



# Wie liest man klinische Studien? –

## P-Wert und Konfidenzintervall



### Klinisches Szenario

In einer doppel-blinden randomisierten Studie wurde der Effekt der Verwendung der quartären Ammoniumbase Sanguinarin in Zahnpasta und Mundspüllösung bei Patienten mit adulter Parodontitis (Parodontaltaschen mit Sondierungstiefe  $\geq 5$  mm) untersucht [9]. Vor der Randomisierung wurde bei allen Patienten eine Vorbehandlung der Wurzeloberfläche (inkl. Wurzelglättung) und eine zweiwöchige Behandlung mit Chlorhexidin (0,2%, 2x/d) durchgeführt.

An diese Anfangsphase schloss sich eine zwölfwöchige Therapiephase an, in der ein Patient zur Zahnpflege entweder eine sanguinarinhaltige Zahnpasta und eine sanguinarinhaltige Mundspüllösung oder Zahnpasta und Mundspüllösung ohne Sanguinarin verwendete. Der Therapieerfolg bezüglich der Gingivitis wurde anhand des Gingival-Index nach Loe und Silness [4] beurteilt.

Grad	Klinisches Bild	Interpretation [nach 5]	
		GI-Werte	Gingivitisgrad
0	Normale Gingiva, keine Entzündung, keine Verfärbung, keine Blutung	0	Keine
1	Leichte Entzündung, leichte Farbveränderungen, leichte Schwellung, keine Blutung bei Palpation	0,1-1	Mild
2	Mäßige Entzündung, Rötung, Schwellung, Blutung bei Abstreichen mit der Parodontalsonde	1,1-2	Mäßig
3	Schwere Entzündung, ausgeprägte Rötung und Schwellung, Tendenz zur Spontanblutung, evtl. Ulzerationen	2,1-3	Schwer

**Tabelle 1** Gingival-Index (GI) nach Loe und Silness [4], und Interpretation nach Meyle [5]

Im Ergebnisabschnitt der Zusammenfassung findet sich die Aussage, dass durch die Verwendung von sanguinarinhaltiger Zahnpasta und Mundspüllösung die erneute Entstehung von Zahnfleischentzündungen signifikant verhindert werden konnte. Dem entsprechend wird in den Schlussfolgerungen die Kombination einer zweiwöchigen Gabe von Chlorhexidin und der dreimonatigen Verwendung von Sanguinarin als Therapie vorgeschlagen. Sollte man aufgrund dieses Ergebnisses bei Patienten mit Erwachsenenparodontitis die Verwendung sanguinarinhaltiger Zahnpasta und Spülung propagieren?

Eine grundlegende Voraussetzung für die Umsetzung der Resultate von Tenenbaum und seinen Straßburger Kollegen [9] in die klinische Praxis ist die interne Validität der Studienergebnisse, d. h. der unverfälschten Schätzung des Therapieeffekts in der Studie. Die detaillierten Angaben zum Studiendesign und zur Studiendurchführung lassen den Schluss zu, dass die interne Validität der Ergebnisse gegeben ist.

Die Art der Ergebnisdarstellung macht es jedoch schwierig, einen zweiten wichtigen Aspekt, die praktische Relevanz der Studienergebnisse, zu beurteilen. In Empfehlungen zur Erstellung von biomedizinischen Manuskripten [3, 7] und

zur Berichterstattung randomisierter Studien [6] findet man den Hinweis, nicht P-Werte, sondern Konfidenzintervalle zur Darstellung der (statistischen) Ergebnisse zu verwenden. Zur Darstellung und Diskussion der Ergebnisse des Gingival-Indexes nach Loe und Silness [4] werden in der Arbeit von Tenenbaum et al. [9] jedoch ausschließlich P-Werte herangezogen. Auf die Vorteile der Verwendung von Konfidenzintervallen wird im Folgenden näher eingegangen.

### P-Wert

Das Ergebnis eines statistischen Signifikanztests wird gewöhnlich in einem P-Wert zusammengefasst. In Tabelle 2 von Tenenbaum et al. (1999) ist für den Gingival-Index nach 14 Wochen (inkl. 2 Wochen Anfangsphase) ein P-Wert  $< 0,01$  angegeben (siehe Tabelle 2).

