

F 20895

# INFORMATIK BIOMETRIE und EPIDEMIOLOGIE IN MEDIZIN UND BIOLOGIE

Offizielles Organ  
der Deutschen Gesellschaft für  
Medizinische Informatik,  
Biometrie und Epidemiologie (GMDS) e.V.

**Persönliches Exemplar für Mitglieder  
der Deutschen Gesellschaft für  
Medizinische Informatik,  
Biometrie und Epidemiologie (GMDS) e.V.,  
darf nicht in öffentlichen Bibliotheken  
eingestellt werden.**

Gustav Fischer Verlag  
Verlag Eugen Ulmer Stuttgart

**Band 28**  
**Heft 2/1997**  
ISSN 0943-5581

# INFORMATIK BIOMETRIE und EPIDEMIOLOGIE

IN MEDIZIN UND BIOLOGIE

Gustav Fischer Verlag Jena  
Verlag Eugen Ulmer Stuttgart

Schriftleitung:  
Prof. Dr. Hans Geidel, Stuttgart  
Prof. Dr. Walter Lehmacher, Köln  
Dipl.-Math. Joachim Vollmar, Mannheim

## Herausgeber

P. Bauer (Wien) · M. Blettner (Heidelberg) · J. Dudeck (Gießen) · U. Feldmann (Homburg) · H. Geidel (Stuttgart)  
R. Haux (Heidelberg) · W. Lehmacher (Köln) · J. Michaelis (Mainz) · W. Stucky (Karlsruhe) · H. Thöni (Hohenheim)  
J. Vollmar (Mannheim) · H.-E. Wichmann (München)

## Wissenschaftlicher Beirat

H. Becher (Heidelberg) · J. Berger (Hamburg) · W. van Eimeren (Neuherberg) · U. Ferner (Basel)  
I. Guggenmoos-Holzmann (Berlin) · H. Haußmann (Hohenheim) · H.-W. Hense (Münster)  
P. Jensch (Oldenburg) · K.-H. Jöckel (Essen) · C. O. Köhler (Heidelberg) · W. Köhler (Gießen)  
W. Maurer (Basel) · R. Mösges (Aachen) · O. Richter (Braunschweig) · H. Rundfeldt (Hannover)  
M. Schumacher (Freiburg) · S. Stiehl (Hamburg) · Th. Tolxdorff (Berlin) · H.-D. Unkelbach (Geisenheim)  
H. F. Utz (Hohenheim) · J. Wahrendorf (Heidelberg)

## Inhaltsverzeichnis 2/1997

Editorial		49
Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme	P. Roebruck, M. Elze, D. Hauschke, F. Leverkus, M. Kieser	51
Sample Size Determination on Survival Time Data – A Review	S. Oellrich, F. Freischläger, A. Benner, M. Kieser	64
Sample size estimation for nanoparametric tests	W. Saueremann	86
An Overview of Software Tools for Sample Size Determination	C. Ortseifen, T. Bruckner, M. Burke, M. Kieser	91
Buchbesprechungen/Book reviews		119

## Redaktion

**Schriftleiter** (verantwortlich im Sinne des Presserechts): Prof. Dr. H. Geidel, c/o Verlag Eugen Ulmer, Postfach 70 05 61, 70574 Stuttgart, Tel. 47 28 80, Verantwortlich für die Mitteilungen der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.: Dipl. rer. pol. F. Stobrawa, Herbert-Lewin-Straße 1, 50931 Köln, Tel. 48 04-4 30.

**Verlag:** Vertrieb und Werbung: Gustav Fischer Verlag GmbH & Co. KG, Niederlassung Jena, PF 100537, D-07705 Jena; Tel. (036 41) 626-3; Fax (036 41) 62 65 00; e-mail: office.j@gfischer.de. Herstellung und Anzeigen: Verlag Eugen Ulmer GmbH & Co., Wollgrasweg 41, 70599 Stuttgart, Tel. (07 11) 45 07-0, Postscheckkonto Stuttgart 74 63-700, Zürich 80-47072, Wien 1083,662 Deutsche Bank AG, Stuttgart, Kto. 14/76 878, Südwesbank AG, Stuttgart, Kto. 21 000, Herstellung: Siegfried Märkisch, Verantwortlich für die Anzeigen: Dieter Boger, Alexandra Signore, Tel. (07 11) 45 07-1 44, z. Z. ist die Anzeigenpreislise Nr. 10 gültig. Anzeigenschluß: am 20. der Monate Januar, April, Juli, Oktober.

**Druck:** Druckhaus "Thomas Müntzer" GmbH, Neustädter Straße 1-4, 99947 Bad Langensalza, Telefon (036 03) 81 33 81/82, 84 23 21/22/23.

**Abonnementsverwaltung:** SFG-Servicecenter Fachverlage GmbH, Zeitschriftenvertrieb: Barbara Dressler, Villengang 2 D-07745 Jena; Telefon: (036 41) 62 64 44; Fax (036 41) 62 64 43.

**Bezugshinweise:** Das Abonnement gilt bis auf Widerruf oder wird auf Wunsch befristet. Die Lieferung der Zeitschrift läuft weiter, wenn sie nicht bis zum 31. 10. eines Jahres abbestellt wird.

**Erscheinungsweise:** (1997) 1 Jahrgang mit 4 Heften.

**Abo-Preise:** (1997) 1 Jahrgang 336,- DM / 2468,- ÖS / 324,50 sFr.; Einzelheftpreis 95,- DM / 694,- ÖS / 91,50 sFr. Alle Preisangaben verstehen sich zuzüglich Versandkosten.

Vorzugspreis für Mitglieder der Deutschen Region der Internationalen Biometrischen Gesellschaft auf Anfrage beim Verlag.

Folgende Kreditkarten werden zur Zahlung akzeptiert: Visa/Eurocard/Mastercard/American Express (bitte Kartenummer und Gültigkeitsdauer angeben).

**Bankverbindung:** Deutsche Bank AG Jena, Konto-Nr. 6 284 707, BLZ 820 700 00.

**Indexed in:** „Biological Abstracts“ und „Current Index to Statistics“.

**Copyright:** Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt: 1.) Mit der Abgabe des Manuskripts versichert der Autor, daß er

allein befugt ist, über die urheberrechtlichen Nutzungsrechte an seinem Beitrag, einschließlich eventueller Bild- und anderer Reproduktionsvorlagen zu verfügen und daß der Beitrag keine Rechte Dritter verletzt. 2.) In Erweiterung von § 38 Abs. 1 UrhG räumt der Autor dem Verlag für die Dauer des Urheberrechts das räumlich und mengenmäßig unbeschränkte Recht der Vervielfältigung und Verbreitung (Verlagsrecht) beziehungsweise der unkörperlichen Wiedergabe der Beiträge ein, auch zur Verwertung außerhalb der Zeitschrift, für die er ursprünglich bestimmt war. Die Übertragung erfolgt auf die Dauer von fünf Jahren ausschließlich, 3.) Im Rahmen von Ziffer 2 räumt der Autor dem Verlag ferner die ausschließlichen Nutzungsrechte am Beitrag ein, so a) das Recht der Übersetzung in Fremdsprachen, das Recht zum ganzen oder teilweisen Vorabdruck und Nachdruck – auch in Form eines Sonderdrucks –, zur Übersetzung in andere Sprachen, zu sonstiger Bearbeitung und zur Erstellung von Zusammenfassungen (Abstracts), b) das Recht zur Veröffentlichung einer Mikrokopie-, Microfiche- und Microforma-Ausgabe, zur Nutzung im Wege von Bildschirmtext, Videotext und ähnlichen Verfahren, zur Aufzeichnung auf Bild- und/oder Tonträger und zu deren öffentlicher Wiedergabe, c) das Recht zur maschinenlesbaren Erfassung und elektronischen Speicherung auf einem Datenträger und in einer eigenen oder fremden Online-Datenbank, das Recht Teile des Beitrages (Beitragskopf mit Zusammenfassungen und Schlüsselwörter in deutsch und englisch) ins Internet zu stellen, zum Download in einem eigenen oder fremden Rechner zur Wiedergabe am Bildschirm, sei es unmittelbar oder im Wege der Datenfernübertragung, sowie zur Bereithaltung in einer eigenen oder fremden Online-Datenbank zur Nutzung durch Dritte, d) das Recht zu sonstiger Vervielfältigung, insbesondere durch fotomechanische oder ähnliche Verfahren und zur Nutzung im Rahmen eines sogenannten Kopier- und/oder Videovertrages, e) das Recht zur Vergabe der vorgenannten Nutzungsrechte an Dritte im In- und Ausland sowie die von der Verwertungsgesellschaft WORT wahrgenommenen Rechte einschließlich der entsprechenden Vergütungsansprüche. Der Verlag wird über die Rechte gemäß Punkt 2 und 3a nur mit Zustimmung des Autors verfügen und sich um eine angemessene Honorierung bemühen. Fotokopien für den persönlichen Gebrauch dürfen nur von den einzelnen Beiträgen oder Teilen daraus als einzelne Kopien erstellt werden.

Printed in Germany  
© 1997 Verlag Eugen Ulmer Stuttgart; Gustav Fischer Verlag Jena

## Editorial

Die Berechnung der Fallzahl spielt eine Schlüsselrolle bei der Planung klinischer Studien. Nur wenn eine ausreichende Zahl von Patienten in eine Studie eingeschlossen wird, besteht eine reelle Chance zur Erreichung des Prüfungsziels. Auf der anderen Seite gebieten es ethische Überlegungen sowie zeitliche und budgetäre Restriktionen, den Stichprobenumfang nicht höher als notwendig zu wählen. Die Verwendung adäquater statistischer Verfahren zur Fallzahlbestimmung ist deshalb essentiell, und die Entwicklung und Verfeinerung entsprechender Methoden ist ein wichtiger und aktiver Zweig biometrischer Forschung. Im Current Index of Statistics finden sich unter dem Stichwort „Sample size“ in den letzten Jahren jeweils zwischen 60 und 80 neue Publikationen. Im Juni 1993 konstituierte sich als Untergruppe der AG Pharmazeutische Forschung der Deutschen Region der Internationalen Biometrischen Gesellschaft die Ad-hoc-Gruppe „Fallzahlplanung“. Ziel dieser Arbeitsgruppe war es, die umfangreiche Literatur zur Fallzahlbestimmung zu sichten und vergleichend gegenüberzustellen. Die nachfolgenden vier Überblicksarbeiten zu den Themen Äquivalenzfragestellung, Überlebenszeitanalysen und nicht-parametrische Verfahren sowie zur Software fassen die Ergebnisse zusammen und sollen den Anwender bei der Auswahl der geeigneten Methodik und Software unterstützen.

Testverfahren zum Äquivalenznachweis waren in den vergangenen Jahren Gegenstand intensiver Forschungsaktivitäten. Die Arbeit von ROEBRUCK et al. gibt eine kommentierte Übersicht über Literatur zur Bestimmung des Stichprobenumfangs für Studien mit Äquivalenzfragestellungen. Es werden Übersichtsarbeiten zu diesem Thema und Verfahren für metrische und für kategoriale Beobachtungen vorgestellt. Die Inhalte der referierten Publikationen werden nach einheitlichen Kriterien tabellarisch dargestellt, und es werden Empfehlungen für häufig verwendete Verfahren gegeben.

Die zahlreiche Literatur zur Fallzahlplanung für Überlebenszeitanalysen ist für Nicht-Spezialisten kaum überschaubar. Dies liegt zu einem Teil daran, daß im Vergleich zu anderen Anwendungssituationen zusätzliche Parameter, wie z. B. die Verteilung der Rekrutierung, der Zensierungen und des Follow-up, die Fallzahl beeinflussen. Daraus resultieren unterschiedliche Verfahren zur Fallzahlbestimmung, die den jeweiligen Kenntnisstand über diese Designparameter zum Zeitpunkt der Planungsphase berücksichtigen. In die Übersichtsarbeit von OELLRICH et al. wurden 105 Publikationen zum Thema Fallzahlplanung bei Überlebenszeitfragestellungen, die zwischen 1967 und 1996 erschienen sind, aufgenommen. Sie wurden entsprechend ihres Inhaltes charakterisiert und mit Deskriptoren versehen. Der Anwender soll damit in die Lage versetzt werden, bei Vorliegen eines speziellen Problems die für ihn relevante Literatur schnell einzukreisen. Anhand eines Reviews der chronologischen Abfolge der Publikationen werden die wichtigsten Arbeiten und ihre wechselseitigen Bezüge sowie neue Entwicklungslinien dargestellt. Damit kann sich der Leser ohne aufwendiges Literaturstudium über den aktuellen Stand der Forschung informieren und seine Kenntnisse bei Bedarf anhand der Originalarbeiten zu einzelnen Themenbereichen vertiefen.

Bei der Auswertung klinischer Studien werden nicht-parametrische Verfahren sehr häufig eingesetzt. Dennoch ist das Methodenspektrum zur Bestimmung der zugehörigen Stichprobenumfänge verglichen mit den vorangehenden Anwendungsbereichen eher schmal. Die Arbeit von SAUERMAN stellt drei unterschiedliche Ansätze vor, die für dieses Problem in der Literatur beschrieben sind: Fallzahlberechnung unter Zugrundelegung parametrischer Modelle, Fallzahlberechnung unter Alternativhypothesen, die in modellunabhängigen Maßzahlen, wie z. B. der stochastischen Überlegenheit, formuliert sind, und Fallzahlberechnung mit dem Bootstrap-Verfahren.

Die Berechnung des Stichprobenumfangs erfolgt heutzutage fast ausschließlich per Computer. Das Spektrum der verfügbaren Software zur Fallzahlbestimmung reicht vom programmierbaren Taschenrechner auf der einen Seite bis zum Expertensystem zur Versuchsplanung, das ein Modul zur Berechnung des Stichprobenumfangs enthält, auf der anderen Seite. Die Arbeit von ORTSEIFEN et al. soll dem Anwender helfen, aus der Vielzahl der Softwareprodukte eine auf die jeweiligen Bedürfnisse zugeschnittene Auswahl zu treffen. Hierzu werden Leistungsumfang und Preis der Programme gegenübergestellt, und für einige Standardsituationen werden die Ergebnisse der Berechnung verglichen. Darüber hinaus werden Internet-Adressen angegeben, über die nicht-kommerzielle Programme angefordert werden können.

Karlsruhe, Februar 1997

Meinhard Kieser

## ORIGINALARBEITEN

# Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme

P. Roebruck<sup>1)</sup>, M. Elze<sup>2)</sup>, D. Hauschke<sup>3)</sup>, F. Leverkus<sup>4)</sup>, M. Kieser<sup>5)</sup>

### Zusammenfassung

*Verfahren zum Äquivalenznachweis bilden eine relativ junge Gruppe von Methoden der angewandten Statistik. Für sie ist, wie für alle statistischen Entscheidungsverfahren, die Bestimmung des adäquaten Stichprobenumfangs im Rahmen der Versuchsplanung von besonderer Bedeutung. Diese Arbeit gibt eine kommentierte Übersicht über die bisher erschienene Literatur zur Fallzahlplanung bei Äquivalenzproblemen. Dem geht eine kurze methodische Einführung voraus.*

### Summary

*Methods for proving equivalence are a relatively young part of applied statistics. The calculation of the adequate sample size for equivalence studies is clearly as important as it is for other empirical investigations. After a short methodical introduction to equivalence problems, this paper presents a survey of the literature on sample size determination for equivalence problems with comments on practical aspects.*

### Einführung

Ziel dieser Arbeit ist es, einen kommentierten Überblick über diejenige Literatur, die dem Biometriker bei der Planung von Studien zum Äquivalenznachweis (z. B. Bioäquivalenz oder therapeutische Äquivalenz) hilfreich sein kann, zu geben. Sie gliedert sich nach den statistischen Charakteristika der Problemstellungen und enthält insbesondere Angaben zu der Art, wie jeweils das Problem der Fallzahlberechnung gelöst worden ist, und zur An-

<sup>1)</sup> Medizinische Informatik, Universität Heidelberg/Fachhochschule Heilbronn

<sup>2)</sup> Zentrallaboratorium Deutscher Apotheker, Eschborn

<sup>3)</sup> BYK Gulden Pharmazeutika, Konstanz

<sup>4)</sup> Pfizer GmbH, Karlsruhe

<sup>5)</sup> Dr. W. Schwabe GmbH, Karlsruhe

wendbarkeit dieser Lösung für den praktisch arbeitenden Biometriker. Publikationen, die sich mit statistischen Verfahren zu Äquivalenzfragestellungen befassen, ohne Angaben zur Berechnung des benötigten Stichprobenumfangs zu machen, sind nicht berücksichtigt. Aufgenommen sind Arbeiten, die einen eigenen Beitrag zum Fallzahlproblem bei Äquivalenzfragestellungen leisten, und zwar im Rahmen der Neyman-Pearson-Theorie, der Bayesverfahren und der Selektionsprozeduren. Andere Arbeiten, wie z. B. Nominogramme zu einer bekannten simplen Formel, die kaum neue Informationen bieten, werden nur im Literaturverzeichnis unter ‚Weitere‘ aufgeführt. Beiträge zu Tagungsbänden, die ähnlich auch in einer Zeitschrift publiziert sind, werden nicht eigens aufgeführt. Da in diesem Heft eine eigenständige Übersicht über die zur Fallzahlplanung verfügbare Software erscheint (ORTSEIFEN et al. (1997)), beschränken wir uns auf die Erwähnung von Programmen, die durch die Autoren der besprochenen Arbeiten angeboten werden. Ziel dieser Arbeit ist ein Überblick über die allgemein zugängliche Literatur bis 1995 einschließlich, ohne jedoch letzte Vollständigkeit anzustreben und „entlegene bibliographische Winkel“ zu erschließen.

Der eigentlichen Literaturübersicht möchten wir eine Einführung in die Fragestellung und allgemeine Lösungsansätze vorausschicken. Besonders ersteres wird dabei sehr kurz gehalten, da unsere Übersicht nicht für statistische Laien gedacht ist. Eine ausführliche Einführung in das Problem der Äquivalenzprüfung findet man z. B. bei WELLEK (1994). Darüber hinaus enthalten viele der zitierten Arbeiten motivierende Einführungen zu der Verwendung spezieller Verfahren für Äquivalenzfragestellungen.

Monographien sowie Übersichtsarbeiten tutorieller oder vergleichender Art werden in einem eigenen Abschnitt behandelt. Sie beinhalten – auch zusammengenommen – keine vollständige Darstellung der Fallzahlprobleme und ihrer Lösungen, so daß wir glauben, mit unserer Aufstellung einen nützlichen Beitrag leisten zu können. Dem Leser wird darüber hinaus auffallen, daß, abgesehen von den Artikeln, die sich mit Selektionsprozeduren befassen, ausnahmslos 1- oder 2-Stichprobenprobleme behandelt werden. Dies hat sicherlich zwei Ursachen. Die Fallzahlkalkulation für Mehrstichprobenprobleme stellt – schon wenn es um den Nachweis von Unterschieden geht – eine erhebliche Herausforderung dar, und zwar eher inhaltlicher als statistischer Art. Die inhaltlichen Probleme werden bei der Betrachtung der Äquivalenz von  $k$  Verteilungen noch größer. Dies beginnt schon bei der Definition der Äquivalenz.

## Methodische Aspekte

### Das Problem des statistischen Äquivalenznachweises

Klassische statistische Testverfahren dienen dem Nachweis von Unterschieden zwischen Verteilungen anhand eines Parameters  $\vartheta$ , der in gewissem Sinn den Abstand zwischen verschiedenen Verteilungen mißt (sogenanntes Effektmaß). Dabei werden die Versuche so konzipiert, daß unter Gleichheit der Verteilungen in jedem Fall die Irrtumswahrscheinlichkeit  $\alpha$  für den Fehler erster Art eingehalten wird, während die Einhaltung einer Mindesttrennschärfe  $1 - \beta$  bei einem bestimmten vorgegebenen Effekt planerischen Unwägbarkeiten unterliegt. Daraus resultiert, daß nichtsignifikante Unterschiede im allgemeinen nicht ausreichend für den Nachweis von Äquivalenz sind. Ursache ist die Forderung, im Gegensatz zum Test auf Unterschiede, die Irrtumswahrscheinlichkeit erster Art nicht für die Gleichheit der Verteilungen (meist bei  $\vartheta = 0$  oder  $\vartheta = 1$ ), sondern an den Rändern  $\vartheta_1, \vartheta_2$  eines sogenannten Äquivalenzbereichs, d. h. eines Intervalls ( $\vartheta_1, \vartheta_2$ ) zu kontrollieren (bei einseitigen Fragestellungen  $\vartheta_1 = -\infty$  oder  $\vartheta_2 = \infty$ ). Dort hängen die Verteilungen aber häufig nicht nur von  $\vartheta$ , sondern auch von Störparametern

ab, die bei der Planung oft nur grob abgeschätzt werden können. Nur wenn die Verteilungen der Teststatistik ausschließlich durch  $\vartheta$  parametrisiert sind, kann bei entsprechender Wahl von  $\alpha$  und  $\beta$  ein nichtsignifikantes Ergebnis des Unterschiedstests als Äquivalenznachweis interpretiert werden. Allerdings sind die resultierenden Verfahren nicht unbedingt optimal.

### Lösungsansätze über Testverfahren

Mathematisch besonders befriedigend ist es, das Effektmaß  $\vartheta$  so zu wählen, daß die Verteilungen der (aus der Anwendung allgemeiner Prinzipien der Mathematischen Statistik) resultierenden Teststatistiken gerade durch  $\vartheta$  parametrisiert sind, um dann in gewissem Sinne optimale Verfahren konstruieren zu können (siehe WELLEK (1994), aber auch schon LEHMANN (1986)). Die Berechnung der Trennschärfe dieser Verfahren geschieht dann über die Verteilungsfunktion und die Quantile der entsprechenden Verteilungsfamilie, womit eine Fallzahlplanung zumindest iterativ möglich ist.

Derartige Effektmaße, wie z. B. standardisierte Erwartungswertdifferenz oder Odds Ratio, entsprechen allerdings inhaltlich nicht immer den Vorstellungen der Anwender, die häufig ihre Probleme eher in Form von z. B. Risikodifferenzen oder Erwartungswertdifferenzen adäquat formulieren können. In diesem Fall führt, insbesondere bei einseitigen Äquivalenzproblemen, die Methode der verschobenen Null-Hypothese zumindest approximativ zum Erfolg. Bei zweiseitigen Problemen hilft das Verfahren der zwei einseitigen Tests (äquivalent dazu die  $(1 - 2\alpha)$ -Konfidenzintervall-Inklusionsregel). Bei zumindest asymptotisch normalverteilten Teststatistiken ist im einseitigen Fall die Approximation von Power und Stichprobenumfang über die nichtzentrale Verteilung möglich. Bei zweiseitigen Testproblemen wie überhaupt bei dichotomen Beobachtungen sind exakte Berechnungen aufwendig (für den Stichprobenumfang meist nur iterativ über die Trennschärfe), und Abschätzungen sind hilfreich.

Über diese allgemeinen Ansätze hinaus sind natürlich auch spezielle Verfahren denkbar. Einige haben sich allerdings auf Dauer nicht durchsetzen können, wie z. B. die Tests von Anderson/Hauck und Patel/Gupta sowie das Inklusionsverfahren für symmetrische Konfidenzintervalle von Westlake.

Verwendet man schon klassische Tests mit verschobenen Nullhypothesen zum Zwecke des Äquivalenznachweises, so liegt es nahe, ebenso klassische gruppensequentielle Verfahren, die für solche Tests existieren, zur sequentiellen Äquivalenztestung zu verwenden. Diese nach den bekannten Grundmustern arbeitenden Verfahren sind in einigen hier zitierten Arbeiten beschrieben. Zur praktischen Durchführung muß der Anwender allerdings auch auf die ursprünglichen Publikationen zu gruppensequentiellen Tests auf Unterschiede zurückgreifen.

Zu den approximativen Verfahren muß noch festgehalten werden, daß diese selbstverständlich nicht eindeutig bestimmt sind, und daß es im allgemeinen auch keine eindeutig besten Approximationen gibt. Demnach macht es auch nicht viel Sinn, ebenfalls approximative Trennschärfe- bzw. Fallzahlberechnungen unabhängig von den approximativen Tests, für die sie entwickelt wurden, zu betrachten. Tatsächlich sollten zur Bewertung solcher Verfahren die Präzision, mit der sie die Irrtumswahrscheinlichkeit erster Art einhalten, ihre Trennschärfe und die Güte der zugehörigen Fallzahlabschätzung berücksichtigt werden.

### Bayesverfahren

Wie WELLEK (1994) treffend formuliert, bietet das Äquivalenzproblem dem Bayesiarer keine besonderen Schwierigkeiten, da seine Interpretation der Versuchsergebnisse im wesentlichen auf der A-posteriori-Wahrscheinlichkeit des Äquivalenzbereichs ba-

siert und er keine Irrtumswahrscheinlichkeiten irgendwelcher Art in seine Überlegungen einbezieht. Der einzige hier zitierte Ansatz zur Fallzahlberechnung für ein Bayesverfahren stützt sich tatsächlich auf ein frequentistisches Gütekriterium. (Siehe unten)

### Selektionsprozeduren

Selektionsprozeduren sind vom Ansatz her geeignet, Gruppen homogener Verteilungen aus einer größeren Anzahl herauszufinden. Einige sind folglich auch für Äquivalenzprobleme zu adaptieren. Die Lösung eines Selektionsproblems besteht nun hauptsächlich in der Fallzahlbestimmung, weswegen Publikationen, die konkrete Lösungen eines Selektionsproblems zur Äquivalenzprüfung angeben, automatisch den benötigten Stichprobenumfang mitliefern.

### Literaturüberblick

#### Tutorielle und vergleichende Übersichtsarbeiten, Monographien

Es gibt eine Reihe von Übersichtsarbeiten tutoriellen oder vergleichenden Charakters, die unter anderem oder auch ausschließlich Fallzahlplanungen für Äquivalenzfragestellungen beinhalten. Die Monographie von WELLEK (1994) behandelt ausschließlich statistische Methoden zum Äquivalenznachweis.

DONNER (1984) greift in einem Äquivalenzabschnitt seines Übersichtsartikels zur Fallzahlplanung die Vorschläge von MAKUCH/SIMON (1978) und BLACKWELDER (1982) auf. FRICK (1987) vergleicht die Trennschärfen der Äquivalenztests nach Anderson/Hauck und Patel/Gupta. Diese haben jedoch nur noch historische Bedeutung und werden hier der Vollständigkeit wegen aufgeführt.

CHOW/LIU (1992) widmen ein Kapitel ihres Buches, das sich mit statistischen Aspekten der Planung und Auswertung von Bioverfügbarkeits-/Bioäquivalenzstudien beschäftigt, der Power- und Fallzahlbestimmung. Powerberechnungen zu SCHUIRMANN'S (1987) 2-einseitige-t-Tests-Prozedur<sup>6)</sup> und zur Prozedur nach Anderson/Hauck werden erläutert und verglichen mit dem Ergebnis, daß Schuirmann's 2-einseitige-t-Tests-Prozedur der Anderson/Hauck-Prozedur vorzuziehen ist. Fallzahlberechnungen für Schuirmann's Prozedur mittels exakter Methode und Approximationsformel, die im wesentlichen auf PHILLIPS (1990) bzw. LIU/CHOW (1992) zurückgehen, werden dargestellt. Stichprobenumfänge können einer Tabelle entnommen werden.

In seiner Monographie gibt WELLEK (1994) einen Überblick über mathematisch gut begründete Verfahren, die größtenteils von ihm selbst entwickelt oder ausgearbeitet wurden. Er gibt Programme und Tabellen zur Bestimmung der kritischen Grenzen und der Trennschärfe für eine Reihe von Problemen an. Im einzelnen werden behandelt: das Einstichprobenproblem für Exponentialverteilungen, Binomialverteilungen und Poissonverteilungen für den Vergleich mit jeweils einem hypothetischen Parameter. Zu diesen Einstichprobentests werden SAS-Makros zur Berechnung der Power angegeben. Da es sich aber um Probleme handelt, die in der Praxis geringe Bedeutung haben und darüber hinaus als einparametrische Probleme bei leicht berechenbaren Verteilungen kaum Schwierigkeiten bereiten, haben wir darauf verzichtet, sie in die tabellarische Übersicht zu übernehmen.

Für das verbundene Zweistichprobenproblem werden der Vorzeichentest, der McNemar-Test, der verbundene t-Test und der asymptotische Vorzeichenrangtest auf Äquivalenz

<sup>6)</sup> Diese ist bekanntlich äquivalent zu einer schon von WESTLAKE (1972) vorgeschlagenen Inklusionsregel für Konfidenzintervalle.



behandelt. Für das Zweistichprobenproblem wird der unverbundene t-Test auf Äquivalenz betrachtet. Die SAS-Makros zu Trennschärferechnungen können iterativ genutzt werden, um eine Stichprobenumfangsplanung durchzuführen. Zum F-Test für den Vergleich zweier Varianzen auf der Basis unverbundener Stichproben gibt es ein Programm und eine Tabelle für die Power. Darüber hinaus gibt Wellek Fortran-Programme zur Berechnung des Stichprobenumfangs für den exakten Test nach Fisher (sowohl einseitig als auch zweiseitig) an. Da hier mit der unbedingten exakten Verteilung gearbeitet wird, ist bei größerem Stichprobenumfang natürlich ein erheblicher Zeitaufwand erforderlich, insbesondere da nur eine iterative Lösung über die unbedingte, exakte Trennschärfe möglich ist.

Die Verfahren der Wellek'schen Monographie für binomiale und normalverteilte Beobachtungen basieren auf der Odds Ratio bzw. der standardisierten Erwartungswertdifferenz als Effektmaß und erfüllen wünschenswerte Optimalitätskriterien. Im Gegensatz dazu befassen sich die meisten im folgenden zitierten Arbeiten eher mit Situationen, in denen keine mathematisch vollständig befriedigenden Lösungen möglich sind.

Im Hinblick auf die Fallzahlplanung muß zu dem Buch von Wellek einschränkend gesagt werden, daß die angegebenen Programme die Trennschärfe der Verfahren im wesentlichen nur für die Situation der exakten Übereinstimmung der zu vergleichenden Verteilungen berechnen, was für die praktische Anwendung eine gewisse Einschränkung darstellt. Nach Angaben des Autors ist es einfach, sie für andere Alternativhypothesen zu modifizieren.

Einen Vergleich von drei approximativen Verfahren (BLACKWELDER (1982), RODARY et al. (1989), FARRINGTON/MANNING (1990)) bieten ROEBRUCK/KÜHN (1995) für die Risikodifferenz bei Binomialverteilungen. Es stellt sich heraus, daß das von Farrington/Manning vorgestellte Verfahren im allgemeinen trotz gelegentlich konservativen Verhaltens die besten Ergebnisse erzielt. Der Test von Blackwelder ist insbesondere für kleinere Stichprobenumfänge stark antikonservativ, die Fallzahlformel für größere Stichprobenumfänge durchaus anwendbar. Der Test von Rodary et al. ist oft stark antikonservativ, und die nach der dazu gehörigen Formel berechneten Fallzahlen sind stets erheblich zu hoch (selbst bei Gesamtfallzahlen von 2000 oder 3000). Offensichtlich ist diese Fallzahlformel für den dort angegebenen Test ungeeignet.

NAM (1995) schließt auf der Basis eines empirischen Vergleichs von sehr geringem Umfang, daß die von NAM (1994) angegebene Fallzahlformel für den Test der Risikodifferenz im Falle ungeschichteter Stichproben der Formel von Farrington/Manning geringfügig überlegen ist. Er zeigt weiter für eine Reihe von Parameterkonstellationen, daß eine neue Fallzahlformel für eine gepoolte Auswertung stratifizierter Stichproben im balancierten Fall größere Werte liefert als die für die stratifizierte Auswertung aus NAM (1994).

FRICK (1994) betrachtet einen auf der Normalapproximation für die Verteilung der beobachteten Risikodifferenz (deren Varianz als bekannt angenommen wird) basierenden Test und berechnet dessen exakte Power über die Binomialverteilungen. Er definiert den benötigten Stichprobenumfang  $N$  so, daß die gewünschten Irrtumswahrscheinlichkeiten für kein  $n \geq N$  überschritten werden. Dies ist eine ungebräuchlich strenge Forderung an stetige Approximationen diskreter Verteilungen. Frick diskutiert dann im Hinblick auf seinen Test und sein Fallzahlkriterium diverse Approximationsformeln zur Berechnung des Stichprobenumfangs und kommt zu dem Schluß, daß selbst die Formel von Rodary et al. zu kleine Stichprobenumfänge liefere. Das verwundert nicht, da diese Fallzahlapproximationen nicht für seinen Ansatz gedacht sind.

Der Diskurs von Frick wird von BOCK (1997) im Äquivalenzkapitel seines Lehrbuchs zur Fallzahlplanung aufgegriffen, schließlich aber mit dem Hinweis beendet, daß der Frick'sche Test ohnehin nicht durchführbar sei, was sicherlich richtig ist, da zu seiner Durchführung die Kenntnis eines Verteilungsparameters erforderlich wäre. Konkret be-

handelt Bock in seinem Buch die Fallzahlapproximation nach WESTLAKE (1988) und die exakte Power der  $(1 - 2\alpha)$  Konfidenzintervall-Inklusionsregel (entspricht zwei einseitigen Tests, siehe oben) für die nicht standardisierte Erwartungswertdifferenz beim Vergleich zweier Normalverteilungen. Die Diskette zum Buch enthält zu diesem Problem ein Programm zur iterativen Fallzahlberechnung basierend auf der numerischen Integration zur exakten Bestimmung der Trennschärfe. Für den Vergleich zweier Binomialverteilungen auf der Basis der Risikodifferenz als Effektmaß verweist Bock auf das stetigkeitskorrigierte approximative Konfidenzintervall von HAUCK/ANDERSON (1986) und gibt ein iteratives Programm zur Fallzahlberechnung basierend auf der exakten Powerkalkulation an.

### Tabellarische Übersicht

Da, wie schon erwähnt, die Möglichkeit der Berechnung der Trennschärfe eines Testverfahrens benutzt werden kann, um iterativ die benötigte Fallzahl für eine zu planende Studie zu bestimmen, sind in der folgenden Übersicht auch solche Arbeiten berücksichtigt, die sich nur mit der Powerberechnung befassen. Bei Arbeiten, die sowohl eine Power- als auch eine Fallzahlkalkulation beinhalten, beschränken wir uns entsprechend dem primären Ziel dieser Übersicht auf letztere.

Bei der Kommentierung der praxisrelevanten Publikationen zur Fallzahlplanung für Äquivalenzfragestellungen erschien es uns für den Leser besonders von Bedeutung, zu charakterisieren, welcher Art die Angaben zur Fallzahlplanung jeweils sind. Eine geschlossene Formel, die relativ leicht berechenbar ist und für alle Parameterkonstellationen anwendbar, ist natürlich von besonderem Nutzen, wobei zwischen exakten und approximativen Formeln unterschieden werden muß. Demgegenüber sind Tabellen und Nomogramme von dem Umfang, wie man sie gewöhnlich in wissenschaftlichen Publikationen findet, in dem Sinne nicht ausreichend, daß die den Anwender gerade interessierenden Parameterkonstellationen oft nicht abgedeckt werden. Aufwendige iterative Verfahren und numerische Integrationen andererseits, die beliebig genaue Ergebnisse liefern können, sind für viele Anwender zu arbeitsintensiv und zeitaufwendig, sofern die Autoren nicht ein entsprechendes Programm zur Verfügung stellen.

### Metrische Beobachtungen (Tabelle 1)

Der größte Teil der Arbeiten befaßt sich mit dem unverbundenen 2-Stichprobenproblem für normalverteilte Beobachtungen oder für Crossover-Versuche unter Verwendung additiver oder multiplikativer Modelle, wobei (für letztere nach Transformation) ebenfalls Normalverteilungen mit gleicher Varianz zugrundegelegt werden. Da letztendlich die Therapievergleiche im Crossover-Modell auf ein unverbundenes 2-Stichprobenproblem, wenn auch mit anderen Varianzen und verdoppeltem Effekt, zurückgeführt werden, erfahren sie in der folgenden Aufstellung keine gesonderte Behandlung. Ausnahmen bilden Arbeiten, bei denen etwa Tabelleneinträge so speziell auf die Crossover-Situation für die Bioäquivalenz zugeschnitten sind, daß sie sich nicht ohne weiteres auf das allgemeine 2-Stichprobenproblem übertragen lassen.

In den beiden Arbeiten von RACINE et al. (1986) und RACINE-POON et al. (1987), die sich mit demselben Bayesverfahren befassen, basiert die Fallzahlberechnung auf der frequentistischen Wahrscheinlichkeit, für den Äquivalenzbereich eine genügend große A-posteriori-Wahrscheinlichkeit (0,95) zu erzielen, um so auf Äquivalenz schließen zu können.

