsten möglichen Fall repräsentieren muss. Wird dieser Abweichung z.B. der Wert  $\pm 2.5\%$ , also die Hälfte der allgemein akzeptierten Gesamtabweichung zugewiesen, dann bleibt die andere Hälfte von  $2.5\%$  für die analytisch bedingte Streuung, entsprechend einer analytisch bedingten (wahren) Standardabweichung von  $0.75 - 1\%$  (s.a. 3.2.3.).

Dieser Wert liegt oberhalb der in Tabelle 2 angegebenen Zielstandardabweichungen; es muss jedoch berücksichtigt werden, dass diese Zielstandardabweichungen Wiederhol-Standardabweichungen repräsentieren, bei denen der Einfluss des Standards nicht berücksichtigt ist, sowie für Gehaltsbestimmungen von Reinsubstanzen gelten, also fast ohne möglicherweise störende Matrixsubstanzen. Da bei Gehaltsbestimmungen aus Formulierungen zusätzliche Matrixeffekte nicht zu vermeiden sind, sind diese Zielstandardabweichungen als Abschätzung für die erreichbare analytische Präzision hier nicht geeignet. Eine  $RSD_R(\%)$  unter  $1\%$  für die Analytik wird oft nicht erreichbar sein.

Dies ist jedoch auch nicht notwendig, da Mehrfachbestimmungen ausdrücklich zugelassen und sogar erwünscht sind. Die Streuung für den Mittelwert aus einer Mehrfachbestimmung ergibt sich aus Gleichung 5:

$$(5) \quad \hat{\sigma}(\bar{x}) = \frac{\hat{\sigma}(x)}{\sqrt{n}}$$

Allein durch Mehrfachbestimmung ist es meist möglich, die Spezifikationsgrenzen  $95 - 105\%$  einzuhalten, sofern der Mittelwert als „reportable result“ definiert wurde, welches mit den Spezifikationsgrenzen zu  gleiches ist. Eine Vergleichs-Standardabweichung von  $2\%$  ist meist erreichbar, für  $n=4$  entspricht dies einem Wert von  $\hat{\sigma}(\bar{x}) = 1\%$ . Mit diesem Wert kann also eine Spezifikation von  $95 - 105\%$  voraussichtlich eingehalten werden. Daher sind auch bei höherer Messwertstreuung Dreifachbestimmungen meist ausreichend, um ein  $\hat{\sigma}(\bar{x})$  von unter  $1\%$  zu erreichen.

Auch wenn die Streuung von einzelnen Messwerten sehr hoch ist, kann durch eine hinreichend hohe Datenzahl die Streuung für den Mittelwert theoretisch beliebig klein erhalten werden. Allerdings sind für  $\hat{\sigma} = 10\%$  bereits  $n = 100$  Messungen notwendig. Solch

hohe Datenzahlen sind normalerweise nicht zu rechtfertigen, es sei denn, dass die verwendete Analytik extrem schnell und kostengünstig ist.

### 3.2. Wann erfordert die analytische Unsicherheit weniger enge Spezifikationsgrenzen?

Diese Überlegungen zeigen, dass man sich um weniger enge, dem Spezialfall angepasste Spezifikationen bemühen muss, wenn die Wiederholstandardabweichung für Einzelbestimmungen 2% übersteigt. Selbst wenn die experimentelle Wiederholbarkeit gerade so eben unter 2% liegt, ist äußerste Vorsicht geboten (s. Tab. 3).

#### 3.2.1. Unsicherheit über die Unsicherheit

Unter diesem Titel stand ein Beitrag von Herrn Dr. Weber (EMPA) [12]. Zunächst ein einführendes Beispiel:

Mit  $n=6$  wird die Wiederholbarkeit entsprechend einer Standardabweichung von 2% bestimmt. Ist diese analytische Methode ausreichend präzise? Der Wert für die Quantile der F-Verteilung für jeweils 5 Freiheitsgrade  $df$  und einer Irrtumswahrscheinlichkeit von 0.1 beträgt  $F_{0.1, 5, 5} = 3.453$ . Erst wenn der Testwert  $T_F$  (Gleichung 6)

$$(6) \quad T_F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

diesen Wert übersteigt, wird ein Unterschied als signifikant erkannt. Aus der Wurzel des jeweiligen F-Werts (hier: 1.858) läßt sich direkt ablesen, wie stark sich die Standardabweichungen unterscheiden müssen, bevor ein Unterschied als signifikant erkannt wird (bezogen auf ein Vertrauensniveau von  $P = 90\%$ ). In diesem Beispiel muss die Standardabweichung um mehr als das 1.8fache überschritten sein! Das heisst umgekehrt, dass die Standardabweichung von 2% eventuell nur zufällig so niedrig ist, da sie sich von 3.5% nicht signifikant unterscheidet! Wenn nun 3.5% die eigentliche Streuung ist, können später in der Routineanalytik die Spezifikationsgrenzen 95 – 105% nur mit hohen Datenzahlen eingehalten werden.

