

Belegexemplar E 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 Stuttgart
Verlag Eugen Ulmer Stuttgart

Band 26
Heft 2/1995
ISSN 0943-5581

INFORMATIK BIOMETRIE und EPIDEMIOLOGIE

IN MEDIZIN UND BIOLOGIE

Gustav Fischer Verlag Stuttgart
Verlag Eugen Ulmer Stuttgart

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

Herausgeber

P. Bauer (Wien) · W. van Eimeren (Neuherberg) · H. Geidel (Stuttgart) · R. Haux (Heidelberg)
W. Lehmacher (Köln) · J. Michaelis (Mainz) · W. Stucky (Karlsruhe) · H. Thöni (Hohenheim)
J. Vollmar (Mannheim) · H.-E. Wichmann (Wuppertal)

Wissenschaftlicher Beirat

H. Becher (Heidelberg) · J. Berger (Hamburg) · 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 2/1995

Einführung in die statistische Methodik von Äquivalenzstudien	S. Wellek	81
Evaluation der Akzeptanz eines Krankenhaus-Informationssystems bei den ärztlichen Mitarbeitern eines Universitätskrankenhauses	H. U. Prokosch, J. Dudeck	107
Die logische Struktur der ICD-10 (Systematik) und ihre Beschreibung mit SGML	M. Schopen	121
Luftverunreinigungen und Lungenkrebs	B. Pesch, K. H. Jöckel, H. E. Wichmann	134
Statistische Ansätze zur Beurteilung seltener Arzneimittelrisiken in einer Fallstudie	N. Holländer, B. Rzany, M. Mockenhaupt, E. Schöpf und M. Schumacher	154
Das wissensbasierte System ADNEXPERT zur Unterstützung der sonografischen Diagnose von Adnextumoren	J. Brüning, R. Becker, M. Entezami, V. Loy, R. Vonk, H. Weitzel, T. Tolxdorff	166
Buchbesprechungen/Bookreviews		181

Redaktion:

Schriftleiter (verantwortlich im Sinne des Presserechts); Prof. Dr. H. Geidel, c/o Verlag Eugen Ulmer, Postfach 700561, 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. 4 00 42 56

Verlag: Vertrieb und Werbung: Gustav Fischer Verlag GmbH & Co. KG, 70599 Stuttgart, Wollgrasweg 49, Tel. (07 11) 45 80 30, Postgirokonto Stuttgart 13 556, Stuttgarter Bank 45 290. Herstellung und Anzeigen: Verlag Eugen Ulmer GmbH & Co., Wollgrasweg 41, 70599 Stuttgart, Tel. (07 11) 45 07 0, Postcheckkonto Stuttgart 74 63 700, Zürich 80 47 072, Wien 1083.662 Deutsche Bank AG, Stuttgart, Kto. 14/76 878, Südwestbank AG, Stuttgart, Kto. 21 000, Herstellung Siegfried Märkisch. Verantwortlich für die Anzeigen: Dieter Boger, Alexandra Signore, Tel. (07 11) 45 07-144, z. Z. ist die Anzeigenpreisliste Nr. 10 gültig. Anzeigenschluß: am 20. der Monate Januar, April, Juli, Oktober.

Druck: Druckhaus »Thomas Mürtzer« GmbH, Neustädter Straße 1-4, 99947 Bad Langensalza, Telefon (036 03) 81 33 81/82, 84 23 21/22/23

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch

Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden.

Informatik, Biometrie und Epidemiologie in Medizin und Biologie ist eine deutsch/englischsprachige Zeitschrift, die theoretische und anwendungsbezogene Arbeiten aus dem Gesamtgebiet der Informatik, Biometrie und Epidemiologie in Medizin und Biologie enthält. Die Zeitschrift veröffentlicht sowohl Beiträge über neu entwickelte Ansätze als auch über 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. Jedes Heft enthält u. a. Editorial, Übersichten zu bestimmten Themenkomplexen, Originalarbeiten, Kurzbeiträge, Stellungnahmen, Fallbeispiele, Informationen über Programme und Programmsysteme, Informationen zu Veranstaltungen und Neuerscheinungen auf dem Buchmarkt.

Die Zeitschrift wird zur Zeit referiert in den »Biological Abstracts« und »Current Index to Statistics«.

Informatik, Biometrie und Epidemiologie in Medizin und Biologie erscheint vierteljährlich.

Der Bezugspreis beträgt jährlich DM 280,- (Einzelheft DM 84,-) zuzüglich Versandkosten. Vorzugspreis für Mitglieder der Deutschen Region der Internationalen Biometrischen Gesellschaft DM 198,- zuzüglich Versandkosten. Bestellungen nehmen jede Buchhandlung sowie die Verlage entgegen.

TUTORIAL

Einführung in die statistische Methodik von Äquivalenzstudien*)

Stefan Wellek

Kurztitel

Äquivalenzstudien

Short title

Equivalence Trials

Schlüsselwörter

Bioäquivalenz, aktiv kontrollierte Therapiestudie, einseitige Äquivalenz, Intervallinklusions-Regel, gleichmäßig bester Test, t-Test auf Äquivalenz, exakter Fisher Test auf Äquivalenz, Log-Rank-Test auf Äquivalenz

Key words

Bioequivalence, active control study, one-sided equivalence, interval inclusion rule, uniformly most powerful test, t-test for equivalence, Fisher's exact test for equivalence, log-rank test for equivalence

Zusammenfassung

Die vorliegende Übersicht über statistische Verfahren zum Nachweis von Äquivalenz richtet sich an Medizinische Biometriker ohne spezifische Vorkenntnisse auf dem Feld der Planung und Auswertung von Studien, die mit dem Ziel durchgeführt werden, die Gleichwertigkeit zweier oder mehrerer Behandlungen zu zeigen. Der Hauptakzent liegt auf der Darstellung solcher Äquivalenznachweis-Prozeduren, welche sich als hinsichtlich der Trennschärfe optimale Tests für Intervallhypothesen herleiten lassen.

*) Schriftliche Fassung eines Tutorials bei der 38. GMDS-Jahrestagung, Lübeck, 19. 9. 1993

Summary

This review article on confirmatory statistical methods of equivalence assessment is primarily meant for medical biometricians without specific experience of the subject, rather than for experts in the field of equivalence testing. The focus is on the presentation of procedures which can be established as optimal tests for interval hypotheses.

1. Grundtypen klinischer Studien mit Äquivalenz-Fragestellung

Eine klinische Studie wird generell immer dann als Äquivalenz- (ÄQ-) Studie bezeichnet, wenn sie mit der Zielsetzung durchgeführt wird, nachzuweisen, daß es zwischen zwei (oder noch mehr) Behandlungen im weitesten Sinne *keine* praktisch relevanten Unterschiede gibt.

Wie stets in sauber kontrollierten Studien wird auch hier grundsätzlich verlangt, daß die Daten einem statistischen *Test* unterzogen werden, der das Risiko eines Fehlers 1. Art auf ein hinlänglich kleines Niveau beschränkt ($\alpha = .05$ in der Mehrzahl der Anwendungen). Allerdings besteht in einer ÄQ-Studie ein *Fehler 1. Art* in einer *irrtümlichen Entscheidung zugunsten der Hypothese, daß die Behandlungen „im wesentlichen identische“ Effekte haben!*

Innerhalb der klinischen Forschung sind hinsichtlich der Häufigkeit der Durchführung die folgenden beiden Typen von ÄQ-Studien vorherrschend:

1. Vergleichende Bioverfügbarkeitsstudien (Bioäquivalenzstudien)
2. Aktiv (positiv) kontrollierte Therapiestudien.

1.1 Bioäquivalenzstudien (Comparative Bioavailability Trials)

Historisch gesehen hat das Bioäquivalenz- (BÄQ-) Problem innerhalb der Medizinischen Biometrie den Anstoß zur Beschäftigung mit Testproblemen gegeben, bei denen die *Alternativhypothese* die Gleichwertigkeit der „Behandlungen“ behauptet.

Die immense gesundheitspolitische Bedeutung des BÄQ-Problems ergibt sich aus der Tatsache, daß für die behördliche Zulassung sog. generischer Arzneimittel ein abgekürztes Antragsverfahren existiert. Dieses erläßt dem Hersteller eines Generikums den direkten Nachweis der therapeutischen Effektivität sowie der Verträglichkeit seines Produkts, indem nur gefordert wird, daß sich an *gesunden Probanden* „im wesentlichen dieselben“ Charakteristiken für die Freisetzung des Wirkstoffs in die Blutbahn ergeben wie beim Originalpräparat. Quantitative Variablen, die den Prozeß der Absorption und des Transports des Wirkstoffs an seine Angriffsorte im Organismus geeignet erfassen, nennt man *Bioverfügbarkeits-Maße* (engl.: Measures of bioavailability). Das Generikum wird als bioäquivalent mit dem Originalpräparat bezeichnet, wenn sich in einer entsprechenden Studie zeigen läßt, daß es hinsichtlich der Bioverfügbarkeit keine relevanten Unterschiede gibt.

Das Standardprotokoll einer BÄQ-Studie sieht die Überprüfung zweier Formulierungen eines Arzneimittels [„Test“ (*T*) versus „Referenz“ (*R*)], die beide denselben Wirkstoff enthalten, in einem 2-Perioden-Crossover-Versuch vor. Dabei ist bei gerader Gesamtzahl $N = 2n$ zur Verfügung stehender Probanden eine randomisierte Aufteilung in zwei gleichgroße Untergruppen vorzunehmen, derart, daß in der 1. Gruppe die Sequenz $T - R$ zur Anwendung kommt, wogegen alle n Probanden der 2. Gruppe umgekehrt in Periode 1 die Formulierung *R*, in Periode 2 die Formulierung *T* erhalten. Die in

festgelegten Zeitpunkten des Beobachtungsintervalls gemessenen Serumkonzentrationen des Wirkstoffs werden nicht als multivariate Daten analysiert, sondern vorab auf skalare Werte reduziert. Als wichtigste eindimensionale Zielkriterien werden dabei die Fläche unter dem gesamten Profil (AUC) und das Maximum der gemessenen Konzentrationen (C_{\max}) betrachtet. Manchmal interessiert außerdem noch die Zeit bis zum Erreichen von C_{\max} . Sie wird als t_{\max} bezeichnet und ist ein Maß für die Geschwindigkeit der Wirkstoffabsorption.

Bei der statistischen Auswertung der Daten einer solchen Studie wird von vielen Biometrikern ein parametrischer Ansatz favorisiert (vgl. [11, Abschn. 2]), der von der Grundannahme ausgeht, daß die *logarithmierten* Bioverfügbarkeitswerte das bekannte, von Grizzle [6] vorgeschlagene parametrische Modell für die Analyse eines 2-Perioden-Crossover-Versuchs erfüllen. Bezeichnet X_i bzw. Y_j für den i -ten Probanden mit Sequenz $T - R$ bzw. für den j -ten Probanden mit Sequenz $R - T$ den logarithmierten Quotienten aus der in Periode 1 und der in Periode 2 gefundenen Bioverfügbarkeit, sind (X_1, \dots, X_n) und (Y_1, \dots, Y_n) unter dem genannten Modell unverbundene Stichproben aus zwei Normalverteilungen mit homogenen Varianzen. Die Differenz zwischen den Erwartungswerten dieser Verteilungen ist die doppelte Differenz der logarithmischen Arzneimittelformulierungs-Effekte.

Das statistische Grundproblem bei der Auswertung einer standardmäßigen BÄQ-Studie besteht also darin, daß ein *Test auf Äquivalenz zweier Normalverteilungen mit gleicher* (wenn auch unbekannter) *Varianz* benötigt wird, aus denen unabhängige Stichproben vorliegen.

1.2 Aktiv (positiv) kontrollierte Therapiestudien

Es handelt sich um einen relativ neuartigen Studientyp, dem aber innerhalb der klinischen Forschung immer größere Bedeutung zukommt.

Klassischerweise sind klinische Therapiestudien im einfachsten, d. h. zweiarmligen Falle so angelegt, daß im ersten der beiden Studienarme die zu prüfende neue Therapie verwendet wird, während die Patienten im Kontrollarm entweder überhaupt keine Medikation oder Plazebo erhalten (Negativkontrolle).

Die moderne Medizin hält für eine beträchtliche Anzahl unterschiedlichster Erkrankungen Behandlungsverfahren von bedrückender oder sogar hoher Effektivität bereit. In Therapiestudien, die sich auf ein derartiges Krankheitsbild beziehen, wäre es aus ethischen Gründen unannehmbar, die zu beurteilende neue Therapie gegen Plazebo oder überhaupt keine Behandlung zu vergleichen. Stattdessen wird in der Kontrollgruppe eines der als wirksam bekannten etablierten Behandlungsverfahren angewandt. Normalerweise wird dabei eine Steigerung der Effektivität gar nicht mehr erwartet. Vielmehr zielt man darauf ab, *unter Verbesserung der Verträglichkeit eine annähernd gleich gute Wirkung* zu erreichen.

Inhaltlich gesehen liegt der Hauptunterschied zu BÄQ-Studien darin, daß das Zielkriterium jetzt das Ansprechen von *Patienten* mit einschlägiger Indikation auf die Behandlung ist, nicht eine pharmakokinetische Variable, die bei gesunden Probanden registriert wird.

In statistischer Hinsicht unterscheiden sich Studien zum Nachweis therapeutischer Äquivalenz von BÄQ-Studien vor allem dadurch, daß sich der Äquivalenztest sehr oft auf Zielvariablen zu beziehen hat, deren Verteilung völlig anders strukturiert ist als eine Gauß- oder eine Lognormal-Verteilung. Besonders häufig sind in aktiv kontrollierten Therapiestudien Situationen, in denen man Vergleiche durchzuführen hat zwischen

- Responderraten (i. e. binomialen Proportionen)
- Kaplan-Meierschen Überlebensfunktionen.

2. Verfahren für die Auswertung von Studien mit einseitiger Problemstellung

Nicht selten wird es als ausreichend angesehen, in einer aktiv kontrollierten Therapiestudie nachzuweisen, daß der Testarm gegenüber der Positivkontrolle nicht in relevantem Ausmaß „abfällt“. Im Sinne dieser Betrachtungsweise leistet die zu beurteilende Therapie auch oder sogar erst recht dann das Gewünschte, wenn sie nennenswert besser abschneidet als die Referenzbehandlung. Obwohl dies terminologisch ziemlich irreführend ist, hat es sich innerhalb der Medizinischen Biometrie eingebürgert, bei einer solchen einseitigen Formulierung der zu prüfenden Hypothesen ebenfalls von einem ÄQ-Problem zu sprechen. Sinnvoller und klarer scheint es, zwischen uneigentlichen einseitigen ÄQ-Problemen und ÄQ-Problemen im strikten Sinne zu unterscheiden. Letztere sind von zweiseitiger Struktur, indem die Alternativhypothese, auf deren Absicherung man abzielt, aussagt, daß die Test-Behandlung weder wesentlich schlechter noch wesentlich besser ist als die Referenz-Behandlung.

Für einseitige ÄQ-Testprobleme gibt es in vielen Fällen eine ausgesprochen einfache Lösung, nämlich immer dann, wenn vorausgesetzt werden kann, daß zwischen den zu beurteilenden Verteilungen ausschließlich Lokationsunterschiede bestehen (Gleichheit der Histogramme bzw. Dichtefunktionen bis auf eventuelle Verschiebungen).

2.1 Einseitige Äquivalenztests für Lokationsprobleme

Wir beschränken uns auf die Darstellung des Vorgehens bei Problemen mit zwei **verbundenen Stichproben**. (Die Übertragung auf Probleme mit unverbundenen Stichproben ist weitgehend selbsterklärend).

Beim *verbundenen Vergleich* bestehen die auszuwertenden Daten bekanntlich aus Paaren $(X_1, Y_1), \dots, (X_n, Y_n)$, wobei X_i bzw. Y_i das Ergebnis der Anwendung von Behandlung A bzw. B auf die i -te Beobachtungseinheit nach einem geeigneten quantitativen Zielkriterium beschreibt. Im übrigen sollen die Bedingungen für die Beobachtung von X_i und Y_i völlig homogen sein, so daß die Differenz $D_i = X_i - Y_i$ direkt als Wirkungsunterschied zugunsten von Behandlung A im i -ten Einzelfall interpretiert werden kann. Von einem Lokationsproblem spricht man hier dann, wenn außerdem vorausgesetzt wird, daß ein etwaiger Wirkungsunterschied zwischen A und B in der zugrundeliegenden Population sich so manifestiert, daß die Verteilung der D_i um eine bestimmte Strecke θ vom Nullpunkt weg verschoben wird, bei ansonsten unveränderter Gestalt der Verteilung ($\theta > 0 \Leftrightarrow$ Tendenz zu „besseren“ Werten unter Behandlung A). In der um 0 symmetrischen Grundform sei die Verteilung der Differenzvariablen D_i stetig.

Als konventionelles einseitiges Testproblem betrachtet man hier natürlich das Hypothesenpaar $H_1^{\circ}: \theta \leq 0$ gegen $K_1^{\circ}: \theta > 0$. Im Unterschied hierzu läßt man beim einseitigen ÄQ-Nachweis zu, daß Behandlung A „etwas schlechter“ sein darf als Behandlung B , daß also die Verteilung der Differenzen D_i in der Grundgesamtheit geringfügig in die negative Richtung verschoben ist. Man formuliert also anstelle von $(H_1^{\circ}, K_1^{\circ})$ das modifizierte Testproblem

$$H_1: \theta \leq -\varepsilon \quad \text{gegen} \quad K_1: \theta > -\varepsilon. \quad (1)$$

Dabei ist ε eine tendenziell kleine positive Konstante, die genau wie das Signifikanzniveau α bereits im Studienprotokoll festgelegt werden muß. Sie markiert sozusagen die Toleranzgrenze für den Effizitätsverlust, den man gegenüber der Referenzbehandlung noch in Kauf zu nehmen bereit ist.

Wenn θ die Bedeutung eines Lokationsparameters hat und ein Test für das herkömmliche einseitige Problem $(H_1^{\circ}, K_1^{\circ})$ ausgewählt worden ist, läuft dessen Umwandlung in einen Test für das einseitige ÄQ-Problem (1) wie folgt. Der Ausgangstest sei bestimmt durch die Vorschrift „Lehne H_1° genau dann ab, wenn $T(D_1, \dots, D_n) > c_{\alpha}$ gefunden wird“, wobei

T eine geeignete Funktion von n Argumenten darstellt (die „Prüfgröße“ des zugehörigen Tests) und c_α den oberen 100α -Prozentpunkt der Verteilung der Zufallsvariablen $T(D_1, \dots, D_n)$ unter $\theta = 0$ bezeichnet. Andererseits liegt, wenn wir $\tilde{D}_i = D_i + \varepsilon$ und $\tilde{\theta} = \theta + \varepsilon$ setzen, für $(\tilde{D}_1, \dots, \tilde{D}_n)$ dieselbe statistische Grundsituation vor wie vorher, jedoch mit $\tilde{\theta}$ anstelle von θ . Außerdem ist unser einseitiges ÄQ-Problem $H_1: \theta \leq -\varepsilon$ gegen $K_1: \theta > -\varepsilon$ offenbar logisch äquivalent mit $\tilde{H}_1: \tilde{\theta} \leq 0$ gegen $\tilde{K}_1: \tilde{\theta} > 0$, so daß die Verschiebung des Hypothesenrandes in dem primär vorgelegten Testproblem (H_1, K_1) rückgängig gemacht wird, indem wir die Variablen D_i einer gegenläufigen Verschiebung derselben Größe unterziehen. D. h., wir erhalten hier einen Test für das einseitige ÄQ-Problem (H_1, K_1) einfach, indem wir mit den gegenläufig verschobenen Variablen $\tilde{D}_i = D_i + \varepsilon$ die zugehörige nichtverschobene Hypothese $\tilde{\theta} \leq 0$ testen. Der so entstehende Test ist evidenterweise bestimmt durch die Entscheidungsregel: „Lehne die Nullhypothese $H_1: \theta \leq -\varepsilon$ dann und nur dann ab, wenn $T(D_1 + \varepsilon, \dots, D_n + \varepsilon) > c_\alpha$ “, wobei T und c_α genauso zu berechnen sind wie vorher.

Beispiel 1. Studie über mögliche Effekte des kalorienfreien Fettersatzes Olestra auf die Absorption hochlipophiler Kontrazeptiva. Doppelblinder, placebokontrollierter Versuch mit 28 gesunden praemenopausalen Probandinnen, die während der vierwöchigen Verumphase in einer Spezialdiät eine tägliche Dosis von 18 gm Olestra zu sich nahmen. In der gleichfalls über 4 Wochen sich erstreckenden Placebophase wurden die gleichen Nahrungsmittel in einer Herstellung verabreicht, bei der Olestra vollständig durch herkömmliche Triglyzeride ersetzt war. Während beider Versuchszyklen nahmen die Probandinnen als orales Kontrazeptivum eine Kombination aus 300 µg Norgestrel und 30 µg Ethynilestradiol ein. In der Mitte beider Zyklen (12.–14. Tag) wurden über ein 12-stündiges Meßintervall hinweg die Serumspiegel dieser Substanzen bestimmt. Die Tabelle 1 enthält die Meßergebnisse für Norgestrel mit C_{max} als pharmakokinetischem Parameter. U. a. sollte gezeigt werden, daß unter der Olestra-Diät die Bioverfügbarkeit von Norgestrel nicht in relevantem Ausmaß reduziert wird. Legt man den noch tolerierbaren medianen Bioverfügbarkeitsverlust auf $\varepsilon = 1.5$ (ng/ml) fest und benutzt einen Test auf der Basis der Vorzeichen-Rang-Statistik (verbundenen Wilcoxon-Test), hat man die folgenden Rechenschritte auszuführen: Man ermittelt für jedes $i = 1, \dots, 28$

Tabelle 1: Maximale Norgestrel-Serumspiegel (ng/ml) von 28 Frauen unter Olestra-Diät (X_i) sowie unter triglyzeridhaltiger normaler Ernährung (Y_i). (Daten aus [13].)
 $\tilde{D}_i = D_i + \varepsilon$; $\tilde{R}_i^+ = \text{Rang von } i\text{-ter Probandin bezüglich } |\tilde{D}_i|$; $\varepsilon = 1.5$

i	X_i	Y_i	\tilde{D}_i	\tilde{R}_i^+	i	X_i	Y_i	\tilde{D}_i	\tilde{R}_i^+
1	6.03	6.62	0.91	12	15	11.81	11.19	2.12	21
2	5.62	6.78	0.34	5	16	8.72	9.55	0.67	9
3	6.93	6.85	1.58	18	17	7.01	5.53	2.98	26
4	5.86	8.09	-0.73	11	18	7.13	6.71	1.92	20
5	8.91	9.18	1.23	15	19	6.56	6.53	1.53	17
6	5.86	7.47	-0.11	1	20	4.22	5.39	0.33	4
7	9.43	9.90	1.03	13	21	4.13	4.92	0.71	10
8	5.30	4.29	2.51	22	22	6.57	9.92	-1.85	19
9	4.99	3.80	2.69	24	23	8.83	10.51	-0.18	2
10	6.12	7.01	0.61	8	24	9.05	10.15	0.40	6
11	12.45	9.53	4.42	28	25	9.31	9.55	1.26	16
12	5.48	6.39	0.59	7	26	7.67	8.95	0.22	3
13	6.04	4.63	2.91	25	27	7.66	6.63	2.53	23
14	8.32	5.54	4.28	27	28	5.45	8.01	-1.06	14

die um $\varepsilon = 1.5$ verschobene intraindividuelle Differenz \tilde{D}_i [s. Tab. 1, 4. Spalte] sowie den Rang \tilde{R}_i^+ der i -ten Probandin bezüglich $(|\tilde{D}_1|, \dots, |\tilde{D}_n|)$ [s. Tab. 1, 5. Spalte]. Für die Summe der Ränge von Probandinnen mit positivem Vorzeichen von \tilde{D} finden wir $\tilde{V}_s^+ = 359$. Unter $\theta = -\varepsilon = -1.5$ ist \tilde{V}_s^+ asymptotisch normalverteilt mit Erwartungswert $E_0(\tilde{V}_s^+) = (1/4) \cdot 28 \cdot 29 = 203$ und Varianz $\text{Var}_0(\tilde{V}_s^+) = (1/24) \cdot 57 \cdot 28 \cdot 29 = 1928.5$. Entsprechend liefert die übliche Normalapproximation mit Stetigkeitskorrektur als p -Wert die Zahl $p = \Phi[(203 - 359 + .5)/\sqrt{1928.5}] = \Phi[-3.5410] = .0002$. Also führt der modifizierte Vorzeichen-Rang-Test auf einseitige ÄQ mit den Daten der Tabelle 1 auf dem 1%-Niveau zu einem positiven Ergebnis. D. h., die Resultate der Studie unterstützen die Annahme, daß es unter Olestra-Diät zu keiner wesentlichen Beeinträchtigung der Bioverfügbarkeit (gemessen als C_{\max}) von Norgestrel kommt.

2.2 Exakter Test auf einseitige Äquivalenz zweier Binomialverteilungen

Der im vorangehenden Abschnitt vorgestellte Ansatz zum Testen auf Äquivalenz im einseitigen Sinne deckt bei weitem nicht alle relevanten Anwendungssituationen ab. Eines der praktisch wichtigsten einseitigen ÄQ-Probleme, die sich nicht durch einfache Translation der Daten auf gewöhnliche einseitige Testprobleme reduzieren lassen, ergibt sich beim Vergleich zweier Binomialverteilungen, aus denen unverbundene Stichproben vorliegen, anhand der *Odds-Ratio als Hauptparameter*. Dieses Problem stellt sich immer dann, wenn im Parallelgruppen-Design die zu vergleichenden Behandlungen A und B nach einem summarischen Erfolgskriterium mit zwei möglichen Ausprägungen beurteilt werden sollen und die Durchführung eines exakten Tests angestrebt wird. Zwar hat sich in der klinischen Forschung anders als in der Epidemiologie der Gebrauch der Odds-Ratio als natürliches Maß für die Unterschiedlichkeit zweier Binomialverteilungen noch nicht einheitlich durchgesetzt. Jedoch gibt es überzeugende Argumente dafür, daß dieses oft als künstliche mathematische Konstruktion kritisierte Effektmaß auch von der Interpretation her letztlich sinnvoller ist als die auf den ersten Blick plausible Differenz der Binomialparameter oder auch das relative Risiko p_1/p_2 (für eine ausführliche Diskussion vgl. [4, pp. 44–45]).

Die Daten, die man dann für die Beantwortung der Frage nach einseitiger Äquivalenz von A mit B zur Verfügung hat, lassen sich ohne statistische Informationsverluste reduzieren auf die unter beiden Behandlungen beobachteten Häufigkeiten von Respondern und Nonrespondern. Eine handliche Darstellung dieser Häufigkeiten ergibt sich im wohlbekannten Schema einer 2×2 -Kontingenztafel, das zwecks Fixierung der Notation in Tabelle 2 nochmals explizit aufgeschrieben ist. Voraussetzungsgemäß ist dabei (X, Y) ein unabhängiges Paar binomialverteilter Zufallsvariablen, wobei die Verteilung von X durch die Konstanten m und p_1 , diejenige von Y durch n und p_2 bestimmt ist. Üblicherweise drückt man diesen Sachverhalt in symbolischer Form aus, indem man $X \sim \mathcal{B}(m, p_1)$ und $Y \sim \mathcal{B}(n, p_2)$ schreibt. Gesucht wird ein exakter Test für

$$H_1: \varrho \leq 1 - \varepsilon \quad \text{gegen} \quad K_1: \varrho > 1 - \varepsilon \quad (0 < \varepsilon < 1, \text{ fest}), \quad (2)$$

mit

$$\varrho = \frac{p_1(1 - p_2)}{(1 - p_1)p_2}. \quad (3)$$

Man erhält eine optimale Lösung dieses Problems, indem man den Konstruktionsansatz des nach R. A. Fisher benannten exakten Tests auf Homogenität der Binomialverteilungen $\mathcal{B}(m, p_1)$ und $\mathcal{B}(n, p_2)$ geeignet verallgemeinert. Der optimale (gleichmäßig beste unverfälschte – siehe Lehmann [9, pp. 154–155]) Test auf einseitige ÄQ von Behandlung A

Tabelle 2: Kontingenztafel-Schema für den Vergleich zweier Behandlungen A und B im Parallelgruppen-Design mit binärem Zielkriterium

Behandlung	Response		Σ
	+	-	
A	X (p_1)	$m - X$ ($1 - p_1$)	m (1.00)
B	Y (p_2)	$n - Y$ ($1 - p_2$)	n (1.00)
Σ	S	$N - S$	N

mit Behandlung B ist wie der exakte Fisher-Test eine bedingte Prozedur, gegeben den jeweiligen Wert $s \in \{0, 1, \dots, N\}$ der Spaltensumme $S = X + Y$. Der bedingte p -Wert $p(x|s)$ der unter A erhaltenen Anzahl x von Respondern berechnet sich anhand einer sogenannten erweiterten hypergeometrischen Verteilung. D. h., es gilt für beliebiges x mit $\max\{0, s - n\} \leq x \leq \min\{s, m\}$ (andere Werte kann X bei gegebenen s nicht annehmen):

$$p(x|s) = \sum_{j=x}^m \binom{m}{j} \binom{n}{s-j} (1-\varepsilon)^j \bigg/ \sum_{j=\max\{0, s-n\}}^{\min\{s, m\}} \binom{m}{j} \binom{n}{s-j} (1-\varepsilon)^j. \quad (4)$$

Die manuelle Berechnung von $p(x|s)$ gemäß (4) ist mit Rücksicht auf die benötigten Binomialkoeffizienten nur für sehr kleine Werte der Stichprobenumfänge m und n praktikabel. Andererseits ist die Erstellung eines Programms für die erweiterte hypergeometrische Verteilung kaum mehr als eine Routineübung, und es enthalten alle Ausgaben der Version 6 von SAS eine vordefinierte Funktion für die exakte Berechnung der zugehörigen Verteilungsfunktion (benannt als „prophyr“).

Beispiel 2. Studie über die Effkazität zweier Antibiotika in der Therapie der Streptokokken-Pharyngitis. Um die Anwendung des optimalen Tests auf einseitige ÄQ zweier Binomialverteilungen im Parallelgruppen-Design zu illustrieren, reanalysieren wir einen Teil der Daten einer vergleichenden Phase-IV-Studie über die Wirksamkeit von Clarithromycin und Erythromycin in der Therapie der Streptokokken-Pharyngitis [16]. Da die Ergebnisse von In-vitro-Analysen und tierexperimentelle Daten von vorneherein für eine verringerte Toxizität von Clarithromycin sprachen, hätte bezüglich des klinischen Behandlungserfolges bereits die einseitige ÄQ mit Erythromycin als Argument für die bevorzugte Anwendung des neuartigen Makrolids ausgereicht. Als Grundlage für die Beurteilung der klinischen Effkazität werden in der Arbeit für die primär interessierende Gruppe der unter 65-jährigen Patienten die in Tabelle 3 aufgeführten beobachteten Responderhäufigkeiten angegeben.

Führt man den Test mit diesen Daten durch unter der Vorgabe $\varepsilon = .50$, errechnet sich der p -Wert anhand von (4) in SAS zu: $p(98|195) = 1 - \text{prophyr}(213, 106, 195, 97, .50) = .049933$. Danach lassen die Daten der Tabelle 3 auf dem 5%-Niveau gesicherter Weise den Schluß zu, daß die „Chancen“ (ausgedrückt in der Form $a:b$) für eine Response bei Therapie mit Clarithromycin nicht um mehr als 50% unter diejenigen absinken, die bei Applikation der „klassischen“ Therapie gegeben sind.

Tabelle 3: Therapieerfolge unter Clarithromycin (Beh. A) und Erythromycin (Beh. B) bei unter 65-jährigen Patienten mit β -hämolytischer Streptokokken-Pharyngitis. (Daten aus [16, Table II].)

Behandlung	Response		Σ
	favourable	nonfavour.	
A	98 (92.45%)	8 (7.55%)	106 (100.0%)
B	97 (90.65%)	10 (9.35%)	107 (100.0%)
Σ	195	18	213

3. Intervallinklusions-Regeln als Tests auf Äquivalenz im engeren Sinne

ÄQ-Probleme im engeren Sinne unterscheiden sich von Testproblemen der im vorangehenden Kapitel behandelten Art durch die Tatsache, daß die *Alternativhypothese*, deren Verifizierung anhand der auszuwertenden Daten angestrebt wird, einem *beidseitig beschränkten Bereich* entspricht.

Bei der Formulierung der Hypothesen geht man auch hier generell aus von einem reellwertigen (skalaren) Verteilungsparameter θ , welcher den interessierenden Effekt in den zugrundeliegenden Populationen sinnvoll quantifiziert. Z. B. scheint für den wichtigen Spezialfall des einfachen Parallelgruppen-Designs (ÄQ-Prüfung von zwei Behandlungen A und B, die in unverbundenen Stichproben appliziert werden), die Festlegung $\theta = \mu_1 - \mu_2$ naheliegend, mit μ_1 bzw. μ_2 als einem Maß für die Lokation der A und B entsprechenden Meßwertverteilung.

Die ÄQ-Hypothese, deren Plausibilität anhand der Studiendaten beurteilt werden soll, behauptet die Gleichheit von θ mit einem geeigneten Sollwert θ_0 „bis auf praktisch irrelevante Abweichungen“. Die Menge solcher noch zu tolerierenden Abweichungen zwischen θ und θ_0 wird dabei spezifiziert als ein genügend enges, θ_0 enthaltendes Intervall $(\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2)$ mit $\varepsilon_1, \varepsilon_2 > 0$ [*hypothetisches Äquivalenzintervall*]. Im Falle des Parallelgruppen-Designs mit $\theta = \mu_1 - \mu_2$ wählt man fast immer $\theta_0 = 0$ und das ÄQ-Intervall sehr oft symmetrisch um θ_0 , i.e. in der Form $(-\varepsilon, \varepsilon)$, mit $\varepsilon > 0$. Analog zur Vorgehensweise beim einseitigen ÄQ-Nachweis sind hier die *beiden ÄQ-Grenzen* $\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2$ im jeweiligen Studienprotokoll, i.e. *ohne Kenntnis der Daten festzulegende Konstanten*.

Beim ÄQ-Nachweis im engeren Sinne wird demnach ein statistisches Entscheidungsverfahren benötigt, welches einen gültigen, d. h. das jeweilige Signifikanzniveau α garantierenden Test ergibt für die *Nullhypothese*

$$H: \theta \leq \theta_0 - \varepsilon_1 \vee \theta \geq \theta_0 + \varepsilon_2 \quad (5.a)$$

der *Nichtäquivalenz*, gegen die *ÄQ-Aussage*

$$K: \theta_0 - \varepsilon_1 < \theta < \theta_0 + \varepsilon_2 \quad (5.b)$$

als *Alternativhypothese*. Ein solches Entscheidungsverfahren braucht nicht primär die Form eines Tests im gewohnten Sinne zu haben, sondern kann z. B. auch eine Bayes-Regel sein, von der sich zusätzlich zeigen läßt, daß die „objektive Wahrscheinlichkeit“ einer falschen Entscheidung für Äquivalenz die Schranke α nicht überschreitet. (Auf Bayessche Tests für ÄQ-Probleme wird allerdings im Rahmen dieser Einführung nicht eingegangen.)

Beachte: Die Alternativhypothese K ist für kein noch so kleines ε_1 bzw. ε_2 *identisch* mit der Nullhypothese $H_0: \theta = \theta_0$ des zugehörigen zweiseitigen Testproblems. Es werden also beim Übergang vom herkömmlichen zweiseitigen Testproblem zum ÄQ-Problem die Hypothesen nicht nur vertauscht, sondern zusätzlich mehr oder minder geringgradig modifiziert. Eine Formulierung von ÄQ-Problemen, bei der unter der Alternativhypothese statt des Intervalls $(\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2)$ die Einpunktmenge $\{\theta_0\}$ für den Hauptparameter spezifiziert würde, wäre statistisch von vornherein sinnlos. In allen Verteilungsfamilien, die für konkrete Anwendungen statistischer Methoden von Interesse sind, hängt nämlich die Gütefunktion jedes Tests auf stetige Weise von θ ab. Für eine stetige Gütefunktion $\theta \mapsto \beta(\theta)$ folgt aber aus $\beta(\theta) \leq \alpha \forall \theta \neq \theta_0$ offensichtlich, daß auch $\beta(\theta_0) \leq \alpha$ erfüllt sein muß. Damit gäbe es beim Testen gegen H_0 als Alternativhypothese keine Möglichkeit, den trivialen Test, der unabhängig von den Daten mit Wahrscheinlichkeit α eine randomisierte Entscheidung gegen die Nullhypothese trifft, durch ein vernünftiges Entscheidungsverfahren zu ersetzen.

Der noch immer *gebräuchlichste Lösungsansatz für statistische Testprobleme von der Form (5) basiert auf einer Bereichsschätzung* (i. e. der Konstruktion eines Konfidenzintervalls) für den Hauptparameter θ . Der Grundidee nach geht dieser Ansatz auf eine vergleichsweise unscheinbare Arbeit (geschrieben in der Form eines „Letter to the editor“ in einer pharmazeutischen Zeitschrift) von Westlake [24] zurück. In der Nachfolge ist das Intervalleinschluß-Prinzip von einer Reihe von Autoren [7, 10–12, 18, 19, 25–27] in mindestens drei verschiedenen Versionen als Lösung für ÄQ-Entscheidungsprobleme empfohlen und vielfach praktisch angewendet worden, hauptsächlich in Zusammenhang mit Bioäquivalenzprüfungen.

In der bezüglich der Trennschärfe effizientesten Version, auf deren Dartellung wir uns hier beschränken wollen, arbeiten solche Intervalleinschluß-Tests wie folgt. Es sei $\underline{\theta}(\mathbf{X}; \alpha)$ bzw. $\bar{\theta}(\mathbf{X}; \alpha)$ eine untere bzw. obere Konfidenzschranke für θ zum *selben einseitigen Konfidenzniveau* $1 - \alpha$. Unter \mathbf{X} verstehen wir dabei die Gesamtheit aller innerhalb des auszuwertenden Experiments vorgenommenen Beobachtungen, d. h. einen zufälligen Vektor, dessen Dimension mindestens so groß ist wie die Summe der Stichprobenumfänge [z. B. gilt im einfachen Parallelgruppen-Design $\mathbf{X} = (X_1, \dots, X_m, Y_1, \dots, Y_n)$]. Die Nullhypothese $H: \theta \leq \theta_0 - \varepsilon_1 \vee \theta \geq \theta_0 + \varepsilon_2$ der Nichtäquivalenz wird genau dann abgelehnt, wenn das Konfidenzintervall mit den Grenzen $\underline{\theta}(\mathbf{X}; \alpha)$, $\bar{\theta}(\mathbf{X}; \alpha)$ vollständig im hypothetischen ÄQ-Intervall enthalten ist, wenn also zugleich $\underline{\theta}(\mathbf{X}; \alpha) > \theta_0 - \varepsilon_1$ und $\bar{\theta}(\mathbf{X}; \alpha) < \theta_0 + \varepsilon_2$ erfüllt ist.