Sämtliche Arbeiten beschränken sich auf die Situation von 2-Stichprobenproblemen mit gleichen unbekanntem Varianzen mit Ausnahme von Selektionsprozeduren (GIANI/FINER (1991), GIANI/STRASSBURGER (1994)). Dort können optional die gemeinsamen Varianzen der  $k$  Stichproben als bekannt vorausgesetzt werden.

Tabelle 1: Literatur zur Fallzahlplanung bei metrischen Beobachtungen

Autor(en)	Modell	Effekt- maß	Verfahren	Power/ Fallz.	Art der Lösung	Anwend- barkeit	Bemerkungen
Anderson/Hauck (1983) <sup>8)</sup>	2-St	Diff	Test	Fallz.	F, a, i	einf	Anders./Hauck- Test
Anderson/Hauck (1990)	COoP		Test	Fallz.	T, a	P	individuelle Bio- äq., 1-Stichpr.-Test für paarige Beob.
Bauer/Bauer (1994)	2-St	Diff + $\sigma_1/\sigma_2$	Test Test	Fallz s. Text	F, a a	einf	Simultane Behandl. beider Effekte
Bock (1997)	CO	Diff	Test	Fallz	F, a, i	P	P zu Westlake (1988)
Bristol (1993)	2-St	Diff	Test	Fallz	F, a	einf	2 1-seit. t-Tests
Diletti/Hauschke/ Steinijans (1991)	CO	Diff	Test	Fallz	T, N, e	einf	2 1-seit. t-Tests
Diletti/Hauschke/ Steinijans (1992)	CO	Diff	Test	Fallz	T, N, e	einf	Ergänzung d. '91er T, N
Esinhart/Chinchilli (1994) (Siehe auch Text)	CO	Diff Quot pro Fall	Tol-Int	Fallz	F, T, e	aufw (T einf)	individuelle Bioäq.: param.+ verteil.freie Toleranzintervalle
Frick (1987) <sup>8)</sup>	2-St	Diff	Test	Power	numInt	sehr aufw	Anders./Hauck- Test
Giani/Finner (1991) (siehe auch Text)	k-St	Diff <sup>7)</sup> SDiff	Test	Fallz	F, T, e	aufw (T einf)	unabhängige oder equikorrelierte Beob.
Giani/Straß- burger (1994) (siehe auch Text)	k-St	Diff <sup>7)</sup> SDiff	Test + Selekt	Fallz	F, T, e	aufw (T einf)	many-one, ähnlich Giani/Finner (1991)
Hauck/Anderson (1984) <sup>8)</sup>	CO	Diff	Test	Fallz	F, a, i	P	Anders./Hauck Test
Hauschke/Steini- jans/Diletti (1992)	CO	Diff	Test	Fallz	F, T, a, i	einf	multiplikativ. Modell 2 1-seit. t-Tests
Henry/Mailhot (1991)	CO	Diff	Test	Fallz	F, T, a, i	einf	additives Modell
Jennison/Turnbull (1993a)	2-St	Diff	GS-Test	Fallz	F, a	aufw	eins., $\sigma$ bek.
Jennison/Turnbull (1993b)	2-St	Diff	GS-Test	Fallz	F, a	aufw	zweis., $\sigma$ bek.
Kanfer/Geertse- ma/Steyn (1988)	2-St	Diff $\sigma$ bek.	Test 2-Schritt	Fallz	F, a	aufw	adapt. Fallz.- schätz.

<sup>7)</sup> Behandlung der Differenzen nur für bekannte Streuungsparameter.<sup>8)</sup> Diese Verfahren sind obsolet und nur der Vollständigkeit und ihrer Bekanntheit wegen aufgeführt.

Tabelle 1 (Fortsetzung)

Autor(en)	Modell	Effektmaß	Verfahren	Power/Fallz.	Art der Lösung	Anwendbarkeit	Bemerkungen
Liu/Chow (1992)	CO	Diff	Test	Fallz	F, T, a, i	einf, P	2 1-seit. t-Tests
Patel/Gupta (1984) <sup>8)</sup>	2-St	Diff	Test	Fallz	A, a, i	aufw, P	Patel/Gupta-Test
Phillips (1990)	CO	Diff	Test	Fallz	T, N, e	einf	2 1-seit. t-Tests
Racine/Grieve/Flühler (1986) und Racine-Poon/Grieve/Flühler/Smith (1987)	CO	Diff	Bayes 2-Schritt	Fallz Power	T, a F	einf aufw	adapt. Fallz.-schätz. zweis.
Schuirman (1990)	CO	Diff	Test	Fallz	F, a, i	einf	2 1-seit. t-Tests
Srinivasan/Langenberg (1986)	2-St	Diff	Test 2-Schritt	Fallz	a	einf	adapt. Fallz.-schätz. 2 1-seit. t-Tests
Steinijans/Hauck/Diletti/Hauschke/Anderson (1992)	CO	Diff	Test	Fallz	T, N, e	einf	2 1-seit. t-Tests Vergl. unterschiedl. Äq.-bereiche
Wellek (1991)	CO	SDiff	Test	Fallz	a, e	P	opt. Tests bei NV
Wellek (1993)	COoP		Test	Fallz	T, a	P	individuelle Bioäq., 1-Stichpr-Test für paarige Beob.
Wellek (1994) (siehe auch Text)	1-St 2-St 2-St	$\mu, \mu/\sigma$ SDiff $\sigma_1/\sigma_2$	Test Test Test	Power Power Power	A, T, e A, T, e A, T, e	P P P	opt. Tests bei NV u. Vorzeichen-Test
Westlake (1988)	CO	Diff	Test	Fallz	F, a, i	einf	2 1-seit. t-Tests
Wie (1993)	2-St	Diff	Test	Fallz	F, T, a	einf	2 1-seit. t-Tests
Yeh/Chen (1993)	CO	Diff	Test	Fallz	T, e	einf	Mehrperioden-CO

## Erläuterungen der Abkürzungen:

2-St	2-Stichprobenproblem	NV	Normalverteilung(en)
CO	Crossover	T	Tabelle
COoP	Crossover ohne Perideneffekte	N	Nomogramm
k-St	k-Stichprobenproblem	a	approximativ
1-St	1-Stichprobenproblem	e	exakt
Diff	unstandardisierte Erwartungswertdifferenz	i	iterativ
SDiff	standardisierte Erwartungswertdifferenz	A	Algorithmus
Quot	Quotient der Erwartungswerte	F	Formel
GS-Test	gruppensequentieller Test	einf	einfach anzuwenden
Bayes	Bayesverfahren	aufw	Anwendung aufwendiger
Tol-Int	Toleranzintervall	P	Autoren bieten ein Programm an
Selekt	Selektionsprozedur		

BAUER/BAUER (1994) schlagen das gleichzeitige Testen auf Lokations- und Streuungsäquivalenz vor. Die Fallzahlberechnung basiert auf zwei einseitigen t-Tests und verwendet (wegen des Tests der Varianzen) die Kontrolle von  $\beta/4$  statt  $\beta/2$ . Es wird eine Approximationsformel für die Wahl des Äquivalenzbereichs der Varianzen angegeben, der für das so berechnete  $n$  gerade noch die jeweilige Power von  $1 - \beta/4$  im Fall der Varianzhomogenität garantiert.

ESINHART/CHINCHILLI (1994) schlagen die Entscheidung für Bioäquivalenz auf der Basis von Toleranzintervallen für individuelle Differenzen oder Quotienten vor. Die Kalkulation der Stichprobenumfänge basiert im verteilungsfreien Fall auf der Forderung, das Intervall möge einen Anteil  $\gamma$  der Werte in der Population mit einer Wahrscheinlichkeit  $\geq 1 - \alpha$  enthalten. Im Fall normalverteilter Beobachtungen wird zusätzlich eine Äquivalenzentscheidung (falls das Toleranzintervall im Äquivalenzbereich enthalten ist) berücksichtigt.

### Kategoriale Beobachtungen (Tabelle 2)

Die vorliegenden Arbeiten behandeln ausschließlich binomialverteilte Beobachtungen und für diese nahezu ausschließlich das unverbundene 2-Stichprobenproblem (Ausnahmen: WELLEK (1994), LU/BEAN (1995)). Sie unterscheiden sich dabei im wesentlichen durch die verwendeten Effektmaße und Approximationsformeln. Die Verfahren für das relative Risiko und die Risikodifferenz sind natürlich auch für Erfolgsraten verwendbar.

### Empfehlungen für häufig verwendete Verfahren

1. Vergleich zweier Normalverteilungen, Effektmaß standardisierte Differenz der Erwartungswerte: Da in dieser Situation ein optimaler Test existiert, kommt kein anderes Verfahren infrage. In WELLEK (1994) findet man die Beschreibung des Tests, Tabellen einiger kritischer Werte und Trennschärfen für diverse Stichprobenumfänge sowie ein SAS-Makro zur Berechnung derselben. Mit letzterem kann die Fallzahlberechnung von SAS-Nutzern iterativ leicht gelöst werden.

2. Vergleich zweier Normalverteilungen, Effektmaß Differenz der Erwartungswerte: Das derzeit bevorzugte Testverfahren ist das der zwei einseitigen t-Tests. Tabellen und Nomogramme zur exakten Fallzahl findet man bei PHILLIPS (1990) und DILETTI et al. (1991) und (1992). Diese basieren auf der numerischen Integration zweidimensionaler t-Verteilungen, für die allerdings z. Zt. kein Programm angeboten oder publiziert wird. Damit dürfte für viele Anwender die exakte Fallzahlberechnung auf die tabellarisch abgedeckten Parameterkonstellationen beschränkt bleiben. Approximationen werden von einigen Autoren angegeben. Eine grobe Abschätzung erhält man über die Fallzahl für einen einseitigen t-Test mit  $\alpha$  und  $\beta/2$ .

In 1. und 2. reicht zur Behandlung der einseitigen Fragestellung das hinlänglich bekannte Verfahren der Datentranslation, das auf einen gewöhnlichen einseitigen t-Test und die zugehörige Berechnung des Stichprobenumfangs führt.

3. Vergleich zweier Binomialverteilungen, Effektmaß Odds Ratio: Der optimale Test verwendet die erweiterte hypergeometrische Verteilung. Fortran-Programme zur exakten Bestimmung der Fallzahlen für ein- und zweiseitige Fragestellungen findet man in WELLEK (1994), ebenso ein SAS-Makro zur Durchführung des zweiseitigen Tests.

4. Vergleich zweier Binomialverteilungen, Effektmaß Differenz der Risiken: Es existiert kein exaktes Verfahren. Unter den Approximationen liefern die einfach anzuwendenden Formeln von FARRINGTON/MANNING (1990) die besten Ergebnisse. Bei großen Stichprobenumfängen sind die Formeln von Blackwelder durchaus verwendbar. Beide betrachten nur einseitige Tests. Für die zweiseitige Fragestellung können analog zu 2. zwei einseitige Tests und für diese eine grobe Fallzahlabschätzung mit  $\alpha$  und  $\beta/2$  verwendet werden.

Tabelle 2: Literatur zur Fallzahlplanung für binomiale Beobachtungen

Autor(en)	Modell	Effektmaß	Verfahren	Power/Fallz.	Art der Lösung	Anwendbarkeit	Bemerkungen
Blackwelder (1982)	$2 \times 2$	RD	Test	Fallz	F, a	einf	unrestringierte Schätzer, einseit.
Bock (1997)	$2 \times 2$	RD	Test	Fallz	e, i	P	zu Hauck/Andersen (1986)
Durrleman/Simon (1990)	$2 \times 2$	RD	GS-Test	Fallz	F, a	aufw	einseitig
Farrington/Manning (1990)	$2 \times 2$	RD, RR	Test	Fallz	F, a	einf	restring. ML-Schätzer, einseitig
Garner (1994)	$2 \times 2$ $2 \times 2$	OR RD, RR	Test Test	Fallz Fallz	F, a F, a	einf aufw	zweiseitige Tests über die Devianz
Jennison/Turnbull (1993a)	2-St	RD	GS-Test	Fallz	F, a	aufw	einseitig
Jennison/Turnbull (1993b)	2-St	RD	GS-Test	Fallz	F, a	aufw	zweiseitig
Lakshminarayanan/Patel/Stager (1994)	$2 \times 2$	RD	GS-Test	Fallz	F, a	aufw	einseitig
Lu/Bean (1995)	$2 \times 2$ verb	RD	Test	Fallz	F, a	einf	einseitig, ähnlich McNemar-Test
Makuch/Simon (1978)	$2 \times 2$	RD	Test	Fallz	F, a	einf	eins., vgl. Blackw.
Nam (1994)	$2 \times 2$ stra	RR	Test	Fallz	F, a	einf	eins., ähnl. Farrington/Manning
Rodary/Com-Nougue/Tournade (1989) <sup>8)</sup>	$2 \times 2$	RD	Test	Fallz	F, a	einf	restring., bedingter Schätzer, einseitig
Wellek (1994) (siehe auch Text)	$2 \times 2$ $2 \times 2$ $2 \times 2$ verb	OR RD	Test Test Test	Fallz Fallz Power	A, e A, e A, e	P P P	auch eins., exakt konserv. über OR McNemar-Situation

Erläuterungen der Abkürzungen: wie zu Tabelle 1, zusätzlich

$2 \times 2$	4-Felder-Tafel (Homogenitätsproblem)
OR	Odds Ratio
RD	Risikodifferenz
RR	relatives Risiko
verb	verbundene Stichproben (paarige Bernoulli-Variable)
stra	stratifiziert

## Referenzen

## Kommentierte Literatur zur Berechnung der Fallzahlen oder der Trennschärfe

- ANDERSON S., HAUCK W. W. (1983): A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communications in Statistics – A: Theory and Methods* 12, 2663–2692.
- ANDERSON S., HAUCK W. W. (1990): Consideration of individual bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* 18, 259–273.
- BAUER P., BAUER M. M. (1994): Testing equivalence simultaneously for location and dispersion of two normally distributed populations. *Biometrical Journal* 36, 643–660.
- BLACKWELDER W. C. (1982): "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials* 3, 345–353.
- BOCK J. (1997): Die Bestimmung des Stichprobenumfangs für biologische Experimente und kontrollierte klinische Studien. G. Fischer Verlag, Stuttgart. Im Druck.
- BRISTOL D. R. (1993): Probabilities and sample sizes for the two one-sided tests procedure. *Communications in Statistics – A: Theory and Methods* 22, 1953–1961.
- DILETTI E., HAUSCHKE D., STEINJANS V. W. (1991): Sample size determination for bioequivalence assessment by means of confidence intervals. *International Journal of Clinical Pharmacology, Therapy and Toxicology* 29, 1–8.
- DILETTI E., HAUSCHKE D., STEINJANS V. W. (1992): Extended tables for the multiplicative model and bioequivalence ranges of 0.9 to 1.11 and 0.7 to 1.43. *International Journal of Clinical Pharmacology, Therapy and Toxicology* 30, 287–290.
- DURRLEMAN S., SIMON R. (1990): Planning and monitoring of equivalence studies. *Biometrics* 46, 329–336.
- ESINHART J. D., CHINCHILLI V. M. (1994): Sample size considerations for assessing individual bioequivalence based on the method of tolerance intervals. *International Journal of Clinical Pharmacology and Therapeutics* 32, 26–32.
- FARRINGTON C. P., MANNING G. (1990): Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine* 9, 1447–1454.
- FRICK H. (1987): On level and power of Anderson and Hauck's procedure for testing equivalence in comparative bioavailability. *Communications in Statistics – A: Theory and Methods* 16, 2771–2778.
- GARNER J. B. (1994): Deviance estimates of sample size for equivalence tests in vaccine trials. *Statistics in Medicine* 13, 1827–1839.
- GIANI G., FINNER H. (1991): Some general results on least favorable parameter configurations with special reference to equivalence testing and the range statistic. *Journal of Statistical Planning and Inference* 28, 33–47.
- GIANI G., STRASSBURGER K. (1994): Testing and selecting for equivalence with respect to a control. *Journal of The American Statistical Association* 89, 320–329.
- HAUCK W. W., ANDERSON S. (1984): A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *Journal of Pharmacokinetics and Biopharmaceutics* 12, 83–91.
- HAUSCHKE D., STEINJANS V. W., DILETTI E., BURKE M. (1992): Sample size determination for bioequivalence assessment using a multiplicative model. *Journal of Pharmacokinetics and Biopharmaceutics* 20, 557–561.
- HENRY E. P., MAILHOT D. W. (1991): Sample size calculation for bioequivalence studies. *American Statistical Association, Proceedings of Biopharmaceutical Section*, 169–174.
- JENNISON C., TURNBULL B. W. (1993a): One-sided sequential tests to establish equivalence between treatments with special reference to normal and binary responses. In: Hoppe F. M. (ed.): *Multiple comparisons, selection, and applications in biometry*. Marcel Dekker, New York, 315–330.
- JENNISON C., TURNBULL B. W. (1993b): Sequential equivalence testing and repeated confidence intervals, with applications to normal and binary responses. *Biometrics* 49, 31–43.
- KANFER F. H. J., GEERTSEMA J. C., STEYN H. S. (1988): A two-stage procedure for testing bioequivalence. *American Statistical Association, Proceedings of Biopharmaceutical Section*, 148–152.
- LAKSHMINARAYANAN M. Y., PATEL H. I., STAGER W. J. (1994): Multistage test procedure for testing Blackwelder's hypothesis of equivalence. *Journal of Biopharmaceutical Statistics* 4, 165–171.

- LIU J.-P., CHOW S.-C. (1992): Sample size determination for the two one-sided tests procedure in bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* 20, 101–104.
- LU Y., BEAN J. A. (1995): On the sample size for one-sided equivalence of sensitivities based upon McNemar's test. *Statistics in Medicine* 14, 1831–1839.
- MAKUCH R., SIMON R. (1978): Sample size requirements for evaluating a conservative therapy. *Cancer Treatment Reports* 62, 1037–1040.
- NAM J. M. (1994): Sample size requirements for stratified prospective studies with null hypothesis of non-unity relative risk using the score test. *Statistics in Medicine* 13, 79–86.
- PATEL H. I., GUPTA G. D. (1984): A problem of equivalence in clinical trials. *Biometrical Journal* 26, 471–474.
- PHILLIPS K. F. (1990): Power of the two one-sided tests procedure in bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* 18, 137–144.
- RACINE A., GRIEVE A. P., FLÜHLER H., SMITH A. F. M. (1986): Bayesian methods in practice: experiences in the pharmaceutical industry. *Applied Statistics* 35, 93–150.
- RACINE-POON A., GRIEVE A. P., FLÜHLER H., SMITH A. F. M. (1987): A two-stage procedure for bioequivalence studies. *Biometrics* 43, 847–856.
- RODARY C., COM-NOUGUE C., TOURNADE M. F. (1989): How to establish equivalence between treatments: A one-sided clinical trial in paediatric oncology. *Statistics in Medicine* 8, 593–598.
- SCHUIRMANN D. J. (1990): Design of bioavailability/bioequivalence studies. *Drug Information Journal* 24, 315–323.
- SRINIVASAN R., LANGENBERG P. (1986): A two-stage procedure with controlled error probabilities for testing bioequivalence. *Biometrical Journal* 28, 825–833.
- STEINJANS V. W., HAUCK W. W., DILETTI E., HAUSCHKE D., ANDERSON S. (1992): Effect of changing the bioequivalence range from (0.80, 1.20) to (0.80, 1.25) on the power and sample size. *International Journal of Clinical Pharmacology, Therapy and Toxicology* 30, 571–575.
- WELLEK S. (1991): Zur Formulierung und optimalen Lösung des Bioäquivalenznachweisproblems in der klassischen Theorie des Hypothesentestens. In: Vollmar (Hg.): *Bioäquivalenz sofort freisetzender Arzneiformen*, G. Fischer, Stuttgart, 91–109.
- WELLEK S. (1993): Basing the analysis of comparative bioavailability trials on an individual statistical definition of equivalence. *Biometrical Journal* 37, 47–55.
- WELLEK S. (1994): *Statistische Methoden zum Nachweis von Äquivalenz*. G. Fischer Verlag, Stuttgart.
- WESTLAKE W. (1988): Bioavailability and bioequivalence of pharmaceutical formulations. In: Peace K. E. (ed.): *Biopharmaceutical Statistics for Drug Development*, Marcel Dekker, New York, 329–352.
- WIE G. (1993): Simple methods of power calculation in bioequivalence study. *American Statistical Association, Proceedings of the Biopharmaceutical Section*, 273–278.
- YEH C. M., CHEN M. (1993): Power and sample size for high-order crossover. *American Statistical Association, Proceedings of the Biopharmaceutical Section*, 279–284.

### Sonstige zitierte Literatur

- CHOW S.-C., LIU J.-P. (1992): *Design and analysis of bioavailability and bioequivalence studies*. Marcel Dekker, New York.
- DONNER A. (1984): Approaches to Sample Size Estimation in the Design of Clinical Trials – A Review. *Statistics in Medicine* 3, 199–214.
- FRICK H. (1994): On approximate and exact sample sizes of equivalence tests for binomial proportions. *Biometrical Journal* 36, 841–854.
- HAUCK W. W., ANDERSON S. (1986): A comparison of large-sample confidence interval methods for the difference of two binomial probabilities. *The American Statistician* 40, 318–322.
- LEHMANN E. L. (1986): *Testing statistical hypotheses*. John Wiley & Sons, New York.
- NAM J. M. (1995): Sample size determination in stratified trials to establish the equivalence of two treatments. *Statistics in Medicine* 14, 2037–2049.
- ORTSEIFEN C., BRUCKNER T., BURKE M., KIESER M. (1997): An overview of software tools for sample size determination. *Informatik, Biometrie u. Epidemiologie in Medizin und Biologie* 28, 91–118.



- ROEBRUCK P., KÜHN A. (1995): Comparison of tests and sample size formula for proving therapeutic equivalence based on the difference of binomial probabilities. *Statistics in Medicine* 14, 1583–1594.
- SCHUIRMANN D. J. (1987): A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 15, 657–680.
- WESTLAKE W. J. (1972): Use of confidence intervals in analysis of comparative bioavailability trials. *Journal of Pharmaceutical Sciences* 61, 1340–1341.

### Weitere Literatur

- ASHTON J. (1990): Hypotheses and sample size for showing equivalence. *SAS.SUGI* 15, 1387–1391.
- BEAL S. L. (1989): Sample size determination for confidence intervals on the population mean and on the difference between two population means. *Biometrics* 45, 969–977.
- BLACKWELDER W. C., CHANG M. A. (1984): Sample size graphs for “proving the null hypothesis”. *Controlled Clinical Trials* 5, 97–105.
- FALK R. W. (1994) : Letter to the Editor. How to establish equivalence between treatments: a one-sided clinical trial in pediatric oncology. *Statistics in Medicine* 13, 2279–2280.
- FLÜHLER H., GRIEVE A. P., MANDALLAZ M., MAU J., MOSER H. A. (1983): Bayesian approach to bioequivalence assessment: an example. *Journal of Pharmaceutical Sciences* 72, 1178–1181.
- GRIEVE A. P. (1991): Two-stage procedure for bioequivalence testing. In: Vollmar J. (Hrsg.): *Biometrie in der chemisch-pharmazeutischen Industrie*, Band 5, Gustav Fischer Verlag, Stuttgart, 45–50.
- METZLER C. M. (1991): Sample sizes for bioequivalence studies. *Statistics in Medicine* 10, 961–970.
- MÜLLER-COHRNS J. (1990): The Power of the Anderson-Hauck test and the double t-test. *Biometrical Journal* 32, 259–266.
- WIJNAND H. P. (1994): Updates of bioequivalence programs (including statistical power approximated by Student's t). *Computer Methods and Programs in Biomedicine* 47, 275–281.

Korrespondenzadresse: Prof. Dr. P. Roebuck, Behringstr. 15, 68789 St. Leon-Rot

## Sample Size Determination on Survival Time Data – A Review

Susanne Oellrich, Frank Freischläger, Axel Benner, Meinhard Kieser

### Abstract

*A review is given on sample size determination methods for planning survival time trials by searching the literature of the last three decades. 105 articles (and books) were screened and classified according to characteristic criteria suitable for survival trial designs. The review covers methods for one sample, two-group or multiple-group trials as well as methods devoted to special designs or applicable given special assumptions on the distribution of the survival time. The historical evolution of survival trial designs is presented and a comprehensive and commented bibliography is included.*

### 1. Introduction

In many clinical trials occurrence of events during the treatment and under follow-up observation are of primary interest. These events may be treatment success, but as well critical events such as myocardial infarction, cancer diagnosis or death. For convenience we consider in the following death as the event of concern.

Already Peto et al. (1976) pointed out, that not only whether or not an event occurred to the patient should be analysed but, as well and even better one should study the time to such an event. With death as the interesting event, this means that the evaluation of a clinical trial should not be restricted to the estimation and statistical testing of death rates and survival rates but has to analyze the survival times directly. Avoiding the loss of information when using rates for the comparison of treatment groups should be worth the benefit to analyze survival times and the consequently higher power of the comparison procedure.

However, survival data are not suitable for the application of standard statistical procedures (as e.g.  $t$ -tests) since they are usually not symmetrically distributed and, which has to be taken very serious, they are often censored, which means that the endpoint could not be observed completely in each patient or investigational subject/object. Censoring happens preferably in clinical trials when the analysis is scheduled for some predefined timepoint and when at that time some of the patients are still alive or have been lost to follow-up at that time or even earlier during the trial. Another reason leading to censored data is given when cause of death is not related to the disease.

Traditional treatment comparison of survival curves  $S_1(t)$  and  $S_2(t)$  is achieved via hypothesis testing. The parameters of the test problem  $H_0 : S_1(t) = S_2(t)$  against the alternative  $H_1 : S_1(t) \neq S_2(t)$  may be ratios of means or medians of survival times or ratios of

hazard rates or even more complicated functions of the survival time (BRESLOW 1975). Often proportional hazards (PH) have been assumed (see COX 1972), especially in the planning stage, where efficient techniques of sample size determination are necessary to guarantee statistical power and where the PH assumption proved to be successful.

For analysis purposes, test statistics as the log rank test, Gehan's generalised Wilcoxon test among many others arised as appropriate test procedures (see e.g. TARONE and WARE, 1977, FLEMING et al., 1984, BELTANGADY and FRANKOWSKI, 1989). Because of comparable asymptotic properties the choice of the test statistic is actually less important for sample size considerations than the type of design and the censoring mechanism (COLLETT 1994).

The conduct of a survival time study is commonly divided into phases. First there is an entry or accrual period, during which patients enter the study. After the end of this accrual a continuation or follow-up period may follow. No further patients enter the study during this period. Patients can be 'administratively censored' at the end of the follow-up period (if no event occurred until then) or they are censored by 'loss to follow-up' before the end of the study.

The target design parameter (sample size, power, or trial duration) given the other design parameters are kept fixed depends on a series of assumptions (see below), which make sample size calculations for survival trials rather complicated. In this context, distributional assumptions on survival time, on accrual time, and on censoring determined partly by the follow-up time have to be considered carefully for all planning purposes.

Because of the bi-phasic nature of the timely conduct of a trial, the computation of sample sizes for survival trials consists of two parts. At first the number of deaths required to detect a certain difference between two treatment groups is determined and, second, one has to account for the fact that the patients usually do not enter the study at one time point and that not all patients will be followed to death (see below).

Many methods can now be found in the literature which deal with sample size determination for survival time data. Even in the classical situation of a two-group single-stage treatment comparison there is a larger number of published papers. But there is also a considerable number of articles which deal with special situations such as the effects of non-compliance and of cure, assumed special distributions and non-proportional hazards. Others deal with stratification by prognostic factors or covariates, multiple-stage designs, or they apply multiple comparison procedures.

The outline of this review is organized as follows:

In Section 2 the methods of review itself are described. In Section 3.1 we provide in chronological order the most important papers on planning survival time trials. The main results are given in Section 3.2 where all relevant papers for the sample size determination in a hypothesis testing set up are listed and classified according to the criteria chosen and defined earlier. In Section 3.3 we discuss papers dealing with special designs as for example factorial or sequential designs. Review papers of the comparison of sample size procedures are discussed in Section 3.4. After the discussion in Section 4 we provide finally a comprehensive and as complete as possible list of commented references which includes also some sample size and survival analysis textbooks and a few other relevant general papers on survival analysis.

## 2. Methods of Review

All papers were reviewed that we became aware of by cross-reference, on-line investigation via CIS and MEDLINE or just by our knowledge of the field.

Each article of the list of references was read, assessed for its relevance for sample size determination in survival time data and, if approved, it was classified into one of eight

Table 1: Definition of categories for classification

Category	Section of paper	Description of Content
A.1	3.1	Two treatment arms: sample size or power as target
A.2	3.1	Two treatment arms: trial duration as target
A.3	3.3	$k$ -treatment ( $k > 2$ ) design, two-stage, factorial, Cox model (incl. covariates)
A.4	3.4	Comparison or review of sample size and power
A.5	3.4	Textbooks on survival analysis, which include sample size estimation methods
A.6	3.3	Special aspects: sample size adjustment, historical controls, cure rate models, screening trials, paired data etc.
B.1	3.4	Sample size textbooks (general) including methods for survival studies
B.2	3.4	Reviews on sample size and power (general)

categories (see Table 1). The *A* categories comprise all references which directly deal with one or more sample designs for survival time studies. The *B* categories consist of more general references on sample size methodology, where survival may be one topic among others.

Three types of design for survival time studies are considered. They are visualised and compared with the common fixed-time design (type A) in Figure 1. For each of the four design types uniformly accrued patients are assumed with the calendar times of their entry into the trial and their times on study. Patients under observation until study termination are administratively right-censored. Patients not under observation until study termination are dead or lost to follow-up. Observed and censored survival times are displayed.

In the design of type A each patient is observed for the same time period. In the design of type B each patient is observed until event, that is, the study ends, when the last patient dies or is lost to follow-up. In the design of type C all patients are observed until the last patient is accrued, that is no follow-up period of continued observation is incorporated. In the design of type D, which is today's common practice, the duration of a survival time trial consists of two distinct periods. First there is an entry or accrual period. After the end of accrual a continuation or follow-up period of a fixed length follows. No further patients enter the study during this period.

In common situations of sample size determination (assuming e.g. normal distribution for the variable of primary interest), type I error  $\alpha$ , type II error  $\beta$ , a clinically relevant treatment difference and an estimation of variance fully determine the sample size. For survival time studies mostly the clinically relevant treatment difference is expressed as the ratio of two hazard rates and the total sample size is calculated depending on the ratio of the number of deaths to be observed and the expected proportion of deaths (see Figure 2).

Note that the sample size as the target design parameter may be replaced by the statistical power of the analysis procedure given sample size. This fact can be used in the planning stage to determine the sample size indirectly or as a post-hoc calculation with sample sizes and number of deaths given. The statistical power only depends on the number of deaths rather than the number of patients in total.

As well trial duration may serve as target design parameter. The combination of length of accrual and length of follow-up is now to be determined in order to guarantee sufficient statistical power.

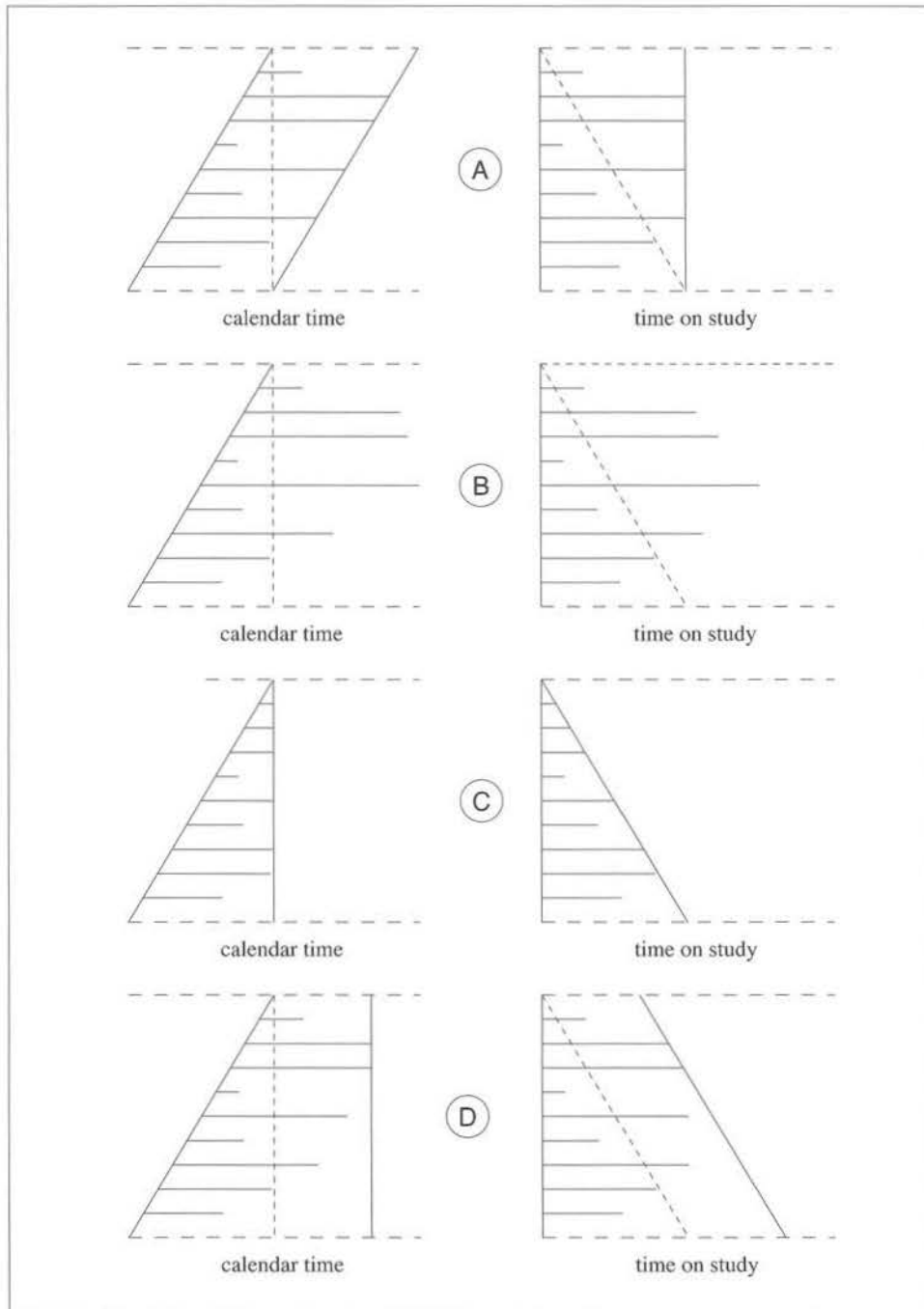


Figure 1: Exemplary patient paths in design types A (fixed-time duration), B (maximum duration), C (accrual only), and D (accrual and follow-up) for trials on survival time data with calendar times (left) and times on study displayed (right) assuming uniform accrual (dashed line = end of accrual, solid lines = begin of study and study termination (administrative censoring))

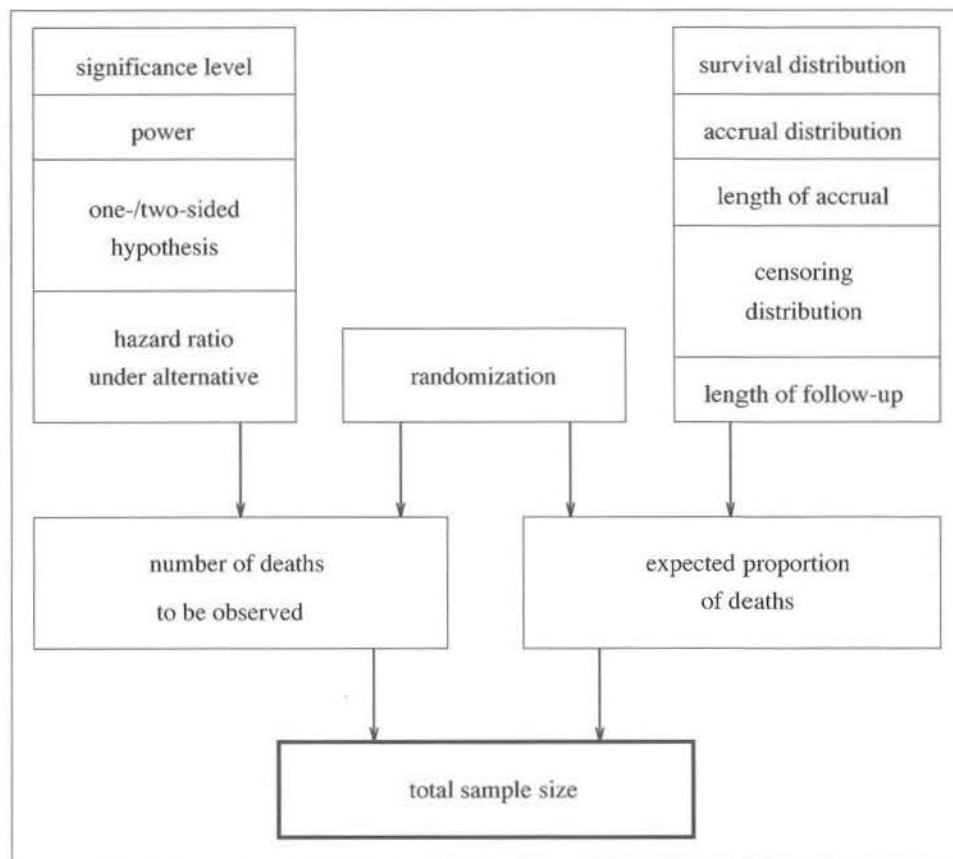


Figure 2: Design parameters implied in survival trials

To simplify the design and to achieve highest power it is optimal to use 1:1 randomisation of patients in clinical trials. In trials with a survival time endpoint the number of events is of main interest which is usually not equal to the number of patients. FREEDMAN (1982) shows that the power of the log rank test is independent of the sample size given that the number of events is kept fixed. Therefore more power may be achieved for equal allocation of the number of events than for the equal allocation of sample sizes. HSIEH (1992) recommends that in unbalanced designs one should try to choose patient allocation in a way to balance the number of events between the two treatment arms and to create an equal event design which can be achieved using a sample size ratio  $k$  with  $k = (1 - p_2)/(1 - p_1)$  where  $p_i$  describe the survival rate in group  $i = 1, 2$  in a fixed sample size design.

