

Zeitschrift:	Mitteilungen aus dem Gebiete der Lebensmitteluntersuchung und Hygiene = Travaux de chimie alimentaire et d'hygiène
Herausgeber:	Bundesamt für Gesundheit
Band:	75 (1984)
Heft:	2
Rubrik:	Erhebung und Vorbereitung von Proben im Hinblick auf aussagekräftige Resultate

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 18.02.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Erhebung und Vorbereitung von Proben im Hinblick auf aussagekräftige Resultate

Im Rahmen der Jahrestagung 1983 der Schweiz. Naturforschenden Gesellschaft führte die Schweiz. Gesellschaft für analytische und angewandte Chemie zusammen mit der Schweiz. Gesellschaft für Instrumentalanalytik und Mikrochemie am 14. Oktober 1983 in Pruntrut ein gemeinsames Symposium durch.

Im folgenden werden die Vorträge von W. Hauert, H. R. Roth, O. Jakob, R. Knutti, J. Solms und B. Ferrini publiziert.

Mitt. Gebiete Lebensm. Hyg. 75, 143–156 (1984)

Praktische Erfahrung bei der mikrobiologischen Qualitätskontrolle

Practical Experience in Microbiological Quality Control

W. Hauert

Wander AG Bern, Neuenegg

Grundsätze – Zielsetzung

Ziel der mikrobiologischen Untersuchungen und Kontrollen ist die ausreichende Sicherung der mikrobiologischen Qualität aller (durch Wander) in den Handel gebrachten Produkte gemäß:

- Gesetzgebung (1)
- den Vorschriften und Richtlinien international anerkannter Organisationen (2–4)
- dem Urteil von Fachspezialisten
- Fachliteratur (5–11) und eigenen Erfahrungen

Dabei ist natürlich ganz wesentlich, daß die Methoden der Untersuchungen den neusten Erkenntnissen entsprechen, zuverlässige, aussagekräftige Resultate ergeben und daß ferner die Arbeitsabläufe zweckmäßig, fachlich optimal und rationell abgewickelt werden.

Die *Grundsätze*, wie wir sie für den ganzen Bereich der Qualitätssicherung und Qualitätskontrolle formuliert haben und denen Bedeutung als Orientierungshilfe bei der Festlegung der Richtlinien sowie bei der Beurteilung praktischer Probleme effektiv zukommt, sind die folgenden:

- Herstellung und Vertrieb von Produkten, die gemäß dem jeweiligen Stand des Wissens eine gute Qualität aufweisen.
- Nur einwandfreie Produkte finden langfristig im Wettbewerb einen zufriedenen Käuferkreis.
- Erfüllung der Vorschriften, Empfehlungen und Richtlinien der Lebensmittelgesetzgebung und der einschlägigen Fachliteratur.

Diese Prinzipien mögen vielleicht sehr theoretisch wirken; dennoch kommt ihnen aber praktische Bedeutung zu, und zwar besonders im weiten Spannungsfeld der Auseinandersetzungen um *Kosten-Nutzen* (insbesondere mit Marketing und Controlling-Organen).

Konzept zur Erreichung der Ziele

Um diese eben doch anspruchsvollen Ziele zu erreichen, setzen wir drei Schwerpunkte, die alle für die Durchführung der mikrobiologischen Untersuchungen und für die Sicherung der mikrobiologischen Produktequalität prinzipiell gleich gewichtet werden (siehe Tabelle 1).

Alle in der Produktion eingesetzten *Rohstoffe* sind erfaßt. Sie werden nur nach der Freigabe durch die Qualitätskontrolle in der Produktion eingesetzt.

Die *Produktionskontrolle* ist primär bezogen auf die Kontrolle der Sauberkeit von Räumen, Anlagen, Apparaten und Einrichtungen.

Die Kontrolle erfolgt optisch und durch stichprobenweise Untersuchung von Abklatschproben sowie vor allem von Tupferproben (die an hygienisch besonders heiklen Stellen erhoben werden, nämlich in Ecken, Ritzen, bei Dichtungen, Verschraubungen usw.).

Die Untersuchung erfolgt auf die An- bzw. Abwesenheit von pathogenen und toxinogenen Keimen sowie von Indikatorkeimen, wie coliforme Keime und Enterokokken.

Ferner umfaßt die Produktionskontrolle (außer den fabrikationshygienischen Kontrollen) noch die personalhygienische Kontrolle und die Prozeßüberwachung (Inprozeßkontrolle).

Bei der personalhygienischen Kontrolle wird durch das vorgesetzte Personal die Sauberkeit am Arbeitsplatz sowie die Sauberkeit der Arbeitskleider und Hände überwacht.

Basis für den Erfolg bildet die Schulung des Personals in Hygiene durch periodisch wiederkehrende Referate, Demonstrationen, Dia-Shows und schriftliche Fragenbeantwortung.

Die Prozeßüberwachung im Sinne von Inprozeßkontrollen hat den Zweck, Prozesse aufzudecken, in denen ungewollte Keimvermehrungen stattfinden können.

Diese Risiken bestehen bei allen Arbeiten mit Wasser und flüssigen Produkten sowie insbesondere bei Eindickungsprozessen, Extraktionen, Trocknungsprozessen usw.

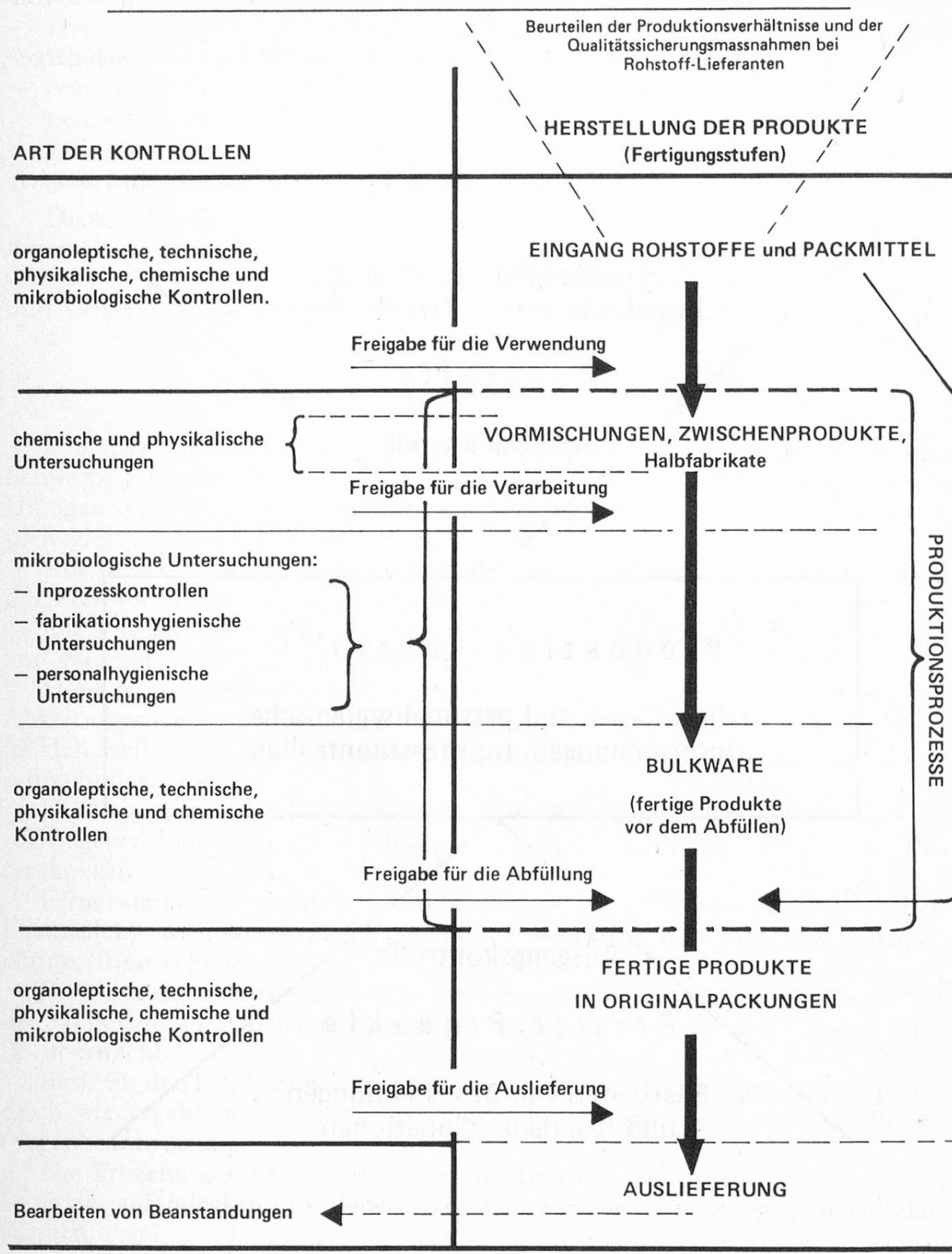
Die fertigen Produkte schließlich werden ebenfalls alle lückenlos erfaßt und nur nach Freigabe durch die Qualitätskontrolle ausgeliefert.

Das Konzept (bezogen auf alle Kontrollen) ist in Tabelle 2 schematisch dargestellt.

Tabelle 1. Bereiche, in denen mikrobiologische Kontrollen durchgeführt werden



Tabelle 2. Kontrollen zur Sicherung der Qualität aller durch Wander hergestellten Produkte



Für alle Produkte (Rohstoffe und fertige Produkte) wird seitens der Qualitätskontrolle ein Entscheid gefällt, wann Produkte freigegeben werden sollen und unter Inkaufnahme welcher Risiken und Konsequenzen! Es darf keine Rohstoffeingänge und keine Produktionen geben, von denen die Qualitätskontrolle nichts weiß.

Es ist für uns (nebst anderen Belangen) von vorrangiger Bedeutung, daß wir eine lückenlose Übersicht und Kontrolle über alle eingehenden Materialien und über alle fertigen Produkte haben.

Durchführung der Untersuchungen

Am Anfang der Kontrolltätigkeit steht (gemäß dem einfachen Konzept der «Rohstoff-Eingangskontrolle – Produktionskontrolle – Fertig-Produkt-Ausgangskontrolle») die Beurteilung der *Risiken* aller in der Produktion verwendeten Ausgangsmaterialien bezüglich ihrer Gefährdung für eine mikrobielle Kontamination sowie die Beurteilung der Risiken aller hergestellten und an den Handel auszuliefernden Produkte.

Kriterien für die Beurteilung der Risiken einer mikrobiellen Kontamination der Rohstoffe und fertigen Produkte

Die Kriterien für die Risikobeurteilung der Rohstoffe sind die folgenden:

Rohstoffe

- Herkunft der Rohstoffe (z. B. tierische, pflanzliche oder synthetische Produkte)
- Herstellung / Verarbeitung
- Möglichkeiten einer Keimvermehrung / antimikrobielle Eigenwirkung
- Erfahrungen (gemäß Erfahrungsaustausch, Publikationen und eigenen Erfahrungen).

Dieser letzte Punkt ist nahezu der wichtigste, da für viele Rohstoffe nur schwer abzuklären ist, welche Prozesse sie effektiv durchlaufen haben.

Rohstoffe, die gemäß diesen Gesichtspunkten nicht zuverlässig beurteilt werden können, werden bis zum Vorliegen genügender Erfahrungswerte der höchsten Gefährdungsklasse zugeteilt.

Entsprechend dieser Risikobeurteilung erfolgt die Einteilung aller Rohstoffe in verschiedene Klassen der Gefährdung für eine mikrobielle Kontamination:

Klasse I: Rohstoffe, die stark gefährdet sind

- Rohstoffe, die aufgrund der Herkunft oder Bearbeitung für eine starke mikrobielle Kontamination gefährdet sind.

- Rohstoffe, die gemäß Literatur, Erfahrungsaustausch und eigenen Erfahrungen stark gefährdet sind.
- Rohstoffe neuer Lieferanten und neue Rohstoffe, bis zum Vorliegen genügender Erfahrungswerte für die Beurteilung der Klassierung.

Beispiele: Milch und Milchprodukte, Eier und Eierprodukte, Fleisch und Fleischprodukte, Hefe und Hefeprodukte, Kakao und Kakaoprodukte, Gemüse und Früchte (Säfte, Extrakte, Konzentrate), Mehle, Gelier- und Verdickungsmittel natürlicher Herkunft, Gewürze, Kräuter u. a.

Klasse II: Rohstoffe, die wenig gefährdet sind

- Rohstoffe, die erfahrungsgemäß aufgrund der Art der Herstellung keimarm sind und die keine mikrobielle Vermehrung ermöglichen.
- Rohstoffe, die während der Herstellung einen ausreichend keimvermindern den Prozeß durchlaufen und anschließend weder mikrobiell rekontaminiert werden noch eine mikrobielle Vermehrung ermöglichen.

Beispiele: Fette, Öle, Aromastoffe, Konzentrate, Extrakte, Nüsse u. ä., Stärke, Vitamine u. a.

Klasse III: Rohstoffe die nicht gefährdet sind

- Rohstoffe, die eine ausreichende antimikrobielle Wirkung aufweisen.

Beispiele: Salze, Säuren, Zucker, Extrakte, synthetische Produkte (wie Konservierungsmittel, Aromastoffe, Farbstoffe, Antioxidantien, Vitamine, Emulgatoren) u. a.

Die Kriterien für die Beurteilung der Risiken einer mikrobiellen Kontamination von fertigen Produkten sind die folgenden:

Fertige Produkte

- Formel der Produkte
(beurteilt gemäß der Zuordnung der in den fertigen Produkten enthaltenen Rohstoffe zu den Gefährdungsklassen)
- Herstellungsverfahren
(Trockenmischung, Naßverfahren mit Pasteurisation, Kochen, Sterilisation usw.)
- Indikation
(Produkte für Kleinkinder, gesundheitlich beeinträchtigte Personen, ältere Personen, gesunde Kinder und Erwachsene usw.)
- Verwendung und Zubereitung zum Konsum
(Verwendung in Spitälern, Altersheimen, Pflegeheimen usw., genußfertig oder nicht genußfertig)
- Erfahrungswerte
(eigene Erfahrungen, Erfahrungen gemäß Publikationen und Erfahrungsaustausch).

Entsprechend dieser Risikobeurteilung erfolgt die Einteilung aller fertigen Produkte in verschiedenen Gefährdungsklassen:

Klasse I: Fertige Produkte, die stark gefährdet sind

Sie enthalten einen oder mehrere Rohstoffe der Klasse I trocken beigemischt oder sind aus anderen Gründen erfahrungsgemäß für eine mikrobielle Kontamination gefährdet.

- *Klasse Ia:* Fertige Produkte für besonders anfällige Personen mit hohem Risiko für eine Kontamination mit Salmonellen.
- *Klasse Ib:* Fertige Produkte für gesunde Kinder (außer Kleinkinder) und Erwachsene mit hohem Risiko für eine Kontamination mit Salmonellen.
- *Klasse Ic:* Fertige Produkte für gesunde Kinder und Erwachsene, die nur einen der gemäß separater Liste aufgeführten Rohstoffe der Klasse I enthalten sowie Produkte, die aus anderen Gründen erfahrungsgemäß für eine mikrobielle Kontamination (außer Salmonellen) gefährdet sind.

Klasse II: Fertige Produkte, die wenig gefährdet sind

Sie enthalten keine Rohstoffe der Klasse I oder haben einen ausreichenden, keimvermindernden Prozeß erfahren.

Eine Rekontamination, ferner eine Vermehrung noch vorhandener Keime und von Keimen einer Rekontamination ist sehr unwahrscheinlich.

- *Klasse IIa:* Fertige Produkte für besonders anfällige Personen.
- *Klasse IIb:* Fertige Produkte für gesunde Kinder und Erwachsene.