Präzisionsmaße, z.B. Standardabweichungen streuen immer sehr stark, viel stärker als Lagemaße (z.B. Mittelwerte). Deshalb sind die Schwellenwerte beim F-Test auch viel weiter gefasst als beim t-Test.

Es muss ausgeschlossen werden, dass die in der Routineanalytik zu erwartende Vergleichs-Standardabweichung 2% übersteigt. Dazu muss die in der Entwicklungsphase erreichte Vergleichspräzision einen ausreichenden Sicherheitsabstand zu 2% aufweisen; je näher die Standardabweichung an 2% liegt, desto mehr Messungen sind notwendig um zu belegen, dass der Wert in der Routine sicher (signifikant) unterhalb 2% liegt (Tabelle 4). Hohe Datenzahlen sind erforderlich, wenn die experimentell bestimmte Standardabweichung nahe am akzeptablen Maximalwert liegt. Der dadurch bedingte Aufwand ist aber vertretbar, da diese Überprüfung nur einmal im Rahmen der Methodenentwicklung durchgeführt werden muss.

In dieser Tabelle wird davon ausgegangen, dass die wahre Standardabweichung angenähert mit einer sehr hohen Datenzahl ( $df=1000$ ) bestimmt wird und bei 2% liegt. Für ein gegebenes

n kann dann das Verhältnis der Varianzen berechnet werden (Gl. 6), ab welchem der Unterschied signifikant wird ( $\alpha = 0.1$ ). Hat die experimentell ermittelte Varianz diesen Wert oder liegt sie darunter, dann ist der Unterschied zu 2% wahrscheinlich nicht nur zufällig bedingt.

Beispiel:  $n = 6$ ,  $df = 5$ , die ermittelte Standardabweichung beträgt 1.5%.  $F_{0.1, 5, 1000} = 1.853$ . Das Verhältnis der Standardabweichungen beträgt  $2/1.5 = 1.333$ , dieser Wert darf nicht größer sein als die Wurzel des F-Werts;  $\sqrt{1.853}$  ist jedoch 1.361. Es kann also nicht ausgeschlossen werden, dass der Unterschied nur zufällig ist. Liegt die Standardabweichung bei 1.47% oder darunter, dann besteht wahrscheinlich ein signifikanter Unterschied ( $\alpha = 0.1$ ).

Tabelle 4: erforderlicher Freiheitsgrad  $df$  (hier:  $n-1$ ), damit die während der Methodenentwicklung bestimmte Vergleichsstandardabweichung  $\hat{\sigma}$  ausreichenden Sicherheitsabstand zu 2% hat ( $\hat{\sigma}(\%) = 2 \cdot \frac{1}{\sqrt{F_{0.1, df, 1000}}}$ ; am besten wäre es, die aus einer

kleineren Stichprobe ermittelten Standardabweichung mit der wahren Standardabweichung der Grundgesamtheit zu vergleichen. Da diese jedoch bei Messwerten in der Regel unendlich groß ist, wird  $df = 1000$  gewählt und mit einer sehr gut charakterisierten Stichprobe verglichen)

