

# Statistische Auswertung von Kommutabilitätsexperimenten : Anwendung eines Äquivalenztests als zweckmäßige Methode

S. Weber<sup>1</sup> und Th. Keller<sup>2</sup>

<sup>1</sup>Fachhochschule Magdeburg-Stendal, Magdeburg, D <sup>2</sup>ACOMED statistik, Leipzig, D

## Einführung

Kommutabilität wird definiert als "Äquivalenz der mathematischen Beziehungen zwischen den Ergebnissen verschiedener Messmethoden, die der Untersuchung ein und derselben Messgröße dienen, für ein Referenzmaterial (RM) und repräsentative Stichproben des designierten Präparates (klinische Proben KP)."

In einem kürzlich veröffentlichten Leitfaden (CLSI C53-P) wurden Aufbau und Analyse von Kommutabilitätsexperimenten dargestellt. Die in diesem Leitfaden vorgeschlagenen statistischen Methoden beziehen sich streng genommen jedoch nicht auf Schlüsse im Sinne von Äquivalenz.

Daraus könnten sich sowohl falsch positive, als auch falsch negative Nachweise von Kommutabilität ergeben. Daher schlagen wir die Anwendung eines Äquivalenztests zur Analyse von Kommutabilitätsexperimenten vor: Die Äquivalenz der Beziehungen des RM und der KP wird durch paar-

weise Analyse der untersuchten Methoden betrachtet. Zum Nachweis der Äquivalenz innerhalb eines Methodenpaares wird zunächst ein die Unsicherheit der RM-Messung beschreibender, 2-dimensionaler Vertrauensbereich berechnet. Dieser muss innerhalb eines prospektiv von Experten festgelegten Bereiches um die KP liegen, damit die Äquivalenz nachgewiesen ist.

Nachdem in Schritt (1) alle Methodenpaarungen einzeln analysiert werden, werden die Resultate in einem zweiten Schritt (2) zusammengefasst. Ungeachtet der Vorgehensweise in Schritt (1) kann die Bildung von Untergruppen von Methoden, für welche das RM als kommutabel anzusehen ist, Schwierigkeiten bereiten.

Neben der Vorstellung des Äquivalenznachweises als unserer Meinung nach adäquate statistische Analyseverfahren für Kommutabilitätsexperimente wird in diesem Poster das bislang offene Problem der Zusammenfassung der Resultate der paarweisen Methodenvergleiche in Schritt (2) thematisiert.

## METHODEN - Statistik

Prinzipiell erfolgt die Analyse von Kommutabilitätsexperimenten in zwei Schritten.

(1) Die Daten jedes Paares von untersuchten Methoden werden analysiert. Die Aussage über die Eigenschaften des RM im Vergleich zu den KP ist Ergebnis dieses Schrittes.

(2) Eine Kontingenztafel, die die Ergebnisse aller paarweisen Untersuchungen beinhaltet, dient der Identifizierung von Gruppen von Methoden, für welche Kommutabilität nachgewiesen wurde. Etwaig auftretende Schwierigkeiten sind im Punkt **RESULTATE** dargestellt.

Der vorgestellte Äquivalenztest bezieht sich auf Schritt(1). Folgende Schritte werden vorgeschlagen:

(1.1) **Prospektive Festlegung von KP-Bereichsgrenzen:** Die Bereiche um die klinischen Proben werden von Klinischen Chemikern vorgegeben. Innerhalb dieser Bereiche wird ein RM als kommutabel angesehen. Die Definition der Bereichsgrenzen kann z.B. durch Vorgabe konstanter Variationskoeffizienten, konstanter Standardabweichungen oder beliebiger Impräzisionsprofile erfolgen.

(1.2) **Mathematische Beziehung zwischen den klinischen Proben:** Zur Ermittlung des Funktionalzusammenhangs zwischen den KP wird die 'Generalisierte Deming Regression' [Martin 2000] genutzt. Das zugehörige Prädiktionsintervall (o. Abb.) kann Aufschluss über die Qualität der Messwerte der KP und somit die Güte der Regression geben.

(1.3) **In Beziehung setzen:** Es erfolgt die Konstruktion der KP-Bereichsgrenzen (1.1) um die Regressionsgerade(1.2).

(1.4) **Ermittlung des Vertrauensbereiches (VB)** für den Mittelwert der RM -Wiederholungsmessungen

(1.5) **Untersuchung der Beziehung des VB für die RM-Messungen und des vorgegebenen Bereiches um die KP:**

- Liegt der VB gänzlich innerhalb des KP-Bereiches, wird Kommutabilität des RM für die Methodenpaarung angenommen.

- Liegt der VB gänzlich außerhalb des KP-Bereiches, wird Kommutabilität des RM für die Methodenpaarung abgewiesen.

- Überlappt der VB den KP-Bereich, so kann keine gesicherte Kommutabilitätsaussage getroffen werden.

