

Normwerte und Zuverlässigkeit  
der Messung von Kälte-, Wärme- und  
Hitzeschmerz-Schwellen für die  
Funktionsprüfung kleiner Nervenfasern

Andreas Möltner<sup>1</sup>, Rupert Hölzl<sup>1</sup>,  
Stefan Lautenbacher<sup>2</sup>, Stephan Roscher<sup>2</sup>  
und Friedrich Strian<sup>2</sup>

F O R S C H U N G S B E R I C H T E

AUS DEM

OTTO - SELZ - INSTITUT

FÜR

PSYCHOLOGIE UND ERZIEHUNGSWISSENSCHAFT

DER

UNIVERSITÄT MANNHEIM (WH)

Normwerte und Zuverlässigkeit  
der Messung von Kälte-, Wärme- und  
Hitzeschmerz-Schwellen für die  
Funktionsprüfung kleiner Nervenfasern

Andreas Möltner<sup>1</sup>, Rupert Hölzl<sup>1</sup>,  
Stefan Lautenbacher<sup>2</sup>, Stephan Roscher<sup>2</sup>  
und Friedrich Strian<sup>2</sup>

<sup>1</sup>FG Klinische Psychophysiologie, Universität Mannheim  
<sup>2</sup>Max-Planck-Institut für Psychiatrie, München

Forschungsbericht Nr. 23  
1993

Gefördert aus Mitteln der  
*Deutschen Forschungsgemeinschaft*  
(Projekte Str 273/1 und Ho 904/4)  
und der  
*Max-Planck-Gesellschaft.*



## ZUSAMMENFASSUNG

Zunächst werden die Begriffe Normbereiche, Toleranzgrenzen, Zuverlässigkeit und Reproduzierbarkeit von diagnostischen Verfahren unter biometrischen Gesichtspunkten analysiert. Daraus leitet sich ein methodisch begründeter und praktisch gangbarer Weg der Evaluation thermästhetischer Neuropathie-Prüfungen ab. Dies wird in einer empirischen Normwert- und Zuverlässigkeitsstudie mit einem kommerziellen Thermästhesiegerät angewendet. Bei 64 gesunden Probanden wurden die Warm- und Kaltschwellen und die Hitzeschmerz-Schwellen an Hand und Fuß bestimmt. Daraus wurden alters- und körpermaßkorrigierte Normbereiche und deren 90%-Vertrauensintervalle ermittelt. Aus der Streuung der Einzeldurchgänge und wiederholten Meßterminen bestimmten sich die interne Reliabilität des Meßverfahrens, seine Wiederholungsstabilität und absolute Reproduzierbarkeit. Die interne Reliabilität (Cronbachs  $\alpha$ ) der thermozeptiven Schwellen ist durchweg sehr hoch ( $\alpha \geq 0.95$ ). Die Wiederholungsstabilität nach 1 Stunde war etwas niedriger, jedoch noch befriedigend ( $r \geq 0.68$ ). Die absolute Reproduzierbarkeit der verschiedenen Schwellenwerte war unterschiedlich, jedoch durchweg gut. Die Untersuchung zeigte ferner, daß die Voraussetzungen der üblichen „Variationskoeffizienten“ im allgemeinen nicht erfüllt sind und andere Reproduzierbarkeitsmaße angewendet werden müssen.

## SUMMARY

Starting from a biometric discussion of normal range, tolerance limits, reliability, and reproducibility of diagnostic tools, a statistically founded and practical approach to evaluate thermesthetic testing for small-fibre neuropathy is derived. This approach is applied in an empirical study of normal ranges and reliabilities of thermal sensitivity measures obtained with a commercial thermoceptive stimulator. 64 healthy subjects were tested for warm, cold and heatpain thresholds at hands and feet. From these data normal ranges corrected for age and body measures as well as corresponding 90%-confidence intervals were estimated. Internal reliabilities (Cronbach's  $\alpha$ ), retest stabilities (after 1 h), and absolute reproducibilities of thermal thresholds were also calculated. Internal reliability is very high ( $\alpha \geq 0.95$ ), while retest reliability is only moderate, but still satisfactory ( $r \geq 0.68$ ). Absolute reproducibilities vary with type of threshold, but were also sufficiently high for practical purposes. Results showed that the assumptions of conventional "coefficients of variation" for reproducibility estimation are often violated. This emphasizes the need for alternative measures of reproducibility.

## 1 Einleitung

### 1.1 Thermozeptive Kennwerte in der Funktionsprüfung kleiner Nervenfasern

Die Quantifizierung der Temperatur- und Hitzeschmerzsensibilität wird wegen ihrer einfachen Durchführung und der geringen Belastung des Patienten zunehmend in der neuropathologischen Diagnostik kleinkalibriger Nervenfasern ( $A\delta$  und C) eingesetzt (z.B. Guy et al. 1985). Hierfür sind inzwischen auch klinisch erprobte Systeme im Handel (z.B. der „PATH-Tester MPI 100“ ®; vgl. Galfe et al. 1990), auf denen die üblichen psychophysikalischen Verfahren zur Bestimmung thermozeptiver Schwellen realisierbar sind (Grenzwert- und Konstantreizmethode). Bei der genannten Anwendung (wie auch bei anderen klinischen Fragestellungen) stellt sich allerdings heute noch immer das zweifache Problem geeigneter Normwerte und des Nachweises hinreichender Stabilität der gewonnenen diagnostischen Kennwerte.

### 1.2 Korrigierte Normbereiche und Toleranzgrenzen

Die Erstellung von Normtabellen zur Korrektur von Alterseinflüssen und anderen Drittvariablen wie Körpermaße (Größe, Gewicht) auf der Basis von Normdaten für unterschiedliche Alters-, Gewichts- und Größenklassen erfordert unrealistisch hohe Fallzahlen. Daher wurde in der Literatur (Claus et al. 1990, Lautenbacher et al. 1988<sup>1</sup>) ein regressionsanalytisches Vorgehen gewählt, das mit wesentlich kleineren Normstichproben auskommt, wenn auch um den Preis schärferer statistischer Voraussetzungen.

Jedoch liegen bisher keine Korrekturdaten vor, die außer dem Alter auch andere relevante Drittvariablen, vor allem Geschlecht, Körpergröße und Gewicht berücksichtigen. Deren zusätzlicher Einfluß wurde in einer vorausgehenden Studie nachgewiesen (Lautenbacher & Strian 1991). Dabei zeigten multiple Regressionsanalysen, daß der Einfluß dieser zusätzlichen Drittvariablen allein durch die Variation in einem der beiden (untereinander hochkorrelierten) Körpermaße erklärt werden kann und die Geschlechtsunterschiede lediglich auf Körpermaßunterschiede zurückgehen. Dieser Einfluß muß also in Normtabellen bzw. Nomogrammen berücksichtigt werden. Im Prinzip kann das mit einer Erweiterung des genannten Regressionsansatzes geschehen, sofern seine statistischen Voraussetzungen erfüllt werden können.

Außer wegen der Beschränkung auf die Alterskorrektur sind die von Claus und Mitarbeitern verwendeten „Normbereiche“ auch statistisch problematisch: Die 90%- bzw. 95%-Perzentile um die Regression der Temperaturschwellen auf das Lebensalter erlauben keine Aussage über die Wahrscheinlichkeit, mit der ein konkreter Meßwert außerhalb des tatsächlichen Normbereichs der Grundbevölkerung liegt. Der Normbereich der Population ist nämlich nicht bekannt und muß aus den Daten der Normstichprobe mit einer bestimm-

---

<sup>1</sup>Lautenbacher S., Galfe G., Möltner A. & Strian F. (1988). Alterskorrigierte Normdaten für die Bestimmung der Temperatur- und Schmerzempfindlichkeit mit dem MPI PATH-Tester 100. Unveröffentlichte Beilage zum Manual des MPI PATH-Tester 100.

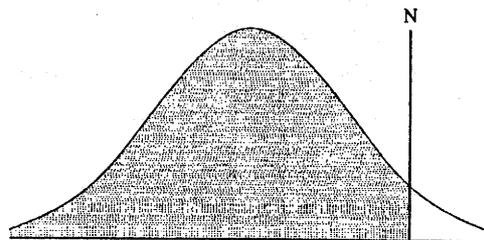
ten Vertrauenswahrscheinlichkeit geschätzt werden (zur Problematik der Definition von Normbereichen für die klinische Diagnostik siehe bes. Ackermann 1985).

Schließlich darf bezweifelt werden, ob die Normalverteilungsannahme, die der regressionsanalytischen Konstruktion von Normbereichen bei Claus et al. (1990) zugrundeliegt, für thermozeptive Schwellen überhaupt erfüllt ist (vgl. Abschnitt 3.1). Im allgemeinen sind auch die Streuungen der Schwellenverteilungen auf verschiedenen Altersstufen nicht homogen. In diesem Fall würde eine weitere Voraussetzung des scheinbar einfachen Ansatzes von Claus et al. (1990) verletzt. Tatsächlich müßte sich dann die Breite des Normbereichs mit dem Alter (z.B. trichterförmig bei zunehmender Streuung) verändern. Prüfungen der genannten Voraussetzungen werden in der Arbeit jedenfalls nicht berichtet.