Klasse III: Konserven und Sterilprodukte, pasteurisierte Produkte, Backwaren ohne Füllung und weitere, auf mindestens 100 °C erhitzte Produkte, ferner andere Produkte

- *Klasse IIIa:* Fertige Produkte, die pasteurisiert oder UHT-behandelt und aseptisch abgefüllt worden sind.
- *Klasse IIIb:* Konserven und Sterilprodukte, die in den verschlossenen Originalpackungen hitzebehandelt worden sind, ferner Backwaren ohne Füllung.
- *Klasse IIIc:* andere Produkte.

Die Beurteilung der Gefährdung bezieht sich ganz allgemein und erfahrungsgemäß bei *unseren* Produkten, in abnehmender Bedeutung der Risiken dargestellt, auf

- Salmonellen
- coliforme Keime und Gesamtzahl Enterobakterien
- Schimmelpilze
- E. coli
- Gesamtzahl mesophile, aerobe Keime
- Staph. aureus
- andere.

Unser Hauptrisiko liegt demgemäß heute immer noch bei den Salmonellen, wogegen Staph. aureus für uns von völlig untergeordneter Bedeutung ist.

Die Gesamtzahl mesophiler, aerober Keime ist nur insofern ab und zu problematisch, als interne Limiten fallweise überschritten werden.

Diese Risikobeurteilung kann sich mit der Verschiebung der Produktpalette, des Produktsortiments, ändern.

Wir haben jetzt vorwiegend Trockenmischprodukte und getrocknete Produkte, führen aber auch schon UHT-behandelte, aseptisch abgefüllte Produkte und Konserven; bei den flüssigen Produkten liegt natürlich das Hauptrisiko nicht bei einer bestimmten Bakterien-Species, sondern bei Sterilität bzw. Unsterilität! Dies bedeutet eine völlig andere Bemusterung und auch Prüfung.

Bemusterung

Als nächster Schritt ist in der Praxis der mikrobiologischen Qualitätskontrolle der Umfang und die Häufigkeit der Bemusterung von Rohstoffen und fertigen Produkten, in Abhängigkeit der Gefährdung für eine mikrobielle Kontamination, festzulegen.

Häufigkeit der Bemusterung

Von den Rohstoffen und fertigen Produkten der *Klasse I* werden alle Eingänge bemustert.

Von den Rohstoffen und fertigen Produkten der *Klassen II und III* müssen (gemäß unseren eigenen Vorschriften) nicht alle Eingänge und Produktionen bemustert werden; die Beurteilung betreffend Bemusterung hängt ab von verschiedenen Faktoren wie Häufigkeit der Eingänge bzw. Produktionen, Konstanz der Befunde, Bedeutung des Produktes im Markt und für das Unternehmen. (Z. B. wird Ovomaltine trotz Zuordnung zu der Klasse II häufiger kontrolliert als es gemäß der Klassenzugehörigkeit erforderlich wäre.)

Umfang der Bemusterung

Rohstoffe

Besonders stark gefährdete Stoffe der Klasse I (z. Z. Hühnereiweißpulver) sowie die fertigen Produkte der Klasse Ia werden gemäß dem Foster-Plan bemustert, in Übereinstimmung mit der Risikobeurteilung (bezogen auf Salmonellen) gemäß FDA. Dies bedeutet, daß von jeder Lieferung von Rohstoffen bzw. von jeder Fabrikationseinheit bei fertigen Produkten 60 Muster zu erheben sind. Als Einheit für die Bemusterung fertiger Produkte gilt die Fabrikation; die Definition einer Fabrikation ist die folgende:

Eine Fabrikation umfaßt die nicht durch einen Produktwechsel (mit oder ohne Reinigung der Anlagen) oder durch ausgedehnte Reinigung (Naßreinigung und Desinfektion) unterbrochene Herstellung eines bestimmten Produktes; sie umfaßt eine Halbtages- oder Tagesproduktion, bei geringen Produktionsmengen jedoch maximal die Produktion einer Woche.

Die übrigen Rohstoffe der Klasse I werden gemäß einer (nicht statistisch erstellten) Tabelle bemustert, die in Anlehnung an bestehende Tabellen erstellt worden ist (siehe Anhang 1).

Von den Rohstoffen der Klasse II werden von z. B. jedem 2. bis jedem 5. Eingang Muster aus mindestens 5 Gebinden erhoben und als Mischmuster mikrobiologisch untersucht.

Fertige Produkte

Die fertigen Produkte der Klasse Ia werden (wie oben erwähnt) gemäß Foster-Plan bemustert.

Die fertigen Produkte der Klasse Ib werden entsprechend der Kategorie II gemäß FDA bzw. ebenfalls gemäß Foster bemustert; d. h. erheben von 30 Mustern pro Fabrikation, sofern alle 3 Risiken für eine Kontamination mit Salmonellen gegeben sind:

- Das Produkt ist potentiell gefährdet für eine Kontamination mit Salmonellen.
- Der Herstellprozeß des Produktes schließt keinen ausreichend keimvermindernden Prozeß ein.
- Die mißbräuchliche Anwendung oder Zubereitung des Produktes kann eine Vermehrung von Mikroorganismen zur Folge haben.

Die anderen Produkte der *Klasse Ib*, mit weniger als den drei Risiken für eine Kontamination mit Salmonellen, werden wiederum gemäß einer speziell (nicht statistisch) erstellten Tabelle bemustert (siehe Anhang 2).

Die Produkte aller übrigen Klassen werden gemäß einer festgelegten Regelung in geringerem Umfang bemustert.

Methoden der Untersuchung

Zur Erfüllung des Auftrages, wie eingangs formuliert (Sicherstellung der mikrobiologischen Qualität aller in den Handel gebrachten Produkte), ist natürlich entscheidend, daß die angewandten Untersuchungsmethoden zu zuverlässigen und aussagekräftigen Ergebnissen führen. Diesem Punkt muß, insbesondere beim Nachweis spezieller Keime, besondere Aufmerksamkeit geschenkt werden!

Der normale Umfang der mikrobiologischen Untersuchung von Rohstoffen und fertigen Produkten umfaßt die Prüfung auf die in Tabelle 3 aufgeführten Keime.

Die effektive Bewertung des Risikos, der Wahrscheinlichkeit für die Kontamination mit einer der erwähnten Species und die Bedeutung einer Kontamination hängen sehr stark von der Art der Rohstoffe sowie von der Formel und dem Herstellverfahren der fertigen Produkte ab! Diese in der Tabelle 3 angegebene Bewertung ist als Durchschnitt aus unserer Produkte-Palette zu verstehen. Für andere Produkte und andere Hersteller kann natürlich die Gewichtung der Risiken (wie schon darauf hingewiesen) ganz anders aussehen.

Tabelle 3. Bewertung der Wahrscheinlichkeit und Bedeutung einer Kontamination

Keimarten	Wahrscheinlichkeit einer Kontamination	Bedeutung einer Kontamination
Gesamtzahl aerobe, mesophile Keime	mittel	klein + I ¹
Schimmelpilze	mittel	groß + I
Hefen	klein	klein + I
Salmonellen	mittel	groß
E. coli	klein	mittel
Coliforme Keime	mittel	klein
Staph. aureus (Plasmakoagulase pos.)	klein	groß
Enterokokken (Streptokokken der Lancefield Gruppe D)	mittel	groß + I
Gesamtzahl Enterobakterien	mittel	klein + I

¹ I = Besondere Bedeutung bezüglich *Indikation* für fehlerhafte Produktionsprozesse oder mangelhafte Hygieneverhältnisse.

In speziellen Fällen werden zusätzliche Untersuchungen durchgeführt auf

- Bac. cereus
- Cl. perfringens
- Ps. aeruginosa
- (u. a.)

In anderen speziellen Fällen wird je nach Art der Risiken eine Reduktion der Untersuchungen vorgenommen; z. B. nur Bestimmung auf

- Gesamtkeimzahl
- Gesamtzahl Enterobakterien
- Staph. aureus

Für alle Untersuchungen werden standardisierte Methoden angewendet, die den jeweils neusten Ergebnissen gemäß Publikationen angepaßt werden.

Anforderungen

Genauso wie wir alle Rohstoffe und alle fertigen Produkte für die Klassierung gemäß der Gefährdung für eine mikrobielle Kontamination erfaßt haben, haben wir mikrobiologische Anforderungen an alle Rohstoffe und an alle fertigen Produkte beziehungsweise an Gruppen von Rohstoffen und fertigen Produkten formuliert.

Wir haben 2 Werte festgelegt, nämlich die Toleranzwerte (wie schon in der Einleitung erwähnt) mit interner Gültigkeit sowie die Grenzwerte, die niemals weniger streng als gesetzliche Anforderungen (sofern vorhanden) sind und die als Limite für die Freigabe beziehungsweise Sperrung gelten.

Beim Überschreiten der Toleranzwerte werden die Produkte intern beanstanden, die Ursachen für die erhöhte Kontamination gesucht und Maßnahmen zur Ausschaltung der Kontamination getroffen.

Beim Überschreiten der Grenzwerte erfolgt die Sperrung der Rohstoffe für die Verarbeitung und der fertigen Produkte für die Auslieferung.

Die Anforderungen an die Produkte der verschiedenen Produkte-Gruppen sind in den Tabellen 4 und 5 wiedergegeben.

Tabelle 4. Anforderungen an die mikrobiologische Qualität von Produkten

Keimarten	Genußfertige Produkte für besonders anfällige Personen		Genußfertige Produkte für gesunde Personen	
	Angestrebte Werte	Grenzwerte	Angestrebte Werte	Grenzwerte
Gesamtzahl aerobe, mesophile Keime in 1 g	<10 000	50 000	<100 000	500 000
Schimmelpilze in 1 g	< 100	1 000	< 100	1 000
Hefen in 1 g	< 100	5 000	< 1 000	---
Salmonellen in 50 g bis 1500 g	n. n.	n. n.	n. n.	n. n.
E. coli in 1 g	< 1	10	< 1	10
Coliforme Keime in 1 g	< 10	100	< 10	100
Gesamtkeimzahl Enterobakterien in 1 g	< 100	1 000	< 100	1 000
Staph. aureus in 1 g	< 10	100	< 10	100
Enterokokken in 1 g	< 1 000	---	< 10 000	---
Bacillus cereus in 1 g	< 100	1 000	< 100	1 000
Cl. perfringens in 1 g	< 10	100	< 10	100
Ps. aeruginosa in 1 g	< 1	10	---	---

In diesem Sinne versuchen wir, in der täglichen Praxis der mikrobiologischen Qualitätskontrolle, die Ergebnisse, Entscheide und Maßnahmen in Übereinstimmung zu bringen mit den eingangs erwähnten Grundsätzen und hoffen dadurch die Zielsetzung, daß nur Produkte in den Handel gelangen, die den mikrobiologischen Anforderungen entsprechen, realisieren zu können.

Tabelle 5. Anforderungen an die mikrobiologische Qualität von Produkten

Keimarten	Nicht genußfertige Produkte für besonders anfällige Personen		Nicht genußfertige Produkte für gesunde Personen	
	Angestrebte Werte	Grenzwerte	Angestrebte Werte	Grenzwerte
Gesamtzahl aerobe, mesophile Keime in 1 g	<50 000	100 000	<100 000	1 Mio
Schimmelpilze in 1 g	< 100	1 000	< 1 000	5 000
Hefen in 1 g	< 1 000	---	---	---
Salmonellen in 50 g bis 1500 g	n. n.	n. n.	n. n.	n. n.
E. coli in 1 g	< 10	100	< 1 000	10 000
Coliforme Keime in 1 g	< 100	1 000	< 1 000	---
Gesamtzahl Enterobakterien in 1 g	< 100	1 000	< 10 000	---
Staph. aureus in 1 g	< 100	1 000	< 1 000	10 000
Enterokokken in 1 g	<10 000	---	---	---
Bacillus cereus in 1 g	< 100	1 000	< 1 000	10 000
Cl. perfringens in 1 g	< 100	1 000	< 1 000	10 000
Ps. aeruginosa in 1 g	---	---	---	---

Zusammenfassung

Zur Begründung und Erreichung der Ziele der mikrobiologischen Qualitätskontrolle (d. i. die ausreichende Sicherung der mikrobiologischen Qualität aller in den Handel gebrachten Produkte) werden Grundsätze (Qualitätspolitik) definiert.

Alle Rohstoffe und fertigen Produkte werden zur Festlegung der erforderlichen Qualitätskontrollen in Gefährdungsklassen für eine mikrobielle Kontamination eingeteilt.

Die Frequenz und der Umfang der Bemusterung sowie die Methoden der Untersuchung und die Anforderungen werden gemäß der Gesetzgebung, internationalen Richtlinien und Empfehlungen, ferner gemäß eigenen Erfahrungen und Publikationen festgelegt.

Die Freigabe oder Sperrung von Rohstoffen für die Verarbeitung und von fertigen Produkten für die Auslieferung erfolgt in Abhängigkeit der Übereinstimmung von Untersuchungsergebnissen mit den Anforderungen.

Résumé

Des principes fondamentaux (politique de qualité) démontrant la nécessité du contrôle de la qualité microbiologique de tous les produits alimentaires mis dans le commerce par l'industrie sont définis et les moyens d'atteindre les objectifs fixés sont décrits.

Dans ce but, les matières premières et les produits finis sont répartis en catégories correspondant chacune aux risques de contamination microbienne.

La fréquence et la taille de l'échantillonnage, tout comme les méthodes d'analyse et les exigences de qualité microbiologique sont conditionnées par la législation, les recommandations et directives internationales, ainsi que par les propres expériences et publications de l'entreprise.

Le déblocage ou le blocage des matières premières pour la fabrication et celui de la livraison des produits finis sont décidés selon que les résultats des analyses répondent ou non aux exigences de qualité microbiologique.

Summary

Principles (qualitative policy) are defined in order to establish and to achieve the aims of microbiological quality control (i. e. to ensure adequately the microbiological quality of all products put on the market).

In order to determine the necessary quality controls, all raw materials and finished products are classified into categories of microbiological contamination risk.

The frequency and the extent of sampling, as well as the methods of examination and the requirements, are stipulated in accordance with legislation, international guidelines and recommendations, together with our own experience and publications.

The release or blocking of raw materials for processing and of finished products for delivery is decided on the basis of the conformance or otherwise of the examination results with the requirements.

Literatur

1. Eidgenössisches Departement des Innern: Verordnung über die hygienisch-mikrobiologischen Anforderungen an Lebensmitteln, Gebrauchs- und Verbrauchsgegenständen, 19. Januar 1983.
2. Bacteriological Analytical Manual for Foods. Food and Drug Administration, Bureau of Foods, Division Microbiology, AOAC, August 1978.
3. Microorganisms in Foods. JCMSC University of Toronto Press 1974.
4. Refai, M. K.: Manuals of food quality control, 4. Microbiological analysis FAO, Rome 1979.
5. Herschdoerfer, S. M.: Quality control in the food industry, Vol. 1 and 3. Academic Press London and New York 1967 and 1972.
6. Bozyk, Z. und Rudzki, W.: Qualitätskontrolle von Lebensmitteln. VEB Fachbuchverlag, Leipzig 1971.
7. Kramer, A. and Twigg, A.: Quality control for the food industry. AVI Publishing Company, Westport, Conn. 1973.
8. Thorner, M. E. and Manning, P. B.: Quality control in Food service. AVI Publishing Company, Westport, Conn. 1976.
9. Puri, S. C., Ennis, D. and Mullen K.: Statistical quality control for food and agricultural scientists. G. K. Hall and Co, Boston 1979.
10. Masing, W.: Handbuch der Qualitätssicherung. C. Hauser Verlag, München 1980.
11. Howard, R.: Food safety. J. Wiley and Sons, New York 1981.

