

# Auswertung longitudinaler Daten in klinischen Studien mit Hilfe von Summary Statistics

Tim Friede, Hans Scheurlen

Abteilung Medizinische Biometrie, Universität Heidelberg  
Im Neuenheimer Feld 305, D-69120 Heidelberg  
eMail: friede@imbi.uni-heidelberg.de

## Abstract

Bei Fragestellungen aus der Medizin sind häufig zeitliche Verläufe von besonderem Interesse, so z.B. in der Pharmakodynamik, bei der Verlaufskontrolle unter Therapie oder bei Nachsorgeuntersuchungen. Es wird eine Übersicht über die sogenannten *summary statistics* gegeben, die den Verlauf eines Patienten in einer Maßzahl widerspiegeln. Das Zusammenfassen des Verlaufs in eine Maßzahl erleichtert dann die Auswertung in erheblichem Maße, was zur Beliebtheit dieser Statistiken beigetragen hat. Die Anwendung der *summary statistics* wird an Beispielen aus der Medizin illustriert und ihre Einsatzmöglichkeiten diskutiert. Die Umsetzung in SAS wird besprochen und ein benötigtes Macro bereitgestellt.

## 1. Einführung

Bei Fragestellungen aus der Medizin sind häufig zeitliche Verläufe von besonderem Interesse, so z.B. in der Pharmakodynamik, bei der Verlaufskontrolle unter Therapie oder bei Nachsorgeuntersuchungen. Datensätze mit wiederholten Messungen zu verschiedenen Zeitpunkten an derselben Versuchseinheit werden als *longitudinale* oder *serielle* Daten bezeichnet. Da wir uns hier mit klinischen Studien befassen, ist die Versuchseinheit im folgenden immer der Patient. In der Literatur werden im wesentlichen zwei Verlaufsformen unterschieden [1]. Es handelt sich um sogenannte *peaked* und *growth curves*. Bei den *peaked curves* handelt es sich um gipflige Verläufe, während die *growth curves* monotone Verläufe darstellen.

Häufig werden longitudinale Daten klinischer Studien durch Gruppenvergleiche (z.B. von Behandlungsgruppen) zu verschiedenen Zeitpunkten ausgewertet. Dieses Vorgehen birgt einige Probleme in sich. Zum einen sind die Ergebnisse in aller Regel schwer zu interpretieren. Was bedeutet z.B. ein signifikanter Unterschied zu einem Zeitpunkt, wenn zu benachbarten Zeitpunkten kein Unterschied nachzuweisen ist? Zum anderen haben wir hier ein multiples Testproblem, das nicht immer als solches erkannt und behandelt wird. Ein ganz wesentliches Problem ist aber die Tatsache, daß individuelle Verläufe einzelner Patienten nicht berücksichtigt werden, obwohl diese von ganz besonderem klinischen Interesse sind.

Statistische Methoden zur Auswertung longitudinaler Daten unter Berücksichtigung individueller Verläufe waren und sind Gegenstand der methodischen Forschung. Einen Überblick über die Verfahren geben z.B. die Bücher von Diggle et al. und Lindsey [2,3] oder die Arbeiten von Albert, Omar et al. und Dunn und Pickles [4-6]. Brunner und Langer [7] beschreiben nichtparametrische Verfahren.

Im folgenden Kapitel wollen wir einen Ansatz zur Auswertung longitudinaler Daten vorstellen, der individuelle Verläufe von Patienten berücksichtigt und dabei mit einfachen statistischen Mitteln auskommt. Das Vorgehen wird an zwei Beispielen illustriert. Daran schließen sich einige Bemerkungen zur Umsetzung des Ansatzes in der Software SAS von SAS Institute, Cary, NC an. Abschließend wird die hier vorgestellte Methode kurz diskutiert.

## 2. Summary Statistics

### Definitionen und Notationen

Im folgenden wollen wir unter einer *summary statistic* eine Funktion verstehen, die die zu verschiedenen Zeitpunkten wiederholten Beobachtungen an einem Patienten auf eine reelle Zahl abbildet. Formal schreiben wir

$$Y_i = \phi(X_{i1}, \dots, X_{iT}),$$

wobei  $X_{it}$  ( $1 \leq i \leq n$ ,  $1 \leq t \leq T$ ) die Beobachtung an Patient  $i$  zum Zeitpunkt  $t$  und  $\phi$  eine reellwertige Funktion ist. Bei der Definition haben wir vorausgesetzt, daß für jeden Patienten gleichviele Beobachtungen, nämlich genau  $T$ , vorliegen. Dies ist zwar für die Definition nicht zwingend notwendig, gestattet aber später die Annahme gleicher Verteilungen der  $Y_i$  und vereinfacht zudem die Schreibweise.