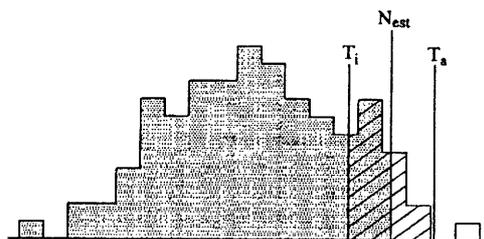
Auch die in Lautenbacher et al. (1988)<sup>1</sup> wiedergegebenen Normwerte sind nur alterskorrigiert. Die dort verwendete Definition einseitiger Normbereiche auf der Basis sog. „äußerer Toleranzgrenzen“ stellt jedoch eine erste Verbesserung dar. Diese geben nämlich mit einer bestimmten Vertrauenswahrscheinlichkeit, z.B. 90%, an, ob ein gegebener Normbereich, z.B. der 95%-Bereich der Population, *unterhalb* dieser Grenzen liegt. Die äußeren Toleranzgrenzen geben jedoch keine Auskunft über das Konfidenzintervall, in dem die Grenzen des Normbereichs liegen. Dazu sind zusätzlich die entsprechenden „inneren Toleranzgrenzen“ erforderlich. Das sind die Grenzen, *oberhalb* deren der betreffende Normbereich (95% der Population) mit der gewählten Vertrauenswahrscheinlichkeit (hier: 90%) liegt (vgl. Abb. 1 und Anhang A1).

Abbildung 1: Veranschaulichung der Begriffe Normbereichs- und Toleranzgrenze.

Die obere Abbildung zeigt die Verteilung eines Parameters mit der einseitigen oberen Normbereichsgrenze  $N$  (z.B. für  $P = 95\%$ ), d.h.  $P\%$  der Verteilungsmasse sind kleiner und  $(1 - P)\%$  größer als  $N$ .



In der unteren Abbildung ist die empirische Verteilung in einer Stichprobe dargestellt. Aus der Stichprobe werden eine Schätzung  $N_{est}$  der Normbereichsgrenze  $N$  sowie mit einer vorgegebenen Vertrauenswahrscheinlichkeit (z.B. 90%) innere und äußere Toleranzgrenzen  $T_i$  bzw.  $T_a$  gewonnen.



Erst beide Angaben gestatten es, für einen bestimmten Patienten eine wohldefinierte Wahrscheinlichkeit anzugeben, mit der er außerhalb des „gesunden“ Normbereichs in Richtung pathologischer Veränderungen (hier: Schwellenerhöhungen) liegt; genau das aber

wird von der üblichen diagnostischen Anwendung der Thermästhesie und ähnlicher Funktionstests verlangt.

Diese Arbeit legt daher zunächst innere und äußere Toleranzgrenzen für die alters- und körpermaßkorrigierten einseitigen Normbereiche thermozeptiver Kennwerte auf der Basis eines multiplen Regressionsansatzes vor. Dabei wird auch auf die Einhaltung der dafür geltenden statistischen Voraussetzungen geachtet.

### 1.3 Aspekte der „Zuverlässigkeit“

Eine Grundvoraussetzung der Brauchbarkeit jedes diagnostischen Verfahrens ist seine befriedigende „Zuverlässigkeit“. Diese hat verschiedene Aspekte, die durch ganz unterschiedliche Testmethoden und Kenngrößen zu bewerten sind. Außerdem wird meist nicht zwischen Zuverlässigkeit im Sinn der Korrelation von wiederholten Messungen und der absoluten Reproduzierbarkeit von Meßwerten unterschieden. Im allgemeinen wird nur ein ganz bestimmter Zuverlässigkeitsaspekt geprüft und berichtet, meist ohne Angabe von Gründen für seine Auswahl. Im folgenden werden diese Aspekte unter dem Gesichtspunkt ihrer Bedeutung für die diagnostische Anwendung der Thermozeptionsmessung und die gegenwärtige Fragestellung kurz diskutiert.

#### 1.3.1 Reliabilität und Stabilität

Es gibt zwei Arten von Zuverlässigkeit im Sinn von Variabilität der Messungen, die in der testtheoretischen Literatur als „Reliabilität“ und „Stabilität“ bezeichnet werden. Unter „Reliabilität“ wird dabei die Zuverlässigkeit des Meßverfahrens selbst verstanden, während „Stabilität“ die Stabilität der tatsächlichen Meßwerte über die Zeit bezeichnet, die außer vom Fehler der Messung noch von der Variabilität des biologischen Merkmals (hier: der thermozeptiven Sensibilität) abhängt und mit ersterem nicht verwechselt werden darf.

Thermozeptive Kennwerte sind einerseits vom allgemeinen Funktionszustand des sensorischen Systems (z.B. andauernde Schwellenerhöhung bei Neuropathie), andererseits von aktuellen Faktoren abhängig (z.B. Aufmerksamkeit, Medikamenteneinfluß). Zu diesen letzten Faktoren gehören auch systematische, biologisch bedingte temporäre Veränderungen wie gewisse Schwankungen im Tagesverlauf (Strian et al. 1989). Unter Zuverlässigkeit im Sinne von „Reliabilität“ versteht man aber die Korrelation der empirischen Meßwerte zu verschiedenen Zeitpunkten bei gleichbleibenden „wahren“ Werten. Da sich die wahren Werte bei tatsächlichen Schwankungen des biologischen Merkmals „thermozeptive Sensibilität“ zwischen verschiedenen Meßzeitpunkten geändert haben, gibt die Korrelation zwischen diesen nur eine untere Grenze für die tatsächliche Reliabilität an (Lord & Novick, 1968). Im Rahmen der Neuropathiediagnostik kleiner Nervenfasern wirken sich die Schwankungen allerdings schwächend auf die diagnostische Validität (Gültigkeit) des Verfahrens aus, nämlich auf seine Fähigkeit, neuropathische Schädigungen zu erkennen.

In den Arbeiten von Galfe et al. (1990), Lautenbacher (1990) und Claus et al. (1990) werden Korrelationen der thermozeptiven Kennwerte zwischen wiederholten Messungen in verschiedenen Zeitabständen angegeben. Die Retest-Korrelationen bei Galfe et al. (1990) beruhen auf Messungen, die innerhalb von ca. 1 Stunde stattfanden. Selbst bei so kurzen Abständen muß, z.B. wegen der allmählichen Abnahme der Konzentrationsfähigkeit der Testpersonen im Lauf vieler Durchgänge, mit einer nicht unwesentlichen Variation der Meßwerte gerechnet werden, die aber auf *Veränderungen des biologischen Merkmals* und *nicht auf verminderte Meßgenauigkeit* zurückgeht. In Galfe et al. (1990) wurden die Begriffe „short-time-“ oder „long-time reliability“ als Bezeichnungen für die Korrelationen der Kennwerte bei Messungen im Abstand von etwa einer Stunde bzw. einer Woche verwendet. Nach den bisherigen Ausführungen sollten sie besser durch den Begriff „Stabilität“ ersetzt werden. Aus den Ausführungen folgt auch, daß die bisher veröffentlichten „Reliabilitäten“ thermozeptiver Messungen zur Abschätzung ihrer tatsächlichen Zuverlässigkeit nur bedingt brauchbar sind.

Eine Reliabilitätsschätzung, die nicht von der Variabilität des biologischen Merkmals konfundiert ist, ergibt sich am ehesten aus dem Vergleich der Einzelbestimmungen, aus denen z.B. in den zitierten Untersuchungen mit dem MPI PATH-Tester die thermozeptiven Kennwerte als Durchschnitt berechnet werden (bei den Warm- und Kaltschwellen je 7, bei den phasischen und tonischen Schmerzschwellen 5 bzw. 6; vgl. Abschnitt 2.2). Diese Reliabilität innerhalb einer Messung (hier als „interne Reliabilität“ bezeichnet) läßt sich mit Standardverfahren der klassischen Testtheorie (Lord & Novick 1968) aus den Interkorrelationen der Einzelreizungen untereinander abschätzen. Entsprechende Ergebnisse werden weiter unten berichtet.

Diese Zerlegung hat noch einen praktisch verwertbaren Aspekt: Definiert man die Korrelation der „wahren“ Kennmaße zu verschiedenen Zeitpunkten als (Merkmals-) Stabilität, so setzt sich die Korrelation der empirisch gewonnenen Meßwerte, also die Stabilität der Messung, aus der Merkmalsstabilität und der Zuverlässigkeit der Messung (Reliabilität) zusammen. Umgekehrt läßt sich dann die Merkmalsstabilität aus der Stabilität der Messungen und ihrer Zuverlässigkeit (wie oben definiert) bestimmen. Dies kann u.U. für diagnostische Zwecke genutzt werden, wenn die thermozeptive Sensibilität selbst als Index des aktuellen Zustands eines Patienten verwendet werden soll (z.B. der Güte der Blutzucker-Einstellung beim Diabetiker).

### 1.3.2 Reproduzierbarkeit

#### a) Ausgangsüberlegungen:

Reliabilität und Stabilität (als Korrelationskoeffizienten) sind abhängig von der Varianz in der Stichprobe, also von deren Zusammensetzung. Je homogener die Stichproben (z.B. gesunde Personen im Alter von 20–30), desto geringer fallen die Korrelationen zwischen aufeinanderfolgenden Messungen aus. In der Praxis ist man aber meist an den tatsächlichen Abweichungen dieser Messungen voneinander, also der Reproduzierbarkeit der Messungen, interessiert. Dafür ist der varianzabhängige Korrelationskoeffizient nicht geeignet.

