

# Sequenzielle Methoden als Entscheidungshilfe bei der analytischen Prüfung von Grenzwerten = Sequential methods in the evaluation of analytical results for regulatory purposes

Autor(en): **Ferrini, B.**

Objektyp: **Article**

Zeitschrift: **Mitteilungen aus dem Gebiete der Lebensmitteluntersuchung und Hygiene = Travaux de chimie alimentaire et d'hygiène**

Band (Jahr): **75 (1984)**

Heft 2

PDF erstellt am: **10.08.2024**

Persistenter Link: <https://doi.org/10.5169/seals-982706>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern. Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden. Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# Sequenzielle Methoden als Entscheidungshilfe bei der analytischen Prüfung von Grenzwerten

Sequential Methods in the Evaluation of Analytical Results  
for Regulatory Purposes

*B. Ferrini*  
SOS Labor, Zürich

## Einleitung

Zweck dieses Beitrages ist die allgemeine Eruiierung der Anzahl erforderlicher Messungen für die Beantwortung analytischer Fragestellungen.

Nach einem Überblick über die allgemeine Problematik werden die sogenannten Folgetestpläne, die Wahl der anzuwendenden statistischen Parameter und die Größe der Standardabweichung als Funktion der Konzentration besprochen. Es wird ferner ein einfaches Mittel für die Überprüfung der Aussagekraft der eingesetzten Anzahl der Stichproben anhand der Entscheidungstabellen der Folgetestpläne angegeben.

Da der Inhalt des gehaltenen Vortrages den üblichen Rahmen dieser Zeitschrift sprengt – es werden insbesondere die speziellen Gebiete der Statistik und des Meßwesens tangiert –, beschränkt sich dieser Beitrag auf eine Übersicht der behandelten Problematik. Die vollständige Abhandlung mit Formeln und graphischen Darstellungen steht jedoch auf Anfrage interessierten Lesern zur Verfügung.

## Allgemeine Einführung

Neben dem Entwurf und der Durchführung analytischer Kontrollen – d. h. eine genaue Befolgung einer analytischen Vorschrift, an deren Ende eine Größe oder ein Satz von Größen erhalten wird, die das Ergebnis darstellt – ist die richtige Auswertung, insbesondere die vollständige und optimale Ausnutzung des gewonnenen Datenmaterials eine der wesentlichen Aufgaben des Analytikers (1). Will man die Tatsache berücksichtigen, daß bei der Wiederholung irgendeiner Messung im allgemeinen numerisch verschiedene Ergebnisse erhalten werden, dann sind statistische Methoden unentbehrlich, um begründete Angaben über die Aussagekraft von Analysenergebnissen zu ermöglichen. Als Endergebnis einer

Kontrolle kann entweder ein Entscheid bezüglich z. B. gemessener Abweichungen von einem Sollwert mit vorgegebenen Risiken oder ein numerisches Resultat mit entsprechenden Vertrauensgrenzen (Minimal- und Maximalwerte, die mit den Meßergebnissen bei vorgegebenen Fehlerrisiken noch kompatibel sind) erhalten werden.

Die Planung eines Experimentes hängt von den Anforderungen ab, die an die Zwischenentscheide und die erwähnten Endergebnisse gestellt worden sind, wobei folgende vier statistische Grundsätze zu beachten sind: das Wiederholen, das zufällige Zuordnen, das Vorsehen von Blöcken mit möglichst gleichartigen Versuchseinheiten, das symmetrische Aufbauen des Versuches (2).

Eine chemisch analytische Kontrolle muß so geplant werden, daß die potentielle Inhomogenität der Proben und die Unsicherheiten in der Richtigkeit der eingesetzten Referenzen (d. h. die Meß- und Eichfehler) einzeln überprüft und eventuell mit Erfahrungswerten verglichen werden können (3). Die Planung der Eichmessungen ist entscheidend für die Vertrauenswürdigkeit der Eichfunktion. Äußerst wichtig ist die Bedingung, daß die eingesetzten Referenzen den in Frage kommenden Konzentrationsbereich umfassen müssen, da außerhalb dieses Bereiches die Unsicherheiten durch Extrapolation der Eichfunktion beträchtlich zunehmen.

Die einfache Absicherung analytischer Schätzungen einer Konzentration verlangt neben dem eigentlichen Meßgut die Anwendung von Referenzen und Kontrollen. Definitionsgemäß besteht ein Unterschied zwischen Kontrollen und aktiven Referenzmessungen; letztere dienen zur Bildung der Eichfunktion, während die Kontrollmessungen passiv zum Vergleich des gefundenen Wertes mit einem erwarteten Wert (Consensus-Bereich) eingesetzt werden.

Die Wiederholung der Messungen der einzelnen Proben ergibt neben einer besseren Schätzung vom unbekanntem Wert durch Anwendung der Eichfunktion einen zusätzlichen Einblick auf die probenbedingte Streuung.

Eine sogenannte Gehaltsbestimmung umfaßt im allgemeinsten Fall (2) folgende Schritte:

1. Prüfen der Linearität der Beziehung zwischen Ansatzgehalt und Meßwert;
2. Prüfen der Parallelität bzw. Identität der Ansatzgehalt-Meßwert-Geraden für Referenz- bzw. Probenansätze;
3. Prüfen der Homogenität der Streuungen der Meßwerte;
4. Schätzen des Gehaltes im Ansatz bzw. der Konzentration in der Probe;
5. Ermitteln der Genauigkeit des Schätzens.