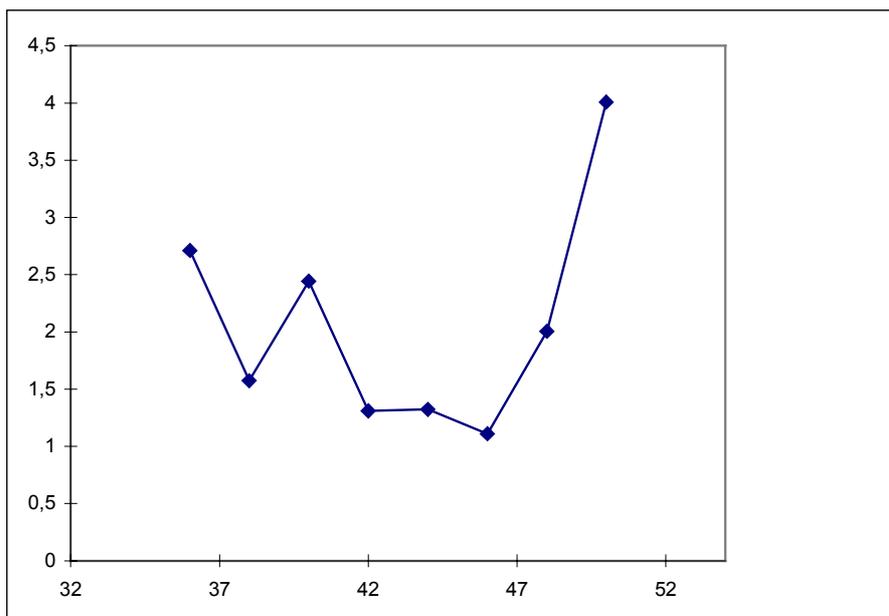
$df = n - 1$	$\hat{\sigma}$ (%)
4	1,43
5	1,47
6	1,5
8	1,54
10	1,58
15	1,64
20	1,67
30	1,72
40	1,75
50	1,77
100	1,83
500	1,9

Noch schlimmer als die vorschnelle Beurteilung einer Messwertstreuung ist der (leider naheliegende) Fehler der Pseudooptimierung. Dazu zunächst ein Beispiel:

Die Präzision einer Kapillarelektrophorese-Methode sollte optimiert werden. Da kurz nach der Einführung der Kapillarelektrophorese noch nicht bekannt war, von welchen Parametern die Präzision abhängt, wurde die Standardabweichung in Abhängigkeit von der Pufferkonzentration untersucht (Tabelle 5, Abb.):

Tabelle 5: Pseudooptimierung: Standardabweichung von Peakflächen bei unterschiedlichen Pufferkonzentrationen ( $n=6$ )

c (in mM)	36	38	40	42	44	46	48	50
$\hat{\sigma}$	2,711	1,575	2,442	1,31	1,324	1,11	2,006	4,006



Auf den ersten Blick scheint es, es gäbe es ein Optimum für eine Pufferkonzentration zwischen 42 und 46 mM – in Wahrheit unterscheiden sich die ermittelten Standardabweichungen aber nicht signifikant! Wäre einfach siebenmal bei gleicher Pufferkonzentration gemessen worden, hätte sich eine ähnliche Verteilung der Streuungen ergeben (mittlerweile ist bekannt, dass die Messwertstreuung in der CE kaum von der Pufferkonzentration abhängt).

Aus sieben Messreihen die beste herauszusuchen ist ein leicht einsehbarer Fehler. Diese Fehlermöglichkeit tritt aber häufig versteckt auf. Bei der Methodenentwicklung werden Parameter variiert; wenn sich eine dadurch eine geringere Messwertstreuung ergibt, dann hofft jeder Analytiker, die Methode verbessert zu haben. Um hier Pseudooptimierungen zu vermeiden, müssen Daten aus Vorversuchen mit kleineren Datenzahlen unbedingt mit höheren Datenzahlen abgesichert werden. Eine Messwertstandardabweichung von 1.1% wäre in der Routineanalytik später nur zufällig jedes siebte Mal erreicht worden.

### 3.2.2. Weniger präzise analytische Methoden

Es ist gar nicht so selten, dass eine Vergleichsstandardabweichung von 2% nicht erreicht werden kann. In diesen Fällen ist es viel besser, dem analytischen Problem ins Auge zu sehen, wenn möglich, die Standardabweichung weiter zu verringern, und wenn nicht möglich, eine Zulassung mit Spezifikationsgrenzen zu beantragen, die mit der erreichbaren Streuung eingehalten werden können. Hier muss die Abteilung Analytische Entwicklung auch innerhalb der eigenen Firma gegenüber der Zulassungsabteilung Überzeugungsarbeit leisten, damit die beantragten Spezifikationen später auch von der Qualitätskontrolle eingehalten werden können. Sicher ist es vordergründig am einfachsten, die 95-105%-Grenzen in den Zulassungsantrag zu schreiben. Aber dies verschiebt nicht nur das Problem, es verschlimmert es, weil es zu sehr hohen Folgekosten durch nichtspezifikationskonforme Ergebnisse (OOS results) oder sehr hohem Messaufwand in der Routinekontrolle führt.