Zur Beurteilung der Reproduzierbarkeit thermozeptiver Kennwerte wie der Kalt-, Warm- und Schmerzschwellen verwendet man besser die absoluten oder relativen Differenzen aufeinanderfolgender Messungen. Da absolute Differenzen oft von der Höhe der Schwellen (den Ausgangswerten) selbst abhängen, nimmt man häufig relative Differenzen (in Prozent der Ausgangswerten) bzw. daraus abgeleitete Maße der Reproduzierbarkeit wie Variationskoeffizienten, so auch Claus et al. (1990) und Galfe et al. (1990). Dabei setzt man aber stillschweigend eine proportionale Abhängigkeit zwischen Meßwertdifferenzen und Meßwerten (im statistischen Mittel) voraus. Diese ist jedoch keineswegs immer gegeben. Es ist daher stets erforderlich, diesen Zusammenhang erst einmal zu analysieren und das gewählte Reproduzierbarkeitsmaß empirisch zu rechtfertigen. Da dies in der Literatur zur Thermästhesie-Diagnostik in der Regel vernachlässigt wird, sind im folgenden dazu einige grundlegende Beziehungen kurz dargestellt, die zum Vorgehen in dieser Studie führten.

b) Fehlermodell mit linearer Ausgangswertabhängigkeit:

Nimmt man statt Proportionalität nur einfache lineare Abhängigkeit zwischen Ausgangswert ( $x$ ) und mittlerem Betrag der Abweichung der zweiten Messung ( $y$ ) an, dann ist der mittlere Meßfehler in einen „absoluten“ ( $a$ ) und einen „relativen“ Anteil ( $b$ ) aufzuspalten:

$$\mathcal{E}\Delta_x = b \cdot x + a \quad (1)$$

$\mathcal{E}\Delta_x$  bedeutet den Erwartungswert der Abweichung  $|x - y|$  unter dem jeweiligen Ausgangswert  $x$ . Zusätzlich wird vorausgesetzt, daß die Verteilungsfunktionen der Meßfehler für verschiedene Ausgangswerte durch eine Streckung ineinander übergeführt werden können. Insbesondere stimmen die Verteilungen bis auf Unterschiede in Mittelwert und Streuung überein (Abb. 2). Die Koeffizienten  $a$  und  $b$  und die Verteilungen der individuellen Abweichungen (in Abhängigkeit vom Ausgangswert) liefern dann die Kennwerte der Reproduzierbarkeit des diagnostischen Verfahrens (s. Anhang A2).

*Fall 1:* Ist die Abhängigkeit zwischen Fehler und Ausgangswert direkt proportional (bei Verdopplung des Ausgangswerts verdoppelt sich auch der zu erwartende Fehler), so ist  $a = 0$  (vgl. Abb. 2c). Zur Beurteilung der Reproduzierbarkeit ist dann die Verteilung der individuellen relativen Abweichungen  $b_i$  oder  $b_i^*$  bzw. deren Kennmaße (Mittelwert, Streuung, Perzentile) geeignet:

$$b_i = \frac{|y_i - x_i|}{x_i} \quad (2)$$

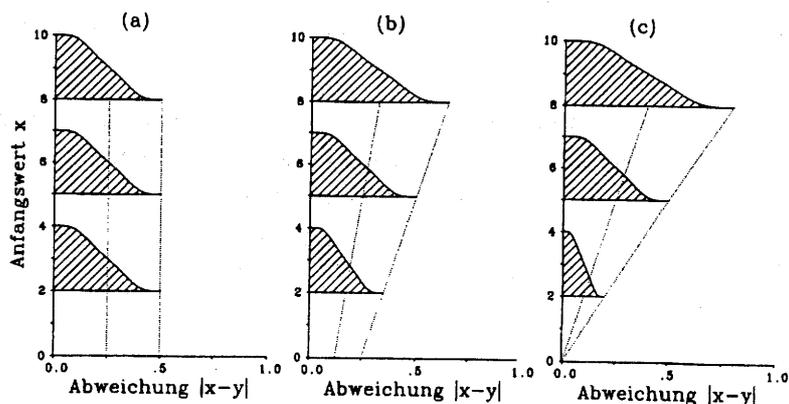
oder

$$b_i^* = b_i \cdot 100\% \quad (3)$$

( $x_i$  bzw.  $y_i$  Meßergebnisse der Person  $i$  zum 1. bzw. 2. Meßtermin)

Die „indices of variability“ bei Galfe et al. (1990) und die prozentualen Verhältnisse von Zweit- und Erstmessungen, die Claus et al. (1990) verwenden, hängen eng mit diesen relativen Differenzen zusammen. Für alle diese Maße gilt gleichermaßen, daß sie nur bei proportionaler Abhängigkeit von Ausgangswert und Meßfehler sinnvoll interpretiert werden können.

Abbildung 2: Beispiele für Abhängigkeiten der Verteilung der Abweichungen  $\Delta_x$  vom Ausgangswert  $x$  nach Gleichung (1):  $\mathcal{E}\Delta_x = bx + a$ . (a) Keine Abhängigkeit ( $b = 0$ ). (b) Allgemeiner Fall ( $b \neq 0, a \neq 0$ ). (c) Proportionale Abhängigkeit ( $a = 0$ ).



*Fall 2:* Sind die Abweichungen von Erst- und Zweitmessungen nicht von den Ausgangswerten abhängig, d.h. ist in Gleichung 1  $b = 0$ , sind die Verteilung der absoluten Abweichungen

$$\delta_i = |x_i - y_i| \quad (4)$$

bzw. deren Kennwerte (Mittelwert, Streuung, Perzentile) zur Bewertung der Reproduzierbarkeit hinreichend (Abb. 2a).

Absolute Fehler können unter anderem durch die begrenzte Auflösung des untersuchten sensorischen Kanals oder des Meßgeräts in den unteren Temperaturbereichen wie bei der Bestimmung der Kaltschwellen bedingt sein. Dann wird der Meßfehler für niedrigere Schwellen nicht kleiner als bei höheren. Auf dieser Überlegung beruht jedoch die Berechnung eines relativen Fehlers, weshalb er in diesem Fall kein sinnvolles Maß der Reproduzierbarkeit mehr darstellt. Hat z.B. eine Person in der Erstmessung eine Kaltschwelle von  $0.6^\circ\text{C}$ , bei der Zweitmessung den Wert  $1.2^\circ\text{C}$ , so entspricht das einem Relativfehler von 100%. Dieser Unterschied ist inhaltlich belanglos; der Relativfehler ist jedoch ebenso hoch wie die inhaltlich bedeutsame Differenz zwischen  $7^\circ\text{C}$  bei der Erst- und  $14^\circ\text{C}$  bei der Zweitmessung.

*Fall 3:* Falls sowohl ein relativer als auch ein absoluter Fehleranteil berücksichtigt werden muß ( $a \neq 0, b \neq 0$ , Abb. 2b), müssen die Abweichungen zwischen Zweit- und Erstmessung bezüglich des Ausgangswertes nach Formel 5 korrigiert werden:

$$k_i = \frac{|x_i - y_i|}{x_i + a/b} \quad (5)$$

Dabei sind  $a$  und  $b$  die Koeffizienten aus (1). Die Verteilung der  $k_i$  dient wie die der  $b_i$  im Fall 1 zur Beurteilung der Reproduzierbarkeit.

c) In der zitierten Arbeit von Galfe et al. (1990) wird das Problem der Skalenabhängigkeit von Reproduzierbarkeitsmaßen aufgeworfen: Auf Quotienten basierende Maße wie ihre „indices of variability“ und die Koeffizienten von Claus et al. (1990) sind nur für „Differenzschwellen“ (Schwellen der Temperaturänderung) von der Temperaturskala unabhängig, weil nur diese einen eindeutigen Nullpunkt haben und damit auch verhältnisskaliert sind. Nur dann sind aber die Quotienten invariant gegenüber Skalentransformationen. Bei Absolutschwellen wie im Fall der Schmerzschwelle, die als Temperatur (und nicht als Temperaturdifferenz) angegeben wird, gilt dies nicht mehr. Das Ergebnis ist, daß Reproduzierbarkeitsmaße, die auf relativen Abweichungen basieren, nicht mehr mit den (Differenz-) Kalt- und Warmschwellen verglichen werden können. Tatsächlich ergeben sich bei diesem Vorgehen unrealistisch hohe Reproduzierbarkeiten für die Schmerzschwellen, obwohl die Schwellen im Durchschnitt um 0.54°C an der Hand und 0.95°C am Fuß schwanken (vgl. Abschnitt 3.3).

Der von Galfe et al. (1990) eingeschlagene Weg der Transformation der Schwellenmaße in die Standardwerte (Z-Scores mit Mittelwert = 100 und Standardabweichung = 10)

$$Z_i = 10 \cdot \frac{x_i - \bar{x}}{s} + 100$$

erweist sich bei genauer Betrachtung jedoch als ungeeignet.

Erstens verwandelt eine lineare Transformation wie die vorliegende eine Intervallskala nicht in die benötigte Verhältnisskala, sondern zerstört sogar im ungünstigen Fall eine bestehende Verhältnisskalierung.