Anhang 1. Tabelle der Bemusterung von Rohstoffen
 Umfang der Bemusterung von Rohstoffen für mikrobiologische
 Qualitätskontrolle

Anzahl Packungen pro Lieferung bzw. Produktionseinheit	Anzahl Muster, zufällig verteilt über die ganze Lieferung bzw. Produktionseinheit entnommen
1	1
2 — 4	2
5 — 8	3
9 — 20	4
21 — 30	5
31 — 40	6
41 — 50	7
51 — 70	8
71 — 100	9
101 — 200	12
201 — 1 000	15
1 001 — 2 000	20
2 001 — 5 000	25
5 001 — 10 000	30
über 10 000	40

Anhang 2. Tabelle der Bemusterung von fertigen Produkten
 Umfang der Bemusterung von fertigen Produkten der *Klasse Ib* («übri-
 ge Produkte») für die mikrobiologische Qualitätskontrolle

Anzahl Packungen pro Fabrikation	Anzahl Muster, gleichmäßig verteilt über die ganze Fabrikation
1 — 4	1
5 — 20	2
21 — 50	3
51 — 100	4
101 — 500	5
501 — 1 000	10
über 1 000	15

Dr. W. Hauert
 Wander AG Bern
 Leiter Qualitätskontrolle Ernährung
 CH-3176 Neuenegg

Allgemeine statistische Überlegungen zur Aussagekraft von Analysenresultaten

Some General Statistical Considerations Concerning the Quality of Analytical Results

H. R. Roth

Institut für Tierproduktion, ETH, Zürich

Einleitung

Wir stoßen laufend auf die Darstellung von Analysenresultaten in der Form

$$22,661 \pm 0,385 \quad (n = 6)$$

wobei 22,661 offenbar den Mittelwert aus $n = 6$ Messungen darstellt und 0,385 dem Leser eine Vorstellung über die Genauigkeit geben soll. Ich will nun die Behauptung wagen, daß in der Praxis eine Darstellung in obiger Form sehr oft irreführend ist. In meinem Vortrag möchte ich aufzeigen, unter welchen Voraussetzungen eine vernünftige Aussage über die Vertrauenswürdigkeit von Analysenresultaten gemacht werden kann.

Absolut nötig ist das Verfolgen des gesamten Werdeganges eines Analysenresultats. Am Anfang steht ein Los, über dessen Qualität etwas ausgesagt werden muß. Daß sich einzelne Lose in ihrer Dimension, ihrem Material und in ihrer Qualitätseigenschaft stark unterscheiden können, geht aus der folgenden Aufstellung hervor:

Los	«Qualität»
Schiffsladung mit Kohle	Verbrennungswert, Aschegehalt
Lastwagen mit Obst	Cu-Rückstand
200 Düngersäcke	N-Gehalt
Honigdosen im Verkaufsregal	Nettogewicht
Brote in Bäckerei	Rohfasergehalt
Wasser in einem Teich	Keimzahl, Temperatur
Luft in der Bahnhofstraße	SO ₂ -Gehalt
1 t Saatgut	Anzahl Unkrautsamen pro 100 g

Das Material kann aus eindeutig erkennbaren Einzelteilen bestehen oder eine unstrukturierte Masse bilden. In beiden Fällen kann eine bereits erfolgte Gruppierung des Materials, zum Beispiel in Form einer Verpackung, vorliegen.

Das interessierende Merkmal kann eine qualitative Eigenschaft des Materials betreffen, welche eindeutig bestimmbar ist. Die Vertrauenswürdigkeit eines entsprechenden Befundes kann in dieser Situation relativ einfach beurteilt werden. Die entsprechende statistische Literatur ist sehr umfassend.

Im folgenden wollen wir uns der wesentlich komplexeren Situation zuwenden, wo das interessierende Merkmal eine quantitative Eigenschaft des Materials betrifft, welche mit einem Meßverfahren mehr oder weniger genau erfaßt werden kann.

Meist sind wir an der mittleren Qualität eines Loses interessiert und nicht direkt an der Variabilität innerhalb des Loses. Sobald wir jedoch eine Aussage über die Vertrauenswürdigkeit eines Analysenresultates haben möchten, sind wir auf die Kenntnis der Art und des Ausmaßes der Streuungsursachen angewiesen. Nur so lassen sich Qualitätskontrollen mit vorgegebenem Produzenten- und Konsumentenrisiko durchführen oder gemessene Abweichungen von einem Sollwert objektiv beurteilen (1-3).

Das einfachste statistische Modell eines Analysenresultats

Wird dem Los eine Probe entnommen und im Laboratorium analysiert, dann setzt sich der Meßwert aus drei Komponenten zusammen:

$$y = Y + \lambda + m$$

wobei y = Meßwert

Y = unbekannte durchschnittliche Qualität des Loses

λ = unbekannte zufällige Abweichung der Laborprobe von Y

m = unbekannte zufällige Abweichung des Meßwertes vom richtigen Wert der Laborprobe.

Der Betrag von λ hängt von der Inhomogenität des Loses ab, welche durch die Varianz σ_I^2 charakterisiert werden kann. Die Varianz von λ muß bei zufälliger Probenahme mit σ_I^2 übereinstimmen.

Den Meßfehler m wollen wir vorläufig nicht in einzelne Komponenten zerlegen. Die Varianz von m bezeichnen wir mit σ_M^2 . Sofern der Meßfehler unabhängig ist vom wahren Wert der Laborprobe, so setzt sich die Varianz des Meßwertes y additiv aus den Komponenten σ_I^2 und σ_M^2 zusammen:

$$\sigma_y^2 = \sigma_I^2 + \sigma_M^2$$

Der Meßwert y kann unter Umständen eine riesige Variabilität (Unsicherheit) aufweisen und ist somit kaum repräsentativ für die Qualität des Loses. Leider gibt uns die eine Messung überhaupt keinen Anhaltspunkt über die Größenordnung der Varianzen σ_y^2 , σ_I^2 und σ_M^2 .

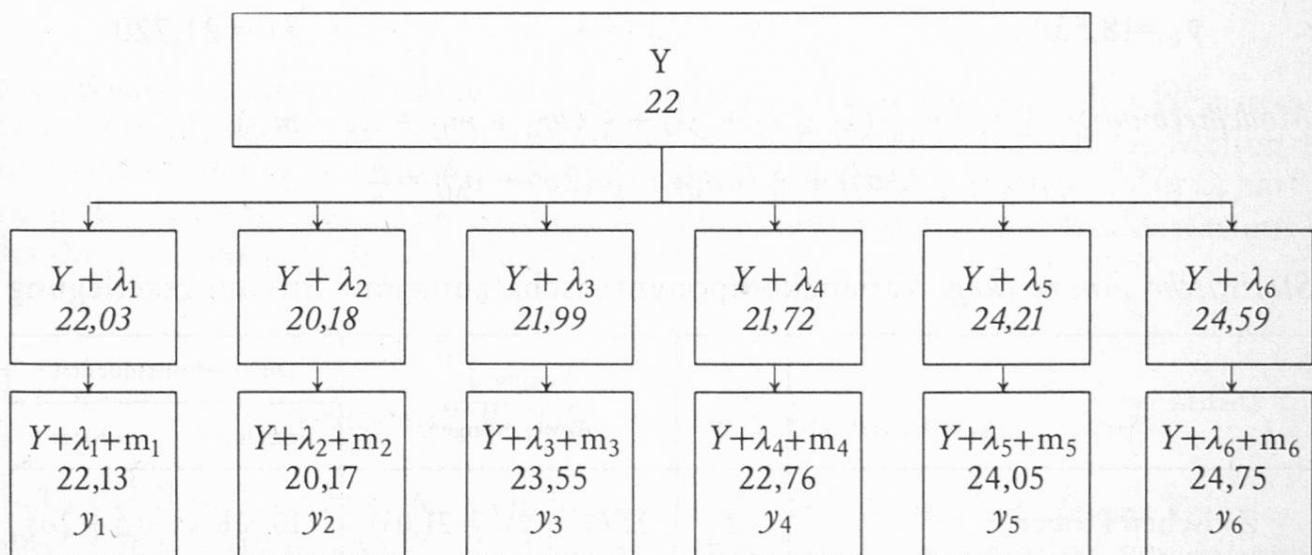
Die folgenden Maßnahmen können zur Steigerung der Aussagefähigkeit des Analysenresultates beitragen:

1. Reduktion von σ_I^2 durch Mischen innerhalb des Loses. Dies ist nicht immer durchführbar und auch nicht immer sinnvoll. Fragwürdig ist ein Mischen dann, wenn innerhalb des Loses eine deutliche Unterteilung in Schichten erkennbar ist, zum Beispiel hervorgerufen durch verschiedene Herkunft einzelner Teile oder durch ein Material, welches selbst zur Schichtbildung neigt. Weisen nämlich die einzelnen Schichten unterschiedliche Qualität oder eine unterschiedliche Heterogenität auf, dann sollte man ein Mischen unterlassen, dafür aber einen entsprechenden Stichprobenplan anwenden.
2. Reduktion von σ_M^2 durch Verbesserung der Meßmethode.
3. Die Zahl der Laborproben erhöhen.
4. Mehrere Analysen pro Laborprobe durchführen.

Die Maßnahmen 3 und 4 führen zu mehreren Meßwerten, welche zu einem Mittelwert \bar{y} zusammenzufassen sind.

Welche Auswirkungen das Stichprobenverfahren auf die Unsicherheit eines berechneten Mittelwertes haben kann, soll anhand von drei möglichen Varianten demonstriert werden. Obwohl in allen Varianten der analytische Aufwand identisch ist, ergeben sich Unterschiede. Die Modellrechnung gibt jeweils ein genaues Bild über die zu erwartende Varianz $\sigma_{\bar{y}}^2$ des Mittelwertes. Die aus den Meßwerten berechnete Größe $s_{\bar{y}}$ heißt Standardfehler. Das Quadrat $(s_{\bar{y}})^2$ ist ein Schätzwert für $\sigma_{\bar{y}}^2$. Simulierte Zahlenwerte sollen zum besseren Verständnis der statistischen Auswertung beitragen. Kursiv gedruckte Zahlen sind hypothetische, für den Analysator aber unbekannte Werte.

Simuliertes Beispiel, Variante A mit 6 einfach analysierten Laborproben ($\sigma_I^2 = 4$, $\sigma_M^2 = 1$)



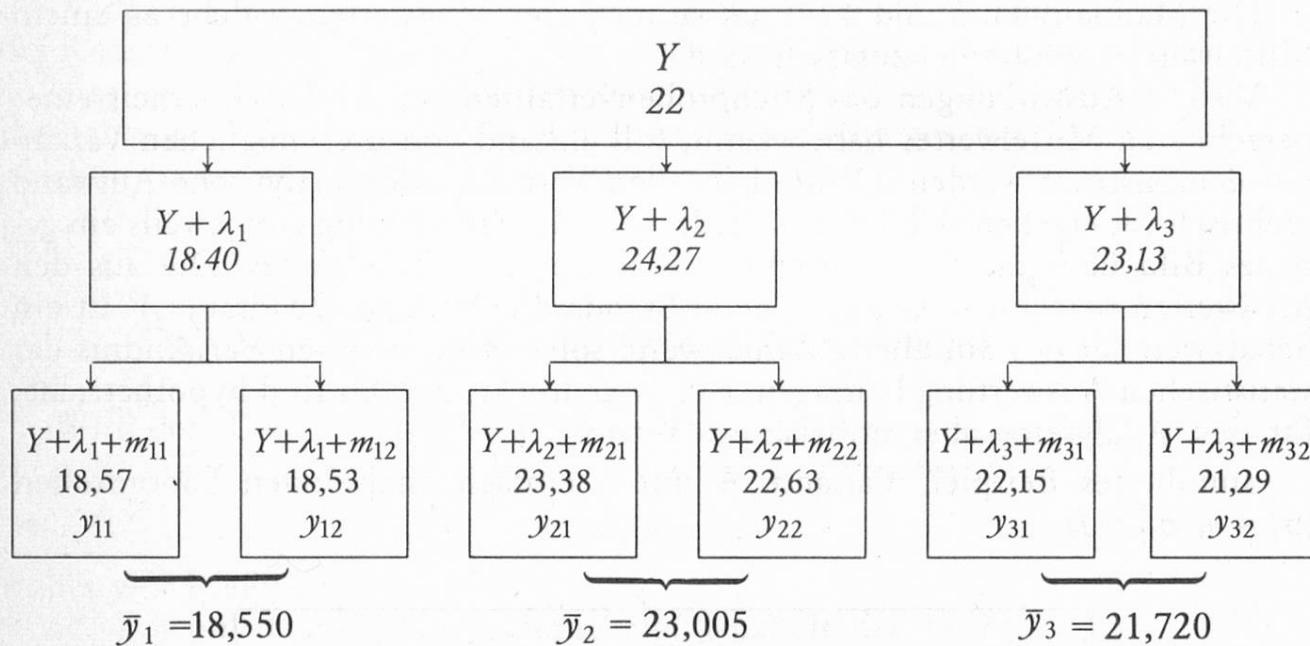
$$\text{Modellrechnung: } \bar{y} = Y + \frac{1}{6} (\lambda_1 + \lambda_2 + \dots + \lambda_6) + \frac{1}{6} (m_1 + m_2 + \dots + m_6)$$

$$\sigma_{\bar{y}}^2 = \frac{1}{36} (6\sigma_I^2) + \frac{1}{36} (6\sigma_M^2) = \frac{1}{6} (\sigma_I^2 + \sigma_M^2) = \frac{s^2}{6}$$

$$\left. \begin{array}{l} \text{Statistische Auswertung: } \bar{y} = 22,902 \\ s = 1,627 \\ s_{\bar{y}} = s/\sqrt{6} = 0,664 \end{array} \right\} Y = 22,902 \pm 0,664$$

Kommentar: Es gibt hier keine Möglichkeit für die Schätzung der einzelnen Varianzkomponenten σ_I^2 und σ_M^2 . Nur die Summe $\sigma_I^2 + \sigma_M^2$ kann mit der berechneten Streuung s^2 geschätzt werden oder $\sqrt{\sigma_I^2 + \sigma_M^2}$ mit der Standardabweichung s . Dafür liefert diese Variante den Mittelwert \bar{y} mit der kleinst möglichen Streuung $s_{\bar{y}}$. In der folgenden Variante lassen sich die einzelnen Varianzkomponenten schätzen. Der Mittelwert \bar{y} wird jedoch eine größere Unsicherheit aufweisen.