## METHODEN - Daten

Die in CLSI C53-P (7 RM, 25 KP, 9 Methoden MA-MJ) wurden genutzt, um die neue Vorgehensweise darzustellen. Die Daten der KP sind geeignet, verschiedene praxisrelevante Situationen darzustellen. Da die RM-Daten des CLSI C53-P nur Mittelwerte darstellen, simulierten wir Mehrfachmessungen der RM. Die Simulation umfasste dabei sowohl unterschiedliche Stichprobengrößen, als auch unterschiedliche Präzisionen.

## METHODEN - Programm

Wir entwickelten ein Software-Tool in 'R' ([www.r-project.org](http://www.r-project.org)) zur statistischen Analyse, welches in eine MS-Excel Umgebung eingebettet ist. In der Excel Umgebung werden sowohl die Datengrundlage, als auch die Parameter organisiert, sowie die aus dem 'R' Programm bezogenen Ergebnisse dargestellt.

Der Großteil der bekannten statistischen Methoden [Vesper 2007] zur Analyse von Kommutabilitätsexperimenten im Sinne von Schritt (1) ist im Software-Tool verfügbar ('Generalisierte Deming Regression'[Martin 2000], Passing Bablok Regression [Passing Bablok 1983], Bland Altman Plots, 'Linear Representation', Korrespondenzanalyse). Zusätzlich ermöglicht die Software die Durchführung des in diesem Poster vorgestellten Verfahrens.

## METHODEN – 2-dimensionaler Vertrauensbereich (VB) des Mittelwertes der RM

Liegen unabhängige Wiederholungen von RM Messungen zweier Methoden vor, so ist der VB des Mittelwertpaares die Höhenlinie einer unkorrelierten 2-dimensionalen t-Verteilung. Die Erscheinung des VB ist diamantförmig für kleine Wiederholungsumfänge N und nähert sich mit steigendem N einer Ellipse an.

Die Höhenlinienniveaus sind nicht exakt berechenbar. Wir tabellierten die Resultate von 10000 Simulationen für praxisrelevante Kombinationen von RM Wiederholungsumfängen und Konfidenzniveaus.

## RESULTATE

Der neue Ansatz ist geeignet, in Schritt (1) im Sinne der Kommutabilitätsdefinition Aussagen zu treffen.

Im zweiten Schritt (2) können möglicherweise mehrere Untergruppen identifiziert werden, in denen Kommutabilität des RM angenommen wird (jedoch nicht zwischen den Gruppen). Solche Fälle sind in unten stehend (Resultate Schritt 2) dargestellt, und stellen ein bislang offenes Problem bei der Analyse von Kommutabilitätsexperimenten dar.

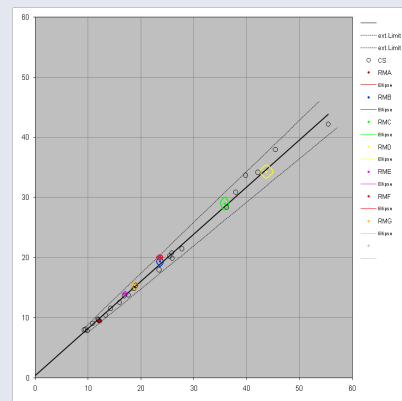
## RESULTATE Schritt(1): Kommutabilitätsprüfung der RM für Methodenpaare

**Abb. 1: Bsp.: Methode E vs. Methode F.**

Vorgegebene Grenzen: CV=5% (Strichlinien)

Annahmen für RM: CV=2%, N = 6

Legende: farbige Symbole: RM mit 95%-VB. schwarz: KP. Für alle RM bis auf RM D (rotes Symbol) ist Kommutabilität für ME/MF nachgewiesen.

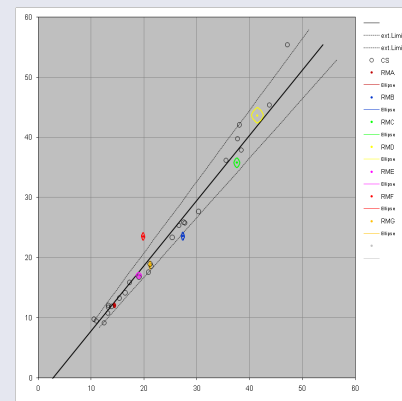


**Abb. 2: Bsp.: Methode B vs. Methode E.**

Vorgegebene Grenzen: CV=5% (Strichlinien)

Annahmen für RM: CV=2%, N = 6

Legende: farbige Symbole: RM mit 95%-VB. schwarz: KP. Bis auf RM A (blau) und RM D (rot) sind alle RM als kommutabel für MB/ME anzusehen.