Zweitens wird die Skala populations- bzw. stichprobenabhängig. Dies kann verheerende Folgen haben, wenn verschiedene klinische Gruppen und Kontrollgruppen verglichen werden, die sich tatsächlich in Bezug auf die Streuung der wahren Werte und/oder der Meßfehler unterscheiden. Veränderte Varianzen und Fehlerstreuungen sind aber bei klinischen Gruppen eher die Regel als die Ausnahme. Dies gilt ganz besonders in den hier einschlägigen Anwendungsgebieten wie der Thermästhesie-Diagnostik von neuropathischen Veränderungen kleiner Nervenfasern. Durch die Z-Transformation wird also die Populationsabhängigkeit des Reproduzierbarkeitsmaßes wieder eingeführt, die man durch die Verwendung von „Variabilitätsscores“ statt von Korrelationskoeffizienten loswerden wollte.

Tatsächlich ist das Problem erst aus der Wahl ungeeigneter bzw. zu voraussetzungsstarker Reproduzierbarkeitsmaße entstanden. Das hier gewählte voraussetzungsärmere lineare Fehlermodell vermeidet die Schwierigkeit, da es von vornherein keine Verhältnisskalen voraussetzt. Stattdessen kann im allgemeinen Fall (Fall 3:  $a \neq 0$ ,  $b \neq 0$ ) die Skala der (Differenz- oder Absolut-) Schwellen auf einen neuen Nullpunkt bezogen und in eine Verhältnisskala transformiert werden. Diese Skalen bzw. die auf ihnen definierten relativen Fehler sind dann miteinander vergleichbar. Wie sich allerdings zeigen wird, löst sich das Problem im Fall der Schmerzschwellen bei empirischem Vorgehen schon dadurch auf, daß für sie keine Ausgangswertabhängigkeit zu finden ist (vgl. Abschnitt 3.3).

## 1.4 Zielsetzung und Fragestellung

Der obigen Analyse entsprechend hatte die vorliegende Studie folgende Ziele:

- 1) Anhand der Daten aus einer früheren Eichstichprobe (vgl. Galfe et al. 1990, Lautenbacher 1990) sollen alters- und körpermaßkorrigierte Norm- und Toleranzbereiche im oben definierten Sinn erstellt werden.
- 2) Bestimmung der „internen Reliabilität“ der Thermästhesie-Messung (Interkorrelationen der Einzelwerte innerhalb eines Durchgangs; vgl. Abschnitt 1.3.1) im Vergleich zur Stabilität der Schwellen über eine Stunde hinweg.
- 3) Überprüfung der Eignung und Bestimmung von absoluten und relativen Abweichungsmaßen aufeinanderfolgender Messungen als Maße der Reproduzierbarkeit im Sinn von Punkt 1.3.2.

## 2 Methodik

### 2.1 Normstichprobe

Die Normdaten entstammen der zitierten früheren Untersuchung von Galfe et al. (1990). Sie umfaßte 32 männliche und 32 weibliche Probanden im Alter von 15–65 Jahren, die sich gleichmäßig auf drei Altersklassen (15–29, 30–44, 45–65) verteilten. Pathologische Veränderungen der Somatosensorik aufgrund von Verletzungen, Neuropathien, radikulären Syndromen und dermatologischen Erkrankungen wurden durch die in Galfe et al. (1990) beschriebenen Voruntersuchungen ausgeschlossen. Medikamenten- und Genußmitteleinnahme am Tag der Untersuchung und mindestens einen Tag davor waren untersagt.

### 2.2 Apparatur und Meßverfahren

Temperatur- und Schmerzschwellen wurden mit dem „MPI PATH-Tester 100“ bestimmt. Die Untersuchungseinheit arbeitet mit einer Marstock-Thermode (Auflagedruck ca.  $0.4 \text{ N/cm}^2$ ; Stimulationsfläche =  $1.6 \times 3.6 \text{ cm}^2$ ). Sie ist in der genannten Arbeit von Galfe et al. (1990) genauer beschrieben, ebenso die verwendeten psychophysikalischen Verfahren (s.u., sowie das Gerätemanual des Herstellers, 1988<sup>2</sup>; zur Verwendung von auf dem Peltierprinzip beruhenden Kontaktthermoden zur Bestimmung von Temperaturschwellen vgl. auch Fruhstorfer 1976).

Warm- und Kaltschwellen (WS, KS) wurden nacheinander mit 7 Warm- und 7 Kaltreizen als mittlere Änderungsschwellen bei einer Ausgangstemperatur von  $32^\circ\text{C}$  bestimmt (PATH-Tester-Routine „Thermästhesie 1“). Dabei werden die rampenförmigen Temperaturänderungen ( $0.7^\circ\text{C/sec}$ ) durch ein akustisches Vorsignal mit variabler Wartedauer

---

<sup>2</sup>PATH Tester MPI 100, Pain and thermal sensitivity tester, Handbuch, PHYWE SYSTEM GmbH (1980).

angekündigt.

Die Hitzeschmerz-Schwelle wurde für phasische und tonische Reize gemessen (PATH-Tester Routinen „Schmerz 1“ bzw. „Schmerz 2“). Die Schmerzschwelle für phasische Reize (PS) wird dabei durch 8 rampenförmige Reize ( $0.7^{\circ}\text{C}/\text{sec}$ ) bei einer Ausgangstemperatur von  $40^{\circ}\text{C}$  als die gerade schmerzhafte Reiztemperatur bestimmt („Absolutschwelle“).

Die Schmerzschwellen für tonische Hitzereizung (TS) werden bei Routine 2 als Mittelwerte aus 6 Durchgängen mit mindestens 40 sec Reizdauer ermittelt. Dabei regelt der Proband zweimal „gerade schmerzhafte“ Temperaturen (anfangs und nach 35 sec konstanter Reizung) selbst ein. Das Mittel aus beiden Temperaturen wird als Meßwert des jeweiligen Durchgangs übernommen.

### 2.3 Untersuchungsablauf

Die beschriebenen Kennwerte WS, KS, PS, TS wurden abwechselnd an der rechten Hand (Thenar, Dermatom: C6) und am rechten Fuß (lateraler Fußrücken, Dermatom: S1) gemessen. Zur Bestimmung der Stabilität wurden KS, WS und PS nach einer Stunde ein zweites Mal durchgeführt („short-time reliability“ nach Galfe et al. 1990). TS wurde nur einmal pro Meßort erhoben (Tabelle 1).

Tabelle 1: Zeitlicher Ablauf der Untersuchung (WS: Warm-, KS: Kaltschwelle; PS und TS: Hitzeschmerzschwelle für phasische bzw. tonische Stimulation).

	Messung 1						Messung 2							
Hand:	WS	KS	PS				WS	KS	PS	TS				
Fuß:				WS	KS	PS					WS	KS	PS	TS

### 2.4 Statistische Analysen

#### 2.4.1 Korrigierte Normbereiche

Die nach Alter, Geschlecht und Körpermaßen korrigierten Normbereiche der thermozeptiven Kennwerte wurden – der Argumentation in 1.2 entsprechend – in folgenden Schritten gewonnen:

1) Schrittweise multiple Regressionsanalyse des Zusammenhangs der Schwellenmaße mit den unabhängigen Variablen Alter, Geschlecht, Körpergewicht und Körpergröße: Man beginnt mit derjenigen unabhängigen Variablen, welche die meiste Varianz der abhängigen Variablen erklärt. Im nächsten Schritt wird die unabhängige Variable in die Regressionsgleichung aufgenommen, die die meiste Varianz der Residuen erklärt usf., bis die verbliebenen Variablen keine signifikanten Varianzanteile mehr erklären ( $p \geq 0.05$ ).

2) Verteilungsprüfung der Residuen: Die Residuen nach Abschluß der schrittweisen Regression wurden auf Symmetrie (abgeschwächte Voraussetzung der Normalverteilung) geprüft. Bei signifikanter Schiefe wird die gesamte Regressionsanalyse nach logarithmischer Transformation der Schwellen wiederholt. Dies war im vorliegenden Datensatz bei den Warm- und Kaltschwellen an den Händen sowie den Kaltschwellen an den Füßen der Fall (s. Abschnitt 3.1).

3) Die Residuen der schrittweisen Regression wurden mit dem Varianzhomogenitätstest nach Hartley (s. z.B. Sachs 1974) auf die Unabhängigkeit ihrer Streuungen von den signifikanten Korrekturvariablen geprüft. Dies ist eine wichtige Voraussetzung der korrekten Bestimmung von Norm- und Toleranzbereichen. Dazu wurde die Ausgangsstichprobe durch Mediandichotomisierung der Korrekturvariablen in  $2^k$  gleich große Untergruppen geteilt ( $k$  = Anzahl der signifikanten Korrekturvariablen) und die Streuung innerhalb dieser Gruppen auf Gleichheit geprüft.

4) Analog zum Vorgehen von Miller (1966) wurden innere und äußere Toleranzgrenzen (Konfidenzbänder der Normbereichsgrenzen) entsprechend den im Anhang A1 aufgeführten Formeln mit einer Vertrauenswahrscheinlichkeit von 90% bestimmt (vgl. Abschnitt 1.2).

5) Zur weiteren Überprüfung der Annahme, daß wenigstens in dem für die Normgrenzen relevanten Bereich die unterstellten statistischen Annahmen (Normalverteilung der Residuen) hinlänglich gut erfüllt sind, wurden die oberen 95%-, 90%- und 80%-Normbereiche aus dem Regressionsmodell bestimmt und die theoretisch zu erwartende Anzahl der Werte inner- und außerhalb der Normbereiche mit den entsprechenden empirischen Häufigkeiten verglichen (zur Notwendigkeit einer solchen Überprüfung vgl. z.B. Ackermann 1983)<sup>3</sup>.