Zuweilen wird der Begriff *summary statistic* enger gefaßt. So verstehen z.B. Brunner und Langer [7] unter einer *summary statistic* einen sogenannten *summarischen Effekt*. Sie definieren somit  $Y_i = \sum_t \lambda_{it} X_{it}$ .

### Einige Summary Statistics

Wir geben nun eine kurze Aufstellung verschiedener *summary statistics*. Vergleichbare Aufstellungen findet der Leser in [8,9].

- **Zeitpunkte**, für das Eintreten eines Ereignisses wie z.B.
  - Erreichen des Maximums (*peak*)
  - Erreichen eines bestimmten, festen Wertes
  - Rückkehr zu Baseline
  - Gleichgewichtspunkt
- **summarische Effekte**
  - Minimum, Maximum
  - Mittelwert
  - *area under the curve* (AUC)
  - *rate of change*: Steigungsparameter einer linearen Regression (1.Ordnung)
- **Verlaufstypen** (nach klinischen Kriterien)

### Der Gleichgewichtspunkt

Man stelle sich die Meßwerte als Kräfte vor, die auf die Zeitachse am Zeitpunkt ihrer Messung einwirken. Der Gleichgewichtspunkt ist dann der Zeitpunkt, an dem die Zeitachse aufgehängt werden müßte, um sich im Gleichgewicht zu befinden. Wie in Abbildung 1 dargestellt, befindet sich ein zweiarmiger Hebel im Gleichgewicht, wenn das Produkt aus Kraft und Kraftarm gleich dem Produkt aus Last und Lastarm ist. Diese Produkte werden auch als Drehmomente bezeichnet. In Abbildung 1 ist die Last durch  $L$ , der Lastarm durch  $l$ , die der Kraft durch  $K$  und der Kraftarm durch  $k$  gekennzeichnet. Hieraus ergibt sich unmittelbar die Definition des Gleichgewichtspunkts. Der Gleichgewichtspunkt  $\theta_i$  ist bestimmt durch  $\sum_j x_{ij}(t_{ij} - \theta_i) = 0$ . Durch algebraische Umformung ergibt sich

$$\theta_i = \frac{\sum_j x_{ij} t_{ij}}{\sum_l x_{il}}.$$

Diese Darstellung macht deutlich, daß es sich beim Gleichgewichtspunkt um ein gewichtetes Mittel der Meßzeitpunkte handelt. Dabei werden die Gewichte durch die Meßwerte bestimmt.

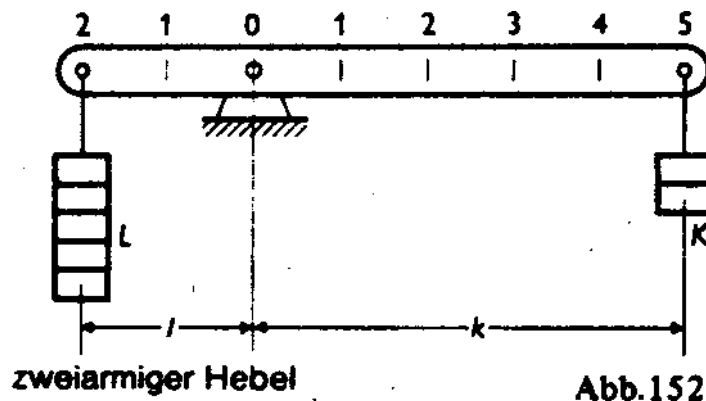


Abbildung 1 Zweiarmiger Hebel im Gleichgewicht (entnommen aus [10])

Wenn die Zeit zwischen der ersten und letzten Beobachtung für verschiedene Patienten variiert, sollte der Gleichgewichtspunkt ins Verhältnis zur Zeit zwischen der ersten und letzten Beobachtung gesetzt werden. Wir bezeichnen das Resultat im folgenden als *relativen* Gleichgewichtspunkt. Weiter unten illustrieren wir den relativen Gleichgewichtspunkt anhand von Tumormarkerverläufen nach kurativer Primärbehandlung.