Die Absicherung der Meßergebnisse wird bei besonders schwerwiegenden Fällen durch den Einsatz einer zweiten Methodik erreicht, wobei die Kompatibilität der Ergebnisse der zwei Methoden für das Spezimen überprüft wird. Die Planung der sogenannten Alternativmethode muß jegliche Wechselwirkung zwischen den beiden Methoden ausschließen können. Dies erfolgt einerseits durch Anwendung von Referenzen verschiedener Konzentrationen oder noch besser durch Anwendung von Referenzen verschiedener qualitativer Eigenschaften, andererseits durch das Messen einer andersartigen Größe, die durch den zu untersu-

chenden Stoff chemisch bzw. physikalisch-chemisch erzeugt und analytisch benützt wird (von Grund auf wesensverschiedene Methoden). Der Einsatz einer zweiten Methodik kann als Wiederholung der Bestimmung durch ein zweites unabhängiges Laboratorium erfaßt werden. Dadurch wird auch die Bedingung der Austauschbarkeit der Ergebnisse zwischen den Laboratorien einigermaßen berücksichtigt, insofern die Alternativmethoden wirklich die Forderung nach Unabhängigkeit erfüllen.

Beim Methodenvergleich kann es vorkommen, daß Inkompatibilitäten in den Ergebnissen festgestellt werden müssen. In einem solchen Fall sind Plausibilitäts-erwägungen (z. B. methodenbedingte, systematische Fehler) als zusätzliches Kriterium für die Wahl des vertrauenswürdigeren Ergebnisses unentbehrlich.

### Statistische Absicherung von Ergebnissen

Die Wahl geeigneter Stichprobenumfänge, d. h. die Anzahl der erforderlichen Messungen zur Beantwortung der Frage, ob ein analytisch zu kontrollierendes Produkt eine bestimmte Qualitätsanforderung erfüllt, setzt nicht nur voraus, daß man die kritische Größe kennt, d. h. die maximal bzw. minimal zugelassene Konzentration eines bestimmten Stoffes, sondern es müssen noch die sogenannten Risiken 1. und 2. Art festgelegt werden, die mit dem darauffolgenden Entscheid, d. h. der Zulassung oder Ablehnung des Produktes aufgrund der Kontrolle, verbunden sind. In dieser Hinsicht muß die Größe des kleinsten noch festzustellenden bedeutsamen Konzentrationsunterschiedes (sog. Trennschärfe) betrachtet werden. Meistens wird dieser Unterschied implizit durch die zahlenmäßige Angabe des Grenzwertes festgelegt. Es ist z. B. nicht gleichgültig, ob man einen Grenzwert mit 1,0 mg/kg oder mit 1,00 mg/kg angibt. Im ersten Fall setzt man eine sogenannte Trennschärfe von 0,1 mg/kg indirekt fest, während im zweiten Fall eine Trennschärfe von 0,01 mg/kg durch die Angabe des Grenzwertes mit zwei Nullen nach dem Komma postuliert wird.

Jede Aussage bzw. Entscheidung ist mit zwei Fehlerrisiken verbunden:

1. Der Fehler 1. Art entsteht bei der unberechtigten Ablehnung der sogenannten Nullhypothese, d. h. es wird aufgrund des Testes entschieden, daß der Grenzwert nicht eingehalten wurde, obwohl in Wirklichkeit das Produkt die gestellte Anforderung erfüllt.
2. Der Fehler 2. Art entsteht beim unberechtigten Beibehalten der Nullhypothese, d. h. es wird aufgrund des Testes entschieden, daß der Grenzwert eingehalten wurde, obwohl in Wirklichkeit das Produkt die gestellte Anforderung nicht erfüllt.

Diese zwei Fehler führen zu einer unberechtigten Ablehnung bzw. einer unberechtigten Zulassung eines Analysengutes, das durch die Aufsichtsbehörde kontrolliert wurde. Will man besonders kritisch vorgehen und als Kontrollbehörde einen gewissen Widerstand gegenüber Fehlentscheidungen zuungunsten der betroffenen Produzenten leisten, ohne jedoch die Anforderungen zugunsten des

Verbrauchers zu übersehen, dann müssen beide Fehlerarten bei der Planung des Versuches, insbesondere bei der Festlegung des Stichprobenumfanges, berücksichtigt werden. Das Risiko 2. Art, d. h. die möglichst kleine Wahrscheinlichkeit, eine falsche sogenannte Nullhypothese beizubehalten, hängt vom Umfang der Stichprobe und vom Grad des noch als bedeutsam anzusehenden Konzentrationsunterschiedes (Trennschärfe) ab.

### Die Problematik der erforderlichen Anzahl Messungen

Die Behörden setzen meist voraus, daß das analytisch erhaltene Meßgut normalverteilt sei. Diese Annahme wurde in der Vergangenheit sehr oft pragmatisch und kritiklos von den Analytikern übernommen. Die heutigen technischen Möglichkeiten der EDV lassen in vielen Arbeitsbereichen beträchtliche Zweifel bezüglich der Normalverteilung von Stammdaten aufkommen (z. B. in der Blutalkoholanalytik (4, 5)). Als Alternative müßten dementsprechend parameterfreie statistische Testverfahren in Betracht gezogen werden.

Die Auswirkungen etwaiger Abweichungen von der Normalität auf die sogenannten statistischen Signifikanztests werden in der Literatur diskutiert (6). Man kann behaupten, daß sich bei vier echt unabhängigen Messungen die befürchteten Konsequenzen auf einen durchaus akzeptablen Rahmen beschränken.