Zwar beträgt das zweiseitige Konfidenzniveau von $(\underline{\theta}(\mathbf{X}; \alpha), \bar{\theta}(\mathbf{X}; \alpha))$ unter den genannten Voraussetzungen lediglich $1 - 2\alpha$. Dennoch hält der zugehörige Test stets das Signifikanzniveau α ein, nicht nur 2α . Trotz der auch für ÄQ-Tests vorhandenen Durchsichtigkeit des Zusammenhangs zwischen Konfidenz- und Signifikanzniveau blieb die Möglichkeit, von einem $(1 - 2\alpha)$ -Konfidenzintervall auszugehen, ohne im zugehörigen Test auf Äquivalenz das Niveau α zu überschreiten, in den ersten Arbeiten zur Anwendung des Intervallinklusions-Prinzips auf das Bioäquivalenznachweis-Problem unbeachtet (vgl. hierzu [27]).

Auf einen anderen für das Verständnis des Intervalleinschluß-Prinzips wichtigen Zusammenhang hat im Spezialfall des Zweistichproben- t -Tests SCHUIRMANN [17] hingewiesen. *Allgemein ist ein Intervalleinschluß-Test im oben erklärten Sinne logisch äquivalent mit einer „Kombination“ von zwei einseitigen Tests.* Wie sich nämlich aus der obigen Definition eines Intervalleinschluß-Tests unmittelbar ergibt, lehnt ein solches Verfahren die Nullhypothese genau dann ab, wenn sowohl der Test von $\theta \leq \theta_0 - \varepsilon_1$ gegen $\theta > \theta_0 - \varepsilon_1$ auf der Basis der Statistik $\underline{\theta}(\mathbf{X}; \alpha)$ als auch derjenige von $\theta \geq \theta_0 + \varepsilon_2$ gegen $\theta < \theta_0 + \varepsilon_2$ mit $\bar{\theta}(\mathbf{X}; \alpha)$ als Prüfgröße signifikant wird.

Beispiel 3. *Bioäquivalenz-Studie mit zwei Formulierungen von Nifedipin.* Die wohl häufigste Anwendung für eine ÄQ-Test nach dem Intervallinklusions-Prinzip ist der Bioäquiva-

lenz-Nachweis auf der Basis von Konfidenzgrenzen für die Differenz der Erwartungswerte zweier Gauß-Verteilungen mit homogenen Varianzen.

Bezeichnet A_{i1} bzw. A_{i2} die beim i -ten Probanden der Gruppe 1 (Sequenz $T - R$) in Periode 1 bzw. in Periode 2 gefundene Bioverfügbarkeit, nimmt das eingangs [vgl. Abschn. 1.1] kurz beschriebene parametrische Modell an, daß die logarithmierten Periodenquotienten $X_i \equiv \log(A_{i1}/A_{i2})$ eine Stichprobe aus einer Normalverteilung mit Erwartungswert $\log(\mu_T^*/\mu_R^*) + \log(\pi_1/\pi_2)$ sowie Varianz σ^2 darstellen. Hierbei beschreiben die Konstanten μ_T^* , μ_R^* den Effekt der Wirkstoffformulierung in der zugrundeliegenden Population von Probanden, π_1 , π_2 denjenigen der Versuchsperiode. In der zweiten Gruppe (Sequenz $R - T$) berechnet man ebenfalls für jeden Probanden den Logarithmus des Quotienten aus dem Ergebnis der Bioverfügbarkeitsmessungen in Periode 1 und Periode 2 (in dieser Reihenfolge). Wählt man für diese n Größen die Bezeichnung Y_1, \dots, Y_n , nimmt das standardmäßige parametrische Modell für die Analyse von BÄQ-Studien weiterhin an, daß die Y_j eine Stichprobe aus einer zweiten Normalverteilung ergeben, welche die gleiche (unbekannte) Varianz σ^2 besitzt wie die erste, aber den Erwartungswert $\log(\mu_R^*/\mu_T^*) + \log(\pi_1/\pi_2)$.

In den neuerdings revidierten behördlichen Richtlinien für die Durchführung vergleichender Bioverfügbarkeitsstudien wird ein statistischer Nachweis dafür gefordert, daß die wahren Formulierungseffekte μ_T^* und μ_R^* im sog. 80–125%-Bereich liegen, d. h. der Bedingung $.80 < \mu_T^*/\mu_R^* < 1.25$ genügen. Dies ist aber gleichbedeutend damit, daß die Differenz der Erwartungswerte der Log-Periodenquotienten X_i und Y_j , die nach dem zuvor Gesagten $2 \cdot \log(\mu_R^*/\mu_T^*)$ beträgt, zwischen den Äquivalenzgrenzen $\theta_0 - \varepsilon_1 = 2 \cdot \log(0.80) = -.4463$ und $\theta_0 + \varepsilon_2 = 2 \cdot \log(1.25) = .4463$ zu liegen hat.

Wenn uns zwei unabhängige Stichproben X_1, \dots, X_n und Y_1, \dots, Y_n aus Normalverteilungen mit gemeinsamer Varianz σ^2 und Erwartungswerten μ_1 , μ_2 vorliegen, ist die Durchführung eines Tests für das zugehörige ÄQ-Problem

$$\begin{aligned} H: \mu_1 - \mu_2 \leq -.4463 \vee \mu_1 - \mu_2 \geq .4463 & \text{ gegen} \\ K: -.4463 < \mu_1 - \mu_2 < .4463 \end{aligned} \quad (6)$$

nach dem Intervallinklusions-Prinzip denkbar einfach.

Mit den in Tabelle 4 zusammengestellten Daten, die sich auf eine standardmäßige Bioäquivalenz-Studie über den Calciumantagonisten Nifedipin mit der AUC als

Tabelle 4: Rohdaten zu einer Bioäquivalenz Studie über den Calciumantagonisten Nifedipin. [Zielgröße: AUC .]

(a) Sequenz T-R				(b) Sequenz R-T			
i	A_{i1}	A_{i2}	X_i	j	B_{j1}	B_{j2}	Y_j
1	106.9	112.9	-.05416	1	217.3	195.2	.10725
2	131.3	124.4	.05398	2	174.4	122.7	.35161
3	81.4	89.5	-.9486	3	155.8	188.2	-.18893
4	154.7	134.9	.13695	4	299.5	309.2	-.03187
5	111.2	108.3	.02643	5	157.6	153.5	.02636
6	85.8	94.0	-.09128	6	121.4	104.7	.14799
7	295.2	418.2	-.34926	7	143.9	119.3	.18748
8	217.0	207.0	.04718	8	157.0	146.8	.06717
9	252.3	239.3	.05290	9	114.5	138.2	-.18813
10	157.9	207.3	-.27221	10	71.0	70.3	.00991
$\bar{X} = -.05448, S_X = .15424$				$\bar{Y} = .04888, S_Y = .16522$			

pharmakokinetischer Zielgröße beziehen, ergibt sich die folgende Rechnung: $\bar{X} - \bar{Y} = -.05448 - .04888 = -.10336$, $S = \sqrt{(.15424^2 + .16522^2)/10} = .07148$. Da das .95-Quantil der (zentralen) t -Verteilung mit $\nu = 18$ Freiheitsgraden den Wert 1.73406 hat, erhalten wir als untere bzw. obere 95%-Konfidenzschranke für $\mu_1 - \mu_2$ die Zahl $-.10336 - 1.73406 \cdot .07148 = -.2273$ bzw. $-.10336 + 1.73406 \cdot .07148 = .0206$. Da das zugehörige 90%-Konfidenzintervall in dem unter der vorgelegten Alternativhypothese (6) spezifizierten ÄQ-Intervall enthalten ist, kann hier somit die Äquivalenz beider Formulierungen von Nifedipin bezüglich des Ausmaßes der Absorption auf dem 5%-Niveau für gesichert deklariert werden.

Bemerkung. Das im obigen Beispiel 3 dargestellte Verfahren ist in der Situation des unverbundenen t -Test ein bemerkenswert vernünftiger Test auf Äquivalenz, sofern in der allgemeinen Formulierung (5) der Hypothesen der bloße Lokationsunterschied zwischen den zugrundeliegenden Normalverteilungen als der ausschlaggebende Parameter θ festgelegt wird. Gerade dies aber ist der wohl kritikwürdigste Punkt des statistischen Vorgehens beim Bioäquivalenznachweis, so wie er bisher routinemäßig in einer sehr großen Anzahl von Anwendungen praktiziert wird. Denn die Einstufung von zwei Gauß-Verteilungen mit gemeinsamer Varianz, deren Mittelwerte nicht allzu stark differieren, als äquivalent, ignoriert die elementare Tatsache, daß die zugehörigen Dichten bei festliegendem Abstand der Mittelwerte bereits rein optisch umso „disparater“ aussehen, je kleiner die gemeinsame Streuung ist. Dies ist einer der Gründe, weshalb die bisher übliche statistische Definition von Bioäquivalenz in den letzten Jahren von mehreren Autoren (vgl. [2, 21]) als unzureichend kritisiert worden ist. Vieles spricht dafür, daß die *standardisierte* Mittelwertsdifferenz $(\mu_1 - \mu_2)/\sigma$ das wesentlich sinnvollere Maß für die Unterschiedlichkeit zweier Normalverteilungen mit gemeinsamer Varianz σ^2 ist als die einfache, die Streuung unberücksichtigt lassende Mittelwertsdifferenz $\mu_1 - \mu_2$. Einen ebenfalls sehr einfach zu handhabenden und überdies auch noch hinsichtlich der Trennschärfe optimalen Test auf Äquivalenz bezüglich der standardisierten Differenz der Mittelwerte werden wir für die Situation des unverbundenen (und auch des verbundenen) t -Tests im nächsten Kapitel kennenlernen.

Zum Abschluß des vorliegenden Kapitels können wir festhalten, daß *Intervallinklusions-Regeln* als Tests auf Äquivalenz im engeren [zweiseitigen] Sinne folgende *Vorzüge* besitzen:

- Anschaulichkeit
 - Einfachheit der praktischen Durchführung in allen Situationen, wo praktikable Verfahren für die Berechnung von Konfidenzgrenzen zur Verfügung stehen.
- Diesen Vorzügen stehen die folgenden, mitunter gravierenden *Nachteile* gegenüber:
- Bei kleinen Fallzahlen kann das Intervallinklusions-Prinzip auch dann zu einem absurden „Test“ führen, der die Nullhypothese der Nichtäquivalenz *nie* ablehnen kann, wenn die zugrundeliegenden Verteilungen extrem starke Regularitätsbedingungen erfüllen.
 - Auch in weniger extremen Fällen, wo wenigstens die maximale Trennschärfe nicht unterhalb des Signifikanzniveaus liegt, sind Tests nach dem Intervallinklusions-Prinzip suboptimal und können durch trennschärfere Verfahren ersetzt werden, die mit geringeren Fallzahlen auskommen.

Trotz der oben angeführten Vorzüge des Intervallinklusions-Prinzips mag man sich fragen, wodurch es gerechtfertigt ist, daß in einer Übersicht, bei der optimale Lösungsansätze für ÄQ-Testprobleme im Zentrum des Interesses stehen, auf Verfahren mit solch klaren Defiziten hinsichtlich der Trennschärfe ausführlich eingegangen wird. Eine tiefgehende Begründung liefert das in [23, Anhang A.3] bewiesene asymptotische Optimalitätsresultat für Intervallinklusions-Tests, dessen exakte Formulierung jedoch über den Rahmen der vorliegenden Einführung weit hinausgehen würde.

4. Klassische Hypothesentests für Äquivalenzprobleme

4.1 Grundsätzliches über optimale Tests für Äquivalenzprobleme

Wie anschaulich unmittelbar einleuchtet, hat jeder vernünftige und damit erst recht ein optimaler Test für ein ÄQ-Problem im Sinne von (5) eine Gestalt, die zu derjenigen eines herkömmlichen zweiseitigen Tests komplementär ist. D. h., die kritische Region ist gegeben als

$$\{\mathbf{x} \mid C_1 < T(\mathbf{x}) < C_2\}, \quad (7)$$

so daß die *Ablehnung* der Nullhypothese der Nichtäquivalenz genau dann zu erfolgen hat, wenn der beobachtete Wert einer geeigneten reellwertigen Teststatistik T innerhalb eines genügend *kurzen Intervalls* bleibt. Die Grenzen C_1, C_2 , des kritischen Intervalls hängen eventuell außer vom Signifikanzniveau α und den ÄQ-Grenzen $\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2$ vom Wert einer weiteren Statistik, sagen wir S ab, auf die man bedingt, um die Eliminierung von Störparametern zu erreichen. Wenn die Familie aller nach den Modellvoraussetzungen zugelassenen Verteilungen der Teststatistik von hinlänglich regulärer mathematischer Struktur ist, ist diejenige Festlegung von C_1, C_2 optimal, für welche die (evtl. bedingte) Ablehnungswahrscheinlichkeit des Tests sowohl unter $\theta = \theta_0 - \varepsilon_1$ als auch unter $\theta = \theta_0 + \varepsilon_2$ exakt den Wert α hat. Im Fall einer stetig verteilten Prüfgröße bedeutet dies, daß (C_1, C_2) zu bestimmen ist als Lösung des Gleichungssystems

$$P_{\theta_1}[t_1 < T(\mathbf{X}) < t_2] = \alpha = P_{\theta_2}[t_1 < T(\mathbf{X}) < t_2], \quad t_1 < t_2, \quad (8)$$

mit $\theta_1 = \theta_0 - \varepsilon_1, \theta_2 = \theta_0 + \varepsilon_2$. Weist die Verteilungsfunktion von $T(\mathbf{X})$ hingegen Sprungstellen auf, braucht eine Lösung von (8) nicht zu existieren. Man hat dann stattdessen das kompliziertere System

$$\begin{aligned} & P_{\theta_1}[t_1 < T(\mathbf{X}) < t_2] + \sum_{v=1}^2 w_v P_{\theta_1}[T(\mathbf{X}) = t_v] \\ & = \alpha = P_{\theta_2}[t_1 < T(\mathbf{X}) < t_2] + \sum_{v=1}^2 w_v P_{\theta_2}[T(\mathbf{X}) = t_v], \\ & t_1 \leq t_2, \quad 0 \leq w_1, w_2 < 1 \end{aligned} \quad (9)$$

zu betrachten. Bezeichnen wir die Lösung von (9), die in Verteilungsfamilien von der hier vorausgesetzten Struktur stets existiert und eindeutig bestimmt ist, als $(C_1, C_2, \gamma_1, \gamma_2)$, hat man im optimalen Test im Falle, daß der beobachtete Wert der Teststatistik auf einen der Randpunkte C_1 bzw. C_2 des kritischen Intervalls fällt, mit Wahrscheinlichkeit γ_1 bzw. γ_2 eine randomisierte Entscheidung zugunsten der Alternativhypothese zu treffen. Jedoch kommen solche randomisierten Entscheidungen für die überwiegende Mehrzahl von Anwendungen statistischer Testverfahren auf reale medizinische Forschungsdaten kaum in Frage. Man modifiziert dann den optimalen Test dahingehend, daß man zwar das durch Lösung von (9) erhaltene kritische Intervall (C_1, C_2) verwendet, aber auf seinen Randpunkten stets die Nullhypothese beibehält, d. h. $\gamma_1 = \gamma_2 = 0$ setzt.

In der Regel gestaltet sich die explizite numerische Berechnung optimaler Tests für ÄQ-Probleme erheblich komplizierter als für die zugehörigen ein- oder zweiseitigen Probleme. Der Aufwand für die Durchführung eines optimalen Tests auf ÄQ reduziert sich aber auf ein Minimum, wenn das unter der Alternativhypothese spezifizierte Intervall gegeben ist als $(\theta_1, \theta_2) = (-\varepsilon, \varepsilon)$ und die Verteilung der Teststatistik $T(\mathbf{X})$ unter $\theta = \varepsilon$ dieselbe ist wie diejenige von $-T(\mathbf{X})$ unter $\theta = -\varepsilon$. Man kann nämlich zeigen (siehe [23, Anhang A.1]), daß dann das kritische Intervall (C_1, C_2) für $T(\mathbf{X})$ des optimalen Tests

selbst ein um 0 symmetrisches Intervall, d. h. von der Form $(-C, C)$ sein muß, wobei zur Bestimmung von C im stetigen Falle nur eine einzelne Gleichung zu lösen ist, nämlich

$$P_\varepsilon[|T(\mathbf{X})| < t] = \alpha, \quad t > 0. \quad (8')$$

Mit anderen Worten bestimmt sich dann C gerade als (unterer!) $100\text{-}\alpha$ -Prozentpunkt der Verteilung von $|T(\mathbf{X})|$ unter $\theta = \varepsilon$ (oder, was dasselbe ergibt, unter $\theta = -\varepsilon$). Ohne die Stetigkeitsvoraussetzung hat man außer $-C_1 = C_2 = C$ auch $\gamma_1 = \gamma_2 = \gamma$ mit $0 \leq \gamma < 1$, wobei C und γ gegeben sind durch

$$\begin{aligned} C &= \max \{t > 0 \mid P_\varepsilon[|T(\mathbf{X})| < t] \leq \alpha\}, \\ \gamma &= (\alpha - P_\varepsilon[|T(\mathbf{X})| < C]) / P_\varepsilon[|T(\mathbf{X})| = C]. \end{aligned} \quad (9')$$

Eine solche vereinfachte Berechnung der kritischen Konstanten eines optimalen Tests auf ÄQ ist insbesondere überall dort möglich, wo die ÄQ-Grenzen symmetrisch festgelegt werden und θ ein Lokationsparameter für die Verteilung von $T(\mathbf{X})$ ist, bei um 0 symmetrischer Verteilung der Teststatistik unter $\theta = 0$.

In der Praxis hat man viele Testprobleme zu betrachten, in denen die gemeinsame Verteilung der zu analysierenden Daten außer von dem Hauptparameter θ , bezüglich dessen Äquivalenz nachgewiesen werden soll, von einem ein- oder mehrdimensionalen Störparameter abhängt (z. B. von μ_2 und σ zusätzlich zu $\theta = (\mu_1 - \mu_2)/\sigma$ in der Situation des unverbundenen t -Tests). In einem Teil dieser Situation führt die Technik des Bedingens auf eine weitere Statistik (zusätzlich zur Prüfgröße) bei der Konstruktion eines optimalen Tests auf ÄQ nicht zum Ziel. Unter günstigen Umständen läßt sich dann das Problem stattdessen dadurch hinreichend vereinfachen, daß man die Invarianz der Hypothese gegenüber bestimmten Transformationen der Daten ausnutzt, indem man nur solche Tests in Betracht zieht, die ihrerseits unter denselben Transformationen unverändert bleiben. Dieser Konstruktionsansatz liegt z. B. den in den folgenden Abschnitten dargestellten ÄQ-Tests auf der Basis der üblichen Prüfgrößen der t -Tests zugrunde.

4.2 Verbundener t -Test auf Äquivalenz

Die Daten mögen wiederum (vgl. Abschn. 2.1) aus n Paaren $(X_1, Y_1), \dots, (X_n, Y_n)$ bestehen, derart, daß X_i bzw. Y_i das Ergebnis von zwei zu vergleichenden Behandlungen A und B bei der i -ten Beobachtungseinheit der Stichprobe bezeichnet. Um den Wirkungsunterschied der Behandlungen bei der i -ten Versuchseinheit zu erfassen, wird für das zugehörige Paar (X_i, Y_i) die Differenz $D_i = X_i - Y_i$ bestimmt.

Nehmen wir vorderhand lediglich an, daß die Grundgesamtheit der intraindividuellen Differenzen der Behandlungsergebnisse eine Verteilung ergibt, die stetig und symmetrisch ist, würde man, wenn die Behandlungen in Wahrheit völlig identische Wirkungen hätten, natürlich erwarten, daß das Symmetriezentrum dieser Verteilung mit der Null zusammenfällt. D. h., die Wahrscheinlichkeit p_+ , bei einem zufällig ausgewählten Probanden oder Patienten eine positive Differenz (\leftrightarrow besseres Ergebnis unter A als unter B) zu erhalten, müßte genauso groß sein wie die Wahrscheinlichkeit für eine negative Differenz, nämlich $1/2$. Ein für die Anschauung natürliches statistisches Kriterium für die Äquivalenz der zu vergleichenden Behandlungen A und B ergibt sich dann, indem man fordert, daß der wahre Wert von p_+ seinem unter identisch wirkenden Behandlungen vorliegenden Referenzwert von $1/2$ hinlänglich nahekommt.

Machen wir nun zusätzlich die für den herkömmlichen verbundenen t -Test spezifische Annahme, daß die Verteilung der D_i eine Gaußsche Normalverteilung mit (unbekannten) Parametern $\delta = E(D_i)$ und $\sigma_D^2 = \text{Var}(D_i)$ darstellt, können wir p_+ explizit ausrechnen

als $p_+ = P[D_i > 0] = 1 - P[(D_i - \delta)/\sigma_D \leq -\delta/\sigma_D] = 1 - \Phi(-\delta/\sigma_D) = \Phi(\delta/\sigma_D)$, wobei $\Phi(\cdot)$ für die Standard-Normalverteilungsfunktion steht. Hieraus folgt unmittelbar, daß die zuvor allgemein formulierte ÄQ-Bedingung

$$1/2 - \tilde{\varepsilon}_1 < p_+ < 1/2 + \tilde{\varepsilon}_2 \quad (10)$$

unter den speziellen Voraussetzungen des verbundenen t -Tests gleichbedeutend ist mit $\Phi^{-1}(1/2 - \tilde{\varepsilon}_1) < \delta/\sigma_D < \Phi^{-1}(1/2 + \tilde{\varepsilon}_2)$ [$\Phi^{-1}(\cdot) \equiv$ Quantilfunktion der Standard-Normalvertlg.]. Suchen wir also einen Test, der es erlaubt zu entscheiden, ob die Daten hinlänglich klare Evidenz zugunsten der Aussage (10) als Alternativhypothese enthalten, führt dies auf das Testproblem

$$H: \delta/\sigma_D \leq -\varepsilon_1 \vee \delta/\sigma_D \geq \varepsilon_2 \quad \text{gegen} \quad K: -\varepsilon_1 < \delta/\sigma_D < \varepsilon_2, \quad (11)$$

vorausgesetzt, wir setzen

$$\varepsilon_1 = -\Phi^{-1}(1/2 - \tilde{\varepsilon}_1), \quad \varepsilon_2 = \Phi^{-1}(1/2 + \tilde{\varepsilon}_2). \quad (12)$$

Anhand von (12) rechnet man insbesondere leicht nach, daß die Verwendung der Grenzen $\varepsilon_1 = \varepsilon_2 = .50$ in einem Test auf ÄQ bezüglich der *standardisierten* Mittelwertsdifferenz bedeutet, daß man als maximale Abweichung zwischen der Wahrscheinlichkeit einer positiven Differenz der Behandlungsergebnisse und $1/2$ rund 20% zulassen möchte. Es dürfte weitgehende Einigkeit darüber bestehen, daß es keine extrem strikte Anforderung an den Grad der Ähnlichkeit zweier Wahrscheinlichkeiten ist, wenn man für die Abweichung zwischen diesen Größen eine „Toleranz“ von $\pm 20\%$ zuläßt. Ausgehend von dieser Beobachtung werden wir im nachfolgenden Beispiel für den optimalen Test für das ÄQ-Problem (11) mit den Grenzen $\varepsilon_1 = \varepsilon_2 = .50$ arbeiten.

Für viele, wenn nicht die meisten praktischen Anwendungen ist es angemessen, von einer *symmetrischen Festlegung der ÄQ-Grenzen* auszugehen, d. h. die Konstanten ε_1 und ε_2 auf einen *gemeinsamen Wert* ε (z. B. eben $\varepsilon = .50$) festzulegen. Unter Anwendung von Sätzen aus Lehmann [9, Ch. 6] kann man zeigen [23], daß man einen optimalen (genauer: unter allen gegenüber Reskalierungen der D_i invarianten Tests gleichmäßig besten) Test für das Problem (11) erhält, indem man die übliche t -Statistik $T = \sqrt{n}\bar{D}/S_D$ als Prüfgröße verwendet und die Nullhypothese der Inäquivalenz genau dann ablehnt, wenn diese dem Betrag nach unterhalb einer geeignet zu spezifizierenden Konstante $C = C(n, \alpha; \varepsilon)$ bleibt. Abbildung 1 veranschaulicht die Berechnungsweise der kritischen Konstante des optimalen Tests für den Fall, daß wir speziell $n = 23$, $\alpha = .05$ sowie $\varepsilon = .50$ (vgl. oben) wählen. Als analytische Darstellung ergibt sich die Gleichung:

$$C(n, \alpha; \varepsilon) = \{100\alpha\text{-Prozentpunkt von } F\text{-Vertlg. mit } 1, n - 1 \text{ Freiheitsgraden} \\ \text{sowie Nichtzentralitätsparam. } NC = n\varepsilon^2\}^{1/2}. \quad (13)$$

Beispiel 4. *Nachweis der Stabilität eines Meßparameters während der Vorbehandlungsphase in einer tierexperimentellen Studie.* In einer experimentellen Untersuchung über Möglichkeiten der medikamentösen Beeinflussung des Flows in der zerebralen Mikrozirkulation von Kaninchen ($n = 23$) sollte sichergestellt werden, daß der Meßparameter [ml/min/100 g Gewicht] während der 15-minütigen Vorbehandlungsphase ausreichend stabil bleibt. Es wurden die folgenden Werte ermittelt (Daten aus [20]): $\bar{D} = \bar{X} - \bar{Y} = 52.65 - 52.49 = 0.16$, $S_D = 3.99$, wobei die X_i bzw. Y_i die im Zeitpunkt $t = -15$ min bzw. $t = 0$ min gemessenen Flows bezeichnen. Mit $\varepsilon = .50$ als ÄQ-Grenze für δ/σ_D ergibt sich $NC = 23 \cdot 0.50^2 = 5.75$. Um den 5%-Punkt einer F -Verteilung mit Zählerfreiheitsgrad 1, Nennerfreiheitsgrad 22 und Nichtzentralitätsparameter 5.75 zu berechnen, genügt es, in einem entsprechenden SAS-Datenschritt die Zeile zu schreiben: „cq = finv (.05, 1, 22, 5.75)“. Man bekommt dann „cq = .576779“ ausgegeben und erhält damit insgesamt für den kritischen Wert des Tests zum 5%-Niveau: $C(23, .05;$

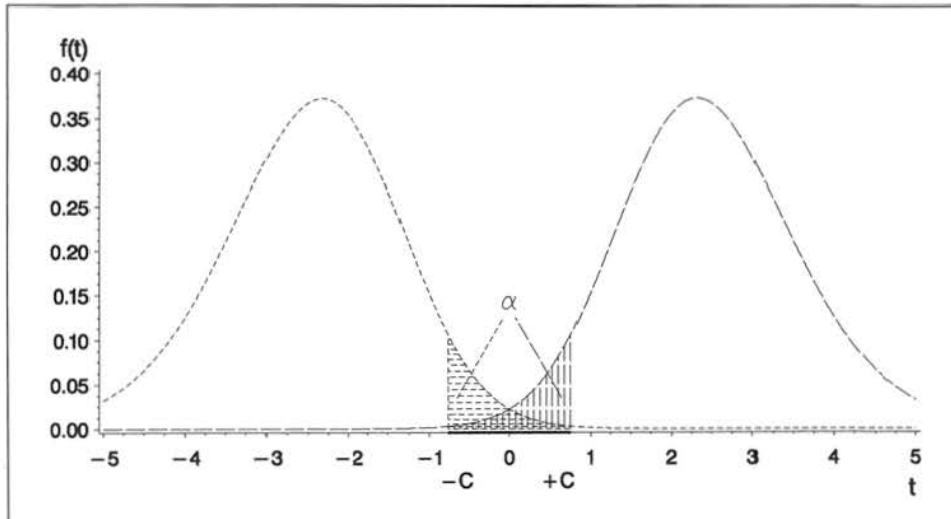


Abbildung 1: Veranschaulichung der Berechnungsgrundlage für die kritische Konstante C des verbund. t -Tests auf ÄQ im Falle $n = 23$, $\alpha = 5\%$, $\varepsilon = .50$ [Li: Vertlg. von T für $\delta/\sigma_D = -.50$; Re: Vertlg. von T für $\delta/\sigma_D = +.50$]

$.50) = \sqrt{.576779} = .7595$. Da sich andererseits der Betrag der t -Statistik hier zu $|T| = .1923$ errechnet, ist die Hypothese, daß sich der Mikrozirkulationsflow innerhalb der Vorbehandlungsphase nicht in relevantem Ausmaß ändert, durch die vorgelegten experimentellen Daten auf dem 5%-Niveau gesichert.

In Anbetracht der Beziehung (13) ist der praktische Aufwand bei der Durchführung des optimalen Tests auf ÄQ bezüglich δ/σ_D im symmetrischen Fall $\varepsilon_1 = \varepsilon_2 = \varepsilon > 0$ kaum größer als im herkömmlichen t -Test für verbundene Stichproben. Bei nichtsymmetrischer Festlegung des hypothetischen ÄQ-Bereichs $(-\varepsilon_1, \varepsilon_2)$ für δ/σ_D läuft der Tests so, daß man nachzuprüfen hat, ob der beobachtete Wert der t -Statistik innerhalb eines geeigneten Intervalls mit Endpunkten $C_1(n, \alpha; \varepsilon_1, \varepsilon_2)$, $C_2(n, \alpha; \varepsilon_1, \varepsilon_2)$ liegt. Leider besitzen diese kritischen Konstanten im allgemeinen Fall keine explizite Darstellung. Man benötigt dann ein iteratives Verfahren für die numerische Lösung desjenigen Gleichungssystems, welches sich aus der Bedingung ergibt (vgl. Abb. 1), daß die Ablehnungswahrscheinlichkeit des Tests sowohl unter $\delta/\sigma_D = -\varepsilon_1$ als auch unter $\delta/\sigma_D = \varepsilon_2$ gerade gleich α zu sein hat. Die detaillierte Darstellung eines hierzu geeigneten Algorithmus sowie dessen Implementierung in SAS findet man in [23].

Schließlich sei noch darauf hingewiesen, daß der verbundene t -Test auf ÄQ unter bestimmten, für die Mehrzahl vergleichender Bioverfügbarkeitsstudien durchaus realistischen Voraussetzungen (Lognormalverteilung der intraindividuellen Bioverfügbarkeitsquotienten; Nichtexistenz von Periodeneffekten) ein empfehlenswertes Verfahren des statistischen Bioäquivalenz-Nachweises ergibt. (Näheres hierzu in [21, Abschn. 4]).

4.3 Unverbundener t -Test auf Äquivalenz

Wird die Überprüfung der Behandlungen A und B im Rahmen eines Parallelgruppen-Versuchs vorgenommen, besteht der Datensatz, anhand dessen die Äquivalenz von A und B zu beurteilen ist, aus $m + n$ wechselseitig unabhängigen Beobachtungen

$X_1, \dots, X_m, Y_1, \dots, Y_n$. Nach wie vor wollen wir unter den X -Werten die Ergebnisse der ersten Behandlung (A) verstehen, während Y_j den beim j -ten Probanden bzw. Patienten der Gruppe B gefundenen Wert des jeweiligen Zielkriteriums bezeichnen soll. Genau wie beim herkömmlichen ein- oder zweiseitigen t -Test für unverbundene Stichproben machen wir in diesem Abschnitt die spezielle parametrische Annahme

$$X_i \sim \mathcal{N}(\mu_1, \sigma^2) \forall i = 1, \dots, m, \quad Y_j \sim \mathcal{N}(\mu_2, \sigma^2) \forall j = 1, \dots, n, \quad (14)$$

wobei alle drei Parameter μ_1, μ_2 und σ^2 als unbekannte Konstanten behandelt werden. Nicht anders als im verbundenen Fall bietet sich als natürliches Maß für die Unterschiedlichkeit der unter Betracht stehenden Gauß-Verteilungen die standardisierte Mittelwertsdifferenz $\theta = (\mu_1 - \mu_2)/\sigma$ an. Entsprechend definieren wir ÄQ der Behandlungen A und B bzw. der Verteilungen $\mathcal{N}(\mu_1, \sigma^2), \mathcal{N}(\mu_2, \sigma^2)$ durch die Forderung, daß der wahre Wert dieses Abstandsmaßes in einem genügend engen Intervall $(-\varepsilon_1, \varepsilon_2)$ um 0 zu liegen hat. Mit anderen Worten formulieren wir analog zu (11) jetzt das Testproblem

$$\begin{aligned} H: (\mu_1 - \mu_2)/\sigma \leq -\varepsilon_1 \vee (\mu_1 - \mu_2)/\sigma \geq \varepsilon_2 \quad \text{gegen} \\ K: -\varepsilon_1 < (\mu_1 - \mu_2)/\sigma < \varepsilon_2 \quad (\varepsilon_1, \varepsilon_2 > 0). \end{aligned} \quad (15)$$

Da man unter Annahme von (14) ersichtlicherweise $(X_i - Y_j) \sim \mathcal{N}(\mu_1 - \mu_2, 2\sigma^2)$ und damit $P[X_i > Y_j] = \Phi((\mu_1 - \mu_2)/\sqrt{2}\sigma)$ schreiben kann für beliebiges (i, j) , ist das Testproblem (15) dasselbe wie

$$\begin{aligned} \tilde{H}: \pi_{XY} \leq (1/2) - \tilde{\varepsilon}_1 \vee \pi_{XY} \geq (1/2) + \tilde{\varepsilon}_2 \quad \text{gegen} \\ \tilde{K}: (1/2) - \tilde{\varepsilon}_1 < \pi_{XY} < (1/2) + \tilde{\varepsilon}_2, \end{aligned} \quad (\tilde{15})$$

vorausgesetzt, wir verabreden, daß

$$\begin{aligned} \pi_{XY} = P[X_i > Y_j]; \quad \tilde{\varepsilon}_1 = \Phi(\varepsilon_1/\sqrt{2}) - 1/2, \\ \tilde{\varepsilon}_2 = \Phi(\varepsilon_2/\sqrt{2}) - 1/2. \end{aligned} \quad (16)$$

In der Situation des unverbundenen t -Tests ist also die Hypothese der Äquivalenz beider Verteilungen bezüglich der standardisierten Mittelwertsdifferenz gleichbedeutend mit der Annahme, daß die Wahrscheinlichkeit, einen X -Wert zu erhalten, der größer („besser“) ist als ein zufällig ausgewählter Y -Wert, nur unwesentlich von $1/2$ abweicht. Natürlich ist dabei $1/2$ derjenige Wert für die Wahrscheinlichkeit des Ereignisses $\{X_i > Y_j\}$, welcher vorliegen müßte, wenn die zugrundeliegenden Verteilungen nicht nur äquivalent, sondern identisch wären. Aus der Gleichung (16) ergibt sich insbesondere, daß eine Festlegung der ÄQ-Grenzen für $(\mu_1 - \mu_2)/\sigma$ auf $\pm .50$ annähernd einer „Toleranz“ von $\pm 15\%$ für die Abweichung zwischen der Wahrscheinlichkeit π_{XY} und deren Sollwert 50% entspricht.

Auch in der Situation des Zweistichproben-Problems mit normalverteilten Beobachtungen gleicher Varianz gibt es einen optimalen Test auf ÄQ bezüglich der standardisierten Differenz der Erwartungswerte. Vorausgesetzt, wir beschränken uns wieder auf den Fall, daß die hypothetischen ÄQ-Grenzen symmetrisch angesetzt werden, ist dieses Verfahren ebenfalls sehr einfach anzuwenden und weist in der praktischen Durchführung weitgehende Ähnlichkeiten mit dem im vorigen Abschnitt dargestellten verbundenen t -Test auf ÄQ auf. Als Teststatistik hat man dieselbe Funktion der einzelnen Beobachtungen zu benutzen wie im Test der zugehörigen ein- oder zweiseitigen Nullhypothese, nämlich

$$T = \sqrt{mn(N-2)/N} (\bar{X} - \bar{Y}) \left/ \left\{ \sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2 \right\}^{1/2} \right. . \quad (17)$$

Alles was sich an der Berechnungsvorschrift für die kritische Konstante, die jetzt natürlich von beiden Stichprobenumfängen abhängt, gegenüber der Beziehung (13) ändert, ist der zweite Freiheitsgrad sowie der Ausdruck für den Nichtzentralitätsparameter. Genauer gesagt ist die kritische Konstante des optimalen Tests für das ÄQ-Problem (15) mit $\varepsilon_1 = \varepsilon_2 = \varepsilon$ gegeben durch

$$C_{m,n}(\alpha; \varepsilon) = \{100\alpha\text{-Prozentpunkt von } F\text{-Vertlg. mit } 1, m + n - 2 \text{ Freiheitsgraden sowie Nichtzentralitätsparam.}\} \\ \text{NC} = (mn/(m + n)) \varepsilon^2\}^{1/2}. \quad (18)$$

Die Entscheidungsregel lautet analog zum verbundenen t -Test auf ÄQ mit symmetrischem ÄQ-Bereich $(-\varepsilon, \varepsilon)$:

$$\text{Entscheidung zugunsten von } \ddot{\text{A}}\text{Q} \Leftrightarrow |T| < C_{m,n}(\alpha; \varepsilon). \quad (19)$$

Eine der naheliegendsten und wichtigsten Anwendungen des unverbundenen t -Tests auf ÄQ ist der *Bioäquivalenznachweis in Crossover-Studien mit nichtvernachlässigbaren Periodeneffekten* sowie modifizierter statistischer Definition der Äquivalenz zweier Normalverteilungen mit gemeinsamer Varianz. Zur Illustrierung dieser Möglichkeit reanalysieren wir die Daten des Beispiels 3.

Beispiel 3 (Fortsetzung). Lassen wir für zwei äquivalente Normalverteilungen diesmal eine Abweichung der Erwartungswerte zu, die höchstens so groß sein darf wie die gemeinsame Standardabweichung, haben wir bei je 10 Fällen pro Gruppe den kritischen Wert für die t -Statistik gemäß (18) zu bestimmen aus der F -Verteilung mit Freiheitsgraden 1 und $2 \cdot 10 - 2 = 18$ sowie Nichtzentralitätsparameter $\text{NC} = (10^2/20) \cdot 1.00^2 = 5.00$. Der 5%-Punkt dieser Verteilung berechnet sich mittels der SAS-Funktion „finv“ zu .376568, so daß wir erhalten: $C_{10,10}(.05; 1.00) = \sqrt{.376568} = .61365$. Andererseits finden wir für die in Tabelle 4 aufgelisteten logarithmierten AUC -Quotienten X_i bzw. Y_j als t -Wert die Zahl $T = -.10336/.07148 = -1.4460$. Also ist der beobachtete t -Wert im Sinne von (19) dem Betrag nach zu groß, so daß die Daten auf dem 5%-Niveau beurteilt *nicht* ausreichen, um die Äquivalenz beider Verteilungen von Log-Bioverfügbarkeits-Quotienten im Sinne von $|\mu_1 - \mu_2|/\sigma < 1.00$ statistisch zu sichern.