### Auswertungen mit Summary Statistics

Es liegen nun unabhängige Beobachtungen vor, die nun wie *gewohnt* ausgewertet werden können. Z.B. durch

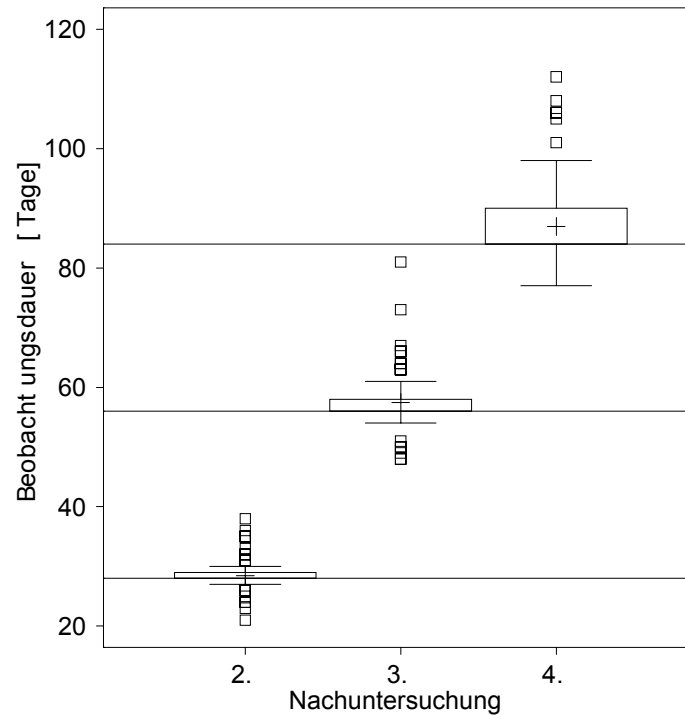
- **graphische Darstellungen** (Histogramme, empirische Verteilungsfunktionen, Streudiagramme etc.)
- **deskriptive Statistiken** (und zugehörige Konfidenzintervalle)
- **Hypothesentests**; z.B. Wilcoxon-Rangsummen- oder Kruskal-Wallis-Test

Es sei an dieser Stelle darauf hingewiesen, daß ungleiche Anzahlen von Beobachtungen oder *missing values* die Annahme gleicher Verteilungen in Frage stellen können. Diese Annahme wird aber z.B. für den Wilcoxon-Rangsummen-Test benötigt (siehe [7], [11]).