Gemäß *Riedwyl* (7) kann die Anzahl der erforderlichen normalverteilten Messungen mit Standardabweichung  $\sigma$  für ein Fehlerrisiko 1. und 2. Art ( $\alpha$  bzw.  $\beta$  mit einer Trennschärfe  $d$  durch die Gleichung

$$n = (z_{\alpha} + z_{\beta})^2 (\sigma / d)^2 = k (\sigma / d)^2 \quad (1)$$

geschätzt werden ( $z_{\alpha}$  und  $z_{\beta}$  sind die entsprechenden Schranken der standardisierten Normalverteilung). Um z. B. eine Zuverlässigkeit von 99% (Wahrscheinlichkeit, keinen falschen Unterschied  $d$  anzunehmen) und eine 99%ige Wahrscheinlichkeit, einen tatsächlichen Unterschied  $d$  zu entdecken, zu erreichen, ist der Faktor  $k$  für sogenannte einseitige bzw. zweiseitige Fragestellungen mit dem Wert 21,65 bzw. 24,03 in der Schätzung von Stichprobenumfängen einzusetzen. Besteht der Zweck der Datenanalyse darin, einen Unterschied in einer bestimmten Richtung zu überprüfen, dann werden sogenannte einseitige Tests eingesetzt. Gibt es über die Richtung eines vermutlichen Größenunterschiedes keine Klarheit, dann sind sogenannte zweiseitige Tests durchzuführen.

Bestehen Unsicherheiten in bezug auf  $\sigma$ , d. h. es ist nur eine Schätzung dieser Größe bekannt, dann sind entsprechende Schranken der  $t$ -Verteilung in Gleichung (1) einzusetzen.

Die Anwendung von Gleichung (1) ergibt eine feste Anzahl von Messungen. Da meistens die Fehlerrisiken und die Standardabweichung vorgegeben sind und  $n$  aus praktischen Gründen nicht zu groß ausfallen darf, ergibt sich, daß die Wahl der Trennschärfe aus Gründen der Praktikabilität nur einem beschränkten will-

kürlichen Rahmen zugesprochen werden kann. Dieser Tatsachenbestand muß bei der Festlegung von Spezifikationen unbedingt berücksichtigt werden.

Eine ungenügende Anzahl Messungen kann je nachdem zugunsten oder zuungunsten auf eine der beiden Parteien abgewälzt werden (Erhöhung eines der beiden Fehlerrisiken). Gleichung (1) zeigt eine Ambivalenz durch die interpretatorischen Unsicherheiten des Faktors  $(z_\alpha + z_\beta)^2$ . Eine ausgewogene Auswertung, die von den beiden Parteien akzeptiert werden kann, verlangt Fehlerrisiken 1. und 2. Art derselben Größenordnung.

Die zentrale Rolle der Anzahl Messungen, die für zuverlässige Entscheidungen erforderlich ist, wurde durch die Anwendung sogenannter Folgetestpläne (Sequential-Analyse) verfeinert. Gute Übersichten der sequentiellen Methoden sind in Standardwerken der angewandten Statistik, z. B. (8) und (9), zu finden. Die schweizerische Gesetzgebung hat dieses Verfahren in der Deklarationsverordnung (10) angewandt.

Das Rationale dieser Methodik besteht in der intuitiven Überlegung, daß bei offensichtlichen Abweichungen von der Spezifikation rascher Entscheidungen getroffen werden können, d. h. mit weniger Messungen, als bei dubiosen Fällen. «Das Grundprinzip der Sequenzanalyse besteht darin, daß nach vereinbarungsmäßiger Festlegung der Fehler 1. und 2. Art,  $\alpha$  und  $\beta$ , bei gegebenem Unterschied betrachteter Grundgesamtheiten der zur Sicherung dieses Unterschiedes erforderliche Stichprobenumfang zur Zufallsvariablen wird. Vorausgesetzt werden Zufallsstichproben aus unendlich großen Grundgesamtheiten. Man sammelt jeweils nur so viele Beobachtungen, wie unbedingt erforderlich sind» (8).

Die Folgetestpläne sind sehr einfach zu berechnen und ermöglichen dem Analytiker, objektiv die angewandten Entscheidungsschemas (auch für festgelegte Stichprobenumfänge) zu beurteilen. Folgetestpläne können für jede beliebige Verteilung mit der Methodik der Maximum Likelihood berechnet werden (11). Als Hypothesen können sowohl Mittelwerte wie auch Streuungen geprüft werden.

Prinzipiell werden kritische Grenzwerte für das Erstellen weiterer Messungen für eine gemessene Größe aufgrund der Anzahl getätigter Beobachtungen (Attribute oder Messungen) festgelegt. Als Prüfgröße von Meßwerten wird meistens entweder die Summe der Werte oder der Mittelwert angewandt. Die Wahl der einzelnen Parameter ( $\alpha$ ,  $\beta$ ,  $d$ ) wird aufgrund der daraus folgenden erforderlichen Anzahl Messungen bestimmt.

In Abbildung 1 wird die graphische Darstellung eines einseitigen Folgetestplanes schematisch wiedergegeben.

Es ist zu betonen, daß die praktische Anwendung von Folgetestplänen echte unabhängige Messungen voraussetzt und daß die angenommene Standardabweichung der Messungen für das Meßgut hieb- und stichfest als gesichert gilt. Hier liegt die größte Schwierigkeit für die Praktikabilität dieser statistischen Methodik, die jedoch als allgemein auftretende Problematik für alle statistischen Tests, die eine Normalverteilung mit bekannter Standardabweichung voraussetzen, zu bezeichnen ist.