Bemerkung. Aus Beispiel 3 ersehen wird insbesondere, daß es unter den Grundvoraussetzungen des unverbundenen t -Tests sehr wohl Situationen gibt, in denen der Intervalleinschluß-Test auf ÄQ bezüglich $\mu_1 - \mu_2$ zu einem positiven Ergebnis kommt, obwohl der optimale Test auf ÄQ bezüglich $(\mu_1 - \mu_2)/\sigma$ zum selben Niveau die Nullhypothese nicht ablehnen kann. Es sollte sorgfältig beachtet werden, daß hieraus nicht etwa ein Widerspruch zur Optimalitätseigenschaft des Zweifstichproben- t -Tests auf ÄQ resultiert. Im Gegenteil sind Trennschärfevergleiche zwischen dem t -Test auf ÄQ und einem beliebigen Test auf Äquivalenz bezüglich $\mu_1 - \mu_2$ von vornherein sinnlos (weitere Verfahren, die für dieses letztere Problem vorgeschlagen worden sind, gehen zurück auf ANDERSON & HAUCK [1] sowie PATEL & GUPTA [14]. Denn die Hypothesen die geprüft werden, sind grundsätzlich verschieden, es sei denn, man betrachtet σ anstatt als freien Störparameter als a priori bekannte Zahl.

4.4 Exakter Fisher-Test auf Äquivalenz

Genau wie im Abschnitt 2.2 setzen wir jetzt voraus, daß die Daten, aufgrund derer die Äquivalenz der Behandlungen A und B beurteilt werden soll, einer 2×2 -Kontingenztafel

mit den festen Zeilensummen m, n entsprechen (vgl. vorn, Tab. 2) und daß als Zielparameter die Odds-Ratio ϱ verwendet wird. Es soll aber hier anstatt auf einseitige Äquivalenz auf ÄQ im engeren Sinne getestet werden, so daß wir das Hypothesenpaar (2) zu ersetzen haben durch

$$\begin{aligned} H: 0 < \varrho \leq \varrho_1 \vee \varrho_2 \leq \varrho < \infty \quad \text{gegen} \\ K: \varrho_1 < \varrho < \varrho_2 \quad (\varrho_1 < 1 < \varrho_2, \text{ fest}), \end{aligned} \tag{20}$$

mit $\varrho := p_1(1 - p_2)/(1 - p_1)p_2$ und p_1 bzw. p_2 als der wahren Responderrate unter Behandlung A bzw. Behandlung B . Bezeichnet X bzw. Y die Anzahl von Respondern in Gruppe A bzw. Gruppe B , wird nicht anders als im herkömmlichen exakten Fisher-Test die Annahme $X \sim \mathcal{B}(m, p_1), Y \sim \mathcal{B}(n, p_2)$ gemacht, bei Unabhängigkeit des Paares (X, Y) .

Der optimale Test für das Hypothesenpaar (20) hat mit dem in Abschnitt 2.2 dargestellten Test auf einseitige ÄQ zweier Binominalverteilungen bezüglich der Odds-Ratio ϱ dies gemeinsam, daß die *bedingte* Verteilung der Anzahl X von Respondern unter Behandlung A betrachtet wird, *gegeben die 1. Spaltensumme* S , i. e. die Gesamtzahl von Respondern in beiden Behandlungsgruppen. Genauer gesagt hat die Entscheidungsregel, die man für s als speziell vorliegendem Wert der 1. Spaltensumme zu benutzen hat, um einen optimalen (gleichmäßig besten unverfälschten) Test für das ÄQ-Problem (20) zu erhalten, folgendes Aussehen:

$$\left\{ \begin{array}{ll} \text{Ablehnung von } H & \text{falls } C_1^{m,n}(s; \varrho_1, \varrho_2, \alpha) < X \\ & < C_2^{m,n}(s; \varrho_1, \varrho_2, \alpha) \\ \text{Abl. v. } H \text{ mit Wkt. } \gamma_1^{m,n}(s; \varrho_1, \varrho_2, \alpha) & \text{falls } X = C_1^{m,n}(s; \varrho_1, \varrho_2, \alpha) \\ \text{Abl. v. } H \text{ mit Wkt. } \gamma_2^{m,n}(s; \varrho_1, \varrho_2, \alpha) & \text{falls } X = C_2^{m,n}(s; \varrho_1, \varrho_2, \alpha) \\ \text{Beibehaltung von } H & \text{falls } X < C_1^{m,n}(s; \varrho_1, \varrho_2, \alpha) \\ & \text{oder } X > C_2^{m,n}(s; \varrho_1, \varrho_2, \alpha) \end{array} \right. \tag{21}$$

Für jedes feste s hängt der Test somit von insgesamt vier verschiedenen kritischen Konstanten ab, nämlich den Grenzen $C_1^{m,n}(s; \varrho_1, \varrho_2, \alpha), C_2^{m,n}(s; \varrho_1, \varrho_2, \alpha)$, zwischen denen die beobachtete Responderzahl X in Gruppe A liegen muß, damit wir die Nullhypothese der Nichtäquivalanz sicher ablehnen dürfen, und den Wahrscheinlichkeiten $\gamma_1^{m,n}(s; \varrho_1, \varrho_2, \alpha), \gamma_2^{m,n}(s; \varrho_1, \varrho_2, \alpha)$, mit denen wir randomisierte Entscheidungen zugunsten von ÄQ zu treffen haben, falls X auf die linke bzw. die rechte dieser Grenzen fällt. Die Berechnung der genannten Konstanten hat dabei zu erfolgen, indem man das Gleichungssystem

$$\begin{aligned} \sum_{x=C_1+1}^{C_2-1} h_s^{m,n}(x; \varrho_1) + \sum_{v=1}^2 \gamma_v h_s^{m,n}(C_v; \varrho_1) = \alpha = \sum_{x=C_1+1}^{C_2-1} h_s^{m,n}(x; \varrho_2) + \\ \sum_{v=1}^2 \gamma_v h_s^{m,n}(C_v; \varrho_2), \quad 0 \leq C_1 \leq C_2 \leq m, \quad 0 \leq \gamma_1, \gamma_2 < 1, \end{aligned} \tag{22}$$

löst. Hierin bezeichnet $h_s^{m,n}(\cdot; \varrho)$ die Wahrscheinlichkeits-Massefunktion der bereits in Abschnitt 2.2 benutzten erweiterten hypergeometrischen Verteilung, welche gegeben ist durch

$$\begin{aligned} h_s^{m,n}(x; \varrho) = \binom{m}{x} \binom{n}{s-x} \varrho^x / \sum_{j=\max\{0, s-n\}}^{\min\{s, m\}} \binom{m}{j} \binom{n}{s-j} \varrho^j, \\ \max\{0, s-n\} \leq x \leq \min\{s, m\}. \end{aligned} \tag{23}$$

Die theoretischen Eigenschaften des Tests (21) ergeben sich aus einem Satz über gleichmäßig beste unverfälschte Tests für mehrparametrische Exponentialfamilien, der in § 4.4 des Buches von Lehmann [9] bewiesen wird. Die praktische Anwendbarkeit des Verfahrens steht und fällt natürlich mit der Verfügbarkeit eines effizienten Algorithmus für die Lösung des Gleichungssystems (22). Bestens bewährt hat sich (auch für analoge Probleme bei anderen diskreten Verteilungen) der folgende in [23] angegebene

Suchalgorithmus:

(i.a) Festlegung eines Startwerts C_1^0 für die linke kritische Grenze C_1 , von dem bekannt ist, daß $C_1^0 \geq C_1$. [Geeignete Wahl: $C_1^0 = ms/(m+n) + 5$.]

(i.b) Aufsuchen der größten positiven Zahl C_2^0 , für die das abgeschlossene Intervall $[C_1^0 + 1, C_2^0 - 1]$ sowohl unter $\varrho = \varrho_1$ als auch unter $\varrho = \varrho_2$ bedingte Wahrscheinlichkeit $\leq \alpha$ hat.

(ii) Behandlung von (22) als lineares Gleichungssystem in (γ_1, γ_2) bei festgehaltenem $C_1 = C_1^0$ und $C_2 = C_2^0$. Berechnung der Lösung (γ_1^0, γ_2^0) dieses Systems.

(iii) Abprüfen der Bedingung $0 \leq \gamma_1^0, \gamma_2^0 < 1$. Wenn erfüllt: Stop. Andernfalls: Verminderung von C_1^0 um 1 und zurück nach (i.b).

Wir illustrieren die Anwendung des bedingten Tests (21), für den die Bezeichnung „Exakter Fisher-Test auf Äquivalenz“ adäquat erscheint, in einem weiteren Beispiel.

Beispiel 5. Nachweis der Homogenität zweier Teilkollektive einer vergleichenden Therapie-studie mittels des exakten Fisher-Tests auf ÄQ. Bei der Auswertung einer Studie zum Vergleich von Verapamil versus Hydrochlorothiazid in der antihypertensiven Therapie sollte ausgeschlossen werden, daß es relevante Unterschiede der Responderraten zwischen einschlägig vorbehandelten und unvorbehandelten Patienten gab. Tabelle 5 gibt die in beiden Teilkollektiven für die Titrationsperiode gefundenen Häufigkeiten von Respondern bzw. Therapieversagern wieder. Als hypothetischer ÄQ-Bereich für die wahre Odds-Ratio $\varrho = p_1(1-p_2)/(1-p_1)p_2$ wurde das Intervall $(\varrho_1, \varrho_2) = (0.5, 1.5)$ spezifiziert, wobei der Test wiederum auf dem 5%-Niveau durchgeführt wurde. Für $m = 225$, $n = 119$, $s = 171$ und die genannte Festlegung von ϱ_1 , ϱ_2 sowie α erhält man mittels des oben formulierten Suchalgorithmus (ein SAS-Macro für die Implementierung dieses Algorithmus ist im Anhang B zu [23] enthalten) das kritische Intervall (106, 113) und die Wahrscheinlichkeiten .8777 bzw. .0257 für eine randomisierte Entscheidung zugunsten von Äquivalenz im Falle $X = 106$ bzw. $X = 113$. Da der beobachtete Wert von X (=108) somit im Ablehnungsbereich liegt, gelangt man für die zu beurteilende Kontingenztafel im exakten Fisher-Test auf ÄQ im Sinne von $0.5 < \varrho < 1.5$ zu einer positiven Entscheidung.

Tabelle 5: Therapieerfolge bei vorbehandelten und unvorbehandelten Patienten der VERDI-Studie [8] am Ende der Titrationsperiode

Antihypertensive Vorbehandlung	Response		Σ
	+	-	
ja	108 (48.00%)	117 (52.00%)	225 (100.00%)
nein	63 (52.94%)	56 (47.06%)	119 (100.00%)
Σ	171	173	344

5. Verfahren für Überlebenszeit-Studien

Tests auf Nichtvorhandensein relevanter Unterschiede zwischen Überlebenskurven sind in der Literatur erst in allerjüngster Zeit beschrieben bzw. entwickelt worden, obwohl sie im onkologischen Bereich für die Auswertung von positiv kontrollierten Therapiestudien häufig benötigt werden. Wir beschränken uns auf die Darstellung eines Verfahrens für zwei unverbundene Stichproben, die auch zensiert sein dürfen. Die Überlebensfunktionen in den zugrundeliegenden Populationen wollen wir als $S_1(\cdot)$, $S_2(\cdot)$ bezeichnen und voraussetzen, daß sie vom stetigen Typ sind und die aus der Cox-Regression gewohnte Annahme *proportionaler Intensitäten* erfüllen. Im vorliegenden Zusammenhang können wir dabei ohne Einschränkung annehmen, daß eine dieser beiden Funktionen, sagen wir $S_1(\cdot)$, die Baseline-Überlebensfunktion des zugrundeliegenden Proportional-hazards-Modells darstellt. Entsprechend gehen wir aus von der Annahme, daß zwischen $S_1(\cdot)$ und $S_2(\cdot)$ die Beziehung besteht:

$$S_2(t) = [S_1(t)]^\theta \quad \text{für alle } t > 0 \quad \text{und ein } \theta > 0. \quad (24)$$

Für die Anschauung ist es plausibel, äquivalente Überlebensfunktionen zu definieren durch die Forderung, daß der vertikale Abstand zwischen $S_1(\cdot)$ und $S_2(\cdot)$ *überall* genügend klein ist. Entsprechend wollen wir die Äquivalenz von $S_1(\cdot)$ und $S_2(\cdot)$ dann als gegeben ansehen, wenn ein genügend kleines $\delta > 0$ existiert, so daß

$$|S_1(t) - S_2(t)| < \delta \quad \text{für alle } t > 0. \quad (25)$$

Man kann nun zeigen (für Details siehe [22]), daß die Bedingung (25) unter Voraussetzung von (24) gleichbedeutend ist mit der Forderung:

$$1/(1 + \varepsilon) < \theta < 1 + \varepsilon \quad \text{für geeignetes } \varepsilon > 0. \quad (26)$$

Hierin ist die Konstante ε durch die für den Abstand von $S_1(\cdot)$ und $S_2(\cdot)$ vorgegebene Schranke δ eindeutig bestimmt vermöge der Gleichung

$$(1 + \varepsilon)^{-1/\varepsilon} - (1 + \varepsilon)^{-(1+\varepsilon)/\varepsilon} = \delta, \quad (27)$$

die sich leicht iterativ lösen läßt. Geben wir, was für die weiteren Betrachtungen bequemer ist, nicht ε selbst an, sondern $\varepsilon' = \log(1 + \varepsilon)$, hat man insbesondere die in Tabelle 6 aufgelisteten Entsprechungen.

Als Testproblem formulieren wir nun das Hypothesenpaar

$$\begin{aligned} H: |S_1(t) - S_2(t)| \geq \delta \quad \text{für mindestens ein } t > 0 \\ \text{gegen } K: |S_1(t) - S_2(t)| < \delta \quad \text{f.a. } t > 0. \end{aligned} \quad (28)$$

Nach den vorangehenden Feststellungen ist dieses Testproblem unter der Proportional-hazards-Annahme (24) dasselbe wie

$$H': \beta \leq -\varepsilon' \vee \beta \geq \varepsilon' \quad \text{gegen } K': -\varepsilon' < \beta < \varepsilon', \quad (28')$$

vorausgesetzt, wir verabreden $\beta = \log \theta$, $\varepsilon' = \log(1 + \varepsilon)$ und berechnen ε als die Lösung der Gleichung (27). Aufgrund von (24) ist weiterhin klar, daß der Parameter β interpretiert

Tabelle 6: Maximale vertikale Abstände δ zwischen den wahren Überlebensfunktionen und zugehörige Äquivalenzgrenzen $\pm \varepsilon'$ für den Parameter $\log \theta$.

δ	0.10	0.15	0.20	0.25
ε'	.2727	.4106	.5505	.6931

werden kann als *Regressionskoeffizient* einer Dummy-Kovariablen, welche für eine beliebige Beobachtung den Wert 0 oder 1 annimmt, je nachdem, ob diese der ersten oder der zweiten Stichprobe entstammt. Demnach läßt sich hier die Standard-Theorie für die Cox-Regression mit einer einzelnen Kovariable anwenden, wonach der Partial-likelihood-Schätzer $\hat{\beta}$ für den Parameter β , auf den sich unser ÄQ-Testproblem in der Formulierung (28') bezieht, asymptotisch normal ist mit einer Varianz $v^2(\hat{\beta})$, die konsistent geschätzt werden kann durch die reziproke beobachtete Fisher-Information $1/I(\hat{\beta})$. Asymptotisch reduziert sich damit das ÄQ-Problem (28') und folglich auch das primär interessierende Problem (28) auf dasjenige des Testens von

$$H_a: |\mu| \geq \varepsilon'/v(\hat{\beta}) \quad \text{gegen} \quad K_a: |\mu| < \varepsilon'/v(\hat{\beta}) \quad (29)$$

in einer Normalverteilung mit Erwartungswert μ und Varianz 1, aus der man $\hat{\beta}/v(\hat{\beta})$ als einzige Beobachtung zur Verfügung hat.

Für das asymptotische Problem (29) existiert ein optimaler Test (Lehmann [9, Sec. 3.7]), dessen kritische Region sich explizit schreiben läßt [23, S. 25] als

$$\{|\hat{\beta}|/v(\hat{\beta}) < C_\alpha(\varepsilon'/v(\hat{\beta}))\}. \quad (30)$$

Unter $C_\alpha(\varepsilon'/v(\hat{\beta}))$ ist hierbei die Wurzel des α -Quantils einer χ^2 -Verteilung mit einem einzelnen Freiheitsgrad und Nichtzentralitätsparameter $\varepsilon'^2/v^2(\hat{\beta})$ zu verstehen. Schließlich können wir aufgrund der Konsistenz von $1/I(\hat{\beta})$ für $v^2(\hat{\beta})$ in (30) $1/v(\hat{\beta})$ durch $[I(\hat{\beta})]^{1/2}$ ersetzen. Insgesamt erhalten wir also einen asymptotischen Test für unser Problem des Nachweises der Äquivalenz der beiden Überlebensfunktionen $S_1(\cdot)$ und $S_2(\cdot)$, indem wir die *Nullhypothese der Nichtäquivalenz ablehnen für Werte des standardisierten geschätzten Regressionskoeffizienten, welche dem Betrag nach die kritische Schranke $C_\alpha(\varepsilon' \cdot [I(\hat{\beta})]^{1/2})$ unterschreiten.*

Beispiel 6. *Nachweis der Äquivalenz der ereignisfreien Überlebenskurven von Patienten aus kleinen und großen Zentren in der II. Medulloblastom-Studie der SIOP.* Für die abschließende Auswertung der Studie (vgl. [3]) standen insgesamt 357 Patienten zur Verfügung, die in der Zeit vom 01. 01. 1984 bis zum 31. 12. 1988 rekrutiert worden waren. Im Hinblick auf die Analyse der Studienergebnisse war es wichtig sicherzustellen, daß es zwischen Patienten, die in „großen Zentren“ (mindestens 8 Studienpatienten rekrutiert) behandelt wurden ($m = 176$), und solchen, die in Zentren mit einem Aufkommen von weniger als 8 rekrutierten Medulloblastom-Fällen („kleinen Zentren“) behandelt wurden ($n = 181$), hinsichtlich der ereignisfreien Überlebenskurven keine wesentlichen Unterschiede gab. Die in beiden Teilkollektiven beobachteten ereignisfreien Überlebenskurven sind in Abbildung 2 wiedergegeben. Als maximaler vertikaler Abstand zwischen äquivalenten wahren Überlebenskurven wurde $\delta = .15$ gewählt. Gemäß (28') sowie den in Tabelle 6 ausgewiesenen Werten bedeutet dies, daß bezüglich des Dummy-Regressionkoeffizienten β als Alternativhypothese die Aussage $-.4106 < \beta < .4106$ spezifiziert wurde.

Anhand der Studiendaten errechnete sich (unter Benutzung der SAS-Prozedur „phreg“) der Maximum-likelihood-Schätzer für β zu $\hat{\beta} = -.0944$ mit $I(\hat{\beta}) = 36.1981$ als zugehöriger beobachteter Fisher-Information, woraus sich für die standardisierte Teststatistik der Wert $[I(\hat{\beta})]^{1/2} \cdot |\hat{\beta}| = .5680$ ergibt. Andererseits hat man $\varepsilon'^2 \cdot I(\hat{\beta}) = .4106^2 \cdot 36.1981 = 6.1027$, und für das .05-Quantil der χ^2 -Verteilung mit Freiheitsgrad 1 und Nichtzentralitätsparameter 6.1027 findet man (z. B. durch Aufruf der SAS-Funktion „cinv“) die Zahl .6892. Entsprechend hat die kritische Konstante $C_\alpha(\varepsilon' \cdot [I(\hat{\beta})]^{1/2})$ hier den Wert $\sqrt{.6892} = .8302$. Da der beobachtete Wert der Teststatistik unterhalb dieser kritischen Schranke bleibt, entscheidet der Test auf dem 5%-Niveau hier zugunsten der Äquivalenz der zu vergleichenden Überlebensfunktionen.

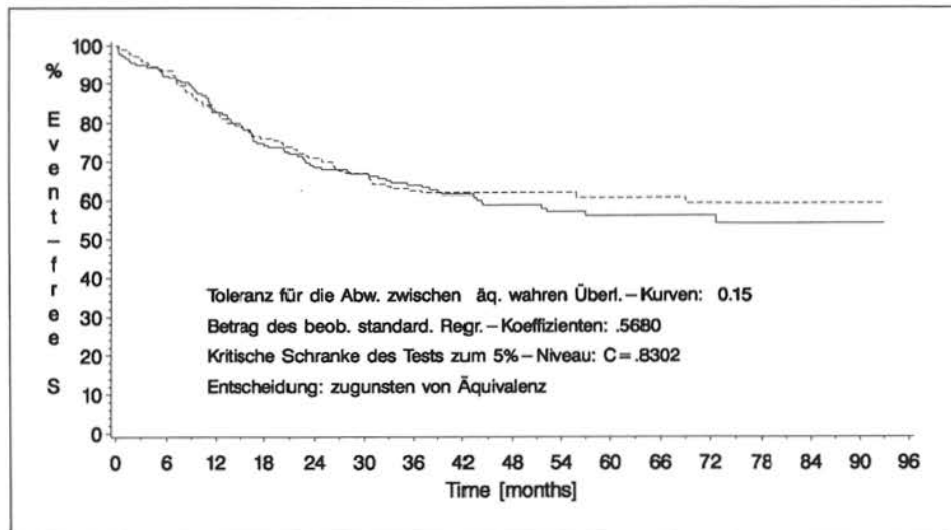


Abbildung 2: Medulloblastoma Trial SIOP/GPO II, beobachtete ereignisfreie ÜL-Kurven nach Zentrumsgröße [Durchgezogene Kurve: Große Zentren mit mindestens 8 Ptn.; Gestrichelte Kurve: Kleine Zentren mit weniger als 8 Ptn.]

Bemerkung. Der im vorliegenden Kapitel behandelte Test mit ÄQ zweier Überlebensfunktionen, für den in der Originalliteratur die Bezeichnung „*Log-Rank-Test auf Äquivalenz*“ eingeführt worden ist, basiert zunächst ausschließlich auf asymptotischen Überlegungen. Dies wirft naheliegender Weise die Frage auf, ob das Verfahren auch in endlichen Stichproben brauchbare Eigenschaften besitzt, insbesondere also in guter Näherung das nominelle Signifikanzniveau α enthält. Nach den Simulationsresultaten zu schließen, die in [22] hierzu vorgelegt worden sind, kommt es auch bei sehr kleinen Stichprobenumfängen nicht zu nennenswerten Niveauüberschreitungen.

6. Power-Berechnung und Fallzahlschätzung in Äquivalenzstudien

Hinsichtlich der Probleme, die mit der Berechnung der Power (Trennschärfe) und minimal benötigter Fallzahlen verbunden sind, genügt es, die Tests auf ÄQ im strikten (zweiseitigen) Sinne zu betrachten. Stehen für ein bestimmtes ÄQ-Testproblem effiziente Algorithmen zur Verfügung, die die Durchführung der erforderlichen Rechnungen auch für extreme Wahlen der ÄQ-Grenzen gestatten, erhält man mit denselben Verfahren Näherungen von jedem gewünschten Grad der Genauigkeit für die entsprechenden Größen beim zugehörigen Test auf einseitige ÄQ, indem man die rechte ÄQ-Grenze genügend groß werden läßt. Im übrigen laufen die Maßnahmen, die zur Adaptierung eines Algorithmus für Trennschärferechnungen oder Fallzahlplanung auf den einseitigen Fall erforderlich sind, stets auf mehr oder minder weitreichende Vereinfachungen hinaus.

Die nachfolgende Graphik (Abb. 3) gibt die exakte Gütefunktion des verbundenen t -Tests auf ÄQ speziell für die Situation des obigen Beispiels 4 wieder, d. h. für $n = 23$, $\alpha = .05$ und $\varepsilon = .50$. Man erkennt, daß es sich um eine streng symmetrische eingipflige Kurve handelt, die grob gesagt an die Dichtefunktion einer Normalverteilung erinnert. Anders

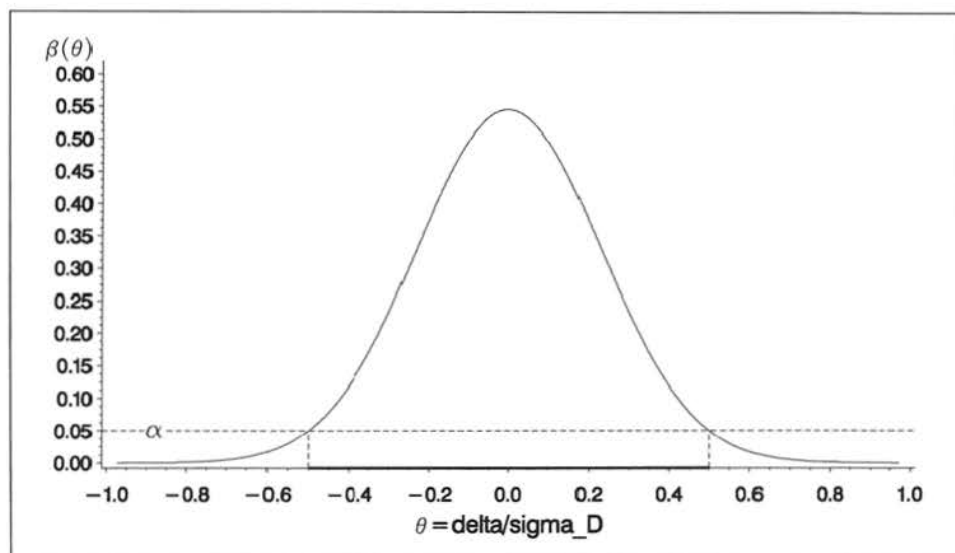
als man es von Tests für herkömmliche ein- oder zweiseitige Probleme kennt, läßt sich hier die Trennschärfe durch genügend extreme Festlegung einer speziellen Alternative *nicht* beliebig nahe an 100% bringen. Vielmehr liegt das Maximum der Gütefunktion, das bei der hier vorliegenden symmetrischen Situation exakt in der Mitte des hypothetischen ÄQ-Intervalls, i. e. in 0 erreicht wird, bei knapp 55%, obwohl es sich ja um einen optimalen Test handelt.

Auch bei nichtsymmetrischer Festlegung des ÄQ-Intervalls unter der Hypothese erfüllt die Gütefunktion beider t -Tests auf ÄQ stets die folgenden Bedingungen:

- (i) Es gibt ein globales Maximum, das im Innern des hypothetischen ÄQ-Intervalls angenommen wird.
- (ii) Die Gütefunktion nimmt links vom Maximum streng monoton zu, rechts davon ist sie streng monoton fallend.
- (iii) Die Kurve schneidet die Parallele zur Abszissenachse mit Abstand α genau in den Endpunkten des hypothetischen ÄQ-Intervalls.

Die für die Berechnung von Mindestfallzahlen relativ übersichtlichste Konstellation ist dann gegeben, wenn man die spezielle Alternative $\theta = \theta_0$ betrachtet, mit θ_0 als dem Mittelpunkt des unter der ÄQ-Hypothese spezifizierten Intervalls (entsprechend der Nullhypothese beim zugehörigen zweiseitigen Testproblem). Mit etwas erhöhtem numerischen Aufwand ist es aber genauso gut möglich, unter irgendeiner anderen speziellen Parameterkonstellation, die im ÄQ-Bereich liegt, die Fallzahlen zu berechnen, die benötigt werden, um eine vorgegebene Trennschärfe zu erreichen. Bei den t -Tests auf ÄQ ist die technische Basis für alle derartigen Rechnungen eine Routine für die nichtzentrale t -Verteilungsfunktion. In SAS hat man eine solche wiederum direkt als vordefinierte Funktion zur Verfügung (unter der Beziehung „probt“). Für den Intervallinklusions-Test auf Äquivalenz zweier Gauß-Verteilungen mit homogenen Varianzen bezüglich der nichtstandardisierten Differenz der Erwartungswerte, der beim Bio-äquivalenz-Nachweis gegenwärtig routinemäßige Verwendung findet (vgl. Beispiel 3), ist

Abbildung 3: Gütefunktion (Power function) des verbund. t -Tests auf ÄQ für $n = 23$, $\alpha = 5\%$ und $-.50 < \delta/\sigma_D < +.50$ als hypothet. ÄQ-Bereich



die exakte Berechnung von Trennschärfewerten und Mindestfallzahlen unter beliebigen Alternativen ebenfalls mit relativ geringem numerischen Aufwand möglich [5, 15]. Die eingehende Untersuchung von ÄQ-Tests für die verschiedensten speziellen Situationen zeigt allerdings übereinstimmend, daß für die Wahl von Alternativen, unter denen die Fallzahlplanung erfolgen soll, nur eine sehr enge Umgebung von θ_0 , d. h. dem identische Verteilungen charakterisierenden Parameterwert in Betracht kommt. Denn bereits unter den bezüglich der Trennschärfe „günstigsten“ Alternativen haben die Mindestfallzahlen für ÄQ-Studien oft eine praktisch nicht zu realisierende Größenordnung.

Wie der Tabelle 7 zu entnehmen ist, trifft dies auch für optimale Tests zu, falls nicht das ÄQ-Kriterium von vornherein so schwach formuliert wird, daß die statistische Absicherung der zugehörigen Hypothese kaum noch von sachlichem Interesse ist. Die Tabelle bezieht sich auf den unverbundenen t -Test auf ÄQ zweier Normalverteilungen bezüglich der standardisierten Differenz der Mittelwerte mit ÄQ-Grenzen $\pm .50$. Das somit unter der Alternativhypothese angesetzte ÄQ-Kriterium kann in Anbetracht der Tatsache, daß die Bedingung $|\mu_1 - \mu_2|/\sigma < .50$ gemäß (15), (15) und (16) gleichbedeutend ist mit $|P[X_i > Y_j] - 1/2| < .1382$ auch hier kaum als extrem restriktiv bezeichnet werden. Trotzdem benötigt man selbst unter einer so nahe am Mittelpunkt des hypothetischen ÄQ-Bereichs gelegenen Alternative wie $|\mu_1 - \mu_2|/\sigma < .125$ (i. e. $P[X_i > Y_j] = .4648$) bereits mehr als 90 Fälle in jeder Gruppe, um eine Trennschärfe von 80% zu erreichen (vgl. Tab. 7, vorletzte Spalte).

Für den *exakten Fisher-Test auf ÄQ* ist bereits die exakte Berechnung einzelner Trennschärfe-Werte sehr aufwendig, wobei der Rechengang hier nur kurz skizziert werden soll. Als erstes hat man für alle möglichen Werte der Spaltensumme S (vgl. das Kontingenztafelschema in Tab. 2) die vier kritischen Konstanten zur Entscheidungsregel (21) zu bestimmen und den Ausdruck auf der linken Seite von (22) auszurechnen, wobei q_1 zu ersetzen ist durch den unter der interessierenden Alternative gültigen Wert q der Odds-Ratio. Dies ergibt die „bedingte Power“, gegeben den jeweiligen Wert von S , der wie gesagt die Menge aller ganzen Zahlen zwischen 0 und N durchlaufen muß. Anschließend müssen die bedingten Trennschärfen gewichtet aufsummiert werden, wobei die Gewichte gegeben sind als die Wahrscheinlichkeiten der verschiedenen Werte von S . Die Verteilung von S ist nur unter einer einzelnen Konstellation, nämlich $q = 1$ (i. e. Identität der Verteilungen von X und Y) wieder eine einfache Binomialverteilung und muß in allen anderen Fällen neu berechnet werden. Der Quellencode eines FORTRAN-Programms, das alle erforderlichen Rechenschritte für beliebige Parameterkonstellatio-

Tabelle 7: Ausgewählte Werte der Gütefunktion des unverbundenen t -Tests auf ÄQ zum Niveau $\alpha = .05$ bei gleichgroßen Stichproben des Umfangs n sowie $\varepsilon_1 = \varepsilon_2 = \varepsilon = .50$.

n	$\beta(3\varepsilon/4)$	$\beta(\varepsilon/2)$	$\beta(\varepsilon/4)$	$\beta(0)$
10	0.06564	0.07972	0.08957	0.09312
20	0.08564	0.12578	0.15840	0.17106
30	0.10920	0.19071	0.26645	0.29786
40	0.13223	0.26399	0.39915	0.45801
50	0.15134	0.32928	0.52149	0.60710
60	0.16735	0.38330	0.61878	0.72319
70	0.18184	0.42965	0.69400	0.80781
80	0.19563	0.47130	0.75282	0.86795
90	0.20903	0.50975	0.79956	0.91005
100	0.22216	0.54565	0.83719	0.93920

nen (p_1, p_2) und Gesamtfallzahlen N bis zu 2000 ausführt, ist im Anhang zu [23] abgedruckt. In unserem *Beispiel 5* findet man auf diesem Wege unter der Annahme, daß die wahren Resonderraten mit den beobachteten Raten übereinstimmen, daß also $(p_1, p_2) = (.4800, .5294)$ gilt, als Trennschärfe des exakten Fisher-Tests auf ÄQ die Zahl $.5466 \approx 55\%$. Läßt man keine randomisierten Entscheidungen zwischen den Hypothesen zu, reduziert sich dieser Wert auf knapp 48%.

Noch wesentlich größer ist bei vorgegebener Trennschärfe der Aufwand für die Berechnung von im exakten Fisher-Test auf ÄQ erforderlichen Stichprobenumfängen, da hier bei festliegendem Verhältnis der beiden Stichprobengrößen beim Iterieren die Power für verschiedene Werte der Gesamtfallzahl N ermittelt werden muß. Auch hierfür findet man in [23] den Quellencode eines FORTRAN-Programms.

Was schließlich den in Kap. 5 behandelten Log-Rank-Test auf ÄQ von zwei Überlebenskurven betrifft, so sind Berechnungen von Trennschärfewerten und Mindestfallzahlen in exakter Form ebensowenig praktikabel wie im herkömmlichen Log-Rank-Test. Man ist hier auf Approximationen auf der Basis asymptotischer Verteilungen angewiesen, und selbst geeignete Näherungsverfahren sind zur Zeit noch nicht in anwendungsreifer Form verfügbar.

Literaturangaben

- [1] ANDERSON, S., HAUCK, W. W. (1983): A new procedure for testing equivalence in comparative bioavailability trials. *Communications in Statistics – Theory and Methods* **A12**, 2663–2692.
- [2] ANDERSON, S., HAUCK, W. W. (1990): Consideration of individual bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* **18**, 259–73.
- [3] BAILEY, C. C., GNEKOW, A., WELLEK, S., JONES, M., ROUND, C., BROWN, J., PHILIPPS, A., NEIDHARDT, M. K. (1994): Prospective randomised trial of chemotherapy given before radiotherapy in childhood medulloblastoma. *International Society of Paediatric Oncology (SIOP) and the (German) Society of Paediatric Oncology (GPO): S.I.O.P. II. Medical and Pediatric Oncology* **22**.
- [4] COX, D.R., SNELL, E. J. (1989): *Analysis of Binary Data*. Second Edition. London: Chapman and Hall.
- [5] 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.
- [6] GRIZZLE, J. E. (1965): The two-period change-over design and its use in clinical trials. *Biometrics* **21**, 467–80.
- [7] HAUSCHKE, D., STEINJANS, V. W., DILETTI, E. (1990): A distribution-free procedure for the statistical analysis of bioequivalence studies. *International Journal of Clinical Pharmacology, Therapy and Toxicology* **28**, 72–78.
- [8] HOLZGREVE, H., DISTLER, A., MICHAELIS, J., PHILIPP, T., WELLEK, S. (1989): Verapamil versus hydrochlorothiazide for the treatment of hypertension. Results of a long-term double-blind comparative trial. *British Medical Journal* **299**, 881–6.
- [9] LEHMANN, E. L. (1986): *Testing Statistical Hypotheses*. 2nd Edition. New York: John Wiley & Sons, Inc.
- [10] MAU, J. (1988): A statistical assessment of clinical equivalence. *Statistics in Medicine* **7**, 1267–77.
- [11] MAU, J. (1991): Vergleich einiger Verfahren zur statistischen Beurteilung der Bioäquivalenz sofort freisetzender Arzneiformen. In J. Vollmar (Hg.), *Bioäquivalenz sofort freisetzender Arzneiformen*. Stuttgart: Gustav Fischer Verlag, 13–44.
- [12] METZLER, C. M. (1974): Bioavailability – a problem in equivalence. *Biometrics* **30**, 309–17.
- [13] MILLER, K. W., WILLIAMS, D. S., CARTER, S. B., JONES, M. B., MISHALL, D. R. (1990): The effect of olestra on systemic levels of oral contraceptives. *Clinical Pharmacology and Therapeutics* **48**, 34–40.

- [14] PATEL, H. I., GUPTA, G. D. (1984): A problem of equivalence in clinical trials. *Biometrical Journal* **26**, 471–4.
- [15] PHILLIPS, K. F. (1990): Power of the two one-sided tests procedure in bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* **18**, 137–44.
- [16] SCAGLIONE, F. (1990): Comparison of the clinical and bacteriological efficacy of clarithromycin and erythromycin in the treatment of streptococcal pharyngitis. *Current Medical Research and Opinion* **12**, 25–33.
- [17] SCHUIRMANN, D. J. (1987): A comparison of the two one-sided tests procedure and the power approach for assessing equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* **15**, 657–80.
- [18] STEINJANS, V. W., DILETTI, E. (1983): Statistical analysis of bioavailability studies: Parametric and nonparametric confidence intervals. *European Journal of Clinical Pharmacology* **24**, 127–36.
- [19] STEINJANS, V. W., DILETTI, E. (1985) Generalisation of distribution-free confidence intervals for bioavailability ratios. *European Journal of Clinical Pharmacology* **28**, 85–8.
- [20] UNGERSBÖCK, K., KEMPSKI, O. (1992): Experimentelle Untersuchungen zur medikamentösen Beeinflussung der zerebralen Mikrozirkulation. *Techn. Bericht, Neurochirurg. Klinik der Univ. Mainz*.
- [21] WELLEK, S. (1993a): Basing the analysis of comparative bioavailability trials on an individualized statistical definition of equivalence. *Biometrical Journal* **35**, 47–55.
- [22] WELLEK, S. (1993b): A log-rank test for equivalence of two survivor functions. *Biometrics* **49**, 877–81.
- [23] WELLEK, S. (1994): *Statistische Methoden zum Nachweis von Äquivalenz*. Stuttgart: Gustav Fischer Verlag.
- [24] WESTLAKE, W. J. (1972): Use of confidence intervals in analysis of comparative bioavailability trials. *Journal of Pharmaceutical Sciences* **61**, 1340–1.
- [25] WESTLAKE, W. J. (1976): Symmetrical confidence intervals for bioequivalence trials. *Biometrics* **32**, 741–4.
- [26] WESTLAKE, W. J. (1979): Statistical aspects of comparative bioavailability trials. *Biometrics* **35**, 273–80.
- [27] WESTLAKE, W. J. (1981): Response to T.B.L. Kirkwood: bioequivalence testing – a need to rethink. *Biometrics* **37**, 591–3.