Die diesem Signifikanztest zugrundeliegende Nullhypothese kann in der Aussage „Die zwölfwöchige Verwendung einer sanguinarinhaltigen Zahnpasta und Mundspüllösung hat keinen Effekt auf den mittleren Gingival-Index“ zusammengefasst werden. Der P-Wert gibt die Wahrscheinlichkeit wieder, in einer Studie das aktuelle bzw. ein noch extremeres Ergebnis zu beobachten, falls die Nullhypothese wahr ist. Je kleiner der P-Wert ist, desto stärker spricht dies gegen die Nullhypothese. In weniger als 1 von 100 Fällen ( $P < 0,01$ ) würde somit der in der Studie von Tenenbaum et al. [9] beobachtete oder ein stärkerer Unterschied im mittleren Gingival-Index eintreten, falls die zwölfwöchige Verwendung einer sanguinarinhaltigen Zahnpasta und Mundspüllösung tatsächlich keinen Effekt auf den mittleren Gingival-Index hätte. Das Ergebnis der Studie spricht somit stark für einen Effekt von Sanguinarin. Anhand der ebenfalls in Tabelle 2 angegebenen Mittelwerte (Sanguinarin-Gruppe: 1,42; Kontrollgruppe: 1,61) kann man erkennen, dass Sanguinarin einen positiven Effekt auf den Gingival-Index hat; allerdings liegen beide Mittelwerte inmitten des Bereichs, dem ein mäßiger Gingivitisgrad entspricht (Tab. 1).

**„Confidence intervals, not P values.“**

Kopfzeile in einem Statistiklehrbuch aus Großbritannien [2]

### Konfidenzintervall

Der Behandlungseffekt nach der zwölfwöchigen Verwendung von sanguinarinhaltiger Zahnpasta und Mundspüllösung kann in der Differenz der mittleren Gingival-Indizes zusammengefasst werden:  $1,42 - 1,61 = -0,19$ . Im Mittel kann der Gingival-Index nach 12 Wochen somit um 0,19 Punkte verringert werden. Diese Aussage basiert auf einer einzelnen Studie mit insgesamt 60 Patienten. Es ist davon auszugehen, dass bei Durchführung einer weiteren Studie ein anderer Wert für den Behandlungseffekt geschätzt würde. Eine Möglichkeit diese Ungenauigkeit zu quantifizieren, besteht in der Berechnung eines Konfidenzintervalls.



	Studienbeginn mit Chlorhexidin		Beginn Therapiephase		6. Woche Therapiephase		12. Woche Therapiephase	
	Ø	(SD)	Ø	(SD)	Ø	(SD)	Ø	(SD)
Experimentelle Gruppe (n= 31)	2,10	(0,27)	1,38	(0,21)	1,41	(0,22)	1,42	(0,23)
Vergleichsgruppe (n= 29)	2,09	(0,28)	1,39	(0,17)	1,58	(0,21)	1,61	(0,19)
Unterschiede zwischen den Gruppen in %	<1		<1		11		12	
P-Wert (zwischen den Gruppen)	NS		NS		<0,01		<0,01	

**Tabelle 2** Mittelwerte (Ø) des Gingival-Index nach Løe und Silness [4] (Standardabweichungen [SD] in Klammern) und prozentuale Unterschiede zwischen den Studiengruppen. Die Angaben entsprechen denen der Tabelle 2 der Arbeit von Tenenbaum et al. [9]). NS: nicht signifikant