Several characteristics of the references were recorded and kept in a SAS database. Those were: target design parameter, technical form, accuracy of the method, type of design, assumption on survival distribution, assumption on accrual distribution, assumption on censoring distribution, number of treatment groups, restriction to equal proportions of patients per treatment group, kind of test statistic, one-sided versus two-sided testing, number of stages of analysis, incorporation of prognostic factors as covariates, availability of software, and use of simulation techniques.

Table 2: Frequencies of relevant papers by categories of Table 1

Category	A.1	A.2	A.3	A.4	A.5	A.6	B.1	B.2	Total
Frequency	19	8	23	13	4	25	7	6	105

### 3. Review of the Literature on Sample Size Calculation

105 articles (and books), published between 1967 and 1996, were decided to be relevant for sample size determination on survival time data. Frequencies according the categories of Table 1 are given in Table 2.

Papers treating equivalence testing, Bayesian methods or epidemiological topics are not included. We excluded also the paper by GAIL (1985) on sample size calculations based on the logistic regression model because of its restrictive assumptions.

#### 3.1 Historical Development

In the following we review the most important references as they appeared chronologically in statistical literature since the early seventies. References to commercial software are included in brackets if these were available for review (cp. ORTSEIFEN et al., 1997).

##### 3.1.1 Exponential distribution assumption

Early papers on planning the duration and the sample size of survival time trials were based on the assumption of exponential survival distributions and they considered at first only two equally sized treatment groups.

PASTERNAK and GILBERT (1971) present an algorithm and tables for estimating the minimum trial duration when a statistically significant increase in median survival time in one group over the other has to be detected at a predetermined degree of confidence. For the determination of the trial duration they first derive the number of failures required. Then, in a second step, they convert this estimate into 'person-years' of follow-up and finally they project it to the required study length. Since they aim to minimise the trial duration, they work under design type C (full accrual until administrative censoring). Patient accrual is modelled by uniformly distributed times of patient entry into the trial.

GEORGE and DESU (1974) expand the ideas of PASTERNAK and GILBERT allowing patient entry to follow a Poisson process. They determine the sample size and the duration of a trial allowing for a follow-up period. They show that in this special situation an optimal length of trial is obtained when there is no follow-up. The sample size is derived from the exact distribution of the likelihood ratio test statistic and it is compared with two normal approximations resulting in slightly smaller sample sizes than those derived by PASTERNAK and GILBERT [Software: PLANUNG, STPLAN].

BERNSTEIN and LAGAKOS (1978) present a FORTRAN subroutine for the sample size or power calculation of a survival time study. Their method is based on that of GEORGE and DESU, allowing for a post-accrual follow-up period, censored observations, stratification into risk groups, and an unequal allocation scheme (design type D: accrual and follow-up with administrative censoring) [Software: PLANUNG].

In his review of sample size methods LACHIN (1981) uses the local power function of the Wald test for equality of exponentially distributed survivor functions. Based on earlier work of GROSS and CLARK (1975) calculations for the case of uniform patient entry followed by a follow-up period of fixed length are presented.

LUSTBADER and LITWIN (1981) point out the power advantages of a design incorporating a follow-up period. They present a sample size formula based on a normal approximation of the time on study, with the aim of minimising total trial duration.

RUBINSTEIN et al. (1981) expand the methods of PASTERNAK and GILBERT and of GEORGE and DESU to the design type D (accrual and follow-up with administrative censoring) for the case of exponential survival, patient entry according to a Poisson process, and exponentially distributed losses to follow-up [Software: NQUERY]. They determine the duration of accrual for a fixed duration of follow-up using computer simulations that show that under certain restrictions the log rank test yields the accurate power. It is also demonstrated that the desired power is approximately reached when the survival distribution is Weibull rather than exponential.

SCHOENFELD and RICHTER (1982) give nomograms for the sample size under the assumptions of design type D with exponential survival, uniform accrual and uniform censoring distribution [Software: EX-SAMPLE, POWER].

MORGAN (1985) concentrates on the trial duration providing methods of determining appropriate combinations of the accrual and follow-up time span under the assumption of a constant accrual rate. Depending on the ratio of the cost of accruing patients to the cost of patient's follow-up he presents the unique cost effective choice of accrual and follow-up period of the many possible choices providing the same expected number of deaths and therefore the same level of significance and power.

LACHIN and FOULKES (1986) expand the equation relating sample size and power of the test for equal hazard rates presented by LACHIN (1981). They allow for nonuniform patient entry, using e.g. a truncated exponential distribution, and exponentially distributed losses of follow-up. They also derive a formula for the adjustment of the sample size if non-compliance occurs or if a stratified analysis is performed [Software: PASS, SOLO-POWER, NSURV].

The criterion of GROSS et al. (1987) is the expected cost of a clinical trial. Their aim is to find the optimal duration of a trial with uniform accrual and exponentially distributed losses to follow-up on the basis of minimising the expected costs. In their model the investigator has to specify a linear cost function as input for the minimisation procedure for the accrual and follow-up period length.

Direct use of simulation techniques to determine sample size is demonstrated by YATEMAN and SKENE (1993).

### 3.1.2 Proportional hazards assumption

SCHOENFELD (1981) provides the theoretical basis for modelling complex studies. He examines the asymptotic properties of TARONE and WARE's (1977) statistics under Lehmann alternatives with an application for the log rank test. That design with proportional hazards survival allows unequally sized treatment groups [Software: SSIZE]. Note, Lehmann alternatives are of the form  $F_2 = F_1^k$ ,  $k > 1$ .

A number of sample size tables for the log rank test are presented by FREEDMAN (1982). Under the assumption of a constant hazard ratio throughout the study, equal censoring in both groups, and small differences in the ratio of patients at risk, he rederives Lachin's method based on the asymptotic expectation and variance of the log rank statistic. Monte Carlo simulations verify the resulting formula. The sample size calculation is based on the survival probabilities at a fixed time after the last patient entered the study. He makes the crucial assumption that all patients are followed the same length of time and that the number of patients required is independent of the accrual rates. Therefore this approach permits an arbitrary accrual distribution. When the accrual period is equal to the follow-up period or longer Freedman recommends to use the survival probabilities at the median length of follow-up. To account for loss to follow-up a simple correction formula is given [Software: NQUERY, SSIZE].



STEIN (1986) comments the method of Freedman and points out that the number of events required is only slightly larger than that of George and Desu's method, but that the number of patients required can be noticeably larger.

MORGAN (1987) develops nonparametric maximum likelihood estimates of duration of accrual and of the total trial duration (accrual and follow-up) under the proportional hazards model and a constant accrual rate. Estimates of the total trial duration and its asymptotic distribution are derived as functions of the accrual duration.

YATEMAN and SKENE (1992) approximate the survival distribution and loss of follow-up distributions by piecewise exponentials, and they approximate patient entry by a piecewise linear distribution, still assuming constant ratios of hazard functions over time. In that way, they obtain a flexible approach at reduced computational costs.

### 3.1.3 Nonproportional hazards / no survival distribution specified

WU et al. (1980) use stepwise exponential functions for modelling event and drop-out rates via a discrete approximation of a waiting-time process. Therefore, their sample size method allows for time-dependent hazard rates.

SCHUMACHER (1981) introduces a distribution-free procedure based on the McNemar test for power calculation in the presence of censoring (design type D), using pairwise preferences with every two consecutive patients forming a pair. He compares his approach with the parametric exponential model of BERNSTEIN and LAGAKOS (1978).

LAKATOS (1986) introduces a Markov chain model to determine the sample size for design type A (usual fixed-time endpoint design) under time-dependent rates of losses and non-compliance. The method can be adjusted to accommodate to a wide range of assumptions, either parametric, or nonparametric. Lag in the effectiveness of treatment is also considered.

The emphasis of the paper of LAKATOS (1988) is to make sample size calculations for the log rank test more sensitive for complicated situations. Under design type D treatment and control group are modelled separately by a nonstationary Markov process to allow for time-dependent rates of losses of follow-up, non-compliance, staggered entry, and stratification. Some hints concerning the planning of the trial duration are also given.

SHIH (1995) describes a SAS macro called SIZE which uses the method of LAKATOS (1988), allowing additionally for unequal allocation between the two experimental groups, heterogeneity in the treatment effects, and computation of the trial duration based on the sample size and power.

HALPERN and BROWN (1987b) present a computer program that allows an interactive determination of the length of accrual and follow-up period. It is designed for use with either the generalised Wilcoxon test or the log rank test. After specification of the assumed survival curves, estimates of the power are given using Monte Carlo simulations.

Concerning unbalanced designs HSIEH (1987) proposes a simple method which multiplies the total number of events estimated from a balanced design by a factor to obtain the total number of events required for an unbalanced design.

In the paper of GAIL (1994) factors influencing the sample size are described. General formulae for sample size calculations are given based on the log rank statistic and for point-wise comparisons of two Kaplan-Meier curves. The impact of various factors, as for example factors that dilute the treatment effect, factors that alter the degree of censoring, and the choice of the test statistic, is illustrated by numerical examples.

The references described above and some more specialised papers mentioned below are displayed in Figure 3. Those are milestones in the development of appropriate methods for sample size determination in survival time trials.

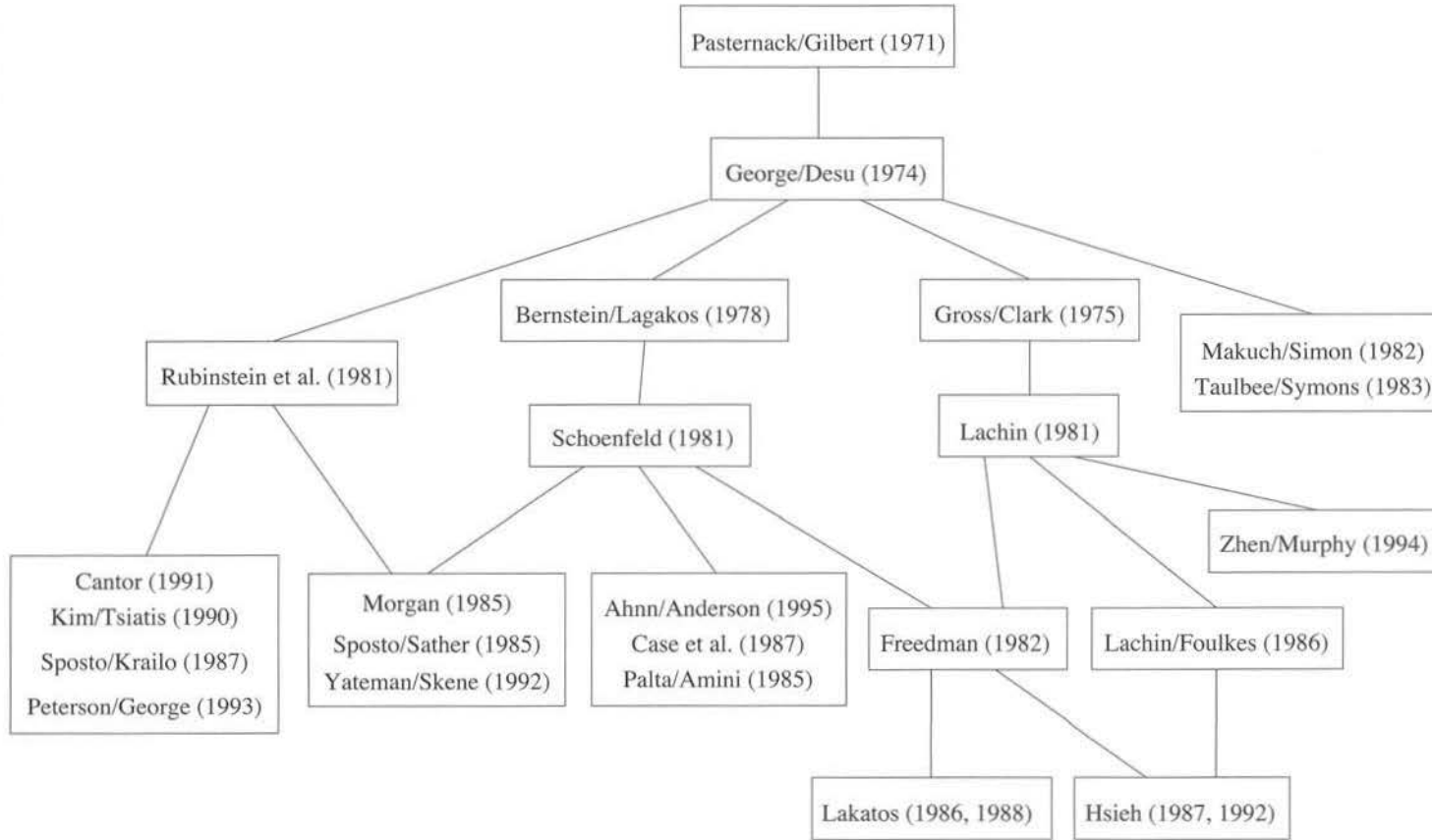


Figure 3: Chronological dependency of methods for sample size determination on survival time data

We should mention that there had been earlier papers which cleared up the problems as e.g. with respect to loss of follow-up or non-compliance e.g. SCHORK and REMINGTON (1967) and HALPERIN et al. (1968). These papers have been fundamental for the development of the methods of LACHIN (1981), WU et al. (1980), and LAKATOS (1986, 1988). By the years of publication it is seen that the development of sample size methods specialised on survival time data became intensive in the 1980's, but that it is still an ongoing field of research and publication.

### 3.2 Characteristics of References

Table 3 refers to articles of categories A.1 and A.2 on the comparison of two treatment arms with sample size and power and/or duration of accrual period and follow-up as targets. Some milestone papers from the other categories are also included. For the majority of the references short reviews and comments on special aspects have been given already in Section 3.1.

Table 3 is organized into category, author(s), year of publication, main target parameter, technical form of presentation, type of design, distributional assumptions on survival, accrual, and follow-up/censoring, balance between groups (equal/unequal), availability of software directly from the authors.

Hypotheses of underlying tests are mostly expressed by equality of hazard rates, mean or median survival times, which are interchangeable under the assumption of proportional hazards. Formulae for two-sided hypotheses can easily be modified to one-sided hypotheses and vice versa. Nearly all methods are approximations to exact solutions due to the use of asymptotic distributions.

The following comments are in order to explain more of the distributional assumptions:

**Accrual distribution:** If it is assumed that patients enter the trial according to a Poisson process with a uniform rate  $a$  per time interval, then the total number  $n = n_c + n_e$  entered on the trial after  $T$  intervals will be distributed as a Poisson variable with mean  $aT$ . Given that  $n$  patients entered according to a Poisson process, the entry times will be iid uniformly distributed over the interval  $[0, T]$ . So, if the sample size is fixed we can assume uniform accrual, but when we assume that the sample size is random we assume that the accrual follows a Poisson process with constant rate. In Table 3 a uniform accrual is taken as a synonym for both situations.

**Follow-up distribution:** This describes the distribution of time to loss to follow-up of a patient.

**Censoring distribution:** An important assumption is that the survival times are independent of any censoring mechanism. Usually it is assumed that censoring is only due to random entry times and random time to the end of the follow-up. If it is assumed that accrual during  $[0, A]$  follows a Poisson process with constant rate and if censoring results only from being alive at the end of the study, then the censoring times are uniformly distributed over  $[T-A, T]$ .

SCHOENFELD (1981) showed that the log rank test is not fully efficient for exponential survival distributions if the censoring mechanism is different in the control and the experimental group.

Another source for loss of efficiency is noncompliance. One has to distinguish between drop-outs, which describe patients in the experimental group that will be noncompliant and will therefore have the same hazard as the control group, and drop-ins, which describe patients in the control group that will cross over to the experimental treatment and will have the same hazard as the experimental group.

Table 3: Characteristics of relevant reviewed papers for sample size determination in single-stage, two-group or multiple-group designs on survival time data

Year	Author(s)	Target parameter	Techn. form	De-sign*	Survival distribution	Accrual distribution	Follow-up distribution	Bal.**	Software
1967	Schork/Remington	sample size	formula	A	none	none	none	yes	no
1968	Halperin et al.	sample size	formula	A	exponential	none	none	yes	no
1971	Pasternack/Gilbert	duration	algorithm	C	exponential	uniform	exponential	yes	no
1972	Pasternack	sample size	formula	C	exponential	uniform	exponential	yes	no
1973	Nam	sample size	formula	A	normal (log)	none	none	no	no
1974	George/Desu	s.s. + dur.	formula	C, D	exponential	uniform	exponential	yes	no
1978	Bernstein/Lagakos	s.s. + pow.	formula	D	exponential	uniform	none	no	yes
1980	Wu et al.	sample size	algorithm	A	piecew. exp.	none	piecew. exp.	no	yes
1981	Lachin	sample size	formula	A-D	exponential	uniform	none	no	no
1981	Lustbader/Litwin	duration	formula	D	exponential	uniform	none	yes	no
1981	Rubinstein et al.	duration	algorithm	D	exponential	uniform	exponential	yes	no
1981	Schoenfeld	sample size	formula	A, C	prop. haz.	none	none	no	no
1981	Schumacher	power	formula	D	none	any	none	yes	no
1982	Freedman	sample size	table	A, D	prop. haz.	any	none	no	no
1982	Schoenfeld/Richter	sample size	nomogram	D	exponential	uniform	uniform	yes	no
1985	Morgan	duration	formula	D	exponential	uniform	uniform	yes	no
1986	Lachin/Foulkes	sample size	formula	D	exponential	any	any	no	yes
1986	Lakatos	sample size	algorithm	A	none	any	none	yes	yes
1987	Gross et al.	duration	formula	D	exponential	uniform	exponential	yes	yes
1987	Halpern/Brown (b)	dur. + pow.	algorithm	D	any	uniform	none	yes	yes
1987	Hsieh	sample size	formula	A-D	any	any	any	no	no
1987	Morgan	duration	formula	D	prop. haz.	uniform	none	no	yes
1988	Lakatos	sample size	algorithm	D	nonprop. hz.	any	any	yes	yes
1991	Frick	sample size	formula	B-D	exponential	uniform	exponential	no	no
1992	Lui	sample size	formula	A-D	exponential	uniform	none	yes	yes
1992	Yateman	sample size	formula	D	piecew. exp.	piecew. lin.	piecew. exp.	no	yes
1992	Yateman/Skene	sample size	formula	D	piecew. exp.	piecew. lin.	piecew. exp.	no	yes
1993	Yateman/Skene	power	table	D	exponential	uniform	exponential	yes	no
1994	Gail	sample size	formula	D	any	any	any	yes	no
1994	Zhen/Murphy	s.s. + pow.	formula	A-D	exponential	uniform	none	yes	no
1995	Brown et al.	sample size	algorithm	C	exponential	uniform	none	yes	yes

\* For type of design see Figure 1

\*\* Bal. = Balance (Restriction to equally sized treatment groups): yes/no

### 3.3 Work regarding Other Designs and Special Aspects of the Design of Survival Studies

#### 3.3.1 Other Designs

##### *k* Treatment Comparisons

MAKUCH and SIMON (1982) generalise the approach of GEORGE and DESU to the *k*-sample problem. They compute the required number of failures for a given power and overall difference between mean failure times of *k* treatment groups in the form of an ANOVA type test. For multiple comparisons Fisher's LSD approach is used. Exponential distribution and no censoring are assumed.

AHNN and ANDERSON (1995) provide a direct generalisation of this taking censoring into account by using an extension of SCHOENFELD's (1981) method. It is shown by simulations that at least for  $k = 3$  this approach gives acceptable accuracy. Also stratification has been investigated for the *k*-sample problem according to the method of PALTA and AMINI (1985).

CHEN and SIMON (1994a, 1994b) present an extension of one-sided tests to the situation of multiple comparisons for normally distributed data. An extension to exponentially distributed data and censoring results in large sample approximation according to RUBINSTEIN et al. (1981).

LIU and DAHLBERG (1995) also follow the basic approach of MAKUCH and SIMON, but they develop their sample size formulae by direct use of the log rank test and the proportional hazards model. Via simulation a close agreement of their method with that of MAKUCH and SIMON for exponentially distributed data is shown and a SAS program for the situation of four treatment groups is given.

##### Methods for Factorial Designs

BERNSTEIN and LAGAKOS (1978) provide a FORTRAN subroutine for the sample size determination in stratified trials comparing two treatments. Uniform accrual, a fixed follow-up period and exponentially distributed survival times are assumed.

PETERSON and GEORGE (1993) provide a method when testing for interactions, which is a direct extension of that given by MAKUCH and SIMON (1982) and GEORGE and DESU (1974). For the calculation of the required duration of the study they use the results of RUBINSTEIN et al.. A simulation study shows that their method has valid power for the score test of an interaction effect if the proportional hazards assumption is appropriate.

XIANG et al. (1994) describe the theoretical relationship between the power of the test and the sample size for testing interaction in a  $2 \times k$  situation assuming a proportional hazards model. The usefulness of their method depends on the method used to estimate the logarithms of the relative failure rates and their variances. A simulation study compares the Mantel-Haenszel (MH), observed/expected, and maximum likelihood estimation (MLE) methods, and shows that both the MH and the MLE behave "reasonably" well.

NATARAJAN et al. (1996) present power calculations for a factorial clinical trial with survival time outcomes using the MAKUCH and SIMON (1982) test statistic. By Monte Carlo simulation they obtain type I and type II error probabilities. Fisher's LSD method is used to adjust for multiple comparisons. A computer program for interactive use is also described.

##### Covariates

TAULBEE and SYMONS (1983) use the method of GEORGE and DESU (1974), extend it to account for censoring, and generalise it to the situation when a single concomitant variable is present. Grouping by dose levels is an explicit example.

SCHOENFELD (1983) shows that the formula for the sample size for the comparison of two groups under the exponential distribution assumption is also valid when the Cox proportional hazards regression model is used to adjust for covariates. The resulting formula for the sample size is the same as that used when two homogeneous patient groups are compared (SCHOENFELD, 1981), but the power changes. The use of covariates can substantially increase the power of the test when the proportional hazards assumption holds.

Another approach to account for stratification and covariates is described by PALTA and AMINI (1985) who extend the method of SCHOENFELD to the case of  $m > 1$  strata or to the inclusion of a single covariate.

The impact of omitting a covariate is analysed by MORGAN and ELASHOFF (1986). The authors demonstrate that not adjusting for covariates results in decreased precision the more observations are censored.

LUI (1992) assumes an exponential model with a binary covariate in the situation of two treatment comparison and he comes up with asymptotic results.

ZHEN and MURPHY (1994) present sample size formulae for detecting the effect of an unrestricted covariate on survival time.

### Sequential Designs

NAGELKERKE et al. (1981) present formulae for the distribution of the sample size and approximate the expected number of failures in a sequential survival trial. They assume proportional hazards assumption, 1:1 randomization and accrual according to a Poisson process. Their sequential schemes are Wald's sequential probability ratio test (SPRT) and Anderson's modified SPRT.

A two-stage group sequential design to minimise the expected sample size (ESS) is given by CASE et al. (1987). Based on SCHOENFELD's formula ESS is determined by iteration.

LAN and LACHIN (1990) investigate a group sequential log rank test procedure in a maximum duration trial (Type B).

KIM and TSIATIS (1990) also propose a design procedure for a maximum duration trial based on the work of RUBINSTEIN et al. (1981).

Under the assumption of exponential survival time and uniform patient entry during accrual KIM (1992) proposes a design for a maximum information trial based on the "use function approach" for determining the study duration or for calculating the power in a group sequential trial with censored survival data. Unequal patient allocation and adjustment for stratified randomisation is allowed. In a second paper KIM (1995) provides two FORTRAN programs for planning sequential survival trials. The program SEQPWR computes the attainable power of group sequential log rank tests in a maximum duration trial (KIM and TSIATIS, 1990) whereas the program SEQOPR computes the operating characteristics of group sequential log rank tests in a maximum information trial. In both programs the user can choose between five different type I error spending functions.

HALPERN and BROWN (1993) extend their program in HALPERN and BROWN (1987b) to a sequential approach. A simulation program, written in C, allows for group sequential testing.

### Selection Designs

SCHAID et al. (1990) propose a two-stage screening design for survival data which enables the search of new treatments better than a standard and permits early termination of those treatments which do not show a minimum prespecified improvement. Their approach assumes that the patient population of interest remains the same for the two

stages of the study. The design minimizes the expected number of patients under the null hypothesis of no effect of the new treatments.

LIU et al. (1993) present a selection design using Cox's proportional hazards model in order to identify in a screening study a best treatment for later application in a comparative study. They provide an iterative solution for the sample size needed to achieve a specified correct selection probability. Their design may be used when the patient population in the comparative study differs from that used in the screening study.

### 3.3.2 Special Aspects

#### Study Continuation (Conditional Power Calculation)

ANDERSEN (1987) investigates the problem of re-examination of the original sample size calculations in an ongoing study. Conditional power calculations are used to support the decision whether a clinical trial comparing two exponential survival distributions should be extended beyond its planned termination or not.

GEBSKI et al. (1987) examine by use of Monte Carlo simulations whether a trial been in progress for some time may be stopped, based on a re-examination of the distributional assumptions.

CANTOR (1991) describes a computer program that estimates through simulations the power of linear rank tests for censored data, which is an extension of the program described in HALPERN and BROWN (1987a). It can also be used to compute the conditional power, given an existing set of survival data.

HENDERSON et al. (1991) describe a simulation program, which is again a modification of that by HALPERN and BROWN (1987b), to calculate the conditional power of the log rank and generalized Wilcoxon test, given an existing set of survival data. A modified Kolmogorov-Smirnov test (FLEMING et al., 1980) is also considered.

#### Historical Controls

DIXON and SIMON (1988) provide formulae and tables to determine the required sample size for the comparison of exponentially distributed survival times, when a control group is already available and probably under continuing follow-up. They use results from RUBINSTEIN et al. (1981) and ANDERSEN (1987).

Independently, EMRICH (1989) obtains similar results using the same assumptions following also the results of RUBINSTEIN et al. (1981).

CANTOR (1996) performs power calculations if historical data are available for the control. He follows the work of SCHOENFELD (1983) and uses the Kaplan-Meier estimate of the historical survival distribution.

#### Subset Randomization

SCHOENFELD and GELBER (1979) show how the sample size must be increased and how the randomization should be weighted to obtain the best power if institutions randomize patients into different subsets of the treatment arms of a clinical trial.

#### Sample Size Adjustment

TAVARÉ et al. (1995) consider situations where a classification variable is required. They present simple sample size and parameter estimate adjustment formulae for a misspecified binary valued explanatory variable using the estimate of the kappa statistic.

#### Paired Data

WILTSE et al. (1980) compare the power of the paired-t, the sign and the signed rank tests for location alternatives when the uncensored observations are correlated and have marginal exponential distributions. They also investigate the empirical power of the tests.

**Cure Rate Models**

GOLDMAN (1984) discusses exponential models with a “cure rate” using Monte Carlo methods.

SPOSTO and SATHER (1985) expand the method of RUBINSTEIN et al. (1981) for determining the length of accrual by replacing the pure exponential model by a cure rate model for exponential survival times.

HALPERN and BROWN (1987a) use the program presented in HALPERN and BROWN (1987b) to compare the power of the log rank and generalised Wilcoxon test in the situation of cure rate models.

CANTOR (1992) uses Gompertz survival functions for modelling cure rates. He gives a program which could be used for this purpose in CANTOR (1991).

**Miscellanea**

SPOSTO and KRAILO (1987) investigate the effect of unequal patient allocation on the power of the log rank test. They conclude that imbalances do not effect the power too much as long as the inferior treatment is given to a smaller proportion of patients.

AKAZAWA et al. (1991) present a SAS/IML simulation program for evaluating the power of specific proportional hazards regression models.

Exemplary data is used to determine sample size by GOLDMAN and HILLMAN (1992).

Methods for planning the duration of a clinical trial when recurrent events occur are described by COOK (1995), modifying the method of RUBINSTEIN et al.

There are a few other papers dealing with special aspects as e.g. screening for chronic disease (STRAATMAN and VERBEEK, 1990) or response surface methods (GOODLOW et al., 1987).

**3.4 Comparisons of Sample Size Formulae – Review Papers****3.4.1 Comparisons or Reviews of Sample Size and Power**

BENEDETTI et al. (1982) analyse the influence of censored observations to the power of the log rank test and the generalised Wilcoxon test. The authors demonstrate that censored observations may increase efficiency compared to uncensored data.

MOUSSA (1988) compares the methods of BERNSTEIN and LAGAKOS (1978), GEORGE and DESU (1974), PASTERNAK and GILBERT (1971), FREEDMAN (1982), SCHOENFELD (1981, 1983), and FLEISS (1981, for proportions) and analyses their dependence on the number of risk groups (strata). FLEISS' method on event rates at the end of study is included in order to compare those formulae with an event-rate design which does not consider the information about time-to-event. MOUSSA (1990) compares approaches for the adjustment of the sample size of clinical trials in the case of noncompliance. Adjustments based on proportions (LACHIN (1981), SCHORK and REMINGTON (1967), HALPERIN et al. (1968), WU et al. (1980)) are compared with adjustment based on the survival distribution (LACHIN and FOULKES (1986)).

HSIEH (1987) uses Monte Carlo simulations for comparing the power of the log rank test in FREEDMAN's (1982), SCHOENFELD's (1981) and his own method for exponentially distributed survival times and unequal allocation.

HSIEH (1992) compares in more detail the formulae of SCHOENFELD (1981), FREEDMAN (1982) and HSIEH (1987), and additionally the method of SHUSTER (1990), for unbalanced designs by use of Monte Carlo simulations. Exponential and lognormal survival distributions are considered.

BELTANGADY and FRANKOWSKI (1989) use a simulation study to examine the effect of random unequal censoring on the size and the power of the two-sample log rank and the Wilcoxon type tests for exponentially distributed survival times.



In the paper of LAKATOS and LAN (1992) the methods of RUBINSTEIN et al. (1981), FREEDMAN (1982) and the Markov method of LAKATOS (1988) are compared under the assumption of exponential survival distribution and general distributions. They suggest that the simple formulae of RUBINSTEIN et al. (1981) and FREEDMAN (1982) can not be accurate in the case of non-proportional hazards. Simulation results for exponential survival, proportional hazards and non-proportional hazards situations are presented.

BONACINA et al. (1993) describe and compare the methods of FREEDMAN (1982), WU et al. (1980), LACHIN and FOULKES (1986), and LAKATOS (1988) by simulations reaching to the conclusion that the method of LAKATOS is the most flexible one.

### 3.4.2 Textbooks on survival analysis, incl. sample size estimation

GROSS and CLARK (1975) derive a formula for the number of deaths required to compare in the two-sample survival case for exponential, Weibull, Rayleigh, and Gamma distributed survival times. In addition, sample size estimates are provided when the population size is known and a description of Nam's procedure (NAM, 1973) on allocation is given.

COLLETT (1994) presents the method and formula (incl. its mathematical deduction) according to GEORGE and DESU (1974) and followers. A short comparison with FREEDMAN's (1982) method results in recommendation of the formula of GEORGE and DESU. A reference to LAKATOS (1988) is made.

MARUBINI and VALSECCHI (1995) follow in their book the approach of FREEDMAN (1982). They give the mathematical background and tables.

PARMAR and MACHIN (1995) also report the formula of FREEDMAN for the proportional hazards assumption. They include short comments on balancing accrual and follow-up, unequal allocation and loss to follow-up. A short paragraph is referring to the exponential distribution assumption with the formula of GEORGE and DESU. They recommend use of the formula of FREEDMAN.

### 3.4.3 Sample size textbooks for clinical trials

MACHIN and CAMPBELL (1987) published a book with tables for the design of clinical trials. For the proportional hazards assumption they included in Section 9 the formula of FREEDMAN (1982), and for the exponential distribution assumption in Section 10 the formula of GEORGE and DESU (1974).

Sample size methodology for comparing survival distributions is also explained in the textbooks of SCHWARTZ et al. (1980), DESU and RAGHAVARAO (1990), SHUSTER (1990), BOCK and TOUTENBURG (1991) and SHUSTER (1993).

An overview of sample size methods for planning survival time trials is given by BOCK (1996 to appear). He compares the methods of GEORGE and DESU, BERNSTEIN and LAGAKOS, SCHOENFELD, RUBINSTEIN et al., FREEDMAN, LACHIN and FOULKES, and LAKATOS, explains the composition of the formulae, presents examples and provides SAS programs to compute the total sample size according to the different methods.

### 3.4.4 Reviews on sample size and power

Power simulations have been used to compare performance and efficiency of test statistics. For early papers see GEHAN and THOMAS (1969), LEE et al. (1975), LININGER et al. (1979), and LATTA (1981).

LACHIN (1981) published a review of methods for sample size determination and power analysis. He deals with t-tests, tests for proportions, survival analysis, and tests for correlation. The methods presented for survival analysis work under the assumption of exponential survival and are discussed with special regard to the problems of limited recruitment and censoring.

Another review on sample size estimation in clinical trials is given by DONNER (1984). He restricts his attention to designs of studies which compare two treatment groups, both with respect to the frequency of events and the time-to-event. Patient drop-out, length of time required to achieve maximum benefit of treatment, and patient accrual by cohorts is discussed.

McHUGH and LE (1984) present sample size determination based upon a confidence estimation approach rather than a significance testing approach. They include a section on censored survival data.

DUPONT and PLUMMER (1990) present a review on power and sample size calculations introducing a computer program (POWER). Their reference is SCHOENFELD and RICHTER (1982) for the exponential distribution assumption.

FAYERS and MACHIN (1995) give a more general discussion about sample size estimation in cancer research. They point out the need for sample size calculation.

#### 4 Discussion

We have found a large number of papers for the design of survival time trials. They range from single and easy to use approximation formulae to very special and detailed algorithms accounting for problems such as loss to follow-up, non-compliance, lag times, drop-in, cure rates, and recurrent events.

We investigated more than one hundred papers on sample size determination in survival time trials. All were characterised and categorised.

Of course, this review may lack a few aspects. So, e.g. we did not investigate Bayesian methods, neither epidemiological problems nor equivalence trials. See e.g. ROEBRUCK et al. (1997) for sample size determination for equivalence trials.

Nevertheless, we are convinced, that the most important papers dealing with methods of sample size determination for survival time data have been included in this review and that the interested reader should find appropriate solutions based on this review at least when he goes into the specific literature.

Some of the papers include source codes or algorithms for computer programming, or programs are available by the authors upon request. There are some public domain or commercial software packages allowing for sample size to be calculated in survival time trial design. For a review on software for sample size determination see ORTSEIFEN et al. (1997).

There is an ongoing development of methods in survival time trial design and specialised trial designs proposals are published. Although, some papers are comparing the methods, from our experience we conclude a strong need of detailed comparisons of methods in order to give well founded recommendations of formulae and algorithms for sample size determination in survival time trial design.