Anschrift des Verfassers: Prof. Dr. S. Wellek, Z1 Mannheim, Abteilung Biostatistik, Postfach 12 21 20 68072 Mannheim

Eingang des Manuskripts 14. 10. 93
Eingang der revidierten Fassung 10. 5. 94

ORIGINALARBEITEN

Evaluation der Akzeptanz eines Krankenhaus-Informationssystems bei den ärztlichen Mitarbeitern eines Universitätskrankenhauses

H. U. Prokosch, J. Dudeck

Zusammenfassung

Das Gießener Krankenhaus-Informationssystem WING befindet sich seit Anfang 1992 in zwei klinischen Abteilungen (Innere Medizin und Chirurgie) im Routineeinsatz. Neben den WING-Funktionen zur Verwaltung klinischer Patientendaten stehen den Anwendern an allen DV-Arbeitsplätzen des Klinikumsnetzwerks (GISNET) verschiedene DOS-Standardprogramme (Textverarbeitung, Graphikprogramme, Tabellenkalkulation, etc.) sowie nationale und internationale Informationsbanken (Rote Liste, Medline, etc.) zur Verfügung. Zur Evaluierung der Zufriedenheit der Benutzer mit den bis zu diesem Zeitpunkt verfügbaren WING/GISNET-Funktionen und zur Unterstützung bei der Planung zukünftiger Erweiterungen wurde im Frühjahr 1993 eine Umfrage unter den ärztlichen Anwendern durchgeführt. Die Ergebnisse dieser Umfrage weisen auf eine mittlerweile sehr positive und aufgeschlossene Einstellung der Ärzteschaft gegenüber dem Computereinsatz im klinischen Umfeld hin. Die Ärzte/Ärztinnen nutzen die angebotenen Funktionen regelmäßig und bestätigen, daß ihnen diese bei der täglichen Arbeit von Nutzen sind. Auch zukünftigen Erweiterungen stehen sie sehr positiv gegenüber, so daß mittlerweile bereits neue Funktionen bereitgestellt worden sind. Allerdings zeigt es sich ebenfalls, daß die flächendeckende Einführung von WING/GISNET im Gießener Klinikum einen Betreuungsaufwand erfordert, der mit der derzeitigen Personaldecke nicht zufriedenstellend abgedeckt werden kann. Um langfristig die Informationsbedürfnisse der klinischen Anwender nicht aus den Augen zu verlieren und eventuelle Probleme frühzeitig erkennen zu können, sind regelmäßige Gespräche mit den klinischen Nutzern auf den Stationen notwendig. In Anbetracht der kontinuierlich steigenden Zahl an GISNET angeschlossener Stationen und klinischer WING/GISNET-Anwender sind hierfür zusätzliche Personalkapazitäten einzuplanen.

Schlüsselwörter

Evaluation, Krankenhaus-Informationssystem, Benutzerumfrage

Key words

Evaluation, Hospital Information System, Physician Attitude, Physician Questionnaire

Summary

Since 1992 the hospital information system WING is routinely used in two clinical departments (internal medicine and surgery) of Gießen University Hospital. Besides the WING-functions to access clinical patient data, DOS-programs (e.g. text processing, graphics, spread sheets) and informationbases (e.g. Medline and drug information systems) are available at every workstation of the hospital network (GISNET). To evaluate the user satisfaction and to support the planning of future system extensions a questionnaire was send to every physician of the two departments in spring 1993. The physician's answers showed a very positive attitude towards the application of computers in the clinical environment. They are regularly using the various WING/GISNET functions and attested them a great value for their daily work. The majority was also convinced that extending the WING/GISNET-functionality would further enhance their working conditions. Consequently some information systems which where heavily requested have already been installed. However, it became also obvious that the continuing expansion of WING and GISNET throughout the whole hospital requires increased maintenance and user education efforts as well as regular visits on the wards. Thus, the group of staff members concerned with keeping in touch with the clinical users to react on their upcoming information needs and system improvement proposals needs to be increased.

1. Einleitung

Am Gießener Universitätsklinikum wurden in den letzten Jahren verschiedene einander ergänzende Anstrengungen unternommen, um das klinische Arbeitsumfeld von Ärzten und Pflegekräften durch die Einführung eines Krankenhaus-Informationssystems (KIS) zu verbessern (PROKOSCH et al. 1991). Um allen Klinikumsmitarbeitern den Zugriff auf die verschiedensten DV-Funktionalitäten dieses KIS zu gewährleisten wurde ein klinikumsweites Netzwerk (GISNET) aufgebaut. Über GISNET werden 25 verschiedene Gebäude im Gelände des Gießener Universitätsklinikums mit Lichtwellenleitern (LWL) verbunden. GISNET umfaßt zur Zeit zwei Tandem Großrechner, einen Tandem Unix-Rechner, eine HP 9000, zwei Modcomp Doppelprozessorsysteme, etwa 20 SUN Sparstations, 15 Novell-Fileserver, ca. 350 discless PCs und die gleiche Anzahl von Druckern, sowie eine wachsende Anzahl von Macintosh-Rechnern. Als Netzwerkprotokoll werden TCP/IP, IPX und Netbios eingesetzt. In den bisher (Frühjahr 1994) an GISNET angebotenen klinischen Abteilungen (Innere Medizin, Allgemeinchirurgie, Herzchirurgie, Unfallchirurgie, Urologie, Neurologie, Hals-Nasen-Ohren-Heilkunde und Gynäkologie) stehen DV-Arbeitsplätze (jeweils ein 386er/486er PC mit 2/4 MB Hauptspeicher ohne lokale Platte und Laufwerk, sowie ein Laserdrucker) in jedem Stationszimmer zur Verfügung. Von jedem dieser Arbeitsplätze aus können die zur Zeit etwa 800 aktiven klinischen Anwender auf eine Vielzahl unterschiedlicher Ressourcen zugreifen.

Die anfangs nur auf einigen Pilotstationen erprobten WING-Funktionalitäten stehen somit seit Anfang 1992 einem immer größer werdenden Kreis klinischer Nutzer zur Verfügung. Damit der zukünftige Ausbau des Gießener Krankenhaus-Informationssystems auch weiterhin eng an die Bedürfnisse der klinischen Anwender angelehnt werden kann, wurden verschiedene Maßnahmen in die Wege geleitet, um die bisherigen Realisierungen zu evaluieren, die Zufriedenheit der Nutzer mit den vorhandenen Funktionen zu überprüfen und ihre Wünsche für den Systemausbau zu eruieren. Ein wesentlicher Schritt in diesem Evaluationsprozeß bestand in der Durchführung einer Befragung des ärztlichen Personals der Medizinischen und Chirurgischen Klinik. Der Inhalt dieser Umfrage und ihre Ergebnisse werden im Folgenden vorgestellt.

2. Methodik

Im Frühjahr 1993 wurde ein 7-seitiger Fragebogen an 149 ärztliche Mitarbeiter der Chirurgischen und der Medizinischen Klinik des Gießener Universitätsklinikums versandt. Der Fragebogen bestand aus

- einem Block demographischer Daten des ausfüllenden Arztes,
- einer Frage nach der Häufigkeit, mit der einzelne WING/GISNET-Funktionen genutzt werden,
- einer Frage zur Bewertung des Nutzens, den einzelne WING/GISNET-Funktionen für den Arzt/die Ärztin haben,
- einem Frageblock, in dem die generelle Einstellung des Arztes/der Ärztin gegenüber der elektronischen Datenverarbeitung im allgemeinen und einigen WING-Funktionen im speziellen erfragt wird,
- einem Frageblock zum WING-Hilfesystem,
- einer Frage zur Bewertung des erwarteten Nutzens der DV-gestützten Befundübermittlung und -präsentation für verschiedene Funktionsbereiche des Klinikums und
- einem Frageblock zur Bewertung möglicher Erweiterungen innerhalb von WING/GISNET.

Die Befragung beschränkte sich auf die ärztlichen Mitarbeiter(innen) der Chirurgischen Klinik (Allgemeinchirurgie, Herzchirurgie und Urologie) sowie der Medizinischen Klinik, da die Stationen dieser zwei Kliniken am längsten an GISNET angebunden sind und die entsprechenden GISNET/WING-Funktionen dort bereits seit über einem Jahr genutzt werden konnten. In den Stationsbereichen steht dort jeweils ein DV-Arbeitsplatz (incl. Drucker) im Stationszimmer und in jedem der zur Station gehörenden Arztzimmer zur Verfügung.

An diesen DV-Arbeitsplätzen konnte zum Zeitpunkt der Befragung auf die folgenden Anwendungen zugegriffen werden: (1) das klinische Informationssystem WING, (2) das Patientendatenverwaltungssystem PILS, (3) ein Abteilungssystem für die Onkologie/Hämatologie (ONCOSYS) sowie (4) eine wachsende Anzahl von elektronischen Informationsdiensten (die Rote Liste, Medline, Drugdex, Poisindex, Emergindex, Drug Interactions, Martindale, Reprorisk und der Bundesbahnfahrplan). Innerhalb von WING standen den Ärzten Module zum Anzeigen demographischer Patientendaten, zum Ausdrucken von Patientenetiketten, zur Präsentation der Untersuchungsergebnisse des Klinisch-Chemischen Labors und zur Diagnosedokumentation und -klassifikation zur Verfügung. Funktionen zur Dokumentation der Arzneimitteltherapie sowie zur Arztbriefschreibung waren als Prototypen in jeweils einer Station der Medizinischen Klinik in Erprobung.

3. Ergebnisse

Von den 149 verschickten Fragebögen wurden innerhalb eines Zeitraums von 2 Wochen 68 ausgefüllt zurückgesandt (Rücklaufquote = 46%). 78% der Antworten stammen von ärztlichen Mitarbeitern männlichen Geschlechts und 22% von weiblichen Mitarbeiterinnen. Die Rücklaufquote bei den Männern lag bei 53%, unter den Frauen bei 38%. Bezogen auf die vier angesprochenen Fachgebiete ergaben sich Rücklaufquoten von 55% für die Allgemeinchirurgie, 39% für die Herzchirurgie, 31% für die Urologie und 46% für die Innere Medizin. Bezogen auf den Ausbildungsstand der Mitarbeiter ergaben sich Rücklaufquoten von 41% für Chef-/Oberärzte, 56% für Assistenzärzte, 56% für Ärzte im Praktikum und 29% für Studenten im Praktischen Jahr. Das Durchschnittsalter lag bei 34 Jahren. 40% der Mitarbeiter(innen) hatten bereits seit

mehr als 12 Monaten mit WING/GISNET gearbeitet (7 bis 12 Monate = 26%, 3 bis 6 Monate = 16%, weniger als 3 Monate = 18%).

Die Fragen 1 und 2 bezogen sich auf die Häufigkeit mit der der Arzt/die Ärztin eine bestimmte WING/GISNET-Funktion benutzt und die Bewertung des Nutzens, den die entsprechende Funktion für ihn/sie hat. Für beide Fragen wurden (1) vier unterschiedliche Darstellungsformen der Labordaten (Kumulativbericht auf dem Bildschirm, Kumulativbericht gedruckt, neueste Werte/Patient und neueste Werte/Station), (2) die WING-Funktion zur Diagnosedokumentation und (3) der Druck von Patientenetiketten sowie (4) die über GISNET verfügbaren Informationsdienste Medline, die Rote Liste, die Micromedix-Informationenbanken (Drugdex, Poisindex, Emergindex, etc.) und der Bundesbahnfahrplan zur Beantwortung aufgeführt. Die Nutzungsfrequenz wurde zwischen ‚mehrmals täglich‘ und ‚überhaupt nicht‘ in fünf Klassen unterteilt. Zur Bewertung des Nutzens einer Funktion konnte ein ganzzahliger Wert zwischen 0 und 5 angekreuzt werden, wobei 0 ‚kein Nutzen‘ bedeutete und 5 für ‚sehr nützlich‘ stand. Die Ergebnisse der entsprechenden Antworten sind in den Abbildungen 1 und 2 zusammengefaßt.

In Frage 3 wurde die Zustimmung bzw. Ablehnung zu bestimmten Aussagen zum Einsatz von Computern in der Medizin und zu einigen WING-Modulen auf einer ganzzahligen Skala zwischen 0 (Ablehnung) und 5 (Zustimmung) erfragt. Zur graphischen Darstellung der Ergebnisse bei der Beantwortung dieser Frage wurde in Abbildung 3 der mittlere Wert von 2,5 als „unentschlossen“ interpretiert und als Basis für den Grad der Ablehnung bzw. Zustimmung gewählt.

Um die generelle Einstellung der Mitarbeiter gegenüber Computern sowie zu dem WING-Benutzerhandbuch und den WING-Schulungen zu bestimmen, mußten 13 Aussagen in Frage 4 mit ‚ja‘ oder ‚nein‘ beantwortet werden. Die jeweilige Verteilung der Antworten ist in Tabelle 1 dargestellt.

Die Antworten bezüglich des über eine Funktionstaste aufrufbaren WING-Hilfesystems ergaben, daß dieses Hilfesystem lediglich 61% der Benutzer bekannt ist (bei 9% ohne Antwort). 67% der Befragten gaben an, daß sie das WING-Hilfesystem nie benutzen (29% selten, 3% häufig und 1% ohne Antwort). Eine Beurteilung des WING-Hilfesystems wurde nur von 44% der Befragten abgegeben. Unter diesen Antworten bezeichneten es 15% als gut, 65% als notwendig aber verbesserungswürdig und 20% als überflüssig. Die Möglichkeit einen Verbesserungsvorschlag für die Gestaltung des WING-Hilfesystems zu machen wurde allerdings von niemandem genutzt.

Frage 6 diente dazu, den von den Ärzten/Ärztinnen erwarteten Nutzen einer DV-basierten Befundübermittlung und -präsentation bezogen auf bestimmte Funktionsbereiche des Klinikums zu bestimmen. Dazu wurden sie gebeten, diesen Nutzen für 10 vorgegebene Funktionsbereiche auf einer Skala von 0 (= nicht notwendig) bis 5 (= sehr hilfreich) anzugeben. Da die Übermittlung und Präsentation der Befunde des Klinisch-Chemischen Labors bereits seit 1989 in WING zur Verfügung steht wurde deren Nutzen an dieser Stelle nicht mehr erfragt. Zusätzlich konnten von den Befragten weitere Funktionsbereiche mit der entsprechenden Bewertung aufgeführt werden. Die durchschnittlichen Bewertungen sind in Abbildung 4 dargestellt.

Abschließend wurden die Ärzte/Ärztinnen gebeten den von ihnen erwarteten Nutzen für 11 mögliche WING/GISNET-Erweiterungen anzugeben. Die durchschnittlichen Bewertungen für die im Fragebogen vorgegebenen WING/GISNET-Funktionen sind in Abbildung 5 dargestellt. Die Möglichkeit, selbst Erweiterungsvorschläge zu machen oder generelle Verbesserungsvorschläge sowie Kritik zu äußern wurde von keinem der Befragten genutzt.

Erfahrungen aus früheren Gesprächen und Anwendungsdemonstrationen für Ärzte haben gezeigt, daß diese oftmals DV-Funktionen positiv bewerten, sich selbst aber nicht als Anwender dieser Funktionen sehen. Vielmehr zeigte die Praxiseinführung dann meist, daß sie der Meinung waren, diese Funktionen müßten vom Pflegepersonal oder von

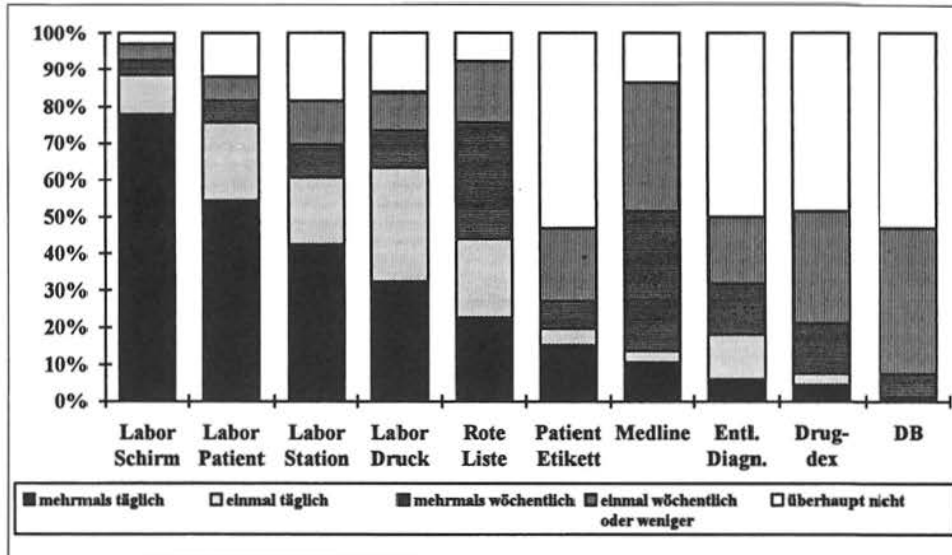
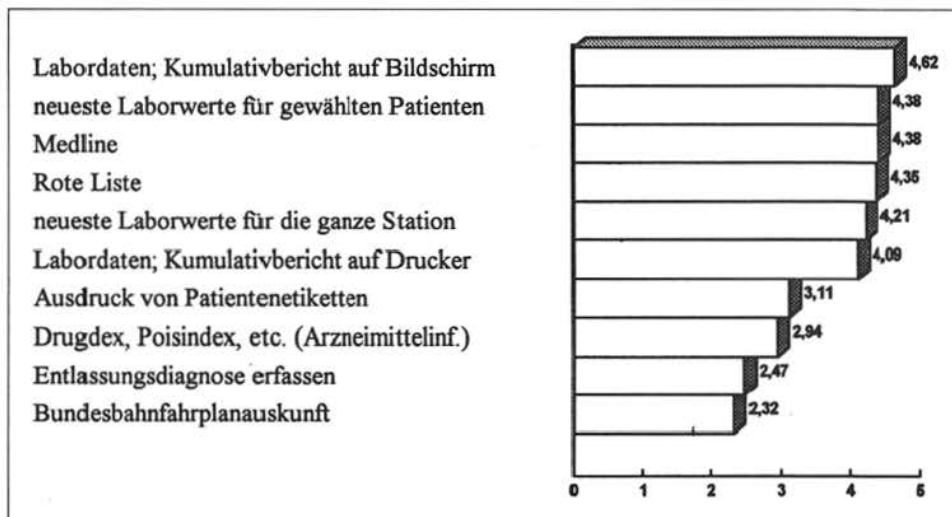


Abbildung 1: Darstellung der Benutzungshäufigkeit der WING-Funktionen zur Darstellung des Labordaten-Kumulativberichts am Bildschirm (Labor Schirm), zum Ausdruck des Kumulativberichts (Labor Druck), zur Anzeige der neuesten Werte für einen Patienten (Labor Patient), zur Anzeige der neuesten Werte für eine Station (Labor Station), zum Ausdrucken von Patientenetiketten (Patient Etikett) und zur Dokumentation der Entlassungsdiagnosen (Entl. Diagn.) sowie der Informationsbanken Rote Liste, Medline, Drugdex und dem deutschen Bundesbahnfahrplan (DB).

Abbildung 2: Durchschnittliche Bewertung des Nutzens verschiedener WING/GISNET-Funktionen auf eine Skala von 0 bis 5 (0 = kein Nutzen, 5 = sehr nützlich).



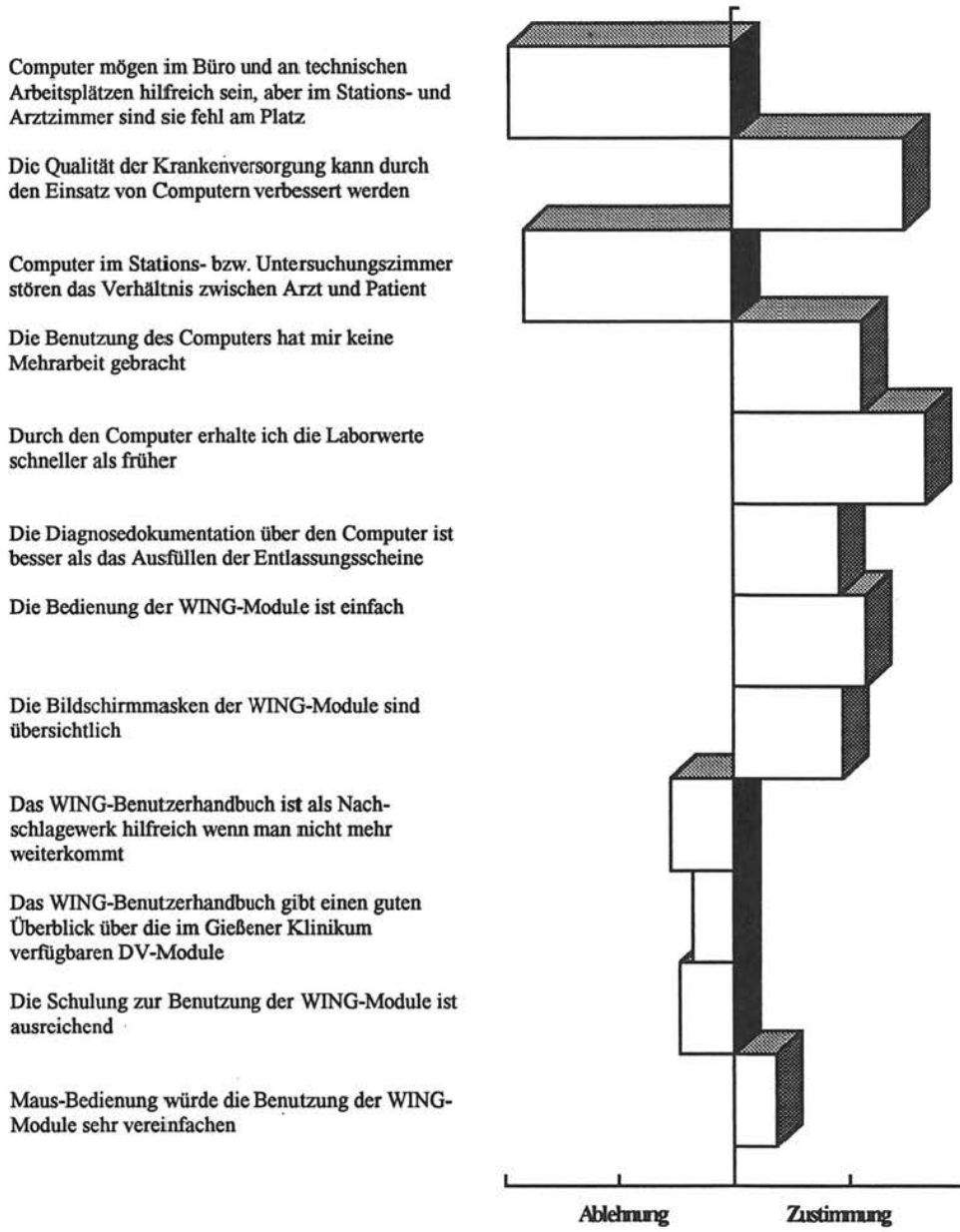


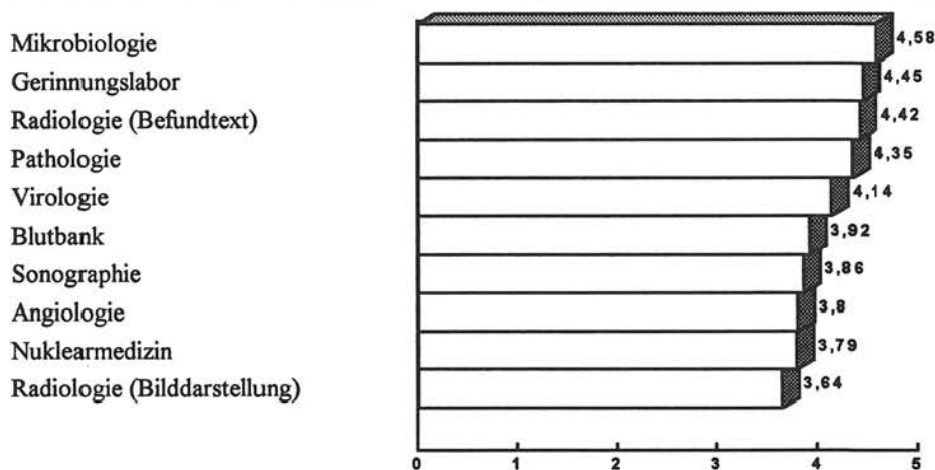
Abbildung 3. Ablehnung bzw. Zustimmung zu Aussagen zum Einsatz von Computern in der Medizin. Es wurden jeweils die durchschnittlichen absoluten Differenzen zu 2,5 (unentschieden) aufgetragen.

Sekretärinnen bzw. Schreibkräften benutzt werden. Aus diesem Grund wurde in Frage 7 (Bewertung eventueller neuer WING/GISNET-Funktionen) zusätzlich gefragt, wen die Ärzte als Nutzer der entsprechenden Funktion sehen würden. Lediglich die DV-

Tabelle 1: Verteilung der Ja/Nein-Antworten zu den Aussagen in Frage 4

Aussage	Prozent Ja-Antwort	Prozent Nein-Antwort
Ich besitze auch privat einen PC	80%	20%
Ich programmiere ab und zu auch mal selbst (z. B. Clipper, Dbase)	24%	76%
Ich kann mit einem Computer überhaupt nichts anfangen	2%	98%
Ich habe bereits einen/mehrere EDV-Kurse am Institut für Medizinische Informatik besucht	20%	80%
Ich habe andernorts bereits an einem/mehreren EDV-Kursen teilgenommen	18%	82%
Das WING-Benutzerhandbuch habe ich vollständig gelesen	5%	95%
Das WING-Bnutzerhandbuch habe ich nur kurz durchgeblättert	44%	56%
Das WING-Benutzerhandbuch benutze ich ab und zu wenn ich etwas nachschlagen will	26%	74%
Das WING-Benutzerhandbuch habe ich noch nie gelesen	47%	53%
Ich habe mich in die Benutzung der WING-Module selbst eingearbeitet	73%	27%
Die Benutzung der WING-Module wurde mir durch Kollegen(innen) gezeigt	70%	30%
Ich wurde durch Mitarbeiter des Instituts für Medizinische Informatik in der Benutzung der WING-Module geschult	14%	86%
Ich würde es begrüßen, wenn noch weitere Tätigkeiten im Umfeld der Krankenversorgung durch EDV unterstützt würden	91%	9%

Abbildung 4: Die durchschnittliche Bewertung des erwarteten Nutzens des DV-basierten Befundübermittlung und -präsentation auf einer Skala von 0 bis 5 (0 = nicht notwendig, 5 = sehr hilfreich)



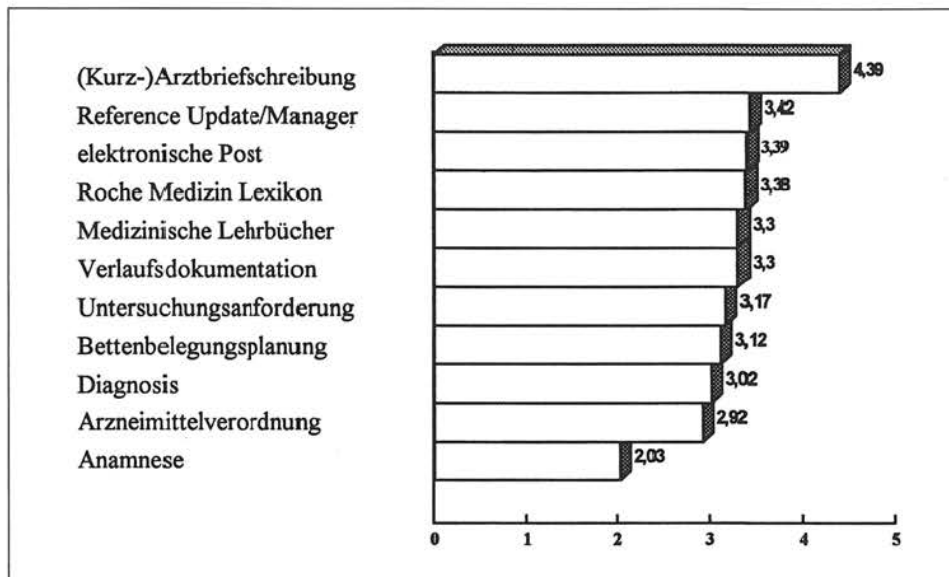


Abbildung 5: Die durchschnittliche Bewertung des erwarteten Nutzens möglicher neuer WING/GISNET-Module auf einer Skala von 0 bis 5 (0 = unnötig, 5 = sehr hilfreich)

Funktionen zur Untersuchungsanforderung und zur Bettenbelegungsplanung wurden eher als pflegerische Funktionen eingestuft. Bei allen anderen aufgeführten Funktionen sahen sich jeweils mindestens 58% der Ärzte als die eigentlichen Nutzer (elektronische Post = 58%, Arzneimittelverordnung = 59%, Verlaufsdokumentation = 62%, Reference Update/Manager = 64%, Medizinische Lehrbücher = 65%, Roche Medizin Lexikon = 71%, Diagnose = 73%, Kurzarztbriefschreibung = 74%).

4. Diskussion

Krankenhaus-Informationssysteme sind dynamische Systeme, deren Einführung sich in der Regel über viele Jahre, meist sogar Jahrzehnte erstreckt. In diesem Zeitraum ergeben sich viele technologische Neuerungen, neue gesetzliche Regelungen sind zu berücksichtigen und auch die Krankenhaus-internen Rahmenbedingungen sind in einem ständigen Wandel begriffen. Die Einführung von Krankenhaus-Informationssystemen darf deshalb nicht nach der Implementationsphase als abgeschlossen betrachtet werden. 1990 kritisierten Safran et al. viele KIS-Installationen mit dem Hinweis darauf, daß vielerorts die Ärzte das installierte System oft noch Jahre nach seiner Installation nicht benutzen. Gründe dafür liegen meist im schlechten Anwendungsdesign und der umständlichen Bedienbarkeit der Anwendungen sowie darin, daß die Anwendungen nicht den ärztlichen Anforderungen und Bedürfnissen angepaßt sind (SAFRAN et al. 1990). Auch andernorts wird wiederholt über Probleme berichtet, Ärzte zur direkten Nutzung von Computern zu bewegen (vgl. z. B. SPILLANE et al. 1990 und ZIBRAK et al. 1990). Andererseits können viele der Ansätze, durch DV-basierte entscheidungsunterstützende Verfahren zur Qualitätssicherung und Kostenreduktion beizutragen, nur dann wirken, wenn der Arzt Hinweise und Vorschläge im direkten Arzt-Computer-Dialog entgegennimmt und unmittelbar darauf reagieren kann. Über entsprechende positive Erfahrungen

berichten z. B. LEPAGE et al. (1992), EVANS et al. (1993) und TIERNEY et al. (1993). Es ist deshalb wichtig, ständig den Kontakt mit den klinischen Nutzern eines Krankenhaus-Informationssystems zu halten, um frühzeitig Schwachstellen des Systems sowie neue Befürfnisse zu erkennen und entsprechende Maßnahmen in die Wege leiten zu können. Es ist notwendig, die Nutzung der implementierten Anwendungssoftware durch die klinischen Mitarbeiter zu evaluieren, aber auch zu überprüfen, welche Auswirkungen die Einführung von DV-Verfahren möglicherweise auf das klinische Arbeitsumfeld hat. SITTIG (1993) beschreibt dazu verschiedene Möglichkeiten der Durchführung von Beobachtungsstudien (vgl. u. a. auch GOUVEIA-OLIVEIRA et al. 1991 und CONNELLY et al. 1993), aber auch die Befragung der Nutzer (vgl. z. B. GILHOOLY et al. 1991, CONNELLY et al. 1993 und SANDS et al. 1994) sowie das in klinische Anwendungen integrierte Monitoring zur Bestimmung des Nutzungsprofils wird häufig eingesetzt (z. B. SAFRAN et al. 1990, CONNELLY et al. 1993 und SANDS et al. 1994).

In der Regel beschränkt sich die Evaluation eines KIS nicht auf eine der erwähnten Methoden, sondern umfaßt verschiedene einander ergänzende Ansätze (wie z. B. bei CONNELLY et al. 1993). Auch im Gießener Universitätsklinikum wurden verschiedene Maßnahmen in die Wege geleitet, um die Akzeptanz des Systems bei den avisierten klinischen Nutzern, die Nutzung der verschiedenen angebotenen Informationsressourcen und den Einfluß der KIS-Einführung auf das ärztliche und pflegerische Arbeitsumfeld zu evaluieren. So wird z. B. die Nutzung der seit 1992 vermehrt angebotenen Online-Informationenbanken in einem Logbuch protokolliert und regelmäßig ausgewertet (PROKOSCH et al. 1994). Zur Vorbereitung einer Beobachtungsstudie im pflegerischen Bereich wurde das Aufgabenspektrum der Pflege analysiert und systematisch beschrieben. Dies wurde durch eine ausführliche Formular- und Informationsflußanalyse ergänzt (BÜRKLE et al. 1994). In der vorliegenden Arbeit wurde eine Befragung der Ärzteschaft zweier klinischer Abteilungen des Gießener Universitätsklinikums ausführlich beschrieben.

Da lediglich die 68 (von 149) zurückgesandten Fragebogen in die Auswertung einbezogen werden konnten, wurde die Subpopulation noch etwas näher analysiert, um Rückschlüsse auf eventuelle Verzerrungen bereits durch die Nichtberücksichtigung der nicht antwortenden Ärzte ziehen zu können. Der Verdacht, daß lediglich Ärzte mit einem bestimmten Ausbildungsstand (z. B. nur die Assistenzärzte) geantwortet hätten, konnte verworfen werden. Erfreulicherweise hatten sogar 41% der Chef- und Oberärzte geantwortet. Bei einer Rückfrage bezüglich der relativ niedrigen Antwortquote unter den Studenten im Praktischen Jahr (29%) erklärte sich dieses Phänomen dadurch, daß diese Studenten im 6 Wochen Rhythmus innerhalb des Klinikums rotieren und deshalb zum Zeitpunkt der Befragung die entsprechenden WING/GISNET-Funktionen größtenteils fast gar nicht benutzt hatten. Auffällig ist aber, daß 80% der Antwortenden angaben selbst einen PC zu besitzen. Nimmt man den Extremfall an, daß von den Ärzten die nicht geantwortet haben keiner einen PC besitzt, so ergäbe sich ein Wert von 36% PC-Besitzern unter allen befragten Ärzten. Aufgrund persönlicher Kontakte mit den ärztlichen Mitarbeitern einiger Stationen und den Ergebnissen einer in einem anderen Zusammenhang durchgeführten Umfrage unter Medizinstudenten kann man davon ausgehen, daß ein Wert von 80% nicht repräsentativ ist, sondern eher ein Wert zwischen 40% und 50% zu erwarten wäre. Aufgrund dieser Tatsache ist zu vermuten, daß der Fragebogen vor allem von denjenigen Ärzten beantwortet wurde, die selbst einen PC besitzen und deshalb einer solchen Umfrage aufgeschlossener gegenüberstehen. Allerdings muß dies nicht unbedingt bedeuten, daß dies diejenigen Ärzte sind, die entsprechende DV-Anwendungen besonders gut bewerten. Vielmehr haben persönliche Kontakte gezeigt, daß Ärzte, die selbst einen PC besitzen, KIS-Anwendungen zum Teil kritischer gegenüberstehen und auch entsprechend höhere Anforderungen an deren Funktionalität stellen als nicht vorbelastete Anwender.

Insgesamt bestätigen die Ergebnisse dieser Umfrage die Erfahrungen durch persönliche Kontakte mit den ärztlichen Mitarbeitern einiger Stationen. Die in WING/GISNET bereitgestellten Anwendungen wurden vollständig angenommen und von den Ärzten in ihren täglichen Arbeitsablauf integriert. Ähnlich wie SLACK (1989) können auch wir bestätigen, daß Ärzte Computer im klinischen Umfeld akzeptieren und auch benutzen, vorausgesetzt die bereitgestellten Funktionen unterstützen tatsächlich ihre Tätigkeiten im Rahmen der Krankenversorgung. So sehen sich 88% der Ärzte z. B. den kommunikativen Laborbericht mehrmals täglich oder mindestens einmal pro Tag am Bildschirm an. Literaturrechercheprogramme (Medline) verwenden 11% der ärztlichen Mitarbeiter mehrmals täglich und mehr als die Hälfte aller Ärzte verwenden sie zumindest mehrmals in der Woche. An einer Erweiterung des Informationsangebots um ein wöchentlich aktualisiertes Literaturinformationssystem (Reference Update/Manager) waren die Ärzte sehr interessiert (Nutzensbewertung = 3,92; vgl. Abbildung 5). Arzneimittelinformationen aus der Roten Liste werden von 44% der Ärzte mehrmals oder mindestens einmal pro Tag und von weiteren 32% zumindest mehrmals pro Woche abgerufen.

Die Frage nach dem Nutzen der WING/GISNET-Funktionen ergab für die 4 Möglichkeiten der Labordatenpräsentation sowie für Medline und die Rote Liste eine relativ hohe durchschnittliche Bewertung (zwischen 4,62 und 4,09 auf einer Skala von 0 bis 5). Selbst die Funktion zum Ausdruck von Patientenetiketten, die in der Regel vom Pflegepersonal benutzt wird (53% der Ärzte gaben an, diese Funktion nie zu benutzen), erhielt eine Bewertung von 3,11. Die etwas niedrigere Bewertung der Micromedix-Arzneimittelinformationsbanken (2,94) ist vermutlich auf ihre sehr starke Spezialisierung zurückzuführen. Betrachtet man nur die Bewertung dieses Arzneimittelinformationssystems durch Mitarbeiter auf den Intensivstationen, so erhält man eine Bewertung von 4,44. Die Bundesbahnfahrplanauskunft, die vermutlich eher in den Sekretariaten genutzt wird, wurde mit 2,32 bewertet.