Die Zulassungsbehörden kennen diese Problematik. Wenn von den Standardspezifikationen 95-105% abgewichen werden muss, wird die Zulassung bei fundierter Begründung ebenfalls erteilt. Die Argumentation dazu muss nach dem Stand der wissenschaftlichen Erkenntnis plausibel und in sich schlüssig sein. Eine Abweichung soll nicht leichtfertig beantragt werden. Würden beispielsweise bei einer relativen Vergleichsstandardabweichung von 1.1% weniger enge Spezifikationsgrenzen beantragt, um später den Aufwand von Mehrfachmessungen in der Routine zu vermeiden, dürfte es fraglich sein, ob dies von den Behörden akzeptiert würde. In der Diskussion während des Workshops ergab sich, dass die Zulassungsbehörden erwarten, dass zunächst Mehrfachmessungen durchgeführt werden, wenn die Wiederholbarkeit über 1% liegt; n=3 Mehrfachmessungen bei Bedarf wurde als vernünftiger Kompromiss aus Zugewinn an Information und Aufwand genannt.

Ein weiterer wichtiger Punkt ist sicher eine eingehende Analyse der wirklichen Standardabweichung mit umfangreichem Datenmaterial (vgl. Tabelle 4), um zu überprüfen, ob die Standardspezifikationen nicht doch eingehalten werden können, oder um zunächst Zulassungsabteilungen und anschließend Behörden überzeugend gegenübertreten zu können. Oft ist die Ursache für eine höhere Streuung bekannt; dann soll dargelegt werden, warum diese nicht abstellbar ist, eventuell unter Bezugnahme auf bekannte ähnliche Produktionsprozesse. Zur Dokumentation sind Kontrollkarten ein wichtiges Instrument.

Antragsteller wünschen sich Regeln, in welchen Fällen bei fundierter Begründung mit einer Zulassung bei individuellen Gehaltsspezifikationen gerechnet werden kann. Allgemeingültige Regeln wären sehr hilfreich, sowohl für die Diskussion zwischen Behörden und pharmazeutischem Unternehmer als auch als Orientierungshilfe für innerbetriebliche Entscheidungsprozesse. Die Fachgruppe Arzneimittelkontrolle/ Pharmazeutische Analytik der DPhG plant daher ein Positionspapier, das als Diskussionsgrundlage für eine solche Leitlinie dienen soll.

### 3.2.3. Gründe, die erweiterte Spezifikationsintervalle erfordern

Es gibt im wesentlichen vier Gründe, warum die Grenzen 95-105% nicht eingehalten werden können:

- ungewöhnlich hohe herstellungsbedingte Chargenvariabilität. Die Hälfte der Gesamtstreuung ist gedanklich für die analytische Unsicherheit vorgesehen (s. 3.1.). Beträgt die produktionsbedingte Variabilität über 2.5%, dann ist nicht mehr genug Spielraum für die analytische Streuung vorhanden. Beispiele hierzu finden sich im Bereich der biotechnisch hergestellten Arzneistoffe und bei einigen Arzneiformen (z.B. bei Dosieraerosolen).
- relativ kleine Arzneistoffkonzentrationen
- Matrixeffekte durch Begleitsubstanzen aus der Formulierung, oder bei relativ komplex aufgebauten Analyten (z.B. Protein-Mikroheterogenitäten, Aufarbeitungen aus biotechnischen Ansätzen, Pflanzenextrakte u.a.)
- Variabilität über längere Zeiträume, besonders in der Stabilitätsanalytik

Aufbauend auf die Rationale aus 3.1. erscheint es sinnvoll, die Komponenten der Variabilität getrennt zu betrachten. Das produktionsbedingte Limit PL und die analytisch bedingte Streuung werden zum Spezifikationslimit SL summiert, die analytisch bedingte Streuung kann aus der Vergleichsstandardabweichung (oder der im jeweiligen Labor etablierten Mehrtagesstandardabweichung)  $\hat{\sigma}_R$ , dem zugehörigen t-Faktor (df ist hier die Zahl der Freiheitsgrade bei der Ermittlung der Standardabweichung, d.h. bei m Serien mit jeweils n Wiederholbestimmungen:  $df=m*(n-1)$ ) und der in der Routineanalytik geplanten Anzahl von Mehrfachmessungen  $n_{\text{assay}}$  berechnet werden (Gleichung 7, nach [13]):

$$(7) \quad SL = 100\% \pm PL \pm \frac{(\hat{\sigma}_R * t_{df,0.95two\ sided})_{validation}}{\sqrt{n_{assay}}}$$