#### Acknowledgements

This work was developed as part of the authors engagement in an Ad Hoc Working Group on Sample Size Determination, a subdivision of the Working Group of Biostatistician in Pharmaceutical Research within the German Region of the International Biometric Society. We thank all members of this Ad Hoc Working Group for many helpful discussions. Especially we thank Dr. Lutz Edler for his many useful comments on the manuscript.

## References

Categorisation of papers is given in brackets, if applicable.

- AHNN, S.; ANDERSON, S. J. (1995): Sample size determination for comparing more than two survival distributions. *Stat Med* **14**, 2273–2282. [A.3].
- AKAZAWA, K.; NAKAMURA, T.; MORIGUCHI, S.; SHIMADA, M.; NOSE, Y. (1991): Simulation program for estimating statistical power of Cox's proportional hazards model assuming no specific distribution for the survival time. *Comp Meth Progr Biomed* **35**, 203–212. [A.3].
- ANDERSEN, P. K. (1987): Conditional power calculations as an aid in the decision whether to continue a clinical trial. *Contr Clin Trials* **8**, 67–74. [A.6].
- BELTANGADY, M. S.; FRANKOWSKI, R. F. (1989): Effect of unequal censoring on the size and power of the log rank and Wilcoxon types of tests for survival data. *Stat Med* **8**, 937–945. [A.4].
- BENEDETTI, J. K.; LIU, P.-Y.; SATHER, H. N.; SEINFELD, J.; EPTON, M. A. (1982): Effective sample size for tests of censored survival data. *Biometrika* **69**, 343–349. [A.4].
- BERNSTEIN, D.; LAGAKOS, S. W. (1978): Sample size and power determination for stratified clinical trials. *J Statist Comput Simul* **8**, 65–73. [A.3].
- BOCK, J. (1996 to appear): *Die Bestimmung des Stichprobenumfanges für biologische Experimente und kontrollierte klinische Studien*. Gustav Fischer Verlag, Stuttgart. [B.1].
- BOCK, J.; TOUTENBURG, H. (1991): Sample size determination in clinical research. In: Rao, C. R. (Ed.): *Handbook of Statistics* **8**, 515–538. [B.1].
- BONACINA, N.; MARIANI, L.; MARUBINI, E. (1993): Metodi di calcolo della dimensione del campione negli studi di sopravvivenza. *Statistica Applicata* **5**, 59–83. [A.4].
- BRESLOW, N. E. (1975): Analysis of survival data under the proportional hazards model. *Int Stat Rev* **43**, 45–57.
- BROWN, B. W.; LOVATO, J.; RUSSEL, K. (1995): Likelihood ratio power and sample size calculations. *Technical Report*. Dept. of Biomathematics, University of Texas, M. D. Anderson Cancer Center. [A.1].
- CANTOR, A. B. (1991): Power estimation for rank tests using censored data: Conditional and unconditional. *Contr Clin Trials* **12**, 462–473. [A.6].
- CANTOR, A. B. (1992): Sample size calculations for the log rank test: a Gompertz model approach. *J Clin Epidemiol* **45**, 1131–1136. [A.6].
- CANTOR, A. B. (1996): Power calculation for the log rank test using historical data. *Contr Clin Trials* **17**, 111–116. [A.6].
- CASE, L. D.; MORGAN, T. M.; DAVIS, C. E. (1987): Optimal restricted two-stage designs. *Contr Clin Trials* **8**, 146–156. [A.3].
- CHEN, T. T.; SIMON, R. M. (1994a): Extension of one-sided test to multiple treatment trials. *Contr Clin Trials* **15**, 124–134. [A.3].
- CHEN, T. T.; SIMON, R. M. (1994b): A multiple decision procedure in clinical trials. *Stat Med* **13**, 431–446. [A.3].
- CHURCH, T. R.; EDERER, F.; MANDEL, J. S.; WATT, G. D.; GEISSER, M. S. (1993): Estimating the duration of ongoing prevention trials. *Am J Epidem* **137**: 797–810. [A.2].
- COLLETT, D. (1994): *Modelling Survival Data in Medical Research*. Chapman and Hall, London, 253–265. [A.5].
- COOK, R. J. (1995): The design and analysis of randomised trials with recurrent events. *Stat Med* **14**, 2081–2098. [A.6].
- COX, D. R. (1972): Regression models and life-tables (with discussion): *J Royal Statist Soc, B* **34**, 187–220.
- DANZIGER, L. (1970): Planning censored life tests for estimation of the hazard rate of a Weibull distribution with prescribed precision. *Technometrics* **12**:408–412. [A.6].
- DESU, M. M.; RAGHAVARAO, D. (1990): *Sample Size Methodology*. Academic Press, San Diego. [B.1].
- DIXON, D. O.; SIMON, R. (1988): Sample size considerations for studies comparing survival curves using historical controls. *J Clin Epidem* **41**, 1209–1213. [A.6].
- DONNER, A. (1984): Approaches to sample size estimation in the design of clinical trials – a review. *Stat Med* **3**, 199–214. [B.2].
- DUPONT, W. D.; PLUMMER, W. D. (1990): Power and sample size calculations: a review and computer program. *Contr Clin Trials* **11**, 116–128. [B.2].

- EMRICH, L. J. (1989): Required duration and power determinations for historically controlled studies of survival times. *Stat Med* **8**, 153–160. [A.6].
- FAYERS, P. M. and MACHIN, D. (1995): Sample size: How many patients are necessary? *Br J Cancer* **72**, 1–9. [B.2].
- FLEISS, J. L. (1981): *Statistical Methods for Rates and Proportions*. John Wiley & Sons, New York.
- FLEMING, T. R.; O'FALLON, J. R.; O'BRIEN, P. C.; HARRINGTON, D. P. (1980): Modified Kolmogorov-Smirnov test procedures with application to arbitrary right-censored data. *Biometrics* **36**, 607–625. [A.6].
- FREEDMAN, L. S. (1982): Tables of the number of patients required in clinical trials using the log rank test. *Stat Med* **1**, 121–129. [A.1].
- FREEDMAN, L. S. (1986): Reply to comments on "Freedman, L. S. (1982). Tables of the number of patients required in clinical trials using the logrank test" (1: 121–129). *Stat Med* **5**, 98–98. [A.4].
- FRICK, H. (1991): On Lachin's formulae for sample sizes of survival tests. *Comm Stat – Ther Meth* **20**, 2267–2280. [A.1].
- GAIL, M. H. (1985): Applicability of sample size calculations based on a comparison of proportions for use with the logrank test. *Contr Clin Trials* **6**, 112–119. [A.4].
- GAIL, M. H. (1994): Sample size estimation when time-to-event is the primary endpoint. *Drug Inf J* **28**, 865–877. [A.1].
- GEBSKI, V.; MCNEIL, D.; COATES, A.; FORBES, J. (1987): Monitoring distributional assumptions and early stopping for a prospective clinical trial using Monte Carlo simulation. *Stat Med* **6**:667–678. [A.6].
- GEHAN, E. A.; THOMAS, D. G. (1969): The performance of some two-sample tests in small samples with or without censoring. *Biometrika* **56**, 127–132. [A.4].
- GEORGE, S. L.; DESU, M. M. (1974): Planning the size and duration of a clinical trial studying the time to some critical event. *J Chron Dis* **27**, 15–24. [A.1].
- GOLDMAN, A. I. (1984): Survivorship analysis when cure is a possibility: A Monte Carlo study. *Stat Med* **3**:153–163. [A.6].
- GOLDMAN, A. I.; HILLMAN, D. W. (1992): Exemplary data: Sample size and power in the design of event-time clinical trials. *Contr Clin Trials* **13**, 256–271. [A.3].
- GOODLOW, J. L.; STABLEIN, D. M.; CARTER, W. H. J. (1987): Sample sizes for cancer therapy survival studies based on response surface methods. *J Stat Comput Simul* **28**, 31–41. [A.6].
- GROSS, A. J.; CLARK, V. A. (1975): *Survival Distributions: Reliability Applications in the Biomedical Sciences*. John Wiley & Sons, New York. [A.5].
- GROSS, A. J.; HUNT, H. H.; CANTOR, A. B.; CLARK, B. C. (1987): Sample size determination in clinical trials with an emphasis on exponentially distributed responses. *Biometrics* **43**, 875–883. [A.2].
- HALPERIN, M.; ROGOT, E.; GURIAN, J.; EDERER, F. (1968): Sample sizes for medical trials with special reference to long-term therapy. *J Chron Dis* **21**, 13–24. [A.1].
- HALPERN, J.; BROWN JR., B. W. (1987a): Cure rate models: power of the log rank and generalised Wilcoxon tests. *Stat Med* **6**, 483–489. [A.6].
- HALPERN, J.; BROWN JR., B. W. (1987b): Designing clinical trials with arbitrary specification of survival functions and for the log rank or generalised Wilcoxon test. *Contr Clin Trials* **8**, 177–189. [A.2].
- HALPERN, J.; BROWN JR., B. W. (1993): A computer program for designing clinical trials with arbitrary survival curves and group sequential testing. *Contr Clin Trials* **14**, 109–122. [A.6].
- HENDERSON, W. G.; FISHER, S. G.; WEBER, L.; HAMMERMEISTER, K. E.; SETHI, G. (1991): Conditional power for arbitrary curves to decide whether to extend a clinical trial. *Contr Clin Trials* **12**, 304–313. [A.6].
- HSIEH, F. Y. (1987): A simple method of sample size calculation for unequal-sample-size designs that use the logrank or t-test. *Stat Med* **6**, 577–581. [A.1].
- HSIEH, F. Y. (1992): Comparing sample size formulae for trials with unbalanced allocation using the logrank test. *Stat Med* **11**, 1091–1098. [A.4].
- KIM, K. (1992): Study duration for group sequential clinical trials with censored survival data adjusting for stratification. *Stat Med* **11**, 1477–1488. [A.3].
- KIM, K. (1995): SEQPWR and SEQOPR: Computer programs for design of maximum information trials based on group sequential logrank tests. *Comp Meth Progr Biomed* **46**, 143–153. [A.3].

- KIM, K.; TSIATIS, A. A. (1990): Study duration for clinical trials with survival response and early stopping rule. *Biometrics* **46**, 81–92. [A.6].
- LACHIN, J. M. (1981): Introduction to sample size determination and power analysis for clinical trials. *Contr Clin Trials* **2**, 93–113. [B.2].
- LACHIN, J. M.; FOULKES, M. A. (1986): Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, non-compliance, and stratification. *Biometrics* **42**, 507–519. [A.1].
- LAKATOS, E. (1986): Sample size determination in clinical trials with time-dependent rates of losses and non-compliance. *Contr Clin Trials* **7**, 189–199. [A.3].
- LAKATOS, E. (1988): Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics* **44**, 229–241. [A.3].
- LAKATOS, E.; LAN, K. K. G. (1992): A comparison of sample size methods for the log rank statistic. *Stat Med* **11**, 179–191. [A.4].
- LAN, K. K. G.; LACHIN, J. M. (1990): Implementation of group sequential logrank tests in a maximum duration trial. *Biometrics* **46**, 759–770. [A.3].
- LATTA, R. B. (1981): A Monte-Carlo study of some two-sample rank tests with survival data. *JASA* **76**, 713–719. [A.4].
- LEE, E. T.; DESU, M. M.; GEHAN, E. A. (1975): A Monte Carlo study of the power of some two-sample tests. *Biometrika* **62**, 425–432. [B.2].
- LININGER, L.; GAIL, M. H.; GREEN, S. B.; BYAR, D. P. (1979): Comparison of four tests for equality of survival curves in the presence of stratification and censoring. *Biometrika* **66**, 419–428. [A.4].
- LIU, P.-Y.; DAHLBERG, S. (1995): Design and analysis of multiarm clinical trials with survival endpoints. *Contr Clin Trials* **16**, 119–130. [A.3].
- LIU, P.-Y.; DAHLBERG, S.; CROWLEY, J. (1993): Selection designs for pilot studies based on survival. *Biometrics* **49**, 391–398. [A.6].
- LUI, K.-J. (1992): Sample size determination under an exponential model in the presence of a confounder and type I censoring. *Contr Clin Trials* **13**, 446–458. [A.3].
- LUSTBADER, E. D.; LITWIN, S. (1981): Design of censored survival studies. *Biometrics* **37**, 697–704. [A.2].
- MACHIN, D.; CAMPBELL, M. J. (1987): *Statistical Tables for the Design of Clinical Trials*. Blackwell Scientific Publications, Oxford. [B.1].
- MAKUCH, R. W.; SIMON, R. M. (1982): Sample size requirements for comparing time-to-failure among  $k$  treatment groups. *J Chron Dis* **35**, 861–867. [A.3].
- MARUBINI, E.; VALSECCHI, M. G. (1995): *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley & Sons, New York. [A.5].
- MCHUGH, R. B.; LE, C. T. (1984): Confidence estimation and the size of a clinical trial. *Contr Clin Trials* **5**, 157–163. [B.2].
- MORGAN, T. M. (1985): Planning the duration of accrual and follow-up for clinical trials. *J Chron Dis* **38**, 1009–1018. [A.2].
- MORGAN, T. M. (1987): Nonparametric estimation of duration of accrual and total study length for clinical trials. *Biometrics* **43**, 903–912. [A.2].
- MORGAN, T. M.; ELASHOFF, R. W. (1986): Effect of censoring on adjusting for covariates in comparison of survival times. *Commun Stat-Theor Meth* **15**, 1837–1854. [A.6].
- MOUSSA, M. A. A. (1988): Planning the size of survival time clinical trials with allowance for stratification. *Stat Med* **7**, 559–569. [A.4].
- MOUSSA, M. A. A. (1990): Planning the size of clinical trials with allowance for patient non-compliance. *Meth Inform Med* **29**, 243–246. [A.4].
- NAGELKERKE, N. J. D.; HART, A. A. M.; STRACKEE, J. (1981): Sample sizes in sequential survival clinical trials. *Biom J./ J Math Meth Biosc* **23**, 709–714. [A.3].
- NAM, J.-M. (1973): Optimum sample sizes for the comparison of the control and treatment. *Biometrics* **29**:101–108. [A.1].
- NATARAJAN, R.; TURNBULL, B. W.; SLATE, E. H.; CLARK, L. C. (1996): A computer program for sample size and power calculations in the design of multi-arm and factorial clinical trials with survival time endpoints. *Comp Meth Progr Biomed* **49**, 137–147. [A.3].
- ORTSEIFEN, C.; BRUCKNER, T.; BURKE, M.; KIESER, M. (1997): An overview of software tools for sample size determination. *Inf Biom Epidem Med Biol* **28**.
- PALTA, M.; AMINI, S. B. (1985): Consideration of covariates and stratification in sample size determination for survival time studies. *J Chron Dis* **38**, 801–809. [A.3].

- PARMAR, M. and MACHIN, D. (1995): *Survival Analysis: A Practical Approach*. John Wiley & Sons, New York. [A.5].
- PASTERNAK, B. S. (1972): Sample size for clinical trials designed for patient accrual by cohorts. *J Chron Dis* **25**, 673–681. [A.1].
- PASTERNAK, B. S.; GILBERT, H. S. (1971): Planning the duration of long-term survival time studies designed for accrual by cohorts. *J Chron Dis* **24**, 681–700. [A.2].
- PETERSON, B.; GEORGE, S. L. (1993): Sample size requirements and length of study for testing interaction in a 1xk factorial design when time-to-failure is the outcome. *Contr Clin Trials* **14**, 511–522. [A.3].
- PETO, R.; PIKE, M. C.; ARMITAGE, P.; BRESLOW, N. E.; COX, D. R.; HOWARD, S. V.; MANTEL, N.; MCPHERSON, K.; PETO, J.; SMITH, P. G. (1976): Design and analysis of randomised clinical trials requiring prolonged observation of each patient – I. Introduction and design. *Br J Cancer* **34**, 585–612.
- ROEBRUCK, P.; ELZE, M.; HAUSCHKE, D.; LEVERKUS, F. (1997): Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme. *Inf Biom Epidem Med Biol* **28**.
- RUBINSTEIN, L. V.; GAIL, M. H.; SANTNER, T. J. (1981): Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *J Chron Dis* **34**, 469–479. [A.2].
- SCHAID, D. J.; WIEAND, S.; THERNEAU, T. M. (1990): Optimal two-stage screening designs for survival comparisons. *Biometrika* **77**, 507–513. [A.6].
- SCHOENFELD, D. A. (1981): The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* **68**, 316–319. [A.1].
- SCHOENFELD, D. A. (1983): Sample-size formula for the proportional-hazards regression model. *Biometrics* **39**, 499–503. [A.3].
- SCHOENFELD, D. A.; GELBER, R. D. (1979): Designing and analyzing clinical trials which allow institutions to randomize patients to a subset of the treatments under study. *Biometrics* **35**, 825–829. [A.6].
- SCHOENFELD, D. A.; RICHTER, J. R. (1982): Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics* **38**, 163–170. [A.1].
- SCHORK, M. A.; REMINGTON, R. D. (1967): The determination of sample size in treatment-control comparisons for chronic disease studies in which drop-out or non-adherence is a problem. *J Chron Dis* **20**, 233–239. [A.1].
- SCHUMACHER, M. (1981): Power and sample size determination in survival time studies with special regard to the censoring mechanism. *Meth Inform Med* **20**, 110–115. [A.1].
- SCHWARTZ, D.; FLAMANT, R.; LELLOUCH, J. (1980): *Clinical Trials*. Academic Press, London. [B.1].
- SHIH, J. H. (1995): Sample size calculation for complex clinical trials with survival endpoints. *Contr Clin Trials* **16**, 395–407. [A.3].
- SHUSTER, J. J. (1990): *CRC Handbook of Sample Size Guidelines for Clinical Trials*. CRC Press, Boca Raton. [B.1].
- SHUSTER, J. J. (1993): *Practical Handbook of Sample Size Guidelines for Clinical Trials: IBM Compatible Version*. CRC Press, Boca Raton. [B.1].
- SPOSTO, R.; KRAILO, M. (1987): Use of unequal allocation on survival trials. *Stat Med* **6**, 119–126. [A.6].
- SPOSTO, R.; SATHER, H. N. (1985): Determination of the duration of comparative clinical trials while allowing for cure. *J Chron Dis* **38**, 683–690. [A.6].
- STEIN, R. A. (1986): Comments on "Freedman, L. S. (1982). Tables of the number of patients required in clinical trials using the log rank test" (1: 121–129). *Stat Med* **5**, 97–97. [A.4].
- STRAATMAN, H.; VERBEEK, A. (1990): Shortcut method to calculate the sample size in trials of screening for chronic disease. *J Clin Epidem* **43**, 1261–1266. [A.6].
- TARONE, R.; WARE, J. (1977): On distribution-free tests for equality of survival distributions. *Biometrika* **64**, 156–160.
- TAULBEE, J. D.; SYMONS, M. J. (1983): Sample size and duration for cohort studies of survival time with covariables. *Biometrics* **39**, 351–360. [A.3].
- TAVARÉ, C.; SOBEL, E.; GILLES, F. (1995): Misclassification of a prognostic dichotomous variable: Sample size and parameter estimate adjustment. *Stat Med* **14**, 1307–1314. [A.6].
- WILTSE, C. G.; LACHENBRUCH, P. A.; WOOLSON, R. F. (1980): A sampling study of the size and power of tests for paired survival data. *Biom J* **22**, 633–638. [A.6].

- WU, M.; FISHER, M.; DEMETS, D. L. (1980): Sample sizes for long-term medical trial with time-dependent dropout and event rates. *Contr Clin Trials* **1**, 109–121. [A.1].
- XIANG, A. H.; SATHER, H. N.; AZEN, S. P. (1994): Power considerations for testing an interaction in a  $2 \times k$  factorial design with a failure time outcome. *Contr Clin Trials* **15**, 489–502. [A.3].
- YATEMAN, N. A. (1992): Choice of sample size for proportional hazards survival studies. In *ASA Proc Biopharm Sec* 152–157. [A.1].
- YATEMAN, N. A.; SKENE, A. M. (1992): Sample sizes for proportional hazards survival studies with arbitrary patient entry and loss to follow-up distributions. *Stat Med* **11**, 1103–1113. [A.1].
- YATEMAN, N. A.; SKENE, A. M. (1993): The use of simulation in the design of two cardiovascular survival studies. *Stat Med* **12**, 1365–1372. [A.1].
- ZHEN, B.; MURPHY, J. (1994): Sample size determination for an exponential survival model with an unrestricted covariate. *Stat Med* **13**, 391–397. [A.1].

## Authors addresses:

S. Oellrich, biodat Gesellschaft für Biometrie und Statistik mbH, Oranienstr. 25, 10999 Berlin, Tel.: +49-(0)30/6 14 80 71, Fax: +49-(0)30/6 14 80 77, E-mail: 101523.664@compuserve.com

F. Freischläger, IMFORM GmbH, Abt. Biometrie, Pallaswiesenstr. 174a, 64293 Darmstadt, Tel.: +49-(0)6151/82 80 30, Fax: +49-(0)6151/82 80 10, E-mail: imform.cro@t-online.de

A. Benner, DKFZ, Abt. Biostatistik, Postfach 10 19 49, 69009 Heidelberg, Tel.: +49-(0)6221/42 23 90, Fax: +49-(0)6221/ 42 23 97, E-mail: benner@dkfz-heidelberg.de

Dr. M. Kieser, Dr. Willmar Schwabe GmbH & Co., 76227 Karlsruhe, Tel.: +49-(0)721/ 4005572, Fax: +49-(0)721/4005333

## Sample size estimation for nonparametric tests

Wilhelm Sauermann

### Summary

*This paper gives an overview about the different approaches used to calculate the power and necessary sample size of rank tests. The classical asymptotic theory and its refinements are explained. Furthermore the bootstrap as a very general method to estimate the power of a test is presented. The relevant literature is described.*

### Keywords

*nonparametric, sample size, power, rank test, bootstrap*

### Zusammenfassung

*In dieser Arbeit wird ein Überblick über die verschiedenen Ansätze zur Berechnung der Macht und des benötigten Stichprobenumfangs von Rangtests gegeben. Es wird die klassische asymptotische Theorie sowie ihre Verfeinerungen erklärt. Weiterhin wird das Bootstrap-Verfahren als sehr allgemeine Methode zur Schätzung der Macht eines Tests vorgestellt. Es wird eine Beschreibung der relevanten Literatur gegeben.*

### Schlüsselwörter

*nichtparametrisch, Stichprobenumfang, Macht, Rangtest, Bootstrap*

### 1. Introduction

This paper is intended to give an overview of two areas of nonparametric statistics: power and sample size formulae for rank tests, which for practical purposes are the most important class of nonparametric tests, and methods to determine the power of nonparametric tests by means of the bootstrap.

The theory of rank tests is nowadays well developed and is described in a number of monographs. It is intended here to present the basic ideas and approaches and to give some hints, which results can be found at which place.



For rank tests, first the classical asymptotic approach is described. Then some refinements of the asymptotic formulae and alternative approximations are reviewed. All these approximations allow to calculate the power and sample size for given parametric models. This is in contrast to the approach of Noether, which expresses the power and sample size in terms of quantities which have a nonparametric interpretation. In contrast, the bootstrap is a new approach, which can be used to estimate the power function of a test, given that some data in the same experimental situation are available.

## 2. Asymptotic theory for rank tests

The classical asymptotic theory for rank tests is characterized by the following features:

- (1) The theory is applicable to a large number of tests.
- (2) The power and sample size can be evaluated for a large class of local alternatives, which only have to fulfill certain regularity conditions.
- (3) The formulae are only valid for "large" sample sizes.

Features (1) and (2) are detailed in the sequel:

- (1) The asymptotic theory is not only applicable to often used tests like the Wilcoxon test, but also to nearly all tests of practical relevance. For example, the monograph of HÁJEK and ŠIDAK (1967) describes among others two-sample tests of location and scale, tests for three or more samples, tests for symmetry, tests for independence, and tests for random blocks.
- (2) The power can be evaluated for local (more precisely: contiguous) alternatives, i.e. for alternatives which converge to the null hypothesis, if the sample size increases to infinity. In an asymptotic framework it is necessary to consider such local alternatives, since if the sample size increases and the null and alternative hypothesis are fixed, the power converges to 1. In regular parametric models local alternatives are obtained if the parameters are of the order  $\theta/\sqrt{n}$ . The class of parametric models, for which the power can be evaluated, is, however, not restricted to simple location or scale models. For example, models, in which only a certain number of subjects responds to a treatment, can be considered as well (CONOVER and SALSBERG (1988)). The models are only required to be "regular". Technically, "regularity" means differentiability in the quadratic mean of the square root of the density as a function of the parameter. It may sound contradictory to evaluate the power of nonparametric tests under parametric models, but this is a consequence of the fact, that nonparametric tests behave differently under different alternatives.

In the following sections some guidance regarding the literature is given. Since the theory is well developed, it can be found in several monographs.

The book of HÁJEK and ŠIDAK (1967) is a comprehensive theoretical account of the theory. After a general introduction to hypothesis tests and distributions (Chapter I) it describes the null hypotheses, which can be tested by means of rank tests, in general and shows, how locally most powerful rank tests can be constructed (Chapter II). In Chapter III selected rank tests are introduced and in Chapter IV methods to calculate the exact distribution of the test statistics under the null hypothesis are described. The asymptotic distribution of the test statistics under the null hypothesis is derived in Chapter V. This chapter covers simple linear rank statistics (which are asymptotically normally distributed), rank statistics of the chisquare-types and statistics of the Kolmogorov-Smirnov types. Chapter VI gives the asymptotic distribution of the test statistics under the alternatives and Chapter VII is concerned with optimality questions.

The book is self-contained and derives all the necessary theory to calculate the asymptotic power and sample size for most rank tests. In Chapter VI the asymptotic tools (e.g.

LeCam's third lemma) are proved. Formulae for the asymptotic power are given but no formulas for the sample size. To use the book, it is necessary to become familiar with the mathematical formalism. No results concerning the quality of the asymptotic approximation are given.

LEHMANN'S (1975) book is a fairly elementary book on rank tests. It covers the most frequently used tests and gives formulas for the asymptotic power (but not the sample size) for some of them (e.g. the Wilcoxon test and the sign test) for simple models. Each section ends with a useful description of further developments such as other approximations of the power function.

SIMPSON and MARGOLIN (1990) consider a class of rank tests for the monotonicity of a dose-response relationship where downturns of the dose-response curve are admissible. The class of tests includes the Jonckheere-Terpstra test as a limiting case. The asymptotic power function (but no sample size formula) is derived and is plotted for some cases. The paper also presents the results from some Monte Carlo simulations to assess the goodness of the approximation. The paper is fairly theoretical and familiarity with asymptotic theory is helpful.

One characteristic of linear rank statistics is, that their power depends not only on the parameter of interest but also on the null distribution. For instance, in a two-sample location model, the Wilcoxon test is locally optimal if the null distribution is the logistic distribution. If, however, the null distribution is the normal distribution, the locally optimal test is the normal scores test. Results in HÁJEK and ŠIDAK (1967) can be interpreted geometrically (see also NEUHAUS (1987)): the asymptotic power of a linear rank test depends on the cosine of the angle (in the space of quadratically differentiable functions) between its score function and the optimal score function under true underlying model. In practice, however, the true underlying model (i.e. the null distribution) is unknown. Therefore the optimal score function and the optimal test are unknown. One approach, already described in HÁJEK and ŠIDAK (1967), is to estimate the optimal score function from the data, e.g. by a suitable kernel estimator. Such "adaptive" tests are an alternative to fixing in advance a test like the Wilcoxon test and thus relying on the logistic distribution. The approach was refined by Behnen, Neuhaus and others in a series of papers of which only NEUHAUS (1987) will be mentioned here. This mathematically written paper develops the asymptotic theory and gives the formulae for the asymptotic power. The theory is extended to models general than the location model.

### 3. Exact power calculations, Edgeworth expansions and other approximations

Although asymptotic approximations are usually sufficient for large sample sizes, they may not be appropriate for small and moderate sample sizes. The exact calculation of the power of rank tests is possible in principle by means of numerical integration, but can be done in practice only for very small sample sizes (say  $N \leq 10$ ).

Only few papers give the exact power of rank tests (regarding the one-sample Wilcoxon test see KLOTZ (1963) and ARNOLD (1965); regarding the two-sample Wilcoxon test see MILTON (1970), BELL, MOSER and THOMPSON (1966), and HAYNAM and GOVINDARAJULU (1966)). One possibility to improve the accuracy of the asymptotic (first order) approximations is to derive (higher order) Edgeworth expansions for the power. For the one-sample symmetry problem ALBERS, BICKEL and VAN ZWET (1976) and for the two-sample problem BICKEL and VAN ZWET (1978) do the necessary calculations.

Another possibility is the quadratic log-likelihood ratio (QLLR) method, described in CHOW and BJERVE (1984), which is related to the average probability (AP) method (CHOW and HODGES (1975)).

The QLLR method requires the complete knowledge of the distribution of the test statistic under the null hypothesis and the knowledge of the moments of the order 0, 1 and 2 of the test statistic under the alternative. The method approximates the log-likelihood ratio of the alternative distribution of the test statistic with respect to the null distribution by a quadratic function. Another way to view this is, that the null and alternative distribution of the test statistic are imbedded in a two-parameter exponential family such that the moments of the order 0, 1 and 2 coincide. The parameter values of the exponential family corresponding to the alternative distribution of the test statistic have to be found by numerical minimization. The approximate power of the test statistic can then be derived from the distribution of the exponential family at these parameter values. CHOW and BJERVE (1984) describe the method for the Wilcoxon test, prove that the minimization always leads to a unique solution, show how to perform the minimization and present a comparison of the power derived by the QLLR method with the exact power and other approximations.

The AP method approximates the log-likelihood ratio by a quartic polynomial, requiring additionally, that certain probabilities coincide. Although this theoretically leads to a more precise approximation, the existence of a suitable quartic polynomial cannot always be guaranteed. CHOW and HODGES (1975) derive the method for the one-sample Wilcoxon test and show some numerical comparisons.

In her diploma thesis, STOLLENWERK (1989) gives an overview about these approximations, presents extensive numerical comparisons including comparisons with first order approximations, and outlines the structure of computer programs to perform the approximations.

#### 4. Noether's approach

The approach of NOETHER (1987) differs from the classical approach in that the power is not evaluated under parametric models but is expressed in terms of quantities independent of a specific model. For example, for the two-sample Wilcoxon test the power and sample size are given as functions of  $\text{Prob}(X < Y)$ , where  $X$  and  $Y$  are independent random variables corresponding to the null and alternative hypothesis. In order to estimate the required sample size, an investigator has to specify  $\text{Prob}(X < Y)$ . The quantity which has to be specified is dependent on the test for which the sample size has to be calculated. The results are valid asymptotically. The paper covers the sign test, the one- and two-sample Wilcoxon test and Kendall's test of independence. Explicit formulae for the sample size are given.

#### 5. Bootstrap procedures

The bootstrap is a very general nonparametric method to estimate the operating characteristics of a given statistical procedure. In the simplest case it can be used to estimate the bias, the variance or the entire distribution of an estimator. It was noted by BERAN (1986), that the bootstrap can also be used to estimate the power function of a statistical test. Beran developed the asymptotic theory for a very general model which has a parameter of interest to be tested and a nuisance parameter which has an influence on the power but is not of interest itself. An example would be a nonparametric location model in which the location parameter is the parameter of interest and the shape of the distribution is the nuisance parameter. It can be shown, that — given a sequence of observations with an underlying (unknown) nuisance parameter — the power as a function of the

parameter of interest can be estimated for the underlying nuisance parameter. Practically, this can be used in a situation, in which a pilot study had been performed and the assumption appears to be valid that the data in the study to be planned have the same distribution except for the parameter of interest (i.e. the nuisance parameter is the same). Another way to view this is that the bootstrap uses the complete data from a previous study and not just an estimate of the variance as is usually done. The method is computer-intensive but does not need distributional assumptions. It lies in the nature of the bootstrap approach that the power is not given by a formula. BERAN's (1986) paper is very general and mathematical. It presents results of a few simulations. In contrast, COLLINGS and HAMILTON (1988) is a more practical paper. Several bootstrap procedures are described for the two-sample Wilcoxon test in a location model. Calculations are carried out for a specific example. Also results from some Monte Carlo simulations regarding the validity of the bootstrap are described.

## 6. References

- ALBERS, W., BICKEL, P. J., VAN ZWET, W. R. (1976): Asymptotic expansions for the power of distribution free tests in the one-sample problem. *Ann Statist* **4**, 108–156.
- ARNOLD, H. J. (1965): Small sample power of the one-sample Wilcoxon test for non-normal shift alternatives. *Ann Math Statist* **36**, 1767–1778.
- BELL, C. B., MOSER, J. M., THOMPSON, R. (1966): Goodness criteria for two-sample distribution-free tests. *Ann Math Statist* **37**, 133–142.
- BERAN, R. (1986): Simulated power functions. *Ann Statist* **14**, 151–173.
- BICKEL, P. J., VAN ZWET, W. R. (1978): Asymptotic expansions for the power of distribution free tests in the two-sample problem. *Ann Statist* **6**, 937–1004.
- CHOW, W. K., BJERVE, S. (1984): A new method of power approximation with application to the Wilcoxon statistics. *Scand J Statist* **11**, 29–37.
- CHOW, W. K., HODGES, J. R. (1975): An approximation for the distribution of the Wilcoxon one-sample statistic. *JASA* **70**, 648–655.
- COLLINGS, B. J., HAMILTON, M. A. (1988): Estimating the power of the two-sample Wilcoxon test for location shift. *Biometrics* **44**, 847–860.
- CONOVER, W. J., SALSBERG, D. S. (1988): Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to "respond" to treatment. *Biometrics* **44**, 189–196.
- HÁJEK, J., ŠÍDAK, Z. (1967): *Theory of rank tests*. New York, San Francisco, London: Academic press.
- HAYNAM, G. E., GOVINDARAJULU, Z. (1966): Exact power of Mann-Whitney tests for exponential and rectangular alternatives. *Ann Math Statist* **37**, 945–953.
- KLOTZ, J. H. (1963): Small sample power and efficiency for the one sample Wilcoxon and normal tests. *Ann Math Statist* **34**, 624–632.
- LEHMANN, E. L. (1975): *Nonparametrics: Statistical methods based on ranks*. San Francisco: Holden-Day.
- MILTON, R. C. (1970): *Rank order probabilities: two-sample normal shift alternatives*. New York: Wiley.
- NEUHAUS, G. (1987): Local asymptotics for linear rank statistics with estimated score functions. *Ann Statist* **15**, 491–512.
- NOETHER, G. E. (1987): Sample size determination for some common nonparametric tests. *JASA* **82**, 645–647.
- SIMPSON, D. G., MARGOLIN, B. H. (1990): Nonparametric testing for dose-response curves subject to downturns: asymptotic power considerations. *Ann Statist* **18**, 373–390.
- STOLLENWERK, G. (1989): *Trennschärfe von Zweistichproben tests*. Diplomarbeit Universität Heidelberg, Fachhochschule Heilbronn, Studiengang Medizinische Informatik.

Anschrift des Verfassers: Wilhelm Sauer mann, DATAMAP GmbH, Munzinger Straße 5a, 79111 Freiburg

## An Overview of Software Tools for Sample Size Determination

Carina Ortseifen, Thomas Bruckner,  
Martin Burke, Meinhard Kieser

### Summary

*This paper provides a survey of currently available software for the calculation of sample size and power. Operating systems, costs and program features are examined, and purchase information is summarized in a table. A comparison of results for some standard designs is given. Further, some non-commercial procedures, such as SAS macros and S-Plus functions, are considered, and references to pages on the World Wide Web (WWW) relating to sample size determination are given.*

### Keywords

*sample size determination, software*

### Zusammenfassung

*Dieser Artikel soll einen Überblick über Software zur Bestimmung des Stichprobenumfanges geben. Das erforderliche Betriebssystem, Kosten und Eigenschaften dieser Programme werden dargestellt, Hersteller- und Bestelladressen angegeben. Die Ergebnisse der Fallzahlschätzungen der Programme werden für einige Standardsituationen verglichen. Nicht-kommerzielle Programme wie SAS-Makros oder S-Plus Funktionen werden genannt und Internetadressen zum Thema Fallzahlschätzung angegeben.*

### Schlüsselworte

*Fallzahlschätzung, Software*

### Introduction

A variety of software is available for experimental design and sample size calculation, varying considerably in type, capabilities, and cost. Software types include source code

for compilation, command-line oriented DOS programs, and Windows applications. Also, a programmable calculator is available with routines for sample size determination. Some programs are available as free- or shareware, while others are commercially available, with prices exceeding DM 1000.—. Most of the programs evaluated are written for sample size determination only, however, some data analysis and training programs also contain this feature. The programs are designed for use by biometricians, clinical staff, and students with a background in statistics or biometrics.

This article examines the features of some of the programs available for sample size determination. Programs and routines restricted to power calculation in specific test situations are not considered.