Es ist offensichtlich, daß die Vereinbarungen, die im voraus getroffen werden müssen, maßgebend für die Entscheidungsregeln sind. Es werden meistens die

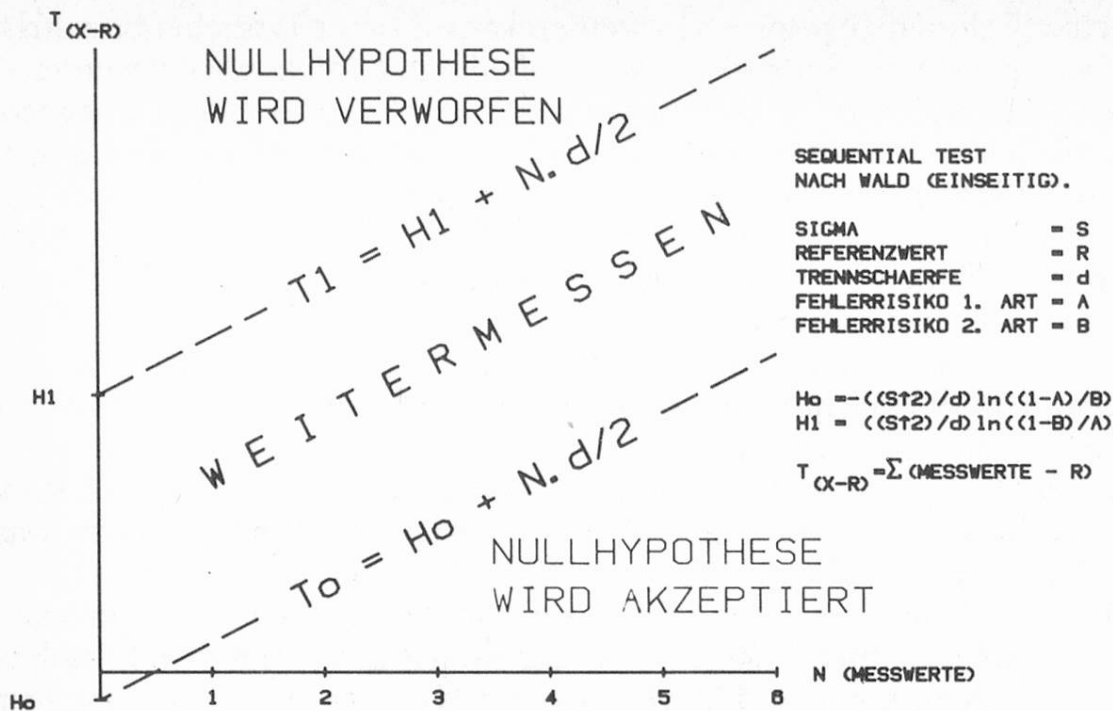


Abb. 1. Schematische Darstellung eines einseitigen Folgetestplanes (Normalverteilung)

Größe der Standardabweichung (innerhalb und zwischen den Laboratorien), die zugelassene Fehlerrisikengröße (die Trennschärfe) und die anzunehmende Verteilung (Normal- bzw.  $t$ -Verteilung) in Diskussion gestellt.

Das Aufstellen von Entscheidungstabellen mit der Methodik der Maximum Likelihood erlaubt, für eine bestimmte Anzahl Messungen die kritischen Grenzen der Prüfgröße festzustellen, die überschritten werden müssen, um einen Entscheid zugunsten einer der beiden Hypothesen zu fällen. Das entspricht der üblichen Praxis, einen sogenannten Sicherheitsabzug bzw. Sicherheitszuschlag in Betracht zu ziehen. Eine wesentliche Eigenschaft dieser Methodik ist jedoch das Verzicht auf die Angabe eines sogenannten Vertrauensbereiches. Dies ist meines Erachtens durchaus gerechtfertigt, wenn man bedenkt, daß die vom Analytiker verlangte beste Schätzung des «wahren» Wertes eine nicht vollumfänglich realisierbare Anforderung darstellt. Was jedoch vom Analytiker verlangt werden kann, ist der Entscheid, ob ein Meßgut bestimmte Bedingungen erfüllt oder nicht.

In Abbildung 2 wird ein Folgetest simuliert. Es sei angenommen, daß es darum geht, durch eine Reihe unabhängiger Analysen (Standardabweichung = 0,028 g/kg) festzustellen, ob eine Probe den kritischen Wert von 0,8 g/kg überschritten hat oder nicht. Es werden Fehlerrisiken 1. und 2. Art von 1% toleriert. Die statistische Prüfung sollte «sicher» eine Probe mit «wahrem» Wert von 0,9 g/kg von einer solchen mit 0,8 mg/kg unterscheiden können (Trennschärfe = 0,9–0,8 = 0,1 g/kg: einseitiger Test). Es ist möglich, mit einem Rechner normalverteilte Ergebnisse mit Mittelwert 0,900 g/kg und Standardabweichung 0,028 g/kg zu erzeugen.