## RESULTATE Schritt (2): Zusammenfassung der Ergebnisse (1) für jedes RM

RM F: Equivalence test: RM CR vs. Fixed Limits

MA MC ME MG MF MD MH MJ MB

MA 1 1 1 1 1 1 1 1 1 1

MC 1 1 1 1 1 1 1 1 1 1

ME 1 1 1 1 1 1 1 1 1 1

MG 1 1 1 1 1 1 1 1 1 1

MF 1 1 1 1 1 1 1 1 1 1

MD 0 0 0 0 0 0 0 0 0 0

MH 0 0 0 0 0 0 0 0 0 0

MJ 0 0 0 0 0 0 0 0 0 0

MB 0 0 0 0 0 0 0 0 0 0

**Ergebnis:**  
Kommutabilität für RM F besteht in einer Untergruppe von Methoden: A,C,E,G

RM D: Equivalence test: RM CR vs. Fixed Limits

MB MG MH MF MJ ME MD MA MC

MB 1 1 1 1 1 1 1 1 1 1

MG 1 1 1 1 1 1 1 1 1 1

MH 1 1 1 1 1 1 1 1 1 1

MF 1 1 1 1 1 1 1 1 1 1

MJ 1 1 1 1 1 1 1 1 1 1

ME 1 1 1 1 1 1 1 1 1 1

MD -1 -1 -1 -1 -1 -1 -1 -1 -1 -1

MA -1 -1 -1 -1 -1 -1 -1 -1 -1 -1

MC 1 1 1 1 1 1 1 1 1 1

**Ergebnis:**  
Kommutabilität für RM D besteht in drei Untergruppen von Methoden: B,E,F,G,H,J; A,D,E und D,E,F,J.

**Tabellen:** Beispiele für Kontingenztafeln, die die Ergebnisse der Paarweisen Kommutabilitätsprüfungen (1) beinhalten. RM F (links) , RM D (rechts)

Legende: 1/grün: kommutabel, 0/rot: nicht kommutabel, -1/gelb: eine Entscheidung ist nicht möglich

## DISKUSSION

Wir wandten in Schritt (1) einen Äquivalenztest zur Analyse von Kommutabilitätsexperimenten an. Das statistische Schließen im Sinne von Äquivalenz ist in Leitfäden der Pharmazeutischen Industrie gängig, falls die Gleichheit zweier oder mehr Prozesse innerhalb festgelegter Grenzen nachgewiesen werden soll (bspw. FDA 2001). In der Literatur der Klinischen Chemie finden Äquivalenztests eher selten Erwähnung (Lung, Gorko et al. 2003). Es ist jedoch Zeit, diese Methoden aufgrund ihrer Vorteile verstärkt anzuwenden.

Die Vorteile des hier vorgestellten Ansatzes sind:

(i) Die neue Methode steht im Einklang mit der Definition von Kommutabilität.

(ii) Die Kommutabilitätsgrenzen werden durch Klinische Chemiker und nicht durch Statistiker vorgegeben. Zusätzlich werden die Prüfungen nicht durch die Messqualität der KP beeinflusst. Insbesondere die Nutzung des Prädiktionsintervalls der Deming Regression mit variablem  $\alpha$ -level (!) zur Kommutabilitätsprüfung (wie in C53-P vorgeschlagen) birgt einerseits die Abhängigkeit von der Messqualität der KP und zusätzlich die Entscheidungen von Statistikern in sich. Darüber hinaus werden dieselben Grenzen für alle Vergleiche verwandt.

(iii) Der Nachweis der Gleichheit mittels Äquivalenztest beruht auf der *Ablehnung einer Nullhypothese auf Unterschied* im Gegensatz zur nichtadäquaten *Annahme der Nullhypothese der Gleichheit* innerhalb eines Tests auf Unterschied. Dies ist mit dem fundamentalen Wechsel von einem *Risikonachweis (proof of hazard)* hin zum *Sicherheitsnachweis (proof of safety)* verbunden. Somit wird im Falle schlechter Messqualität bzgl. der KP falsch positiven Kommutabilitätsaussagen vorgebeugt. Ebenso sind falsch negative Kommutabilitätschlüsse, die durch extrem präzise, unverzerrte KP Messungen hervorgerufen würden, nicht möglich.

Bezüglich **Schritt (2)** könnten mehrere Untergruppen identifizierbar sein, für welche der Kommutabilitätsnachweis des RM erbracht wird. Dieser Tatbestand bleibt bislang ungelöst.

## Literaturnachweise

CLSI EP53-P: Characterization and Qualification of Reference Materials for Laboratory medicine (Proposed Guideline). Clinical and Laboratory Standards Institute, Wayne, USA, 2008  
Lung, K. R., M. A. Gorko, et al. (2003). "Statistical method for the determination of equivalence of automated test procedures." J Autom Methods Manag Chem 25(6): 123-7.  
Martin, R. F. (2000). "General deming regression for estimating systematic bias and its confidence interval in method-comparison studies." Clin Chem 46(1): 100-4.  
Passing, H. and Bablok (1983). "A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry. Part 1." J Clin Chem Clin Biochem 21(11): 709-20.  
Vesper, H. W., W. G. Miller, et al. (2007). "Reference Materials and Commutability." Clin Biochem Rev 28(4): 9.

## Danksagung

Die Software wurde für das 'EU Institute for Reference Materials and Measurements' (IRMM), Gent, Belgien entwickelt. Wir danken Frau Ingrid Zeegers und Herrn Heinz Schimmel für die hilfreichen Diskussionen und Anregungen.