

**Medizinische Informatik,
Biometrie und
Epidemiologie**

79

**H. KUNATH U. LOCHMANN R. STRAUBE
K. H. JÖCKEL C. O. KÖHLER (HRSG.)**

**Medizin
und
Information**

**39. Jahrestagung
der Deutschen Gesellschaft für Medizinische Informatik,
Biometrie und Epidemiologie (GMDS) e. V.**

**Tagung Epidemiologie
der Deutschen Arbeitsgemeinschaft für Epidemiologie**

Dresden, September 1994

**MMV Medizin Verlag
München**

Vorzüge von Äquivalenztesten mit standardisierter Mittelwertsdifferenz demonstriert an einer Anwendung

E. Schuster¹

Einführung

Äquivalenzteste werden bei Studien angewandt, um nachzuweisen, daß zwischen zwei Behandlungen keine praktisch relevanten Unterschiede bestehen. Deshalb besteht der Fehler 1. Art bei Äquivalenztesten in einer irrtümlichen Entscheidung zugunsten der Hypothese, daß die Behandlungen „im wesentlichen identische“ Effekte haben. Bei der Äquivalenz im engeren Sinne (im Gegensatz zur „einseitigen“ Äquivalenz, auf die hier nicht eingegangen wird) besteht die Alternativhypothese, deren Verifizierung anhand der auszuwertenden Daten angestrebt wird, aus einem beidseitig beschränkten Bereich. Sei Θ ein Verteilungsparameter, der den interessierenden Effekt in den Grundgesamtheiten sinnvoll quantifiziert. Die Äquivalenzhypothese behauptet die Gleichheit von Θ mit einem geeigneten Sollwert Θ_0 „bis auf praktisch irrelevante Abweichungen“. Diese werden durch ein meistens symmetrisches Intervall um Θ_0 beschrieben, d.h. $(\Theta_0 - \varepsilon, \Theta_0 + \varepsilon)$ mit $\varepsilon > 0$ klein. Häufig ist $\Theta_0 = 0$. Beim Äquivalenztest steht die *Nullhypothese der Nichtäquivalenz* $H_0: \Theta \leq \Theta_0 - \varepsilon$ oder $\Theta \geq \Theta_0 + \varepsilon$ gegen die *Alternativhypothese der Äquivalenzannahme* $H_A: \Theta_0 - \varepsilon < \Theta < \Theta_0 + \varepsilon$.

Entscheidend ist die Wahl des Verteilungsparameters θ . Für zwei unverbundene Stichproben mit gleicher Varianz σ^2 empfiehlt sich, die standardisierte Mittelwertsdifferenz $\Theta := (\mu_1 - \mu_2) / \sigma^2$ zu wählen. Die andere Möglichkeit $\Theta := (\mu_1 - \mu_2)$ zu wählen, ist aus statistischer Sicht *grundsätzlich abzulehnen*, da die Festlegung eines Äquivalenzbereichs ohne Berücksichtigung der Varianz nicht sinnvoll ist.

Das Anwendungsbeispiel

Der Sachverhalt wird am folgenden Beispiel einer Bioäquivalenzstudie mit zwei Formulierungen von Nifedipin an Daten aus [6] demonstriert. Es handelt sich um einen 2-Perioden-Crossover-Versuch. Dabei erhalten die zufällig ausgewählten Probanden der ersten Gruppe zuerst das Test- und dann das Referenzpräparat, die Probanden der zweiten Gruppe erst Referenz- und dann Testpräparat. Nach Gabe des Präparates wird zu festgelegten Zeitpunkten die Serumkonzentration des Wirkstoffes gemessen. Die Fläche (AUC) unter dem Polygonzug der Meßpunkte dient als Maß für die Bioverfügbarkeit. Die AUC-Werte A_i und B_i sind in Tabelle 1 dargestellt.