Ein Konfidenzintervall wird derart konstruiert, dass der wahre, unbekannte Behandlungseffekt mit einer vorgegebenen Wahrscheinlichkeit im Konfidenzintervall enthalten ist. Die Wahrscheinlichkeit, dass der wahre, unbekannte Therapieeffekt in einem 95%-Konfidenzintervall enthalten ist, ist dementsprechend 95%. Zur Berechnung eines 95%-Konfidenzintervalls für die Differenz der mittleren Gingival-Indizes wird zusätzlich die Anzahl von Patienten und die Standardabweichung in den beiden Behandlungsgruppen benötigt [8]; diese Angaben sind ebenfalls in Tabelle 2 der Publikation von Tenenbaum et al. [9] enthalten. Das 95%-Konfidenzintervall berechnet sich zu [-0,299; -0,081]; alle Werte in diesem Bereich sind plausible Schätzungen für den Behandlungseffekt nach zwölfwöchiger Behandlung mit Sanguinarin. Im günstigsten Fall kann der mittlere Gingival-Index nach 12 Wochen somit um 0,299 Punkte gesenkt werden; ein Senkung um lediglich 0,081 Punkte ist aber auch denkbar. Man kann jedoch durchaus behaupten, dass im vorliegenden Fall selbst eine Senkung um 0,3 im Gingival-Index nach Løe und Silness [4] klinisch kaum relevant ist.

## Schlussbemerkungen

In obigem Beispiel zeigt sich deutlich der Unterschied zwischen statistischer Signifikanz und klinischer Relevanz, der von Barnett [1] in treffender Weise als Tyrannei des P-Wertes bezeichnet wurde. Die klinische Relevanz eines Studienergebnisses kann nicht ausschließlich anhand des Ergebnisses eines Signifikanztests, d.h. des P-Wertes, beurteilt werden. Eine sinnvolle Zusammenfassung des Unsicherheit eines Studienergebnisses ist durch die Angabe eines Konfidenzintervalls möglich. Die Angabe eines genauen P-Wertes als zusätzliche Information ist sinnvoll, wohingegen eine bloße Unterteilung in signifikante ( $P < 0,05$ ) und nicht signifikante Ergebnisse ( $P > 0,05$ ) vermieden werden sollte. In der Arbeit von Tenenbaum et al. [9] geht die ungenaue, kategorisierende Unterteilung von P-Werten sogar noch weiter: „NS“,  $P < 0,0001$ ,  $P < 0,01$ ,  $P < 0,02$ ,  $P > 0,20$  und schließlich sogar  $P < 0,35$ . Lediglich die Angabe  $P < 0,0001$ , die häufig als Untergrenze zur Angabe von exakten P-Werten verwendet wird, erscheint sinnvoll. In allen anderen Fällen sollten die P-Werte mit ausreichender Genauigkeit berichtet werden.

## Literatur

1. Barnett, M.L., Mathisen, A.: Tyranny of the p-value: the conflict between statistical significance and common sense. *J Dent Res* 76, 534 (1997).
2. Gardner, M.J., Altman, D.G.: Confidence intervals rather than P values. In: Altman, D.G., Machin, D., Bryant, T.N., Gardner, M.J. (Hrsg.): *Statistics with Confidence*. Second edition. BMJ Books, Bristol 2000.
3. International Committee of Medical Journal Editors: Uniform requirements for manuscripts submitted to biomedical journals. *N Engl J Med* 336, 309 (1997).
4. Løe, H., Silness, J.: Periodontal disease in pregnancy: prevalence and severity. *Acta Odontol Scand* 21, 532 (1963).
5. Meyle, J.: Indizes. In: Heidemann, D. (Hrsg.): *Parodontologie (Praxis der Zahnheilkunde, Band 4)*. 3 Auflage. Urban & Schwarzenberg, München 1997, 53.

6. Moher, D., Schulz, K.F., Altman, D.G.: The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *Ann Intern Med* 134, 657 (2001).
7. Newcombe, R.G.: The controversy over how to present research findings. *Br J Orthod* 26, 233 (1999).
8. Newcombe, R.G.: Statistical Applications in Orthodontics. Part I. Confidence Intervals: an introduction. *J Orthod* 27, 270 (2000).
9. Tenenbaum, H., Dahan, M., Soell, M.: Effectiveness of a sanguinarine regimen after scaling and root planing. *J Periodontol* 70, 307 (1999).

Guido Schwarzer, Freiburg  
Jens C. Türp, Freiburg / Basel  
Gerd Antes, Freiburg