Die in Abbildung 2 aufgeführte Tabelle der kritischen Mittelwertgrenzen als Funktion der vorhandenen Anzahl Messungen zeigt z. B., daß mit vorgegebenen Fehlerrisiken der entsprechende Entscheid getroffen werden kann, daß die Probe

SEQUENTIAL TEST  
 SIGMA = 0.028 REFERENZWERT = 0.8 ; TRENNSCHAERFE = 0.100  
 FEHLERRISIKO 1. ART = 0.0100; FEHLERRISIKO 2. ART = 0.0100  
 100 SIMULIERUNGEN MIT WAHEM WERT = 0.900

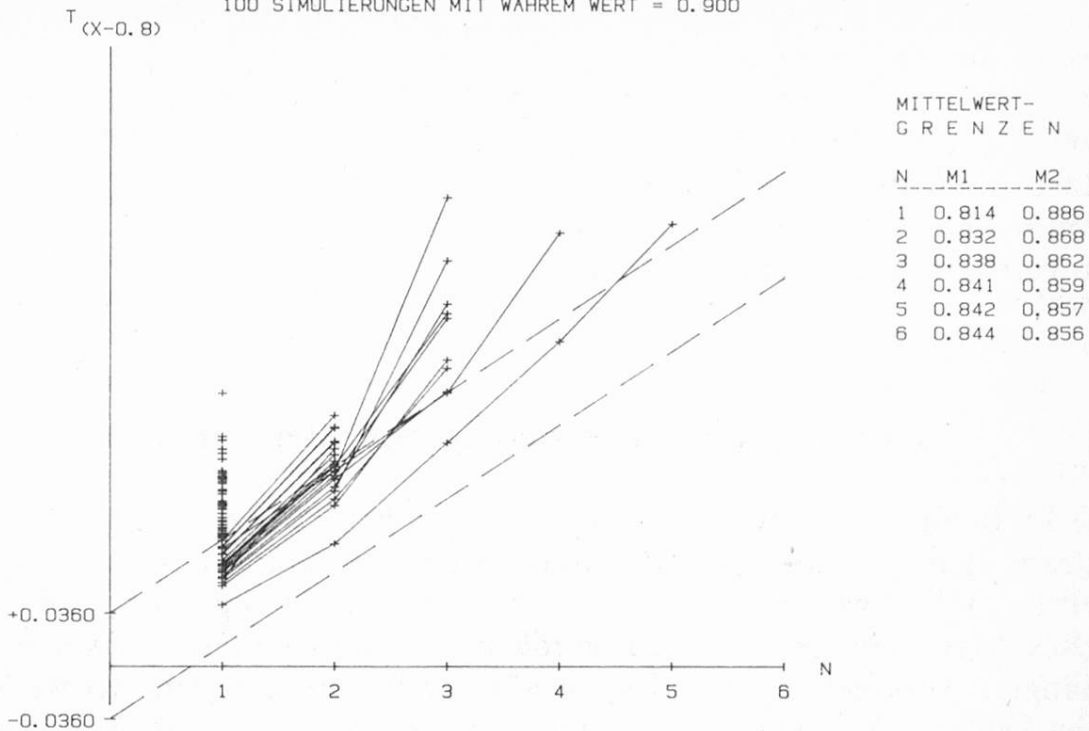


Abb. 2. Monte-Carlo-Simulationen eines einseitigen Folgetestes

die kritische Konzentrationsgrenze von 0,8 g/kg überschritten hat, wenn nach drei Messungen die Mittelwertgrenze von 0,862 g/kg erreicht bzw. überschritten wird. Wird bei drei Messungen die untere Grenze von 0,838 g/kg nicht erreicht, dann darf angenommen werden, daß die Probe die kritische Grenze nicht überschritten hat. Liegt der Mittelwert nach drei Messungen innerhalb der Grenzen 0,838 und 0,862 g/kg, dann muß, um die festgelegten Fehlerrisiken nicht zu überschreiten, weiter gemessen werden.

Abbildung 2 zeigt, wie oft die Mittelwertgrenzen für eine hypothetische Probe mit «wahrem» Wert 0,9 g/kg überschritten werden, d. h. wieviele unabhängige Messungen nach diesem Folgetestplan erforderlich sind. In ca. 10% der simulierten Fälle dieses Beispiels sind zwei Messungen ungenügend, um den Entscheid einer Abbrechung der Meßreihe zu treffen. In Anbetracht des relativ hohen Wertes 0,9 g/kg ist es selbstverständlich, daß sehr oft schon nach der ersten Messung die Meßreihe mit vorgegebenen Fehlerrisiken und Trennschärfe abgebrochen werden kann.

Unter Umständen kann es vorkommen, daß es praktisch nicht möglich ist, mehr als z. B. vier Messungen auszuführen. Wenn ein Analytiker unter einem solchen Zwang steht und die bis anhin erhaltene Mittelwertgrenze ihn auffordert, trotzdem weiter zu messen, dann sieht sich der Analytiker gezwungen, willkürlich zwischen einer der beiden Entscheidungsmöglichkeiten zu wählen, da durch die Fragestellung von ihm eine Antwort erwartet wird. Es ist offensichtlich, daß beim Unterbrechen einer Meßreihe vor dem Erreichen der erforderlichen



Anzahl Messungen die Antwort so gestaltet werden muß, daß etwaige Nachteile, die durch Fehlentscheide entstehen könnten, für die Betroffenen minimisiert werden müssen. Eine vorzeitige Abbrechung der Meßreihe entspricht einer Vergrößerung der Fehlerrisiken.

Die praktische Anwendung dieser Methodik setzt voraus, daß man durch eine Vereinbarung die Anforderungen bezüglich Fehlerrisiken und Trennschärfe festsetzen muß. Ferner muß betont werden, daß die vorliegende Methodik der Maximum Likelihood annimmt, daß die Eichfehler und andere systematische Fehler vernachlässigt werden können.