Simuliertes Beispiel, Variante B mit 3 doppelt analysierten Laborproben ($\sigma_I^2 = 4$, $\sigma_M^2 = 1$).



$$\begin{aligned} \text{Modellrechnung: } \bar{y} &= Y + \frac{2}{6} (\lambda_1 + \lambda_2 + \lambda_3) + \frac{1}{6} (m_{11} + m_{12} + \dots + m_{32}) \\ \sigma_{\bar{y}}^2 &= \frac{4}{36} (3\sigma_I^2) + \frac{1}{36} (6\sigma_M^2) = \frac{1}{6} (2\sigma_I^2 + \sigma_M^2) = \frac{2}{6} \end{aligned}$$

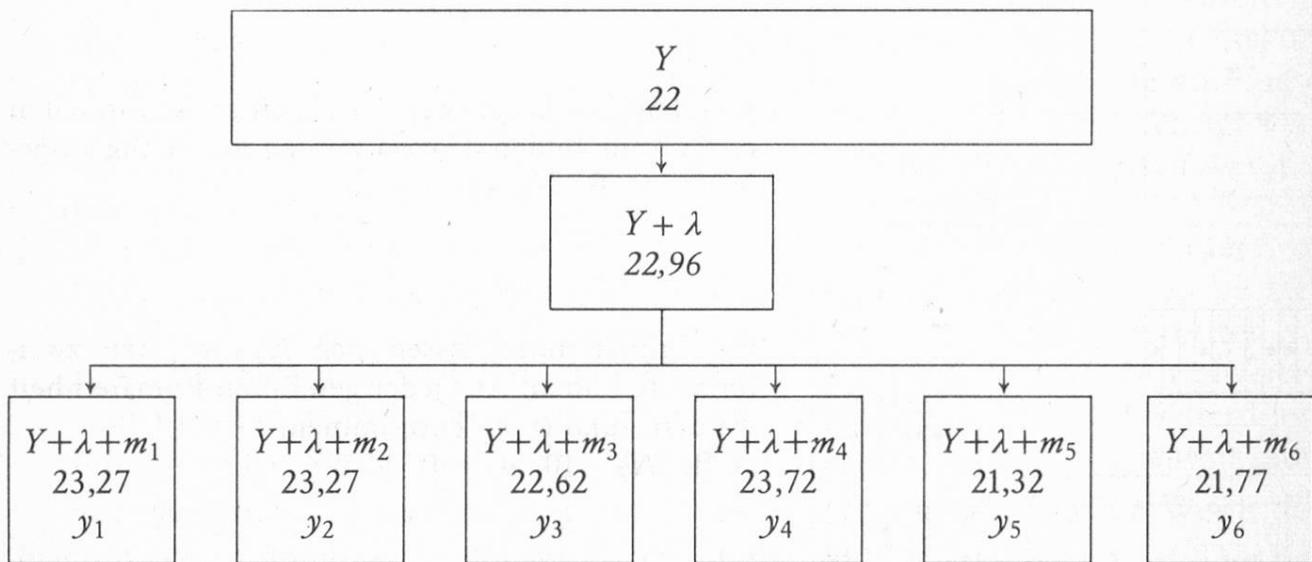
Statistische Auswertung: Varianzkomponentenschätzung mit Streuungszerlegung

Ursache	FG	Summe der quadrierten Abweichungen	Durchschnittsquadrat	
			berechnet	schätzt
Zwischen Proben	2	$2\sum_i (\bar{y}_i - \bar{y})^2 = 21,031$	10,516	$\sigma_M^2 + 2\sigma_I^2$
Innerhalb Proben	3	$\sum_{i,j} (y_{ij} - \bar{y}_i)^2 = 0,652$	0,217	σ_M^2
Total	5	$\sum_{i,j} (y_{ij} - \bar{y})^2 = 21,683$		

$$\begin{aligned} \bar{y} &= 21,092 \\ s_{\bar{y}} &= \sqrt{\frac{1}{6} (2\hat{\sigma}_I^2 + \hat{\sigma}_M^2)} = \sqrt{\frac{1}{6} (10,516)} = 1,324 \end{aligned} \quad \left. \right\} \quad Y = 21,092 \pm 1,324$$

geschätzte Varianzkomponenten: $\hat{\sigma}_M^2 = 0,217 \cdot \hat{\sigma}_I^2 = 5,150$

Simuliertes Beispiel, Variante C mit einer 6fach analysierten Laborprobe ($\sigma_I^2 = 4$, $\sigma_M^2 = 1$)



$$\text{Modellrechnung: } \bar{y} = Y + \lambda + \frac{1}{6} (m_1 + m_2 + \dots + m_6)$$

$$\sigma_{\bar{y}}^2 = \sigma_I^2 + \frac{1}{36} (6\sigma_M^2) = \sigma_I^2 + \frac{1}{6} \sigma_M^2 = \frac{25}{6}$$

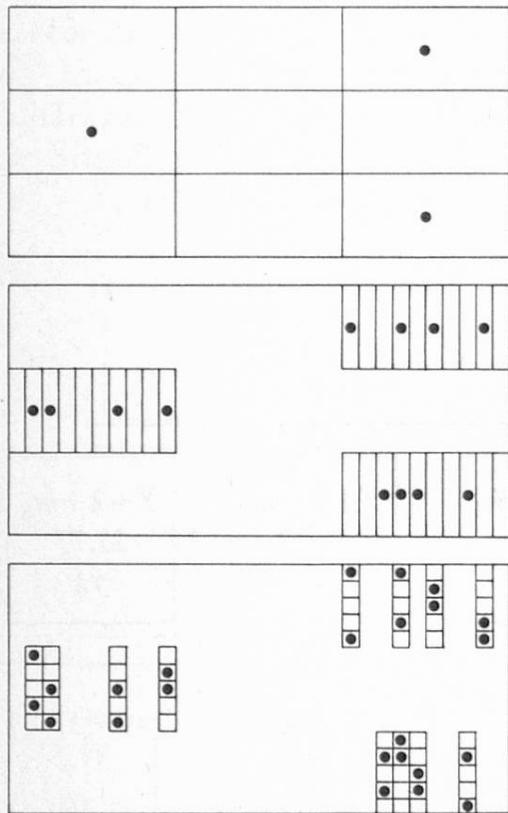
$$\text{Statistische Auswertung: } \bar{y} = 22,661 \quad \left. \right\} \quad s = 0,944 \quad s_{\bar{y}} = 0,385 \quad Y = 22,661 \pm 0,385$$

Kommentar: In dieser Variante haben wir auf rechnerischem Weg den kleinsten Standardfehler $s_{\bar{y}}$ ermittelt. Diese Zahl bezieht sich aber hier nur auf die Meßgenauigkeit und verleitet somit zu völlig falschen Schlußfolgerungen. Zur Schätzung der Qualität des Loses ist diese Variante immer die ungünstigste, was aus der Modellrechnung für $\sigma_{\bar{y}}^2$ klar hervorgeht.

Schrittweise Reduktion des Probenmaterials

In der Regel wird bei Massengütern nur ein verschwindend kleiner Teil des Loses einer Prüfung unterzogen. Wer sich diese Größenverhältnisse vor Augen hält, stellt unwillkürlich die Frage nach der Aussagefähigkeit eines Analysenresults. Auf keinen Fall darf man sich von der im Laboratorium erreichbaren hohen Analysengenaugkeit blenden lassen, denn die Art der Probengewinnung, die Anzahl und die Größe der Proben und ihre Reduktion sind ebenfalls wichtig.

Die Reduktion des Materials erfolgt meist in mehreren Stufen eventuell parallel mit einem Zerkleinerungsprozeß. Für jede Stufe gilt es Entscheidungen zu treffen, welche von der Variabilität des Materials in der jeweiligen Stufe, von den Kosten und von der geforderten statistischen «Sicherheit» der Aussage abhängen. Bei der stufenweisen Reduktion wählt man zweckmäßigerweise ein systematisches Vorgehen, wie es in Abbildung 1 angedeutet wird.



Aus dem Los lassen sich N_1 gleich große Einheiten erster Stufe bilden. Davon werden n_1 zufällig ausgewählt. (z. B. $N_1 = 9$, $n_1 = 3$)

Pro Primäreinheit lassen sich N_2 -Einheiten zweiter Stufe bilden. Aus jeder gewählten Primäreinheit werden zufällig n_2 entnommen.
(z. B. $N_2 = 10$, $n_2 = 4$)

Pro Sekundäreinheit lassen sich N_3 Einheiten dritter Stufe bilden. Aus jeder gewählten Sekundäreinheit werden zufällig n_3 entnommen.
(z. B. $N_3 = 5$, $n_3 = 2$)

Abb. 1. Vorgehen bei der stufenweisen Reduktion des Probenmaterials

Die Aufteilung kann analog bis zu einer beliebigen Stufenzahl fortgesetzt werden. Wir wollen jedoch annehmen, daß die eigentlichen Messungen an den Einheiten der dritten Stufe durchgeführt werden.

Stufe	Anzahl der vorhandenen Einheiten	Davon ausgewählt	Streuung zwischen den Einheiten
0	1	1	—
1	N_1	n_1	σ_1^2
2	$N_1 N_2$	$n_1 n_2$	σ_2^2
3	$N_1 N_2 N_3$	$n = n_1 n_2 n_3$	σ_3^2

Mit y_{ijk} bezeichnen wir den Meßwert an der k -ten Einheit dritter Stufe innerhalb der j -ten Sekundäreinheit innerhalb der i -ten Primäreinheit. Die Herkunft jeder Probe muß also festgehalten werden. Die mittlere Qualität des Loses schätzen wir mit dem Mittelwert aller Meßwerte

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} y_{ijk}$$

Nach Stange (4, 5) gehört zu diesem Schätzwert die Varianz

$$\sigma_{\bar{y}}^2 = \frac{1}{n_1} \left(1 - \frac{n_1}{N_1}\right) \sigma_1^2 + \frac{1}{n_1 n_2} \left(1 - \frac{n_2}{N_2}\right) \sigma_2^2 + \frac{1}{n_1 n_2 n_3} \left(1 - \frac{n_3}{N_3}\right) \sigma_3^2$$

Wir wollen aus dieser Formel einige wichtige Schlußfolgerungen ableiten:

1. Die in Klammern gesetzten Ausdrücke bezeichnet man als Endlichkeitskorrekturen. Sie geben den nicht geprüften Anteil der jeweiligen Stufe an. Diese Faktoren müssen in der Praxis meist nur für die erste Stufe berücksichtigt werden, da in den höheren Stufen die Zahl der vorhandenen Einheiten wesentlich größer ist als die Zahl der ausgewählten Einheiten.
2. Die Streuung jeder Stufe wird im Verhältnis zur Gesamtzahl der in dieser Stufe durchgeführten Beobachtungen herabgesetzt. Es lohnt sich also, die Gesamtzahl der Beobachtungen $n = n_1 n_2 n_3$ so in die einzelnen Häufigkeiten n_1 , n_2 und n_3 aufzuteilen, daß die zu erwartende Ungenauigkeit des Mittelwertes \bar{y} in Relation zum erforderlichen Aufwand möglichst klein wird. Allerdings muß dazu das Verhältnis der einzelnen Streuungen σ_1^2 , σ_2^2 und σ_3^2 bekannt sein.

Sofern man es nicht mit Routineproblemen zu tun hat, sind die Werte der Streuungen σ_1^2 , σ_2^2 und σ_3^2 unbekannt, so daß der Wert $\sigma_{\bar{y}}^2$ zu schätzen ist. Die Berechnung der Streuung der Einzelwerte nach der Formel

$$s^2 = \frac{1}{n-1} \sum_i \sum_j \sum_k (y_{ijk} - \bar{y})^2$$

hilft uns dabei sehr wenig. Wir berechnen zuerst die Mittelwerte auf allen Stufen. Zuerst die $n_1 n_2$ Mittelwerte auf der zweiten Stufe

$$\bar{y}_{ij} = \frac{1}{n_3} \sum_{k=1}^{n_3} y_{ijk}$$

und daraus die n_1 Mittelwerte erster Stufe

$$\bar{y}_i = \frac{1}{n_2} \sum_{j=1}^{n_2} \bar{y}_{ij}$$

Die Summen der quadrierten Abweichungen

$$\left. \begin{array}{l} S_1 = n_2 n_3 \sum_{i=1}^{n_1} (\bar{y}_i - \bar{y})^2 \\ S_2 = n_3 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\bar{y}_{ij} - \bar{y}_i)^2 \\ S_3 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} (y_{ijk} - \bar{y}_{ij})^2 \end{array} \right\} S = S_1 + S_2 + S_3 = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y})^2$$

werden in der bekannten Art zusammengestellt:

Stufe	Summe der Quadrate	Anzahl Freiheitsgrade	Quotient	Der Quotient ist ein Schätzwert für
1	S_1	$n_1 - 1$	$S_1 / (n_1 - 1)$	$\sigma_3^2 + n_3 \sigma_2^2 + n_3 n_2 \sigma_1^2$
2	S_2	$n_1 (n_2 - 1)$	$S_2 / [n_1 (n_2 - 1)]$	$\sigma_3^2 + n_3 \sigma_2^2$
3	S_3	$n_1 n_2 (n_3 - 1)$	$S_3 / [n_1 n_2 (n_3 - 1)]$	σ_3^2
Total	S	$n_1 n_2 n_3 - 1$	—	—

Sofern die Endlichkeitskorrekturen vernachlässigt werden können, vereinfacht sich die Formel für $\sigma_{\bar{y}}^2$ erheblich. In diesem Fall liefert der Quotient $S_1 / [(n_1 - 1) n_1 n_2 n_3]$ eine erwartungstreue Schätzung für die Varianz $\sigma_{\bar{y}}^2$ des Mittelwertes \bar{y} mit $n_1 - 1$ Freiheitsgraden.

Aus der Zahl der Freiheitsgrade kann entnommen werden, daß mindestens 2 Untereinheiten pro Obereinheit gewählt werden müssen, damit alle Streuungskomponenten geschätzt werden können. Bei mehr als drei Stufen, was in der Praxis sehr bald benötigt wird, steigt der Analysenaufwand ganz erheblich an. Kann hier die Bildung von Mischproben eine vernünftige Abhilfe schaffen?

Die Bildung von Mischproben

Der Analysenaufwand lässt sich unter Umständen stark einschränken, wenn mehrere Proben derselben Stufe zu einer Mischprobe zusammengefaßt werden. Sofern der Mischvorgang zu homogenen Mischproben führt, genügen dann einige Laborproben (welche eventuell mehrfach analysiert werden) für eine Schätzung der mittleren Qualität. Während diese Technik unbestreitbare Vorteile aufweist, verführt sie doch oft zu Genauigkeitsaussagen, welche als zu optimistisch bezeichnet werden müssen.

Der besonders häufige Fall, wo nur eine Mischprobe hergestellt wird, ist aus statistischer Sicht oft problematisch. Das Analysenresultat kann durchaus sehr

genau sein, aber eine Abschätzung der Genauigkeit ist nur möglich, wenn die einzelnen Streuungskomponenten bereits aus früheren Untersuchungen bekannt sind. Die Information über die Variabilität innerhalb des Loses geht nämlich verloren. Wir können mit anderen Worten nicht entscheiden, ob die Mischprobe repräsentativ ist oder nicht. Dieses Verfahren ist deshalb nur bei Routineüberwachungen mit bekannten Streuungskomponenten zu empfehlen.

Wenn man eine Aussage über die Variabilität eines Analysenresultats haben möchte und über keine Vorinformation verfügt, dann muß man mindestens 2 gleichwertige Mischproben herstellen. Andernfalls wird die Aussage immer nur für die Mischprobe Gültigkeit haben und nicht für das Los.

Die theoretische Ableitung der Eigenschaft von Schätzwerten, welche aus Mischproben bestimmt worden sind, ist recht aufwendig. Der klassische Ansatz wird von *Cameron* (6) und *Duncan* (7, 8) aufgezeigt. Untersuchungen von *Elder et al.* (9) basieren auf komplizierteren Modellen, bringen aber für die Praxis keine wesentlich neuen Entscheidungshilfen.

Die Berechnung von Vertrauensintervallen

Ein Vertrauensintervall wird so konstruiert, daß es den unbekannten Wert Y mit einer Wahrscheinlichkeit von $1-\alpha$ einschließt. Bei einer vorgegebenen statistischen Sicherheit $1-\alpha$ ist man natürlich an möglichst kurzen Intervallen interessiert. Für normalverteilte Mittelwerte \bar{y} berechnet man die Grenzen nach der Formel

$$\bar{y} \pm t_{\alpha/2;v} \cdot s_{\bar{y}}$$

wobei $s_{\bar{y}}$ der aus den Messungen berechnete Standardfehler des Mittelwertes \bar{y} ist. Der Zahlenfaktor $t_{\alpha/2;v}$ berücksichtigt neben der geforderten Sicherheit die Tatsache, daß $s_{\bar{y}}$ nur mehr oder weniger gut mit dem wahren Wert $\sigma_{\bar{y}}$ übereinstimmt. Die Zahl v ist der Freiheitsgrad, welcher zum Standardfehler $s_{\bar{y}}$ gehört und beim soeben betrachteten hierarchischen Modell den Wert $n_1 - 1$ aufweist.

Die Länge des Intervalls richtet sich somit nach dem Standardfehler $s_{\bar{y}}$ und nach dem t -Wert, welcher bei kleinem Freiheitsgrad $v = n_1 - 1$ stark ins Gewicht fallen kann. Aufgrund dieser Überlegungen sollten mindestens drei Primäreinheiten dem Los entnommen oder mindestens drei Mischproben hergestellt werden.