Für ein Szenario mit 2 Serien a 6 Werten ( $df = 10$ ,  $\alpha = 0.975$  einseitig bzw. 0.95 zweiseitig) errechnet sich:

$$SL = 100\% \pm 2.5\% \pm \frac{2.23 * \hat{\sigma}_R}{\sqrt{n_{assay}}}$$

Es liegt nahe, diesen Ansatz gedanklich zu erweitern. Das Spezifikationslimit bleibt die Summe aus den Komponenten Produktion und Analytik. Für die Analytik wird hier üblicherweise eine Streuung von 2.5% (entsprechend  $\hat{\sigma}_R / \sqrt{n_{\text{assay}}} \approx 1\%$ ) angesetzt, ein Wert, der in den meisten Fällen gewährleistet werden kann. Mit dieser Erweiterung können erhöhte produktionsbedingte Streuungen einfach in das Gesamtkonzept einbezogen werden.

Beispiel: Es ist dokumentiert, dass die produktionsbedingte Variabilität 4% beträgt. Die maximal zulässige Abweichung wird auf  $4 + 2.5 = 6.5\%$  festgelegt, die Spezifikation des Gehalts muß danach zwischen 93.5 und 106.5% des deklarierten Wertes betragen.

Der Standardansatz geht von einem Anteil von 2.5% für die analytisch verursachte Streuung aus, entsprechend 1% RSD% für Mittelwerte (s. 3.1.). Höhere Streuungen müssen genau dokumentiert und begründet werden. Überprüfungen durch Ringversuche wären wünschenswert, aber nur in Ausnahmefällen ausführbar, denn dies würde ein gemeinsames Interesse von verschiedenen Labors erfordern. In den meisten Fällen geht es aber um individuelle Fragestellungen.

Die Gesamtstreuung ist auch nach dem EURACHEM-Konzept abschätzbar [12, 14]. Dieses basiert auf der Fehlerfortpflanzungsrechnung und ermöglicht zuverlässige Abschätzungen der Gesamtstreuung, wenn alle relevanten Elemente der Streuung bekannt sind. Viele Komponenten lassen sich tatsächlich gut abschätzen, wie etwa Pipettier- oder Verdünnungsfehler. Leider sind diese bekannten Bestandteile der Gesamtstreuung die meist unkritischen; größere, aber unbekannte Komponenten dominieren im Gesamtergebnis.

Ist es trotzdem möglich, allgemein akzeptable Regeln zu finden, in welchen Fällen eine höhere Streuung als entsprechend  $RSD\%(\bar{x}) \approx 1\%$  akzeptiert werden muß? Die Auswirkungen von Matrixeffekten sind bisher schwer klassifizierbar. Das gleiche gilt für Langzeiteffekte, die durch Referenzstandards nicht ausgeglichen werden können. Minimale Veränderungen der Selektivität, etwa durch Veränderungen der stationären Phase, können nach sehr langen Zeiträumen zu zusätzlicher Mitelution führen und damit die Richtigkeit beeinflussen. Hier muss noch weitere Erfahrung gesammelt und systematisch aufgearbeitet werden, bevor allgemeine Regeln abgeleitet werden können. Ringversuche zu diesem Thema wären sicher sehr informativ, aber auch aufwändig.

#### 3.2.4. Abschätzung der Messwertstreuung mit der Horwitz-Funktion

Wenn aber ungewöhnlich geringe Wirkstoffkonzentrationen die Ursache für eine erhöhte Messwertstreuung sind, dann bietet die Horwitz-Funktion eine gute Möglichkeit, die zu erwartende Messwertstreuung abzuschätzen. Horwitz stellte fest, dass die prozentuale Messwertstandardabweichung  $RSD_R(\%)$  im wesentlichen von der Konzentration  $c$  des Analyten (in g/g) in der Ausgangssubstanz (also z.B. in einer Tablette oder einer Infusionslösung) abhängt, relativ unabhängig von der verwendeten analytischen Technik oder der Probenvorbereitung (Gleichung 8, Tabelle 6; [15-17]).

$$(8) \quad RSD_R(\%) = 2^{(1-0.5\log_{10} c)}$$

In vielen Fällen ist der Zusammenhang zwischen Präzision und Probenkonzentration bereits untersucht worden (z.B. [18]). Das unerwartete und attraktive an der Horwitz-Beziehung ist jedoch, dass diese unabhängig von der verwendeten analytischen Technik gültig ist. Möglicherweise ist die Probenvorbereitung eine so dominante Fehlerquelle, dass sie alle anderen Streueffekte überdeckt (vgl. [12]). Bei höherem Gehalt in der Urschubstanz sind meist Verdünnungsschritte erforderlich, unabhängig von der anschließend verwendeten analytischen Technik. Niedrige Konzentrationen erfordern teils Aufarbeitungsschritte zur Erhöhung der Empfindlichkeit, wiederum relativ unabhängig von der anschließend verwendeten Messtechnik.