Die Tatsache, daß die WING-Funktion zur Dokumentation der Entlassungsdiagnose insgesamt nur von der Hälfte der Befragten benutzt wird, ist darauf zurückzuführen, daß es sich hierbei um eine Routinefunktion handelt, die für jeden Patienten lediglich einmal (und zwar nach seiner Entlassung) durchzuführen ist. Diese mehr administrative Tätigkeit wird innerhalb einer Station in der Regel auf einen Arzt (oft einen Arzt im Praktischen Jahr) delegiert. Aus diesem Grund muß hier nicht nur die durchschnittliche Bewertung aller Befragten (2,47), sondern insbesondere die Bewertung derjenigen Mitarbeiter, die diese Funktion auch tatsächlich benutzen (und meist auch vorher für das Ausfüllen des Entlassungsscheins verantwortlich waren) herangezogen werden. Diese lag mit 3,39 deutlich höher. Auch zu der Aussage, daß die Diagnosedokumentation über den Computer besser ist als das Ausfüllen der Entlassungsscheine, ergab sich unter den tatsächlichen Nutzern dieser WING-Funktion mit 4,0 eine deutliche Zustimmung. Berücksichtigt man, daß mit dieser Funktion erstmals auch die direkte Eingabe von klinischen Daten durch ärztliches Personal erreicht wurde, und daß sich die ärztliche Nutzung von KIS-Funktionen vielerorts noch immer auf das Abrufen von Daten (vgl. z. B. SPILLANGE et al. 1990, ZIBRAK et al. 1990) beschränkt, sind die entsprechenden Antworten sehr hoch einzustufen. Die Zustimmung der Ärzte zu der Aussage, daß die Diagnosedokumentation über den Computer besser ist als das früher übliche Ausfüllen eines Entlassungsscheins, bestätigt uns darin, daß die Bemühungen, dieses WING-Modul optimal auf die ärztlichen Bedürfnisse anzupassen (MICHEL et al. 1992), erfolgreich waren. Dies wurde zusätzlich noch durch erste Analysen der Diagnosedatenbank untermauert (DUGRENIER et al. 1993).

Um in Zukunft noch detailliertere Aussagen über die Nutzung der einzelnen WING-Funktionen durch die verschiedenen Typen klinischer Mitarbeiter machen zu können wurde mittlerweile, analog zu der Logbuch-Funktion bei den bereitgestellten Informationsdiensten (vgl. (PROKOSCH et al. 94)), auch in fast jede WING-Funktion eine

automatische Aufrufprotokollierung integriert. Bei der Vergabe von Benutzernummern ist hier jedoch noch eine spezifischere Untergliederung in Benutzerklassen notwendig als sie heute zur Verfügung steht. Dies würde für spätere Analysen noch objektivere Aussagen liefern als dies lediglich aufgrund einer Befragung möglich ist.

Die Fragen zur generellen Einstellung der Ärzte gegenüber dem Computereinsatz im Krankenhaus zeigten, daß sich hier in den letzten Jahren ein Wandel vollzogen hat, der mit einer immer größeren Aufgeschlossenheit der Mediziner gegenüber Computern einhergeht. Die Aussagen, daß Computer in Stations- und Arztzimmern fehl am Platz seien und daß sie das Verhältnis zwischen Arzt und Patienten stören, wurden deutlich abgelehnt. Vielmehr waren die Ärzte der Meinung, daß die Qualität der Krankenversorgung durch den Einsatz von Computern verbessert werden kann, daß ihnen die Benutzung des Rechners keine Mehrarbeit gebracht hat und daß sie z. B. die Ergebnisse der angeforderten Laboruntersuchungen schneller als früher erhalten. Bemerkenswert ist auch, daß 80% aller ärztlichen Mitarbeiter angaben, privat einen PC zu besitzen und 24% der Befragten ab und zu selbst programmieren. 20% hatten bereits an einem EDV-Kurs im Institut für Medizinische Informatik und 18% außerhalb des Klinikums teilgenommen. 91% der Ärzte würden es begrüßen, wenn noch weitere Tätigkeiten im Umfeld der Krankenversorgung durch EDV unterstützt würden. Lediglich einer der Befragten gab an, mit einem Computer überhaupt nichts anfangen zu können. Da zu vermuten ist, daß vor allem die der EDV positiv gegenüberstehenden Ärzte den Fragebogen ausgefüllt und zurückgeschickt haben, dürfen diese Zahlen allerdings nicht überbewertet werden.

Ein großes Problem für die ärztlichen Mitarbeiter ist immer noch die traditionelle, meist mit einer deutlichen Zeitverzögerung und mit hohem Nachfrageaufwand verbundene, Befundrückmeldung von den verschiedenen Funktionsbereichen des Krankenhauses. Dies zeigt sich deutlich an der hohen Nutzungsfrequenz und der sehr positiven Nutzensbewertung der entsprechenden WING-Funktionen zur Präsentation der Befunde des Klinisch-Chemischen Labors. Aber auch für alle anderen Funktionsbereiche würden die ärztlichen Mitarbeiter in der DV-basierten Befundübermittlung und -präsentation einen hohen Nutzen sehen (vgl. Abbildung 4). Durch die verstärkte Integration von funktionsorientierten Subsystemen in GISNET und deren funktionelle Integration soll in Zukunft diesen Bedürfnissen Rechnung getragen werden.

Auch wenn die Bewertungen von heute noch nicht verfügbaren KIS-Funktionen noch etwas vorsichtig zu interpretieren sind (da sich die Ärzte die Funktionalität einer solchen Funktion in der Regel nur schwer vorstellen können), bestätigen die Antworten bezüglich des zu erwartenden Nutzens eventueller WING/GISNET-Erweiterungen zumindest die positive Einstellung der Ärzteschaft gegenüber dem Ausbau der DV-Funktionalitäten im klinischen Umfeld (vgl. Abbildung 5). Außer für die Anamnese sahen die Ärzte in der Einführung aller vorgeschlagenen DV-Funktionen einen deutlichen Nutzen für ihre tägliche Arbeit. Hervorzuheben ist hier die sehr hohe Bewertung einer DV-Unterstützung bei der (Kurz-)Arztbriefschreibung. Auf dieses Bedürfnis kann in 1994 durch die flächendeckende Bereitstellung einer bisher als Pilotinstallation erprobten WING-Funktion zur Unterstützung der Kurzarztbriefschreibung eingegangen werden. Speziell bei der Bereitstellung von Informationsdiensten ist die Erweiterung des Informationsangebots mit relativ geringem Aufwand und zu vergleichsweise niedrigen Kosten realisierbar. Auf einige der in den Antworten geäußerten Wünsche (Medizin Lexikon, Diagnosis) konnte dementsprechend schon kurzfristig reagiert werden.

Neben diesen sehr positiven Ergebnissen traten allerdings auch einige Schwachstellen zu Tage. So wurde es deutlich, daß die langfristige Betreuung und Schulung der Krankenhausmitarbeiter bei einem ständig an Funktionalität zunehmenden Krankenhaus-Informationssystem zu einem immer größeren Problem wird. Der Aussage, daß die Schulung zur Benutzung der WING-Module ausreichend ist, konnte sich nur eine

Minderheit der Befragten anschließen (Bewertung = 2,03). Lediglich 5% der Ärzte hatten das WING-Benutzerhandbuch vollständig gelesen, fast die Hälfte der Befragten (47%) hatten es noch nie gesehen. 44% der Ärzte haben das Handbuch lediglich kurz durchgeblättert und nur 26% benutzen es, um bei Problemen etwas nachzuschlagen. Dementsprechend erhielt das WING-Benutzerhandbuch auch in Frage 3 eine eher schlechte Bewertung (vgl. Abbildung 3).

Als Erklärung für das Schulungsproblem ist vor allem die starke Fluktuation und die Zeitknappheit der ärztlichen Mitarbeiter zu nennen. Mit der Neueinführung eines GISNET-Arbeitsplatzes auf einer Station ist jedesmal eine ausführliche Schulung (ca. 60 bis 90 Minuten) der betroffenen Mitarbeiter verbunden. Zu einer solchen Schulung sind in der Regel alle pflegerischen Mitarbeiter anwesend. Vom ärztlichen Personal ist trotz langfristiger Vorankündigungen häufig nur der PJler oder ein Arzt im Praktikum zur Schulung anwesend und auch dieser wird oft noch während der Schulung wieder zu Patienten gerufen. Angebote auch in den „ruhigeren Abendstunden“ Extraschulungen für ärztliche Mitarbeiter durchzuführen werden leider nur sporadisch aufgegriffen. Durch die Rotation innerhalb der Abteilungen und durch Neueinstellungen kommen auch nach der erstmaligen Schulung neue Mitarbeiter auf die Station ohne daß die Mitarbeiter des Instituts für Medizinische Informatik davon Kenntnis erlangen, so daß diese Mitarbeiter in der Regel keine offizielle Schulung erhalten. Dies zeigt sich z. B. auch darin, daß lediglich 14% der Befragten von Institutsmitarbeitern geschult wurden. Ein Großteil der Ärzte hatte sich autodidaktisch in die Programmbedienung eingearbeitet (73%) oder ist durch Kollegen/innen eingearbeitet worden (70%). Vorteilhaft hierfür hat sich die leichte Bedienbarkeit der WING-Funktionen und die übersichtliche Gestaltung der Bildschirmmasken (Bewertungen von 3,64 und 3,44; vgl. Abbildung 3) ausgewirkt.

Ob eine bessere Unterstützung des Arbeitens mit den WING-Funktionen durch ein verbessertes Hilfesystem zu erreichen wäre, konnte durch die Antworten nicht geklärt werden. Zwar bezeichneten es 65% derjenigen die eine Beurteilung abgaben (lediglich 44%) als verbesserungswürdig (15% bezeichneten es als gut und 20% fanden es ganz überflüssig), doch zeigte es sich auch, daß dieses Hilfesystem lediglich 61% der Benutzer bekannt ist, obwohl in der Hinweisleiste jeder WING-Maske darauf hingewiesen wird.

Auch wenn das Problem des internen Mitarbeiterwechsels mit der wachsenden Verbreitung von GISNET/WING im gesamten Gießener Klinikum (spätestens Ende 1994 soll jede klinische Abteilung an GISNET angeschlossen sein) reduziert wird, bleibt es doch offen, wie z. B. in Zukunft bei der Neueinführung einer WING-Funktion verfahren werden soll. Der Zeitaufwand der mit der Schulung der Mitarbeiter einer jeden einzelnen Station verbunden ist, ist immens und erfordert einen Personalbedarf, den in Deutschland heute wohl noch kein Institut für Medizinische Informatik aufweisen kann.

Im Gießener Universitätsklinikum macht sich dies erstmals im Rahmen der flächendeckenden Verbreitung einer Funktion für die Kurzarztbriefschreibung bemerkbar. Diese Funktion wurde auf einer Station der Medizinischen Klinik pilothaft erprobt und soll nun nach einigen Verbesserungen allen WING-Anwendern zur Verfügung gestellt werden. Der dringliche Bedarf für eine solche Funktion wird auch durch die hohe positive Bewertung dieser Funktion in Frage 7 (vgl. Abbildung 5) bestätigt. Innerhalb der Medizinischen Klinik hat sich die Nutzung dieser Funktion auch ohne spezielle Schulung (obwohl diese immer noch wünschenswert wäre), lediglich aufgrund der Mund-zu-Mund-Propaganda innerhalb der Ärzteschaft verbreitet. Die notwendige Ausweitung auf andere klinische Abteilungen konnte jedoch aus Personalmangel noch nicht durchgeführt werden.

Auch nach einer inhaltlichen Überarbeitung des WING-Benutzerhandbuchs ist nicht unbedingt zu erwarten, daß dies das Problem der Schulungen deutlich vermindern wird. Obwohl dieses Handbuch bei jeder GISNET/WING-Neuinstallation an alle bei der Schulung anwesenden Mitarbeiter verteilt wurde und eine Anzahl zusätzlicher Exemplare

auf der Station verblieben, gaben 47% der Befragten an, dieses Handbuch nie gesehen zu haben und lediglich 5% hatten es ganz gelesen. Weder die Bereitstellung von schriftlichen Schulungsunterlagen noch von Hilfsmasken in der KIS-Anwendung selbst kann den unmittelbaren menschlichen Kontakt und das erklärende Gespräch ersetzen. Dies erfordert jedoch Personalressourcen die zur Zeit in den wenigsten Kliniken vorhanden sind. Je mehr ein Krankenhaus-Informationssystem in das klinische Umfeld vordringt und je umfangreicher seine Funktionalität wird, um so mehr müssen in Zukunft Personalkapazitäten nicht nur im Bereich der DV-Entwicklung und Systempflege, sondern auch für kontinuierliche Anwenderbetreuung geschaffen werden.

Danksagung

Die Autoren danken allen Mitarbeitern des Instituts für Medizinische Informatik der Justus-Liebig-Universität Gießen, ohne deren unermüdliche Arbeit die Bereitstellung der beschriebenen und evaluierten KIS-Funktionalitäten nicht möglich gewesen wäre.

Literatur

- BÜRKLE, T., SCHMITZ, M., PROKOSCH, H. U. und DUDECK, J. (1994). A Systematic Approach for Evaluation of Nursing Work in an University Hospital. MIE 94, Lisboa, Portugal, Mai 1994, in Druck.
- CONNELLY, D. P., WERTH, G. R., DEAN, D. W., HULTMAN, B. K. und THOMPSON, T. R. (1993). Physician Use of an NICU Laboratory Reporting System. In: FRISSE, M. E. (Hrsg.) *Proceedings Sixteenth Annual Symposium on Computer Applications in Medical Care*, McGraw-Hill, New York, 8–12.
- DUGRENIER, C., MICHEL, A., MARQUARDT, K. und DUDECK, J. (1993). Erfahrungen mit der online Erfassung von Entlassungsdiagnosen am Universitätsklinikum der Justus-Liebig-Universität Gießen. *Tagungsband der 38. Jahrestagung der GMDS*, Lübeck, September 1993, in Druck.
- EVANS, R. S., PESTOTNIK, S. L., CLASSEN, D. C. und BURKE, J. P. (1993). Development of an Automated Antibiotic Consultant. *M.D. Computing* 10: 17–22.
- GILHOOLY, K. J., LOGIE, R., ROSS, D., RAMAYYA, P. und GREEN, C. (1991). Users' Perceptions of a Computerised Information System in Intensive Care (ABICUS) on Introduction and after 2 Months Use. *International Journal of Clinical Monitoring and Computing* 8: 101–106.
- GOUVEIA-OLIVEIRA, A., RAPOSO, V. D., SALGADO, N. C., ALMEIDA, I., NOBRE-LEITAO, C. und GALVAO DE MELO, F. (1991). Longitudinal Comparative Study on the Influence of Computers on Reporting of Clinical Data. *Endoscopy* 23: 334–337.
- LEPAGE, E. F., GARDNER, R. M., LAUB, R. M. und GOLUBJATNIKOV, O. K. (1992). Improving Blood Transfusion practice: Role of a Computerized Hospital Information System. *Transfusion* 32: 253–259.
- MICHEL, A., PROKOSCH, H. U. und DUDECK, J. (1992). Some Experience With the Implementation and Operation of a Hospital Wide Online Diagnosis Documentation and Coding System. In: LUN, K. C., DEGOULET, P., PIEMME, T. E. und RIENHOFF, O. (Hrsg.). *MEDINFO 92 – Proceedings of the Seventh World Congress on Medical Informatics*, North-Holland, Amsterdam London New York Tokyo, 1311–1315.
- PROKOSCH, H. U., DUDECK, J., JUNGHANS, G., MARQUARDT, K., SEBALD, P. und MICHEL, A. (1991). WING-Entering a New Phase of Electronic Data Processing at the Gießen University Hospital. *Methods of Information in Medicine* 30: 289–298.
- PROKOSCH, H. U., PUHLE, B. und DUDECK, J. (1994). One-Stop-Information-Shopping: Do We Meet the Information Needs of the Hospital Staff? MIE 94, Lisboa, Portugal, Mai 1994, in Druck.
- SAFRAN, C., HERRMANN, F., RIND, D., KOWALOFF, H. B., BLEICH, H. L. und SLACK, W. V. (1990). Computer-Based Support for Clinical Decision Making. *M.D. Computing* 7: 319–322.

- SANDS, D. Z., SAFRAN, C., SLACK, W. V., BLEICH, H. L. (1994). Use of Electronic Mail in a Teaching Hospital. In: SAFRAN, C. (Hrsg.). *Proceedings Seventeenth Annual Symposium on Computer Applications in Medical Care*, 306–310.
- SITTIG, D. F. (1993). Work-Sampling: A Statistical Approach to Evaluation of the Effect of Computers on Work Patterns in Healthcare. *Methods of Information in Medicine* 32: 167–174.
- SLACK, W. V. (1989). Remembrance, Thanks and Welcome: Assessing the Use of Computers. *M.D. Computing* 6: 183–185.
- SPILLANE, M. J., McLAUGHLIN, M. B., ELLIS, K. K., MONTGOMERY, W. L. und DZIUBAN, S. (1990). Direct Physician Order Entry and Integration: Potential Pitfalls. In: MILLER, R. A. (Hrsg.). *Proceedings Fourteenth Annual Symposium on Computer Applications in Medical Care*, IEEE Computer Society Press, 774–778.
- TIERNEY, W. M., MILLER, M. E., OVERHAGE, J. M. und McDONALD, C. J. (1993). Physician Inpatient Order Writing on Microcomputer Workstations: *JAMA* 269: 379–383.
- WIECZOREK, D., PROKOSCH, H. U., NEIDHART, B., KRECKEL, H. und DUDECK, J. (1993). One Year of Experience with EDP Support for Drug Therapy at Gießen University Hospital. In: REICHERT, A., SADAN, B. A., BENGTTSSON, S., BRYANT, J. und PICCOLO, U. (Hrsg.). *Proceedings of MIE 93*, Freund Publishing House, Ltd., 322–325.
- ZIBRAK, J. D., ROBERTS, M. S., NELICK-COHEN, L. und PETERSON, M. (1990). Creating an Environment Conducive to Physician Participation in a Hospital Information System. In: MILLER, R. A. (Hrsg.). *Proceedings Fourteenth Annual Symposium on Computer Applications in Medical Care*, IEEE Computer Society Press, 779–783.

Anschrift des ersten Verfassers: Dr. H. U. Prokosch, Institut für Medizinische Informatik, Heinrich-Buff-Ring 44, 35392 Gießen

Eingang des Manuskripts 10. 1. 94
Eingang der revidierten Fassung 23. 3. 94

Die logische Struktur der ICD-10 (Systematik) und ihre Beschreibung mit SGML

M. Schopen

Kurztitel:

ICD-10-Struktur und SGML

Short title:

ICD-10-Structure and SGML

Schlüsselwörter:

Internationale Klassifikation der Krankheiten – Standard Generalized Markup Language

Keywords:

International Classification of Diseases – Standard Generalized Markup Language

Zusammenfassung

Das Deutsche Institut für medizinische Dokumentation und Information, DIMDI, ist beauftragt, die amtliche deutschsprachige Ausgabe der ICD-10 herauszugeben. Ausgehend von den Anforderungen an die maschinenlesbare Version der ICD-10 (Systematik) wird die Standard Generalized Markup Language (SGML) vorgestellt als ein Formalismus, mit dem die logische Struktur der Klassifikation beschrieben werden kann. Der im DIMDI für die ICD-10 verfolgte Ansatz – Reduktion auf die logische Struktur und Verzicht auf Layoutinformation – macht die maschinenlesbare Fassung unabhängig von spezifischer Hard- und Software und hält sie offen für unterschiedlichste Anwendungen.

Die vorliegende Arbeit beschreibt in SGML die Grobstruktur der ICD-10-Kapitel sowie den Aufbau der Krankheitsklassen und einiger ihrer Elemente. An einem Beispiel wird gezeigt, wie die SGML-Daten für spezielle Anwendungen der Klassifikation restrukturiert werden können.

Abstract

The German Institute for Medical Documentation and Information, DIMDI, has been authorized to publish the official German language edition of ICD-10. Based on the requirements for a machine-readable version of ICD-10 (Tabular List), SGML – the

Standard Generalized Markup Language — is introduced as a formalism to describe the logical structure of this classification. By specifying the mere logical structure and abandoning layout information, DIMDI's concept for ICD-10 makes the machine-readable version independent of any hardware or software and keeps it open to a broad range of applications.

This paper uses SGML to describe the structure of ICD-10 — the chapters, the disease categories and some of their elements. An example is given how to rearrange SGML-data for specific applications of the classification.

1. Einleitung und Fragestellung

Die Internationale Klassifikation der Krankheiten (International Classification of Diseases, ICD) wird von der Weltgesundheitsorganisation in 10-jährigen Zyklen revidiert und gepflegt (World Health Organization, 1992). DIMDI, das Deutsche Institut für medizinische Dokumentation und Information, ist vom Bundesministerium für Gesundheit beauftragt, die amtliche deutschsprachige Ausgabe herauszugeben und zu pflegen. Die 9. Revision der ICD (ICD-9) wird in der Bundesrepublik Deutschland seit 1979 zur Verschlüsselung der Todesursachen eingesetzt; seit 1986 ist sie durch die Bundespflegeverordnung zur Grundlage der Diagnosenstatistik in Krankenhäusern geworden; die Änderungen im Sozialgesetzbuch V durch das Gesundheitsstrukturgesetz 1992 verlangen darüberhinaus erstmals die Diagnoseverschlüsselung in der vertragsärztlichen Praxis. Zur Zeit wird im DIMDI die 10. Revision der ICD (ICD-10) ins Deutsche übertragen. Sie soll in der Bundesrepublik nicht vor dem Jahre 1997 eingesetzt werden.

Die Erfahrungen des DIMDI mit der Abgabe der ICD-9 auf Datenträgern zeigen, daß bei der 10. Revision dieser Klassifikation ein anderer Weg beschritten werden muß. Als die 9. Revision übersetzt wurde, diente die maschinenlesbare Fassung primär als Grundlage für den Buchdruck und zur Klassifikationspflege; die Abgabe auf Datenträgern an Anwender war zu dieser Zeit die Ausnahme. Wer die ICD-9 EDV-basiert benutzen wollte, mußte daher zunächst mit Hilfe der Steuerzeichen in der maschinenlesbaren Fassung sinnvolle Sätze mit Verschlüsselungsangaben erstellen — ein Weg, der sich in der Praxis als mühsam und durchaus fehleranfällig erwiesen hat. Inzwischen haben sich die Anforderungen an eine maschinenlesbare Version der ICD erheblich gewandelt — sei es bedingt durch den Fortschritt der elektronischen Datenverarbeitung, sei es aber auch bedingt durch den breiten Einsatz der ICD zur Morbiditätsverschlüsselung: Immer mehr Kliniken benutzen für ihre Leistungsstatistiken die maschinenlesbare Fassung der ICD; mit der Auflage durch das Gesundheitsstrukturgesetz, in Zukunft auch in der vertragsärztlichen Praxis Diagnosen zu verschlüsseln (Bundesministerium für Gesundheit, 1993) wird die ICD verstärkt in Arztpraxis-Informationssysteme eingebunden werden.

An eine maschinenlesbare Version der ICD-10 sind — im Hinblick auf ein breites Anwendungsspektrum — daher folgende Anforderungen zu stellen:

- Die Klassifikation muß in einer Zentraldatei gepflegt werden, denn nur so ist die Datenkonsistenz zu gewährleisten. Da die WHO beabsichtigt, die ICD in Zukunft jährlich zu aktualisieren, ist gegenüber der ICD-9 mit einem wesentlich höheren Pflegeaufwand zu rechnen.
- Sie muß als Grundlage für den Buchdruck dienen und gleichzeitig die Erstellung elektronischer Folgeprodukte ermöglichen (z. B. Online-Datenbank, CD-ROM, EDV-gestützte Verschlüsselungshilfen). Sie soll daher einerseits das Layout für den Buchdruck unterstützen, andererseits darf sie jedoch nicht durch die reine Beschreibung des Druckbildes die Übernahme logischer Klassifikationselemente in die Felder einer Datenbank behindern.

- Sie muß leicht restrukturierbar sein, denn die Strukturanforderungen an die Datenbasis eines EDV-gestützten Verschlüsselungssystems lassen sich über die 10- bis 15-jährige Einsatzzeit der Klassifikation kaum voraussagen.
- Sie muß sich eignen als eine der Quellen für das Alphabetische Verzeichnis zur deutschsprachigen ICD-10.
- Sie muß sich aus Gründen der Arbeitsökonomie und der Datenkonsistenz weitgehend automatisch aus jenen Textdateien erzeugen lassen, in denen die deutschsprachige Fassung zur Zeit vorliegt. Weiterhin sollte sie mit Standardsoftware gepflegt und weiterverarbeitet werden können.
- Es versteht sich von selbst, daß sie leicht portabel sein muß, also nicht an bestimmte Hardware oder Software gebunden sein darf.

Nachdem die Übertragung der ICD-10-Systematik ins Deutsche nun fast abgeschlossen ist, stellt sich die Frage nach einem geeigneten Format für die maschinenlesbare Version.

2. SGML – die Standard Generalized Markup Language

Die Standard Generalized Markup Language (SGML), ein ISO-Standard (International Standards Organization, 1986), unterstützt diese Anforderungen weitestgehend. SGML beschreibt die logische Struktur eines Dokumentes und nicht dessen physische Merkmale (Druckformate, Layout), da Satzsteuerzeichen immer systemspezifisch und selten portabel sind. SGML unterscheidet sich daher grundlegend von eher Layout-orientierten Dokument-Beschreibungssprachen wie z. B. TeX. SGML-konforme Dokumente sind unabhängig von jeder Hard- und Software und können zwischen unterschiedlichen EDV-Umgebungen problemlos ausgetauscht werden. SGML ist vor allem im angloamerikanischen Raum weit verbreitet, setzt sich inzwischen als Standard aber auch in der Bundesrepublik mehr und mehr durch.

SGML ermöglicht eine Trennung zwischen der Beschreibung der Dokumentstruktur (Grammatik, Dokumenttyp: Vorgabe des logischen Aufbaus aus Einzelementen für eine Klasse von Dokumenten) und der Auszeichnung des eigentlichen Dokumentes (Markup: Kennzeichnung der einzelnen Elemente im Text eines Dokumentes). Das folgende Beispiel einer Referenz aus dem Literaturverzeichnis dieser Arbeit soll das Prinzip der Auszeichnung eines Textes und der Beschreibung seiner Struktur zeigen¹⁾. Die einzelnen Elemente des Dokumentes werden von Start-Tags (z. B. **<Autor>**) und End-Tags (z. B. **</Autor>**) umschlossen:

```

<!DOCTYPE Ref SYSTEM "REF.DTD">
<Ref>
<Autor> Bryan, M. </Autor>
<Jahr> 1988 </Jahr>
<Titel> SGML. An Author's Guide to the Standard Generalized Markup
Language. </Titel>
<Verlag> Addison-Wesley Publishing Company </Verlag>
<Ort> Reading, Massachusetts </Ort>
</Ref>

```

Der logische Aufbau des Dokumentes aus Einzelementen wird in der Document Type Definition (DTD) beschrieben. Sie wird in der DOCTYPE-Declaration zu Beginn des Dokumentes mit diesem verknüpft und meist in einer separaten Datei abgelegt (im

¹⁾ Um die folgenden Beispiele lesbar zu halten, sind SGML-Auszeichnungen stets fett gedruckt; wo im Text auf Elemente eines SGML-Dokumentes Bezug genommen wird, sind deren Namen fett und kursiv gedruckt.

Beispiel die Datei "REF.DTD"). Die DTD des einfachen Beispiels könnte den Dokumenttyp *Ref* wie folgt beschreiben:

```
<!ELEMENT Ref -- (Autor+, Jahr, Titel, Verlag, Ort+) >
<!ELEMENT Autor -- (#PCDATA) >
<!ELEMENT Jahr -- (#PCDATA) >
<!ELEMENT Titel -- (#PCDATA) >
<!ELEMENT Verlag -- (#PCDATA) >
<!ELEMENT Ort -- (#PCDATA) >
```

Eine Referenz *Ref* besteht aus mindestens einem Autor *Autor*, gefolgt vom Publikationsjahr *Jahr*, dem Titel *Titel*, dem Verlag *Verlag* und mindestens einer Ortsangabe *Ort*. Die Grundelemente, z. B. *Ort* oder *Jahr*, bestehen aus Parsed Character Data (#PCDATA), im Beispiel aus einfachen Zeichen.

Die Syntax von SGML kann hier nicht vollständig beschrieben werden, dazu sei auf entsprechende Literatur verwiesen (z. B. BRYAN 1988, oder VAN HERWIJNEN 1990); über weitere Informationsquellen zu SGML orientiert ein Anhang. Zum weiteren Verständnis erforderlich sind lediglich einige Konventionen aus der DTD.

Ein unärer Operator in Postfix-Schreibweise bezeichnet die Häufigkeit der einzelnen Elemente: „*“ hinter einem Elementnamen bedeutet, daß dieses Element beliebig oft erscheint, aber auch fehlen kann; „+“ besagt, daß dieses Element beliebig oft, aber mindestens einmal auftritt; „?“ bezeichnet ein höchstens einmal vorhandenes Element. Fehlt eine Häufigkeitsangabe, so kommt das betreffende Element genau einmal vor.

Die Reihenfolge der einzelnen Elemente wird durch einen binären Operator in Infix-Stellung beschrieben: „&“ zwischen Elementnamen bedeutet, daß alle so verbundenen Elemente vorhanden sind und daß jede Permutation erlaubt ist (reines logisches AND); „,“ gibt an, daß alle so verbundenen Elemente auftreten und daß zusätzlich ihre Reihenfolge festgelegt ist (restriktiver als logisches AND); „|“ besagt, daß genau eines der so verbundenen Elemente erscheint (logisches XOR). Im Prinzip beschreibt SGML also den Aufbau des Dokumentes aus Einzelementen mit regulären Ausdrücken. Kommentare in der DTD werden durch „--“ ein- und ausgeleitet.

Ein Problem beim Austausch von Dokumenten zwischen unterschiedlichen EDV-Umgebungen bereiten Umlaute und Sonderzeichen, da ihre Position in den systemspezifischen Zeichensatztabellen wechselt. SGML kann Sonderzeichen mit sogenannten Entity References darstellen: „ämie“ wäre zu lesen als „Anämie“, wobei „ä“ mit „ä“ kodiert ist. Dies kommt den zahlreichen Elementen nationalsprachlicher Zeichensätze in der ICD-10 sehr entgegen.²⁾

Das einfache Beispiel einer Referenz *Ref* zeigt, wie ein Dokument unter Verzicht auf Formatierungsangaben ausgezeichnet wird. Beim Druck werden den einzelnen logischen Elementen dann die entsprechenden Druckformate zugeordnet. So könnte im Beispiel das Element *Autor* fett oder kursiv gesetzt werden, das *Jahr* in Klammern ausgegeben und der *Titel* in einer neuen Zeile gedruckt werden. Viele Druckereien können heute bereits SGML-Daten mit Standardsoftware bearbeiten.

Die logischen Elemente eines SGML-Dokumentes können leicht in die Felder einer Datenbank übernommen werden. Manche Datenbanksysteme verfügen heute schon über ein SGML-Import-Filter. Besonders komfortable Systeme können nicht nur die SGML-Auszeichnungen erkennen, sondern legen auf der Grundlage der DTD bereits eine Datenbankstruktur fest. Fehlt eine solche Schnittstelle, so lassen sich SGML-Auszeichnungen mit einem Umsetzprogramm in Steuerzeichen für vorhandene Software transformieren.

²⁾ Im folgenden wird auf diese Kodierung jedoch stets verzichtet, damit die SGML-Beispiele leichter zu lesen sind.

SGML-Dokumente können mit Hilfe eines SGML-basierten Editors gut gepflegt werden. Bei der Bildschirmdarstellung können die SGML-Auszeichnungen unterdrückt und mit speziellen Ausgabeformaten verbunden werden, so daß ein übersichtlich strukturiertes Bild des Gesamtdokumentes entsteht. Ein integrierter Parser prüft die SGML-Dateien auf Normkonformität und die Folge der einzelnen Elemente auf ihre logische Konsistenz mit den Vorgaben der DTD. Änderungen oder Ergänzungen des Dokumentes werden nur akzeptiert, wenn sie mit der logischen Dokumentstruktur übereinstimmen.

SGML eignet sich vorzüglich zur Beschreibung hierarchischer Strukturen mit sich vererbenden Eigenschaften, wie sie in Klassifikationen auftreten. Die DTD erinnert an frame-artige Formate der Wissensrepräsentation (REIMER 1991): Elemente, die aus Parsed Character Data bestehen, entsprechen den terminalen Slots; Elemente, die wiederum aus Elementen aufgebaut sind, entsprechen den nicht-terminalen Slots.

3. DREAM – ein Werkzeug zur strukturellen Anreicherung von Texten

SGML erwies sich erst als ein effizientes Format für die maschinenlesbare Fassung der ICD-10, als ein Software-Tool zur automatischen Auszeichnung der vorhandenen Textdateien verfügbar wurde. Am Institut für Integrierte Publikations- und Informationssysteme der Gesellschaft für Mathematik und Datenverarbeitung in Darmstadt wurde mit DREAM ein solches Programm entwickelt (GÖTTKE und FANKHAUSER 1992). DREAM (Document Structure Recognition And Markup) nutzt die Erfahrung, daß die logische Struktur eines Dokumentes in Textdateien in der Regel durch Layoutelemente (Leerzeilen, Einrückungen, Tags) dargestellt wird. DREAM arbeitet auf ASCII-Dateien und benutzt diese Layoutinformation, um die logische Struktur des Textes zu erkennen. Dabei werden die Layoutelemente entfernt und die logischen Elemente des Textes mit SGML-Auszeichnungen abgegrenzt. Der Weg von der physischen hin zur logischen Struktur wird vom Anwender in einer DREAM-Grammatik mit regulären Ausdrücken beschrieben; sie lehnt sich in ihrer Syntax eng an eine SGML-DTD an. In dieser Grammatik lassen sich inkonsistente Formatierungskonventionen recht gut abfangen, etwa durch die Angabe von Bereichen für die Leerzeichenzahl oder von Alternativen für Tags. Mit der Komponente DREAMGEN wird diese Grammatik zunächst in die Übergangstabelle eines nichtdeterministischen endlichen Automaten übersetzt; diese arbeitet der DREAM-Parser in einer Tiefensuche systematisch ab, um den Text auszuzeichnen. DREAM steht in C++ und läuft unter MSDOS oder UNIX.

4. Logische Struktur der ICD-10

Vor der eigentlichen Umsetzung der ICD-Textdateien stand eine eingehende Analyse der logischen Klassifikationsstruktur. Sie wird im folgenden mit SGML dargestellt. Dabei wird auf Details insoweit verzichtet, als sie zum prinzipiellen Verständnis nicht erforderlich sind.

Gegenüber der 9. Revision der ICD gibt es keine wesentlichen Änderungen der Struktur. Der grundlegende Aufbau aus logischen Elementen läßt sich daher an einer deutschsprachigen Ausgabe der ICD-9 nachvollziehen. In dieser Arbeit sind die Beispiele zur ICD-10 bewußt so gewählt, daß die Vorzüge dieser neuen Revision deutlich werden. Bezüglich eines Vergleiches der beiden Revisionen sei auf die Literatur verwiesen (BRAUN und DIEKMANN 1991; WIESNER 1993).

4.1. Grobstruktur

Die ICD-10 besteht in ihrem Kern aus 21 Kapiteln. Jedes Kapitel *Kap* gliedert sich in Gruppen von Krankheiten *Gruppe*, in denen die Drei- und Viersteller (*D* und *V*) als die

elementaren Krankheitsklassen zusammengefaßt sind. Abbildung 1 veranschaulicht diese hierarchische Struktur, die auf einer durch Hyponymie (Ober-/Unterbegriffsrelationen) induzierten Begriffsordnung beruht, mit einer absteigenden klassifikatorischen Kette aus Kapitel XIII; dem Text ist die zugehörige DTD und eine Fassung mit SGML-Auszeichnungen beigegeben.

Es wird deutlich, daß nicht unmittelbar zur logischen Struktur gehörende Teile des Textes in der SGML-Fassung fortfallen können, so z. B. die Klammern um den Verschlüsselungsrahmen bei Kapitel- und Gruppenüberschriften. Sie können den entsprechenden Elementen bei der Ausgabe wieder hinzugefügt werden: *KRahm* bzw. *GrRahm* erhalten bei der Formatierung das Präfix „(,“ und das Suffix “)“, ihr Unterelement *bis* erhält das Präfix „-“.

4.2. Klassenaufbau

Bei den Kapiteln, Gruppen, Dreistellern und Vierstellern können zur Beschreibung des Klasseninhaltes und zur Definition der Klassengrenzen bestimmte Eigenschaften angegeben werden. Es sind dies

- Inklusiva: Bezeichnungen von Krankheiten, die in dieser Klasse eingeschlossen sind (*Inkl1* mit einleitendem „Inkl.“, *Inkl2* ohne);
- Exklusiva: Bezeichnungen von Krankheiten, die in dieser Klasse ausgeschlossen sind (*Exkl*, in Klammern wird jeweils auf die zutreffende Schlüsselnummer verwiesen);
- Hinweise und Erklärungen (*Hinw* und *Erkl*);
- Hinweise zur Verwendung weiterer Schlüsselnummern (*Benutze*);
- Hinweis auf eine vorangestellte Subklassifikation (*Siehe*);
- Angabe einer Subklassifikation für die betreffende Klasse (*Sub*).

Erklärungen kommen im wesentlichen in Kapitel V „Psychische und Verhaltensstörungen“ vor und streben eine intensionale Definition des Klasseninhaltes an; Inklusiva und Exklusiva hingegen finden sich in allen Kapiteln und können als extensionale Beschreibung des Klassenumfanges verstanden werden. Die Eigenschaften einer Klasse werden in entsprechenden logischen SGML-Elementen dargestellt, die zu einem Element *Inhalt* zusammengeführt werden; es ist in Abbildung 2 beschrieben. Die Struktur der einzelnen Elemente von *Inhalt* ist dabei zunächst nicht weiter ausgeführt. Neben der Elementdefinition zeigt Abbildung 2 – mit und ohne SGML-Auszeichnungen – einen Ausschnitt aus Kapitel XIII, der einige dieser Elemente enthält.

Dieses Beispiel zeigt, wie infolge der hierarchischen Struktur der Klassifikation die Klasseneigenschaften vererbt werden: so gilt der Ausschluß „Autoimmunkrankheit eines einzelnen Organs“ für alle Klassen der Gruppe „Systemkrankheiten des Bindegewebes“ und der Einschluß „Sklerodermie“ für den Dreisteller „Systemische Sklerose“ und alle ihm untergeordneten Viersteller. Diese Vererbung wird durch die Schachtelung der SGML-Elemente repräsentiert: Der Dreisteller umschließt alle zugehörigen Viersteller, die Gruppe ihrerseits alle zugehörigen Dreisteller – dies ist am Schluß des SGML-Beispiels an der Folge der End-Tags gut zu erkennen.