In the following section, a detailed description of the programs' features and documentation is given. The next section is devoted to the presentation of results obtained from a study comparing computed sample size estimates for the independent t-test; comparison of two proportions; survival analysis, and equivalence tests. Literature references related to the software and theoretical aspects of sample size determination are given in separate sections. Tabular surveys of software included in the comparison are given in Appendix A, and references to a variety of internet resources related to the topic are summarized in Appendix B.

The selection of the programs examined in this article is not comprehensive, and not meant to serve as a recommendation.

We thank all participating software companies and distributors for support and provision of complete programs and documentation for this study. Two product beta versions were included in the study.

### Description of the Programs

Programs for sample size determination have been developed for more than 10 years. A full list of the programs investigated is given in Table A.1, Appendix A. Programs may be classified as commercial, public domain, freeware, or shareware. Freeware is any software provided by a developer free of charge, often distributed via the internet or similar media. Shareware is usually distributed similarly and is initially free of charge for an evaluation period, after which the user is expected to register and pay fees, in case of continued use. Public domain programs are similar to freeware, the difference being that the developer waives any rights to the code. Distributor of commercial software were asked to provide evaluation copies using a standard letter. Programs available for the comparison are marked 'C' (or 'D', for demo version) in Table A.1. A summary of suppliers, hardware requirements, and prices is presented in Table A.2 (current: January 1997).

The programs SEPARATE, TRIQ, EAST and PEST are not included in the study because they are restricted to special designs: SEPARATE is used in design and analysis of multiple decision strategies for selection and partitioning in parallel designs, crossover and block designs, and for comparison of more than two treatment groups. Users can choose between

- (a) comparison of treatments with the best,
- (b) comparison of treatments with a control or known standard,
- (c) partitioning into good and poor treatments, and
- (d) partitioning into good, equivalent, and poor treatments.

To run SEPARATE, installation of DBOS (distributed with the software) is required. SEPARATE is very fast for the intensive calculations performed. Of TRIQ, a software tool for planning sequential designs, only a demo version was available, and no copies of either

EAST or PEST could be obtained. Sample size determination for sequential designs was not included in the comparison. For more information on TRIQ, see SCHMIDTKE (1996), for a detailed review of both EAST and PEST see EMERSON (1996). The programs SOLO-POWER and PASS for DOS only differ in name, thus, only one SOLO-POWER is referred to in the analysis. For PASS 6.0 for Windows, EGRET SIZ and POWER & EFFECT, only informational material could be obtained. A total of 20 programs is evaluated in detail.

Table 1: Designs supported by the programs

Program	Means <sup>1)</sup>				Proportions <sup>1)</sup>					Others <sup>1)</sup>				
	1	2 diff	2 equiv	k anova	1	2 diff	2 equiv	k kxc	k trend	Survival	Correlation	Non-par.	Confidence	Distr. Funct
CADEMO <sup>2)</sup>	•	•		•	•									
CADEMO LIGHT	•	•		•	•	•		•			•			
DATASIM	•	•		•				•			•			•
EX-SAMPLE	•	•		•	•	•		•		•	•	•	•	
GPOWER	•	•		•				•			•			
IFNS	•	•	•			•						•	•	
INSTAT		•				•								•
N	•	•	•		•	•	•							
NQUERY	•	•	•	•	•	•	•		•	•	•		•	•
NSURV										•				
PC-SIZE	•	•		•		•					•		•	
PLANUNG	•	•	•		•	•	•		•	•		•		
POWER	•	•	•		•	•				•				
POWERPACK	•	•		•	•			•			•	•		
POWERT	•	•												
SMPLSIZE		•		•		•					•		•	
Solo-Power	•	•	•	•	•	•	•	•		•	•			•
SSIZE		•				•	•			•				
STAT-POWER	•	•		•	•	•		•			•		•	•
STPLAN	•	•			•	•		•		•	•			

<sup>1)</sup> 1: one group comparison versus a fixed value; 2 diff: test for difference of two groups; 2 equiv: test for equivalence of two groups; k anova: comparison of more than two groups (anova); k kxc:  $\chi^2$ -Test; k trend: test on linear trend of proportions.

<sup>2)</sup> Analysis of CADEMO is restricted to the designs supported by the modules *MIWA* and *Versuchsanlagen*, which were available for evaluation.

An overview of the designs supported by the programs investigated in this study is given in Table 1. The statistical designs are separated into three categories: 'Comparison of means', 'Comparison of proportions' and 'Others'. The third category provides information on support of sample size calculations for survival analysis and correlation analysis, nonparametric tests, and confidence intervals, and determination of quantiles and probabilities for given distribution functions.

Table 2: Important characteristics of the programs

Program	Language <sup>1)</sup>		Program features <sup>2)</sup>						Additional Tools <sup>3)</sup>					Calculated Parameters <sup>4)</sup>		
	E	G	C	M	H	P	S	O	R	G	T	D	V	S	P	O
CADEMO/ CADEMO LIGHT	•	•		•	•		•	•	•			•	•	•		•
DATASIM	•		•		•					•				•	•	
EX-SAMPLE	•			•	•		•	•	•				•	•		
GPOWER	•			•	•	•		•	•	•		•	•	•	•	
IFNS	•		•											•	•	
INSTAT	•		•		•							•		•	•	
N/NSURV		•		•	•		•	•	•			•	•	•	•	•
NQUERY	•			•	•		•	•	•	•	•		•	•	•	•
PC-SIZE	•		•			•		•	•			•	•	•		
PLANUNG		•	•			•		•			•	•		•	•	•
POWER	•		•		•	•	•	•	•			•	•	•	•	•
POWERPACK	•		•		•	•		•						•		
POWERT	•		•		•		•							•		
SEPARATE	•	•		•	•	•	•	•	•			•	•	•		
SMPLSIZE	•		•				•							•		
SOLO-POWER	•			•	•		•	•	•	•	•		•	•	•	•
SSIZE	•		•									•	•	•		
STAT-POWER	•			•	•	•		•	•		•	•	•	•	•	
STPLAN	•		•			•	•	•			•			•	•	•

<sup>1)</sup> G: German, E: English.

<sup>2)</sup> C: Command line, M: Mask, H: Online help, P: Protocol or logfile, S: Summary of results, O: Output on file or printer.

<sup>3)</sup> R: Report to file, G: Graphs, T: Tables, D: Default values for parameters, V: History of parameter values.

<sup>4)</sup> S: Sample size, P: Power, O: Others.



As displayed in Table 1, some programs support a wide range of the selected designs, while others are more or less restricted in their scope, but almost all programs are capable of calculating sample size for standard situations, in particular the comparisons of means and proportions.

Overall functionality of the programs is briefly described in Table 2. The following information is tabulated: language, type of user interface (c.g. command dialogue or parameter input via input-masks), availability of on-line help and protocol facility, and capability for generation of a design summary report after estimating the sample size.

Further, program facilities for output processing, graphics and report generation, provision of default values for input parameters, and storage of parameter sets were examined. Other items in Table 2 provide information on the design parameters that can be determined, on literature references pertaining to statistical tests underlying the designs examined in more detail, and on other features (one- or two-sided tests, unbalanced designs).

The following programs permit a graphical representation of the required sample size for given power or effect size: GPOWER, NQUERY, POWERPACK and SOLO-POWER. A valuable feature is the option of generating tables of required sample sizes over a given parameter range, as included in NQUERY, PLANUNG, SOLO-POWER, STAT-POWER and STPLAN.

### Command-line oriented DOS-Programs

The following programs are command-line oriented: PC-SIZE, PLANUNG, POWER, SMPLSIZE, SSIZE and STPLAN. With the exception of SMPLSIZE, all are available as either public domain or shareware.

PC-SIZE allows the determination of sample size or power for both hypothesis testing and confidence interval estimation. During parameter input, previously entered values are displayed for user reference. For all supported designs, default parameter values are provided. Calculations not only include the requested sample size but also the power values corresponding to some of the next smaller sample sizes. Results can be simultaneously displayed on the screen and stored into an ASCII-file containing additional information. No references regarding the statistical algorithms used for calculations are given. Figure 1 shows a typical screen display of PC-SIZE.

In addition to sample size and power, PLANUNG can calculate any design parameters, e.g. the minimum difference between means, given all other parameters. For each individual computation, parameter values deviating from the defaults need to be re-entered, as no user input is stored by the program. Occasionally, the input disappears from the screen and can only be restored by scrolling. For some test designs, the program facilitates the creation of a table of sample sizes over a given range of values for the input parameters, which can be printed.

POWER is one of the few DOS programs offering specific on-line help in addition to an interactive reference to the algorithms implemented. The defaults are not well chosen (e.g. significance level  $\alpha = 2.0$ , power  $1 - \beta = 0$ ). Sample size, power, and the minimum detectable difference for mean comparison can be calculated.

SSIZE begins with a list of supported designs for selection. In most cases, a reference is given. Default values are reasonable and may be changed by the user. For given power, the sample size is computed, and some sample sizes corresponding to smaller values for power are displayed.

STPLAN allows the calculation of sample size, power, and other design. Attention must be given to the order in which parameters are entered: A question mark is used to indicate the parameter of interest, and a table of sample sizes can be generated by enter-

```

This program operates in seven modes :
  1 - single factor design
  2 - two factor design
  3 - randomized blocks design
  4 - paired t-test
  5 - generic F
  6 - correlation coefficient
  7 - proportions
Choose a mode [ 1 ] :

Enter the level of the test [ .050 ] : .01
Enter the required power [ .800 ] : .9
Enter the number of groups [ 2 ] :

There are three ways to specify the alternative :
  1 - Specific individual effects.
  2 - Specify a range throughout which group
      means are uniformly distributed.
  3 - Specify average squared effect divided
      by the error variance.
Choose a method [ 2 ] :

```

Figure 1: Screen display of PC-SIZE

ing a 't' in a specified location. There are no specific error messages, and in case of an error, all values need to be re-entered. A report containing a list of input values and results can be generated.

### Screen-oriented DOS-Programs

EX-SAMPLE, GPOWER, N, NSURV, SOLO-POWER, and STAT-POWER are all DOS programs with screen-oriented input masks. All required parameters can be seen on one screen. All of these programs feature on-line help.

The input screen of SOLO-POWER is overloaded with options (see Figure 2).

GPOWER is based on the book by Cohen (1982) and covers many designs where sample size calculations make use of F-distributions. It allows the graphical representation of the results, and a protocol of the calculations performed can be requested.

For calculations with EX-SAMPLE, the underlying formula and intermediate results are displayed upon entering parameter values. References are given, and the algorithms are well documented.

Both N and NSURV address only a small range of designs, however, the designs are well explained, permitting an easy choice of the required settings. N is restricted to the calculation of sample size and power for two group designs, NSURV for designs in survival analysis.

STAT-POWER (and DESIGN-POWER, the previous release) provides a wide range of designs. Design selection is made easy using a set of menus. All necessary values can be entered in one column which can be copied for repeated calculations using similar sets of parameter values.

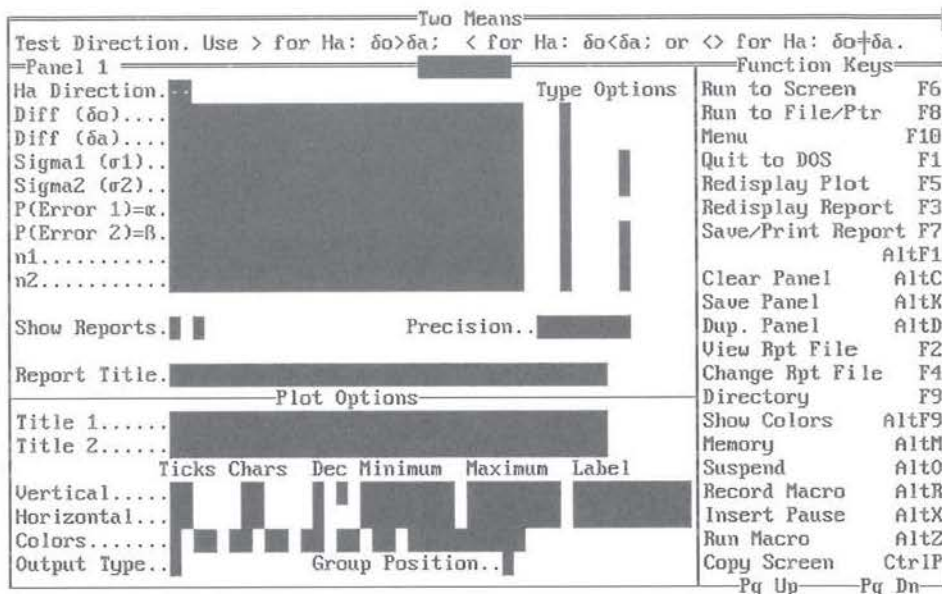


Figure 2: Screen display of SOLO-POWER/PASS

### Windows-based Programs

The only Windows-based programs with typical Windows features like cut and paste, printing support etc., are CADEMO and NQUERY.

CADEMO was originally developed under DOS, but a Windows version is also available. It provides a dictionary explaining the most important statistical terms and a bibliography containing the articles and books upon which the algorithms are based.

NQUERY (see Figure 3) supports the computation of tables for different parameter sets and graphical representation. Whenever a new design is selected, the program requires some time for hard disk read- and write-operations.

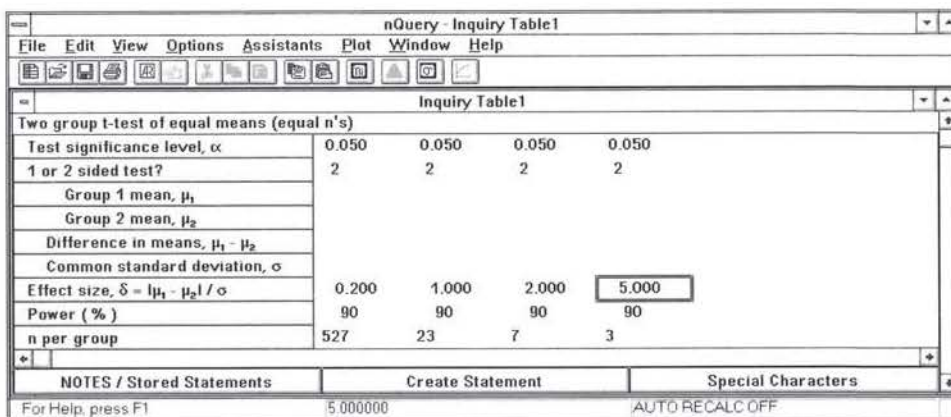


Figure 3: Screen display of NQUERY

### Others Programs

The IFNS Sharp PC-1270 sample size calculator is the only programmable pocket calculator evaluated. Eight function keys enable design selection and the activation of hidden dialogues. Program functions are menu-driven, but input requires some practice, and reference to the manual is inevitable. In addition to sample size calculations, statistical tests can be performed.

Both INSTAT and DATASIM are programs that were not primarily developed for the purpose of sample size determination, but rather for statistical theory and evaluation (INSTAT) and simulation of statistical models and verification of test procedures under a variety of designs (DATASIM). Thus, their functionality with respect to sample size calculation is limited. They are command-line oriented programs running under DOS, and each has subroutines for the calculation of the required sample size (*SIZ*, *NSAMP*) and power (*POWER*). The programs are executed via command-line function calls and parameter input. Using the command "help function name", a list of all required parameters is displayed, along with their default values and results derived thereof. For a two-sided comparison of two means of normally distributed data with effect size 1, with a power of 0.9, the commands to obtain the minimum total sample size of 46 are 'siz; nor; ndf 1;' (INSTAT) and 'nsam t 2 0.05 0.9 1' (DATASIM). Due to the main purposes of these programs, their capabilities include the calculation of probabilities and quantiles for common distribution functions.

POWERPACK is more of a programming language than an application program. As in SAS or S-Plus, distribution functions and other statistical functions are available for the creation of individual routines for sample size determination, by programming given algorithms or formulae. Some routines for commonly used queries are provided as examples, assisting in learning the language's syntax.

POWER<sub>T</sub> was developed exclusively for theoretical purposes. Test statistics and required sample sizes are calculated only for dependent and independent t-test.

Source code offered in journal publications is often available via email or the internet upon request to the publication's author. Appendix B contains a list of programs available only as source code, which must be compiled appropriately. Due to space restrictions, program descriptions and the tests covered are not included. Some of the code was used by the authors as reference code in the comparison study.

### Documentation of the Programs

The following aspects regarding program documentation are examined, and summarized in Table 3:

- a) physical format of manual (printed manual, text file to be printed),
- b) availability of on-line help,
- c) documentation of algorithms implemented, or literature reference,
- d) documentation of statistical theory for hypotheses, tests, an related sample size calculations,
- e) availability of examples

The presence of examples allows the user to understand program usage more rapidly, and permits control of proper program usage through comparison of calculated and published results.

Table 3: Documentation of the programs

Program	Documentation		Language		Algorithms	Statistical Theory	Examples	Bibliography
	Paper	Disc	English	German				
CADEMO	•		•	•	•	•		•
DATASIM	•		•					
EX-SAMPLE	•		•			•	•	•
GPOWER	•	•	•	•	•	•		•
IFNS	•			•		•	•	
INSTAT-PC		•	•				•	
N/NSURV	•			•	•	•	•	•
NQUERY	•		•		•	•	•	•
PC-SIZE		•	•			•	•	•
PLANUNG	•			•	•	•	•	•
POWER		•	•					•
POWER-PACK	•		•		•	•	•	
POWERT		•	•					
SMPLSIZE		•	•					•
SOLO- POWER	•		•		•	•	•	•
SSIZE								
STAT-POWER	•		•		•	•	•	•
STPLAN		•	•		•	•	•	•

### Sample Size Calculation

In addition to the comparison of the programs' features and their documentations, a comparison of the results for the following experimental designs was performed: Comparison of two means and two proportions with respect to difference and equivalence, and survival analysis. For the comparison of means and proportions, sample sizes were calculated for the following significance levels:  $\alpha = 1\%$ ,  $5\%$  and  $10\%$ . The following values for power were investigated:  $1 - \beta = 80\%$  and  $90\%$ . Equal allocation between groups was assumed. The tables show the sample sizes *per group*, as calculated by the programs. With GPOWER, IFNS and INSTAT, only the total sample size (i.e. both groups) could be determined; that number was divided by two and rounded up. Calculations were performed on an COMPAQ 410C notebook computer equipped with a 486 DX processor and 8 MB RAM. For the programmable environments POWERPACK, INSTAT and DATASIM, only results obtained with routines provided with the software were included in the comparison.

### Test for Difference of Two Means

Under the assumption of two independent samples from normally distributed data with means  $\mu_1$  and  $\mu_2$  and common variance  $\sigma^2$ , the required sample size was to be determined for the following testing problem:

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_{1,\Delta} : \mu_1 - \mu_2 = \Delta \neq 0.$$

The specification of an alternative  $H_{1,\Delta}$  is required for power calculation. In the rather hypothetical event that  $\sigma_2$  is known, the so-called Gauss-test is used for analysis, and the required sample size per group can be calculated as

$$n = 2(u_{1-\alpha} + u_{1-\beta})^2 (\sigma/\Delta)^2$$

where  $u_\gamma$  denotes the  $\gamma$ -quantile of the normal distribution. If the population variance has to be estimated from the data, the common two-sample t-test situation is given, and the sample size has to be calculated using an iterative process, since the quantiles of the t-distribution depend on the sample size to be determined.

Nineteen programs allow calculation of sample size for the comparison of two independent means; three of these programs are based on the Gauss-test: EX-SAMPLE, INSTAT, and SEMPLSIZE. All other programs employ the iterative approach referred to. Three programs permit a choice between the two methods: CADEMO, PLANUNG, and SOLO-POWER.

For the comparison of the programs, the common variance was set to  $\sigma^2 = 1$ , and the following differences in means were used:  $\Delta = 0.2, 1, 2$  and  $5$ . Three programs require values for the two means as input parameters rather than their difference: EX-SAMPLE, STAT-POWER, and STPLAN. In these cases,  $\mu_1$  was set to zero and  $\mu_2$  was set  $\Delta$ . Results for  $\Delta = 1$  are given in Table 4; for the other values of  $\Delta$  see Table A.3 (Appendix A). The tables include reference values calculated using the author's SAS macros based on the respective algorithms.

All programs indicate whether the t- or Gauss-test was used, except for four programs: CADEMO LIGHT, EX-SAMPLE, INSTAT and PLANUNG.

There are slight discrepancies between the different versions of CADEMO (DOS, Windows and LIGHT version). Often the DOS version results for the t-test exceed the other versions' results by one. The Windows version provides a choice between t- and Gauss-test, while in the LIGHT version, only the t-test can be used. All tables contain the results obtained with LIGHT version.

DATASIM employs an iterative algorithm and requires long computing times for sample sizes greater than approximately 100. Results are rounded, which may lead to a decrease in power below the required level.

The routine NORMAL.PWK supplied with POWERPACK is designed for the one-sample t-test; results were multiplied by two for comparison with the other programs and for presentation in the table.

SSIZE also employs an iterative algorithm. The power for a given sample size is calculated and then compared to the required power.

STPLAN reports the sample size with up to three decimal places, however, for the comparison, they were rounded to one decimal place.

Apart from differences due to rounding, discrepancies in calculated sample sizes are minor, with results varying by not more than two over the range of parameter sets investigated (Tables 4 and A.3). For large sample sizes, this difference is of no practical relevance, but sample size calculation using normal approximation for small sample sizes may result in power considerably below that required. For all designs considered in the comparison, the loss in power does not exceed 0.01.

Table 4: Two-sided test for difference of two means (sample sizes per group, effect size  $\Delta/\sigma = 1$ )

Program	$\alpha/1 - \beta$					
	0.1/0.8	0.1/0.9	0.05/0.8	0.05/0.9	0.01/0.8	0.01/0.9
t-test						
Reference values <sup>1)</sup>	13.2	18.0	16.8	22.1	25.1	31.4
CADEMO LIGHT	14	18	17	23	25	32
DATASIM	13	18	17	22	25	32
GPOWER	14	18	17	23	26	32
IFNS	14	18	17	22	26	32
N	14	18	17	23	26	32
NQUERY	14	18	17	23	26	32
PC-SIZE	14	18	17	23	26	32
PLANUNG	14	18	17	23	26	32
POWER	13	18	17	22	25	32
POWERPACK	14	18	16	22	24	30
POWERT	13	18	17	22	24	31
SOLO-POWER	14	18	17	22	25	32
Ssize	14	18	17	23	25	32
STAT-POWER	13.1	17.8	16.7	22.0	25.1	31.5
STPLAN	13.2	18.0	16.7	22.0	24.9	31.3
Gauss-test						
Reference values <sup>1)</sup>	12.4	17.1	15.7	21.0	23.4	29.8
EX-SAMPLE	12.5	17.0	15.5	21.0	23.5	30.0
INSTAT	12.4	17.2	15.7	21.0	23.4	29.8
PLANUNG	13	18	16	22	24	30
SMPLSIZE	13	18	16	21	24	30
SOLO-POWER	13	18	16	22	24	30

1) Calculated using the author's SAS macros.

### Test for Equivalence of Two Means

Under the assumption of two independent samples from normally distributed data with means  $\mu_1$  and  $\mu_2$  and common variance  $\sigma^2$ , the required sample size was to be determined for the following testing problem:

$$H_0: |\mu_1 - \mu_2| > \delta \quad \text{vs.} \quad H_{1,\Delta}: |\mu_1 - \mu_2| \leq \delta \quad |\Delta| < \delta$$

The null hypothesis  $H_0$  can be tested controlling a level of  $\alpha$  by performing two one-sided t-tests at level  $\alpha$  for the hypotheses  $H_{0>}: \mu_1 - \mu_2 > \delta$  and  $H_{0<}: \mu_1 - \mu_2 < -\delta$ . Then,  $H_0$  is rejected if both  $H_{0>}$  and  $H_{0<}$  are rejected.

For the comparison, the following parameter values were used:  $\sigma^2 = 1$ , and  $\delta = 0.2, 1, 2$  and  $5$ . This choice of equivalence boundaries corresponds to the choice of effect sizes in the discussion on differences of means above.

The program N is restricted to the case  $\Delta = 0$  (equality of means), and the comparison was restricted accordingly. The results are summarized in Table 5.

Three programs allow sample size calculation for the situation described above: IFNS, N, and PLANUNG. In all of them, the algorithm presented PATEL and GUPTA (1984) is implemented. However, caution is indicated in that for certain situations, the actual significant-

Table 5: Two-sided test for equivalence of two means (sample sizes per group, equivalence range  $\delta$ ,  $\sigma = 1$ ,  $H_{1,\Delta} : \Delta = \mu_1 - \mu_2 = 0$ )

Program	$\delta$	$\alpha/1 - \beta$					
		0.1/0.8	0.1/0.9	0.05/0.8	0.05/0.9	0.01/0.8	0.01/0.9
Reference values <sup>1)</sup>	5	2	2	2	3	2	3
	2	4	6	5	7	7	9
	1	14	18	18	23	27	33
	0.2	329	429	429	542	651	790
IFNS	5	—	—	—	—	4	4
	2	5	6	6	7	9	10
	1	15	19	19	23	29	34
	0.2	330	430	430	543	653	791
N	5	2	2	2	3	2	3
	2	4	6	5	7	8	9
	1	14	18	18	23	27	33
	0.2	329	429	429	542	652	790
PLANUNG <sup>2)</sup>	5	2	2	2	3	2	3
	2	4	6	5	7	8	9
	1	14	18	18	23	27	33
	0.2	329	429	429	542	652	—

<sup>1)</sup> Calculated using SAS macro EQUATL.SAS by M. Wargenau (see Appendix B) based on PATEL and GUPTA (1984)

<sup>2)</sup> Calculation was performed using the dialogue "one-sided" in order to get comparable results.

ce level may exceed  $\alpha$  (FRICK 1991). Until now, no correct method for sample size determination has been published for the approach using two one-sided t-tests as indicated above. Therefore, the reference values given in Table 5 are also based on the work of PATEL and GUPTA. It is expected that the sample size necessary to obtain the required power using the t-test approach is slightly higher than the values given in Table 5, as the algorithm by PATEL and GUPTA is anticonservative.

For the range of parameters investigated, both N and PLANUNG yielded identical results. The maximum deviation from IFNS was two, and the power difference to the reference value is not greater than 0.01.

### Test for Difference of Two Proportions

A situation frequently encountered in data analysis is the comparison of rates for a binary outcome variable, such as 'treatment success', between two groups. Let  $\pi_1$  and  $\pi_2$  be the probabilities of 'success' for the respective groups. The following testing problem is considered:

$$H_0: \pi_2 - \pi_1 = 0 \quad \text{versus} \quad H_{1,\Delta}(\pi_1): \pi_2 - \pi_1 = \Delta \neq 0.$$

Not only the difference of the rates  $\Delta = \pi_2 - \pi_1$  has to be fixed, but also one of the success probabilities, since the underlying random variable's variance is dependent on the location of the rates. For the comparison, the following parameters were used:  $\pi_1 = 0.5$  and  $0.9$ , and  $\Delta = 0.05$  and  $0.1$ .

The standard methods for the analysis of the corresponding  $2 \times 2$ -table are Fisher's exact test and the  $\chi^2$ -test. Despite the continuing debate about these tests, an adequate sample size should be calculated based on the test designated for analysis. For the  $\chi^2$ -test, the arcsine approximation is appropriate, and for Fisher's exact test, the algorithm documen-



ted in CASAGRANDE, PIKE and SMITH (1978) is used. The uncorrected version of the formula in CASAGRANDE et. al. is also used for the calculation of sample size calculation for the  $\chi^2$ -test.

Of the 14 programs supporting the design, five are based on Fisher's exact test: CADEMO LIGHT, IFNS, N, PC-SIZE and PLANUNG. Five programs are based on the  $\chi^2$ -test: INSTAT and SMPLSIZE use the uncorrected formula given in CASAGRANDE et. al., and EX-SAMPLE, SOLO-POWER, and STAT-POWER use the arcsine approximation; the other four support both methods, with NQUERY, POWER, and SSIZE using the uncorrected formula in CASAGRANDE et. al., and STPLAN using arcsine approximation. However, only six programs explicitly refer to the test for which the sample size is calculated: IFNS, N, NQUERY, POWER, SOLO-POWER, and STPLAN.

The results of the comparison are shown in Table 6 for the  $\chi^2$ -test and in Table 7 for Fisher's exact test. Reference values for Fisher's exact test were taken from FLEISS (1981), reference values for the  $\chi^2$ -test were determined by the authors with a SAS implementation of the formulae from the literature.

Within each group, all corresponding sample sizes are equal, except for some minor differences caused by the various rounding mechanisms, and resulting in deviations from the prescribed power by less than 0.01. Due to the conservativeness of Fisher's exact test, the sample sizes are always higher than those calculated with a  $\chi^2$  approach.

Table 6: Two-sided test for difference of two proportions (sample sizes per group,  $H_{1,\Delta} : \Delta = \pi_2 - \pi_1$ )

Program	$\pi_1$	$\Delta$	$\alpha/1 - \beta$					
			0.1 / 0.8	0.1 / 0.9	0.05 / 0.8	0.05 / 0.9	0.01 / 0.8	0.01 / 0.9
$\chi^2$ -test – arcsine approximation								
Reference values <sup>1)</sup>	0.5	0.1	305.0	422.4	387.2	518.3	576.1	734.0
	0.9	0.05	333.8	462.3	423.3	567.3	630.5	803.3
EX-SAMPLE	0.5	0.1	305	422	387	518	576	734
	0.9	0.05	334	462	424	567	631	803
SOLO-POWER	0.5	0.1	305	423	388	519	577	734
	0.9	0.05	334	463	424	568	631	804
STAT-POWER	0.5	0.1	305.0	422.4	387.2	518.3	576.1	734.0
	0.9	0.05	333.8	462.3	423.3	567.3	630.5	803.3
STPLAN	0.5	0.1	305.0	422.4	387.2	518.3	576.1	734.0
	0.9	0.05	333.8	462.3	423.7	567.3	630.5	803.3
$\chi^2$ -test – uncorrected Casagrande et. al. formula								
Reference values <sup>1)</sup>	0.5	0.1	305.0	422.0	387.3	518.0	576.7	734.1
	0.9	0.05	342.1	473.4	434.4	581.1	646.7	823.3
INSTAT	0.5	0.1	305.0	422.1	387.4	518.1	576.7	734.1
	0.9	0.05	342.1	473.5	434.5	581.1	646.8	823.4
NQUERY	0.5	0.1	305	423	388	519	577	735
	0.9	0.05	343	474	435	582	647	824
POWER	0.5	0.1	305	422	387	518	577	734
	0.9	0.05	342	474	435	581	647	824
SMPLSIZE	0.5	0.1	305	422	387	518	577	734
	0.9	0.05	342	473	434	581	647	824
SSIZE	0.5	0.1	305	423	388	519	577	735
	0.9	0.05	343	474	435	582	647	824

<sup>1)</sup> Calculated using the authors' SAS macros.

Table 7: Two-sided test for difference of two proportions (sample sizes per group,  $H_{1,\Delta} : \Delta = \pi_2 - \pi_1$ )

Program	$\pi_1$	$\Delta$	$\alpha/1-\beta$					
			0.1 / 0.8	0.1 / 0.9	0.05 / 0.8	0.05 / 0.9	0.01 / 0.8	0.01 / 0.9
Fisher's exact test								
Reference values <sup>1)</sup>	0.5	0.1	325	442	407	538	597	754
	0.9	0.05	381	513	474	621	686	863
CADEMO LIGHT	0.5	0.1	325	442	407	538	596	754
	0.9	0.05	381	513	474	620	686	863
IFNS	0.5	0.1	325	442	408	538	597	754
	0.9	0.05	382	513	474	621	687	863
N	0.5	0.1	325	442	408	538	597	754
	0.9	0.05	580	787	726	957	1061	1340
	0.6	-0.1	325	442	408	538	597	754
NQUERY	0.95	-0.05	382	513	474	621	687	863
	0.5	0.1	325	442	408	538	597	754
	0.9	0.05	382	513	474	621	687	863
PC-SIZE	0.5	0.1	325	442	408	538	597	754
	0.9	0.05	382	513	474	621	687	863
PLANUNG	0.5	0.1	325	442	408	538	597	754
	0.9	0.05	382	513	474	621	687	863
POWER	0.5	0.1	325	442	407	538	597	754
	0.9	0.05	381	513	474	620	686	864
SSIZE	0.5	0.1	325	442	408	538	597	754
	0.9	0.05	382	513	474	621	687	863
STPLAN	0.5	0.1	324.7	441.8	407.1	537.9	596.5	763.9
	0.9	0.05	381.0	512.6	473.6	620.4	686.2	862.9

<sup>1)</sup> Taken from FLEISS (1981).

With the exception of N, all programs yields results symmetrical in the proportions for experimental group and control. In contrast, N computes the sample sizes for two one-sided tests with half the significance level, and takes the maximum of the two resulting values. Therefore, the difference  $\Delta$  between the success probabilities for control and experimental group is taken to be the distance from the control group probability. In the case of  $\pi_1 = 0.5$  and  $\pi_2 = 0.6$ , the calculated sample sized do not deviate since the binomial distribution is symmetrical about 0.5, and the situation is identical to the case  $\pi_1 = 0.5$  and  $\pi_2 = 0.4$  (see Table 7).

### Test for Equivalence of Two Proportions

For testing equivalence of two proportions  $\pi_1$  and  $\pi_2$ , the one-sided case is investigated. In clinical trials, this relates to the concept of therapeutic equivalence. The testing problem is:

$$H_0 : \pi_2 - \pi_1 > \delta \quad \text{versus} \quad H_{1,\Delta}(\pi_1) : \pi_2 - \pi_1 = \Delta \leq \delta,$$

where  $\delta > 0$  is an upper bound on the range of equivalence. A variety of tests and corresponding methods of sample size calculation has been proposed. For a review of these methods see ROEBRUCK et al. (1997). Five programs support sample size determination for testing the one-sided equivalence of proportions: NQUERY, PLANUNG, SOLO-POWER, and SSIZE employ an algorithm described in BLACKWELDER and CHANG (1982),

Table 8: One-sided test for equivalence of two proportions (sample sizes per group, equivalence range  $\delta$ ,  $H_{1,\Delta}(\pi_1) : \Delta = \pi_2 - \pi_1 = 0$ )

Program	$\pi_1$	$\delta$	$\alpha/1 - \beta$					
			0.1/0.8	0.1/0.9	0.05/0.8	0.05/0.9	0.01/0.8	0.01/0.9
Reference values <sup>1)</sup>	0.5	0.1	226	329	310	429	502	651
	0.9	0.05	325	474	446	617	723	938
NQUERY <sup>2)</sup>	0.5	0.1	221	322	303	420	492	638
	0.9	0.05	248	362	341	472	552	716
PLANUNG	0.5	0.1	226	329	310	429	502	651
	0.9	0.05	325	473	446	617	723	938
SOLO-POWER	0.5	0.1	226	329	310	429	502	651
	0.9	0.05	325	474	446	617	723	938
SSIZE	0.5	0.1	226	329	310	429	502	651
	0.9	0.05	325	474	446	617	723	938
N	0.5	0.1	242	344	324	442	513	660
	0.9	0.05	288	402	381	513	593	758

<sup>1)</sup> Calculated using the authors' SAS macros, using the algorithm in Blackwelder and Chang (1982)

<sup>2)</sup> Sample sizes set in italics would produce power values less than 0.775 or 0.875, respectively.

and N employs an algorithm described in FRICK (1994). For the comparison, the following parameters were used:  $\pi_1 = 0.5$  and  $0.9$ ,  $\Delta = 0$  and  $\delta = 0.05$  and  $0.1$ . The results are summarised in Table 8.

N and SOLO-POWER explicitly name the test for which the sample size is calculated, and NQUERY and SOLO-POWER are the only programs that allow parameter values of  $\Delta \neq 0$ . When calculating sample sizes as above with  $\Delta = \delta/2$ , both programs yield the same correct results.

The sample sizes calculated by SOLO-POWER and SSIZE are identical to the reference values; those by PLANUNG differ by one in the case  $\alpha/1 - \beta = 0.1/0.9$  and  $\pi_1 = 0.9$ ,  $\Delta = 0.05$ . NQUERY computes lower values, with major differences for  $\pi_1 = 0.9$ ,  $\Delta = 0.05$  and  $\pi_1 = 0.05$ ,  $\delta = 0.1$ . In the former case, the power requirements of 80% and 90% could not be met. In contrast, N computes higher values for  $\pi_1 = 0.5$ ,  $\delta = 0.1$  and lower values for  $\pi_1 = 0.9$ ,  $\delta = 0.05$ , compared to the reference values based on the method by BLACKWELDER as indicated in FRICK (1994).