## 3. Beispiele aus der klinischen Forschung

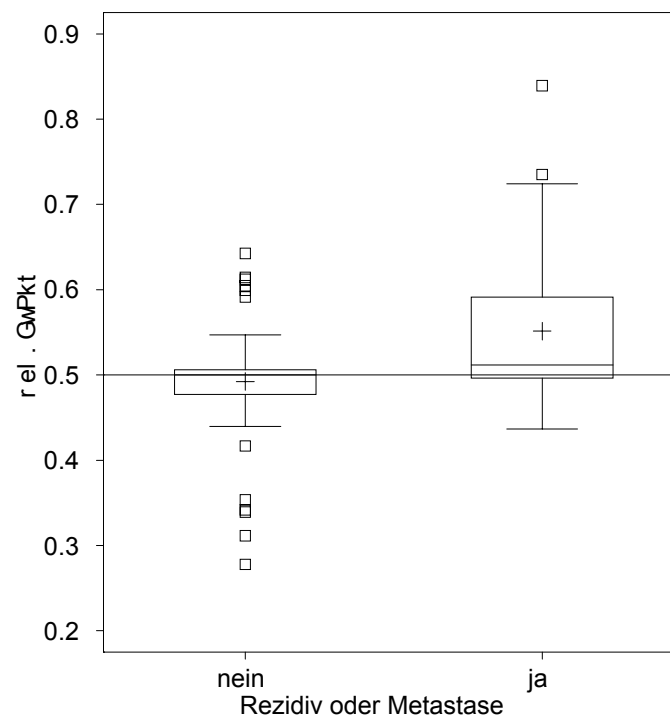
### Tumormarker in der Nachsorge

In einer multizentrischen Studie wurden insgesamt 156 Patienten nach kurativer Therapie eines Ovarial- oder Margenkarzinoms nachbeobachtet [12]. Die Planung der Studie sah vor, zu vier Zeitpunkten, nämlich 28, 56 und 84 Tage nach der ersten Untersuchung, Blutproben abzunehmen und für diese einen Tumormarker (CA72-4) zu bestimmen. Zusätzlich sollte erfaßt werden, ob ein Rezidiv oder Metastasen unter der Beobachtung diagnostiziert werden konnten. In zwei Zentren kam es zu deutlichen Abweichungen vom Nachbeobachtungsschema, so daß sie von der hier durchgeführten Auswertung ausgeschlossen wurden. Bei 110 Patienten lagen vollständige Angaben über vier Tumormarkerbestimmungen sowie über das Auftreten eines Rezidivs oder von Metastasen vor. Diese 110 Patienten wurden zwischen 77 und 112 Tagen nachbeobachtet (siehe Abbildung 2).



**Abbildung 2** Nachbeobachtungsschema. Die Referenzlinien markieren die geplanten Beobachtungen nach 28, 56 und 84 Tagen (n=110 bei jeder Nachuntersuchung).

Obwohl das Nachbeobachtungsschema im wesentlichen eingehalten wurde (siehe Abbildung 2), kam es zu kleineren Abweichungen von den geplanten Beobachtungszeitpunkten und insbesondere der Gesamtbeobachtungsdauer. Deshalb wurde der *relative* Gleichgewichtspunkt als *summary statistic* gewählt. Bei 24 Patienten kam es zu einem Rezidiv oder einer Metastasierung im Verlauf der Beobachtung.



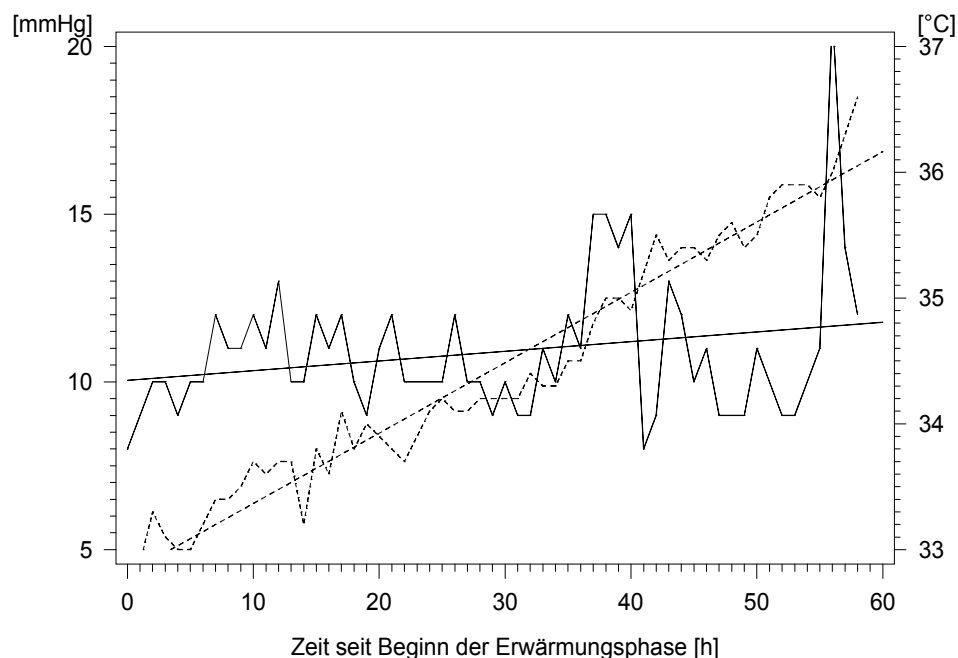
**Abbildung 3** Boxplots der relativen Gleichgewichtspunkte (rel.GwPkt) für Patienten ohne und mit Rezidiv oder Metastasierung (n=86 bzw. n=24)

Abbildung 3 zeigt die relativen Gleichgewichtspunkte für Patienten mit und ohne Rezidiv oder Metastasierung. Aus Abbildung 3 ist zu erkennen, daß die Patienten mit Rezidiv

tendentiell in der zweiten Hälfte der Beobachtungszeit höhere Tumormarkerwerte aufweisen. Führt man einen Wilcoxon-Rangsummen-Test für die relativen Gleichgewichtspunkte durch, so ergibt sich ein zweiseitiger approximativer p-Wert von 0.008. Dies Ergebnis ist umso erstaunlicher, da es sich um einen sehr kurzen Beobachtungszeitraum von etwa zwölf Wochen handelt, den man aus praktischen Erwägungen heraus gewählt hat.