Gleichung 8 kann auch in 8a umgeformt werden [17]:

$$(8a) \quad RSD_R(\%) = 2 \cdot c^{-0.1505}$$

Diese Form der Gleichung ist interessant, weil sie die durchschnittliche Abhängigkeit der Varianzfunktion von der Konzentration beschreibt und damit auch wichtige Information zur richtigen Auswahl von Wichtungsfunktionen liefert.

Tabelle 6: Zusammenhang zwischen Konzentration  $c$  des Analyten in der Ausgangssubstanz und prozentualer Messwertstreuung  $RSD_R(\%)$ . Ergebnisse der Horwitz-Gleichung (8) [15-17]; in diese Gleichungen wird  $c$  in der Einheit g/mL eingesetzt, aus Gründen der Übersichtlichkeit wurde für diese Tabelle eine gebräuchlichere Einheit für die Konzentration gewählt.

Der erhaltene Wert für  $RSD_R(\%)$  ist ein Durchschnittswert, der für Ringversuche gilt, in der Praxis werden sowohl halb als auch doppelt so große Werte gefunden. Durchschnittliche Wiederhol-Standardabweichungen innerhalb eines Labors betragen etwa  $\frac{1}{2}$  bis  $\frac{2}{3}$  dieses Wertes [17].

$c(\text{g/kg})$	$RSD_R(\%)$
1000	2
100	2,8
10	4
1	5,7
0,1	8
0,01	11,3
0,001	16
1,E-04	22,6
1,E-05	32
1,E-06	45,3
1,E-07	64

Es ist offensichtlich, dass die Horwitz-Gleichung nur grobe Schätzungen von Durchschnittswerten ermöglicht, da sie allgemein für die verschiedensten analytischen Methoden gültig ist. Die aus der Horwitz-Gleichung zunächst erhaltenden Werte  $RSD_R(\%)$  gelten für Ringversuche. Aus diesem Wert kann leicht die Standardabweichung für Versuchsreihen innerhalb eines Labors berechnet werden; diese ist aber nur etwa  $\frac{1}{2}$  bis  $\frac{2}{3}$  so hoch wie die Ringversuchsstandardabweichung [17].

Auch bei der Interpretation der Horwitz-Daten muss die große Unsicherheit bei der Bestimmung von Standardabweichungen berücksichtigt werden. Deshalb müssen die Daten mit weiten Toleranzgrenzen interpretiert werden, real erhaltene können doppelt so hoch oder halb so hoch sein. Die Präzision analytischer Methoden hat sich im Laufe der Jahrzehnte geringfügig verbessert [17].

Tabelle 7: Spezifikationsgrenzen, die durch Einsetzen der Horwitz-Daten (Tabelle 6) in Gleichung 7 (PL = 2.5%) erhalten werden

		daraus resultierende Spezifikationsgrenzen:			unteres (lower) Spezifikationslimit	oberes (upper) Spezifikationslimit
A)	Szenario I					
	c (g/kg)	RSD <sub>R</sub> (%) (vgl. Tabelle 6)	n	analytische Variabilität	LSL	USL
	100	2,8	3	4,2	93,3	106,7
	10	4	3	5,9	91,6	108,4
	1	5,7	3	8,5	89	111
	0,1	8	3	11,9	85,6	114,4
B)	Szenario Ia (n angepasst)					
	c	RSD <sub>R</sub> (%)	n		LSL	USL
	100	2,8	8		2,5	95
	10	4	17		2,5	95
	1	5,7	32		2,6	94,9
	0,1	8	66		2,5	95
C)	Szenario II (bestenfalls erreichbare Vergleichspräzision, die Hälfte der Horwitz-Streuung RSD <sub>R</sub> (%) wird angesetzt)					
	c	RSD <sub>R</sub> (%)	n		LSL	USL
	100	1,4	3		2,1	95,4
	10	2	3		3	94,5
	1	2,9	3		4,3	93,2
	0,1	4	3		5,9	91,6
D)	Szenario IIa (n angepasst)					
	c	RSD <sub>R</sub> (%)	n		LSL	USL
	100	1,4	3		2,1	95,4
	10	2	5		2,3	95,2
	1	2,9	9		2,5	95
	0,1	4	17		2,5	95