#### 2.4.2 Reliabilität und Stabilität

Die „interne Reliabilität“ der Messung wurde durch Cronbachs  $\alpha$  aus den Einzelmeßwerten innerhalb eines Durchgangs ( $N_1 = 7$  für WS und KS,  $N_2 = 8$  für PS,  $N_3 = 6$  für TS) bestimmt. Der Koeffizient wird mit der aus der Theorie diagnostischer Tests bekannten Spearman-Brown-Formel für einen aus mehreren Prüfitems (Fragen, Einzelmessungen etc.) zusammengesetzten Test bestimmt und gibt eine untere Grenze für die Reliabilität in dem in 1.3.1 verwendeten Sinn an, also der Korrelation zwischen empirischen Meßwerten zu verschiedenen Zeitpunkten bei gleichbleibenden „wahren“ Werten (vgl. z.B. Lord & Novick 1968).

Für die in einem Abstand von etwa 1 Stunde wiederholten Messungen von WS, KS und PS wurden die Stabilitäten anhand der sog. „Attenuationsformel“ abgeschätzt (Lord & Novick 1968).

---

<sup>3</sup>Ein nicht-parametrisches Verfahren zur Bestimmung von Normbereichen in Abhängigkeit von Drittvariablen wäre theoretisch befriedigender, steht unseres Wissens aber noch nicht zur Verfügung. Desweiteren wurde aus praktischen Gründen auf die Konstruktion multivariater Normbereiche verzichtet. Für eine inhaltlich orientierte Diskussion der gesamten Problematik wird auf Ackermann (1985) verwiesen.

### 2.4.3 Reproduzierbarkeit

Zur Untersuchung der Ausgangswertabhängigkeit der Abweichungen  $\delta_i = |x_i - y_i|$  zweier aufeinanderfolgender Messungen (nach Abschnitt 1.3.2) wurde eine lineare Regression der  $\delta_i$  auf ihre Ausgangswerte  $x_i$  durchgeführt. Im Fall einer signifikanten Steigung, durch die ein proportionaler Fehleranteil nachgewiesen ist, wurde der Meßfehler nach Gleichung (1) in den Relativanteil  $b$  und (falls vorhanden) den Absolutanteil  $a$  aufgeteilt.

Neben diesen globalen (durchschnittlichen) Maßen der Reproduzierbarkeit wurden außerdem in Anlehnung an das Vorgehen in Claus et al. (1990) ausgewählte Perzentilpunkte der  $\delta_i$ - bzw.  $b_i$ -Verteilungen (Formeln 4 und 2) bestimmt. Wie erwähnt geben diese Perzentile Anhaltspunkte für die Größenordnung der zu erwartenden Fehlerbreite.

## 3 Ergebnisse

### 3.1 Normbereiche der thermozeptiven Kennwerte

Die multiple Regressionsanalyse der untransformierten Warm- und Kaltschwellen an der Hand sowie der Kaltschwellen am Fuß ergab Residuen mit deutlich schiefen Verteilungen ( $p < 0.05$ ; Prüfung unter der Annahme der Normalverteilung nach Abschnitt 2.4.1), die durch eine logarithmische Transformation hinreichend symmetrisiert wurden. Die im folgenden berichteten Normbereiche dieser thermozeptiven Kennwerte wurden daher aus der multiplen Regression der logarithmierten Schwellen auf Alter und Körpermaße gewonnen. Die entsprechenden Residuen der Warmschwellen am Fuß sowie der beiden Hitzeschmerzschwellen waren annähernd symmetrisch verteilt und blieben untransformiert.

Die zweite Voraussetzung der Normbereichsbildung durch multiple Regression, die Homoskedastizität (gleiche Streuungen der Meßwerte bei verschiedenen Prädiktorwerten), wurde nur für die Prädiktoren geprüft, die sich in der Analyse als relevante Korrekturvariablen herausgestellt hatten (s.u.: Alter und Körpergewicht). Hierzu wurde die Stichprobe durch Medianhalbierung bezüglich der beiden Korrekturvariablen in 4 gleich große Untergruppen aufgeteilt. Der Varianzhomogenitätstest nach Hartley (vgl. 2.4.1) lieferte für keines der erhobenen thermozeptiven Maße eine signifikante Abweichung von der Varianzhomogenität der Residuen.

Die Ergebnisse sind in den Tabellen 2 (Regressionsanalyse) und 3 (Normtabellen) sowie den Abbildungen 3–5 (Nomogramme zur kombinierten Alters- und Gewichtskorrektur) zusammengefaßt.

Für die Warm- und Kaltschwellen an Hand und Fuß ließ sich danach ein signifikanter Einfluß des Körpergewichts nachweisen (Schwellenänderungen:  $0.9$  und  $0.5^\circ\text{C}$  bzw.  $4.0$  und  $1.8^\circ\text{C}$  im Gewichtsbereich von  $55$ – $95$  kg). Am Fuß waren alle thermozeptiven Schwellenmaße, also sowohl Warm-/Kalt- als auch phasische und tonische Schmerzschwellen, signifikant vom Alter abhängig ( $2$  und  $1.2^\circ\text{C}$  bzw.  $1.6$  und  $1.3^\circ\text{C}$  im Altersbereich von  $20$ – $60$  Jahren). Dagegen hatte das Alter auf die thermozeptiven Schwellenwerte an der Hand keinen Einfluß. Umgekehrt hatte das Körpergewicht keinen Einfluß auf die Schmerzpara-

Tabelle 2: Regressionskoeffizienten, multiple Korrelationskoeffizienten  $R$  und Standardabweichungen der Residuen  $s$  für die Schwellen in Abhängigkeit von Alter und Gewicht ( $n = 64$ ).

Kennwert		Regressionsgewichte			$R$	$s$
		Gewicht $b_G$	Alter $b_A$	const		
log WS	Hand	0.015		-0.624	0.333	0.477
log KS	Hand	0.011		-0.941	0.357	0.348
PS	Hand			45.495		2.254
TS	Hand			45.270		1.528
WS	Fuß	0.098	0.048	-2.630	0.485	2.472
log KS	Fuß	0.019	0.014	-1.401	0.467	0.568
PS	Fuß		0.041	43.601	0.329	1.438
TS	Fuß		0.032	43.567		1.000

meter. Die Schmerzmaße an der Hand waren weder vom Körpergewicht noch vom Alter abhängig.

Die Körpergröße hatte interessanterweise keinen zusätzlichen Einfluß, der über den Effekt des mit ihr hochkorrelierten Körpergewichts hinausging. Da dieses in allen Fällen, in denen die Körpermaße überhaupt einen Einfluß hatten, einen höheren Varianzanteil der thermozeptiven Schwellen erklärte, wurde es als erster Prädiktor in die multiple Regressionsgleichung aufgenommen. Wegen seiner Korrelation mit der Körpergröße bringt deren Berücksichtigung in der Regressionsgleichung keine weitere Verbesserung der Korrektur.

Bemerkenswert ist auch, daß das Geschlecht nach Berücksichtigung des Körpergewichts keinen nachweisbaren Effekt mehr auf die Höhe der Kennmaße besitzt (vgl. Lautenbacher & Strian 1991). Dieses zunächst überraschende Ergebnis wird auch dadurch erhärtet, daß die nach Geschlechtern getrennten Regressionen der thermozeptiven Kennwerte auf Alter und Körpermaße praktisch identische Regressionsgewichte ergaben.

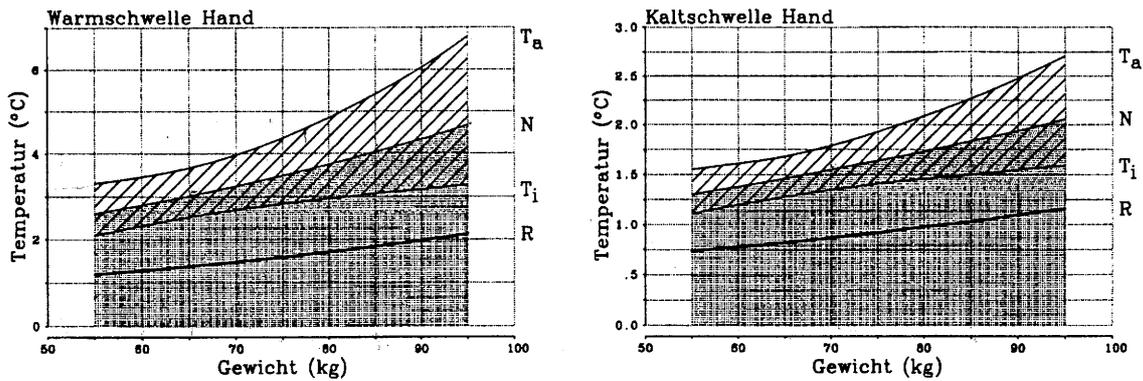
Aus diesen Gründen wurden Körpergröße und Geschlecht in den Nomogrammen der Abbildungen 3–5 nicht mehr dargestellt. Zur Korrektur für Alters-, Geschlechts- und Körpermaß-Unterschiede reichen also die Variablen *Alter* und *Körpergewicht* völlig aus.

Tabelle 3: Einseitige 95%-Norm- und Toleranzbereiche der Schmerzschwellen für phasische und tonische Reize an der Hand:  $N$ : Normbereichsgrenze,  $T_i$ ,  $T_a$ : innere bzw. äußere Toleranzgrenzen zur Konfidenz 90%.