### Statistische Charakterisierung von Meßverfahren

Die Problematik der Wahl der zugelassenen Fehlerrisiken kann für juristisch bedeutsame Fälle in der Schweiz aufgrund der Deklarationsverordnung (10) als geregelt betrachtet werden, da in dieser Verordnung verlangt wird, daß — um ein amtliches Verfahren gegen einen Verdächtigen einzuleiten — zwei statistisch unabhängige Aussagen mit mindestens 95% Treffsicherheit verlangt werden, womit ein Fehlerrisiko gegen den «Verdächtigen» von  $0,05 \times 0,05 = 25 \times 10^{-4} = 0,0025$  noch toleriert wird. Aus Gründen der Rechtsgleichheit sollten entsprechende Fehlerrisiken für andere Vergehen eingehalten werden. Dies ist z. B. bei Radargeschwindigkeitsmessungen im Straßenverkehr (12) der Fall.

Die Größe der anzunehmenden Standardabweichung der analytischen Schätzung eines «wahren» Wertes kann prinzipiell nur durch Ausführung von hieb- und stichfest geplanten Rundversuchen eruiert werden. Es wurde darauf hingewiesen (13), daß seitens der Analytiker die Präzision oft überschätzt wird und daß etliche Schwierigkeiten in der Planung und Ausführung von hieb- und stichfesten Rundversuchen in dieser Hinsicht bestehen. Es gibt jedoch eine offizielle Aussage der Food and Drug Administration (14), woraus aufgrund einer umfangreichen statistischen Auswertung der Rundversuche dieser Behörde als allgemeine Regel im voraus angenommen werden kann, daß folgende empirische Beziehung zwischen dem Rundversuchsvariationskoeffizienten und der Konzentration  $C$  des untersuchten Stoffes in einer Matrix besteht:

$$\text{Rundversuchsvariationskoeffizient } VK \text{ (Rel.\%)} = 2^{(1-0,5 \log C)} \quad (2).$$

Gemäß der erwähnten Stellungnahme kann ferner angenommen werden, daß innerhalb eines einzelnen Laboratoriums ca. die Hälfte des geschätzten Fehlers stattfindet. Für z. B. die Bestimmung eines Stoffes im Konzentrationsbereich von 1 g/kg kann gemäß angegebener Formel ein Rundversuchsvariationskoeffizient von ca. 5–6 Rel.% vorausgesehen werden. In den einzelnen Laboratorien sind Messungen im angegebenen Konzentrationsbereich mit einem von Tag zu Tag Variationskoeffizienten von ca. 2–3 Rel.% vorauszusehen. Innerhalb einzelner Meßreihen kann ein Variationskoeffizient von ca. 1–1,5 Rel.% abgeleitet werden.

Für Konzentrationen im mg/kg-Bereich (ppm) kann ein Rundversuchsvariationskoeffizient von ca. 16 Rel.% postuliert werden. Es wurde auch darauf hingewiesen (14), daß bei der Beurteilung eines Verfahrens die Anfälligkeit auf Ausreißer berücksichtigt werden muß.

Die FDA-Schätzungen müssen als optimale Grenzwerte betrachtet werden, womit nicht ausgeschlossen werden kann, daß in einem einzelnen Laboratorium mit größeren Standardabweichungen als den postulierten gearbeitet wird. Es besteht also das Bedürfnis, die Varianz innerhalb des Laboratoriums sowie auch die Richtigkeit der Ergebnisse durch interne Qualitätskontrollen im Griff zu haben. Insbesondere muß die Varianz der einzelnen Proben überprüft bzw. die Möglichkeit des Vorkommens von Ausreißern berücksichtigt werden. Dies kann selbstverständlich nur mit einer Wiederholung der Messungen erreicht werden – bei der Interpretation der Ergebnisse unter Anwendung von *t*- bzw. *F*-Verteilungen bzw. sogenannter Ausreißertests. Die Entscheidungstabellen von Folgetestplänen mit *t*-Verteilungen zeigen, daß eine vernünftige Schätzung der Varianz einer Probe erst ab 4 Messungen möglich ist.

Wird die Zerlegung der Rundversuchsstreuung berücksichtigt, dann kann man entsprechende Sicherheitsmargen schätzen. In Abbildung 3 wird angenommen, daß die Rundversuchsstandardabweichung für eine Konzentration im 1 g/kg-Bereich gemäß FDA (d. h. 0,056 g/kg) in zwei Anteilen zu je 0,028 g/kg für die Streuung innerhalb und zwischen den Laboratorien (bzw. Methoden) vorliegt. Werden zwei Methoden (d. h. *i* = 2) eingesetzt und Fehlerrisiken 1. und 2. Art von 0,0025 gemäß Deklarationsverordnung (10) postuliert und wird angenommen,