Das vorgeschlagene parametrische Modell wird dabei in der Regel angewandt. Es geht davon aus, daß der Logarithmus des Periodenquotienten normalverteilt ist, d.h. $X_i = \ln(A_{i1}/A_{i2}) \sim N(\mu_1, \sigma^2)$ mit $\mu_1 = \ln(\mu_T / \mu_R) + \ln(\pi_1/\pi_2)$. Analoges wird für die zweite Gruppe vorausgesetzt, d.h. $Y_i = \ln(B_{i1}/B_{i2})$ sei normalverteilt $Y_i \sim N(\mu_2, \sigma^2)$ mit $\mu_2 = \ln(\mu_R / \mu_T) + \ln(\pi_1/\pi_2)$. μ_T ist dabei der Effekt des Testpräparates, μ_R der Effekt der Referenz und die π_i sind die Periodeneffekte. Zur Verdeutlichung des Sachverhaltes wurden zwei Werte geeignet verändert - wie in Tab. 1 angegeben.

¹ Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig

Tab. 1: Rohdaten zu einer Bioäquivalenzstudie über den Calciumantagonisten Nifedipin

i	Gruppe 1				Gruppe 2		
	$A_{i1}(\text{Test})$	$A_{i2}(\text{Ref.})$	$X_{i\text{org}} = \ln(A_{i1}/A_{i2})$	X_i	$B_{i1}(\text{Ref.})$	$B_{i2}(\text{Test})$	$Y_i = \ln(B_{i1}/B_{i2})$
1	106,9	112,9	-,05461	-,35461	217,3	195,2	,10725
2	131,3	124,4	,05398	-,25398	174,4	122,7	,35161
3	81,4	89,5	-,09486		155,8	188,2	-,18893
4	154,7	134,9	,13695		299,5	309,2	-,03187
5	111,2	108,3	,02643		157,6	153,5	,02636
6	85,8	94,0	-,09128		121,4	104,7	,14799
7	295,2	418,6	-,34926		143,9	119,3	,18748
8	217,0	207,0	,04718		157,0	146,8	,06717
9	252,3	239,3	,05290		114,5	138,2	-,18813
10	157,9	207,3	-,27221		71,0	70,3	,00991
			$\bar{X}_{\text{org}} = -0,05448$	$\bar{X} = -0,1153$			
			$S_X = 0,1542$	$S_X = 0,181$			
					$\bar{Y} = 0,04888$		
					$S_Y = 0,1652$		

t-Test

Die Ergebnisse eines normalen t-Tests sind in Tab. 2 angegeben und zeigen, daß sich die Behandlungen bei 5% Irrtumswahrscheinlichkeit statistisch gesichert ($p = 0,049$) unterscheiden, d.h. das Testpräparat ist schlechter!

Äquivalenztest für $\Theta = \mu_1 - \mu_2$

Für μ_T / μ_R ist ein 80-120% Äquivalenzbereich üblich: $0,8 < \mu_T / \mu_R < 1,2$. Da $\Theta := (\mu_1 - \mu_2) = \ln(\mu_T / \mu_R) + \ln(\pi_1 / \pi_2) - [\ln(\mu_R / \mu_T) + \ln(\pi_1 / \pi_2)] = 2 \ln(\mu_T / \mu_R)$ ist, folgt aus obiger Ungleichung als Äquivalenzbereich für Θ : $-0,4463 < \Theta < 0,3646$.

Nach dem **Intervallinklusionstest** (siehe[6]) sind 2 Behandlungen äquivalent, wenn das $(1-2\alpha)$ -Konfidenzintervall im Äquivalenzbereich liegt. Das 90% Konfidenzintervall $(-0,299; -0,030)$, das in Tab. 2 angegeben ist, liegt im Äquivalenzbereich, so daß bei Benutzung von $\Theta = (\mu_1 - \mu_2)$ Äquivalenz vorliegt, obwohl sich beide Behandlungen signifikant unterscheiden.

Äquivalenztest für $\Theta := (\mu_1 - \mu_2) / \sigma$

Für einen ÄQ-Test zu einer Testgröße $T(x)$ ergibt sich nach [6] als kritische Region:

$$\{ x \mid C_1 < T(x) < C_2 \}.$$

Im Beispiel ist die Testgröße $T = -2,12$ aus Tab. 2 entnehmbar. Die Freiheitsgrade der F-Verteilung sind 1 und $2 \cdot 10 - 2 = 18$. Mit Nichtzentralitätsparameter $NC = (10^2/20) 1^2 = 5$ für $\varepsilon = 1$ ergibt sich $C_{10,10}(0,05; 1) = 0,61365 < 2,12$, indem man in einem SAS_Datenschritt $C = \text{sqrt}(\text{finv}(.05, 1, 18, 5))$ schreibt, sowie für $\varepsilon = 1,5$: $C_{10,10}(0,05; 1,5) = 1,66969 < 2,12$. Die Behandlungen sind also selbst für $\varepsilon = 1,5$ nicht äquivalent.