	Mittelwert	Normbereich		
		$T_i$	$N$	$T_a$
PS	45.5	48.5	49.2	50.0
TS	45.3	47.1	47.8	48.6

Abbildung 3 stellt die einseitigen 95%-Normbereiche der Warm-/Kaltschwellen an der Hand mit ihren Toleranzgrenzen bei 90% Konfidenz in Abhängigkeit vom Gewicht dar (zum Begriff der „einseitigen“ Normbereiche vgl. Abschnitt 1.2). Die Abbildung 4 enthält

Abbildung 3: Geschätzte einseitige Normbereiche ( $N$ ) sowie innere und äußere Toleranzgrenzen zur Konfidenz 90% für den 95%-Normbereich der Warm- und Kaltschwellen an der Hand in Abhängigkeit vom Körpergewicht.  $R$  gibt die Regression der Schwellen auf das Gewicht an (vgl. Tab. 2).



die entsprechenden Normbereiche und Toleranzgrenzen der Warm- bzw. Kaltschwellen am Fuß, wegen der gleichzeitigen Altersabhängigkeit aufgliedert nach drei Altersgruppen. Abbildung 5 stellt die Altersabhängigkeit der phasischen und tonischen Schmerzschwellen am Fuß dar. Bei den Schmerzschwellen an der Hand bestand weder Alters- noch Körpermaßabhängigkeit; für sie genügt daher die Angabe der univariaten Normbereiche in Tabelle 3.

Abschließend wurden zur Post-hoc-Überprüfung des multiplen Regressionsmodells die theoretisch zu erwartenden Häufigkeiten innerhalb der 80%-, 90%- und 95%-Normbereiche mit den empirisch gefundenen Häufigkeiten verglichen. Die Übereinstimmungen waren zufriedenstellend.

Die Korrektur der Warmschwelen am Fuß nach Alter und Körpergewicht führt man wie folgt durch (die übrigen Maße werden analog korrigiert):

(1) Man wählt das Nomogramm für die Altersgruppe, in die der Patient fällt (z.B. 15–29, Abbildung 4a).

(2) Dann bestimmt man aus der Abbildung die einseitige 95%-Normbereichsgrenze  $N$  und deren innere ( $T_i$ ) und äußere Toleranzgrenzen ( $T_a$ ) für das Körpergewicht des Patienten (z.B. 75 kg). Daraus ergibt sich:  $N = 9.8$ ;  $T_i = 8.2$ ;  $T_a = 11.6^\circ\text{C}$ .

(3) Anschließend wird der gemessene Wert mit diesen Grenzen verglichen: (a) Ist die WS am Fuß größer als  $11.6^\circ\text{C}$ , so liegt der Wert des Patienten mit mindestens 90%-iger Sicherheit (d.i. die Konfidenz der Toleranzgrenze) oberhalb des einseitigen 95%-Normbereichs, d.i. die Grenze, unterhalb der 95% aller Normwerte bei dem zugehörigen Körpergewicht liegen. Der Wert wäre also mit hoher Sicherheit auffällig erhöht. (b) Ist die WS kleiner als  $8.2^\circ\text{C}$ , ist der Wert mit 90%-iger Sicherheit normal. (c) Werte zwischen  $8.2$  und  $11.6^\circ\text{C}$

Abbildung 4: Geschätzte einseitige Normbereiche ( $N$ ) sowie innere und äußere Toleranzgrenzen zur Konfidenz 90% für den 95%-Normbereich der Warm- und Kaltschwellen am Fuß in Abhängigkeit vom Körpergewicht für die Altersklassen 15–29, 30–44 und 45–65 Jahre.  $R$  gibt die Regression der Schwellen auf das Gewicht an (vgl. Tab. 2).

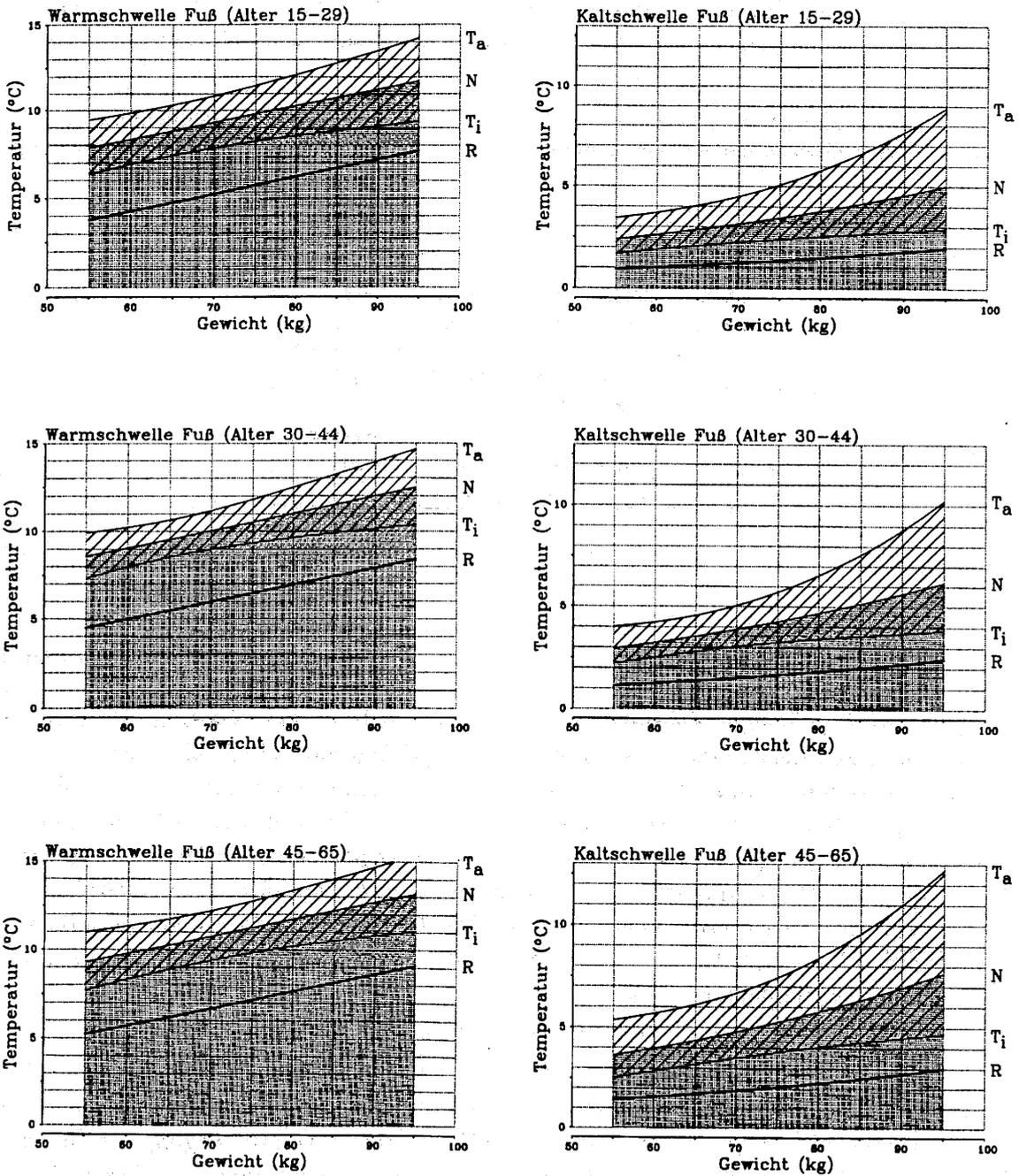
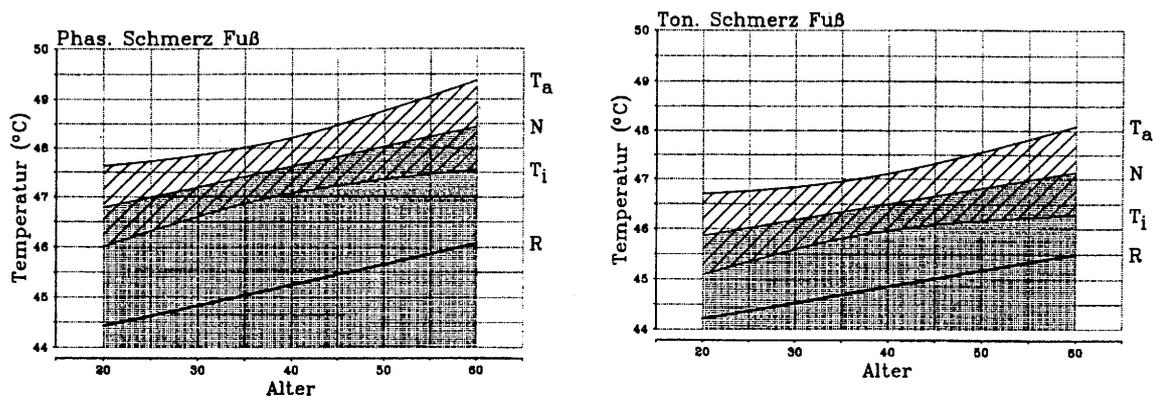


Abbildung 5: Geschätzte einseitige Normbereiche ( $N$ ) sowie innere und äußere Toleranzgrenzen zur Konfidenz 90% für den 95%-Normbereich der Schmerzwellen für phasische und tonische Hitzezeize am Fuß in Abhängigkeit vom Alter.  $R$  gibt die Regression der Schwellen auf das Alter an (vgl. Tab. 2).



sind „fraglich“ (in Bezug auf die verfügbaren Normdaten), und zwar oberhalb von  $N = 9.8^\circ\text{C}$  „fraglich erhöht“, unterhalb „fraglich normal“.