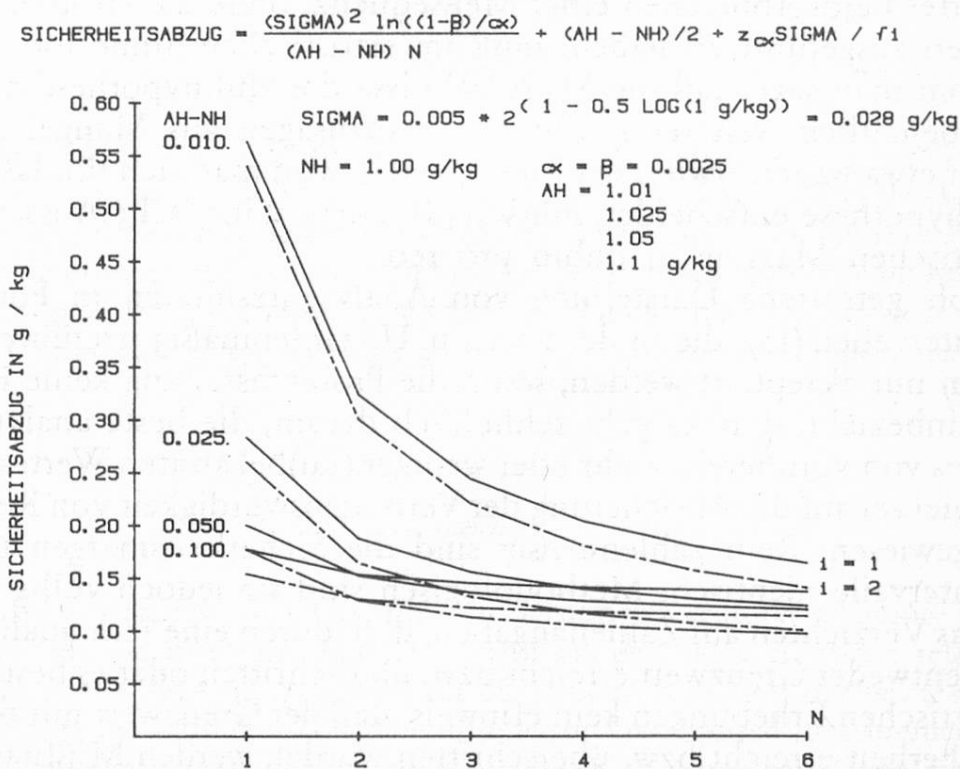


Abb. 3. Sicherheitsmargen gemäß Methodik der Maximum Likelihood für  $\bar{x} = 1,00 \text{ g/kg}$  ( $\sigma$  nach Horwitz mit Zerlegung). Trennschärfe 0,01; 0,025; 0,05 und 0,1 g/kg

daß die einzelnen Messungen in verschiedenen Meßreihen erfolgen (d. h. 0,028 g/kg ist die Standardabweichung von Tag zu Tag bzw. 0,014 g/kg ist die Standardabweichung innerhalb der Meßreihen), so sind für z. B. Trennschärfen von 0,01 g/kg und zwei Messungen Sicherheitsabzüge bzw. -zuschläge der Größenordnung von 0,3 g/kg vorauszusehen. Bei insgesamt vier Messungen sind dagegen ca. 0,2 g/kg als Sicherheitsmarge zu schätzen. Diese Betrachtungen zeigen eindeutig die Rolle der Anzahl Messungen bei der Festlegung von Sicherheitsmargen.

### Schlußfolgerungen

Werden die angeführten Erwägungen vollumfänglich bei der Planung und Auswertung von analytischen Bestimmungen zur Überprüfung des Einhaltens gesetzlicher Anforderungen berücksichtigt, dann ermöglicht die Methodik der Maximum Likelihood klare Antworten auf klare Fragestellungen. Folgetestpläne erlauben nur, entweder mit ja oder nein eine Frage zu beantworten, sofern genügend Messungen und die gestellten Anforderungen bezüglich Fehlerrisiken und Trennschärfe erfüllt worden sind. Als einzige Alternative bleibt die Möglichkeit offen, entweder weiter zu messen oder (willkürlich) zugunsten einer der möglichen Entscheidungen Stellung zu nehmen. Dabei müssen jedoch – wie schon erwähnt – im voraus entsprechende Vereinbarungen getroffen werden, um die prinzipiell willkürliche Entscheidungsstrategie zu rechtfertigen.

Das Beibehalten der Nullhypothese, d. h. z. B. das Nichtüberschreiten eines Grenzwertes beim Abbrechen einer Meßsequenz, ohne die erforderliche Anzahl Messungen ausgeführt zu haben, muß im statistischen Sinne interpretiert werden, indem man sagt, daß die Meßergebnisse die Nullhypothese nicht widerlegen. Vorbehaltlich weiterer Prüfungen – sozusagen aus Mangel an Beweisen und nicht etwa wegen erwiesener Richtigkeit – wird man sich für das Beibehalten der Nullhypothese entscheiden müssen (8). Diese statistische Aussage entspricht der juristischen *Maxime in dubio pro reo*.

Die oft getroffene Darstellung von Analysenergebnissen in Form von Vertrauensintervallen (15), die in der Praxis u. U. zahlenmäßig irreführend sein können, kann nur akzeptiert werden, wenn die Problemstellung keine Entscheidungen miteinbezieht, d. h. es geht schließlich darum, die beste analytische Schätzung eines von vornherein mehr oder weniger (un)bekannten Wertes festzulegen. Als Beispiel sei auf die Absicherung der Vertrauenswürdigkeit von Referenzlösungen hingewiesen. Rein zahlenmäßig sind die Sicherheitsmargen und die Vertrauensintervalle identisch: Methodologisch sind sie jedoch völlig verschieden. Durch das Verzicht auf Zahlenangaben, d. h. durch eine rein qualitative Aussage (z. B. entweder Grenzwert erreicht bzw. überschritten oder es besteht aufgrund der analytischen Erhebungen kein Hinweis, daß der Grenzwert mit der erforderlichen Sicherheit erreicht bzw. überschritten wurde), werden Mißbräuche, wie sie oft im Rahmen eines freien Ermessens seitens des Benützers analytischer Ergebnisse vorkommen, verhindert. Die Gültigkeit einer subjektiven Überzeugung fin-

det ihre Grenze dort, wo diese Überzeugung gegen objektive Sachverhalte und gegen die Logik verstößt. Einen Verstoß gegen die Folgerungen der Statistik in gerichtlichen Verfahren kann auch derjenige nicht gutheißen, der den Spielraum der richterlichen Überzeugungsbildung möglichst weit bemessen will (16).