Tabelle 7 zeigt, dass sich mit diesen vorläufig abgeschätzten Werten und Gleichung 7 die folgenden Spezifikationsgrenzen ergeben, wenn man n=3 belässt und PL = 2.5% annimmt (A). Eine Konzentration von 100 g/kg entspricht einer Tablette, die etwa 10% Wirkstoff enthält. Die Grenzen 95-105% sind nur zu halten, wenn n drastisch erhöht wird (B). Die Kompensation zwangsläufig fehlender Präzision durch sehr hohe Datenzahlen ist aber nicht sinnvoll, es sei denn, es steht eine besonders schnelle Analytik zur Verfügung. Die grundsätzliche Forderung nach 95-105% bei kleinen Konzentrationen würde einen unverhältnismäßig hohen analytischen Aufwand und Kosten verursachen.

Die Szenarien (A) und (B) sind möglicherweise etwas pessimistisch, da in der pharmazeutischen Analytik besonders leistungsfähige und gut charakterisierte Analysemethoden zur Verfügung stehen. Die Rechnungen werden daher auch mit der halben Horwitzstreuung, also der im günstigen Fall erreichbaren Vergleichspräzision, wiederholt (C, D). Aber auch hier werden schnell unverhältnismäßig hohe Datenzahlen erreicht.

Wir wollen versuchen, im Positionspapier unserer Fachgruppe Eckwerte für zu erwartende Standardabweichungen bei Gehaltsbestimmungen von Arzneistoffen festzulegen. Daraufhin ist es möglich, auch für Zwischenwerte (z.B. 1.5 g/L) einen Sollwert für die Standardabweichung zu berechnen. Mit diesem Sollwert können dann durch Gleichung 7 die an die Wirkstoffkonzentration angepasste Spezifikationsgrenzen festgelegt werden.

#### **4. Spezifikationen für Nebenkomponenten: ein an ICH Q6A orientierter Vorschlag**

Zum Abschluss wurde noch kurz als Diskussionsgrundlage vorgestellt, wie Spezifikationen für Verunreinigungen von Arzneistoffen festgelegt werden können [19]. Das Limit für eine einzelne Substanz SL wird hier nach einer einfachen Formel festgelegt (9):

$$(9) \quad SL = \bar{x} + k \cdot \hat{\sigma}$$

Der erhaltene Wert wird auf eine Dezimale gerundet. Mittelwert und Standardabweichung sollen aus mindestens 5, besser 10 Chargen aus den klinischen Phasen II und III bestimmt werden. In Übereinstimmung mit der ICH-Richtlinie Q6A (S. 83052) [20] wurde die Verwendung von  $k=3$  vorgeschlagen und erschien allen Diskussionsrednern sinnvoll. Ein Vorteil dieses Konzept ist die gemeinsame Berücksichtigung der produktionsbedingten Chargenvariabilität und der Messunsicherheit bei der Berechnung der Standardabweichung. Allerdings liegen Messwerte bei Reinheitsprüfungen häufig in der Nähe der Bestimmungsgrenze. Daher kann nicht davon ausgegangen werden, dass diese Werte normalverteilt sind. Die vorgestellte Vorgehensweise ist aber trotzdem einsetzbar. Bei einer symmetrisch monomodalen Verteilung liegen mindestens 95,1 % aller Werte im Bereich von Mittelwert  $\pm 3 \sigma$ , bei einer beliebigen Verteilung sind es nach Tschebyschew mindestens 88,9 %.

#### **5. Erstellung eines Positionspapier zum Thema „Analytische Unsicherheit und Rationale Spezifikationsfindung“**

Dieser Bericht enthält eine Reihe von Vorschlägen, die wir zur Erstellung eines Positionspapiers nutzen wollen. Wir wollen Ihnen ein ausgereiftes Konzept anbieten und benötigen dazu jetzt Ihre Hilfe. Bitte schreiben Sie uns, wenn Sie Einwände oder Verbesserungsvorschläge haben. Wir werden die hier vorgestellten Inhalte auf unserer Haupttagung vom 8.-10.10.2002 in Berlin erneut diskutieren und im Anschluss unser Positionspapier verabschieden und veröffentlichen.

Wir bedanken uns sehr herzlich bei den Herren Doktoren Manfred Fischer, Bernd Renger und Holger van Lishaut für die kritische Durchsicht des Manuskripts und ihre zahlreichen konstruktiven Vorschläge und Ergänzungen, die wir an vielen Stellen berücksichtigt haben.