4.3. Feinstruktur von Listen

In Abbildung 2 deutet sich bereits an, daß die Listen der Inklusiva und Exklusiva (z. B. in den Gruppeneigenschaften) weiter zu granularisieren sind – nicht nur wegen des exakten Druckbildes, sondern auch, weil sie für EDV-gestützte Anwendungen gegebenenfalls restrukturiert werden müssen (siehe dazu 5.2.).

Um das Prinzip darzustellen, ist in Abbildung 3 eine relativ einfache Liste von Inklusiva weiter aufstrukturiert. Der Eintrag *Eintrag* einer solchen Inklusiv-Liste kann einspaltig, im komplexen Fall auch mehrspaltig mit geschweiften Klammern sein. An dieser Stelle sollen jedoch nur einspaltige Einträge behandelt werden.


```

XIII:    Krankheiten des Muskel-Skelett-Systems und des
         Bindegewebes
         (M00-M99)

         Systemkrankheiten des Bindegewebes
         (M30-M36)

M34     Systemische Sklerose

M34.0   Progressive systemische Sklerose
    
```

```

<!DOCTYPE Kap SYSTEM "KAP.DTD" >
<Kap> <KNr> XIII </KNr> <KTitel> Krankheiten des
Muskel-Skelett-Systems und des Bindegewebes </KTitel>
<KRahm> <von> M00 </von> <bis> M99 </bis> </KRahm>
<Gruppe> <GrTitel>
Systemkrankheiten des Bindegewebes </GrTitel>
<GrRahm> <von> M30 </von> <bis> M36 </bis> </GrRahm>
<D> <DCode> M34 </DCode>
<DTitel> Systemische Sklerose </DTitel>
<V> <VCode> M34.0 </VCode>
<VTitel> Progressive systemische Sklerose </VTitel>
</V> </D> </Gruppe> </Kap>
    
```

```

<!-- ICD-Kapitel: -->
<IELEMENT Kap -- (KNr,KTitel,KRahm,
Inhalt?,Gruppe +) >
<IELEMENT KNr -- (#PCDATA) >
<IELEMENT KTitel -- (#PCDATA) >
<IELEMENT KRahm -- (von,bis) >

<!-- Krankheitsgruppe: -->
<IELEMENT Gruppe -- (GrTitel,GrRahm,Inhalt?,D +) >
<IELEMENT GrTitel -- (#PCDATA) >
<IELEMENT GrRahm -- (von,bis?) >
<IELEMENT von -- (#PCDATA) >
<IELEMENT bis -- (#PCDATA) >

<!-- dreistellige Krankheitsklasse: -->
<IELEMENT D -- (DCode,DTitel,Inhalt?,V*) >
<IELEMENT DCode -- (#PCDATA) >
<IELEMENT DTitel -- (#PCDATA) >

<!-- vierstellige Krankheitsklasse: -->
<IELEMENT V -- (VCode,VTitel,Inhalt?) >
<IELEMENT VCode -- (#PCDATA) >
<IELEMENT VTitel -- (#PCDATA) >
<!-- Das ELEMENT Inhalt wird später beschrieben. -->
    
```

Abbildung 1: Absteigende Klassifikatorische Kette aus Kap. XIII: Text mit Layout, Text mit SGML-Anzeichnungen, Document Type Definition

Systemkrankheiten des Bindegewebes (M30-M36)

Inkl.: Autoimmunkrankheit:
 ■ systemisch
 ■ o.n.A.
 Kollagen- (Gefäß-)Krankheit:
 ■ systemisch
 ■ o.n.A.

Exkl.: Autoimmunkrankheit eines einzelnen Organs oder eines einzelnen Zelltyps (Verschlüsselung des betreffenden Zustandes)

M34 Systemische Sklerose

Inkl.: Sklerodermie
Exkl.: Sklerodermia circumscripta (L94.0)
 Sklerodermie beim Neugeborenen (P83.8)

M34.0 Progressive systemische Sklerose

M34.1 CR(E)ST-Syndrom

Kombination von Kalzinose, Raynaud-Phänomen, Ösophagusdysfunktion, Sklerodaktylie, Teleangiectasie.

M34.2 Systemische Sklerose, durch Arzneimittel oder chemische Substanzen induziert

Soll das Arzneimittel angegeben werden, ist zusätzlich eine Schlüsselnummer aus Kapitel XX zu benutzen.

< IELEMENT Inhalt -- (Siehe?, Hinw?, Erkl?, (Inkl1 | Inkl2)?, Benutze *, Exkl?, Sub?) >

<!-- Siehe, Hinw, Erkl, Benutze, Sub sind hier nicht weiter strukturiert. Zu inkl1, inkl2, Exkl siehe weiter unten. -->

```
<Gruppe>
<GrTitel>Systemkrankheiten des Bindegewebes</GrTitel>
<GrRahm><von>M30</von><bis>M36</bis>
</GrRahm>
<Inhalt><Inkl1>
Autoimmunkrankheit:
■ systemisch
■ o.n.A.
Kollagen- (Gefäß-)Krankheit:
■ systemisch
■ o.n.A.
</Inkl1><Exkl>
Autoimmunkrankheit eines einzelnen Organs oder eines
einzelnen Zelltyps (Verschlüsselung des betreffenden
Zustandes)
</Exkl></Inhalt>
<D><DCode>M34</DCode>
<DTitel>Systemische Sklerose</DTitel>
<Inhalt><Inkl1>
Sklerodermie
</Inkl1><Exkl>
Sklerodermia circumscripta (L94.0)
Sklerodermie beim Neugeborenen (P83.8)
</Exkl></Inhalt>
<V><VCode>M34.0</VCode>
<VTitel>Progressive systemische Sklerose</VTitel></V>
<V><VCode>M34.1</VCode>
<VTitel>CR(E)ST-Syndrom</VTitel>
<Erkl>Kombination von Kalzinose, Raynaud-Phänomen,
Ösophagusdysfunktion, Sklerodaktylie, Teleangiectasie.
</Erkl></Inhalt></V>
<V><VCode>M34.2</VCode>
<VTitel>Systemische Sklerose, durch Arzneimittel oder
chemische Substanzen induziert</VTitel>
<Benutze>Soll das Arzneimittel angegeben werden, ist
zusätzlich eine Schlüsselnummer aus Kapitel XX zu
benutzen.</Benutze></Inhalt>
</V></D></Gruppe>
```

Abbildung 2: Klassenaufbau: Text mit Layout, Text mit (groben) SGML-Auszeichnungen, Definition der Elemente
 Informatik, Biometrie und Epidemiologie in Medizin und Biologie 2/1995

```

Systemkrankheiten des Bindegewebes
(M30-M36)

    Inkl.: Autoimmunkrankheit:
           ■ systemisch
           ■ o.n.A.
    Kollagen- (Gefäß-)Krankheit:
           ■ systemisch
           ■ o.n.A.

...
    
```

```

<IELEMENT Inkl1 -- (Eintrag) >
<IELEMENT Inkl2 -- (Eintrag) >
<IELEMENT Exkl -- (Eintrag) >
<IELEMENT Eintrag -- (EinSp | MehrSp)+ >
<IELEMENT EinSp -- (Atom | (Kopf,Liste))+ >
<IELEMENT Liste -- (BEinSp | BMehrSp)+ >
<IELEMENT BEinSp -- (BAtom | (BKopf,BListe))+ >
<IELEMENT BListe -- (SBAtom)+ >
<IELEMENT Kopf -- (#PCDATA) >
<IELEMENT Atom -- (#PCDATA) >
<IELEMENT BKopf -- (#PCDATA) >
<IELEMENT BAtom -- (#PCDATA) >
<IELEMENT SBAtom -- (#PCDATA) >

<!-- MehrSp, BMehrSp, SBMehrSp sind nicht weiter strukturiert. -->
    
```

```

<Gruppe>
<GrTitel> Systemkrankheiten des Bindegewebes </GrTitel>
<GrRahm> <von> M30 </von> <bis> M36 </bis>
</GrRahm> <Inhalt> <Inkl1>
<Eintrag> <EinSp>
<Kopf> Autoimmunkrankheit: </Kopf>
<Liste> <BEinSp>
<BAtom> systemisch </BAtom>
<BAtom> o.n.A. </BAtom> </BEinSp> </Liste>
<Kopf> Kollagen- (Gefäß-)Krankheit: </Kopf>
<Liste> <BEinSp>
<BAtom> systemisch </BAtom>
<BAtom> o.n.A. </BAtom> </BEinSp> </Liste>
</EinSp> </Eintrag> </Inkl1> </Inhalt>

...
    
```

Abbildung 3: Listenstrukturen: Text mit Layout, Text mit SGML-Auszeichnungen, Definition der Elemente

Ein einspaltiges Element *EinSp* besteht in seiner einfachsten Form aus logisch selbständigen Einheiten (Element *Atom*, logische Stufe 0), die nicht weiter strukturiert sind. Sie sind im Layout nicht eingerückt und nicht durch ein Bullet „■“ eingeleitet (so z. B. in Abbildung 2 der Inklusiv-Eintrag „Sklerodermie“ bei M34).

Die typische Struktur eines einspaltigen Elementes *EinSp* umfaßt jedoch logische Einheiten, die sich über mehrere Zeilen erstrecken: sie werden durch einen Begriff mit Doppelpunkt eingeleitet (Element *Kopf*), auf den das Element *Liste* mit weiteren Zeilen folgt. Im Layout des Beispiels in Abbildung 3 ist *Kopf* nicht eingerückt, jede mit *Kopf* zu einer logischen Einheit zu verknüpfende Zeile in *Liste* wird durch ein Bullet eingeleitet. Bei den Eigenschaften der Gruppe „Systemkrankheiten des Bindegewebes“ ist dies der *Kopf* „Autoimmunkrankheit“ und eine *Liste* mit „systemisch“ und „o. n. A.“ (= ohne nähere Angabe). Logische Einheiten aus der Verknüpfung von *Kopf* und *Liste* sind dann „Autoimmunkrankheit: systemisch“ und „Autoimmunkrankheit: o. n. A.“.

In diesem Beispiel sind die Einträge in *Liste* atomar (Element *BAtom*, das einleitende B im Elementnamen deutet das im Druckbild vorangestellte Bullet an, logische Stufe 1). Listeneinträge können jedoch als rekursive Struktur wieder aus ein- oder mehrspaltigen Einträgen bestehen, die ihrerseits wiederum einen Kopf und eine Liste aufweisen können: *BEinSp* ist daher im Prinzip die Wiederholung von *EinSp* auf einer tieferen Ebene. An dieser Stelle sei auf mehrstufige Beispiele verzichtet (siehe bei 5.2.).

5. Restrukturierung

SGML-Dateien lassen sich restrukturieren: Umsetzprogramme können neue Strukturen und andere Formate erzeugen. Dies sei an zwei knappen Beispielen gezeigt.

5.1. Dreistellige allgemeine Systematik

In Band I der ICD-10 geht der „Vierstelligen ausführlichen Systematik“ die „Dreistellige allgemeine Systematik“ voraus. Sie wird für die deutschsprachige Ausgabe aus der ausführlichen Systematik erzeugt, indem alle Viersteller (das Element *V* mit allen Unterelementen) und auf den höheren Ebenen die Beschreibung der Kategorieeigenschaften (das Element *Inhalt* mit allen Unterelementen) gelöscht werden.

5.2. Aufbereitung von Listenstrukturen

Die Listenstrukturen der Inklusiva und Exklusiva sind für EDV-gestützte Verschlüsselungshilfen in ihrer ursprünglichen Form wenig geeignet. Abbildung 4 zeigt ein komplexeres Beispiel aus Kapitel III „Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems“. Für den Leser einer gedruckten Fassung sind die logischen Zusammenhänge der einzelnen Zeilen evident, ein Suchprogramm würde jedoch auf diesem Text wegen der mangelhaften logischen Satzstruktur eine Reihe sinnloser Treffer produzieren, so z. B. bei der Suche

„FIND (Alder AND Hypersegmentation)/SAME SENTENCE“.

Für den Einsatz in Verschlüsselungssystemen, die auf einer Textsuche basieren, müssen diese Listen daher restrukturiert und zu logisch sinnvollen Sätzen zusammengefügt werden. Aufgrund der feinen Granularisierung in der SGML-Datei ist dies relativ leicht möglich. Abbildung 4 zeigt – neben dem ursprünglichen Text – diese Liste in restrukturierter Form als String-Kreuzprodukt; die einzelnen Einträge sind zu Sätzen zusammengefügt, die jeweils durch ein Semikolon von einander getrennt werden.

Ein Programm zur effektiven Suche in Texten (Einsatz von Rechts- und Linkstrunkierung, Kontextoperatoren, SAME-SENTENCE-Suche) wird auf dieser neuen

Abbildung 4: Listenstrukturen: Text mit Layout, Text mit SGML-Auszeichnungen, restrukturierter Text

D72.0 Genetisch bedingte Leukozytenanomalien

Anomalie (Granulation) (Granulozyten) oder Syndrom:

- Alder-
- May-Hegglin-
- Pelger-Huët-

Hereditär:

- Leukomelanopathie
- leukozytär:
 - Hypersegmentation
 - Hyposegmentation

Anomalie (Granulation) (Granulozyten) oder Syndrom: Alder-;

Anomalie (Granulation) (Granulozyten) oder Syndrom: May-Hegglin-;

Anomalie (Granulation) (Granulozyten) oder Syndrom: Pelger-Huët-;

Hereditär: Leukomelanopathie;

Hereditär: leukozytär: Hypersegmentation;

Hereditär: leukozytär: Hyposegmentation;

```
<V><VCode>D72.0</VCode><VTitel>Genetisch bedingte Leukozytenanomalien</VTitel>
<Inhalt><Inkl2><Eintrag><Einsp>
<Kopf>Anomalie (Granulation) (Granulozyten) oder Syndrom:</Kopf><Liste><BEinSp>
<BAtom>Alder-</BAtom>
<BAtom>May-Hegglin-</BAtom>
<BAtom>Pelger-Huët-</BAtom></BEinSp></Liste>
<Kopf>Hereditär:</Kopf><Liste><BEinSp>
<BAtom>Leukomelanopathie</BAtom>
<BKopf>leukozytär:</BKopf><BListe>
<SBAtom>Hypersegmentation</SBAtom>
<SBAtom>Hyposegmentation</SBAtom>
</BListe></BEinSp></Liste></EinSp></Eintrag></Inkl2></Inhalt></V>
```

Struktur wesentlich bessere Ergebnisse erzielen können als auf den ursprünglichen Listen.

In der leichten Restrukturierbarkeit der Daten liegt aber auch eine Gefahr: Keinesfalls darf dabei die Klassifikation selbst verändert werden. Die Position jeder Krankheit innerhalb der Systematik muß invariant und erkennbar bleiben.

6. Schlußfolgerung

Mit SGML steht ein adäquates Format zur Repräsentation der Struktur der ICD-10-Systematik bereit. Da die Daten auf einer logischen Ebene beschrieben werden und zwischen beliebigen EDV-Umgebungen austauschbar sind, erhält der Anwender eine maschinenlesbare Version, aus der er vom Buch über Datenbanken bis hin zu EDV-gestützten Verschlüsselungshilfen die unterschiedlichsten Benutzersichten ableiten kann. Sofern SGML-basierte Standardsoftware vorhanden ist, kann dies ohne aufwendige Programmierarbeit geschehen; ansonsten ermöglicht die feinkörnige SGML-Auszeichnung eine Anpassung an die Erfordernisse des jeweiligen Programmpaketes.

Die Offenheit des beschriebenen Ansatzes wird erkaufte durch eine nicht gerade geringe Vergrößerung der Dateien: die Entity References für die zahlreichen Umlaute und Sonderzeichen, aber auch die SGML-Auszeichnungen selbst führen zu einem Anwachsen des Speicherplatzbedarfes gegenüber der reinen Textdatei auf etwa das Anderthalbfache.

Die dargelegten Vorteile überwiegen diesen Nachteil jedoch bei weitem, wenn man berücksichtigt, daß die SGML-Datei in erster Linie eine Distributionsdatei ist, die für die konkrete EDV-Anwendung – selbst bei SGML-basierten Systemen – in ein kompaktes internes Format umgesetzt wird.

7. Danksagung

Besonderer Dank gilt Herrn Dipl.-Inf. Peter Fankhauser (Institut für Integrierte Publikations- und Informationssysteme der Gesellschaft für Mathematik und Datenverarbeitung, Darmstadt) für seine unermüdliche Hilfe beim Einsatz von DREAM.

8. Anhang

Da im Rahmen dieser Arbeit kein Überblick gegeben werden kann über die vielfältigen Anwendungen von SGML und die Fülle der verfügbaren Werkzeuge, dürfen einige Hinweise auf weitere Informationsquellen nicht fehlen.

Die umfassendste Darstellung von SGML mit zahlreichen Kommentaren, Querverweisen und Anwendungshinweisen stammt von Goldfarb (GOLDFARB 1990), der an der Entwicklung von SGML maßgeblich beteiligt war. Eine Neuauflage seines Handbuchs, das den vollständigen Text der ISO-Norm enthält, steht kurz vor dem Erscheinen.

Einen regen Erfahrungsaustausch mit SGML-Anwendern in aller Welt bietet die Newsgroup „comp.text.sgml“ im UseNet des INTERNET. Einen guten Überblick über die Literatur zu SGML verschafft die „Annotated Bibliography and List of Resources“ von R. Cover, die in der Datei „./pub/SGML/bibliography“ auf dem FTP-Server „ftp.ifi.uio.no“ im INTERNET bereitsteht. Dort und auf vielen weiteren FTP-Servern

steht auch Public-Domain-Software für SGML-Anwendungen zur Verfügung. Eine Liste von freien und kommerziellen SGML-Werkzeugen (Parser, Editoren, Datenbanken etc.) mit Bezugsquellen ist ebenfalls auf „ftp.ifi.uio.no“ als Datei „/pub/SGML/SGML-Tools/SGML-Tools.text“ vorhanden. Sie wird in unregelmäßigen Intervallen gepflegt.

9. Literatur

- BRAUN, R. und DIEKMANN, F. (1991): ICD-Ziele und Veränderungen der 10. Revision. Praxis Medizinischer Dokumentation 11: 59–62.
- BRYAN, M. (1988): SGML. An Author's Guide to the Standard Generalized Markup Language. Addison-Wesley Publishing Company, Reading, Massachusetts.
- Bundesministerium für Gesundheit (1993): Bekanntmachung über die Einführung eines Diagnose-schlüssels zur Angabe auf den Abrechnungsunterlagen für die Krankenkassen. Bundesanzeiger 45: 6253.
- GÖTTKE, T. und FANKHAUSER, P. (1992): DREAM 2.0. User Manual. Arbeitspapiere der GMD, 660. Gesellschaft für Mathematik und Datenverarbeitung mbH, Sankt Augustin.
- GOLDFARB, C. F. (1990): The SGML Handbook. Oxford University Press, Oxford.
- International Standards Organization (Hrsg.) (1986): Information Processing – Text and Office Systems – Standard Generalized Markup Language (SGML). (ISO 8879). International Standards Organization, Geneva.
- REIMER, U. (1991): Einführung in die Wissensrepräsentation. Teubner-Verlag, Stuttgart.
- VAN HERWIJNEN, E. (1990): Practical SGML. Kluwer Academic Publishers, Dordrecht, Boston, London.
- WIESNER, G. (1993): Die 10. Revision der Internationalen Statistischen Klassifikation von Krankheiten – Änderungen und Neuerungen. Bundesgesundheitsblatt 36: 193–197.
- World Health Organization (Hrsg.) (1992): International Statistical Classification of Diseases and Related Health Problems. Volume 1. World Health Organization, Geneva.

Anschrift des Verfassers: Dr. med. Michael Schopen, Deutsches Institut für medizinische Dokumentation und Information, DIMDI, Weißhausstraße 27, 50939 Köln

Eingang des Manuskripts 1. 3. 94
Eingang der revidierten Fassung 18. 5. 94

Luftverunreinigungen und Lungenkrebs

B. Pesch¹⁾, K. H. Jöckel^{2,3)}, H. E. Wichmann^{4,5)}

Stichworte:

Lungenkrebs, Luftverunreinigung, Risikoabschätzung, Übersichtsarbeit

Key words:

lung cancer, air pollution, risk assessment, review

Zusammenfassung

Eine höhere Sterblichkeit an Lungenkrebs, der häufigsten Krebstodesursache beim Mann, in Ballungsräumen und Großstädten gab Anlaß zu der Hypothese, daß dieser urbane Faktor durch Luftverunreinigungen verursacht sein könnte. Ökologische Korrelationen, die regionale Mortalitätsraten mit den dort gemessenen Schadstoffkonzentrationen verknüpfen, liefern jedoch keine verlässliche Risikoabschätzung, da sie Risikofaktoren wie Rauchen und berufliche Belastungen nicht ausreichend berücksichtigen. Betrachtet man statt dessen epidemiologische Studien mit individuellen Belastungsdaten, dann zeigt sich, daß für Luftverunreinigungen ein vergleichsweise geringes relatives Risiko verbleibt, dessen Größenordnung unterhalb von 1,5 anzusetzen ist. Die Autoren schließen sich der Schätzung von 5% bis 10% der Lungenkrebsfälle an, die in stark belasteten Gebieten dem Einfluß der Luftverschmutzung zuzuschreiben wären. Auf der Grundlage konservativer Risikoextrapolationen liegt das Attributivrisiko für die alten Bundesländer bei ca. 2%.

Summary

Reasons for believing that air pollution might be an important risk factor in the development of lung cancer, the leading cancer among males, were supported by pronounced urban/rural gradients of mortality. But ecological correlations of lung cancer death rates with measurements of pollutants are an inappropriate approach for risk assessment, because smoking and occupational exposition, the known risk factors for lung cancer, have to be

¹⁾ Medizinisches Institut für Umwelthygiene an der Universität Düsseldorf

²⁾ Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen

³⁾ Bremer Institut für Präventionsforschung und Sozialmedizin

⁴⁾ Institut für Epidemiologie, GSF Neuherberg

⁵⁾ Bergische Universität, Wuppertal

controlled, which cannot be done sufficiently at an aggregated data level. Risk estimations with individual exposure data from epidemiological studies suggest odds ratios of less than 1.5. The authors agree with the estimation that 5 to 10% of the lung cancer cases in highly polluted areas could be attributed to the ambient air pollution. Based on conservative risk assessment about 2% of the cancer cases of the West German population could be attributed to the ambient air pollution.

Einleitung

Der Lungenkrebs ist in allen Industrienationen die führende Krebstodesursache beim Mann. Die Mortalität zeigt ein ausgeprägtes Stadt-Land-Gefälle und einen deutlichen Geschlechtsunterschied. Während beim männlichen Geschlecht in den 80er Jahren die Lungenkrebssterblichkeit sich auf einem hohen Niveau eingependelt hat, ist beim weiblichen Geschlecht ein kontinuierlicher Aufholtrend zu beobachten, für den keine Anzeichen zu einer Abflachung zu erkennen sind (PESCH & POTT, 1992).

Da für diese Erkrankung auch heute noch keine Therapie mit günstiger Prognose existiert, liegt der Schwerpunkt der Bemühungen auf der Vermeidung von Risiken, die die Entstehung von Lungenkrebs beeinflussen. Es gibt kaum eine Krebsart mit einem vergleichbar deutlichen Einfluß des Tabakrauchens, ferner erhöhen berufliche Expositionen gegenüber Asbest, Chromatstaub, Nickel- und Arsenverbindungen, polyzyklischen aromatischen Kohlenwasserstoffen (PAH) und Radon das Risiko, ein Bronchialkarzinom zu entwickeln. Als dritter Bereich sind die möglicherweise umweltbedingten Lungenkrebsrisiken zu nennen. Besonders in den Jahren des industriellen Aufschwunges nach dem Zweiten Weltkrieg waren in Ballungsräumen beträchtliche Verunreinigungen der Außenluft zu verzeichnen. Während in den alten Bundesländern für die Mehrzahl der Schadstoffe deutliche Konzentrationssenkungen durch verschärfte Auflagen für industrielle Emissionen und Umstellung des Hausbrandes auf emissionsarme Heizungstypen seit Mitte der 70er Jahre verzeichnet werden können, sind die neuen Bundesländer bis Ende der 80er Jahre in hohem Maße durch Luftverunreinigungen belastet gewesen. Noch 1989 lagen in Großstädten der neuen Bundesländer die SO_2 -Immissionswerte im Jahresmittel zwischen 60 und $380 \mu\text{g}/\text{m}^3$ im Vergleich zu $30 \mu\text{g}/\text{m}^3$ im Ruhrgebiet (BfLR, 1990). Weiterhin enthalten die Emissionen durch den stetig ansteigenden Fahrzeugverkehr mit Rußpartikeln und PAH ein kanzerogenes Potential, das nach derzeitigen Risikoabschätzungen zur Entwicklung von Bronchialkarzinomen beitragen kann.

Hohe Lungenkrebsraten bei chinesischen Frauen, die beträchtlichen Schadstoffkonzentrationen durch Kochen in engen und schlecht belüfteten Räumen ausgesetzt sind, haben gezeigt, daß zusätzlich zu den Luftverunreinigungen der Außenluft auch Innenraumbelastungen berücksichtigt werden müssen. Dabei ist von Bedeutung, daß der überwiegende Teil des Tages im Innenraum verbracht wird. Während zum Einfluß des Tabakrauchens und zu beruflichen Risiken eine große Zahl epidemiologischer Erhebungen vorliegt, ist der Einfluß von Luftverunreinigungen auf die Entstehung des Bronchialkarzinoms bislang nur unzureichend untersucht worden, obwohl das ausgeprägte Stadt-Land-Gefälle der Lungenkrebssterblichkeit bereits in den 50er und 60er Jahren Anlaß gegeben hat, einen Einfluß von Luftverunreinigungen anzunehmen. 1977 gab die International Agency for Research on Cancer (IARC) eine Monographie über den möglichen Einfluß von Luftverunreinigungen auf das Krebsgeschehen heraus (MOHR et al., 1977), eine neuere Monographie erschien zu diesem Thema von der European School of Oncology (TOMATIS, 1990). Die Umweltministerkonferenz hat sich nach Veröffentlichung des ersten Krebsatlasses für die alten Bundesländer (FRENTZEL-BEYME et al., 1979) erstmals mit dem Problem des Krebsrisikos durch Luftverunreinigungen befaßt und vorsorglich Gegenmaßnahmen eingeleitet. In den Folgejahren

wurden Beurteilungsmaßstäbe für das kanzerogene Potential von Luftverunreinigungen erstellt (LAI, 1992). Weiterhin wurde im Auftrag des Umweltbundesamtes Berlin gemeinsam vom Bremer Institut für Präventionsforschung und Sozialmedizin, dem Medizinischen Institut für Umwelthygiene an der Universität Düsseldorf und der Wissenschaftlichen Einheit Statistik der Universität Bremen eine epidemiologische Pilotstudie zur Frage des möglichen Lungenkrebsrisikos von Luftverunreinigungen unter Berücksichtigung der Risikofaktoren Rauchen und berufliche Belastung durchgeführt (GREISER et al., 1984, WICHMANN et al. 1991, JÖCKEL et al. 1992). Im folgenden sollen die in dieser sowie in vergleichbaren Studien erzielten Ergebnisse genutzt werden, eine Bewertung des Einflusses von Luftverunreinigungen auf die Entstehung des Bronchialkarzinoms zu treffen. Hierzu werden die Daten zusammengestellt, die zu einer Bewertung des Einflusses von Luftverunreinigungen auf die Entstehung des Bronchialkarzinoms herangezogen werden müssen. Als Ergebnis dieser Untersuchung ist die Entwicklung methodischer Ansätze zur Expositions- und Risikoabschätzung für Luftverunreinigungen zu sehen.

Fall-Kontroll-Studien liefern vorrangig Abschätzungen für relative Risiken, die angeben, um wieviel höher oder niedriger das Erkrankungsrisiko in einer exponierten Personengruppe ist im Vergleich zu den nicht exponierten Personen. Um eine Hochrechnung auf den Anteil der Krebsfälle zu erhalten, der in der Allgemeinbevölkerung dem betrachteten Risikofaktor zugeschrieben werden kann, sind Kenntnisse über die Verteilung der Exposition notwendig. Für das bevölkerungsbezogene Attributivrisiko sind somit die Häufigkeit der Erkrankung, die Stärke des betrachteten Risikofaktors und seine Verbreitung in der Bevölkerung (Expositionsraten) von Bedeutung. Der Einfluß von anderen möglichen Risikofaktoren muß dabei entsprechend berücksichtigt werden. Dies ist ein wesentlicher Mangel älterer Studien zum Einfluß der Luftverunreinigungen. Daher soll zuerst das Vorkommen des Lungenkrebses in der Bundesrepublik Deutschland im Kontext des Risikoprofils beschrieben werden.

1. Untersuchungen zur Lungenkrebssterblichkeit in der Bundesrepublik Deutschland

Angaben über die jährlichen Neuerkrankungsfälle an Lungenkrebs liegen für das Gebiet der alten Bundesländer nicht vor, da nur im Saarland verlässliche Daten zur Krebshäufigkeit dokumentiert werden. Das Krebsregister Saarland gibt für 1988 die nach SEGI (1960) altersstandardisierte Lungenkrebsinzidenz bei Männern mit 68,7/100000 und bei Frauen mit 8,8/100000 an (STATISTISCHES LANDESAMT SAARLAND, 1991), die Mortalitätsraten lagen bei 68,8/100000 für Männer und 7,3/100000 für Frauen. Etwas geringer werden die Inzidenzraten mit 60,3/100000 für Männer und 6,9/100000 für Frauen in den Jahren 1983 bis 1987 für das Gebiet der ehemaligen DDR berichtet (ZENTRALINSTITUT FÜR KREBSFORSCHUNG, 1991).

Aufgrund des Fehlens von geeigneten Inzidenzdaten im Bereich der alten Bundesländer muß auf die amtliche Mortalitätsstatistik zurückgegriffen werden. Durch die schlechte Prognose beim Bronchialkarzinom spiegeln die Mortalitätsdaten hinreichend die Inzidenzen wider, auch wenn die Validität der Angaben auf Todesbescheinigungen kritisch beurteilt wird. MÜLLER und BOCTER (1991) haben hierzu jedoch festgestellt, daß die Nennung von bösartigen Neubildungen auf Todesbescheinigungen in den meisten Fällen ausreichend gut begründet ist.

In der Bundesrepublik Deutschland verstarben 1991 27785 Männer und 7252 Frauen (STATISTISCHES BUNDESAMT, 1993) an dieser Erkrankung. Unter Zugrundelegung der Relation von Inzidenz zu Mortalität anhand der Daten des Krebsregisters Saarland, wonach für Männer Neuerkrankungsfälle und Sterbefälle etwa gleichhäufig registriert werden und bei Frauen etwa ein Fünftel mehr inzidente Fälle als Sterbefälle erfaßt

werden, kann angenommen werden, daß etwa 9000 Neuerkrankungsfälle beim weiblichen Geschlecht auftreten. Für die meisten Industrienationen wird in Stadtgebieten und Ballungsräumen eine bis zu 60% höhere Lungenkrebssterblichkeit beim männlichen Geschlecht und eine bis zu 90% höhere Lungenkrebsmortalität bei Frauen im Vergleich zu ländlichen Regionen festgestellt (MUIR et al., 1987, DOLL, 1992). Auch für die Bundesrepublik Deutschland kann eine auffällige Häufung des Bronchialkarzinoms in den Industrieregionen und Großstädten gezeigt werden (BECKER et al., 1984) (Abbildung 1), welche etwa 50% über dem Bundesdurchschnitt liegt (BECHER, 1992). Betrachtet man die Situation am Beispiel von Nordrhein-Westfalen genauer, so zeigt sich ein Stadt-Land-Verhältnis der Lungenkrebssterblichkeit zwischen nördlichem Ruhrgebiet und Nordost-Westfalen von 1,4:1 für Männer und 1,6:1 für Frauen (PESCH & POTT, 1992). In der zeitlichen Entwicklung ergibt sich nach dem Zweiten Weltkrieg ein starker Anstieg der Lungenkrebssterblichkeit beim Mann, die in den 70er und 80er Jahren ein Plateau erreicht hat (BECKER et al., 1984, HOFFMEISTER & JUNGE, 1984). Eine nach Altersgruppen gegliederte Darstellung zeigt, daß bei Männern eine Zunahme der Lungenkrebssterblichkeit in den Jahren 1965 bis 1990 nur in den oberen Altersklassen festzustellen ist (BECHER, 1992, PESCH & POTT, 1992). Demgegenüber sieht die Situation für Frauen vor allem in den älteren Jahrgangsklassen zwar bezüglich der Rate noch erheblich günstiger aus, jedoch ist ein stetig ansteigender Trend in allen Altersgruppen vorhanden, und Zeichen für eine Abflachung sind – im Gegensatz zum männlichen Geschlecht – nicht erkennbar. Besonders hervorzuheben ist der rapide Anstieg der weiblichen Lungenkrebserkrankungen in Ballungsräumen und Stadtstaaten.

2. Risikofaktoren des Bronchialkarzinoms

Neuere Übersichten zur Epidemiologie des Lungenkrebses in Deutschland geben BECHER (1992) und WICHMANN et al. (1991). Sichere und wahrscheinliche Risikofaktoren des Bronchialkarzinoms sind in der folgenden Tabelle 1 zusammenfassend dargestellt:

Tabelle 1: Risikofaktoren des Lungenkrebses

Biologische Faktoren	Alter Geschlecht ¹⁾ genetische Faktoren
Starke nichtnatürliche Einflußgrößen	Rauchen berufliche Belastung
Weitere Risikofaktoren	Radon und Radonzerfallsprodukte Passivrauchen Luftverunreinigung Ernährung, insbesondere Vitamin-A-Mangel
¹⁾ Wahrscheinlich kein eigenständiger Risikofaktor	

Der mit Abstand wichtigste exogene Risikofaktor für den Lungenkrebs ist das Tabakrauchen, dessen Kanzerogenität bereits 1939 von MÜLLER erkannt und durch verschiedene Studien in den USA und in England in den 50er und 60er Jahren nachgewiesen wurde (DOLL, 1953, HAMMOND, 1966, DOLL & PETO, 1976). Daher muß in erster Linie geprüft werden, welche zeitlichen Trends und regionalen Unterschiede im Rauchverhalten vorliegen.

In den Industrienationen sind über 90% der Lungenkrebspatienten Raucher (MORTON & TREYVE, 1982). Die Abschätzungen von DOLL und PETO (1981) und HAMMOND und SEIDMAN (1980) ergeben für die USA, daß ca. 80% bis 90% der Lungenkrebsfälle bei Männern und ca. 30% bis 60% bei Frauen durch das Rauchen bedingt sind. Das relative Risiko von Zigarettenrauchern liegt für Männer nahezu 10fach höher als bei Nichtrauchern (LUBIN et al., 1984, JÖCKEL et al., 1992).

1986 wurde von der International Agency for Cancer Research (IARC) eine Übersicht zum Risikofaktor Tabakrauchen herausgegeben. Danach sind zu Beginn der 80er Jahre in den Industrieländern durchschnittlich 40% der erwachsenen Männer und 30% der Frauen Raucher gewesen, wobei die Raucherrate der Bundesrepublik Deutschland mit 40% für Männer und 29% für Frauen angegeben wurde. In der Altersgruppe der 20- bis 45jährigen Männer haben in den alten Bundesländern in den 80er Jahren knapp die Hälfte geraucht, bei gleichaltrigen Frauen liegt dieser Anteil mit etwa 30% bis 40% etwas niedriger, weitere rund 20% waren ehemalige Raucher (DER BUNDESMINISTER FÜR GESUNDHEIT, 1991). Bei älteren Menschen nimmt der Anteil der Exraucher beim männlichen Geschlecht zu, etwa die Hälfte der Männer über 60 haben das Rauchen aufgegeben. Mit zunehmendem Lebensalter steigt dagegen bei Frauen der Anteil der Nieraucherinnen deutlich an. Die 1978 im Rahmen des Mikrozensus durchgeführte Erhebung zum Rauchverhalten in den alten Bundesländern zeigt, daß in Ballungsräumen und Stadtstaaten mehr geraucht wird als auf dem Land (WICHMANN et al., 1991).

Nach dem Zweiten Weltkrieg kam es in den alten Bundesländern zu einem steilen Anstieg des pro-Kopf-Verbrauchs an Zigaretten. Von 1955 (mit damals noch unter 1000 Zigaretten pro Kopf) bis 1970 erfolgte eine Verdopplung des Zigarettenkonsums. Ab 1970 hat sich der Verbrauch auf einem hohen Niveau von etwa 2000 Zigaretten pro Erwachsenen eingependelt. Dieser Trend verläuft zeitverschoben parallel zu dem Verlauf der Lungenkrebssterblichkeit bei Männern. Da das Rauchen insbesondere bei jüngeren Frauen jedoch noch weiter zunimmt, ist zu erwarten, daß der Anstieg wegen der langen Latenzzeit der Erkrankung bei Frauen noch für Jahrzehnte anhält.

In den neuen Bundesländern wurde ein zeitlich verzögerter Anstieg im Verbrauch festgestellt (STATISTISCHES AMT DER DDR, 1990). Nach Ergebnissen aktueller Studien rauchen dort derzeit die Frauen besonders in den jüngeren Altersgruppen noch weniger als in den alten Bundesländern. Bei den Männern im Alter zwischen 25 und 64 Jahren raucht ein ähnlicher Prozentsatz wie in Westdeutschland, der Zigarettenkonsum pro Raucher ist aber etwas niedriger (BORMANN et al., 1991).

Der nächste bedeutende Risikofaktor für das Bronchialkarzinom besteht in einer beruflichen Exposition gegenüber einer Reihe von Karzinogenen wie Asbest, polyzyklischen aromatischen Kohlenwasserstoffen (PAH), Nickel-, Chrom- und Arsenverbindungen. Insgesamt nehmen DOLL und PETO (1981) für die USA an, daß 15% der männlichen und 5% der weiblichen Lungenkrebsfälle beruflich bedingt sind. Für die alten Bundesländer liegen die Schätzungen des attributablen Risikos der beruflichen Exposition von JÖCKEL et al. (1992) in einer vergleichbaren Größenordnung von 7% bis 12% für Männer, das relative Risiko liegt für Männer in der Größenordnung von unter 2. Eine ausführliche Übersicht über Risikoberufe und Risikoabschätzungen geben JÖCKEL et al. (1992b). In den neuen Bundesländern ist wegen der teilweise deutlich höheren Belastungen durch kanzerogene Arbeitsstoffe ein höheres Attributivrisiko anzunehmen, ohne daß sich dieses derzeit quantifizieren ließe. Hinzu kommt eine sehr große Zahl berufsbedingter Lungenkrebserkrankungen aus dem Uranbergbau der ehemaligen DDR, die vermutlich alle Erkrankungen durch chemische Kanzerogene bei weitem übertrifft. In Sachsen und Thüringen waren in der Zeit zwischen 1946 und 1990 mindestens 200000 Personen im Uranbergbau tätig, für die aufgrund der hohen Radonbelastung von einem deutlich erhöhten Lungenkrebsrisiko ausgegangen werden muß (BRÜSKE-HOHLFELD et al., 1993, MEHNERT et al., 1992).