### Survival Analysis

For the comparison of sample size programs for survival trials, the programs are divided into two groups, according to the assumptions regarding the distribution of the survival times. The first group, containing the majority of the programs, requires exponential survival: EX-SAMPLE, NSURV, PLANUNG, POWER, SOLO-POWER, and STPLAN. The second group allows calculations of sample sizes under the proportional hazards assumptions. SSIZE is a representative of this group. NQUERY supports both assumptions.

Algorithms implemented in programs of the first group are derived from GEORGE and DESU (1974) for PLANUNG and STPLAN, BERNSTEIN and LAGAKOS (1978) for PLANUNG, RUBINSTEIN et al. (1982) for NQUERY, SCHOENFELD and RICHTER (1982) for EX-SAMPLE and POWER, and LACHIN and FOULKES (1986) for NSURV and SOLO-POWER. In the second group, the algorithms implemented are derived from FREEDMAN (1982) and SCHOENFELD (1981) for NQUERY and SSIZE. For more information on the algorithms see OELRICH et al. (1997).

A difficulty impeding a true comparison is the variety of parameters required for the various algorithms, such as median or mean survival times, hazard rates or survival rates. A valuable utility in NQUERY permits the transformation of one set of parameters into another. STPLAN allows input parameter choice, and SOLO-POWER has a feature to transform hazard rate  $\lambda$  into a survival probability using the exponential distribution assumption. In all other programs, that transformation has to be made manually, if necessary.

For the comparison of sample size calculations for survival analysis, the same set of parameter values was used as in the first part of the simulation study in LAKATOS and

Table 9: Parameter values used for the comparison of programs for sample size determination in survival analysis

Total Study Time T	10	10	10	10	10	10
Proportion $\pi_C$ at time T	0.8000	0.8000	0.8000	0.2000	0.2000	0.2000
Median survival $\text{med}_C$	31.0630	31.0630	31.0630	4.3070	4.3070	4.3070
Exponential parameter $\lambda_C$	0.0223	0.0223	0.0223	0.1609	0.1609	0.1609
Proportion $\pi_E$ at time T	0.8620	0.8940	0.9460	0.3420	0.4470	0.6690
Median survival $\text{med}_E$	46.5710	62.1260	124.2510	6.4570	8.6140	17.2270
Exponential parameter $\lambda_E$	0.0149	0.0112	0.0056	0.1073	0.0805	0.0402
Total Study Time T	9.5	9.5	9.5	9.5	9.5	9.5
Proportion $\pi_C$ at time T	0.809	0.809	0.809	0.217	0.217	0.217
Median survival $\text{med}_C$	31.083	31.083	31.083	4.308	4.308	4.308
Exponential parameter $\lambda_C$	0.0223	0.0223	0.0223	0.1609	0.1609	0.1609
Proportion $\pi_E$ at time T	0.868	0.899	0.948	0.361	0.465	0.683
Median survival $\text{med}_E$	46.520	61.888	123.776	6.460	8.611	17.242
Exponential parameter $\lambda_E$	0.0149	0.0112	0.0056	0.1073	0.0805	0.0402
Total Study Time T	7.5	7.5	7.5	7.5	7.5	7.5
Proportion $\pi_C$ at time T	0.846	0.846	0.846	0.299	0.299	0.299
Median survival $\text{med}_C$	31.083	31.083	31.083	4.308	4.308	4.308
Exponential parameter $\lambda_C$	0.0223	0.0223	0.0223	0.1609	0.1609	0.1609
Proportion $\pi_E$ at time T	0.894	0.919	0.959	0.447	0.547	0.740
Median survival $\text{med}_E$	46.520	61.888	123.776	6.460	8.611	17.242
Exponential parameter $\lambda_E$	0.0149	0.0112	0.0056	0.1073	0.0805	0.0402
Total Study Time T	5.5	5.5	5.5	5.5	5.5	5.5
Proportion $\pi_C$ at time T	0.885	0.885	0.885	0.413	0.413	0.413
Median survival $\text{med}_C$	31.083	31.083	31.083	4.308	4.308	4.308
Exponential parameter $\lambda_C$	0.0223	0.0223	0.0223	0.1609	0.1609	0.1609
Proportion $\pi_E$ at time T	0.921	0.940	0.970	0.554	0.642	0.802
Median survival $\text{med}_E$	46.520	61.888	123.776	6.460	8.611	17.242
Exponential parameter $\lambda_E$	0.0149	0.0112	0.0056	0.1073	0.0805	0.0402
Hazard ratio $\delta = \ln(\pi_C)/\ln(\pi_E)$ = $\text{med}_C/\text{med}_E = \lambda_E/\lambda_C$ $\Delta = 1/\delta$	2/3	1/2	1/4	2/3	1/2	1/4
	1.5000	2.0000	4.0000	1.5000	2.0000	4.0000

LAN (1992, Table 1), where an exponential model with the following restrictions was used: a balanced design (1:1 randomization), uniformly distributed accrual times, identical censoring probabilities, and no loss to follow-up. Further parameters and their values are:

- a) survival probability at the end of the trial for the control group:  $\pi_C(T) = 0.2$  and  $0.8$
- b) hazard ratio  $\delta = \lambda_E(T)/\lambda_C(T) = 2/3, 1/2$  and  $1/4$ ,
- c) total study length  $T = 10$  [years],
- d) length of accrual period  $A = 1, 5$  and  $9$  [years], and
- e) length of follow-up period  $F = 9, 5$  and  $1$  [years].

In all cases two-sided tests with a significance level of  $\alpha = 0.05$  and a power of  $0.9$  were used. Table 9 gives the sets of transformed parameter values: proportion  $\pi(T) = \exp(-\lambda T)$ , median survival time  $\text{med} = \ln(2/\lambda)$  and hazard rate  $\lambda$ , for the control and treatment groups, indexed by C and E, respectively, at the times T.

All programs designed for sample size determination for survival analysis calculate the sample sizes for the logrank test. With the exception of POWER, both one- and two-sided versions of the test are provided; POWER only supports the two-sided test.

Where the algorithms are based on FREEDMAN (1982) or SCHOENFELD (1981), no accrual and follow-up times are required as input parameters. The survival probability needs to be calculated for the study mid-point  $T - A/2 (= F + A/2)$  under the assumption of uniformly distributed accrual times, rather than at study end T, in order to obtain appropriate sample sizes in Table 10. For the corresponding hazard rates, see Table 9 at times 9.5, 7.5, and 5.5. All other programs take the lengths of accrual and follow-up periods directly into account.

Table 10: Survival analysis (sample sizes per group)<sup>1)</sup>

$\pi_C(T)$	$\delta$	A	Exponential distribution							Proportional hazards			
			Ref. val. <sup>2)</sup>	PLA-NUNG <sup>3)</sup>	NSURV	SOLO-POWER	NQUE-RY <sup>4)</sup>	POWER	STPLAN	NQUERY <sup>5)</sup>	Ssize <sup>6)</sup>	Ssize <sup>7)</sup>	
0.8	2/3	1	803	726.9	804	806	821	822	819.1	821	799	824	
		5	985	904.9	1003	1005	1024	1025	1021.3	1036	1012	1039	
		9	1347	1220.8	1355	1357	1383	1385	1379.3	1425	1392	1428	
	1/2	1	318	267.4	312	327	334	332	341.7	328	305	329	
		5	389	333.3	389	407	418	415	426.8	414	388	418	
		9	538	450.0	526	550	565	562	576.9	560	520	560	
	1/4	1	107	83.9	106	119	135	135	143.3	122	95	124	
		5	132	104.9	133	148	170	169	179.5	150	113	154	
		9	184	141.9	179	200	230	230	243	201	152	207	
	0.2	2/3	1	183	170.4	180	185	182	182	184.2	185	181	186
			5	203	194.0	207	211	210	209	212	210	205	211
			9	267	245.7	264	269	286	267	270.5	257	250	257
1/2		1	68	60.8	67	71	69	68	71.8	73 (71)	68 (66)	73 (71)	
		5	79	69.7	77	82	81	80	83.5	82	77	83	
		9	102	88.6	99	105	104	103	107.6	101	95	102	
1/4		1	22	17.7	21	26	25	24	27.8	27 (26)	22 (20)	28 (27)	
		5	25	20.5	24	30	29	29	33	39	23	32	
		9	33	26.4	32	38	39	38	43.2	38	29	39	

<sup>1)</sup> Two-sided,  $\alpha = 0.05, 1 - \beta = 0.90$ , total study length  $T = 10$  years, follow-up time =  $10 - \text{accrual length (A)}$ ,  $\pi_C(T)$  is the survival rate of the control group,  $\delta$  is the hazard ratio.

<sup>2)</sup> Calculated using sample sizes of Table 1 and formula on page 183 of Lakatos and Lan (1992) and therefore only valid for underlying exponential distribution.

<sup>3)</sup> Sample sizes set in italics would produce power values differing in more than 2.5% to the required 90% of the calculation.

<sup>4)</sup> exponential survival, <sup>5)</sup> Logrank test, <sup>6)</sup> Schoenfeld, <sup>7)</sup> Freedman.

The results are displayed in Table 10. The total sample sizes resulting from SOLO-POWER and the accrual rates resulting from STPLAN are transformed to correspond to sample sizes per group. With NQUERY, sample size based on exponential survival was calculated assuming a maximum follow-up period of 10 years. With NSURV, the settings for the median of the treatment group (Median Test), the median of the control group (Median Standard), and symmetry in log differences were selected, implying that a test for  $\delta$  will require the same sample size as a test for  $\Delta = 1/\delta$ . The median survival time was selected with POWER, and with PLANUNG, selected parameters were median survival of the control group and  $\Delta$  for the required parameter Delta.

Sample sizes calculated with PLANUNG are smaller than required if assessed in light of the power values determined in LAKATAOS and LAN (1992). Sample sizes calculated with NSURV always yield the required power. The other programs, SOLO-POWER, NQUERY, POWER and STPLAN, calculate higher sample sizes, which results for small hazard ratios in a too high power (power > 0.925). Apparently, for the assumptions of proportional hazards, the sample sizes calculated with SSIZE according to SCHOENFELD (1981) are always smaller than those calculated according to FREEDMAN (1982). EX-SAMPLE also supports sample size determination for survival analysis, however, the authors did not succeed in this application.

In addition to the programs assuming exponentially distributed survival times in the above comparison, STPLAN offers two design variants: a comparison with a historical control group as described in DIXON and SIMON (1988), and a two-step hazard function with a given breakpoint (SCHOENFELD, 1981). Due to the incompatibility with the designs supported by the other programs, these variants were not used in the comparison. NSURV also permits sample size calculation for hypotheses of equivalence in survival analysis.

### Acknowledgements

This work was originated in the authors' participation in the Ad Hoc Working Group on Sample Size Determination, a subdivision of the Working Group of Biostatisticians in Pharmaceutical Research within the German Region of the International Biometric Society. The authors thank all members of working group for the many helpful discussion, Christoph Edel for carefully reading the manuscript, and Manfred Wargenau for calculation of the reference values in Table 5.

### Literature

The following literature on software for sample size calculation is divided into three groups: (a) summary articles comparing programs, (b) article's by programmers, and reviews of individual programs, and (c) articles containing program source codes and an evaluation.

### Summary Articles

- GOLDSTEIN, R.: Power and sample size via MS/PC-DOS computers. *The American Statistician*, 1989, Vol. 43, No. 4, pp. 253–260.
- SAYN, H, and MERKLE, W.: Statistical software for sample size estimation — Power, Design-Power and IFNS. *Statistical Software Newsletter*, 1989, Vol. 15, No. 2, pp. 56–59.
- WOODWARD, M.: A computer-based method for determining optimal size with medical applications. *The Statistician*, 1989, Vol. 38, pp. 221–226.

## Reviews and Articles by Programmers

- BAILEY, T.: Statistical Software Reviews: INSTAT. *Applied Statistics*, **37**, pp. 273–277, 1988.
- DALLAL, G. E.: PC-SIZE: A Program for Sample Size Determinations. *The American Statistician*, **40**, p. 52–52, 1986.
- DALLAL, G. E.: PC-SIZE: Consultant – A Program for Sample Size Determinations. *The American Statistician*, **44**, p. 243, 1990.
- DUPONT, W. D.; PLUMMER, WALTON D.: Power and sample size calculations – A Review and Computer Program. *Controlled Clinical Trials*, **11**, pp. 116–128, 1990.
- EMERSON, S.: Statistical Packager for Group Sequential Methods. pp. 183–192. *The American Statistician*, **50**, 1996.
- ENAS, N.: PASS: A quick and easy power and sample-size calculation program. *American Statistician*; **42**, pp 229–229, 1988.
- ERFELDER, E.; FAUL, F. & BUCHNER, A. (1996): GPOWER: A general power analysis program. *Behavioral Research Methods, Instruments, & Computers*, **28**(1), 1–11.
- GIANI, G., STRASSBURGER, K.: Selektionsverfahren bei normalverteilten Beobachtungen und ihre Realisierung in SEPARATE. Aus: Internationale Biometrische Gesellschaft, Deutsche Region, Abstracts zum 42. Biometrischen Kolloquium, 12.-15. 03. 1996 in Magdeburg.
- GRAF, R. G. and ALF, Jr.; E. F.: BASIC Programs to Determine Sample Size and Exact Power in ANOVA Designs. *EdPsMeas*, **50**, p. 117–121, 1990.
- HEISELBETZ, C.; EDLER, L.: A sample size program for 'Proving the null hypothesis'. *Controlled Clinical Trials*, **8**, 45–48, 1987.
- LATOUR, K.: POWER: A Simple Macro for Power and Sample Size Calculations. *SUGI*, **17**, p. 1173–1177, 1992.
- LENTH, R. V.: PowerPack: A microcomputer package for power and sample-size problems. *The American Statistician*, **41**, p. 239–240, 1987.
- MUHM, J. M. and OLSHAN, A. F.: A Program to Calculate Sample Size, Power, and Least Detectable Relative Risk Using a Programmable Calculator. *American Journal of Epidemiology*, **129**, p. 205–211, 1989.
- NATARAJAN, R. TURNBULL, B.W., SLATE, E.H. and CLARK, L.C.: A computer program for sample size and power calculations in the design of multi-arm and factorial clinical trials with survival time endpoints. *Computer Methods and Programs in Biomedicine*, **49**, p. 137–147, 1996.
- RAHLFS, V. W.: Nsurv – A Program for Planning and Interpretation of Studies Concerning Sample Size and Related Topics for Time-to-occurrence Data. *Applied Statistics*, Vol. **43**, pp. 285–293, 1994.
- RASCH, D.; GUIARD, V.; NÜRNBERG, G.; RUDOLPH, E. and TEUSCHER, F.: The expert system CA-DEMO – computer-aided design of experiments and modelling. *Statistical Software Newsletter*, 1987, Vol. **13**, pp. 107–114.
- SCHMIDTKE, J.: TRIQ – A PC-Program for Design and Analysis of Triangular Sequential Designs. Aus: Faulbaum, F., Bandilla, W. (Eds.): *SoftStat '95 – Advances in Statistical Software 5*, Stuttgart: Lucius & Lucius, pp. 509–515, 1996.
- SSIZE: A sample size program for clinical and epidemiologic studies. *The American Statistician*, 1991, Vol. **45**, p. 338.
- TURIEL, T. P.: A FORTRAN Program to Determine Sample Sizes and Generate Power Curves for Completely Randomized and Randomized Complete Block Designs. *Jornal of Qual. Tek.*, **21**, p. 277–286, 1989.
- Woodward, M.: Formulae for sample size, power and minimum detectable relative risk in medical studies. *The Statistician*, 1992, Vol. **41**, pp 185–196.

## Articles with Source Code

- BERGSTRALH, E.J.: SAS macros for sample size and power calculation. *SUGI* 9, pp 633–638, 1984.
- ENAS, G. G.; OFFEN, W. W. and TAULBEE, J. D.: An interactive, menu-driven SAS system for sample size and power computations in clinical trial design. *SUGI*, 10; pp 930–935, 1985.
- JEANPRÉTRE, N. and KRAFTSIK, R.: A program for the computation of power and determination of sample size in hierarchical experimental designs. *Computer Methods and Programs in Biomedicine*, **29**, pp. 179–190, 1989.

## References

- BERNSTEIN, D., LAGAKOS, S. W.: Sample size and power determination for stratified clinical trials. *Journal of Statistical Computation and Simulation* **8**, 65–73, 1978.
- BLACKWELDER, W. C., CHANG, M. A.: Sample size graphs for “Proving the null hypothesis”. *Controlled Clinical Trials* **5**, 97–105, 1984.
- BLACKWELDER, W.C.: “Proving the null hypothesis” in clinical trials. *Controlled Clinical Trials* **3**, 345–353, 1982.
- CASAGRANDE, J. T., PIKE M. C., SMITH P. G.: An improval approximative formula for calculating sample sizes for comparing two binomial distributions *Biometrics* **34**, 483–486, 1978.
- COHEN, J.: *Statistical power analysis for the behavioral sciences*. Hillsdale, New York, Erlbaum, 1988.
- DIXON, D. O., SIMON, R.: Sample size considerations for studies comparing survival curves using historical controls. *Journal of J Clin Epidemiology* **41**, 1209–1213, 1988.
- FLEISS, J. L.: *Statistical methods for rates and proportions*. Wiley & Sons, New York, 1981.
- FREEDMEN, L. S.: Tables of the number of patients required in clinical trials using the logrank test. *Statistic in Medicine* **1**, 121–129, 1982
- FRICK H.: On some properties and asymptotic results of Anderson-Hauck’s procedure and Patel-Gupta’s noncentral t-method for testing equivalence. In: Vollmar J. (Ed.): *Biometrie in der chemisch-pharmazeutischen Industrie*. Band 5, Gustav-Fischer Verlag, Stuttgart, 51–62, 1991.
- FRICK, H.: On approximate and exact sample size for equivalence tests for binomial proportions. *Biometrical Journal* **36**, 7841–854, 1994.
- GEORGE, S. L., DESU M. M.: Planning the size and duration of a linical trial studying the time to some critical event. *Journal of Chronic Disease* **27**, 15–24, 1974.
- LACHIN, J. M., FOULKES, M. A.: Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, non-compliance, and stratification. *Biometrics* **42**, 507–519, 1986.
- LAKATOS, E., LAN, K. K. G.: A comparison of sample size methods for the logrank statistic. *Statistic in Medicine* **11**, 179–191, 1992.
- OELLRICH, S., FREISCHLÄGER, F., BENNER, A., KIESER, M: Sample size determination on survival time data. *Informatik Biometrie und Epidemiologie in Medizin und Biologie*, **28**, 64–85, 1997.
- PATEL, H. I., GUPTA, G. D: A problem of equivalence in clinical trials. *Biometrical Journal* **26**, 471–474, 1984.
- ROEBRUCK, P., ELZE, M., HAUSCHKE, D., LEVERKUS, F.: Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme. *Informatik Biometrie und Epidemiologie in Medizin und Biologie*, **28**, 51–63, 1997.
- RUBINSTEIN, L. V., GAIL, M. H., SANTNER, T. J.: Planning the duration of a comparative clinical trial with losses to follow-up and a period of continued observation. *Journal of Chronic Disease* **34**, 469–479, 1981.
- SCHOENFELD, D. A.: The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* **68**, 316–319, 1981.
- SCHOENFELD, D. A., RICHTER, J. R.: Nomograms for calculating the number of patients needed for a clinical trials with survival as an endpoint. *Biometrics* **38**, 163–170, 1982.
- SILLITO, G. P.: Note on approximation to the power function of the “2 × 2 comparative trial” *Biometrika* **36**, 347–352, 1949.



**Appendix A – Listing of the Programs**

Table A.1: Software for sample size determination - Availability for the investigation

Commercial programs		Availability <sup>1)</sup>
CADEMO	CADEMO – Computer aided design of experiments and modelling/CADEMO - Ein Programm zur statistischen Versuchsplanung und Modellwahl	C/D
CADEMO LIGHT	CADEMO <i>light</i>	C
DATASIM	DATASIM	C
EAST	EaSt – A Software Package for the Design and Interim Monitoring of Group Sequential Clinical Trials	–
EGRET SIZ	EGRET SIZ Program – Advanced Tool for Power & Sample Size Estimation	–
EX-SAMPLE	EX-Sample+ – An expert system which helps determine sample size for research projects	C
IFNS	Fallzahlenrechner – Statistische Fallzahlenplanung und Signifikanztests auf dem Sharp PC-1270	C
N	N – Das Instrument zur Planung und Interpretation von Studien	C
NQUERY	nQuery – Software for planning research studies	C
NSURV	Nsurv – a tool for the planning and interpretation of studies for time-to-occurrence data	C
PASS	PASS – A quick and simple power and sample size calculation program	C
PEST	Pest – Planning of sequential studies	–
POWEREFFECT	Power & Effect: A statistical utility for Macintosh & Windows Systems	–
POWERPACK	PowerPack	C
SMPLSIZE	SMPLSIZE	C
SOLO-POWER	SOLO Power Analysis	C
STAT-POWER	STAT-POWER	C
TRIQ	TRIQ – Ein Programm zur Versuchsplanung und Auswertung von sequentiellen Dreiecksversuchen/Triangular Sequential Designs	D
Shareware and Public domain programs		
GPOWER	Gpower – A general power analysis program	C
INSTAT	INSTAT Shareware Package	C
PC-SIZE	PC-SIZE – A program for sample size determinations	C
	PC-SIZE: Consultant – A program for sample size determinations	
PLANUNG	PLANUNG – Programmpaket zur Versuchsplanung	C
POWER	POWER	C
POWERT	POWERT	C
SEPARATE	SEPARATE	C
SSIZE	SSIZE: A sample size program for clinical and epidemiological studies	C
STPLAN	STPLAN – Calculations for sample sizes and related problems	C

<sup>1)</sup> Availability: C - complete version, D - demo version, - not available

Table A.2: Software for sample size determination – Author, Pricing, Address

<p>CADEMO für DOS, 2.1 (15.02.91), DM 340.– to DM 1800.– + fees, MS-DOS          BIORAT GmbH, Zentrum für Statistische Beratung und Datenverarbeitung, Joachim-          Jungius-Str. 9, 18059 Rostock, BRD          iec ProGAMMA, P.O. Box 841, 9700 AV Groningen, The Netherlands          CADEMO für Windows, 1.0, DM 450.– to DM 2500.– + fees, MS-Windows 3.x</p>
<p>CADEMO <i>light</i>, DM 200.– + fees, MS-Windows 3.1</p>
<p>DATASIM, 1.1 (Dos), 1.2 (Mac), US \$ 30.–, MS-DOS, Macintosh          Drake R. Bradley, Desktop Press, 90 Bardwell, St. Lewiston, Maine 04240, USA, Phone:          +207-786-6180 (4113), Fax: +207-786-6123, E-mail: dbradly@abacus.bates.edu</p>
<p>EaSt CyTeL Software Corporation, 675 Massachusetts Avenue, Cambridge, MA 02139, USA,          Tel. +617 661-2011, Fax +617 661-4405</p>
<p>EGRET SIZ, US \$ 465.–, MS-DOS 3.1          Statistics and Epidemiology Research Corporation, EGRET Software Division, 909 North-          east 43rd Street, Suite 202, Seattle, WA 98195, USA, Phone: +206 632 3014, Fax:          +206 547 4140, E-mail: rhm@ms.washington.edu</p>
<p>EX-SAMPLE, US \$ 250.–, MS-DOS          The Idea Works, Inc., 100 West Briarwood, Colombia, MO 65203, USA, +314 445          4554 USA, +314 445 4554</p>
<p>GPOWER, 2.0 (24. 03. 94), free of charge, MS-DOS 3.3, Macintosh System 6.05          Edgar Erdfelder, Franz Faul, Axel Bucher, Psychologisches Institut der Universität Bonn,          Römerstr. 164, 53117 Bonn, BRD, E-mail: erdfelder@uni-bonn.de, ftp.uni-trier.de pub/          pc/msdos-local, Gpower2i.exe (english version), Gpower2d.exe (german version)</p>
<p>IFNS-Fallzahlenrechner, DM 1400.–          Dr. W. Haase, IFNS GmbH, Institut für numerische Statistik, Max-Planck-Str. 37, 50833          Köln, Phone: +2234 1 85 20, Fax: +2234 18 52 65–</p>
<p>INSTAT, 6.1, Autumn 1994, UK Pfd. 35.–, MS-DOS, BBC microcomputers, Beta version of          sample size tools          Statistical Services Center, Harry Pitt Building, University of Reading, Whiteknights Rd.,          P.O. Box 217, Reading, RG6 2AN, U.K., Tel. +44 1734 318028, Fax. +44 1734          753169, E-mail: j.knock@reading.ac.uk</p>
<p>N, 2.2, DM 980.– + fees, MS-DOS          idv – Datenanalyse und Versuchsplanung, Dr. Volker W. Rahlfs, Wessobrunnerstr. 6,          82131 Gauting, BRD, Phone: +89/ 8 50 80 01, Fax: +89/850 36 66</p>
<p>nQuery, £ 295.– / DM 595, MS Windows 3.1          Statistical Solutions Ltd., 8 South Bank, Crosse's Green, Cork, Ireland, +353 21 319630,          Sales@StatSol.IE or: GWD mbH, Siemensstraße 21, 63165 Mühlheim/Main, BRD,          +61 08 73066</p>
<p>Nsurv, 2.2, DM 980.– + fees, MS-DOS          see N for address</p>

Table A.2 (continued)

<p>PASS, 6.0, US \$ 249.—, MS-Windows 3.1  Methodist Hospital of Indiana Inc., 1701 North Senate Boulevard, P.O. Box 1367, Indianapolis, IN 46206, USA, +317 924-6411, or  NCSS Statistical Software, 329 North 1000 East, Keyssville, Utah 84037, Phone: (801) 546-0445, Fax: (801) 546-3907, Email: ncss@ix.netcom.com, WWW:  <a href="http://www.ncss.com/pass.html">http://www.ncss.com/pass.html</a></p>
<p>PC-SIZE, 2.13 (1986), SIZE 1.01 (1990), US \$ 15.— (both), MS-DOS  Gerard E. Dallal, 54 High Plain Road, Andover, MA 01810, USA</p>
<p>PEST 3.0  John Whitehead, Hazel Brunier, Caroline Ellwood, The PEST Project, Department of Applied Statistics, University of Reading, United Kingdom</p>
<p>PLANUNG II.5 (1991), free of charge, MS-DOS  Deutsches Krebsforschungszentrum, Abt. Biostatistik, Dr. Lutz Edler, Im Neuenheimer Feld 280, 69120 Heidelberg, BRD, Fax: (+62 21) 42-23 97, E-mail: edler@krebs-heidelberg.de</p>
<p>POWER, 1.4, MS-DOS  William Dupont, Walton D. Plummer, S-3301 Medical Center North, Department of Preventive Medicine, Vanderbilt University School of Medicine, Nashville, TN 37232-2367, USA</p>
<p>Power &amp; Effect, Single copy US \$ 50.—, Up to 12 copies US \$ 210.—, MS Windows 3.0 or Macintosh and Hypercard 2.1  Glenn E. Meyer, Department of Psychology, Trinity University, 715 Stadium Drive San Antonio, Texas 78212-7200, USA, E-mail: gmeyer@trinity.edu</p>
<p>PowerPack, US \$ 195.— (Educ.: US \$ 95.—), MS-DOS  Russell V. Lenth, PowerPack, 3061 Hastings Avenue, Iowa City, Iowa 52240, USA, +319 337-8549</p>
<p>POWER, MS-DOS  Ph.D. Marcy Petrini, Pulmonary Division, Department of Medicine, University of Mississippi Medical School, Jackson, MS 39216-4505, Phone-B: 601-984-5656, Phone-H: 601-957-1423</p>
<p>SEPARATE, 1.1, US \$ 650.— (for universities free), MS-DOS 3.3  Prof. Dr. Guido Giani, Diabetes Forschungsinstitut, Abteilung Biometrie &amp; Epidemiologie, Auf'm Hennekeamp 65, D-40225 Düsseldorf, Phone: (+49) 211 338 22 59, Fax: (+49) 211 338 26 77, E-mail: giani@dfi.uni-duesseldorf.de</p>
<p>SMPLSIZE, 1994, MS-DOS  Victor L. Landry, Mercy Hospital, Trauma Center, 1400 Locust Street, Pittsburgh PA 15219-5166 USA, +412 232-8414 (Work), +412 744-7931 (Home), E-mail: landry@duq3.cc.duq.edu</p>
<p>SOLO-POWER 1.0, 1992, US \$ 195.—, MS-DOS  BMDP Statistical Software, Inc., 1440 Sepulveda Blvd., Suite 316, Los Angeles, CA 90025, USA, +310 479-7799 or: Cork Technology Park, Model Farm Road, Cork, Ireland or: GWD mbH, Siemensstraße 21, 63165 Mühlheim/Main, BRD, +61 08 730 66</p>

Table A.2 (continued)

SSIZE, US \$ 10.-, MS-DOS 3.3 F. Y. Hsieh, Computing and Statistical Services, ANAQUEST, BOC Health, Care100 Mountain Ave Murray Hill, NJ 07974, USA
STAT-POWER, US \$ 195.-, MS-DOS Drs. K. Peter Koning, interuniversity expertise center ProGAMMA, P.O. Box 841, 9700 AV Groningen, The Netherlands, +50 63 69 60, E-mail: p.koning@gamma.rug.nl
STPLAN 4.0 (06/93), MS-DOS Barry W. Brown, The University of Texas, M.D. Anderson Cancer Center, Department of Biostatistics, Box 237, 1515 Holcombe Boulevard, Houston, Texas 77030, USA, ftp:// odin.mdacc.tmc.edu (or: odin.mda.uth.tmc.edu, 129.106.3.17), /pub/unix/stplan-4.0-tar.Z (source code), /pub/stplan40.exe (MS-DOS executable),
TRIQ, 1.1, DM 1099.-, + fees, MS-Windows BIORAT GmbH, s. CADEMO für DOS

Table A.3: Two-sided test for difference of two means (sample sizes per group, effect size  $\Delta/\sigma$ )

Program	$\Delta/\sigma$	$\alpha/1 - \beta$					
		0.1/0.8	0.1/0.9	0.05/0.8	0.05/0.9	0.01/0.8	0.01/0.9
t-test							
Reference values <sup>1)</sup>	5	2.0	2.0	2.1	2.3	2.9	3.1
	2	4.1	5.3	5.1	6.4	7.6	9.1
	0.2	309.9	429.0	393.4	526.4	585.5	745.5
CADEMO LIGHT	5	2	2	3	3	3	4
	2	5	6	6	7	8	10
	0.2	310	430	394	527	586	746
DATASIM	5	2	2	2	2	2	3
	2	4	5	5	7	8	9
	0.2	310	429	394	526	586	746
GPOWER	5	2	2	2	2	2	4
	2	4	6	6	7	8	10
	0.2	310	429	394	527	586	746
IFNS	5	2	2	2	2	3	3
	2	4	5	5	7	8	10
	0.2	310	429	394	527	586	746
N	5	2	2	3	3	3	4
	2	4	6	6	7	8	10
	0.2	310	429	394	527	586	746

Table A.3: (continued)

Program	$\Delta/\sigma$	$\alpha/1 - \beta$					
		0.1/0.8	0.1/0.9	0.05/0.8	0.05/0.9	0.01/0.8	0.01/0.9
NQUERY	5	2	2	3	3	3	4
	2	4	6	6	7	8	10
	0.2	310	429	394	527	586	746
PC-SIZE	5	2	2	3	3	3	4
	2	4	6	6	7	8	10
	0.2	310	429	394	526	586	744
PLANUNG	5	2	2	3	3	3	4
	2	4	6	6	7	8	10
	0.2	310	429	394	527	586	746
POWER	5	—	2	2	2	3	3
	2	4	5	5	6	8	9
	0.2	309	428	392	525	584	744
POWERPACK	5	2	2	2	2	2	2
	2	4	6	4	6	6	8
	0.2	310	428	396	526	584	744
POWERT	5	8	8	8	8	8	8
	2	8	8	8	8	8	9
	0.2	311	430	395	528	587	747
SSIZE	5	—	2	2	2	3	3
	2	4	6	5	7	8	9
	0.2	310	429	394	527	586	746
STAT-POWER	5	2.0	2.0	2.5	2.5	3.0	3.5
	2	4.0	5.1	5.1	6.4	7.7	9.3
	0.2	309.8	429.8	393.4	526.3	585.6	745.6
STPLAN	5	—	—	2.1	2.3	2.8	3.0
	2	4.0	5.2	5.1	6.4	7.5	9.1
	0.2	309.4	429.1	393.5	526.2	585.4	745.3
Gauss-test							
Reference values <sup>1)</sup>	5	0.5	0.7	0.6	0.8	0.9	1.2
	2	3.1	4.3	3.9	5.3	5.8	7.4
	0.2	309.1	428.2	392.4	525.4	584.0	744.0
EX-SAMPLE	5	0.5	0.5	0.5	1.0	1.0	1.0
	2	3.0	4.5	4.0	5.5	6.0	7.5
	0.2	309.0	428.0	392.5	525.5	584.0	744.0
INSTAT	5	0.5	0.7	0.7	0.9	0.9	1.2
	2	3.1	4.2	3.9	5.3	5.9	7.5
	0.2	309.2	428.2	392.5	525.4	584.0	744.0

Table A.3: (continued)

Program	$\Delta/\sigma$	$\alpha/1 - \beta$					
		0.1/0.8	0.1/0.9	0.05/0.8	0.05/0.9	0.01/0.8	0.01/0.9
PLANUNG	5	1	1	1	1	1	2
	2	4	5	4	6	6	8
	0.2	310	429	393	526	584	744
SMPLSIZE	5	1	1	1	1	1	2
	2	4	5	4	6	6	8
	0.2	309	428	392	525	584	744
SOLO-POWER	5	2	2	2	2	2	2
	2	4	5	4	6	6	8
	0.2	310	429	393	526	584	744

<sup>1)</sup> Calculated using the authors' SAS macros

## Appendix B – Source Code Based Programs and WWW-Pages related to Sample Size Determination

In addition to complete programs, there is a variety of programs for sample size determination programs which require a compiler such as Fortran or Basic, or a program environment such as SAS or S-Plus. These programs are not included in the comparative tests, but are listed below, by environment, with their source address.

Application of these programs requires an appropriate compiler or program environment and familiarity with the programming language. An advantage is that the source code is accessible, for modification or embedding in other programs.

### C Routines

- MHGSEQ: A computer program for designing clinical trials with arbitrary survival curves and group sequential testing. Author: Jerry W. Halpern (funn@osiris.stanford.de), B.Wm. Brown
- Program makes use of the Makuch and Simon test statistic to calculate the power for one-way or multi-way factorial designs. Monte Carlo method is used to obtain type I and II error probabilities. Survival distribution can be exponential, Weibull or lognormal. Author: R. Natarajan (ranjini@stat.brown.edu)

### Fortran Routines

- MUL4.F, MUL4XC.F: Backward and Forward Elimination, based on Chen and Simon. Author: Timothy Chen (tchen@helix.nih.gov)
- SEQPWR.F, SEQOPR.F: Power calculations for sequential designs, based on Kim (1995). Author: Dr. Kim (kmkim@umich.edu)

### SAS Macros and Programs

- EQUATL.SAS: SAS-macro for sample size calculation for the assessment of (at least) equivalence of two means and proportions. Author: Manfred Wargenau, FOCUS, Clinical Drug Development GmbH (100534.3056@compuserve.com).