## Hirndruck während der Erwärmung nach Hypothermie

Patienten mit malignem Media-Infarkt wurden mit moderater Hypothermie therapiert. Bei der sogenannten moderaten Hypothermie wird die Körpertemperatur auf 33°C abgesenkt und auf diesem Niveau für etwa zwei bis drei Tage gehalten. In der anschließenden Erwärmungsphase kehrt der Patient zur normalen Körpertemperatur von 36.5°C zurück. Bei der Erwärmung wurden verschiedene Protokolle verwendet. Zehn von insgesamt 22 Patienten wurden passiv wiedererwärmt. Bei den übrigen zwölf Patienten wurde eine kontrollierte Wiedererwärmung durchgeführt, d.h., daß die Erwärmung mit einer bestimmten Geschwindigkeit durchgeführt wurde, was durch Oberflächenkühlung erreicht wurde.



**Abbildung 4 Körpertemperatur (gestrichelte Linie) und ICP (durchgezogene Linie) eines Patienten mit malignem Media-Infarkt im Verlauf der Erwärmungsphase nach moderater Hypothermie**

Während der Erwärmung wurden die Körpertemperatur und der intrazerebrale Druck (ICP) stündlich gemessen. Da beschrieben worden war, daß es während der Erwärmung zu kontinuierlichen Hirndruckanstiegen kommt [13], sollte untersucht werden, ob höhere Erwärmungsgeschwindigkeiten zu schnelleren Hirndruckanstiegen führen. Als Maß für die Geschwindigkeit der Erwärmung und des Druckanstiegs wurden die Kleinste-Quadrate-Schätzer für die Steigungsparameter linearer Regressionen erster Ordnung gewählt. Abbildung 4 zeigt die Körpertemperatur und den ICP für einen Patienten im Verlauf der Erwärmung. Die geschätzten Geraden sind ebenfalls eingezeichnet. In Abbildung 5 sind die Steigungen von ICP und Körpertemperatur für alle 22 Patienten dargestellt. Es ergibt sich ein Spearman-Rangkorrelationskoeffizient von 0.6 ( $p=0.003$ ). Für eine ausführliche Darstellung der Ergebnisse verweisen wir auf [14].

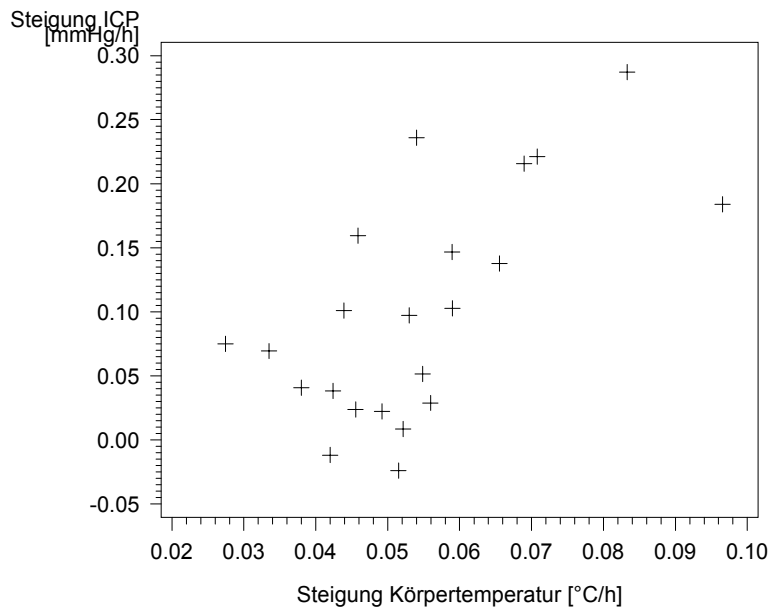


Abbildung 5 Steigungen von ICP und Körpertemperatur für alle 22 Patienten

## 4. Umsetzung in SAS

Viele *summary statistics* können mit entsprechenden SAS-Prozeduren berechnet werden. So wurden im Hirndruck-Beispiel die Steigungsparameter mit PROC GLM geschätzt. Andere *summary statistics* wie die *area under the curve* (AUC) oder der von uns vorgeschlagene Gleichgewichtspunkt allerdings müssen in Data steps, SAS/IML oder SQL programmiert werden.