Die Abgrenzung der Praktikabilität dieser Erwägungen liegt in der nicht zu unterschätzenden erforderlichen Arbeit zur Gewinnung der benötigten Information durch die analytischen Bestimmungen.

Diese theoretisch pragmatischen Betrachtungen sollten sachdienlich eingesetzt werden, um die Rahmenbedingungen jeglichen Entscheides, der nach Meßverfahren folgt, rationell zu gestalten und insbesondere in Grenzfällen entsprechend abzusichern.

### *Zusammenfassung*

Die Auswertung von Meßergebnissen zuhanden der Aufsichtsbehörde verlangt Treffsicherheiten, die durch eindeutige Festlegung der Spezifikationen erreicht werden können. Es bestehen bei vorgegebenen Sicherheitsabzügen direkte Zusammenhänge zwischen Trennschärfe als Funktion der Standardabweichung der Messungen und Anzahl der Messungen. Durch Folgetestpläne kann die Anzahl der erforderlichen Messungen optimiert werden. Eine Formel zur Berechnung des sogenannten Minimalwertes wird angegeben. Die Anwendung von Vertrauensintervallen wird im Zusammenhang mit den Folgetestplänen eruiert.

### *Résumé*

L'évaluation des résultats analytiques à l'intention des autorités de contrôle exige une sûreté pouvant être garantie par des spécifications claires. Pour des marges de sécurité données, il existe une relation directe entre le taux de discrimination, fonction de l'écart type des mesures, et le nombre des résultats obtenus. Par des tests séquentiels le nombre des mesures nécessaires peut être optimisé. Une formule pour le calcul de la «valeur minimale» est présentée. L'utilisation d'intervalles de confiance est discutée par rapport aux méthodes séquentielles.

### *Summary*

The evaluation of analytical results for official purposes requires reliability which can be realised by clear specifications. For given error margins there are direct relations between sharpness of the decision as a function of standard deviation of the measurements and the number of the acquired data. The number of the required measurements can be optimised by sequential test plans. A formula for the calculation of the «minimal value» is given. The use of confidence limits is discussed from the point of view of the sequential methods.

## Literatur

1. *Brandt, S.*: Datenanalyse. Mit statistischen Methoden und Computerprogrammen. Wissenschaftsverlag, Bibliographisches Institut Mannheim, Wien, Zürich 1975.
2. *Linder, A.*: Planen und Auswerten von Versuchen. Eine Einführung für Naturwissenschaftler, Mediziner und Ingenieure, dritte erweiterte Auflage. Birkhäuser Verlag, Basel und Stuttgart 1969.
3. *Henschler, D.*: Analytische Methoden zur Prüfung gesundheitsschädlicher Stoffe. Verlag Chemie, Weinheim 1978.
4. *Zink, P.*: Persönliche Mitteilung (Gerichtsmedizinisches Institut Bern).
5. *Ulmer, F.*: Persönliche Mitteilung (Universität Freiburg im Breisgau).
6. *Davies, O. L.*: Statistical methods in research and production. With special reference to the chemical industry, third edition revised. Oliver and Boyd, London 1958.
7. *Riedwyl, H.*: Angewandte mathematische Statistik in Wissenschaft, Administration und Technik. Haupt, Bern 1975.
8. *Sachs, L.*: Angewandte Statistik. Statistische Methoden und ihre Anwendungen. Zugleich fünfte, neubearbeitete und erweiterte Auflage der «Statistischen Auswertungsmethoden» mit neuer Bibliographie. Springer-Verlag Berlin, Heidelberg, New York 1978.
9. *Ireson, W. G.*: Reliability handbook. Executive Head Department of Industrial Engineering Stanford University. McGraw-Hill Book Company New York, San Francisco, Toronto, London, Sydney 1966.
10. Deklarationsverordnung: Verordnung über verbindliche Angaben in Handel und Verkehr mit meßbaren Gütern. SR 941.281 (15. 7. 1970).
11. *Davies, O. L.*: The design and analysis of industrial experiments, second edition. Oliver and Boyd, London 1963.
12. *Koch, P.*: Persönliche Mitteilung (Eidg. Amt für Meßwesen, Bern).
13. *Zender, R.* und *Linder, A.*: Interlaboratory comparison of both accuracy and precision by a two-sample method. In quality control in clinical chemistry, Transactions of the VI<sup>th</sup> International Symposium, Geneva, April 23–25, 1975. Walter de Gruyter Berlin, New York 1975.
14. *Horwitz, W.*: Evaluation of analytical methods used for regulation of foods and drugs. Analytical Chemistry, Vol. 54, No 1, 67 A (January 1982).
15. *Roth, H.-R.*: Allgemeine statistische Überlegungen zur Aussagekraft von Analysenergebnissen. Mitt. Gebiete Lebensm. Hyg. 75, 157–166 (1984).
16. *Lundt, P. V.* und *Jahn, E.*: Gutachten des Bundesgesundheitsamtes zur Frage Alkohol bei Verkehrsstraftaten. Kirschbaum-Verlag, Bad Godesberg 1966.

Dr. B. Ferrini  
SOS Labor  
CH-8033 Zürich