- NSUJ: based on method by Whitehead, but is adapted for Wilcoxon rangsum test and logistic regression, *Statistics in Medicine*, 1993. Author: Alain Duhamel (duhamel@saphir.univlille2.fr).
- POWER.MAK: SAS-macros for sample size calculation for the one- and two-sample t-test and test on proportions. Author: Patrick Onghena, University Leuven (patrick=onghena%psl%psy@cc3.kuleuven.ac.de).
- POWER.SAS is included in SAS/STAT software since release 6.09 of the SAS software, power.tst contains some examples for using. Author: Kerstin Latour (ftp.sas.com techsup/download/technotes TS272. PS, 12.03.92).
- SAMPLESIZE allows calculation of sample sizes for the two-sample t-test on difference and equivalence in means and two-sample test on equivalence in proportions. It is a set of SAS/IML-moduls, written in English with a menu- and mask-based user interface (realized with SAS/AF). It supports creation of graphs and tables. The documentation is on paper available. Authors: Dieter Hauschke, Martin Burke, Edgar Diletti, BYK Gulden Pharmazeutika, Abt. Biometrie, BYK Gulden Str. 2, 78467 Konstanz, +7531842301.
- SAMSIZE: Power/Sample size calculation for one-way anova. Send Email to archive@jse.stat.ncsu.edu with the message 'send software/sas/survrisk.pak'. Author: Frank Harrell (feh@biostat.mc.duke.edu).
- SAMSIZEB: Power/Sample size for Fisher's exact test following CASAGRANDE and PIKE (1982). Send Email to archive@jse.stat.ncsu.edu with the message 'send software/sas/survrisk.pak'. Author: Frank Harrell (feh@biostat.mc.duke.edu).
- SAMSIZEC: SAS-macro for power and sample size calculation for survival analysis based on the Cox model. Send Email to archive@jse.stat.ncsu.edu with the message 'send software/sas/survrisk.pak'. Author: Frank Harrell (feh@biostat.mc.duke.edu). See also S-plus functions ciapower and cpower.
- SAMSIZER: Power/Sample size calculation for linear correlation. Send Email to archive@jse.stat.ncsu.edu with the message 'send software/sas/survrisk.pak'. Author: Frank Harrell (feh@biostat.mc.duke.edu).
- SIZE: SAS-program based on J. Shih (1995, *Controlled Clinical Trials*, Vol. 16, 395–407): Sample size calculations for complex clinical trials (jshih@helix.nih.gov).
- UNIFYPOW.SAS is a module of SAS data step statements that perform power computations for the classical one-group and two-group t-tests, and the one-way ANOVA. It replaces ONEWYPOW.SAS which is described in O'Brien, R. G. and Muller, K. E. (1993). Author: R. G. O'Brien (robrien@bio.ri.ccf.org)

### S-Plus Functions and Packages

All the following Splus-functions and Splus-packages are located on the STATLIB server which could be access via anonymous ftp to lib.stat.cmu.edu or via the Web-page <http://lib.stat.cmu.edu/S/>.

- ASYPOW: Asymptotic sample size, power, and detectable effect calculations for a variety of generalized linear models and survival models (include Fortran code). Author: Barry W. Brown (bwb@odin.mda.uth.tmc.edu).
- CIAPOWER compute power of interaction test for exponential survival (Peterson and George, 1993). Author: Frank Harrell (feh@biostat.mc.duke.edu).
- CPOWER: Power for two-sample Cox or logrank survival test. Frank Harrell (feh@biostat.mc.duke.edu).
- POWER: A suite of Splus functions to compute power and sample size for tests of the general linear hypothesis (Muller and Peterson, 1984). Author: Daniel F. Heitjan (dheitjan@biostats.hme.psu.edu).

- SAMPLESIZE-BETABIN: Calculates optimal sample sizes for a beta-binomial mixture model in a situation where both priors and utilities differ for two parties in a negotiation process. Based on the paper Wolfson, L. J., Kadane, J. B., and Small, M. J. (1995), "Expected Utility as a Policy-Making Tool: An Environmental Health Example", Author: Lara Wolfson (ljwolfson@setosa.uwaterloo.ca).
- SAMPLESIZE-MEAN: An S-plus program to calculate sample sizes based on highest posterior density intervals for normal means using three different Bayesian approaches. Authors: Lawrence Joseph (joseph@binky.epi.mcgill.ca) and Patrick Belisle.
- SAMPLESIZE-PROP: A S-plus program to calculate sample sizes based on highest posterior density intervals when comparing two binomial proportions using three different Bayesian approaches. Authors: Lawrence Joseph (joseph@binky.epi.mcgill.ca), Roxane du Berger and Patrick Belisle.

### WWW Pages

In addition to the complete programs and source codes, the growing publicity of the WWW also makes sample size programs available on the Internet. Two examples are:

- <http://hedwig.mgh.harvard.edu/size.html>  
Web page of David A. Schoenfeld, Ph.D.: Statistical considerations for clinical trials and scientific experiments (schoenfe@stat3.mgh.harvard.edu)
- <http://www.interchg.ubc.ca/cacb/power>  
gives 'A comprehensive list of power analysis software for microcomputers'. The page is maintained by Len Thomas, Center for Applied Conservation Biology, University of British Columbia (lthomas@unixg.ubc.ca)
- <http://www.stat.ucla.edu/bond/HTMLPOWER/>  
calculates sample size and power for one- and two-sample test situations with normal, binomial, exponential and poisson distribution (jbond@stat.ucla.edu).

#### Authors' addresses:

Dr. C. Ortseifen, Universitätsrechenzentrum, INF 293, 69120 Heidelberg  
T. Bruckner, Statistik und DV, Lessingstr. 3, 68723 Schwetzingen  
M. Burke, BYK Gulden Pharmazeutika, BYK Gulden Str. 2, 78467 Konstanz  
Dr. M. Kieser, Dr. Willmar Schwabe GmbH & Co., Willmar-Schwabe-Str. 4, 76227 Karlsruhe



## BUCHBESPRECHUNG/BOOKREVIEWS

DIRSCHIEDL, P., OSTERMANN, R. (Eds.)

**Computational Statistics – Papers Collected on the Occasion of the 25th Conference on Statistical Computing at Schloß Reisenburg**

Physica-Verlag, Heidelberg 1994, ISBN 3-7908-0813-X, 552 S., 82 Abb., kt.

Dieses Buch entstand anlässlich des 25. Jahrestages der jährlich im Juni stattfindenden Reisenburg-Arbeitskonferenzen über Statistical Computing. Es enthält eine Sammlung von 30 Aufsätzen zu den unterschiedlichsten Themen im Zusammenhang mit Statistical Computing bzw. den Reisenburg-Tagungen. Es deckt ein breites Themen-Spektrum ab, mit der Idee (wie die Herausgeber in ihrem Vorwort selber schreiben) als eine Art Tutorial vertieften Ein- bzw. Überblick über die Schwerpunkt-Themen zu geben, die auf Reisenburg-Tagungen wichtig waren bzw. sind. Von historischen Rückblicken sowohl über die Reisenburgtagung als auch die europäische Geschichte des Statistical Computing, über die Eigenschaften verfügbarer Softwarepakete, wie z. B. deren numerische Verlässlichkeit, bis hin zu Spezialthemen wie Bootstrapping, Multivariate Graphics, Cluster-Algorithmen, Multiples Testen etc. etc. bietet dieses Buch sicher für jeden, der mit Statistical Computing befaßt ist, neben interessanten Artikeln zum persönlichen Arbeitsschwerpunkt einen guten Ein- und Überblick in andere Facetten dieser Disziplin.

K. Schulz, Hannover

MAOR, R

**Die Zahl e – Geschichte und Geschichten**

1996, 225 S., DM 48,-, ISBN 3-7643-5093-8, Birkhäuser Verlag AG, Basel

Die Zahl  $e$  hat in allen Bereichen der Mathematik und speziell auch in den Anwendungen der Mathematik in der Physik, Biologie, Musik und Kunst eine große Bedeutung. Der Autor hat es verstanden, die Geschichte der Entdeckung von  $e$  sowie u. a. die Verquickung mit der Zahl  $\pi$  anschaulich darzustellen. Dabei schildert er u. a. Einzelheiten aus dem Leben der Persönlichkeiten, die bei der Erforschung der Zahl  $e$  entscheidend mitgewirkt haben. So entstand ein Buch, das auch für Leser mit geringen mathematischen Vorkenntnissen eine interessante und anregende Lektüre sein kann, da es Zusammenhänge deutlich werden läßt, die heute bei dem Streben nach Spezialwissen vielfach nicht beachtet werden.

H. Geidel, Stuttgart

BRAUER W., MÜNCH S.

**Studien- und Forschungsführer Informatik**

Dritte, völlig neubearbeitete Auflage, 1996, 388 S., DM 28, ISBN 3-540-60417-0  
Berlin: Springer Verlag

Der Studien- und Forschungsführer bietet eine Übersicht über die wissenschaftlichen Hochschulen und öffentlich-rechtlichen Forschungseinrichtungen auf dem Gebiet der Informatik. W. Brauer und S. Münch haben das Buch herausgegeben im Auftrag der Gesellschaft für Informatik, des Fakultätentags Informatik und der Hochschulrektorenkonferenz. Wesentlicher Unterschied zu früheren Auflagen ist, daß nun auch die Informatik-Studiengänge der Universitäten der neuen Bundesländer aufgeführt sind. Im Gegensatz sind jedoch die Fachhochschulstudiengänge nicht mehr erwähnt, da für diese ein eigener Studienführer herausgegeben wird. Insgesamt werden 59 (früher 25) Studiengänge und 35 (früher 14) Forschungseinrichtungen beschrieben.

Nach einer Einleitung über die Informatik aus verschiedenen Sichten, wurden unterschiedliche Informationsmaterialien zusammengetragen. Diese reichen von der Rahmenordnung über Ausbildungsempfehlungen bis zur Beschreibung einzelner Gesellschaften.

Der Abschnitt über Studiengänge, Hochschullehrer und Forschungsaktivitäten macht den größten Teil des Buches aus. Er ist nach den Städten alphabetisch sortiert. Dabei wird jeder Studiengang durch eine Kurzbeschreibung des Curriculums und auch durch die beteiligten Professoren (bzw. Forschungsgruppenleitern) mit ihren Forschungsgebieten und -projekten beschrieben. Ebenso werden Kooperationen mit anderen Forschungsinstitutionen und Firmen aufgeführt. Schade, daß die anwendungsorientierten Fächer wie Wirtschafts- und Medizininformatik nicht mit aufgenommen worden sind.

Im letzten Kapitel werden 35 nichtuniversitäre Forschungsinstitutionen vorgestellt mit ihren Aufgaben, ihren Forschungs- und Entwicklungsvorhaben und den Kooperationen mit anderen Forschungsinstitutionen bzw. Firmen.

Das Schlagwortregister beschränkt sich wie in der zweiten Auflage auf ein Namensregister, hier wäre sicherlich für manche Zielgruppen auch ein Verzeichnis von Forschungsgebieten wünschenswert gewesen.

Das Buch ist vor allem ein wertvolles Nachschlagewerk für Studierende der Informatik, solche, die es werden wollen und Wissenschaftler auf dem Gebiet der Informationsverarbeitung, die sich über Forschungsschwerpunkte an anderen Einrichtungen informieren wollen. Petra Knaup, Heidelberg

Coy, W. et al (ed)

**Sichtweisen der Informatik**

Braunschweig; Wiesbaden: Vieweg, 1992, ISBN 3-528-05263-5, 409 Seiten, DM 59,-

Das Buch ist der erste Band einer neuen Reihe der „Theorie der Informatik“ – nicht der Theoretischen Informatik, wohlgermerkt! Wolfgang Coy, Informatik-Professor aus Bremen und Sprecher der GI-Fachgruppe „Theorie der Informatik“, möchte den starren Blick von der „Maschine Computer“ abwenden und hinwenden zu einer breiteren, mehr geisteswissenschaftlich orientierten Theoriebildung der Informatik. Philosophische, sozialwissenschaftliche, rechtliche, politische wie kulturelle Ansätze sollen gleichberechtigt neben den physikalischen, technischen, mathematischen und logischen Grundlagen dargestellt werden.

Die Informatik scheint eine Disziplin im Umbruch, ihre Konturen sind sehr verschwommen. Einige fordern gar einen Rückzug „back to the roots“ – Informatik als „Science of Computing“, als Theorie der Algorithmen und ihrer Komplexität. Diese Theorie leistet mit der Entwicklung des Berechenbarkeitsbegriffs einen bedeutenden Beitrag zu Mathematik und Philosophie – aber reicht das für die PRAXIS des Rechnereinsatzes? Es fehlt eine Theorie der Informatik, die eine Fundierung der Informatikpraxis ermöglicht!

Arno Rolf und Dirk Siefkes plädieren im Buch für eine umfassende Theorie der Informatik – einer Theorie im ursprünglichen Sinne des Wortes: Eine Theorie, die eine Sicht auf die Dinge bietet, welche Verstehen, und damit verantwortungsvolles Handeln in der Informatikpraxis ermöglicht. „Betroffen vom Einsatz von Rechnern sind nicht die Rechner, sondern die Menschen, oft auch die Natur. Warum also so bescheiden? Informatiker konstruieren nicht Rechnersysteme, sondern sie manipulieren soziale und ökologische Systeme, indem sie Rechner hineinkonstruieren. Nur mit dieser weiteren Sicht können sie versuchen, zu verstehen, was sie herstellen.“

Informatik als Gestaltungswissenschaft: „Gestaltend balancieren Informatiker zwischen den Humanisten, die interpretieren, und den Ingenieuren, die konstruieren. Nur so können sie die menschlichen Belange wahren.“ Folgt man Coy und den vielen anderen Autoren, muß sich die (Medizinische) Informatik schleunigst in Richtung Gestaltungswissenschaft entwickeln, will sie für die heutigen und zukünftigen praktischen und wissenschaftlichen Aufgaben gerüstet sein.

Die Architektur könnte für die Informatik ein Vorbild sein: Für viele Architekten spielen neben dem funktional Konstruktiven ästhetische und gesellschaftliche Kategorien für ihr professionelles Handeln eine Rolle. Sie sind sich bewußt, daß sie Mitgestalter gesellschaftlicher Prozesse sind und in Lebens-, Arbeits- und Wohnprozesse eingreifen. Obwohl die Informatik nicht weniger stark als die Architektur am Wandel der Gesellschaft beteiligt ist (Es wird ja sogar behauptet, wir lebten im „Informationszeitalter“ – und das wird auch kritiklos akzeptiert!), haben viele Informatiker diese erweiterte Sichtweise noch nicht für sich entdeckt. In ihrem Selbstverständnis verstehen sie sich als Bauingenieur oder Klempner.

Um diese, etwas andere, Sichtweise von Informatik kennenzulernen, ist dieser Band Pflichtlektüre für alle an der Weiterentwicklung unseres Faches interessierten Medizininformatiker – zumal, wenn sie sich als „angewandte Informatiker“ begreifen.

Das Werk besteht aus vier großen Teilen: So werden auf den ersten 100 Seiten Gedanken zu den „Grundlagen einer Theorie der Informatik“ formuliert. Ein weiteres Viertel des Buches steht unter dem Thema „Computer und Arbeit“. 100 Seiten beschäftigen sich mit „Kultur – Anthropologie – Computer“. Am Ende des Buches stehen Aufsätze zum Themenkomplex „Informatik – Ethik – Verantwortung“.

21 Autoren haben Beiträge in diese Diskussion beigesteuert.

Claus Stark, Heilbronn

**Mitteilungen der  
Deutschen Gesellschaft für Medizinische Informatik,  
Biometrie und Epidemiologie**

Brief des Präsidenten

Lieben Kolleginnen und Kollegen,

Die Vorbereitungen für die nächste GMDS-Tagung laufen auf vollen Touren. Bisher sind über 300 Beiträge eingereicht worden.

Das Begutachtungsverfahren ist noch nicht ganz abgeschlossen. Das Programmkomitee wird am 8. Mai das endgültige Tagungsprogramm festlegen, so daß Anfang Juni das Programm verschickt werden kann. Wie bereits angekündigt wird der Proceedingsband zur Tagung vorliegen. Auf einige Sonderveranstaltungen möchte ich schon jetzt hinweisen:

- Unter dem Titel „Brennpunkt MI“ wird eine Podiumsdiskussion zwischen einem Kliniker, einem Verwaltungsdirektor und Vertretern der MI zur Frage „Medizinische Informatik - DV-Abteilung der Krankenhausverwaltung oder Wissenschaft“ stattfinden.
- Die Herren Feinstein (Vortragsthema „Evidence based medicine“) und Altmann („How to develop unreliable models“) werden unserer Einladung, an der Tagung vorzutragen, nachkommen.
- Als fachfremder Redner wird Herr Rommel (ehem. OB Stuttgart) zum Thema „Die Schwaben und die Welt“ sprechen.
- Zum Thema „Bessere Heilungschancen bei Krebs - Ein Erfolg von Ärzten und Statistikern, Epidemiologen und Informatikern“ wird ein öffentlicher Nachmittag mit mehreren Vorträgen für die Ulmer Bevölkerung und die regionale und überregionale Presse angeboten.

Die bereits in meinem letzten Brief angesprochene Neustrukturierung der GMDS-Geschäftsstelle ist weiter fortgeschritten. Inzwischen hat ein Auswahlverfahren für den hauptamtlichen Geschäftsführer der GMDS stattgefunden. Das Präsidium ist der festen Überzeugung mit Herrn Dipl.-Volksw. Thomas Banasiewicz einen hervorragenden Kandidaten ausgewählt zu haben. Herr Banasiewicz wird seine Tätigkeit am 1. Juni 1997 beginnen.

Der Aufruf zum Vorschlag von Kandidaten für die Wahl zum 1. Vizepräsidenten der GMDS hat zur Nennung einer Reihe von potentiellen Bewerbern geführt. Naturgemäß konnten einige der Genannten aus verschiedenen Gründen nicht zur Verfügung stehen. Um so erfreuter bin ich, daß sich zwei in jeder Beziehung ausgewiesene Bewerber bereit erklärt haben, für das Amt des 1. Vizepräsidenten zu kandidieren. Ich bitte alle Mitglieder der GMDS von Ihrem Briefwahlrecht Gebrauch zu machen und sich so aktiv an der weiteren Gestaltung der GMDS zu beteiligen.

Im Dezember 1996 hatten in einer gemeinsamen Stellungnahme die GMDS Arbeitsgruppen „Medizinische Dokumentation und Klassifikation“ und „Qualitätssicherung in der Medizin“ zur medizinischen Dokumentation für die Qualitätssicherung bei den Fallpauschalen und Sonderentgelten (SGB V § 137 i. V. m. § 112) darauf aufmerksam gemacht, daß die Qualitätsdokumentation in der vorgesehenen Form methodische Mängel aufweist, der Dokumentationsaufwand in keinem Verhältnis zu dem erwarteten Nutzen steht und daß das klinische Qualitätsmanagement eher behindert als gefördert wird. Anlässlich der Ankündigung der baden-württembergischen Krankenhausgesellschaft, daß an der Qualitätskontrolle in

bisherigen Umfang festgehalten wird, haben die baden-württembergischen Fachvertreter für Medizinische Informatik in einer gemeinsamen Erklärung vom 17. Februar 1997 noch einmal auf die verschiedenen Mängel hingewiesen. Inzwischen hat der Deutsche Bundestag am 20. März 1997 in zweiter und dritter Lesung mit dem 2. GKV-Neuordnungsgesetz die Bedürfnisse der Ärzteschaft bei der Gestaltung und Umsetzung der Maßnahmen zur Qualitätssicherung ärztlicher Leistungen im Krankenhaus erweitert. Sowohl die Bundesärztekammer als auch die Landesärztekammern und die Kassenärztliche Bundesvereinigung werden damit zukünftig neben den Spitzenverbänden der gesetzlichen Krankenversicherung und der Deutschen Krankenhausgesellschaft vertreten sein. Mit dieser politischen Neuordnung ergibt sich die Chance, methodisch sauber fundierte Qualitätssicherungsmaßnahmen zu etablieren. Die GMDS mit ihren fachkundigen Arbeitsgruppen ist auf alle Fälle bereit, über die entsprechenden Schienen (Bundesärztekammer, AWMF) ihr Know-how zur Verfügung zu stellen.

Im zweiten Quartal des Jahres 1997 möchte ich den folgenden Kollegen im Namen der GMDS ganz herzlich zum Geburtstag gratulieren:

**65. Geburtstag**

Herr Günter Kunkel (Bremen)

**60. Geburtstag**

Herr Dipl. Phys. Jürgen Pachaly (Berlin)

Herr Prof. Dr. Karl-Heinz Höhne (Hamburg)

Ihr

Prof. W. Köpcke

### Weitere Informationen

#### GMDS Geschäftsstelle:

**Achtung:** Neue Telefonnummer (0221) 4004-865: Frau Büttner, unsere neue Sekretärin, ist i.d.R. von 8.00 bis 13.00 Uhr zu erreichen.

\*\*\*\*\*

### Ankündigungen

#### Buchankündigung

#### Leiner, Gaus, Haux: Medizinische Dokumentation.

Die 2. Auflage dieses einführenden Lehrbuchs ist soeben erschienen. In ihr sind neuere Entwicklungen in der Gesetzgebung und bei medizinischen Ordnungssystemen berücksichtigt. Ein zusätzliches Schlagwortverzeichnis rundet die Nutzungsmöglichkeiten ab. Das Lehrbuch führt in die Grundlagen einer professionellen medizinischen Dokumentation ein. Ein umfangreicher Thesaurus macht es zu einem nützlichen Nachschlagwerk. Zielgruppe sind Studierende ebenso wie alle, die in der Medizin dokumentieren müssen: Ärzte, Pflegekräfte, Dokumentare und Verwaltungsangestellte.

Leiner, F., Gaus, W., Haux, R.: Medizinische Dokumentation - Einführendes Lehrbuch. Stuttgart: Schattauer Verlag, 2. Auflage 1997, ISBN 3-7945-1796-2, 191 Seiten, Preis: DM 49,-.

#### Medizinische Informatik, Biometrie und Epidemiologie

Herausgegeben von Hans-Jürgen Seelos

Unter Mitarbeit von S. Bürsner, H. Dickhaus, B. Graubner, I. Guggenmoos-Holzmann, J. Ingenerf, R. Klar, E. Pelikan, B. Pietsch-Breitfeld, R. Reppes, B. Schneider, H. K. Selbmann, K. Spitzer, C. Spreckelsen, T. Tolxdorff und K. Ulm  
24,0 x 17,0 cm. XXV, 299 Seiten. 1997. Broschiert DM 58,-/öS 423,-/sFr 53,-  
ISBN 3-11-014317-8 (de Gruyter Lehrbuch)

Die Medizinische Informatik, Biometrie und Epidemiologie wendet sich an Studierende der Medizin und ist nach dem aktuellen Gegenstandskatalog für den zweiten Abschnitt der ärztlichen Prüfung „Medizinische Statistik und Informatik“ gestaltet.

Im wesentlichen vermittelt das knapp bemessene Buch Prüfungswissen, auch für Studierende der Medizinischen Informatik und benachbarter Disziplinen wie „Gesundheitswissenschaften“ und „Wirtschaftsinformatik“ und hilft bei der Bewältigung praktischer Probleme, z.B. bei einer statistisch angelegten Dissertation.

\*\*\*\*\*

#### Fachbereich Medizinische Informatik

CBT in der Medizin

6./7. Juni 1997 in Aachen

Institut für Medizinische Informatik und Biometrie in Zusammenarbeit mit der AG Computergestützte Lehr- und Lernsysteme in der Medizin (CBT) der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmDS) wird ein Workshop „CBT in der Medizin“ ausgerichtet. Beiträge sind vorgesehen in Form von: wissenschaftlichen Vorträgen (30 min), Kurzvorträgen / Vorstellung von Arbeitsgruppen (15 min), Systemdemonstrationen und Postern. Die Deadline für die Abstracts ist der 15. Februar 1997. Aktuelle Informationen finden sich unter

[http://www.imib.rwth-aachen.de/www/cbt\\_workshop](http://www.imib.rwth-aachen.de/www/cbt_workshop)

Anfragen können gestellt werden unter: Tel.: 0241 / 80-88790, Fax 0241 / 8888-426, mail [cbt@imib.rwth-aachen.de](mailto:cbt@imib.rwth-aachen.de)

Für die Organisatoren: Helmar Conradi, email: [helmar@silmaril.imib.rwth-aachen.de](mailto:helmar@silmaril.imib.rwth-aachen.de)  
 imib RWTH Aachen, <http://www.imib.rwth-aachen.de/www/mitarb/helmar>  
 Pauwelsstrasse 30, Tel.: +49 241 8088794, D-52057 Aachen, Fax: +49 241 8888246

Advances in Quantitative Laryngoscopy using Motion-, Image- and Signal Analysis  
 18.-19. Juli 1997 in Erlangen

Leitung: Prof. Dr. Dr. Ulrich Eysholdt, Abt. Phoniatrie und Pädaudiologie, Erlangen

Prof. Dr. Hanspeter Herzel, Lehrstuhl f. Theoretische Biologie, Berlin

Informationen: Thomas Wittenberg, Abteilung für Phoniatrie und Pädaudiologie  
 HNO-Klinik, Uni. Erlangen, Bohlenplatz 21, D-91054 Erlangen, Germany,

Tel.: +49- (0)9131-85 2603 Fax: +49- (0)9131-85 9272

Email: [wittenberg@phoni.med.uni-erlangen.de](mailto:wittenberg@phoni.med.uni-erlangen.de), <http://www.rrze.uni-erlangen.de/docs/FAU/fakultaet/med/kli/kphno/phon/rt2main.htm>

Themen: Bewegungsanalyse und Bildverarbeitung

Biomechanische Modellierung der Stimme

Signalanalyse von Stimmlippenschwingungen

Medizinische Aspekte und Ergebnisse

Aspekte der Phonetik

Spracherkennung

GEOMED '97

International Workshop on Geomedical Systems

4. - 6. September 1997 in Rostock

Angesichts globaler Veränderungen der Umwelt und der Dynamik der Weltbevölkerung braucht das Gesundheitssystem regionale und transnationale computergestützte Systeme zur Kontrolle und Prävention insbesondere von sich rasch ändernden und neuen Krankheiten. Die Diskussion um BSE zeigt, daß mangelnde Kontrollsysteme sogar die politische Stabilität einer Region gefährden können.

Interessierte wenden sich bitte an:

M. Bull, Institut für Medizinische Informatik und Biometrie,

Universität Rostock, Rembrandtstr. 16/17, 18055 Rostock

Tel: 0381/494-7308, Fax: 0381/494-7203,

E-mail: [mathias.bull@medizin.uni-rostock.de](mailto:mathias.bull@medizin.uni-rostock.de).

New Technologies in Hospital Information Systems

Conference on Architectural Concepts for Hospital Information Systems

20.-22. September 1997 in Ulm

Leitung: Prof. Dr. J. Dudeck, Gießen, Dr. B. Blobel, Magdeburg, Herr W. Lordieck, Bielefeld,  
 Dr. T. Bürkle, Gießen

Die Veranstaltung ist noch bis zum 20. Mai offen für Beitragsanmeldungen zu den Themen

Referenzarchitekturen für Krankenhausinformationssysteme

Middleware für Krankenhausinformationssysteme

Schnittstellen und Kommunikationsserver

Workflow Management

Objektorientiertes Design im KIS

Intranet - Technologie

**Distributed Hospital Environment**

Organisation, Call für Papers und nähere Auskünfte: Thomas Bürkle, Institut für Medizinische Informatik, Heinrich-Buff-Ring 44, 35392 Gießen, Tel. 0641/99-41370, Fax 0641/99-41359, Email: thomas.buerkle@informatik.med.uni-giessen.de

**Workshop Medizinische Informatik auf der Tagung der GI in Aachen**

27. September 1997 in Aachen

Auf der GI Tagung in Aachen wird am Freitag, dem 27. 9. 1997 ein Workshop mit dem Thema "Wissensbasierte Funktionen als Innovationsfaktor in Klinischen Informationssystemen" durchgeführt werden

Organisation und nähere Auskünfte: Prof.Dr. Joachim Dudeck, Institut für Medizinische Informatik, Heinrich-Buff-Ring 44, 35392 Gießen, Tel.: 0641 - 99.441350, Fax: 0641 - 99.41359, E-Mail: joachim.dudeck@informatik.med.uni-giessen.de

**Gemeinsames Treffen der CEN - TC 251 Working Groups**

16. - 18. Oktober 1997 in Berlin

Die für die Standardisierung zuständigen Working Groups des Technischen Komitees 251 (TC 251) des CEN veranstalten ihre diesjährige gemeinsame Tagung vom 16. - 18. Oktober 1997 in Berlin. Die Organisation hat das DIN gemeinsam mit dem Ausschuß C7 (Medizinische Informatik) übernommen.

Nähere Informationen: Frau Dr.-Ing. Vera Sattelmayer NA Medizin (NAMed) im DIN, Burggrafenstr. 6, 10787 Berlin, Tel. 030-2601-0/2740, Fax 030-2601-1231

**Veranstaltungen der Akademie Medizinische Informatik**

**1. Seminar "Verschlüsselung von Diagnosen und medizinischen Prozeduren  
Kodierkurs für Dokumentare und andere Interessierte"**

Termin, Ort: 6.6.97, DKFZ, Heidelberg

Zielgruppe: Med. Dokumentationspersonal, Med. Informatiker und andere an der med. Dokumentation beteiligte Mitarbeiter (einschl. Ärzten) in Krankenhäusern und sonstigen mit med. Dokumentation und Abrechnungsfragen befaßten Institutionen der Leistungserbringer und Kostenträger.

Referenten: Dr. med. Bernd Graubner, Prof. Dr. Rudolf Thurmayr

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,-DM  
(Sonstige 800,-DM, Studierende 160,-DM)

**2. Seminar "Online-Dienste für Ärzte"**

Termin, Ort: 11.6.97, CAM-Labor, Hygiene-Institut, Universität Heidelberg

Zielgruppe: Ärzte aus dem Klinik- und niedergelassenen Bereich, Ärzteberater, Medizinjournalisten, med. Dokumentare, Angehörige von Heil- und Hilfsberufen, wiss. Mitarbeiter des Pharma-Marketings.

Referentin: Edda Bhattacharjee

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,- DM  
(Sonstige 800,-DM, Studierende 160,-DM)

**3. Seminar "Wirkungsvolle Gestaltung von Vortragsmitteln"**

Termin, Ort: 12.6.97, DKFZ, Heidelberg

Zielgruppe: Referenten, Redner, Trainer. Alle, die Vorträge und Seminare halten oder Produktpräsentationen durchführen.

Referentin: Dipl.-Inform. Med. Cornelia Vosseler

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,- DM  
(Sonstige 800,-DM, Studierende 160,-DM)

**4. Seminar "Wissensbasierte Systeme in der Medizin - eine Einführung"**

- Termin, Ort: 16.6.97, Medis-Institut der gsf, Neuherberg  
 Zielgruppe: Medizinische Informatiker, Informatiker, Ärzte mit Informatikkenntnissen  
 Referent: Dr. sc. hum. Gerd Mann  
 Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,--DM  
 (Sonstige 800,--DM, Studierende 160,--DM)
5. Seminar "Wissensbasierte Systeme in der Medizin - Praxis der Entscheidungsunterstützung"\*  
 Termin, Ort: 17.6.97, Medis-Institut der gsf, Neuherberg  
 Zielgruppe: Medizinische Informatiker, Informatiker, Ärzte mit Informatikkenntnissen  
 Referent: Prof. Dr. Christian Ohmann  
 Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,--DM  
 (Sonstige 800,--DM, Studierende 160,--DM)
6. Seminar "Krankenhausinformationssysteme (für Med. Dokumentare)"  
 Termin, Ort: 19.6.97, Universität Gießen  
 Zielgruppe: Medizinische Dokumentare, Medizinisch-Technische Assistenten  
 Referenten: Dr. Thomas Bürkle, Prof. Dr. Hans-Ulrich Prokosch  
 Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,--DM  
 (Sonstige 800,--DM, Studierende 160,--DM)
7. Seminar "Serving von Informationen und Wissen im World Wide Web"  
 Termin, Ort: 23.6.97, CAM-Labor, Hygiene-Institut, Heidelberg  
 Zielgruppe: Ärzte, Med. Informatiker, EDV-Verantwortliche und -Beauftragte, die eigene Inhalte ins World Wide Web bringen wollen.  
 Referenten: Dipl.-Inform. Med. Martin Haag, Prof. Dr. Franz-Josef Leven, Dipl.-Inform. Med. Leonhard Maylein  
 Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,--DM  
 (Sonstige 800,--DM, Studierende 160,--DM)
8. Seminar "Internet und WWW: Browsing und andere Methoden der Informationsrecherche"  
 Termin, Ort: 23.6.97, CIP-Pool 1 der Med. Fakultät, Universität Münster  
 Zielgruppe: Das Seminar richtet sich an alle, die die verschiedenen Dienste des Internets und die Möglichkeiten, diese zur Informationsrecherche zu nutzen, kennenlernen wollen.  
 Referenten: Prof. Dr. Hans-Ulrich Prokosch, T. Ganslandt, Dr. Oliver Obst  
 Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,--DM  
 (Sonstige 800,--DM, Studierende 160,--DM)
9. Seminar "Serving von Informationen und Wissen im World Wide Web"  
 Termin, Ort: 24.6.97, CIP-Pool 1 der Med. Fakultät, Universität Münster  
 Zielgruppe: Ärzte, Med. Informatiker, EDV-Verantwortliche und -Beauftragte, die eigene Inhalte ins World Wide Web bringen wollen. Interessenten, die auch etwas über das Recherchieren von Informationen im Internet (und speziell im WWW) lernen wollen, können dieses Seminar mit dem Seminar "Internet und WWW: Browsing und andere Methoden der Informationsrecherche" verbinden.  
 Referenten: Prof. Dr. Hans-Ulrich Prokosch, T. Ganslandt, Dr. Oliver Obst  
 Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,--DM  
 (Sonstige 800,--DM, Studierende 160,--DM)  
 bei Buchung von Seminar 8. und 9.: Mitglieder von BVMI, DVMD, GMDS 500,--DM (Sonstige 1.200,--DM, Studierende 300,--DM)
10. Seminar "Elektronische Dokumentenmanagement- und Archivierungssysteme - ein Weg zur elektronischen Patientenakte"



Termin, Ort: 25.- 27.6.97, DKFZ, Heidelberg

Zielgruppe: Med. Informatiker in Industrie, Forschung und Krankenhaus, leitende Kräfte in Klinikrechenzentren, Ärzte mit Kenntnissen der Med. Informatik, Med. Dokumentare, Pflegedienst, Funktionsbereiche, Verwaltungsfachkräfte.

Referenten: Dipl.-Inform. Paul Schmücker, Dr. Carl Dujat, Dipl.-Inform. Med. Oliver Reinhard

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 800,--DM  
(Sonstige 1.800,--DM, Studierende 350,--DM)

11. Seminar "Kommunikationsstandards im Krankenhaus (incl. Internet, Multimedia, Bilder, DOXX)"\*)

Termin, Ort: 3.-4.7.97, IVMed, Universität Erlangen-Nürnberg

Zielgruppe: Dieses Seminar richtet sich an die für die Kommunikation im Gesundheitswesen verantwortlichen Personen.

Referenten: Prof. Dr. Joachim Dudeck, Dipl.-Inform. Med. Christof Seggewies, Dr. Bernhard Wentz

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 500,--DM  
(Sonstige 1.200,--DM, Studierende 300,--DM)

12. Seminar "Stand Übergang ICD9-ICD10 und Änderungen durch ICD10"\*)

Termin, Ort: 7.7.97, DKFZ, Heidelberg

Zielgruppe: Ärzte im Krankenhaus, niedergelassene Ärzte, Medizinische Dokumentare

Referent: Dr. Albrecht Zaiß

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 320,- DM  
(Sonstige 800,- DM, Studierende 160,- DM)

\*) Diese Kurse werden für den Erwerb des Zertifikats „Medizinische Informatik“ besonders empfohlen

Auskunft : Frau P. Skalecki, Universität Heidelberg, Abt. Medizinische Informatik, Im Neuenheimer Feld 400, 69120 Heidelberg, Tel.: 06221/56-7398, Fax: 06221/56-4997,  
e-mail: Petra.Skalecki@kkrzmail.krz.uni-heidelberg.de

\*\*\*\*\*

#### Zentrum Biometrie - Veranstaltungen 1997/98

Veranstaltung: Instrumente der Datenverarbeitung (SAS I und SAS II)\*

Termin, Ort: 22. - 24.05.97 und 26. - 28.06.97, Ruhr-Universität Bochum

Zielgruppe: Biometriker, Mediziner, Statistiker, Mathematiker, Med. Dokumentare

Referentin: Dipl.-Stat. Annette Pfahlberg, Göttingen

Teilnahmegebühren: Mitglieder: 1000,--DM, Studierende 200,--DM, Sonstige 1500,--DM

Veranstaltung: Zeitreihenanalyse

Termin, Ort: 11. - 13.06.97, Ruhr-Universität Bochum

Zielgruppe: Biometriker, Statistiker und andere Berufsgruppen, die mit der Erfassung, Auswertung und Interpretation von zeitlich geordneten Beobachtungen befaßt sind

Referent: Prof. Dr. Rainer Schlittgen, Hamburg

Teilnahmegebühren: Mitglieder: 900,--DM, Studierende 150,--DM, Sonstige 1200,--DM

Tutorium: Einführung Diabetes

Termin, Ort: 14.09.97 nachmittags, Ulm

Zielgruppe: Biometriker und Med. Dokumentare, die mit der Planung und Auswertung klinischer Studien betraut sind

Referent: Dr. med. Marcus Hompesch, Wuppertal

Teilnahmegebühren: Mitglieder: 150,--DM, Studierende 50,--DM, Sonstige 300,--DM

Tutorium: Einführung KHK

Termin, Ort: 14.09.97 vormittags, Ulm

Zielgruppe: Biometriker und Med. Dokumentare, die mit der Planung und Auswertung klinischer Studien betraut sind

Referentin: Dr. med. Claudia Hänel, Düsseldorf

Teilnahmegebühren: Mitglieder: 150,--DM, Studierende 50,--DM, Sonstige 300,--DM

Tutorium: Cross-over-designs

Termin, Ort: 14.09.97 vormittags, Ulm

Zielgruppe: Biometriker und Mediziner in der präklinischen und klinischen Forschung

Referenten: Prof. Dr. Walter Lehmacher, Dr. Ralf-Dieter Hilgers, Köln

Teilnahmegebühren: Mitglieder: 150,--DM, Studierende 50,--DM, Sonstige 300,--DM

Tutorium: Guidelines

Termin, Ort: 14.09.97 nachmittags, Ulm

Zielgruppe: Biometriker und Ärzte, die an der Planung, Durchführung und Berichterstattung klinischer Prüfungen beteiligt sind und sich über die rechtlichen Grundlagen (Gesetze bis Guidelines) informieren möchten.