Zum simulierten Beispiel berechnen wir nun noch die 95% Vertrauensintervalle.

$$\text{Variante A: } Y = \bar{y} \pm t_{0,025;5} \cdot s_{\bar{y}} = 22,902 \pm 2,571 \cdot 0,664 \Leftrightarrow \begin{matrix} 24,609 \\ 21,195 \end{matrix}$$

$$\text{Variante B: } Y = \bar{y} \pm t_{0,025;2} \cdot s_{\bar{y}} = 21,092 \pm 4,303 \cdot 1,324 \Leftrightarrow \begin{matrix} 26,789 \\ 15,395 \end{matrix}$$

Variante C: Keine Aussage möglich

Zusammenfassung

Anhand von fiktiven Beispielen wird aufgezeigt, daß die Aussagekraft von Analysenresultaten nicht nur von der im Laboratorium erzielbaren Meßgenauigkeit, sondern oft er-

heblich von der Art der Probengewinnung abhängt. Probleme bei der Verwendung von Mischproben werden kurz diskutiert.

Résumé

A l'aide d'exemples fictifs nous démontrons que la qualité de résultats d'analyses ne peut être attribuée uniquement à la précision des mesures effectuées dans les laboratoires. Souvent nous constatons que celle-ci dépend considérablement de la manière dont les échantillons ont été récoltés. Une brève discussion touche aux problèmes posés par l'utilisation d'échantillons moyens (mélanges).

Summary

The discussion of some fictive examples shows that the quality of analytical results is not only dependent on the precision reached in the laboratories, but is sometimes considerably influenced by the sampling process. A short discussion is given on the problems arising when composite samples are used.

Literatur

1. *Bingham, R. S., Jr.*: Bulk Sampling – A common sense viewpoint. *Tappi* **46**, 288–296 (1963).
2. *Kolthoff, I. M.* and *Elwing, P. J.*, (eds.): Treatise on analytical chemistry. Part I, Vol. 1, 2nd ed. Wiley, New York 1978. Chapter 4 by *Currie, L. A.*: Sources of error and the approach to accuracy in analytical chemistry, pp. 95–242. Chapter 5 by *Mandel, J.*: Accuracy and precision: Evaluation and interpretation of analytical results, pp. 243–298. Chapter 6 by *Bicking, C. A.*: Principles and methods of sampling, pp. 299–359.
3. *Youden, W. J.*: The sample, the procedure, and the laboratory. *Anal. Chem.* **32** (13), 23A–36A (1960).
4. *Stange, K.*: Stichprobenverfahren bei der Beurteilung von Massengütern, insbesondere von Kohle. *Mitteilungsblatt für mathematische Statistik* **6**, 204–220 (1954).
5. *Stange, K.*: Statistische Methoden für den Verfahrens-Ingenieur. *Chemie-Ingenieur-Technik* **32**, 143–154 (1960).
6. *Cameron, J. M.*: The use of components of variance in preparing schedules for sampling of baled wool. *Biometrics* **7**, 83–96 (1951).
7. *Duncan, A. J.*: Bulk sampling: Problems and lines of attack. *Technometrics* **4**, 319–344 (1962).
8. *Duncan, A. J.*: Bulk Sampling. In: *Juran, J. M.*, *Gryna, F. M.* and *Bingham, R. S.* (eds.): Quality control handbook, 3rd ed. pp. 25A1–25A13. McGraw-Hill, New York 1974.
9. *Elder, R. S.*, *Thompson, W. O.* and *Myers, R. H.*: Properties of composite sampling procedures. *Technometrics* **22**, 179–186 (1980).

Dr. H. R. Roth
Institut für Tierproduktion
Gruppe Biometrie und
theoretische Populationsgenetik
ETH-Zentrum
CH-8092 Zürich

Probenerhebung aus forensisch-toxikologischer Sicht

Sampling and Forensic Analysis

O. Jakob, J. Bäumler und S. Rippstein

Gerichtlich-medizinisches Institut der Universität, Basel

Einleitung

Während auf der Apparateseite der Analytik ein geradezu ungeheuerer Aufwand zur Qualitätsverbesserung betrieben wird, ist dies bei der Probenerhebung nicht der Fall, ja man kann durchaus feststellen, daß Kontrolle und Qualität der Probenerhebung gegenüber der eigentlichen Analytik vernachlässigt werden. Fehler jedoch, die im frühen Zeitpunkt der Probenerhebung gemacht werden, können – auch durch eine noch so ausgeklügelte Analytik – weder rückgängig gemacht noch erfaßt werden.

Durch die Art der Angabe von Analysenresultaten oder durch die Absicherung mittels Statistik werden derartige Fehler sehr oft vordergründig bagatellisiert oder gar überspielt. Etwas provokatorisch ausgedrückt kann man sagen, daß mit einem immer größer werdenden Aufwand bei der Analytik Resultate produziert werden, die den Anschein machen, genauer zu sein, deren Bezug zur materiellen Wahrheit jedoch immer unbestimmter wird. Dieses Auseinanderklaffen wird immer brisanter, da die Verfügbarkeit von «Black-box»-Apparaten, die solche vordergründig genaue Daten liefern können, immer leichter wird und deren Handhabung ohne jede Kenntnis der eigentlichen Analysen- oder Meßvorgänge erfolgen kann. Darauf hat auch H. W. Nürnberg in seiner Eröffnungsansprache des Jülich-Symposiums im Zusammenhang mit der Umweltanalytik hingewiesen: Es sei zu befürchten, meint er, daß die bereits bestehende Zahl der falschen Daten durch eine zusätzliche Schwemme falscher Daten vermehrt werde, da sich Dilettanten Instrumente kaufen und diese – ohne einen analytischen Background zu besitzen – bedienen könnten.

Die permanente Verbesserung der Analysentechnik führt damit – wegen der Vernachlässigung des Umfeldes und Hintergrundes der Analytik – lediglich zu einer Pseudo-Genaugigkeit und bringt eine zunehmende Verunsicherung mit sich.

Spezielle Problematik der forensischen Toxikologie

Aus forensischer Sicht hat bei der Analytik im allgemeinen und bei der Probenherhebung im besonderen zusätzlich die Frage der Übereinstimmung von Meßresultat und materieller Wahrheit besondere Bedeutung.

Das Beurteilen im allgemeinen und insbesondere im Sinne von «Rechtsprechen» setzt voraus, daß ein Zustand oder eine Aktion entsprechend der kritischen, d. h. der zu beurteilenden Situation «rekonstruiert» bzw. dargestellt wird. Anders formuliert: Für den Rechtsprechenden ist die Kenntnis der «materiellen Wahrheit» zumindest theoretisch eine wesentliche Voraussetzung.

Dieser Dimension der Problematik kommt eine Brisanz zu, die immer größer wird, in einer Zeit, wo verbale Kommunikation und Gebrauch der Sprache immer undisziplinierter werden, indem Begriffsverwischungen erfolgen, Begriffe mit unterschiedlichem Inhalt synonym gebraucht oder gegeneinander ausgetauscht werden, usw. Im Umgang mit dem Strafrecht ist in diesem Zusammenhang vor allem die Ebene der Kausalität zu nennen: So ist ein «nicht bewiesener» Kausalzusammenhang nicht identisch mit einem «nicht existenten» Kausalzusammenhang, «nicht schuldig» ist nicht identisch mit «kausal nicht verantwortlich», «nicht schuldig gesprochen» ist nicht identisch mit «unschuldig» usw., usw.

Verfahrensablauf

Nicht meßbare Größen können durch Aussagen von direkt oder indirekt Beteiligten «faßbar» gemacht werden, und da der Richter das entsprechende Erhebungsinstrument (die Befragung) selbst handhaben kann, ist er weitgehend unabhängig, d. h. auf sich selbst gestellt im Spannungsfeld von Subjektivität und Objektivität.

Bei meßbaren Größen ist theoretisch zwar eine «objektive» Bestimmung möglich, in allen Fällen, wo die Komplexität der Materie an sich oder das Prozedere des Messens jedoch einen gewissen Schwierigkeitsgrad übersteigen, wird der Richter abhängig von der Mitarbeit von Fachleuten und im Spannungsfeld Objektivität/Subjektivität ist er gewissermaßen ein Seiltänzer ohne Netz:

Ein Resultat – berechnet auf x Stellen nach dem Komma und versehen mit einem Streubereich – präsentiert sich ihm als so genau, daß er sich im Glauben befinden mag, es sei wohl geradezu blasphemisch, die «banale Frage» nach der Verwechslung oder dergleichen zu stellen. Es kann davon ausgegangen werden, daß zur «Rekonstruktion» der materiellen Wahrheit eine gewisse Kenntnis der Hintergrundproblematik der zur Diskussion stehenden Materie notwendig ist. Daß – genau wie die Analytik selbst – auch die Beurteilung des Umfeldes berücksichtigt werden muß, sei hier am Rande erwähnt, ist jedoch nicht Gegenstand des Tagungsthemas.

Ausgangslage und geforderter Zuverlässigkeitsgrad

Es liegt im Wesen der Sache, daß die eigentliche Analyse nicht anlässlich des kritischen Ereignisses vorgenommen werden kann und daß in der Regel der Ort, wo die Analyse ausgeführt wird, nicht identisch ist mit dem kritischen Ort.

Bei der Durchführung einer forensisch-toxikologischen Analyse ist auch stets zu beachten, daß das zur Untersuchung kommende Material in der Regel nur einmal vorhanden ist und – anders als bei der medizinischen Analyse – auf keine Weise wieder beschafft werden kann.

Soll nun ein Analysenresultat eine geforderte Zuverlässigkeit auch nur annähernd erreichen, so sind bereits bei der Probenentnahme bestimmte Voraussetzungen notwendig. Diese können in einige Schwerpunktkomplexe gruppiert werden:

- Die zweckmäßige Probenerhebung setzt voraus, daß der Dynamik des zu analysierenden Stoffes Rechnung getragen wird. Aufnahme, Verteilung, Metabolismus und Ausscheidung sind damit bereits zum Zeitpunkt der Probenentnahme von Bedeutung.
- Die zweckmäßige Probenerhebung setzt voraus, daß weder im Verlauf der Probenentnahme noch während Transport oder Lagerung, aber auch nicht anlässlich der Analysenvorbereitung, das Material unkontrollierbar bzw. nicht replizierbar verändert wird.
- Zu einer zweckmäßigen Probenerhebung sind zudem entsprechende organisatorische Voraussetzungen notwendig.

Diese drei Komplexe sollen in der Folge etwas detailliert dargestellt und an einigen Beispielen erläutert werden:

Aufnahme, Verteilung, Metabolismus und Ausscheidung

Die hohe Empfindlichkeit moderner Analysengeräte ermöglicht die Vornahme von Analysen an sehr kleinen Probenmengen. In diesem Zusammenhang erhalten Inhomogenitäten eine zunehmende Bedeutung, da mit zunehmender Verkleinerung der Probe die Umrechnungsfaktoren größer werden. Soll ein Analysenresultat daher aussagekräftig werden, so muß bereits bei der Probenentnahme darauf geachtet werden, ob zur Untersuchung ein homogenes oder ein inhomogenes Material vorliegt. Je nach Fragestellung – nicht aber nach der Voreingenommenheit der Sache gegenüber, aber auch nicht der Bequemlichkeit des Untersuchers – muß beim Vorliegen eines inhomogenen Materials eine größere Zahl von Proben sichergestellt und entsprechend beschriftet werden. Es mag trivial klingen, ist aber gerade in diesem Zusammenhang von wesentlicher Bedeutung, daß jede einzelne Probe so beschriftet werden muß, daß eine Verwechslung ausgeschlossen ist. Eine noch so exakte Analyse taugt nämlich nichts, wenn eine Verwechslung des Analysenmaterials erfolgt ist.

Bei der Probenentnahme ist darauf zu achten, daß auch bei späteren Manipulationen eine Verwechslung ausgeschlossen wird, wobei vor allem dafür gesorgt werden muß, daß das Nichtvorliegen einer Verwechslung dokumentiert werden muß, damit es überprüft werden kann.

Fragestellung

Als Beispiel für die Wichtigkeit der Berücksichtigung der Fragestellung kann der Nachweis gewisser Insektizide genommen werden: Es ist bekannt, daß eine Speicherung im Fettgewebe erfolgen kann, daß die Entfaltung der biologischen Aktivität jedoch nicht im Fettgewebe erfolgt. Damit erfordert die Fragestellung nach der momentan biologisch aktiven Menge die Entnahme eines anderen Materials als die Frage nach dem Vorhandensein des Stoffes im Körper überhaupt.

Hierhin gehört im Zusammenhang mit forensisch-toxikologischen Analysen bei Todesfällen die Problematik der agonalen und frühpostmortalen Veränderungen von autoptisch erhobenen Proben.

Eine zusätzliche Komplikation ergibt sich, wenn nicht nur die Frage nach der biologisch aktiven Menge zur Diskussion steht, sondern wenn auch eine Interpretation der aufgetretenen Wirkung erfolgen soll. In dieser Hinsicht ergeben sich vor allem im Zusammenhang mit Opiatkonsumenten (sogenannte «Drogentote») Probleme, da die Angabe einer letalen Dosis bei einem Stoff mit Gewöhnung nur mit Bedenken erfolgen kann.

Organisatorische Voraussetzungen

Als Beispiel, wie ein eingespieltes Prozedere mit der notwendigen Organisation eine optimale Probenerhebung ermöglicht, kann die Dopingkontrolle gelten.

Die Fragestellung ist eindeutig: Es soll festgestellt werden, ob in der kritischen Zeit eine unerlaubte Substanz im Körper vorhanden war oder nicht. Der Mechanismus und das Ausmaß stehen *nicht* zur Diskussion.

Damit der ganze Analysenvorgang replizierbar ist, wird die Urinprobe sofort in zwei Chargen aufgeteilt und mit A und B bezeichnet. Während Charge A analysiert wird, wird Charge B aufbewahrt, damit eine Kontrollanalyse vorgenommen werden kann. Über das ganze Prozedere muß ein Protokoll geführt werden, das auch vom Betroffenen Exploranden unterzeichnet werden muß. Da es sich um eine Routineangelegenheit handelt und der Ablauf uniform ist, ist eine optimale Vorbereitung möglich (Behälter, Etiketten, Protokolle, eingespieltes Prozedere usw.). Im Zusammenhang mit Dopingkontrollen ergeben sich denn auch praktisch nie Probleme — unter der Voraussetzung allerdings, daß das vorgeschriebene Prozedere eingehalten wird.

Sind irgenwelche Arbeitsgänge notwendig, um eine Analyse auch später zu ermöglichen (Zentrifugieren von Blut usw.), so hat das plombierte Gefäß in Anwesenheit eines Funktionärs geöffnet und anschließend wieder verschlossen zu werden.

Schwieriger und vor allem mit größerem Aufwand verbunden, ist die Probenerhebung immer dann, wenn die Richtung, in die die Fragestellung später gehen wird, nicht zum vornherein klar ist, und wenn bei der Probenerhebung ein vielfältiges Spektrum hinsichtlich Analytik abgedeckt werden soll; d. h. Aufwand und Schwierigkeit im Zusammenhang mit der Probenerhebung nehmen mit dem Grad der Unvoreingenommenheit und Komplexität der Fragestellung zu.

Als Beispiel kann das Problemfeld eines teilweise verbrannten menschlichen Körpers dienen. Hier steht in der Regel die Frage der Todesursache zur Diskussion. Darüber hinaus kann sich jedoch bereits vor oder anlässlich der Autopsie die Frage der Anwendung von Brandbeschleunigungssubstanzen oder Narkotisierungsmitteln stellen. Der analytische Nachweis solcher – meist sehr flüchtiger – Substanzen hat in verschiedenen Materialien zu erfolgen, da sich daraus Hinweise auf einen möglichen Aufnahmeweg ergeben können und damit dahin, ob die Substanzen bereits vor oder erst nach dem Todeseintritt appliziert wurden. Gerade im Zusammenhang mit hochflüchtigen Substanzen ist es notwendig, unter Umständen auch Gase in Hohlorganen vor deren Eröffnen zu asservieren. Unterbleibt dies, so wird dadurch die Aussagemöglichkeit eines mit großem analytischen Aufwand erhobenen Resultates zum vornherein vernichtet.