Abschließend soll die Wahl von ε untersucht werden. Da $X_i \sim N(\mu_1, \sigma^2)$ und $Y_i \sim N(\mu_2, \sigma^2)$ beide normalverteilt sind, ist die Differenz auch normalverteilt mit doppelter Varianz ($X_i - Y_i \sim N(\mu_1 - \mu_2, 2\sigma^2)$), woraus $P(X_i > Y_i) = \Phi((\mu_1 - \mu_2)/(\sigma\sqrt{2}))$ folgt. Dividiert man den ÄQ-Bereich $K: -\varepsilon < (\mu_1 - \mu_2)/\sigma < \varepsilon$ durch $\sqrt{2}$ und geht über die Ungleichungen mit der monoton wachsenden Standardnormalverteilung Φ , so ergibt sich ein neuer ÄQ-Bereich:

$1/2 - \varepsilon' < P(X_i > Y_i) < 1/2 + \varepsilon'$ mit $\varepsilon' = \Phi(\varepsilon/\sqrt{2}) - 1/2$ für obige Wahrscheinlichkeit. Keinem Unterschied entspricht die Wahrscheinlichkeit $1/2$, deshalb ist ε' so normiert, daß es die Abweichung von $1/2$ angibt. Aus der Tab. 3 ist ersichtlich, daß ε keinesfalls größer als 1 sein sollte, denn das bedeutet schon, daß die Wahrscheinlichkeit statt bei $1/2$ zwischen $1/4$ und $3/4$ liegen kann.

Schlußfolgerungen

Auf die Problematik von ÄQ-Testen für $\Theta := (\mu_1 - \mu_2)$ wurde in [5,6] und auch von anderen Autoren hingewiesen. Das ausgeführte Beispiel belegt meiner Ansicht nach überzeugend, daß in der Regel auf unstandardisierte ÄQ-Teste verzichtet werden muß. Eine Ausnahme kann höchstens dann gemacht werden, wenn der Äquivalenzbereich für einen Lageparameter dadurch definiert wird, daß kleinere Unterschiede aus fachwissenschaftlicher Sicht unbedeutend sind. Dann testet der Intervallinklusionstest auf Übereinstimmung der Lageparameter jedoch nicht auf Übereinstimmung der zugrunde liegenden Verteilungen. Die in [3] vorgeschlagene Kombination eines einseitigen Tests auf Unterschiede, der ein schlechteres Testpräparat aufdeckt, und eines einseitigen ÄQ-Tests ist beachtenswert. Allerdings wird in [3] für den ÄQ-Test keine Standardisierung vorgenommen. Eine Alternative zur Standardisierung liegt meiner Meinung nach nur in zwei simultanen Äquivalenztests für Lage- und Dispersionsparameter, wie kürzlich in [1] vorgeschlagen.

Literatur

1. Bauer, P.; Bauer, M. M.: Testing equivalence simultaneously for location and dispersion of two normally distributed populations: *Biom. J.* 36, 1994, 643-660.
2. Hauschke, D.; Steinijans, V. W.; Diletti, E.: A distribution-free procedure for the statistical analysis of bioequivalence studies: *Int. J. of Clinical Pharmacology, Therapy and Toxicology*, 28, 1990, 72-78.
3. Hilgers, R. A.: Tests auf „therapeutische Äquivalenz“. In: Pöpl, S. J. u.a. (Hrsg): *Medizinische Informatik*. München: MMV Medizin 1994.
4. Munk, A.: An improvement on commonly used tests in bioequivalence assessment: *Biometrics*, 49, 1993, 1225-1230.
5. Wellek, S.: Basing the analysis of comparative bioavailability trials on an individualized statistical definition of equivalence: *Biom. J.* 35, 1993, 47-55.
6. Wellek, S.: *Statistische Methoden zum Nachweis von Äquivalenz*. Stuttgart: Gustav Fischer 1994.