- [1] Wätzig, H., Neben- und Abbauprodukte in Wirkstoffen und Fertigarzneimitteln – Identifizierung, Qualifizierung, Bilanzierung, Deutsche Apotheker Zeitung, 141(9), 1073 (2001)
- [2] Wolin, A. M., United States vs. Barr Laboratories Inc., 812 F Supp.458 (New York District Federal Court, February 1993)
- [3] Daas, A. G. J., Miller, J. H. McB., Relationship Between Content Limits, System Suitability for Precision and Acceptance/Rejection Criteria for Assays Using Chromatographic Methods, Pharmeuropa 11(4), 571 (1999)
- [4] Daas, A. G. J., Miller, J. H. McB., Content Limits in the European Pharmacopoeia (Part 2), Pharmeuropa 10(1), 138 (1998)
- [5] Miller, J. H. McB., European Directorate for the Quality of Medicines (EDQM), Strasburg, Setting of content limits in Pharm. Eur., Beitrag zum erwähnten Workshop
- [6] Renger, B., System Performance and Variability of Chromatographic Techniques Used in Pharmaceutical Quality Control, J. Chromatogr. B 745, 167 (2000)
- [7] Küppers, S., Renger, B., Meyer, V. R., Autosamplers – a Major Uncertainty Factor in HPLC Analysis Precision, LC-GC Eur. 13, 114 (2000)
- [8] Schirm, B., Wätzig, H., Peak recognition imitating the human judgement, Chromatographia 48, 331 (1998)
- [9] Schirm, B., Neue Möglichkeiten der Datenauswertung in der Kapillarelektrophorese durch Einsatz chemometrischer Verfahren, Dissertation Würzburg 2000.
- [10] Daas A. G. J., Miller J. H. McB., Content limits in the European Pharmacopoeia, Pharmeuropa 9, 148 (1997)
- [11] Richtlinie 2001/83/EG zur Schaffung eines Gemeinschaftskodexes für Arzneimittel vom 06.11.2001, Anhang I, Teil 1.2.
- [12] Weber, M., Eidgenössische Materialprüfungs- und Forschungsanstalt (EMPA), St. Gallen, Unsicherheit über die Unsicherheit, Beitrag zum erwähnten Workshop
- [13] Ermer, J., A Refined Concept for Setting / Verifying Acceptance Limits in Pharmaceutical Content Specifications, Vortrag auf 4<sup>th</sup> World Meeting on Pharmaceutics, Biopharmaceutics, Pharmaceutical Technology, Florenz, 8.-11.04.2002; Kurzfassung im Tagungsband
- [14] [www.measurementuncertainty.org](http://www.measurementuncertainty.org); [www.uncertaintymanager.com](http://www.uncertaintymanager.com)
- [15] Horwitz, W., Evaluation of Analytical Methods Used for Regulation of Foods and Drugs, Anal. Chem. 54, 67A (1982)
- [16] Horwitz, W., Albert, R., Reliability of the Determinations of Polychlorinated Contaminants (Biphenyls, Dioxins and Furans), J. AOAC Int. 79, 589 (1996)
- [17] Albert, R., Horwitz, W., A Heuristic Derivation of the Horwitz Curve, Anal. Chem. 69, 789 (1997)
- [18] Wätzig, H., Dette C., Precise Quantitative Capillary Electrophoresis (CE): Methodological and Instrumental Aspects, J. Chromatogr. 636, 31 (1993)
- [19] Köller, G., Boehringer Ingelheim, Biberach, Practise on Specification Finding for the Impurities of a Drug Substance, Beitrag zum erwähnten Workshop
- [20] International Conference on Harmonisation, Notes for Guidance, Q6A.  
<http://www.fda.gov/cder/guidance/ichq6a.pdf>
- [21] International Conference on Harmonisation, Notes for Guidance, Q2A.  
<http://www.fda.gov/cder/guidance/ichq2a.pdf>
- [22] DIN 58936 Teil 2, zitiert in Funk, W., Dammann, V., Donnevert, G., Qualitätssicherung in der analytischen Chemie, VCH Weinheim 1992

Korrespondenz: Prof. Dr. Hermann Wätzig, Institut für Pharmazeutische Chemie, Beethovenstr. 55, 38106 Braunschweig.

[h.waetzig@tu-bs.de](mailto:h.waetzig@tu-bs.de)

<http://www.tu-bs.de/institute/pharmchem/phchem-waetz.htm>