Ganz allgemein wäre damit zu fordern, daß eine weitere Entwicklung analytischer Apparate in Richtung Genauigkeit nur dann sinnvoll angewandt werden kann, wenn auch der Probenerhebung die notwendige Aufmerksamkeit und Sorgfalt geschenkt wird.

Zusammenfassung

Auf der Apparateseite der Analytik wird – im Gegensatz zur Probenerhebung – ein geradezu ungeheurer Aufwand zur Qualitätsverbesserung betrieben. Dabei können sich sogar Dilettanten Instrumente kaufen und diese – ohne analytischen Background zu besitzen – bedienen. Deshalb ist die Probenerhebung der schwache Punkt des ganzen analytischen Ablaufes.

Die Übereinstimmung von Meßresultat und materieller Wahrheit hängt somit also vor allem von einer zweckmässigen Probenerhebung ab. Daraus wäre zu folgern, daß eine weitere Entwicklung analytischer Apparate in Richtung Genauigkeit nur dann sinnvoll angewandt werden kann, wenn auch der Probenerhebung die notwendige Aufmerksamkeit und Sorgfalt geschenkt wird.

Résumé

Dans le domaine analytique, les efforts portent trop sur le perfectionnement de l'appareillage, comparé à l'attention donnée au prélèvement des échantillons. Ce dernier apparaît le point faible dans le procédé analytique.

La concordance des résultats d'analyse avec la réalité dépend essentiellement du prélèvement des échantillons. Le perfectionnement de l'appareillage en vue d'en augmenter la précision n'a donc de sens que si l'on donne au prélèvement des échantillons l'importance primordiale qui lui revient.

Summary

The quality of chemical analysis has much improved, at least as far as the laboratory equipment is concerned. By means of fool-proof units for analytical working even nonprofessional people can produce results that show a remarkable precision. On the other hand, sampling is the weak point of the whole analytical procedure.

Therefore today the accuracy of results is rather dependent on sampling than on the analytical technique. This aspect has to be considered when dealing in forensic matters with results of chemical analysis.

Dr. med., Dr. med. dent. O. Jakob
Dr. phil. J. Bäumler
S. Rippstein
Gerichtlich-medizinisches Institut
der Universität Basel
Postfach 282
CH-4012 Basel

Wie zuverlässig lässt sich die Bleibelastung des Menschen früherer Zeiten aus archäologischen Knochenfunden abschätzen?

Is the Lead Concentration of Archeological Bone Finds a Reliable Method
for the Determination of Lead Exposure in Ancient Times?

R. Knutti

Institut für Toxikologie der ETH und Universität Zürich, Schwerzenbach

Einleitung

Blei und alle anderen Metalle gehörten schon immer zur Umwelt. Ein Teil dieser Elemente wurde von den Lebewesen unseres Planeten während ihrer Entwicklung nutzbringend eingesetzt; es bildeten sich aber auch Abwehrmechanismen gegen schädliche Auswirkungen von Elementen in der Umwelt. Für eine toxikologische Beurteilung der heutigen Belastung der Bevölkerung durch Blei ist es deshalb von Interesse zu wissen, wie groß diese Belastung durch unsere zivilisatorischen Tätigkeiten gegenüber der naturbedingten ist. Da das vom Körper aufgenommene Blei zu über 90 % in den Knochen gespeichert wird (1, 2), ergibt sich im Prinzip die Möglichkeit, aus der Bleikonzentration im Skelett die Bleibelastung der Bevölkerung in der heutigen Zeit mit derjenigen früherer Epochen und Kulturen zu vergleichen.

Die Feststellung, daß ein solcher Vergleich der Bleibelastung nur sinnvoll ist, wenn vergleichbares Probenmaterial zur Verfügung steht, mag trivial erscheinen. Den Beweis zu führen, daß die untersuchten Knochenproben dieser Anforderung genügen, stellt sich bei der hier besprochenen Problemstellung gerade als einer der schwierigsten Punkte heraus.

Im folgenden soll anhand verschiedener Arbeiten anderer Autoren, aber auch aufgrund eigener, unangenehmer Erfahrungen dargelegt werden, von welchen Faktoren die Bleikonzentration in Knochenproben abhängt. Schließlich werden noch kurz einige Arbeiten zum Thema Bleibelastung der Bevölkerung in früheren Zeiten diskutiert. Wer sich ausführlicher mit dem Thema befassen möchte, sei auf die entsprechenden Original- und Übersichtsartikel verwiesen (3–13).

Faktoren, die die Bleikonzentration im Knochen beeinflussen

Bei der zu beurteilenden Belastung eines Individuums durch Blei gibt es eine ganze Reihe von Faktoren, die zu berücksichtigen sind:

- Richtigkeit der Analysenresultate
- Todesalter des untersuchten Individuums
- Geschlecht
- Knochentyp
- Verteilung von Blei innerhalb des untersuchten Knochens
- Bezugsgrößen für die Bleikonzentration
- Post-mortem-Veränderungen (Kontamination durch Lagerung im Boden)

Richtigkeit der Analysenresultate

Die Präzision moderner spurenanalytischer Meßmethoden verdeckt leider häufig systematische Fehlerquellen bei Bestimmungen in komplex zusammengesetzten biologischen Probenmaterialien. Eine gute Präzision und in sich konsistente Analysendaten sind leider noch lange kein Beweis für die Richtigkeit der Resultate. Mit der Methode der Standardadditions-Eichung können Matrixeffekte, die sich nur auf die Empfindlichkeit des Meßsignals auswirken, eliminiert werden. Dies ist aber auch nur dann der Fall, wenn sich das in den Proben zu bestimmende Element und der Eichzusatz identisch verhalten, was meistens gar nicht so leicht zu beweisen ist. Dabei darf aber nicht vergessen werden, daß Matrixeffekte, die sich *additiv* auf das Meßsignal auswirken, mit der Standardadditions-Eichung nicht erfaßt werden können! Bei der Bestimmung von Blei in Knochen ist deshalb z. B. zu beachten, daß bei Verwendung der Atomabsorptionsspektrometrie der hohe Calciumgehalt, auch bei einer Atomisierung der Probe in der Flamme, ein zusätzliches, unspezifisches Absorptionssignal verursachen kann.

Wird die Messung mit der Graphitrohr-Atomabsorptionsspektrometrie oder der Inversvoltammetrie durchgeführt, dann können bei der Bestimmung von Blei in Knochen die Probelösungen dank der extrem hohen Empfindlichkeit dieser zwei Meßmethoden so stark verdünnt werden, daß Matrixeffekte weitgehend wegfallen, dies allerdings auf Kosten einer erhöhten Kontaminationsgefahr für die Meßlösungen, wenn im tiefen ng/g-Bereich gemessen werden soll. Die Beurteilung der Richtigkeit publizierter Analysendaten für Blei in Knochen ist meistens schwierig, da auch heute noch kaum laborinterne oder -externe Möglichkeiten einer effektiven Analysenqualitätskontrolle bestehen. Die in den folgenden Abschnitten diskutierten großen Streuungen von publizierten Analysenresultaten für verschiedene Knochenproben des gleichen Individuums sind möglicherweise in manchen Fällen eher auf die hier aufgezeigten analytischen Probleme als auf einen biologischen Effekt zurückzuführen.

Abhängigkeit der Bleikonzentration der Knochen vom Lebensalter

Ein Teil der vom Körper aufgenommenen Bleimenge wird in die Knochenstruktur eingebaut. Da wir Blei — vor allem mit der Nahrung — mehr oder weniger kontinuierlich aufnehmen, ist zu erwarten, daß die Bleikonzentration in den Knochen mit dem Lebensalter ansteigt. *Rabinowitz* (14) hat in Metabolismusstudien mit stabilen Bleisotopen gezeigt, daß die Abgabe von Blei aus dem Skelett so langsam verläuft, daß bei einer konstanten Exposition mit einer lebenslangen Zunahme der Bleikonzentration in den Knochen zu rechnen ist. Dies wurde in einer Reihe von Untersuchungen an menschlichen Knochen bestätigt (12, 15, 16). *Steenhout* (17) gibt in ihrer Zusammenfassung von Literaturdaten über die Akkumulation von Blei in den Knochen bei beruflich nicht bleiexponierten Personen eine jährliche Zunahme der Bleikonzentration von 0,5 bis 1 µg/g an. Auch wenn sich diese Untersuchungen auf Probenmaterial von beruflich nicht bleiexponierten Personen beziehen, können die individuellen Bleibelastungen noch beträchtlich schwanken, was zu einer recht großen Streuung der Konzentrationswerte für die einzelnen Individuen führt. Ein recht typisches Bild gibt Abbildung 1, in der die alters- und geschlechtsabhängige Verteilung der von *Grandjean* (6) gemessenen Bleikonzentrationen in Wirbeln von dänischen Autopsieproben aus dem Jahre 1972 dargestellt ist.

Abhängigkeit der Bleikonzentration der Knochen vom Geschlecht

Verschiedene Studien bei beruflich nicht belasteten Personen (18, 19) haben gezeigt, daß Frauen gegenüber Männern im Mittel eine um 10 bis 15 % tiefere Blutbleikonzentration aufweisen. Dies widerspiegelt sich offenbar auch im Bleigehalt der Knochen, wie wir aus den Daten von *Grandjean* in Abbildung 1 gesehen haben. *Barry* (20) fand bei einer Untersuchung an Skeletten von 36 weiblichen und 58 männlichen, beruflich nicht bleiexponierten Personen, bei den Frauen im Mittel sogar um bis zu 40% tiefere Konzentrationen als bei den Männern. Aus der Arbeit geht allerdings nicht hervor, ob diese großen Differenzen nur durch einen biologisch bedingten Unterschied zu erklären sind.

Bleigehalt verschiedener Knochentypen

Beim gleichen Individuum können verschiedene Knochenarten recht unterschiedliche Bleikonzentrationen aufweisen. *Weinig* und *Börner* (3) fanden bei 30 Skeletten von beruflich nicht belasteten Personen eine ziemlich konstante Reihenfolge der Bleigehalte in der Knochenasche von Wirbel > Rippen > Schienbein > Oberschenkelknochen. Das Verhältnis der Bleikonzentration von Wirbel zu Oberschenkelknochen schwankte dabei zwischen einem Faktor 1,1 und 2,2. *Barry* (20) fand bei 41 erwachsenen Individuen in der Asche von Rippen, Schienbein und Schäeldach im Mittel etwa die gleichen Werte, während die Konzen-

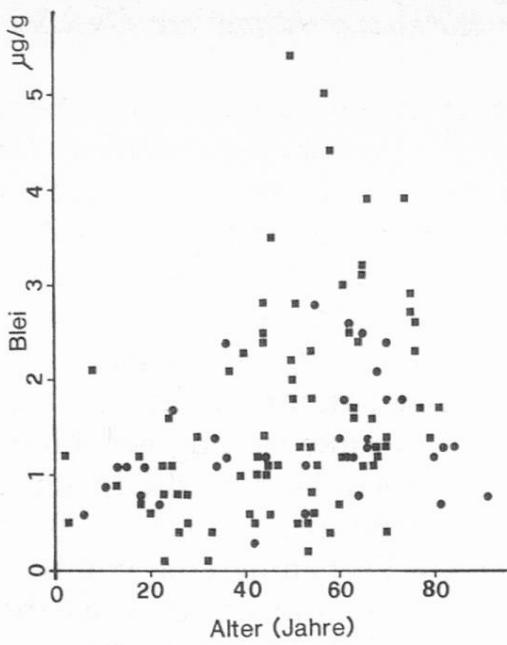


Abb. 1. Abhängigkeit der Bleikonzentration in Rückenwirbeln vom Todesalter bei Personen ohne berufliche Bleibelastung. ■ Männer, ● Frauen (nach Grandjean (6), dänische Autopsieproben 1972, Konzentrationen bezogen auf Trockengewicht)

tration im Felsenbein etwa doppelt so hoch war. Drasch (11) untersuchte Schädel, Brustbein, Beckenknochen und Oberschenkelknochen und fand bei den einzelnen Individuen sehr unterschiedliche Konzentrationsverhältnisse zwischen den einzelnen Knochenarten, die z. B. für Oberschenkelknochen/Brustbein zwischen 0,8 und 9 lagen. Diese extrem großen Schwankungen lassen allerdings vermuten, daß sehr wahrscheinlich ein beträchtlicher Teil dieser Streuung auf analytische Fehlerquellen zurückzuführen ist.

Verteilung des Bleigehaltes innerhalb des gleichen Knochens

Eine ausführliche Untersuchung der Verteilung des Bleigehaltes in menschlichen Rückenwirbeln und Rippen stammt von Ulrich (8). In allen Fällen waren die Konzentrationen in den Rippen tiefer als in den zugehörigen Wirbeln, wobei das Konzentrationsverhältnis zwischen 1,05 und 2,5 lag. Abbildung 2 zeigt die Bleiverteilung innerhalb von 4 Rippen des gleichen Individuums. Die geringsten Schwankungen wiesen die Rippen im ventralen Bereich auf. Vergleichsmessungen zur Beurteilung der Bleibelastung einer Person sollten deshalb mit Vorteil an Knochenproben aus diesem Bereich durchgeführt werden.

Beträchtliche Konzentrationsunterschiede bestehen auch über den Querschnitt der Kompakta von Rippenknochen. Es ist seit langem bekannt (3), daß bei Opfern einer akuten Bleivergiftung in den äußersten Knochenschichten ein Vielfaches der mittleren Bleikonzentration im Knochen gefunden wird. Ein –

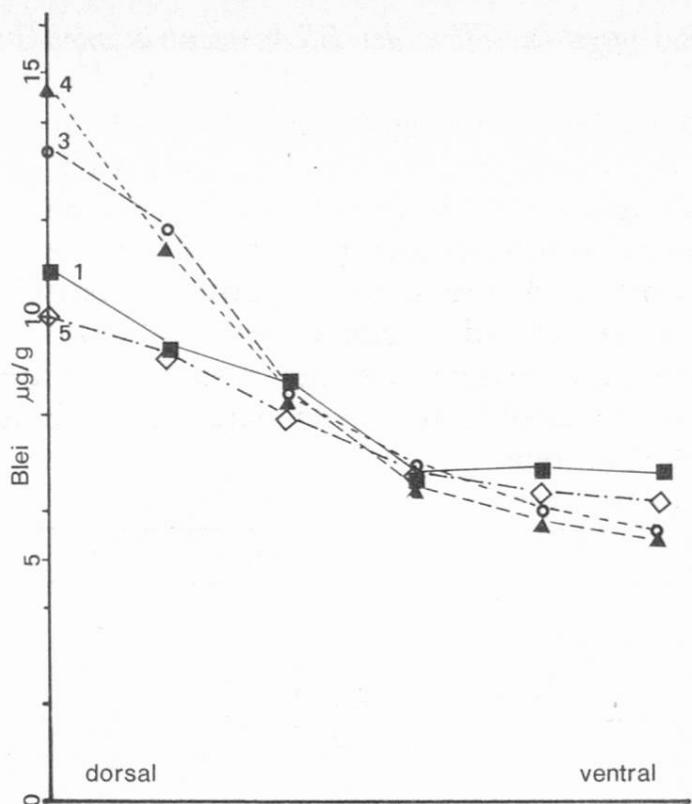


Abb. 2. Verteilung der Bleikonzentration in den Rippen Nr. 1, 3, 4 und 5 bei einem 1975 verstorbenen Erwachsenen (nach Ulrich (8), Konzentration bezogen auf Knochenasche)

zwar geringerer — Konzentrationsgradient quer durch den Knochen besteht aber auch bei Personen, die weder akut, noch beruflich einer erhöhten Bleibelastung ausgesetzt waren (Abb. 3). In dieser Darstellung zeigt sich auch deutlich der unterschiedliche Bleigehalt der Knochen von Kindern und Erwachsenen.