Die Tabelle 2 erlaubt schließlich auch eine (aufgrund der statistischen Unsicherheit etwas vorsichtig zu interpretierende) Bestimmung eines Standardscores. Für eine z.B. 22 Jahre alte Person (Alter= $A$ ) mit 75 kg Gewicht (=  $G$ ) und einer Warmschwelle  $WS$  am Fuß von  $9^\circ\text{C}$  ergibt sich die Korrekturformel

$$\begin{aligned}
 S &= \frac{WS - (\text{const} + b_A \cdot A + b_G \cdot G)}{s} \\
 &= \frac{9.0 - (-2.63 + 0.048 \cdot 25 + 0.098 \cdot 75)}{2.472} \\
 &= \frac{9.0 - 5.78}{2.472} = 1.30
 \end{aligned}$$

Das zugehörige einseitige Perzentil der Normalverteilung ist 0.097, d.h. daß etwa 10% aller gesunden Personen dieser Alters- und Gewichtsgruppe eine höhere Schwelle aufweisen.

### 3.2 Reliabilität und Stabilität

Die auf der Basis der Einzelreizungen bestimmten Zuverlässigkeiten innerhalb einer Meßserie („interne Reliabilität“ nach Abschnitt 1.3.1) sämtlicher thermozeptiven Maße waren mit Reliabilitätskoeffizienten von über 0.95 (Cronbachs  $\alpha$ ) sehr hoch (s. Tabelle 4). Die Zuverlässigkeit der Messung selbst ist also sehr hoch und erfüllt damit die Anforderungen eines diagnostischen Instruments.

Mithilfe dieser Schätzungen der Zuverlässigkeit der Messung selbst läßt sich aus den veröffentlichten Retestreliabilitäten thermozeptiver Kennwerte, also den Korrelationen zwischen wiederholten Messungen, die Stabilität des biologischen Merkmals („Merkmalsreliabilität“ nach Lienert 1969) bestimmen (s. Abschnitt 2.4.2). Diese sog. „disattenuierten“ oder „minderungskorrigierten“ Werte für den Datensatz von Galfe et al. (1990) nebst den zugehörigen Retestkorrelationen bei Messungen im Abstand von einer Stunde sind ebenfalls in Tabelle 4 dargestellt. Die beiden Koeffizienten unterscheiden sich wegen der hohen internen Zuverlässigkeit der Messungen nur wenig voneinander.

Tabelle 4: Reliabilitäten (Cronbachs  $\alpha$ ) der Schwellen WS, KS, PS und TS, Korrelationen)<sup>a</sup> und Stabilitäten („minderungskorrigierte“ Korrelationen) der wiederholt bestimmten Schwellen WS, KS und PS ( $n = 64$ ).

	WS	KS	PS	TS
Reliabilität				
Hand	0.968	0.950	0.989	0.984
Fuß	0.973	0.970	0.984	0.980
Korrelation				
Hand	0.658	0.810	0.847	
Fuß	0.790	0.722	0.887	
Stabilität				
Hand	0.680	0.853	0.856	
Fuß	0.812	0.744	0.901	

<sup>a</sup> aus Galfe et al. 1990

### 3.3 Reproduzierbarkeit

Die Überprüfung des Zusammenhangs zwischen den Abweichungen  $\delta_i$  und den Ausgangswerten  $x_i$  ergab, daß außer bei den Kaltschwellen am Fuß kein Zusammenhang zwischen Meßfehler und Ausgangswert nachweisbar war. Für diese Parameter liefern also die üblichen relativen Differenzen und ihre Verteilungen *keine* geeignete Abschätzung der Reproduzierbarkeit. Stattdessen ist die Verteilung der absoluten Abweichungen  $\delta_i$  zwischen den beiden Messungen zu wählen. Aus deren Kennwerten (Mittelwert, Perzentile, Streubreich) können der durchschnittliche Meßfehler und seine Streubreite abgelesen werden. Sie sind in Tabelle 5 dargestellt. Aus ihr ergibt sich, daß z.B. bei der Bestimmung der Schmerzschwelle für phasische Hitzereize am Fuß 2/3 aller Meßfehler zwischen 0.10 und 1.09°C liegen, also zwischen 16%- und 84%-Perzentil (entsprechend  $\bar{\delta}_i \pm 1s$  bei normal verteilten Daten). Im Mittel ist die Messung auf etwa 0.54°C genau reproduzierbar ( $\bar{\delta}_i$ ).

Nur bei den Kaltschwellen am Fuß ergab sich eine beträchtliche Korrelation zwischen Meßfehler und Ausgangswerten ( $r = +0.79$ ). Der Achsenabschnitt der zugehörigen Regression war nicht signifikant von 0 verschieden. Daraus folgt die proportionale Abhängigkeit zwischen Abweichungen und Ausgangswerten. Nur hier sind also Reproduzierbarkeitsmaße auf der Basis von relativen Differenzen angebracht, wie sie Claus et al. (1990) verwendete

Tabelle 5: Mittelwerte, 10%, 16%)<sup>a</sup>, 50%, 84%)<sup>a</sup> und 90%-Perzentile der Verteilung der absoluten Abweichungen  $\delta_i$  bzw. der relativen Abweichungen  $b_i^*$  der Schwellen ( $n = 64$ ).

	$\bar{x}$	$P_{10\%}$	$P_{16\%}$	$P_{50\%}$	$P_{84\%}$	$P_{90\%}$
$\delta$ in °C						
WS Hand	0.54	0.03	0.09	0.33	0.86	1.63
WS Fuß	1.46	0.24	0.43	1.09	2.71	3.38
KS Hand	0.16	0.00	0.01	0.10	0.30	0.39
PS Hand	0.95	0.14	0.25	0.70	1.49	1.84
PS Fuß	0.54	0.10	0.10	0.38	1.09	1.18
$b_i^*$ in %						
KS Fuß	32.6	1.7	5.5	24.7	59.2	71.5

)<sup>a</sup>: Die 16%- und 84%-Perzentile entsprechen bei normalverteilten Daten  $\bar{x} - s$  bzw.  $\bar{x} + s$ .

und wie sie auch den „Variabilitätsindizes“ von Galfe et al. (1990) zugrundeliegen. Für die Kaltschwellen am Fuß sind daher in Tabelle 5 die Verteilungskennwerte der Reproduzierbarkeitsmaße  $b_i^*$  nach Formel 2 angegeben, die jenen Koeffizienten äquivalent sind (vgl. Abschnitt 1.3.2). Aus der Tabelle (letzte Zeile) entnimmt man, daß der Median (=  $P_{50\%}$ ) der Abweichungen für dieses spezielle Maß etwa 25% bzw. ihr arithmetisches Mittel 33% ( $\bar{b}_i^*$ ) beträgt, also relativ schlecht ist. Bei einer mittleren Kaltschwelle von 2°C sind dies allerdings nur ca. 0.5 bzw. 0.7°C. Dies ist bei der Beurteilung des diagnostischen Werts des Verfahrens zu beachten, da hierbei die absolute Größe der pathologischen Schwellenerhöhung mit eingeht.

## 4 Diskussion

### 4.1 Normbereiche

Die ermittelten Konfidenzintervalle oder „Toleranzbereiche“ für die Normbereichsgrenzen der Thermästhesie-Messungen mit dem PATH-Tester sind nach diesen Ergebnissen noch sehr weit. Dies liegt daran, daß Varianzschätzungen dieser Art bei kleinen Stichproben instabil sind. Eine Verbesserung ist nur durch eine erhebliche Vergrößerung der Eichstichprobe zu erreichen und für die weitere diagnostische Nutzung unumgänglich.

Aus unseren Ergebnissen wird auch deutlich, daß die in der bisherigen Literatur übliche Praxis, nur  $(\bar{x} \pm 2s)$ -Bereiche anzugeben und auf die Berechnung der zugehörigen Konfidenzintervalle zu verzichten, die Beurteilung der diagnostischen Verfahren irreführen. Aus solchen einfachen Angaben über Streubereiche (ob alterskorrigiert oder nicht) läßt sich nämlich nicht abschätzen, wie verläßlich diese Normbereiche überhaupt sind. Damit wird aber auch die Klassifikation eines gegebenen Werts als „pathologisch“ oder „nicht pathologisch“ in der Nähe der Normbereichsgrenzen fraglich. So gesehen, sind die hier berichteten Bereiche mit ihren Konfidenzintervallen noch die beste verfügbare Lösung.

Die Normwertbereiche wurden in Abhängigkeit von mehreren biologischen Merkmalen und dem Alter bestimmt, da sich bei der Mehrzahl der thermozeptiven Parameter Abhängigkeiten von Alter und Körpermaßen nachweisen ließen. Außer Alterungsprozessen scheinen also auch andere Variablen eine Rolle zu spielen. Dies könnten Faserlänge und/oder Rezeptordichte sein, die beide mit Körpermaßen korrelieren.