Referent: Dr. Horst Nowak, Bernd Cromme, Frankfurt, Prof. Dr. Joachim Röhmel, Berlin

Teilnahmegebühren: Mitglieder: 150,--DM, Studierende 50,--DM, Sonstige 300,--DM

Tutorium: Sequentialpläne und adaptive Designs

Termin, Ort: 14.09.97 vormittags, Ulm

Zielgruppe: Interessierte an Versuchsplanung, insbesondere für klinische Versuche

Referent: Prof. Dr. Peter Bauer, Wien

Teilnahmegebühren: Mitglieder: 150,--DM, Studierende 50,--DM, Sonstige 300,--DM

Tutorium: Stabile Multivariate Auswertungsverfahren

Termin, Ort: 14.09.97, nachmittags Ulm

Zielgruppe: Biometriker mit Erfahrungen in der multivariaten Datenauswertung

Referent: Prof. Dr. Jürgen Läuter, Dr. Siegfried Kropf, Magdeburg

Teilnahmegebühren: Mitglieder: 150,--DM, Studierende 50,--DM, Sonstige 300,--DM

Veranstaltung: ICH

Termin, Ort: 29.09.97, Ruhr-Universität Bochum

Zielgruppe: Biometriker, Ärzte aus der pharmazeut. Industrie und Universitäten

Referent: Prof. Dr. Joachim Röhmel, Berlin

Teilnahmegebühren: Mitglieder: 350,--DM, Studierende 70,--DM, Sonstige 450,--DM

Veranstaltung: Resampling

Termin, Ort: 29. - 30.09.97, Ruhr-Universität Bochum

Zielgruppe: Für Statistiker und Anwender statistischer Methoden, die sich dafür interessieren, wie man anhand moderner computerintensiver Verfahren u.a.

Konfidenzintervalle berechnen und statistische Tests durchführen kann, ohne auf Verfahren der klassischen Interferenz zurückzugreifen.

Referent: Prof. Dr. Iris Pigeot-Kübler, München

Teilnahmegebühren: Mitglieder: 600,--DM, Studierende 100,--DM, Sonstige 700,--DM

(\* in Kooperation mit der postgraduellen Ausbildung „Medizinische Biometrie“ der Universität Heidelberg)

Auskunft: Walter Dieckmann, Akademie für öffentliche Gesundheit, Ruhr-Universität, D-44780 Bochum, Tel.: 0234-700-5162, Fax: 0234/709-4325, e-mail: Walter.Dieckmann@rz.ruhr-uni-bochum.de <http://www.amib.ruhr-uni-bochum.de>

\*\*\*\*\*

**Der Fachbereich Epidemiologie teilt folgende Veranstaltungstermine mit:**

Kurs: CONFERENCE: „DISEASE PREVENTION: CURRENT CONTROVERSIES“

Zeit: 24. bis 29. Mai 1997, Ort: Castelvechio Pascoli (Lucca) Italy

Teilnahmegebühr: 3600 French francs (Approx US\$ 690)

Anmeldung: Rodolfo Saracci  
European Science Foundation  
1 qual Lszay - Marnesia, 67080 Strasbourg, France  
Tel.: +33-388-767138, Fax: +33-388-366987

Kurs: SUMMER COURSE IN EPIDEMIOLOGY AND FIELD RESEARCH METHODOLOGY

Zeit: 26. Mai bis 13. Juni 1997, Ort: Umea, Sweden

Teilnahmegebühr: Approx: \$ 150 US (not including board and lodging)

Sprache: englisch  
Anmeldung: Kjerstin Dahlblom, Lena Mustonen  
Department of Epidemiology and Public Health  
Umea University, S-90185 Umea, Sweden  
Tel.: +46-90-102928, Fax: +46-90-138977

Kurs: COURSE: ECORISQUE '97 (A 3-WEEK COURSE IN ENVIRONMENTAL HEALTH SIMILAR TO THE EIS COURSES AT CDC)

Zeit: 26. Mai bis 14. Juni 1997, Ort: France - Chateau de Bierville

Teilnahmegebühr: 12000 FTTC (Approx US\$ 2300)

Sprache: French  
Anmeldung: Dr. William Dab, ENSP, 14 rue du Val d'Osne,  
94415 Saint-Maurice cedex, France  
Tel.: 01 4396-6470, Fax: 01-4396-6495

Kurs: NUTRITIONAL AND LIFESTYLE EPIDEMIOLOGY

Zeit: 29. Mai bis 6. Juni 1997, Ort: Rotterdam, The Netherlands

Teilnahmegebühr: Dfl 1750 (Approx US\$ 1000)

Anmeldung: Mrs. Gerda Bathoorn, Admissions Coordinator

The Netherlands Institute for Health Sciences, Erasmus University Medical School, Room Ee 2122  
 PO Box 1738, 3000 DR Rotterdam, The Netherlands  
 Tel.: +31-10-4 08-82 88, Fax: +31-10-4 36-59 33  
 e-mail: NIHES@nihes.fgg.eur.nl

Kurs: INTENSIVE COURSE IN EPIDEMIOLOGY AND MEDICAL STATISTICS

Zeit: 23. Juni bis 11. Juli 1997, Ort: London

Teilnahmegebühr: 1300 pounds (Approx US\$ 2690)

Anmeldung: Registry, Department of Epidemiology and Population Sciences  
 London School of Hygiene and Tropical Medicine  
 Keppel Street, London WC1E 7HT England  
 Tel.: 44-171-927-2074, Fax: 44-171-323-0638

Kurs: EUROPEAN EDUCATIONAL PROGRAMME IN EPIDEMIOLOGY:  
 10th RESIDENTIAL SUMMER COURSE

Zeit: 23. Juni bis 11. Juli 1997, Ort: Lyon (France)

Sprache: englisch

Anmeldung: Rodolfo Saracci, MD  
 European Educational Programme in Epidemiology  
 c/o International Agency for Research on Cancer  
 150 Cours Albert Thomas  
 69008 Lyon Cedex 087 (France)  
 Tel.: +33-47-27-38483, Fax: +33-47-27-3 85 75

Kurs: 8. INTERNATIONALER KURS DER EPIDEMIOLOGIE

Introduction to epidemiological methods, Advanced Epidemiological Methods, Infectious Disease Epidemiology, Genetic Epidemiology

Kursleiter: Prof. Dr. med. Hermann Brenner

Zeit: 14. bis 18. Juli 1997, Ort: Universität Ulm

Sprache: englisch/deutsch

Teilnahmegebühr: DM 600,--

Anmeldung: Renate Decker - Sekretariat Prof. Brenner  
 Universität Ulm/Abteilung Epidemiologie  
 Albert-Einstein-Allee 43  
 89081 Ulm, Germany  
 Tel.: (0731)502-64 51, Fax: (0731)502-64-53

Kurs: EPIDEMIOLOGY IN ACTION

Zeit: September bis Dezember 1997, Ort: Liverpool

Teilnahmegebühr: 3,300 Pounds Sterling (Approx US \$ 5500)

Anmeldung: Dr. Bernard J.M. Schlecht, Course covener  
 EILA, Unit for Statistics and Epidemiology  
 Liverpool School of Tropical Medicine  
 Pembroke Place, Liverpool L3 5QA UK  
 Tel.: 44-151-708-9393, Fax: 44-151-7085233  
 e-mail: schlecht@liv.ac.uk or jhowley@liv.ac.uk

Kurs: ADVANCED EPIDEMIOLOGICAL METHODS

Zeit: September (2 Weeks not yet determined), Ort: London, England

Anmeldung: Registry Short Course  
 London School of Tropical Medicine & Hygiene  
 Keppel Street, London, England WC1E7 7H7  
 Tel.: 44-61-267-6066, Fax: 41-61-267-6190

Kurs: 16th EUROPEAN COURSE IN TROPICAL EPIDEMIOLOGY

Zeit: 08. bis 18. September 1997, Ort: Heidelberg, Germany

Teilnahmegebühr: 2000,-DM (Approx US\$ 1285)

Anmeldung: Anke Nitschke  
 Dept. of Tropical Hygiene and Public Health  
 Im Neuenheimer Feld 324, 69120 Heidelberg, Germany  
 Tel.: +49-6221-564905, Fax: +49-6221-564918  
 e-mail: anke.nitschke@urz.uni-heidelberg.de

### Veranstungskalender

Diese Termin- und Thementübersicht enthält nur Veranstaltungen, von denen die Bearbeiter glauben, daß sie das Interesse der Mitglieder der GMDS und aller Leser der Mitteilungen finden. Die Übersicht dient der Information und Terminkoordination. Für die Richtigkeit der Angaben, die uns von den Veranstaltern übermittelt werden, können wir keine Gewähr übernehmen.

Jahrestagung der Gesellschaft für Medizinische Ausbildung

\* 29.05.1997 \* in Homburg/Saar, Klinik der Universität des Saarlandes

Information: Prof. Dr. F. Eitel, Chir. Klinik, Klinikum Innenstadt, Nußbaumstr. 20, 80336 München, Tel.: 089-5160-2580, Fax (089) 5160-4493

1. Fortbildungstagung der Deutschen Gesellschaft für Krankenhaushygiene

\* 05.06.1997 \* in Düsseldorf

Information: Prof. Dr. U. Junghannß, Fachhochschule Anhalt, Abt. Köthen, Bernburger Straße 52-57, 06366 Köthen, Tel.: (03496) 65-583, Fax: (03496) 212081

5. Deutscher Kongress der Deutschen Gesellschaft für Schlafforschung und Schlafmedizin

\* 12.06. - 14.06.1997 \* in Berlin

Information: Dr. I. Fietze, Schlafmedizin. Zentrum der Charité, Luisenstr. 13 a, 10117 Berlin, Tel.: (030)2802-5286, Fax: (030)2802-1441, e-mail: Fietze@rz.charite.hu-berlin.de

Arbeitstagung der Deutschen Gesellschaft für Ärztliche Hypnose und Autogenes Training (DGÄHAT)

\* 20.06. - 22.06.1997 \* in Schmitten/Taunus

Information: Tagungsbüro DGÄHAT c/o Frau D. Meyer-Schmeling, 23227 Wensin, Fax: (04559) 13 49

26. Jahrestagung der Deutschen Gesellschaft für Angiologie

\* 03.09. - 06.09.1997 in München

Information: Prof. Dr. Wolfram Theiss, 1. Med. Klinik der TU, Ismanninger Str. 22, 81675 München, Tel.: (089) 4140-2660, Fax: (089) 4140-4904

THE HEALTH OF POPULATIONS IN AN CHANGING EUROPE

IEA European Epidemiological Group und Deutsche Arbeitsgemeinschaft für Epidemiologie

\* 04.09. - 06.09.1997 \* in Münster

Information: Prof. Dr. Ulrich Keil, Prof. Dr. Hans-Werner Hense, Institut für Epidemiologie und Sozialmedizin, Domagkstr. 3, D-48129 Münster, Tel.: 49(0)251 83 5399, Fax: 49(0)251 83 5300, e-mail: HENSE@UNI-MUENSTER.DE

29. Jahrestagung der Gesellschaft für Nephrologie

\* 07.09. - 10.09.1997 \* in Aachen, Eurogress

Information: PD Dr. T.H. Ittel, Tel.: (0241) 80-89658, Fax: (0241) 8888 486

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

\* 15.09. - 18.09.1997 \* in Ulm (Donau)

Information: Tagungssekretariat GMDS '97, Abt. Biometrie und Med. Dokumentation, Universität Ulm, 89070 Ulm, Telefon: 0731-5026905, Fax: 0731-5026902, E-mail: gm97@medizin.uni.-ulm.de

WWW:<http://www.uni-ulm.de/uni/fak/medizin/biodok/gm97.html>

Tage der Elektrotechnik

Fachtagung „Automatisierungstechnik“

Fachtagung „Biomedizinische Technik“

Fachtagung „Regenerative Energien“

\* 14.10. - 15.10.1997 in Köthen

Information: Frau Dagmar Becker, Tel.: (03496)67- 487, Fax: (03496) 212081

COMMON SECURITY SOLUTIONS FOR COMMUNICATING PATIENT DATA

IMIA Working Conference on Security

\* 22.11. - 25.11.1997 \* in Osaka/Kobe, Japan

Information: Ab Bakker, Schipholweg 97, 2316 XA Leiden, The Netherlands, Fax: 31 71 5216571, e-mail: ABAKKER@HISCOM.NL (chairman SPC)

OMNICARD 1998

\* 14.01. - 16.01.1998 \* in Berlin

Information: inTIME Berlin, Seesener Straße 53, D-10711 Berlin, Tel.: 49(0)30 890 92-852, Fax: 49(0)30 890 92-854

25 Jahre Medizinische Informatik Heidelberg/Heilbronn

\* 16.01.1998 \* in Heilbronn

Festveranstaltung zum 25-jährigen Bestehen des Studiengangs

Information: Sekretariat Medizinische Informatik, Heilbronn, Tel.: (07131)504416 oder Sekretariat Medizinische Informatik, Heidelberg Tel.: 06221/567483

6. Jahrestagung Epidemiologie

44. Biometrisches Kolloquium

Deutsche Arbeitsgemeinschaft für Epidemiologie (DAE)

\* 16.03. - 19.03.1998 \* in Mainz

**Information:** Institut für Med. Statistik und Dokumentation der Johannes Gutenberg-Universität, 55101 Mainz, Tel.: 06131-173252, Fax: 06131-172968, e-mail: mainz98@imsd.uni-mainz.de

#### 6. INTERNATIONALES KOLLOQUIUM

Epidemiologie und berufliche Risiken

\* 22.04. - 24.04.1997 \* in Graz, Österreich

**Information:** Kongressbüro, Adalbert-Stifter-Straße 65, A - 1200 Wien, Tel.: 43-1-33111-537, Fax: 43-1-33111-469

### Ausschreibungen

#### **GMDS-Förderpreis für Studierende 1997**

Zur Förderung des wissenschaftlichen Nachwuchses schreibt die GMDS 1997 erneut einen Preis für herausragende Abschlußarbeiten auf den Gebieten Medizinische Informatik, Medizinische Biometrie und Epidemiologie aus.

Der GMDS-Förderpreis für Studierende ist mit 500,- DM dotiert. Der/die Gewinner können die Ergebnisse ihrer Arbeit auf der bevorstehenden Jahrestagung der GMDS präsentieren. Die Teilnahme an der Jahrestagung ist kostenlos. Kriterien zur Beurteilung der Arbeiten sind sowohl die Relevanz der Ergebnisse für die Medizin als auch die erzielten methodischen Fortschritte auf einem der Fachgebiete Medizinische Informatik, Medizinische Biometrie oder Epidemiologie.

Wer kann sich bewerben?

Absolventinnen und Absolventen eines Studiengangs einer wissenschaftlichen Hochschule (z.B. Informatik, Mathematik, Medizin, Medizinische Informatik, Statistik) mit ihrer Abschlußarbeit (z.B. Informatik-Diplomarbeit, Medizin-Dissertation), die 1995, 1996 oder 1997 abgegeben wurde.

Bitte senden Sie Ihre Bewerbung bis zum **30. Juni 1997** mit

- Ihrer Arbeit,
- Ihrem Abschlußzeugnis,
- einem Beurteilungsschreiben der Betreuerin bzw. des Betreuers Ihrer Arbeit,

in dreifacher Ausfertigung an die Geschäftsstelle der GMDS,  
Herbert-Lewin-Straße 1, 50931 Köln.

#### **GMDS-Posterpreis 1997**

Das Programmkomitee der GMDS Jahrestagung wird auch 1997 in Ulm die ausgestellten Poster bewerten und einen Posterpreis verleihen. Der Posterpreis ist mit 500,-DM dotiert.

Die Kriterien zur Beurteilung der Poster gliedern sich in zwei Gruppen:

Zum einen wird ähnlich wie beim GMDS Förderpreis für Studierende (s.o.) der Inhalt des Posters bewertet. Zum anderen wird jedoch gleichrangig die Präsentationsform beurteilt. Hier

spielen das graphische Layout, die Konzentration auf Wesentliches, der geeignete Verweis auf weiterführende Quellen, etc. eine Rolle. Poster (Plakate) müssen aus bis zu 10 m Entfernung auf sich aufmerksam machen und zum Lesen auf ca. 1 m Entfernung einladen.

### Personalia

#### Als neue Mitglieder begrüßen wir:

Dipl.-Inform. med. Monika Adamowicz Habsburgerallee 90 60385 Frankfurt Tel.: 069/496567	Dipl.-Kfm. Heribert Balks Im Mediapark 13 50670 Köln Tel.: 0221/4545430
Dr. med. Hagen Barlag Fliegel Data GmbH zur Lüne 44 37671 Höxter Tel.: 05271/68080 Fax: 05271/689844	Dipl.-Ing. Rainer Binder Hewlett-Packard GmbH Medizinelektronik Schickardstr. 4 71034 Böblingen Tel.: 07031/14-3940
Dipl.-Ing. Andreas Buchholz Niebuhrstr. 39 E 10629 Berlin Tel.: 030/3275037	Dr. med. Tilo Dahmann Im Grün 29 76199 Karlsruhe Tel.: 0721/888767
Dr. med. Jürgen Dobmeyer Alois Eckert Str. 24 60528 Frankfurt/Main Tel: 069/6666-374 Fax: 069/6666-346	Dr. med. Thomas Dobmeyer Richard-Wagner-Str. 15 63263 Neu-Isenburg Tel: 06102/31667
Dipl.-Phys. Günter Drewes Städt. Krankenhaus Wismar Radiologie Friedrich-Wolf-Straße 23966 Wismar Tel.: 03841/331360	Dipl.-Inform. Hans-Peter Eich Zweite Gewanne 24 41540 Dormagen Tel.: 02133/82453
Jörn Erdmann M.A. Gabrielenstr. 55 13507 Berlin Tel.: 030/4333707	Alexander Feidel Albert-Einstein-Str. 38 60437 Frankfurt/Main Tel.: 069/507-6279
Dr. med. Reinhard Friedl Universität Ulm Steinhövelstr. 9 89075 Ulm Tel.: 0731/502-7322	Dr. med. Veronika Gradel mibeg Institut Kaiser-Wilhelm Ring 40 50672 Köln Tel.: 0221/136662
Eva Heiler Roekstr. 7 23568 Lübeck Tel.: 0451-36035	Dr. med. Frank Hoffmann Brabanter Str. 9 50674 Köln Tel.: 0221-5101899



<p>Michael Kilian Ev. Krankenhaus Gelsenkirchen Chirurgische Klinik Menckelstr. 27 45879 Gelsenkirchen Tel.: 0209/1601109 Fax: 0209/1602678</p>	<p>Prof. Dr. Hans-Dieter Klein Adlerweg 11 88489 Wain Tel: 07353/91359</p>
<p>Dr. med. Helmut König Am Kornacker 70 35041 Marburg Tel: 06421/84376</p>	<p>Dipl.-Inform. med. Gerhard Mayer Schülerstr. 52 89522 Heidenheim Tel.: 07321/23520</p>
<p>Dipl.-Inf. med. Gerhard Mönnich Weilerstr. 11 71409 Schwaikheim Tel.: 07195/57310</p>	<p>Dr. med. Stefan Muffert Austraße 10 97222 Rimpfing Tel.: 09365/2683</p>
<p>Dr. rer. nat. Dipl.-Math. Ullrich Munzel Abt. Medizinische Statistik Georg-August-Universität Göttingen Humboldtallee 32 37073 Göttingen Tel.: 0551/399670 Fax: 0551/394995</p>	<p>Dr. med. Lara Niemeyer Mittelstraße 44 22869 Schenefeld Tel.: 040/8300568</p>
<p>Dipl.-Inform. med. Ute M. Nolte Planetenstr. 44 40223 Düsseldorf Tel.: 0211/347896</p>	<p>Johannes Rosenberger Friedrich-Ebert-Str. 20 72793 Pfullingen Tel: 07121-799501 Fax: 07121-799502</p>
<p>Dr. rer. nat. Rita Rossol Alois-Eckert Str. 24 60528 Frankfurt Tel: 069/6666374 Fax: 069/666346</p>	<p>Dr. rer. nat. Bärbel Schack Tachover Ring 7 07646 Stadtroda Tel.: 036428/41923</p>
<p>Dipl.-Stat. Sandra Schläfke Med. Institut für Umwelthygiene Heinrich-Heine-Universität Düsseldorf Auf m Hennekamp 50 40211 Düsseldorf Tel.: 0211/3389-299</p>	<p>Guido Schmidt Arthur-Strobel-Str. 12 09127 Chemnitz Tel.: 0371/741730</p>
<p>Dr. med. Birthe Schmitt-Thomas, M.P.H. Krebsregister Rheinland-Pfalz Institut f. med. Statistik und Dokumentation Universitätsklinik Mainz Obere Zahlbacher Str. 69 55131 Mainz Tel.: 06131/176710 Fax: 06131-172968</p>	<p>Dr. rer. nat. Dipl.-Stat. Norbert Schmitz Rhein. Landes- und Hochschulklinik Psychosomatische Medizin Bergische Landstr. 2 40629 Düsseldorf Tel.: 0211/9224723</p>

Dipl.-Inform. med. Peter Schurr Leichtensternstraße 27 50937 Köln Tel.: 0221-416059	Dipl.-Phys. Cord Spreckelsen Institut für Med. Informatik und Biometrie der RWTH Aachen Pauwelsstr. 30 52074 Aachen Tel.: 0241/8088870 Fax: 0241/888426
Dr. med. Dietmar Stephan GSS-Gesundheits-Service Siegen gem. GmbH Institut für Qualitätsmanagement Organisation, Fort- und Weiterbildung Friedrichstr. 128 57072 Siegen Tel.: 0271/2313313	Dipl.-Inform. Birgit Wacker Hewlett Packard gmbH Medical Products Group Europe Marketing Center Schickardstr. 4 71034 Böblingen Tel.: 07031/141367
Dipl.-Math. Sandra Ziegler Institut für Med. Biometrie und Informatik Abt. Med. Biometrie Universität Heidelberg Im Neuenheimer Feld 305 69120 Heidelberg Tel.: 06221/564371 Fax: 06221/564195	

Als studentische Mitglieder begrüßen wir

Klaus Andreas Ganser Dammstr. 70 74076 Heilbronn Tel.: 07131/166146	Okan Gür Ute Str. 14 53179 Bonn Tel.: 0228/341515
--	--

# Autorenmerkblatt der Zeitschrift „Informatik, Biometrie und Epidemiologie“

## Thematik und Veröffentlichungssprachen

1. Die Zeitschrift „Informatik, Biometrie und Epidemiologie in Medizin und Biologie“ dient der Dokumentation von theoretischen und anwendungsbezogenen Arbeiten aus dem Gesamtgebiet der Informatik, Biometrie und Epidemiologie in Medizin und Biologie. Die Zeitschrift veröffentlicht sowohl Beiträge über neu entwickelte Ansätze als auch neue Anwendungsmöglichkeiten bekannter Methoden. Zusätzlich soll dem tutoriellen Aspekt durch Übersichtsartikel Rechnung getragen werden. Der Bezug zur Medizin oder Biologie darf dabei jedoch nicht vernachlässigt werden. Die Arbeiten dürfen nicht zuvor in anderen in- und ausländischen Zeitschriften veröffentlicht worden sein. Über die Annahme entscheidet die Schriftleitung, unterstützt von Mitgliedern des Herausgeberbeirats.

2. Veröffentlichungssprache ist deutsch. Beiträge mit allgemeiner Bedeutung können auch in englischer Sprache aufgenommen werden.

## Form und Manuskripte

1. Manuskripte sind auf Diskette, sowie zusätzlich in doppelter Ausfertigung auf Papier einzureichen. Siehe auch Abschnitt „Verwendung von Disketten“. Disketten und Manuskripte sind bitte eindeutig zu kennzeichnen (Autor, Beitrag, Programm, Version). Die ausgedruckte Variante soll auf DIN A 4-Bogen einseitig mit breiten Rändern und  $1\frac{1}{2}$ -fachem Zeilenabstand in gut lesbarer Schriftgröße (i. d. R. 12 Punkt) geschrieben sein. Der linke Rand sollte mindestens 10 Anschläge breit sein.

2. Manuskripte sollen 20 Seiten nicht überschreiten. Arbeiten mit weniger als 8 Seiten können bevorzugt in Druck gegeben werden. Arbeiten mit weniger als 3 Seiten können als „Kurzmitteilung“ veröffentlicht werden. Alle Tabellen und Abbildungen sowie das Literaturverzeichnis sind in die Seitenzahlen einzubeziehen.

3. Originalarbeiten sind in die Abschnitte Zusammenfassung, Stichworte, Summary, Key-words, Einleitung (z. B. Literaturübersicht, Problemstellung), Fragestellung, Material und Methoden, Ergebnisse, Diskussion/Schlußworte und Literaturverzeichnis zu gliedern. Die Bedeutung der Ergebnisse für die Informatik, Biometrie und Epidemiologie sind in der Diskussion hervorzuheben. Bei deutschsprachigen Arbeiten ist dem Titel der Arbeit die englische Übersetzung beizufügen, bei englischsprachigen Arbeiten die deutsche. Die Schriftleitung ist für ihre Formulierung nicht verantwortlich.

4. Autorennamen werden in KAPITÄLCHEN geschrieben.

5. Abkürzungen sind nur dann zulässig, wenn sie international bekannt sind. Gegebenenfalls (nicht ggf.!) müssen sie im Text oder durch Fußnote erklärt werden.

6. Der Autorennamen mit Initialen der Vornamen bei Herren, mit voll ausgeschriebenen Vornamen bei Damen stehen unter dem Titel.

7. Am Schluß der Arbeit ist das Institut sowie eine Korrespondenzadresse (Verfasser oder Mitautor) anzugeben, die Auskünfte über die Arbeit geben kann.

8. Die Tabellen und die Legenden zu den Abbildungen sind auf getrennten Seiten, jeweils in Deutsch (und Englisch), einzureichen. Die Beschriftungen von Tabellen und Grafiken sollen ausreichend groß gewählt werden, damit sie auch bei einer auf Spaltenbreite verkleinerten Abbildung lesbar bleiben.

## Darstellungen und Tabellen

1. Die Ergebnisse können entweder in Form einer Tabelle oder als Grafik dargestellt werden, eine doppelte Wiedergabe ist unerwünscht.

2. Als Vorlage für Zeichnungen und Fotos können nur scharfe, kontrastreiche Originale angenommen werden. Fotonegative können leider nicht verwendet werden. Bei grafischen Darstellungen sind die Originale, Fotoabzüge, Dias oder gleichwertige Reproduktionen einzureichen. Bei Fotos sollte mit Aufkleber auf der Rückseite, bei Dias auf dem Rahmen, der Name des Fotografen und die Bildnummer vermerkt sein. Die dazugehörigen Bildunterschriften sind, auf einem gesonderten Blatt unter der betreffenden Nummer aufzuführen.

3. Farbbildungen können auf Kosten der Autorin/des Autors aufgenommen werden. Preisliste auf Anfrage beim Verlag.

## Verwendung von Disketten

1. Es können DOS-formatierte (3,5") und falls eine DOS-Formatierung nicht möglich ist, auch Macintosh-formatierte Disketten eingelesen werden. Bitte genau kennzeichnen!

2. Die Texte sind vorrangig in MS Word oder Tech zu erfassen, falls dies nicht möglich ist, bitte das Textprogramm mit Versions-Nummern z. B. Word Perfect 6.0, DOS oder Windows angeben und zusätzlich die Texte im reinen Textformat bzw. ASCII-Format abspeichern.

3. Auf den Disketten nur die zur Verwendung vorgesehenen Texte mitliefern, nicht benötigte Texte vorher löschen. Die Zeilen eines Absatzes sind stets endlos zu erfassen (d. h. ohne Zeilende) ohne Tab, ohne Einzüge am Anfang, ohne Trennungen bzw. Trennhilfe. Jeder Absatz ist durch einen Return zu beenden.

4. Nach Möglichkeit sollten die Autoren oder deren Schreibkräfte bereits bei Erfassung Formatierungsbeefehle eingeben. Dadurch sind Auszeichnungen (z. B. halbfett, gesperrt, unterstrichen o.ä.) bereits im Grundtext erfaßt. Ein Mustertext und die Stilvorlagen für Word können bei Verlag und Schriftleitung angefordert werden. Sofern Sonderzeichen verwendet werden, sollten diese in einer separaten Sonderzeichendatei abgespeichert werden. In ihr sollten alle in der Arbeit aufgetretenen Sonderzeichen in der Art ihrer Kodierung aufgelistet sein. Durch ein Tab abgetrennt, sollte dann die verbale Bedeutung erläutert werden. Bspl.:

&gr.a%	griechischer Alpha
&x%	Malkreuz
&zm%	Zeichen für männlich . . . .

5. Wurden Texte und Disketten an die Autoren zur Überarbeitung zurückgegeben, so sollten die Autoren ausdrücklich bestätigen, daß auch die Diskette auf den neuesten Stand gebracht wurde.

6. Folgende Bildformate können verarbeitet werden: EPS und TIFF (jeweils auch mit Macintosh-Bildschirmdarstellung abspeichern).

7. Tabellen in eine separate Datei setzen. Sie sind einzeln abzuspeichern. Im Dateinamen sollte bereits zu erkennen sein, wohin die Tabelle gehört. So könnte man zum Beispiel die 3. Tabelle vom Kapitel 1 K1\_TAB3 benennen. Tabellen werden ohne Linien gesetzt, Spalten sind durch Tab's zu trennen. Nach jeder Tabellenzeile wird ein Return gegeben. Die Anzahl der Tabulatoren in jeder Tabellenzeile muß gleich sein. (Für jede leere Tabellenspalte ist ein Tabulator zu vergeben).

#### Literatur

1. Nur wesentliche und allgemein zugängliche Literatur sollte zitiert werden, z. B. Bücher, Zeitschriftenbeiträge und zum gleichen Thema abgefaßte Dissertationen (keine Diplomarbeiten!)

2. Die zitierte Literatur wird im Literaturverzeichnis am Ende der Arbeit alphabetisch nach den Autorennamen zusammengefaßt. Im Text der Arbeit wird auf das Literaturverzeichnis durch Angabe des Autorennamens mit Erscheinungsjahr der Publikation verwiesen.

3. Bei Zeitschriften gilt folgendes Schema: Verfasserin/Verfasser (Vorname abgekürzt) – Jahr der Veröffentlichung – Titel der Abhandlung – Zeitschrift – Bandzahl oder Jahrgang (unterstrichen, im Satz halbfett) – Heft-Nr. (in Klammern) – Seitenzahlen der Arbeit oder der Seiten, auf die das Zitat direkt bezogen ist.

Beispiel: WASMUS, A., P. KINDEL, S. MATTUSSEK, H. H. RASPE, 1989: Activity and severity of rheumatoid arthritis in Hannover/FRG and in one regional referral center. *Scand. J. Rheumatol. Suppl.* 79 (5), 33–44.

4. Bei Buchveröffentlichungen gilt folgendes Schema: Verfasserin/Verfasser – Jahreszahl – Buchtitel – Verlag – Erscheinungsort – Auflage (von der 2. Auflage an) – Seitenzahl.

Beispiel: HORN, M. und VOLLANDT, R., 1995: Multiple Tests und Auswahlverfahren, G. Fischer Verlag, Stuttgart und Jena, 289 S.

5. Normalerweise wird der Titel der Arbeit in der Originalsprache zitiert, bei Sprachen mit eigenen Schriftzeichen (z. B. russisch, arabisch, japanisch) besser die englische Übersetzung und die Angabe, aus welchem Referateblatt ihr Inhalt dem Verfasser bekannt wurde.

#### Honorar/Sonderdrucke

Anstelle eines Honorars erhalten die Verfasser kostenlos 30 Sonderdrucke. Weitere Sonderdrucke können gegen Berechnung geliefert werden. Preisliste auf Anfrage beim Verlag. Wir empfehlen unseren deutschsprachigen Autoren, einen Wahrnehmungsvertrag mit der Verwertungsgesellschaft WORT, Goethestr. 49, 80336 München, abzuschließen.

#### Verlagsrechte

1. Bestand der Rechte: Der Autor versichert, daß er allein befugt ist, über die urheberrechtlichen Nutzungsrechte an seinem Beitrag, einschließlich eventueller Bild- und anderer Reproduktionsvorlagen, zu verfügen und daß der Beitrag keine Rechte Dritter verletzt.

2. Dauer der Rechte: In Erweiterung von § 38, Abs. 1, UrhG, räumt der Autor dem Verlag für die Dauer

des Urheberrechtsschutzes das räumlich und mengenmäßig unbeschränkte Recht der Vervielfältigung und Verbreitung (Verlagsrecht) bzw. der unkörperlichen Wiedergabe des Beitrags ein, auch zur Verwertung außerhalb der Zeitschrift, für die er ursprünglich bestimmt war. Die Übertragung erfolgt für die Dauer von 5 Jahren ausschließlich.

3. Umfang der Rechte: Im Rahmen von Ziffer 2 räumt der Autor dem Verlag ferner die ausschließlichen Nutzungsrechte am Beitrag ein

a) das Recht zum ganzen oder teilweisen Vorabdruck und Nachdruck – auch in Form eines Sonderdrucks –, zur Übersetzung in andere Sprachen, zu sonstiger Bearbeitung und zur Erstellung von Zusammenfassungen (Abstracts),

b) das Recht zur Veröffentlichung einer Mikrokopie-, Microfish- und Microforma-Ausgabe, zur Nutzung im Wege von Bildschirmtext, Videotext und ähnlichen Verfahren, zur Aufzeichnung auf Bild- und/oder Tonträger und zu deren öffentlicher Wiedergabe,

c) das Recht zur maschinenlesbaren Erfassung und elektronischen Speicherung auf einem Datenträger und in einer eigenen oder fremden Online-Datenbank, zum Download in einem eigenen oder fremden Rechner zur Wiedergabe am Bildschirm, sei es unmittelbar oder im Wege der Datenfernübertragung, sowie zur Bereithaltung in einer eigenen oder fremden Online-Datenbank zur Nutzung durch Dritte,

d) das Recht Teile des Beitrages (Beitragskopf, Zusammenfassung, Stichworte, Summary und Keywords) ins Internet zu stellen,

e) das Recht zu sonstiger Vervielfältigung, insbesondere durch fotomechanische oder ähnliche Verfahren und zur Nutzung im Rahmen eines sogenannten Kopienversandes auf Bestellung,

f) das Recht zur Vergabe der vorgenannten Nutzungsrechte an Dritte im In- und Ausland sowie die von der Verwertungsgesellschaft WORT wahrgenommenen Rechte einschließlich der entsprechenden Vergütungsansprüche.

Der Verlag wird über die Rechte gemäß Punkt 2 und 3a nur mit Zustimmung des Autors verfügen und sich um eine angemessene Honorierung bemühen.

#### Manuskripteinsendung

Der Manuskripteingang wird bestätigt; vor Drucklegung wird (im Inland) ein Probeabdruck versandt, in dem nur Druckfehler berichtigt werden können. Weitergehende Änderungen oder Zusätze gehen zu Lasten der Autoren.

Manuskripte sind an die Schriftleiter/Editoren zu richten, z. Zt. an:

Informatik, Biometrie und Epidemiologie  
Prof. Dr. H. Geidel,  
c/o Verlag Eugen Ulmer  
Postfach 70 05 61  
D-70574 Stuttgart, Fed. Rep. of Germany

Prof. Dr. W. Lehmacher  
Institut für Medizinische Statistik,  
Informatik und Epidemiologie  
D-50924 Köln

Dipl.-Math. Joachim Vollmar  
Pharmaceutical Research Ass. GmbH  
Besselstraße 2–4  
D-68219 Mannheim