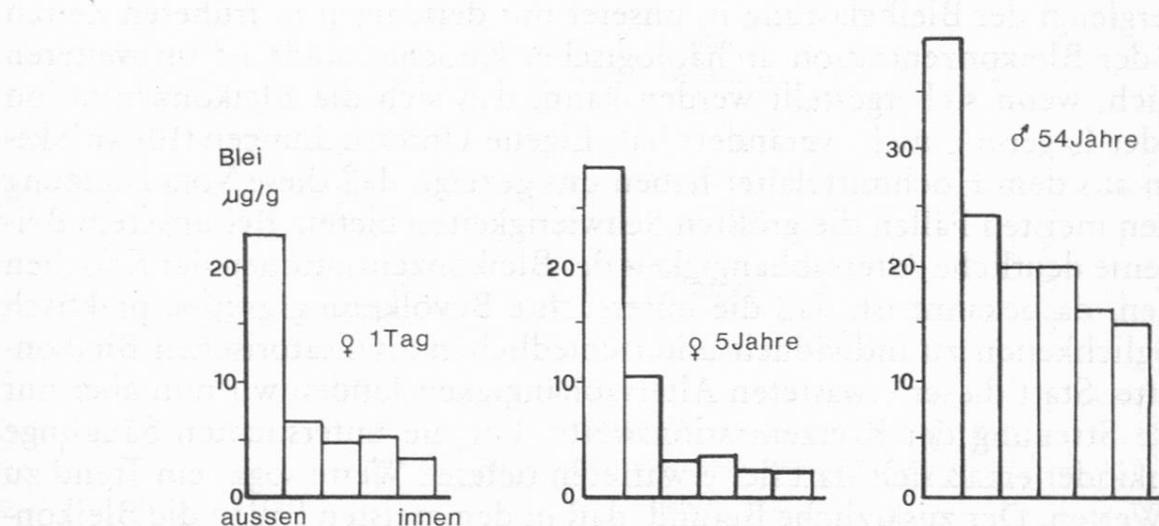


Abb. 3. Konzentrationsprofil quer durch die Kompakta von Rippen bei 1978 verstorbenen Personen ohne berufliche Bleibelastung (nach Knutti (10), Konzentration bezogen auf Knochenasche)

Auf welche Bezugsgröße sollen die Bleikonzentrationen bezogen werden?

In der Literatur findet man Angaben über die Bleikonzentration von Knochen, die sich auf das Frischgewicht, das Trockengewicht, die fettfreie Trockenmasse oder die Knochenasche beziehen. Für einen Vergleich der heutigen Bleibelastung mit derjenigen in früheren Zeiten eignet sich wegen des Abbaus der organischen Knochensubstanz während der Lagerung praktisch nur der Bezug auf die Knochenasche. Dies hat zudem den Vorteil, daß dabei die Unterschiede zwischen verschiedenen Knochenarten geringer werden, wie dies *Barry* (20) für Rippen, Schienbein und Schädeldach gezeigt hat. Für die Umrechnung empfiehlt *Barry* (12) folgende Faktoren:

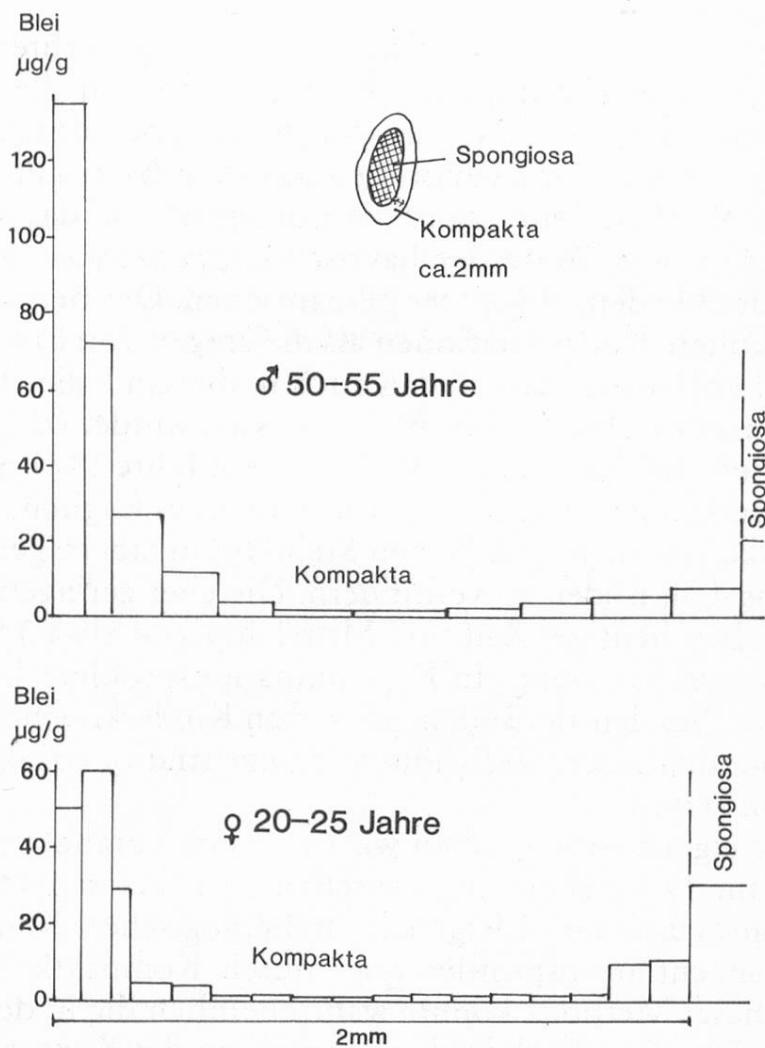
	Asche: Feucht	Asche: Trocken
Wirbel	6 : 1	3 : 1
Rippen	4 : 1	2 : 1
Kompakte Knochen	1,6 : 1	1,6 : 1

Das Verhältnis für Kompaktknochen gilt auch für den kompakten Anteil von Röhrenknochen. Diese Angaben sind Mittelwerte, die von Publikation zu Publikation leicht schwanken. Für den uns interessierenden Vergleich der Bleibelastung in früheren Zeiten spielen diese Streuungen allerdings keine große Rolle.

Post-mortem-Veränderungen der Bleikonzentration im Knochen

Ein Vergleich der Bleibelastung in unserer mit derjenigen in früheren Zeiten aufgrund der Bleikonzentration archäologischer Knochenfunde ist im weiteren nur möglich, wenn sichergestellt werden kann, daß sich die Bleikonzentration während der Lagerung nicht verändert hat. Eigene Untersuchungen (10) an Skelettfunden aus dem Hochmittelalter haben uns gezeigt, daß diese Voraussetzung wohl in den meisten Fällen die größten Schwierigkeiten bieten. Bei unserem Beispiel war eine deutliche Altersabhängigkeit der Bleikonzentrationen der Knochen zu erwarten, da bekannt ist, daß die untersuchte Bevölkerungsgruppe praktisch keine Möglichkeiten zu individuell unterschiedlichen zivilisatorischen Bleikontakten hatte. Statt dieser erwarteten Altersabhängigkeit fanden wir nun aber nur eine große Streuung der Konzentrationswerte. Für die untersuchten Säuglinge und Kleinkinder ergab sich statt der erwarteten tieferen Werte sogar ein Trend zu höheren Werten. Der zusätzliche Befund, daß in den meisten Fällen die Bleikonzentration der anhaftenden Erde über der mittleren Konzentration der Knochen lag, machte uns eindeutig klar, daß wir nicht die zu Lebzeiten aufgenommene Bleimenge, sondern die Kontamination durch die fast eintausendjährige Lage-

rung der Skelette im Boden bestimmt hatten . . . Dank dem hohen Nachweisvermögen und der geringen benötigten Probenmenge war es dann aber doch noch möglich, zumindest einen Hinweis auf die uns interessierende Bleibelastung zu Lebzeiten zu erhalten. Der Querschnitt durch die Kompakta der untersuchten Rippenstücke (Abb. 4) zeigt einen extrem großen Konzentrationsgradienten von außen nach innen. Die Form dieser Konzentrationsprofile mit einem breiten Konzentrationsplateau in der Mitte der Kompakta lässt vermuten, daß der Plateauwert dem zu Lebzeiten aufgenommenen Blei entspricht. Diese Methode konnte allerdings nur für die Erwachsenenknochen eingesetzt werden, da bei den dünnen Kinderknochen dieser Plateaueffekt noch nicht vorhanden war. Der Vergleich mit den entsprechenden Konzentrationswerten von frischen Knochenproben von Erwachsenen (Abb. 3) gibt einen Hinweis darauf, daß die Bleibelastung der Bevölkerung in Zürich im Hochmittelalter gegenüber heute etwa um einen Faktor 10 geringer war.



Beurteilung der Bleibelastung des Menschen in früheren Zeiten

Patterson (21, 22) schätzt aufgrund vor allem geochemischer Überlegungen den von Natur aus bedingten, gesamten Bleigehalt eines Erwachsenen auf etwa 2 mg. Umgerechnet auf die Bleikonzentration in der Knochenasche ergibt dies im Mittel einen Wert von etwa 0,5 bis 1 $\mu\text{g}/\text{g}$. Voraussetzung für eine experimentelle Bestätigung solch tiefer Werte in jahrtausendealten Knochenfunden von Kulturen, von denen angenommen werden kann, daß sie keinen zivilisatorischen Bleibelastungen ausgesetzt waren, ist das vollständige Fehlen einer Kontamination während der langen Lagerung im Boden. Grandjean et al. (7) untersuchten im Wüstensand natürlich konservierte Mumien von 5000 Jahren alten nubischen Kulturen. Die gemessene mittlere Konzentration von 0,6 $\mu\text{g}/\text{g}$ stimmt gut mit dem von Patterson postulierten Wert überein. Ebenfalls in Übereinstimmung mit diesem natürlichen Basiswert sind die von Drasch (11) gemessenen mittleren Konzentrationen von ebenfalls 0,6 $\mu\text{g}/\text{g}$ bei prähistorischen peruanischen Knochenfunden, die wie die nubischen ebenfalls in trockenem Wüstensand gelagert waren.

Schwieriger zu beurteilen sind Knochenfunde, die während Jahrhunderten in Kontakt mit einem feuchten Boden waren, wie dies für die meisten archäologischen Funde aus Europa der Fall ist. Drasch hat neben den peruanischen Proben eine große Anzahl von Knochenproben aus dem bayrischen Raum untersucht, die von prähistorischen Zeiten über die Römerzeit und das Mittelalter bis in unsere Tage reichen. Die Werte der bayrischen prähistorischen Funde liegen mit 2 $\mu\text{g}/\text{g}$ deutlich über denjenigen der peruanischen. Der Befund, daß die Knochen von Kindern höhere Konzentrationen als diejenigen von Erwachsenen aufweisen, ist ein deutlicher Hinweis darauf, daß auch in diesem Falle eher eine Kontamination als zu Lebzeiten absorbiertes Blei gemessen wurde. Ulrich (8) vergleicht Rippen und Wirbel von Autopsiematerial aus dem Jahre 1975 mit frühmittelalterlichen Skelettfunden. Auch in diesem Fall scheint es fraglich, ob die Lagerung der Erwachsenenskelette in geschlossenen Steinkistengräbern genügte, um eine Kontamination aus dem Boden zu verhindern. Die eher geringen Konzentrationsverhältnisse zwischen heutiger Zeit und Mittelalter von etwa 3:1 für Wirbel und 5:1 für Rippen deuten auch auf ein Kontaminationsproblem hin. Noch deutlicher ist die Situation bei den im Boden gelagerten Kinderknochen, welche wie in unserer eigenen Arbeit, zum Teil höhere Konzentrationen als diejenigen der Erwachsenen aufweisen.

Auf unsere eigene Arbeit haben wir bereits im vorangehenden Abschnitt verwiesen (10). Eine Differenzierung zwischen zu Lebzeiten aufgenommenem und kontaminationsbedingtem Bleigehalt archäologischer Knochenfunde ist aufgrund des Konzentrationsprofiles quer durch Kompaktknochen möglich. Die Anwendung dieser Methode könnte wahrscheinlich die in den verschiedenen Arbeiten aufgezeigte Unsicherheit der Beurteilung des Kontaminationsanteiles am gemessenen Bleigehalt beseitigen.

Zusammenfassung

Der Bleigehalt der Knochen ist ein Maß für die lebenslange, kumulative Bleibelastung einer Person. Daraus ergibt sich die Möglichkeit, die Bleibelastung der Bevölkerung verschiedener Kulturen und Zeiten zu vergleichen. Die Arbeit diskutiert die zusätzlichen Faktoren, von denen die gemessenen Bleikonzentrationen abhängen: Analytische Probleme, Todesalter und Geschlecht, Knochenart, Verteilung innerhalb eines Knochens, Bezugsgröße und Post-mortem-Veränderungen der Bleikonzentration. Dieser letzte Punkt, die Kontamination durch die Lagerung im Boden, stellt bei archäologischen Knochenfunden meistens das größte Problem dar. Zur Unterscheidung zwischen dem zu Lebzeiten aufgenommenen und dem kontaminationsbedingten Bleigehalt wird die Bestimmung des Konzentrationsprofiles quer durch Kompaktknochen empfohlen.

Résumé

Le taux de plomb des os est en corrélation plus au moins étroite avec la quantité de plomb accumulée par les personnes durant leur vie, d'où la possibilité de comparer la charge de populations de différentes époques et civilisations. Dans cet article, les divers paramètres dont dépendent les concentrations en plomb mesurées sont discutés: problèmes analytiques, mortalité, sexe, espèce de l'os, distribution à l'intérieur de l'os, point de référence de la concentration et modification par contamination de la teneur après la mort. Ce dernier point, la contamination due au contact avec le sol, représente le plus souvent le problème majeur au sujet des os provenants de fouilles archéologiques. Pour distinguer au mieux le taux d'accumulation durant la vie du taux dû à la contamination, il est recommandé de déterminer la concentration à travers les couches des parties compactes des os.

Summary

As the skeleton contains more than 90% of the body burden of lead, the exposure of populations of different cultures can thus be compared by measuring the lead content of bones. Additional factors determining the measured concentrations are: analytical problems, age of death and sex, type of bone, distribution of lead within a piece of bone, reference of concentration and post-mortem changes. In most cases, this last point, e.g. contamination from soil, is the main problem in estimating lead exposure. The concentration profile across compact bones allows a distinction between contamination from soil and lead accumulated during life.

Literatur

1. Schroeder, H. A. und Tipton, I. H.: The human body burden of lead. Arch. Environ. Health 17, 965–978 (1965).
2. Barry, P. S. I. and Moßmann, D. B.: Lead concentrations in human tissues. Br. J. Ind. Med. 27, 339–351 (1970).
3. Weinig, E. und Börner, B.: Über den normalen Bleigehalt der menschlichen Knochen. Arch. Toxicol. 19, 34–48 (1961).