Für die Berechnung des Gleichgewichtspunktes haben wir das SAS-Macro GWPKT entwickelt, das auf Anfrage von den Autoren zur Verfügung gestellt wird. Das Macro benötigt nur SAS/BASE, da es in Data steps programmiert wurde. GWPKT wurde für die Berechnung der Gleichgewichtspunkte im Tumormarker-Beispiel verwendet.

## 5. Diskussion

Longitudinale Daten sind von großem klinischen Interesse. Aufgrund ihrer Abhängigkeitsstruktur sind sie aber häufig nicht einfach auszuwerten. Durch die gezielte Wahl von *summary statistics* kann die Auswertung stark vereinfacht und der klinische Informationsgehalt bewahrt werden. Dies veranschaulichen die beiden angeführten Beispiele aus der klinischen Praxis.

Mit dem Gleichgewichtspunkt wurde eine neue *summary statistic* vorgeschlagen, deren Anwendbarkeit am Beispiel demonstriert wurde. Für die Berechnung wurde das SAS-Macro GWPKT bereitgestellt. Die Verwendung von *summary statistics* bei stark variierenden Anzahlen von Beobachtungen für verschiedene Patienten ist Gegenstand weiterer Untersuchungen.

## Danksagungen

Wir danken Herrn Rolf Spitzer (Abteilung Medizinische Biometrie, Universität Heidelberg) für die technische Unterstützung bei der Erstellung des Tumormarker-Beispiels. Darüberhinaus sind wir Herrn Dr.med.Thorsten Steiner (Neurologische Klinik, Universität Heidelberg) dankbar für die fachlichen Erläuterungen und die Bereitstellung der Daten zum Hirndruck-Beispiel.

## Literatur

1. Matthews JNS, Altman DG, Campbell MJ, Royston P (1990). Analysis of serial measurements in medical research. *British Medical Journal* 300, 230-235.
2. Diggle PJ, Liang KY, Zeger SL (1994). *Analysis of longitudinal data*. Oxford University Press, Oxford.
3. Lindsey JK (1999). *Models for repeated measurements*. Oxford University Press, Oxford, 2.Auflage.
4. Albert PS (1999). Tutorial in biostatistics - longitudinal data analysis (repeated measures) in clinical trials. *Stat Med* 18:1707-1732.
5. Omar RZ, Wright EM, Turner RM, Thompson SG (1999). Analysing repeated measurements data: a practical comparison of methods. *Stat Med* 18:1587-1603.
6. Dunn G, Pickles A (1998). Longitudinal data analysis, overview. in: Armitage P, Colton T (eds). *Encyclopedia of Biostatistics*. Wiley, Chichester.
7. Brunner E, Langer F (1999). *Nichtparametrische Analyse longitudinaler Daten*. Oldenbourg Verlag, München.
8. Matthews JNS, Altman DG, Campbell MJ, Royston P (1990). Analysis of serial measurements in medical research. *BMJ* 300:230-235.
9. Altman DG (1991). *Practical statistics for medical research*. Chapman and Hall, London.
10. Schüler-Duden - *Die Physik* (1974). BI, Mannheim:170.
11. Matthews JNS (1993). A refinement to the analysis of serial data using summary measures. *Stat Med* 12:27-37.
12. Scheurlen H, Fateh-Moghadam A, Fiedler H et al. (1996). Die Varianzanalyse als statistisches Verfahren zur Beurteilung von Tumormarkertests. in: Spanuth E (ed.) *Malignome und Hämostase – 11.Heidelberger Symposium – Neue Entwicklungen in der Hämostaseologie*, Springer, Heidelberg:78-90.
13. Schwab S, Schwarz S, Bertram M, Spranger M, Hacke W (1999). Moderate Hypothermie zur Behandlung des malignen Mediainfarktes. *Der Nervenarzt* 70(6):539-546.
14. Steiner T, Friede T, Manara R, Hacke W (2000). Controlled rewarming after moderate hypothermia improves outcome in stroke patients and hemispheric edema (in Vorbereitung).