Zu den Unterschieden in der räumlichen Verteilung von Berufen und Branchen kommen Unterschiede im Rauchverhalten zwischen den verschiedenen Berufsgruppen hinzu. Während z. B. über die Hälfte der im Mikrozensus 1989 befragten männlichen Chemikarbeiter, Metallarbeiter und Kraftfahrer Raucher waren, betrug der Raucheranteil bei Ingenieuren, Chemikern, Ärzten und Lehrern nur etwa ein Viertel dieser Berufsgruppen (DER BUNDESMINISTER FÜR GESUNDHEIT, 1991). Berufe mit hohem Risikopotential kommen in industriellen Ballungsräumen gehäuft vor, und in ländlichen Regionen wird im allgemeinen weniger geraucht als in Großstädten. Dies muß bei der Beurteilung der regionalen Auffälligkeiten in der Sterblichkeit an Lungenkrebs unter dem Gesichtspunkt, inwieweit Luftverunreinigungen zu einem solchen Stadt-Land-Faktor beigetragen haben, berücksichtigt werden.

Ferner ist die radioaktive Belastung durch Radon und seine Zerfallsprodukte in Wohnungen von Bedeutung. Das radioaktive Edelgas gelangt aus dem Gesteinsuntergrund ins Haus, wobei die entstehende Radonkonzentration vom Zustrom und von der Lüftung abhängt. Gegenden mit hoher Radonbelastung sind der Süden Thüringens und Sachsens, Ostbayern, das Saarland und die Eifel. Epidemiologische Studien zum Lungenkrebsrisiko durch Radon in Wohnungen werden derzeit weltweit durchgeführt, auch in der Bundesrepublik (KREIENBROCK et al., 1992, WICHMANN et al., 1993). Derzeit wird das Risiko durch Extrapolation aus Studien an Uranbergarbeitern abgeschätzt (WHO, 1993). Auf dieser Grundlage kommt die STRAHLENSCHUTZKOMMISSION (1987) zu der Aussage, daß 4 bis 12% der Bronchialkarzinomerkrankungen in den alten Bundesländern möglicherweise durch Radon bedingt sind. Diese Abschätzung wird von JACOBI (1993) bestätigt.

Während die Wirkung externer Faktoren auf das Krebsgeschehen seit längerem untersucht wird, haben Studien über endogene Faktoren erst mit der zu Beginn der 80er Jahre erkannten Funktion der Onkogene deutliche Impulse erhalten (HELMAN & THIELE, 1991). Genetische Transformationen von Normalzellen führen zur Aktivierung von Onkogenen und in der Folge zur Dedifferenzierung des Karyotyps. So wurde geschätzt, daß 50% aller Tumoren Punktmutationen im p53-Gen aufweisen (HUGHES, 1994). Neoplastische Veränderungen von Zellen können jedoch von sogenannten Suppressorgenen unterdrückt werden. Durch Inaktivierung dieser Gene, u. a. durch Verlust von genetischem Material, das diese Gene trägt, wird die Tumorentstehung begünstigt. Altersabhängige Einflüsse müssen hierbei noch evaluiert werden.

Auch das Immunsystem scheint bei der Krebsentstehung eine Schlüsselrolle zu spielen. So bestehen strukturelle Ähnlichkeiten zwischen bestimmten biochemischen Signalen wie den ‚growth factors‘ und Onkogen-Produkten (ULLRICH, 1992). Für die Untersuchung der Frage, ob die großen Unterschiede in der Krebsinzidenz zwischen Männern und Frauen auch mit der unterschiedlichen Chromosomenausstattung verbunden sind, wird festzustellen sein, welche Immunfaktoren auf den Geschlechtschromosomen kodiert werden.

Diese Prozesse stehen in einem noch nicht im einzelnen erkannten Zusammenspiel mit dem Metabolismus von krebserzeugenden Substanzen. Nach WYNDER (1991) überfordert ein ‚metabolic overload‘ an exogen zugeführten Stoffen die Regelmechanismen des Organismus, und in der Folge können Schadstoffe oder ihre Stoffwechselprodukte zur Krebsentstehung beitragen. Die Metabolisierung von Schadstoffen ist jedoch individuell verschieden, wobei diesem eine genetische Veranlagung zugrundeliegen kann, welche insbesondere die Kodierung für ‚drug metabolizing enzymes‘ (DME) steuert (u. a. GONZALES et al., 1988). Diese Veranlagungen stehen über Einflüsse auf den Nitrosamin- und Benzpyren-Stoffwechsel im Verdacht eines erhöhten Risikos für Lungenkrebs (CAPORASO, 1992).

Auch die regionale Verteilung von bestimmten Genotypen kann unterschiedlich sein. Erste Ergebnisse zur Verbreitung von CYP1A1-Allelen, die bei der Entstehung von

Lungenkrebs eine Rolle zu spielen scheinen, belegen ethnische Unterschiede (NEBERT & PUGA, 1992), wonach die Häufigkeit des Msp I 1,9-kb-Fragmentes in Japan etwa doppelt so hoch ist wie in Europa. Für Japan wird von der IARC (1986) ein hoher Anteil von 70% männlicher Raucher angegeben. Dazu liegen die nach SEGI standardisierten Inzidenzraten mit durchschnittlich 35 bis 40 Lungenkrebsfällen auf 100000 Einwohner vergleichsweise niedrig (IARC, 1992).

3. Verunreinigungen der Außenluft

„Luftverunreinigungen“ bezeichnen gasförmige, partikelförmige oder faserige Stoffe in der Außenluft, die durch anthropogene Prozesse, vorrangig industrielle Produktion, Hausbrand, Energieerzeugung, Abfallentsorgung und Kraftverkehr, oder deren Folgen (z. B. Verwitterung, photochemische Reaktionen) entstanden sind. Erst in der zweiten Hälfte dieses Jahrhunderts wurde das Gefahrenpotential von Luftverunreinigungen erkannt. Messungen zum Vorkommen von Luftverunreinigungen liegen deshalb erst für die neuere Zeit vor, und diese sind in verschiedener Hinsicht unbefriedigend. Das Spektrum dieser Stoffe ist umfangreich und im Schadenspotential verschieden. Meßtechnisch kann nur eine Teilmenge observiert werden. In der Vergangenheit hat sich zum Beispiel Schwefeldioxid als eine solche Leitsubstanz erwiesen, für die die bislang umfangreichsten Meßdaten vorliegen, obwohl auch hier zeitliche und räumliche Lücken vorhanden sind. Für eine Bewertung des kanzerogenen Potentials der Luftverunreinigungen sind jedoch Schwefeldioxid und einige andere Meßgrößen nicht unmittelbar geeignet. Sie können nur bedingt als indirektes Maß für das Vorkommen von Prozessen angesehen werden, in deren Folge auch kanzerogene Substanzen emittiert werden. Im Hinblick auf den Lungenkrebs sind polyzyklische aromatische Kohlenwasserstoffe (PAH) (POTT & HEINRICH, 1992) und Dieselrußpartikel (WICHMANN & BRÜSKE-HOHLFELD, 1991) von besonderer Bedeutung. Ferner spielen anorganische Arsen- und Cadmiumverbindungen sowie Asbest eine gewisse Rolle (Tabelle 2). Eine Charakterisierung der Immissionsbelastung durch diese Stoffe gelingt erst seit den 80er Jahren und ist für einige wichtige Komponenten wie Dieselrußpartikel bis heute nicht zufriedenstellend gegeben. Auf der

Tabelle 2: Emission von kanzerogenen Luftverunreinigungen (nach LAI (1992))

Schadstoff	Emissionsart	Hauptemissionsquellen	Jahresemission
Dieselruß	partikulär	Kraftfahrzeugverkehr (Dieselnutzfahrzeuge)	70000 t/a
PAH (BaP)	gasförmig bzw. absorbiert an Partikeln	unvollständige Verbrennung von org. Material (Hausbrand, Kfz-Verkehr, Kokereien)	4 – 11 t BaP/a
anorgan. Arsen- und Cadmium-Verbindungen	an Staubpartikel gebunden	Industrie (Metallerzeugung, Steine/Erden, Glasherstellung)	20 – 70 t/a As 25 – 50 t/a Cd
Asbest	Fasern	Asbestverschleiß (Verwitterung, Abrieb) Asbestindustrie	100 – 120 t/a

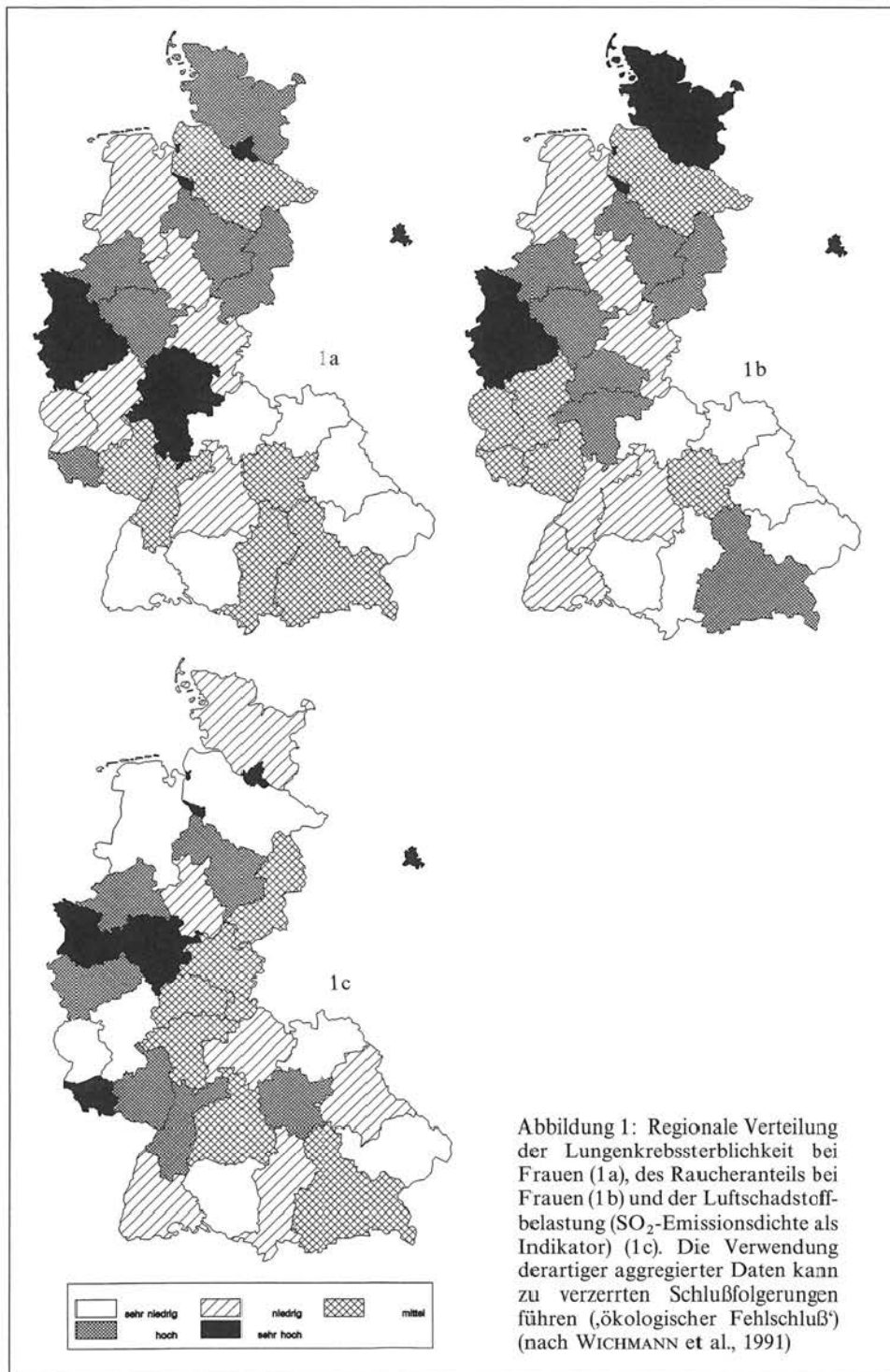


Tabelle 3: Zusätzliches Lungenkrebsrisiko²⁾ in der Bundesrepublik Deutschland (alte Bundesländer) durch Luftschadstoffe (nach LAI (1992))

	Ballungsgebiete ¹⁾		ländlich geprägte Gebiete ¹⁾	
	Immissionskonz.	Risiko ($\times 10^{-5}$) ²⁾	Immissionskonz.	Risiko ($\times 10^{-5}$)
Dieselrußpart.	7,2 $\mu\text{g}/\text{m}^3$	50	0,9 $\mu\text{g}/\text{m}^3$	6,3
PAH	1,8 ng/m^3	13	0,7 ng/m^3	5,0
Arsen	11 ng/m^3	4,3	0,9 ng/m^3	1,1
Cadmium	3,3 ng/m^3	3,9	2,7 ng/m^3	1,1
Asbest	110 F/m^3	2,1	38 F/m^3	0,8
Gesamt		73,3		14,3

¹⁾ Bevölkerungsanteil: Ballungsgebiete (incl. West-Berlin): 55,5%, kleinstädtische Gebiete: 28,7%, ländliche Gebiete: 15,8%

²⁾ zusätzliches lebenslanges Lungenkrebsrisiko (70 Jahre), konservative Abschätzung anhand von unit risk-Werten

Grundlage dieser Angaben ergibt sich in Ballungsgebieten ein 4- bis 5mal so hohes Zusatzrisiko durch die Außenluft wie in ländlichen Regionen (Tabelle 3).

Für die Vergangenheit liegen nur für PAH in einigen Regionen Immissionsmeßwerte vor (Abbildung 2). Diese zeigen, daß die Belastung in industriellen Ballungsräumen wie dem Ruhrgebiet erheblich höher war als heute, während sie sich in ländlichen Gebieten kaum geändert hat. Die Messungen stammen überwiegend aus den 70er Jahren, doch man kann davon ausgehen, daß die Situation in den 60er Jahren ähnlich war. Für frühere Zeiten und für andere Kanzerogene liegen keine Daten vor.

In Hinblick auf die epidemiologische Untersuchung des Zusammenhangs zwischen Luftschadstoffbelastung und dem Auftreten von Lungenkrebskrankungen ist dies eine schwierige Ausgangssituation. Wegen der langen Latenzzeiten des Bronchialkarzinoms von 20 und mehr Jahren wäre es wünschenswert, für die Untersuchung der heute auftretenden Erkrankungsfälle die individuelle Exposition gegenüber den wichtigen Kanzerogenen mindestens für die Zeit nach dem Krieg zu kennen. Davon ist man weit entfernt, so daß nur ein näherungsweise Zugang möglich ist.

4. Belastungen im Innenraum

Bei den epidemiologischen Untersuchungen zum Lungenkrebsrisiko durch Luftverunreinigungen wurden bisher meist nur Außenluftkonzentrationen berücksichtigt. Dies ist berechtigt, wenn für die relevanten Kanzerogene keine Innenraumquellen vorhanden sind, weil dann die Innenraumkonzentration allein von der Außenluftkonzentration bestimmt wird.

In vielen Fällen ist diese Annahme aber nicht zutreffend. Die wichtigste Schadstoffgruppe, die im Innenraum freigesetzt wird, stellen PAH dar. Als Quelle sind vor allem Feuerstätten im Wohnbereich wie Einzelöfen, offene Kamine und Küchenherde zu nennen. Insbesondere Holz und Kohle sind Brennstoffe mit hohen Emissionsfaktoren, wobei bei Brikett das verwendete Bindemittel (früher Pech, später Bitumen) die Emission stark beeinflusst. Bis Mitte der 60er Jahre wurden in den untersuchten Regionen (Ruhrgebiet und

Norddeutschland) noch überwiegend Einzelöfen mit Kohle als Brennstoff benutzt (WICHMANN et al., 1991).

Als zweiter Innenraumfaktor für den Lungenkrebs ist das Passivrauchen zu nennen. Da in deutlich mehr als der Hälfte aller deutschen Haushalte geraucht wurde und geraucht wird, ist auch ein großer Teil der Nichtraucher gegenüber den kanzerogenen Stoffen, darunter PAH und N-Nitroso-Verbindungen, im Nebenstromrauch exponiert. Im Mikrozensus 1989 wurde für 22,4% der Befragten eine Passivrauchexposition festgestellt, wobei insbesondere jedes zweite Kind dem Rauch von Familienmitgliedern ausgesetzt war (DER BUNDESMINISTER FÜR GESUNDHEIT, 1991). In der von uns untersuchten Altersgruppe der 50–75jährigen waren 10–20% der männlichen Nichtraucher und 20–30% der weiblichen Nichtraucher in der Wohnung gegenüber Tabakrauch exponiert (WICHMANN et al., 1991). Hinzu kommt das Passivrauchen am Arbeitsplatz und in der Freizeit. PERSHAGEN (1989) schätzt das relative Lungenkrebsrisiko aufgrund einer Metaanalyse von etwa 20 epidemiologischen Studien auf 1,43 (95% C.I.: 1,27; 1,60), wenn ein Nichtraucher dem Nebenstromrauch eines Rauchers mit 30 packyears ausgesetzt ist.

Tabelle 4 gibt eine Übersicht über die tägliche Zufuhr der relevanten Kanzerogene, soweit sie anhand der heutigen Datenlage beurteilt werden kann.

Neben den chemischen Kanzerogenen kann im Innenraum die Strahlenbelastung für das Lungenepithel durch Alpha-Strahlung eine Rolle spielen, die beim Zerfall von Radon und Radontöchtern freigesetzt wird. Nach einer älteren Studie wird in ca. 1% der Wohnungen der westlichen Bundesländer die ‚obere Grenze des Normbereichs‘ von 250 Bq/m³ überschritten (SSK, 1992). In einer laufenden Studie in den Radon-Belastungsgebieten der alten und der neuen Bundesländer überschreiten ca. 3% der Meßwerte dieser Grenze (WICHMANN et al., 1993).

Abbildung 2: Zeitliche Entwicklung der mittleren Immissionsbelastung durch Benzo(a)pyren (BaP) als Indikator für polyzyklische aromatische Kohlenwasserstoffe (PAH) in Nordrhein-Westfalen (aus WICHMANN et al., 1991)

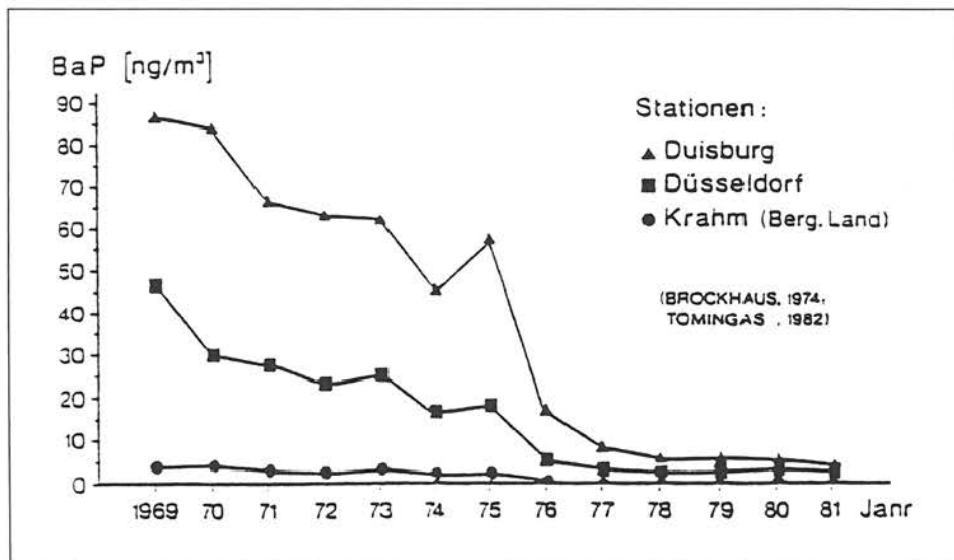


Tabelle 4: Tägliche Zufuhr von ausgewählten Kanzerogenen (nach LAI (1992))

Zufuhr	Dieselsruß µg/d	PAH ng/d	Cadmium ng/d	Arsen µg/d	Asbest Fasern/d
<i>inhalativ</i>					
Außenluft (2 m ³ /d) + Innenluft (18 m ³ /d)					
ländlich	< 30				< 2000
städtisch	< 200				< 4000
emittentennah	< 500				< 6600
Außenluft (2 bzw. 20 m ³ /d)*					
ländlich		1–21	10–30	< 0.1	
städtisch		2–12	40–100	0.1–0.4	
emittentennah		6–100	100–600	0.2–0.6	
Innenluft (18 m ³ /d)					
ländlich		5–10			
städtisch		10–50			
emittentennah		30–450			
Raucherhaushalt		110			
Rauchen (20 Zig./d)		400	2000–4000		
<i>oral</i>					
Trinkwasser (2 l/d)		4	700	< 10	–20000
Nahrung		2–500	30000**)	100–150*)	

*) fehlen Angaben zum Innenraum, werden 20 m³ für Außenluft angenommen

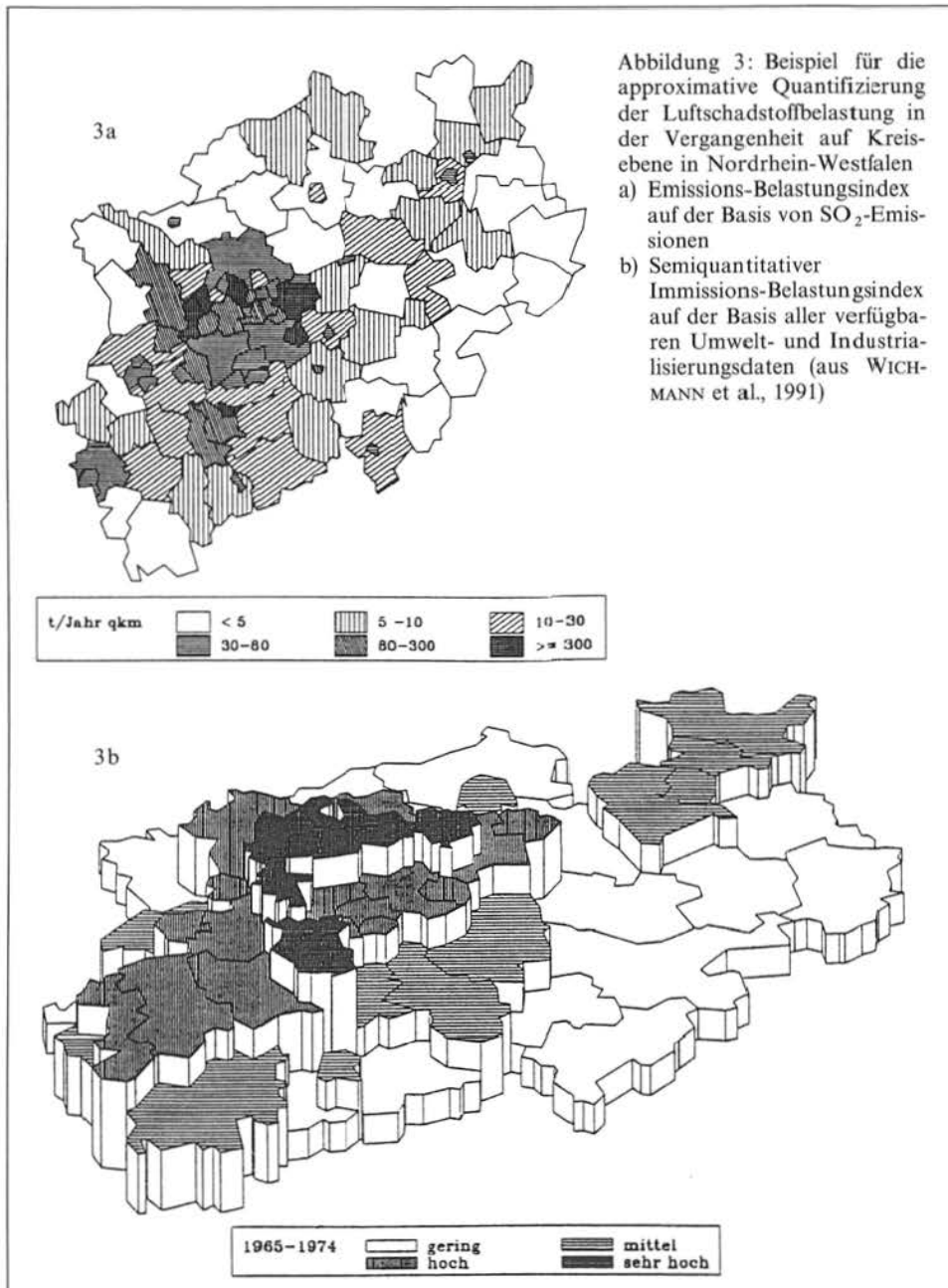
***) bei oraler Aufnahme andere Bindungsformen und Wirkungsmechanismen möglich

5. Belastungsindizes zur Expositionsabschätzung

Ein vereinfachtes und robustes Verfahren zur Expositionsabschätzung gegenüber Luftschadstoffen liefert die Ermittlung der Lebensjahre in Gebieten mit hoher Luftverunreinigung. Für die Kreise und kreisfreien Städte in Nordrhein-Westfalen und Norddeutschland (Bremen, Hamburg, Niedersachsen) wurden auf der Grundlage von Emissionsdaten sowie unter Berücksichtigung weiterer Datenquellen in WICHMANN et al. (1991) zwei Belastungsindizes vorgeschlagen. Die lebenslange Belastung einer Person ergibt sich dann aus der Summe der über die Lebensjahre gewichteten Einstufung des Wohnortes in die jeweiligen Belastungskategorien.

Da hinreichende Meßdaten zu kanzerogenen Luftverunreinigungen fehlen, wurde ersatzweise die Belastung mit Schwefeldioxid als Indikator für hohe Luftverunreinigungen bewertet. Da auch hier keine räumlich und zeitlich lückenlosen Immissionsdaten vorliegen, wurde aus vorhandenen Emissionsdaten für SO₂ für die Emittentengruppen Industrie, Verkehr, Hausbrand und Kleingewerbe ein ‚Emissionsindex‘ gebildet. Die Herleitung dieses Maßes erfolgt in 5-Jahres-Intervallen ab 1955. Es wurde eine Einteilung der Orte in 6 Belastungskategorien von unter 5 t/km² SO₂ bis über 300 t SO₂/km² für jedes Zeitintervall getroffen (Abbildung 3a).

Wegen der begrenzten Aussagekraft von Emissionsdaten und der Vielzahl von Informationen, die für die Entwicklung von Luftverunreinigungen zur Verfügung stehen, sich jedoch nicht algorithmisch umsetzen lassen, wurde weiterhin ein ‚semiquantitativer Umweltindex‘ entwickelt, der sowohl Emissions- und Immissionsdaten als auch Informationen über die Entwicklung der Industrialisierung, der Energieerzeugung, des Berg-



baus und anderer Branchen berücksichtigt. Danach werden 5 Regionalklassen gebildet und deren Entwicklung in Dekaden mittels 8 Stufen beurteilt (Abbildung 3b). GRÜGER und SCHÄFER (1990) haben die lufthygienische Situation im Rahmen einer regionalstatistischen Untersuchung im Saarland ebenfalls anhand von kleinräumig gegliederten Emissionsdaten der Schadstoffgruppen CO, SO₂, NO_x, Kohlenwasserstoffe

und Stäube beurteilt. Unter Einbeziehung der meteorologischen und orographischen Situation wurden vereinfachte Ausbreitungsrechnungen durchgeführt, um Immissionswerte zu schätzen. Für jede Gemeinde des Saarlandes wurden mittlere Immissionsbelastungen ermittelt, gewichtet nach Siedlungsdichte. Durch Trendrechnungen wurde eine Abschätzung der Luftbelastung bis zum Anfang der 60er Jahre getroffen.

6. Ökologische Untersuchungen zum Risikofaktor Luftverunreinigung

Ausgehend von dem in Industrienationen vorhandenen Stadt-Land-Gefälle in der Lungenkrebsmortalität wurden verschiedene ökologische Analysen mit regionalaggregierten Daten durchgeführt, um einen Bezug zu den dort vorliegenden Luftverunreinigungen herzustellen. Dieser Ansatz war naheliegend, da fast überall eine korrespondierende räumliche Verteilung der Luftschadstoffe vorhanden ist. Dem überwiegenden Teil der Studien mangelte es jedoch an geeigneten Daten zur Beurteilung der Luftbelastung. Auch die Qualität der Daten zum Rauchverhalten und zur Berufsbelastung ist häufig unzureichend. Wichtiger jedoch als die Frage der Datenqualität sind die methodischen Bedenken, ob bei dem dominierenden Einfluß des Rauchens vergleichsweise kleine andere Einflußgrößen auf der Grundlage aggregierter Daten überhaupt verlässlich quantifiziert werden können (BRENNER et al., 1992).

Eine der ersten umfangreichen Untersuchungen wurde von HAENSZEL et al. (1962) durchgeführt. Für eine Stichprobe von 10% der Krebstodesfälle der weißen Bevölkerung im Jahr 1958 wurden Daten zum Rauchverhalten aus dem Nationalen Gesundheitsurvey der USA herangezogen. Nach Berücksichtigung des Rauchens verblieb ein Einfluß, der besonders bei männlichen Rauchern vorhanden war und der der Luftverunreinigung zugeschrieben wurde.

In der ländlichen Region von Xuan Wei, China, untersuchten MUMFORD et al. (1987) den Zusammenhang von Innenraumbelastungen durch Kohleverbrennung und der Lungenkrebssterblichkeit, die zu der höchsten von China zählt. Die Assoziation zwischen den Mortalitätsraten der einzelnen Kommunen und dem prozentualen Anteil von Haushalten, die stark rauchende Kohle als Brennstoff verwenden, ist stärker als der Zusammenhang mit dem Tabakrauchen. Zusätzlich wurden Innenraummessungen durchgeführt und Staubextrakte toxikologisch untersucht. Im Ames-Test zeigen die Proben aus Kommunen mit hoher Lungenkrebssterblichkeit im Vergleich zu Extrakten anderer Brennmaterialien die höchste Mutagenität. Die B(a)P-Innenraumkonzentrationen lagen etwa 1000fach über den für Innenräume in den USA üblichen Werten (CHUANG et al., 1991).

Von Bedeutung für die Bundesrepublik Deutschland ist die Studie von ULMER (1982). Aufgrund der Daten von 16429 Todesbescheinigungen (davon 726 mit der Diagnose Bronchialkarzinom) aus Bochum und den ländlichen Vergleichsgebieten Brilon und Meschede sowie weiteren anamnestic Angaben von Angehörigen und einer Bevölkerungsbefragung zum Rauchverhalten wird in dieser Untersuchung der Schluß gezogen, daß der ermittelte Stadt-Land-Unterschied von 1,57 im wesentlichen auf Unterschiede in den Rauchgewohnheiten und zum Teil auf Unterschiede in der beruflichen Belastung zurückzuführen ist. Altersstandardisierte Raten wurden jedoch nicht berechnet.

Die Autoren haben eine ökologische Analyse auf der Basis der Regierungsbezirke der Bundesrepublik Deutschland durchgeführt (Abbildung 1, WICHMANN et al., 1991). Eine feinere Auflösung konnte aus Datenschutzgründen für die hierzu verwendeten Daten aus dem Mikrozensus nicht erzielt werden. Es zeigte sich für die Faktoren Rauchen und Luftverunreinigung bei beiden Geschlechtern eine signifikante Korrelation zur Lungenkrebssterblichkeit. Wurden in einer multivariaten Regressionsanalyse Rauchen und Beruf

kontrolliert, so wurde der verbleibende Beitrag der Luftverunreinigungen deutlich abgesenkt.

Ein Beispiel für eine sehr differenzierte Regionalanalyse liegt für das Saarland vor. Einen Schwerpunkt dieser Untersuchung bildete eine ökologische Studie über den Zusammenhang zwischen Lungenkrebs und Luftverunreinigungen (GRÜGER und SCHÄFER, 1990). Hierzu wurden die Inzidenzraten auf der Ebene der Gemeinden anhand der Daten des Saarländischen Krebsregisters für den Zeitraum 1975 bis 1984 ermittelt (GRÜGER et al., 1991). Die lufthygienische Situation wurde über geschätzte Immissionsdaten anhand eines feinräumigen Emissionskatasters bewertet. Weiterhin erfolgte anhand einer faktoren analytischen Untersuchung regionalstatistischer Merkmale eine Bewertung des Industrialisierungsgrades, der Urbanität und der sozialökonomischen Situation. Das berufliche Risiko wurde über eine zweistufige Variable anhand des Anteils von Risikoberufen nach der Berufs- und Volkszählung 1970 beurteilt. Zur Charakterisierung der Rauchgewohnheiten erfolgte eine Telefonumfrage, wobei der Anteil derzeitiger Zigarettenraucher in den Gemeinden in die statistische Auswertung (Pfadanalyse und Regressionsanalyse) aufgenommen wurde. Als Ergebnis fanden die Autoren, daß die lufthygienische Situation auch nach Adjustierung für Rauchen und Beruf einen nachweislichen Einfluß auf die Lungenkrebsinzidenz hatte. Jedoch gilt auch für diese sorgfältige Studie, daß die methodischen Schwächen der ökologischen Analyse bei der Interpretation berücksichtigt werden müssen. Eine verlässliche quantitative Risikoschätzung lassen diese sicherlich nicht zu (BRENNER et al., 1992, WICHMANN & KREIENBROCK, 1992).

7. Analytische epidemiologische Studien zur Abschätzung des Risikos von Luftverunreinigungen

Valide Risikoschätzungen sind letztlich nur durch epidemiologische Studien mit Individualdaten über die Exposition gegenüber den vermuteten Risiken möglich. Wegen des relativ geringen Einflusses von Luftverunreinigungen auf das Risiko an Lungenkrebs zu erkranken ist jedoch ein erheblicher Studienumfang notwendig, um diesen Einfluß statistisch sichern zu können.

Follow-up-Studien mit dem vorrangigen Ziel, den Einfluß der Luftverunreinigungen zu schätzen, wurden nicht durchgeführt. Es gibt jedoch sehr umfangreiche Untersuchungen, die primär den Einfluß des Rauchens zum Gegenstand haben und dabei eine Unterteilung nach Großstädten und ländlichen Gebieten vornehmen, so daß auch Aussagen über die Größe eines „urbanen Faktors“ möglich sind. Die umfangreichste Studie dieser Art wurde an über einer Million Männer und Frauen aus 25 Bundesstaaten der USA durchgeführt. HAMMOND und GARFINKEL (1980) analysierten die Daten in Hinblick auf einen möglichen Einfluß von Luftverunreinigungen. Dabei kommen die Autoren zu dem Schluß, daß es keinen überzeugenden Beweis dafür gibt, daß die allgemeine Luftverunreinigung das Lungenkrebsrisiko erhöht.

Eine der wenigen Untersuchungen mit dem vorrangigen Ziel der Schätzung des Einflusses der Luftverunreinigungen auf die Entstehung von Lungenkrebs ist die Fall-Kontroll-Studie von VENA (1982). Sie wurde in den Jahren 1957 bis 1965 an 417 männlichen Lungenkrebspatienten und 752 Hospitalkontrollen in der Region Erie County durchgeführt. Das Rauchverhalten wurde über Packungsjahre (packyears), die berufliche Belastung durch die Beschäftigungsdauer in Berufen mit einem hohen Lungenkrebsrisiko quantifiziert. Zur Bewertung der Luftverunreinigung wurden Schwebstaubdaten von 21 Meßstationen des Jahres 1960 zugrunde gelegt. Weiterhin wurden lokale Emittenten identifiziert und bewertet. Nach diesen Angaben erfolgte eine Klassifizierung der Wohnorte in hoch, mittel und niedrig belastete Gebiete. Zur Quantifizierung des Einflusses der Luft-

verunreinigung wurden die Lebensjahre in den stark und mittel belasteten Regionen verwendet. Ein Matching nach dem Alter erfolgte jedoch nicht; Unterschiede in der Altersverteilung von Fällen und Kontrollen sind vorhanden. Das unkorrigierte relative Lungenkrebsrisiko stieg mit der Dauer des Wohnens in belasteten Regionen an und erreichte einen Wert von 1.56 für die länger als 50 Jahre in Belastungsgebieten Wohnenden. Nach Adjustierung für Alter, Rauchen und Beruf sank es in dieser Gruppe auf 1,26 ab. Für die kürzer in Belastungsgebieten Wohnenden war das Risiko nicht erhöht.

In einer Fall-Kontroll-Studie in Krakau, Polen (JEDRYCHOWSKI et al., 1990), wurden im Zeitraum von 1980 bis 1985 Angehörige von 901 an Lungenkrebs verstorbenen Männern und 198 Frauen über Rauchgewohnheiten, Berufe und Wohnorte der Fälle befragt. Als Kontrollen dienten 875 Männer bzw. 198 Frauen der gleichen Region, deren Todesursache keine Erkrankung der Atemwege war. Bezüglich des Faktors Luftverunreinigung wurden drei Gruppen gebildet. Als Indikatoren dienten Schwebstaub und SO₂. Dabei wurde für Männer im Vergleich von hoch zu niedrig belastet ein Odds Ratio von 1,42 ($p < 0,05$) ermittelt. Daraus folgern die Autoren, daß 5% bis 10% des Risikos, an Bronchialkarzinom zu erkranken, der allgemeinen Luftverunreinigung zugeschrieben werden kann. Diese Aussage wird allerdings dadurch relativiert, daß beim weiblichen Geschlecht kein entsprechender Trend erkennbar ist.