4. Jaworowski, Z.: Stable lead in fossil ice and bones. *Nature* **217**, 152–153 (1968).
5. Crawford, M. D. and Crawford, T.: Lead content of bones in a soft a hard water area. *Lancet* **1**, 699–701 (1969).
6. Grandjean, Ph.: Lead in Danes. Historical and toxicological studies. In: Griffin, T. B. and Knelson, J. H. (ed.): *Lead. Suppl. Vol. II of Environmental Quality and Safety*. Thieme, Stuttgart 1975.
7. Grandjean, Ph., Nielson, O. and Shapiro, I. M.: Lead retention in ancient Nubians and contemporary populations. *J. Environ. Pathol. Toxicol.* **2**, 781–787 (1979).
8. Ulrich, L.: Untersuchungen über den Bleigehalt in Wirbeln und Rippen. *Arch. Toxicol.* **41**, 133–148 (1978).
9. Ericson, J. E., Shirahata, H. and Patterson, C. C.: Skeletal concentrations of lead in ancient peruvians. *New Engl. J. Med.* **300**, 946–951 (1979).
10. Knutti, R.: Die Bleibelastung der Zürcher Bevölkerung im Hochmittelalter (9.–11. Jahrhundert). *Mitt. Gebiete Lebensm. Hyg.* **73**, 186–195 (1982).
11. Drasch, G. A.: Lead burden in prehistorical and modern human bones. *Sci. Total Environ.* **24**, 199–231 (1982).
12. Barry, P. S. I.: Distribution and storage of lead in human tissues. In: Nriagu, J. O. (ed.): *The biogeochemistry of lead in the environment. Part B*. Elsevier, Amsterdam 1978.
13. Steenhout, A.: Kinetics of lead storage in teeth and bones. An epidemiologic approach. *Arch. Environ. Health* **37**, 224–231 (1982).
14. Rabinowitz, M. B., Wetherill, G. W. and Kopple, J. D.: Kinetic analysis of lead metabolism in healthy humans. *J. Clin. Invest.* **58**, 260–270 (1976).
15. Gross, S. B., Pfitzer, E. A., Yeager, D. W. and Kehoe, R. A.: Lead in human tissues. *Toxicol. Appl. Pharmacol.* **32**, 638–651 (1975).
16. O'Connor, B. H., Kerrigan, G. C., Taylor, K. R., Morris, P. D. and Wright, C. R.: Levels of temporal trends of trace element concentrations in vertebral bone. *Arch. Environ. Health* **35**, 21–27 (1980).
17. Steenhout, A. and Pourtois, M.: Lead accumulation in teeth as a function of age with different exposures. *Br. J. Ind. Med.* **38**, 297–303 (1981).
18. Mahaffey, K. R., Annest, J. L., Roberts, J. and Murphy, R. S.: National estimates of blood lead levels: United States, 1976–1980. *New Engl. J. Med.* **307**, 573–579 (1982).
19. Berlin, A.: Assessment of exposure to lead of the general population in the European Community trough biological monitoring. *Environ. Monit. Assess.* **2**, 225–231 (1982).
20. Barry P. S. I.: A comparison of lead concentrations in human bones and soft tissues. International symposium environmental health aspects of lead: Amsterdam, 2.–6. Oct. 1972. Luxembourg: Commission of the European Communities 415–426, 1973.
21. Patterson, C. C.: Contaminated and natural lead environments of man. *Arch. Environ. Health* **11**, 344–360 (1965).
22. Patterson, C. C.: An alternative perspective – Lead pollution in the human environment: Origin, extent and significance. In: *Lead in the human environment. A report prepared by the Committee on lead in the human environment*. National research Council, National Academy of Sciences, Washington, D.C. 1980.

Dr. R. Knutti
 Institut für Toxikologie der
 Eidg. Techn. Hochschule
 und der Universität Zürich
CH-8603 Schwerzenbach

Probenvorbereitung in der Sensorik am Beispiel der Bestimmung von Geschmacksschwellenwerten von Kupfer in Trinkwasser

Determination of Threshold Concentrations of Copper in Drinking Water – An Example for Sample Treatment in Sensorial Analysis

J. Solms, Y. Béguin-Bruhin, F. Escher

Institut für Lebensmittelwissenschaft, ETH, Zürich

und *H. R. Roth*

Institut für Tierproduktion, ETH, Zürich

Kupfer kommt in vielen Trinkwasserproben in gelöster, möglicherweise auch in nur partiell gelöster Form vor. Dabei dürfte das Kupri-Ion (Cu^{++}) für einen Metallgeschmack von Bedeutung sein. Diese Geschmackseffekte können mit sensorischen Schwellenwerts-Bestimmungen erfaßt werden.

Solche Schwellenwerte für Kupfer wurden unter genau kontrollierten Bedingungen ermittelt. Als Kupfer wurde Kupfer-(II)-sulfat verwendet.

Bei der Sensorik zeigen Kupferlösungen einen stark haftenden Geschmack, der sich zudem nur sehr langsam entwickelt. Aus diesen Gründen wurde ein spezieller Fünf-Proben-Test angewendet, der je Serie nur eine Kupferprobe enthält. Der Test-Panel bestand aus ca. 30 Personen beiderlei Geschlechts im Alter von 17 bis 61 Jahren. Es war Aufgabe jedes Prüfers, die abweichende Probe je einer Serie zu erkennen. Die statistische Auswertung erfolgte mit zwei verschiedenen Computer-Programmen mittels Probit-Analyse.

Drei Wasserproben wurden in den Untersuchungen berücksichtigt. Destilliertes Wasser ist für sensorische Untersuchungen nicht geeignet; es besitzt einen variablen und oft unangenehmen Eigengeschmack. Daher wurde eine wässrige Lösung von 0,005 m Saccharose gewählt (I). Ferner wurde Mineralwasser aus Bad Pyrmont (Deutschland) verwendet (II). Schließlich wurde auch Zürcher Trinkwasser getestet (III). Die Systeme (I) und (II) ergaben stabile Kupferlösungen. Wasser (III) ergab schon nach kurzer Zeit kupferhaltige Niederschläge; hier konnten keine Schwellenwerte ermittelt werden.

Für die Probenvorbereitung und Auswertung der Resultate wurden die Lösungen mit Atomabsorptions-Spektrophotometrie (AAS) und mit spezifischen Cu-Elektroden (CuE) analytisch charakterisiert. Die AAS-Methode ergab keine zu-

verlässigen, reproduzierbaren Resultate im Zusammenhang mit der Sensorik; die CuE-Methode ergab dagegen auswertbare, reproduzierbare Werte.

Die ermittelten Schwellenwerte betragen für die wässrige Lösung (I) 2,4 bis 3,2 ppm Kupfer (als Cu⁺⁺) und für das Mineralwasser (II) 0,8 bis 1,0 ppm (als Cu⁺⁺). Der tiefere Wert für (II) kann eventuell darauf zurückgeführt werden, daß das Kupfer in der Versuchslösung durch einen Überschuß an CO₂ in anderer Form gelöst vorliegt.

Die unterschiedliche Eignung der Bestimmungsmethoden (AAS und CuE) kann wohl darauf zurückgeführt werden, daß die CuE-Methode gelöstes Kupfer im System im gleichen Maße erfaßt, wie dies bei den sensorischen Rezeptoren der Fall ist. Dagegen erfaßt die AAS-Methode Kupfer weitgehend unabhängig vom jeweiligen Lösungszustand.

Zusammenfassend kann festgestellt werden, daß reproduzierbare sensorische Schwellenwerte nur mit genau definierten Lösungen und geeigneten Analysenmethoden erhalten werden. Die Untersuchungen wurden in vollem Umfang an anderem Ort (1) veröffentlicht.

Zusammenfassung

Fragen der Probenvorbereitung und Probenanalyse werden am Beispiel der Bestimmung von Geschmacksschwellenwerten von Kupfer in Trinkwasser behandelt.

Résumé

La détermination d'un seuil de perception gustative du cuivre dans de l'eau de boisson est utilisée comme exemple pour discuter les problèmes de la préparation et de l'analyse des échantillons à cet égard.

Summary

Problems of sample preparation and sample analysis are discussed with the determination of sensory thresholds of copper in drinking water as an example.

Literatur

1. Béguin-Bruhin, Y., Escher, F., Roth, H. R. and Solms, J.: Threshold concentration of copper in drinking water. Lebens. Wiss. Technol. **16**, 22–26 (1983).

Prof. Dr. J. Solms
Institut für Lebensmittelwissenschaft
ETH-Zentrum
CH-8092 Zürich

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 Eruierung 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 Analysenresultaten 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 umfangen 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 unbekannten 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 benutzt 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 Stichprobenumfangs, 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 σ für ein Fehlerrisiko 1. und 2. Art (α bzw. β 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_α und z_β 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 σ , 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, α und β , 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 (α , β , 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

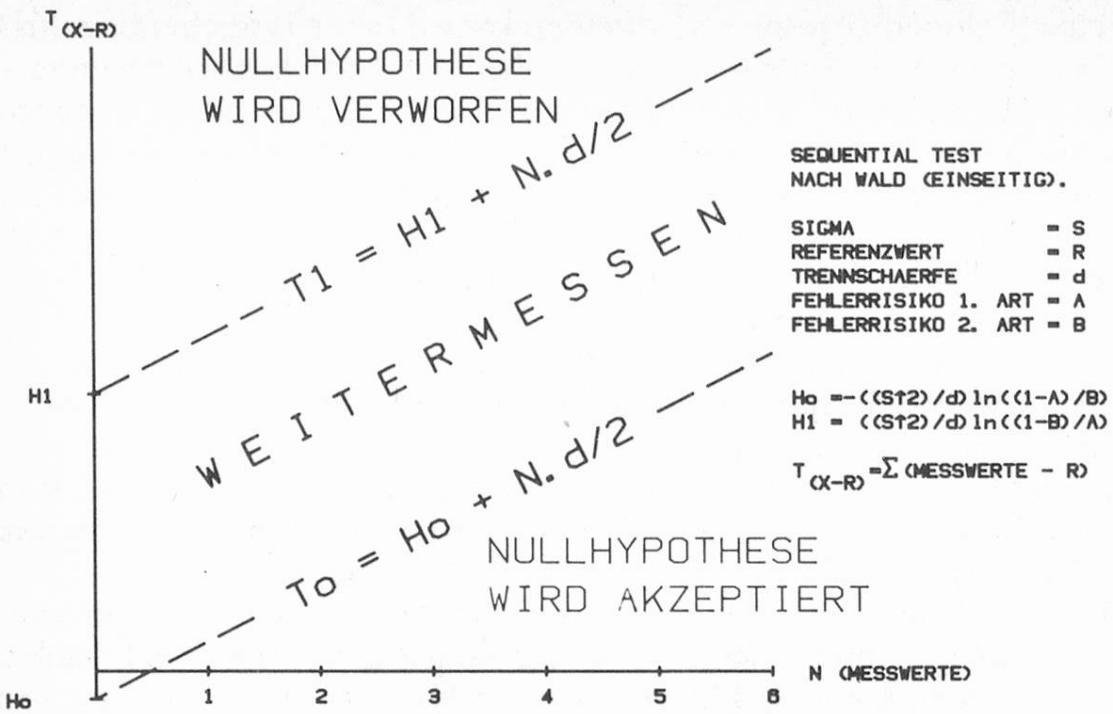


Abb. 1. Schematische Darstellung eines einseitigen Folgetestplanes (Normalverteilung)

Größe der Standardabweichung (innerhalb und zwischen den Laboratorien), die zugelassene Fehlerrisikogröße (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 Verzichten 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

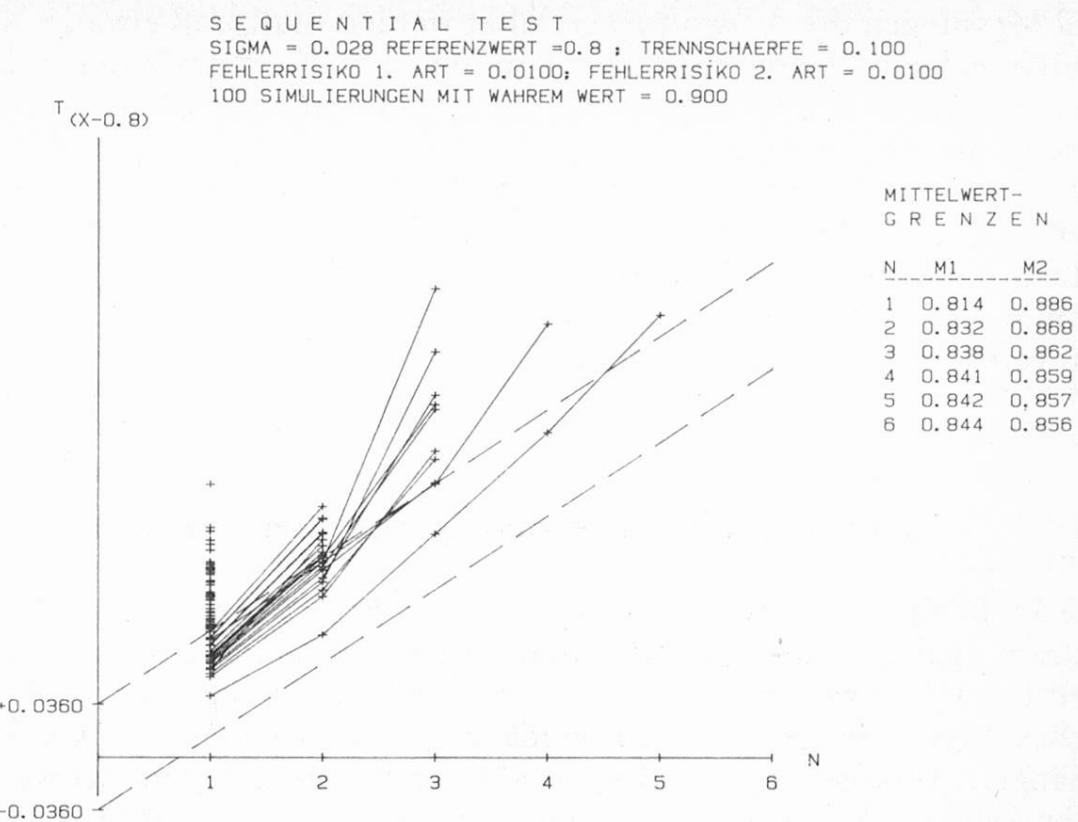


Abb. 2. Monte-Carlo-Simulationen eines einseitigen Folgetestes

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 minimiert 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 umfänglichen 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:

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

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,

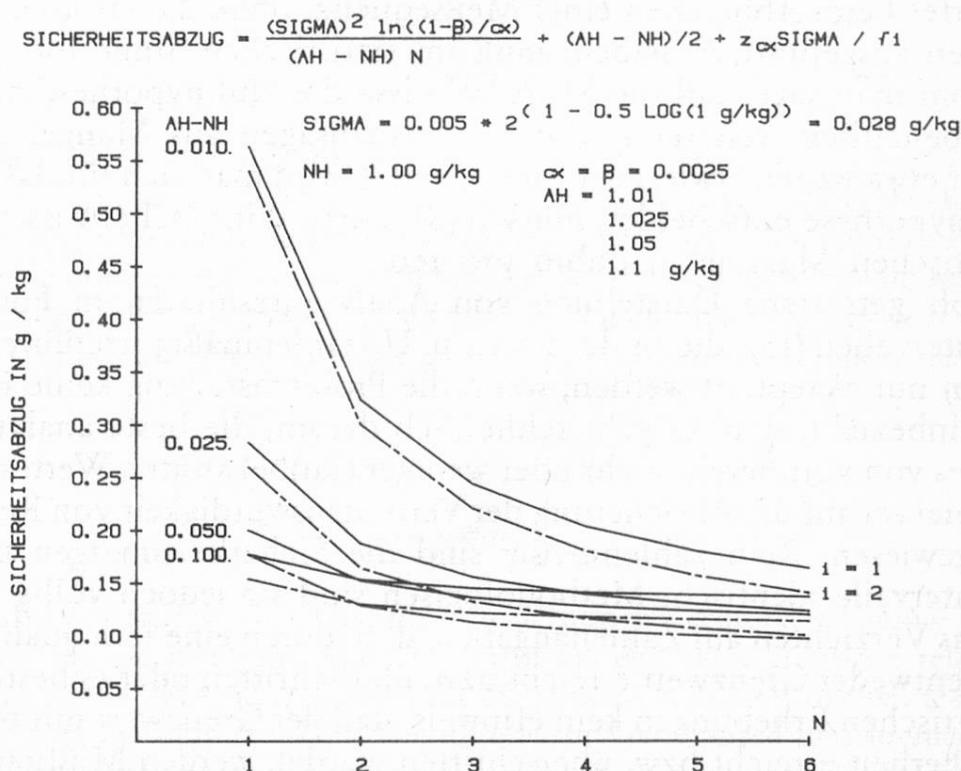


Abb. 3. Sicherheitsmargen gemäß Methodik der Maximum Likelihood für $\bar{x} = 1,00 \text{ g/kg}$ (σ nach Horwitz mit Zerlegung). Trennschärfe 0,01; 0,025; 0,05 und 0,1 g/kg

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 Analysenresultaten 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 Verzichten 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 VIth 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 Analysenresultaten. 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