Das Ergebnis, daß es anscheinend genügt, nur für das Körpergewicht zu korrigieren, und die Körpergröße keine zusätzliche Varianz mehr erklärt, ist allerdings mit Vorsicht zu interpretieren. Nach der Logik der schrittweisen multiplen Regression, mit der diese Korrekturen vorgenommen wurden, bestimmt nämlich die Reihenfolge, in der die Varianzanteile der einzelnen Prädiktorvariablen extrahiert werden, wann abgebrochen wird. In unserem Fall wurde die Reihenfolge wie üblich durch die Größe der erklärten Varianz bestimmt. Dabei ergab sich, daß das Körpergewicht für sich genommen einen etwas höheren Varianzanteil erklärte als die Körpergröße. Daher wurde erstere zuerst extrahiert. Der Unterschied ist jedoch nicht so groß, als daß nicht die Vertauschung der Reihenfolge zu rechtfertigen wäre, falls dies aus inhaltlichen Gründen plausibler zu machen ist, z.B. weil man die Faserlänge als Ursache der Körpermaßeffekte vermutet.

## 4.2 Zuverlässigkeit

Die interne Zuverlässigkeit der thermozeptiven Kennwerte war mit Reliabilitäten von über 0.95 sehr hoch. Die demgegenüber relativ niedrige Stabilität und Reproduzierbarkeit über verschiedene Zeitpunkte hinweg gehen also vor allem auf den Einfluß zeitlich und situativ veränderlicher Faktoren zurück. So wurden in der vorliegenden Untersuchung weder Tagesschwankungen (Strian et al. 1989) noch Änderungen der Hauttemperatur (Kenshalo 1976) berücksichtigt. Eichstudien, die diese Variablen in die Korrektur der Normwerte miteinbeziehen, dürften die Stabilitäten erheblich verbessern und sollten durchgeführt werden.

Für die klinische Anwendung zur Neuropathie-Diagnose bei Diabetikern ist außerdem zu bedenken, daß temporäre Schwellenerhöhungen auch von akuten Blutzuckerschwankungen bedingt sein können und bei Wiederherstellung der Normoglykämie teilweise wieder verschwinden (Haslbeck et al. 1993, in prep.). Sie dürfen daher nicht mit echten neuropathischen Veränderungen verwechselt werden.

## ANHANG

### A1: Bestimmung von Norm- und Toleranzgrenzen

Unter den üblichen Annahmen des Standardmodells der linearen Regression (Normalverteilung der Residuen mit gleicher Varianz) gewinnt man in Verallgemeinerung des Vorgehens von Miller (1966) äußere und innere Toleranzintervalle  $T_a$  bzw.  $T_i$ . Dabei kombiniert man das Konfidenzband der linearen Regression mit dem Konfidenzintervall für die Residualvarianz mittels der Bonferroni-Ungleichung. Die äußeren bzw. inneren Toleranzintervalle mit Überdeckung  $P$  (üblicherweise 95%) zur Konfidenz  $1 - 2\alpha$  lauten:

$$T_a(x_2, \dots, x_p) = b_1 + \sum_{i=2}^p b_i x_i \pm s \cdot \left\{ \sqrt{p F_{p, n-p}^\alpha \cdot \mathbf{x}^T \mathbf{V} \mathbf{x}} + N(P) \sqrt{\frac{n-p}{\chi_{n-p, \alpha}^2}} \right\}$$
$$T_i(x_2, \dots, x_p) = b_1 + \sum_{i=2}^p b_i x_i \pm s \cdot \left\{ -\sqrt{p F_{p, n-p}^\alpha \cdot \mathbf{x}^T \mathbf{V} \mathbf{x}} + N(P) \sqrt{\frac{n-p}{\chi_{n-p, 1-\alpha}^2}} \right\}$$

In den beiden Formeln bezeichnen die  $b_i$  die Schätzungen der Regressionsgewichte ( $b_1$  ist darin die Konstante),  $s^2$  den üblichen erwartungstreuen Schätzer für die Residualvarianz,  $n$  die Zahl der Daten,  $p$  die Zahl der unabhängigen Variablen *einschließlich* der Konstanten,  $N(P)$  den  $P$ -Perzentil-Punkt der Normalverteilung und  $F_{p, n-p}^\alpha$  und  $\chi_{n-p, \alpha}^2$  die  $\alpha$ -Quantile der  $F$ - bzw.  $\chi^2$ -Verteilung mit den zugehörigen Freiheitsgraden. Die Matrix  $V$  ist die aus der Designmatrix zu berechnende Varianz-Kovarianzmatrix für die geschätzten Regressionsgewichte  $b_1, \dots, b_p$  und schließlich  $\mathbf{x}^T = (1, x_2, \dots, x_p)$  der Vektor der Prädiktorwerte (einschließlich der Konstanten).

Miller (1966) gibt lediglich Formeln für die äußeren Toleranzbereiche im Falle der einfachen Regression an, die Erweiterung auf mehrere Prädiktoren ist jedoch offensichtlich. Die Ableitung der Formeln für innere Toleranzbereiche, die von Miller nicht behandelt werden, ergibt sich analog zu dieser durch Abschätzung der Residualvarianz nach unten und Verwendung des unteren Konfidenzbandes der Regressionsgerade/-ebene.

### A2: Modell der Fehlerverteilung

Die Verteilungsfunktion der absoluten Abweichungen der Zweitmessung von der Erstmessung bei einem Ausgangswert  $x$  werde mit  $\Delta_x(\delta)$  bezeichnet.

Die beiden in Abschnitt 1.3.2 (b) getroffenen Annahmen lauten:

(1) Für zwei Ausgangswerte  $x$  und  $x'$  lassen sich die zugehörigen Verteilungsfunktionen durch

$$\Delta_{x'}(\delta) = \Delta_x(c\delta)$$

mit einer von  $x$  und  $x'$  abhängigen Konstanten  $c$  ineinander überführen.

(2) Der Erwartungswert der Abweichungen  $\Delta_x$  mit Ausgangswert  $x$  ist eine lineare Funktion von  $x$ :

$$\mathcal{E}\Delta_x = bx + a$$

Aus (1) und (2) folgt für die von  $x$  und  $x'$  abhängige Konstante  $c$ :

$$c = \frac{bx + a}{bx' + a}$$

Wählt man  $x' = x^* := \frac{b-a}{b}$ , so erhält man für gegebenes  $x$  und  $b \neq 0$  das zugehörige  $c = x + a/b$ . Dann besitzt die durch die Transformation

$$K = \frac{\Delta_x}{x + a/b}$$

definierte Zufallsvariable die vom Ausgangswert  $x$  *unabhängige* Verteilungsfunktion  $\Delta_{x^*}$ ,  $x^* = \frac{b-a}{b}$ . Die transformierten Rohwerte können somit zur Schätzung dieser Verteilung oder spezieller Kennmaße (Mittelwert, Streuung o.ä.) verwendet und durch Rücktransformation für beliebige Ausgangswerte  $x$  die entsprechenden Kenngrößen von  $\Delta_x$  bestimmt werden.

## Literatur

- Ackermann H. (1983). Sind „ $\bar{x} \pm 2s$ “-Bereiche nützliche diagnostische Hilfsmittel? *Med. Welt*, **34**, 212–215.
- Ackermann H. (1985). *Mehrdimensionale nicht-parametrische Normbereiche*. Berlin: Springer.
- Claus D., Hilz M.J. & Neundörfer B. (1990). Thermal discrimination thresholds: a comparison of different methods. *Acta Neurol Scand*, **81**, 533–540.
- Fruhstorfer H., Lindblom U. & Schmidt W.G. (1976). Method for quantitative estimation of thermal thresholds in patients. *J Neurol Neurosurg Psychiatry*, **39**, 1071–1075.
- Galfe G., Lautenbacher S., Hölzl R. & Strian F. (1990). Diagnosis of small-fibre neuropathy: Computer assisted methods of combined pain and thermal sensitivity determination. *Hospimedica*, **12**, 38–48.
- Guy R.Y., Clark C.A., Malcolm P.N. & Watkins P.J. (1985). Evaluation of thermal and vibration sensation in diabetic neuropathy. *Diabetologia*, **28**, 879–913.
- Haslbeck M., Möltner, A., Hölzl, R., Strian, F. (1993). Die Bewertung der diabetischen Neuropathie kleiner Nervenfasern durch Thermosensitivitätsmessung (in prep.).
- Kenshalo, D.R. (1976). Correlations of temperature sensitivity in man and monkey, a first approximation. In: Y. Zotterman (Ed.) *Sensory functions of the skin in primates*. Oxford: Pergamon Press, 305–330.
- Lautenbacher (1990). *Einflüsse des Alters auf die Schmerzwahrnehmung*. Peter Lang, Frankfurt a. Main.
- Lautenbacher, S. & Strian, F. (1991). Sex differences in pain and thermal sensitivity: The role of bodysize. *Perception & Psychophysics*, **50**, 179–183.
- Lienert G.A. (1969<sup>3</sup>). *Testaufbau und Testanalyse*. Weinheim: Julius Beltz.
- Lord F.M. & Novick M.R. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.
- Miller R.G. (1966). *Simultaneous statistical inference*. New York: McGraw-Hill.
- Sachs L. (1974<sup>4</sup>). *Angewandte Statistik*. Berlin: Springer.
- Strian F., Lautenbacher S., Galfe G. & Hölzl R. (1989). Diurnal variations in pain perception and thermal sensitivity. *Pain*, **131**, 125–131.