Die Bremen-Düsseldorfer Fall-Kontroll-Studie der Autoren (GREISER, 1984, WICHMANN et al., 1991, JÖCKEL et al., 1992) hatte einen vergleichsweise geringen Umfang von 194 Fällen (146 Männer, 48 Frauen) und der doppelten Zahl von Kontrollen. Es war auch nicht das Ziel, eine Abschätzung des Risikos von Luftverunreinigungen vorzunehmen, sondern als Pilotprojekt die Durchführbarkeit einer großen Fall-Kontroll-Studie zum Lungenkrebs in der Bundesrepublik Deutschland zu überprüfen und das notwendige Instrumentarium vorzubereiten und zu testen. Daher wurde der Schwerpunkt auf die Entwicklung methodischer Ansätze zur Quantifizierung der Exposition gegenüber den Risikofaktoren Luftverunreinigung, Beruf und Rauchen gelegt. Verwendet man die entwickelten Umweltindizes für eine orientierende Risikoabschätzung, dann liefert der semiquantitative Index, welcher die lebenslange kumulative Belastung berücksichtigt, leicht erhöhte Risikowerte (Tabelle 5), während der Index, der sich auf Emissionsdaten stützt, keine Risikoerhöhung ergibt. In beiden Rechnungen ist aufgrund der geringen Fallzahlen der Vertrauensbereich der Schätzung so groß, daß sich hieraus keine Aussagen über die Größe des Effektes ableiten lassen. Beachtung verdient allerdings die Tatsache,

Tabelle 5: Relative Risiken (Odds Ratios) für den Einfluß der allgemeinen Luftverunreinigung auf die Entstehung von Lungenkrebs (WICHMANN et al., 1991, JÖCKEL et al., 1992)

	Luftverunreinigung	Fälle	Kontrollen	Odds Ratio adjustiert nach	
				Alter (95% Konfidenzintervall)	Alter, Rauchen und Beruf (95% Konfidenzintervall)
Emissionsindex	niedrig	41	85	1,0	1,0
	hoch	39	75	1,10 (0,60; 2,04)	1,01 (0,53; 1,91)
Semiquant. Umwelt-Index	niedrig	36	80	1,0	1,0
	hoch	44	80	1,25 (0,71; 2,19)	1,15 (0,63; 2,11)

daß bei einer Beschränkung auf Expositionen vor 1966 das Odds Ratio 2,35 betrug, das nach Adjustierung für Rauchen und berufliche Belastung auf 1,82 (nicht signifikant) gesenkt wurde (JÖCKEL et al., 1992).

Aufgrund der großen Konfidenzintervalle erscheint es nicht sinnvoll, anhand dieser Ergebnisse ein Attributivrisiko zu berechnen.

Neben der Frage des Einflusses der allgemeinen Luftverschmutzung gibt es Hinweise auf Lungenkrebshäufungen im Umkreis lokaler Schadstoffemittenten, z. B. von arsenemittierenden Betrieben und Kupferschmelzen in den USA (BROWN et al., 1984) und Schweden (PERSHAGEN, 1985) sowie in der Nachbarschaft schottischer Gießereien (LLOYD et al., 1985).

Diskussion

Lungenkrebs ist die häufigste Krebstodesursache beim Mann und zeigt beim weiblichen Geschlecht einen ausgeprägten ansteigenden Trend. Deshalb bedeuten auch geringe Risiken relevante Erkrankungszahlen. CEDERLÖF et al. (1978) schätzen, daß etwa 5% bis 10% der Lungenkrebsfälle in stark belasteten Gebieten der Luftverunreinigung zuzuschreiben sind. Das relative Risiko des Einflusses von Luftverunreinigungen dürfte auch in Regionen der Bundesrepublik, die früher stark belastet waren, deutlich unter 1,5 liegen. Legt man die derzeitigen Konzentrationen kanzerogener Luftschadstoffe für die alten Bundesländer sowie die unit risk-Annahmen der Arbeitsgruppe „Krebsrisiko durch Luftverunreinigungen“ des Länderausschusses für Immissionsschutz (LAI) (1992) zugrunde, dann ergibt sich, daß ca. 2% aller Lungenkrebsfälle im Westen der Bundesrepublik den kanzerogenen Schadstoffen der Außenluft zuschreibbar sind. Damit dürften die durch Luftverunreinigung verursachten zusätzlichen Lungenkrebstodesfälle in den alten Bundesländern in der gleichen Größenordnung liegen wie die dem Passivrauchen zuschreibbaren Fälle (FISCHER, 1992, BECHER, WAHRENDORF, in Druck). Eine Übertragung von Risikoschätzungen aus den USA ist problematisch, da Deutschland historisch gewachsene Ballungsräume mit hoher Siedlungsdichte hat. In deutschen Großstädten wurden aufgrund der traditionellen Kohleheizungen in der Vergangenheit weitaus höhere PAH-Konzentrationen gemessen. Auch spielen im deutschen Kraftfahrzeugverkehr Dieselfahrzeuge eine größere Rolle, und Katalysatoren haben später Einsatz gefunden.

Ein geeignetes Maß für die Bewertung des kanzerogenen Potentials von Luftverunreinigungen zu finden, erweist sich als schwierig: Es werden vorwiegend Stoffe wie SO_2 und Schwebstaub gemessen, die zwar die mengenmäßig bedeutendsten Bestandteile an Außenluftverunreinigungen darstellen, jedoch selbst nicht kanzerogen sind und daher nur als Korrelate angesehen werden können. Meßnetze sind im allgemeinen nicht flächendeckend vorhanden, und für die Vergangenheit liegen nur unzureichend Meßwerte vor.

Auch kann nicht erwartet werden, mit Hilfe einfacher Ansätze einen ausreichend verallgemeinerbaren Einfluß der Luftverunreinigungen auf die Entstehung des Lungenkrebses zu finden, da sowohl weltweit als auch regional die Expositionsbedingungen erheblich variieren. Mitteleuropa weist stark verdichtete Ballungsräume schon seit Beginn des Jahrhunderts auf. In den westlichen Industrienationen fanden in den 60er und 70er Jahren gravierende Umstellungen im Energiesektor bei den verwendeten Brennstoffen statt, und moderne Industrieanlagen wurden aufgrund verstärkter Auflagen zunehmend emissionsärmer. Dadurch konnte in den letzten beiden Jahrzehnten ein starker Rückgang der Emission krebserzeugender Stoffe durch Industrie, Kleingewerbe und Hausbrand erreicht werden. Andererseits nahm die Freisetzung von Kanzerogenen durch den Straßenverkehr zu. Diese Entwicklung verlief in den Ostblockländern stark verzögert,

Tabelle 6: Zuschreibbarer Risikoanteil (attributable risk percent) bei wichtigen Risikofaktoren des Bronchialkarzinoms

Risikofaktor	USA ¹⁾	Bundesrepublik Deutschland ²⁾
Luftverunreinigungen ³⁾	2% ¹⁰⁾	1,7–2,3% ⁴⁾
Radon in Wohnungen		4–12% ⁵⁾
Passivrauchen		4,3% ⁶⁾
berufliche Exposition		
– Männer	15%	7–12% ⁷⁾
– Frauen	5%	– ⁸⁾
Rauchen		
– Männer	80–90%	89% ⁹⁾
– Frauen	30–60%	– ⁸⁾

¹⁾ DOLL, PETO (1981)

²⁾ alte Bundesländer

³⁾ in stark belasteten Gebieten 5–10% (CEDERLÖF et al., 1978)

⁴⁾ nach LAI (1992) konservative Abschätzung

⁵⁾ STRAHLENSCHUTZKOMMISSION (1987), JACOBI (1993)

⁶⁾ JÖCKEL (1991)

⁷⁾ JÖCKEL et al. (1992b)

⁸⁾ in den vorliegenden Studien sind für Frauen aufgrund geringer Fallzahlen keine verlässlichen Angaben möglich

⁹⁾ nach JÖCKEL et al. (1992)

¹⁰⁾ alle Krebsformen durch Luft- und Wasserverunreinigungen

so daß dort auch noch bis zum Beginn der 90er Jahre Schadstoffkonzentrationen vorlagen wie vor 20 bis 30 Jahren im westlichen Europa.

Eine zusammenfassende Darstellung des Beitrages der Luftverunreinigung, der Radonbelastung und anderer wichtiger Risikofaktoren des Bronchialkarzinoms ist in Tabelle 6 versucht. Danach ergibt sich sowohl für die USA (auf der Grundlage epidemiologischer Studien) als auch für die Bundesrepublik (auf der Grundlage von Risikoextrapolationen) ein Attributivrisiko von etwa 2% für die Luftverunreinigung. Dies ist kein Widerspruch zur Aussage von CEDERLÖF et al. (1978), daß in stark belasteten Gebieten 5 bis 10% der Lungenkrebskrankungen den Luftschadstoffen zuzuschreiben sind.

Zum Radon in Wohnungen liegen Risikoextrapolationen vor, die epidemiologischen Daten sind bisher nicht ausreichend. Das Attributivrisiko wird auf 4 bis 12% geschätzt. Hierbei ist allerdings zu bedenken, daß Radon den Hauptbeitrag zur natürlichen Strahlenbelastung darstellt und das ‚Minderungspotential‘ durch Sanierung hochbelasteter Häuser erheblich niedriger ist.

Die berufliche Belastung trägt in den USA zu 15% beim Mann und zu 5% bei der Frau zum Lungenkrebsrisiko bei. Die Schätzung für Männer in der Bundesrepublik (alt) liegt mit 7% bis 12% in der gleichen Größenordnung. Für die deutschen Frauen ist die Datenlage nicht ausreichend für eine entsprechende Abschätzung.

Das Rauchen schließlich wird in den USA für 80 bis 90% der Bronchialkarzinome beim Mann und für 30 bis 60% bei der Frau verantwortlich gemacht. In der Bundesrepublik dürfte die Situation ähnlich sein. Die Abschätzung des Attributivrisikos anhand der Daten der Fall-Kontroll-Studie von WICHMANN et al. (1991) liegt mit 84% für Männer in der gleichen Größenordnung. Für Frauen sind auch hier die Fallzahlen zu gering, um eine zuverlässige Aussage treffen zu können.

Abschließend sei darauf hingewiesen, daß die Prozentzahlen in Tabelle 6 unter Zugrundelegung des Konzeptes einfacher Attributivrisiken berechnet wurden. Insbesondere ist es wegen der Wechselwirkung zwischen den Kanzerogenen möglich, daß die Summe der Risikoanteile größer als 100% wird.

Literatur

- BECHER, H. (1992): Epidemiologie des Bronchialkarzinoms. *MMW* **134**, 569–574.
- BECHER, H., WAHRENDORF, J. (1994): Passivrauchen und Lungenkrebs – Gegenwärtiger epidemiologischer Kenntnisstand und Abschätzung des Effektes in der Bundesrepublik Deutschland. *Deutsches Ärzteblatt* (in press).
- BECKER, N., FRENTZEL-BEYME, R., WAGNER, G. (1984): Krebsatlas der Bundesrepublik Deutschland. Springer Verlag, Berlin, Heidelberg, New York, Tokyo.
- BORMANN, C., HEINEMANN, L., HOELTZ, J. (Hrsg.) (1991): Kardiovaskuläre Risiken in Deutschland-Ost und -West. Infratest München.
- BRENNER, H., SAVITZ, D. A., JÖCKEL, K.-H., GREENLAND, S. (1992): The Effect of Nondifferential Misclassification in Ecological Studies. *Am. J. Epid.* **135**, 85–95.
- BROWN, L. M., POTTERN, L. M., BLOT, W. J. (1984): Lung Cancer in Relation to Environmental Pollutants Emitted from Industrial Sources. *Env. Research* **34**, 250–261.
- BRÜSKE-HOHLFELD, I., KREIENBROCK, L., HEINRICH, J., KREUZER, M., SCHNELLER, H., BAUDREX, I., WEBER, J., WIESNER, B., WICHMANN, H. E. (1993): Beschreibung von 93 ehemaligen Wismut-Beschäftigten, die 1992 an Lungenkrebs erkrankten. Tagung Epidemiologie. Berlin 15.–16. 3. 1993 (Vortrag).
- BUNDESFORSCHUNGSANSTALT FÜR LANDESKUNDE UND RAUMORDNUNG (BfLR) (Hrsg.) (1991): Arbeitsbericht 1990. Bonn.
- DER BUNDESMINISTER FÜR GESUNDHEIT (Hrsg.) (1991): Daten des Gesundheitswesens. Ausgabe 1991. Schriftenreihe des Bundesministeriums für Gesundheit **3**, Nomos Verlagsgesellschaft Baden-Baden.
- CAPORASO, N. (1992): Genetic susceptibility and common malignancies: promise and pitfalls of interdisciplinary studies. 83rd Annual Meeting of the American Association for Cancer Research. San Diego (USA) 20.–23. 5. 1992.
- CEDERLÖF, R., DOLL, R., FOWLER, B. et al. (1978): Air Pollution and Cancer: Risk Assessment Methodology and Epidemiology and Epidemiological Evidence. *Environm. Health Persp.* **22**, 1–12.
- CHUANG, J. C., CAO, S. R., XIAN, Y. L. et al. (1990): Chemical Characterization of Indoor Air of Homes from Communes in Xuan Wei, China, with high Lung Cancer Mortality Rate. Proceedings of the 5th International Conference on Indoor Air and Climate, Toronto 29. 7.–3. 8. 1990, 103–108.
- DOLL, R. (1953): Mortality from Lung Cancer among Nonsmokers. *Br. J. Cancer* **7**, 303–373.
- DOLL, R., PETO, R. (1976): Mortality in Relation to Smoking: 20 Years' Observations on Male British Doctors. *Brit. Med. J.* **273**, 1525–1535.
- DOLL, R., PETO, R. (1981): The Causes of Cancer – Quantitative Estimates of Avoidable Risks in the United States Today. *JNCI* **66**, 1191–1308.
- DOLL, R. (1992): Urban Life and Cancer, An Introduction. In: Behörde für Arbeit, Gesundheit und Soziales der Freien und Hansestadt Hamburg (Hrsg.): Urban Life and Cancer – A Selection of Essays – Contributions to the Annual Meeting of the IARC. Hamburg 13.–15. 8. 1990, 4–22.
- FISCHER, M. (1992): Kanzerogene Luftschadstoffe und ihre Bedeutung für die Krebsmortalität. *Bundesgesundheitsblatt* **35**, 184–189.
- FRENTZEL-BEYME, R., LEUTNER, R., WAGNER, G., WIEBELT, H. (1979): Cancer Atlas of the Federal Republic of Germany. Springer Verlag, Berlin, Heidelberg, New York.
- GONZALES, F. J., SKODA, R. C., KIMURA, S., UMENO, M., ZANGER, U. M., NEBERT, D. W., GELBOIN, H. V., HARDWICK, J. P., MEYER, U. A. (1988): Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. *Nature* **331**, 442–446.
- GREISER, E., JÖCKEL, K. H., MOLIK, B., TIMM, J., WICHMANN, H. E. (1984): Überlegungen zu einem Studiendesign zur Erfassung der Wirkung von Luftverschmutzung, Arbeitsplatzexposition und Rauchen auf die Entstehung des Bronchialkarzinoms. Unveröffentlichtes Manuskript. Bremen.

- GRÜGER, J., PEULING, G., SCHÄFER, T., TENHAGEN, P. (1991): Krebsatlas Saarland – Atlas der Morbidität, Mortalität und Prävalenz bösartiger Neubildungen in den Gemeinden des Saarlandes für den Zeitraum 1975 bis 1984. UBA (Hrsg.) Texte 91/18. Berlin.
- GRÜGER, J., SCHÄFER, T. (1990): Regionalstatistische Analyse Saarland – Luftverunreinigungen und Lungenkrebs. Tagung „Umwelt und Krebs“ der Arbeitsgemeinschaften der Großforschungseinrichtungen 13.–14. 12. 1990, Bonn.
- HAENSZEL, W., LOVELAND, D. B., SIRKEN, M. (1962): Lung Cancer Mortality as Related to Residence and Smoking Histories. I. White Males. *JNCI* **28**, 947–1001.
- HAMMOND, E. C. (1966): Smoking in Relation to the Death Rates of One Million Men and Women. *Natl. Cancer Inst. Monogr.* **19**, 127–204.
- HAMMOND, E. C., GARFINKEL, L. (1980): General Air Pollution and Cancer in the United States. *Prev. Med.* **9**, 206–211.
- HAMMOND, E. C., SEIDMAN, H. (1980): Smoking and Cancer in the United States. *Prev. Med.* **9**, 169–173.
- HOFFMEISTER, H., JUNGE, B. (1984): Was sagt die Mortalitätsstatistik aus? *Z. Allg. Med.* 240–252.
- HUGHES, J. (1994): Environmental exposure: evaluating potential cancer risks. I. Internationaler Kongreß für Umweltmedizin. Duisburg 23.–26. 2. 1994.
- IARC/WHO (1986): IARC Monographs Vol. 28. Tobacco Smoking. IARC Lyon.
- IARC/WHO (1992): Cancer incidence in five continents. Vol. VI. IARC Scientific Publications 120. IARC Lyon.
- JACOBI, W. (1993): Radon und Lungenkrebs: Historie und aktuelle Probleme. In: Rundgespräche der Kommission für Ökologie, Bd. 7, S. 125–137, Verlag Dr. Friedrich Pfeil, München.
- JEDRYCHOWSKI, W., BECHER, H., WAHRENDORF, J., BASA-CIERPIALEK, Z. (1990): A case-control study of lung cancer with special reference to the effect of air pollution in Poland. *J. Epidemiol. Community Health* **44**, 114–120.
- JÖCKEL, K. H. (1991): Passivrauchen – Bewertung der epidemiologischen Befunde. In: VDI-Berichte Nr. 888.
- JÖCKEL, K. H., AHRENS, W., WICHMANN, H. E. et al. (1992): Occupational and Environmental Hazards Associated with Lung Cancer. *Int. J. Epidemiol.* **21**, 202–213.
- JÖCKEL, K. H., AHRENS, W., BOLM-AUDORFF, U. (1992b): Lungenkrebsrisiko am Arbeitsplatz: Die Public Health Perspektive. In: Behörde für Arbeit, Gesundheit und Soziales der Freien und Hansestadt Hamburg (Hrsg.): *Urban Life and Cancer – A Selection of Essays – Contributions to the Annual Meeting of the IARC.* Hamburg 13.–15. 8. 1990.
- KREIENBROCK, L., WICHMANN, H. E., GERKEN, M., HEINRICH, J., GOETZE, H.-J., KREUZER, M., KELLER, G. (1992): The German Radon Project – Feasibility of Methods and First Results. *Radiation Protection Dosimetry* **45**, 643–649.
- LAI (Länderausschuß für Immissionsschutz) (Hrsg.) (1992): Krebsrisiko durch Luftverunreinigungen. MURL Düsseldorf.
- HELMAN, L., THILE, C. (1991): New insights into the causes of cancer. *Pediatr. Clin. North Am.* **38**, 201–221.
- LLOYD, O. L. et al. (1985): Raised mortality from lung cancer and high sex ratios of birth associated with industrial pollution. *Br. J. Ind. Med.* **42**, 475–480.
- LUBIN, J. H., BLOT, W. J., BERRINO, F. et al. (1984): Patterns of Lung Cancer Risk According to Type of Cigarette Smoked. *Int. J. Cancer* **33**, 569–576.
- MEHNERT, W. H., LAUSSMANN, D., ARNDT, D. (1992): Lungenkrebsinzidenz in den Südbezirken der ehemaligen DDR. *ST-Berichte* **3**, 88–104.
- MOHR, U., SCHMÄHL, D., TOMATIS, L. (eds.) (1977): *Air Pollution and Cancer in Man.* IARC Scientific Publications No. 16. Lyon.
- MORTON, W. E., TREYVE, E. L. (1982): Histologic Differences in Occupational Risks of Lung Cancer Incidence. *Am. J. Ind. Med.* **3**, 441–457.
- MÜLLER, F. H. (1939): Tabakmißbrauch und Lungenkrebs. *Z. Krebsforschung* **49**, 57–85.
- MÜLLER, W., BOCTER, N. (1989): Beitrag zur Abschätzung der Aussagekraft der amtlichen Todesursachenstatistik. Schriftenreihe des BMJFG **253**, Kohlhammer, Stuttgart, Berlin, Köln.
- MUIR, C., WATERHOUSE, J. MACK, T. et al. (eds.) (1987): *Cancer Incidence in Five Continents, Vol. V.* IARC Scientific Publications No. 88, Lyon.
- MUMFORD, J. L., HE, X. Z., CHAPMAN, R. S. et al. (1987): Lung Cancer and Indoor Air Pollution in Xuan Wei, China. *Science* **235**, 217–220.

- NEBERT, D. W., PUGA, A. (1992): Human drug metabolism polymorphisms: effects on cancer risk. 83rd Annual Meeting of the American Association for Cancer Research. San Diego (USA) 20. – 23. 5. 1992.
- PERSHAGEN, G. (1985): Lung cancer mortality among men living near an arsenic emitting smelter. *Am. J. Epid.* **122**, 684–694.
- PERSHAGEN, G. (1989): Environmental tobacco smoke and cancer. In: Mohr, U., Bates, D. V., Dungworth, D. L., Lee, P. N., McCellan, R. O., Roe, F. J. C. (eds.): *Assessment of inhalation hazards*. Springer Verlag Berlin, Heidelberg, New York, 275–284.
- PESCH, B., POTT, F. (1992): Entwicklungen der Krebssterblichkeit in Nordrhein-Westfalen zwischen 1970 und 1989. In: Gesellschaft zur Förderung der Lufthygiene und Silikoseforschung e. V. (Hrsg.): *Umwelthygiene. Jahresbericht 1991/92*. **24**. Albers Verlag, Düsseldorf, 178–225.
- POTT, F., HEINRICH, U. (1992): Staub und Staubinhaltsstoffe/Polyzyklische aromatische Kohlenwasserstoffe (PAH). Umweltschadstoffe VI-2. In: Wichmann, H. E., Schlipkötter, H.-W., Fülgraff, G. (Hrsg.): *Handbuch der Umweltmedizin*. Ecomed-Verlag, Landsberg.
- SEGI, M. (1960): Cancer Mortality for Selected Sites in 24 Countries (1950–1957). Dep. of Public Health, Tohoku University School of Medicine, Sendai (Japan).
- STATISTISCHES AMT DER DDR (Hrsg.) (1991): *Statistisches Jahrbuch der Deutschen Demokratischen Republik 1990*. Rudolf Haufe Verlag Berlin.
- STATISTISCHES BUNDESAMT (Hrsg.) (1992): Sterbefälle nach Todesursachen im früheren Bundesgebiet (Gebietsstand vor dem 3. 10. 1990). Einzelnachweis **1**, 1990. Wiesbaden.
- STRAHLENSCHUTZKOMMISSION (SSK) (1987): Strahlenexposition und mögliches Lungenkrebsrisiko durch Inhalation von Radon-Zerfallprodukten in Häusern. In: *Veröffentlichungen der Strahlenschutzkommission Band 6*, G. Fischer Verlag Stuttgart.
- TOMATIS, L. (ed.) (1990): *Air Pollution and Human Cancer*. European School of Oncology Monographs. Springer Berlin usw.
- ULMER, W. T. (1982): *Das Bronchialkarzinom im Stadt-/Landfaktor*. Epidemiologische Studie zur Abgrenzung anderer Einflußgrößen. Thieme Stuttgart.
- ULLRICH, A. (1992): Growth receptors in oncogenesis. 83rd Annual Meeting of the American Association for Cancer Research. San Diego (USA) 20. – 23. 5. 1992.
- VENA, J. (1982): Air Pollution as a Risk Faktor in Lung Cancer. *Am. J. Epidem.* **116**, 42–56.
- WHO (1993): *Indoor Air Quality: A Risk-based Approach to Health Criteria for Radon Indoors*. World Health Organisation, Regional Office for Europe, Copenhagen EUR/ICP/CEH 108 (S).
- WICHMANN, H. E., JÖCKEL, K.-H., MOLIK, B. (1991): Luftverunreinigungen und Lungenkrebsrisiko – Ergebnisse einer Pilotstudie. UBA-Berichte 7/91, Erich Schmidt Verlag Berlin.
- WICHMANN, H. E., KREIENBROCK, L. (1992): *Umweltepide miologie*. In: Wichmann, H. E., Schlipkötter, H. W., Fülgraff, G. (Hrsg.): *Handbuch der Umweltmedizin*, Kapitel III-1.2, Ecomed-Verlag Landsberg.
- WICHMANN, H. E., MOLIK, B., POTT, F., JÖCKEL, K.-H. (1986): Luftverunreinigungen und andere Risikofaktoren des Bronchialkarzinoms. In: Wichmann, H. E. (Hrsg.): *Methodische Aspekte der Umweltepide miologie*. Medizinische Informatik und Statistik. **65**, Springer-Verlag Berlin.
- WICHMANN, H. E., KREIENBROCK, L., KREUZER, M., GOETZE, H. J., HEINRICH, J., GERKEN, M. (1993): Radon und Lungenkrebs – Kenntnisstand und erste Erfahrungen mit einer Studie in der Bundesrepublik Deutschland. In: *Veröffentlichungen der Strahlenschutzkommission*, G. Fischer Verlag Stuttgart.
- WICHMANN, H. E., BRÜSKE-HOHLFELD, I. (1991): Epidemiologische Befunde zum Krebsrisiko durch Dieselmotorabgase. *VDI-Berichte* **888**, VDI-Verlag Düsseldorf, 171–209.
- WYNDER, E. (1991): Listen to nature. The challenge of lifestyle medicine. *Soz. Präventivmed.* **36**, 137–146.
- ZENTRALINSTITUT FÜR KREBSFORSCHUNG (Hrsg.) (1991): *Krebsinzidenz in der DDR 1987*. Berlin.

Anschrift des ersten Verfassers: B. Pesch, Medizinisches Institut für Umwelthygiene an der Universität Düsseldorf

Eingang des Manuskripts 22. 3. 94

Eingang der revidierten Fassung 16. 6. 94

Statistische Ansätze zur Beurteilung seltener Arzneimittelrisiken in einer Fallstudie

Norbert Holländer¹⁾, Berthold Rzany²⁾, Maja Mockenhaupt²⁾, Erwin Schöpf²⁾
und Martin Schumacher¹⁾

Zusammenfassung

Diese Arbeit stellt drei statistische Ansätze zur Beurteilung seltener Arzneimittelrisiken dar, wenn lediglich das Datenmaterial von „Fällen“ vorliegt. Die Methoden werden am Beispiel von Patienten mit verschiedenen schweren Hautreaktionen illustriert. Anhand der Berechnung arzneimittelbezogener Kennzahlen wird mit dem ersten Ansatz versucht, die Anzahl der nach Einnahme eines bestimmten Arzneimittels auftretenden unerwünschten Reaktionen mit der zugehörigen exponierten Population ins Verhältnis zu setzen. Mit der zweiten Methode, einem statistischen Modell für intermittierende Expositionen, lassen sich relative Risiken auf der Basis eines prospektiven Falldesigns schätzen. Ein Vergleich zwischen verschiedenen Diagnosekategorien kann in einem dritten Ansatz anhand eines polychotomen logistischen Regressionsmodells durchgeführt werden. Bei der Anwendung wird insbesondere die Eignung der Ansätze zur Beantwortung der vorliegenden Fragestellung und die Problematik fehlender Kontrollen diskutiert.

Schlüsselwörter

Unerwünschte Arzneimittelreaktion, arzneimittelbezogene Kennzahlen, intermittierende Exposition, Falldesign, relatives Risiko, polychotomes logistisches Regressionsmodell.

Einleitung

Seltene unerwünschte Reaktionen auf die Einnahme von Arzneimitteln können erst durch die Behandlung großer Patientenpopulationen, d. h. nach der Zulassung der Präparate, entdeckt werden. Da in der Bundesrepublik Deutschland die Meldung unerwünschter Arzneimittelwirkungen an das Bundesgesundheitsamt auf freiwilliger Basis erfolgt, muß von einer hohen Dunkelziffer nicht gemeldeter Fälle ausgegangen werden. Besonders bei schweren Erkrankungen ist aber eine intensive und möglichst vollständige Fallerfassung, wie sie beispielsweise durch das Dokumentationszentrum schwerer Hautreaktionen

¹⁾ Institut für Medizinische Biometrie und Medizinische Informatik, Albert-Ludwigs-Universität, Freiburg

²⁾ Dokumentationszentrum schwerer Hautreaktionen, Universitäts-Hautklinik der Albert-Ludwigs-Universität, Freiburg

(dZh) der Universitäts-Hautklinik Freiburg vorgenommen wird, erstrebenswert. Das dZh registriert seit 1990 in West-Deutschland und Berlin Patienten mit schweren, überwiegend arzneimittelinduzierten Hautreaktionen (RZANY et al. 1990).

Die statistische Auswertung solcher Fallstudien ist im Hinblick auf die Beurteilung von Arzneimittelrisiken äußerst problematisch. Da keine Kontrollgruppe mit gesunden Individuen vorhanden ist, lassen sich gebräuchliche epidemiologische Methoden zur Risikoschätzung nicht anwenden. Daneben basieren – unabhängig von der Problematik fehlender Kontrollen – viele epidemiologische Ansätze auf der Annahme, daß das durch eine bestimmte Exposition (z. B. Bluthochdruck) bedingte Risiko permanent besteht und somit Beobachtungszeit und Expositionszeit identisch sind. Eine solche Betrachtungsweise erscheint bei der Exposition durch Arzneimittel oftmals nicht gerechtfertigt. Beispielsweise reicht die Anwendungspalette bei Analgetika von der einmaligen Einnahme bis hin zur Dauertherapie. Kontroverse Diskussionen (z. B. FELDMANN et al. 1987) lösten deshalb die im Jahr 1986 veröffentlichten Resultate der 'INTERNATIONAL AGRANULOCYTOSIS AND APLASTIC ANAEMIA STUDY', einer Fall-Kontrollstudie, aus, welche den Mangel an geeigneten statistischen Methoden zur Beurteilung von Arzneimittelrisiken offenbart haben und zur Entwicklung mehrerer Ansätze (z. B. GUESS 1989; FELDMANN et al. 1989; MIETTINEN und CARO 1989) führten, bei denen der zeitliche Aspekt von Arzneimittelexpositionen berücksichtigt wird.

In der vorliegenden Arbeit werden drei Methoden zur Beurteilung seltener Arzneimittelrisiken beschrieben, für die lediglich das Datenmaterial der Patienten benötigt wird. Alle Ansätze werden am Beispiel von Patienten mit schweren Hautreaktionen, die durch das dZh erfaßt wurden, illustriert. Abschnitt 1 dokumentiert das Konzept arzneimittelbezogener Kennzahlen, die mit Hilfe von Verkaufs- bzw. Verordnungszahlen berechnet werden. Die Anwendung einer zweiten Methode, einem statistischen Modell von FELDMANN (1989, 1993 a), bei dem die Expositionsdauer für jeden einzelnen Patienten berücksichtigt wird und welches die Schätzung relativer Risiken auf der Basis eines prospektiven Falldesigns ermöglicht, erfolgt in Abschnitt 2. Mit Hilfe eines polychotomen logistischen Regressionsmodells werden in Abschnitt 3 verschiedene Diagnosekategorien miteinander verglichen (vgl. THOMAS et al. 1986). Bei der Anwendung sowie in der abschließenden Diskussion wird insbesondere die Eignung der aufgeführten Ansätze zur Beurteilung von Arzneimittelrisiken kritisch hinterfragt und die Problematik fehlender Kontrollpersonen erörtert.

1. Arzneimittelbezogene Kennzahlen

Werden für alle innerhalb eines bestimmten Beobachtungszeitraumes auftretenden Krankheitsfälle die vor der Reaktion eingenommenen Medikamente registriert, läßt sich für jedes Arzneimittel die Anzahl derjenigen Patienten bestimmen, die exponiert waren und bei denen sich die unerwünschte Reaktion entwickelt hat. Ein Vergleich der absoluten Zahlen ist nicht möglich, da einige Arzneimittel häufig verabreicht werden, andere dagegen sehr selten. Wünschenswert wäre es, die unter Exposition erkrankten Personen mit der jeweils exponierten Population ins Verhältnis zu setzen. Diese Population läßt sich allerdings nur durch Hilfsgrößen wie der Verkaufs- bzw. Verordnungszahl annähern. In der Bundesrepublik Deutschland ist die Angabe der Verordnungsmengen wie bei SCHWABE und PAFFRATH (1992) in definierten Tagesdosen (DDD) üblich. Mit diesen bekannten Größen lassen sich für verschiedene Präparate bzw. Wirkstoffe sogenannte arzneimittelbezogene Kennzahlen (AK)

$$AK_{\Delta t} = \frac{I}{V} 10^6 \quad (1)$$

berechnen, wobei I die Anzahl der im Zeitraum Δt nach vorheriger Einnahme des Arzneimittels auftretenden unerwünschten Reaktionen und V die zugehörige Verkaufs- bzw. Verordnungszahl in DDD bezeichnet. Es ist evident, daß sich Zähler und Nenner von (1) auf dieselbe Grundgesamtheit beziehen müssen.

Unter der Annahme, daß die exponierte Population sehr groß und die Erkrankungswahrscheinlichkeit sehr gering ist, kann man die in einem definierten Beobachtungszeitraum auftretenden Fälle als annähernd Poissonverteilt mit Parameter $\lambda = E(I)$ auffassen. Damit läßt sich anhand von Verteilungstabellen bzw. einer Normalverteilungsapproximation ein Konfidenzintervall $[I_u, I_o]$ für die erwartete Anzahl von Krankheitsfällen in der Grundgesamtheit ermitteln. Bei einer Vollerhebung können die beobachteten Reaktionen I als Stichprobe über die Zeit aufgefaßt werden, d. h. das berechnete Konfidenzintervall bezieht sich auf die in Beobachtungszeiträumen der Länge Δt erwartete Anzahl neuer Krankheitsfälle.

Das zugehörige Konfidenzintervall für die arzneimittelbezogene Kennzahl ergibt sich aus

$$\left[\frac{I_u}{V} 10^6, \frac{I_o}{V} 10^6 \right]. \quad (2)$$

Durch die in (1) und (2) definierten Größen lassen sich auf eine einfache Art die Zusammenhänge zwischen einer unerwünschten Reaktion und dem vorher eingenommenen Arzneimittel beschreiben. Bei der Anwendung ist allerdings zu berücksichtigen, daß der aufgeführte Ansatz mehrere Probleme mit sich bringt.

Ein entscheidender Kritikpunkt besteht darin, daß die Verkaufs- bzw. Verordnungszahlen nicht mit der tatsächlichen Einnahmemenge übereinstimmen. Wie sich zum Beispiel aus einer Studie der AOK Essen (HEEKE und GÜNTHER 1993) ableiten läßt, existieren dabei teilweise erhebliche Unterschiede zwischen verschiedenen Medikamenten. In Tabelle 1 wird exemplarisch veranschaulicht, welche Resultatsverzerrungen sich bei unterschiedlichen Differenzen zwischen tatsächlicher (hypothetischer!) Einnahmemenge und bekannter Verordnungszahl ergeben können. Betrachtet werden Patienten mit Stevens-Johnson-Syndrom (SJS), Toxisch epidermaler Necrolyse mit Maculae (TEN) und einer Übergangsform zwischen beiden Erkrankungen (SJS/TEN), die innerhalb eines zweiwöchigen Zeitraumes vor Auftreten einer dieser drei Reaktionen die in der Tabelle aufgeführten Wirkstoffe eingenommen haben. Die Grundgesamtheit bilden alle Einwohner der alten Bundesländer und aus Berlin, wobei durch das dZh eine Vollerhebung der genannten Hautreaktionen angestrebt wird. Aufgrund der im Beispiel illustrierten Resultate erscheint ein Vergleich der arzneimittelbezogenen Kennzahlen nur innerhalb

Tabelle 1: Auswirkungen der Abweichung zwischen Verordnungszahlen und tatsächlicher (hypothetischer!) Einnahmemenge auf arzneimittelbezogene Kennzahlen

Wirkstoff	I ¹⁾	V ²⁾	AK ³⁾	T ⁴⁾	AK ⁵⁾
Doxycyclin	5	214,1	0,02 [0,00; 0,04]	100	0,05 [0,01; 0,09]
Phenytoin	10	23,7	0,42 [0,17; 0,68]	20	0,50 [0,20; 0,80]
Amoxicillin	4	16,6	0,24 [0,06; 0,48]	8	0,50 [0,13; 1,00]

¹⁾ Anzahl der 1991 erfaßten Patienten mit SJS, SJS/TEN oder TEN, die innerhalb eines zweiwöchigen Zeitraumes vor Auftreten der Reaktion exponiert waren

²⁾ Verordnungen V in Millionen DDD (1991)

³⁾ arzneimittelbezogene Kennzahl [95%-Konfidenzintervall] auf der Basis von V

⁴⁾ hypothetische tatsächliche Einnahmemenge T in Millionen DDD für 1991

⁵⁾ arzneimittelbezogene Kennzahl [95%-Konfidenzintervall] auf der Basis von T

einer Arzneimittelgruppe (z. B. Antibiotika) sinnvoll, da hier zu erwarten ist, daß die Differenzen zwischen tatsächlichen Einnahmemengen und Verordnungszahlen für verschiedene Präparate etwa gleich sind.

Eine weitere Schwierigkeit des beschriebenen Ansatzes stellt die fehlende Möglichkeit zur Kontrolle potentieller Confounder dar. Durch den Vergleich der arzneimittelbezogenen Kennzahlen über mehrere Jahre sowie einer simultanen Betrachtung zusätzlicher Faktoren, wie zum Beispiel bestimmter Komedikationen, läßt sich eine systematische Überlagerung durch eine dieser Begleitvariablen allerdings leicht feststellen. Da keine Kontrollgruppe mit gesunden Individuen vorhanden ist, bleibt aber unbekannt, ob alle Patienten spezifische, nicht kontrollierbare endogene Charakteristika besitzen, ohne deren Vorliegen keine Reaktion aufgetreten wäre.

Ergänzend sei kurz angemerkt, daß im Nenner von (1) und (2) weder die Dauer der Exposition noch die Dosierung des Arzneimittels berücksichtigt wird. Weitere Probleme werden bei BÉGAUD et al. (1993) diskutiert.

Das in diesem Abschnitt beschriebene Konzept stellt eine sinnvolle Methode zur deskriptiven Auswertung einer Fallstudie dar. Risikounterschiede zwischen verschiedenen Arzneimitteln lassen sich erst dann vermuten, wenn die berechneten arzneimittelbezogenen Kennzahlen um ein Vielfaches voneinander abweichen und die zugehörigen Konfidenzintervalle weit auseinanderliegen. Aufgrund der offensichtlichen Schwächen dieses Ansatzes ist eine sichere Beurteilung von Arzneimittelrisiken aber kaum möglich.

2. Modell für intermittierende Expositionen

Das individuelle Einnahmeverhalten von Arzneimitteln läßt sich durch das von FELDMANN (1989, 1992, 1993a, 1993b) entwickelte Modell für intermittierende Expositionen darstellen. Entsprechend Abbildung 1 ist postulierbar, daß die Hazardfunktion während Perioden der Exposition den Wert λ_1 und unter Nichtexposition den Wert λ_0 annimmt. Ferner wird in dem Modell zugrundegelegt, daß eine Kovariable x sowohl einen Effekt auf die Hazardfunktion unter Exposition als auch auf die Hazardfunktion unter Nichtexposition ausüben kann. Zur Kontrolle mehrerer potentieller Confounder kann x als Vektor von Kovariablen aufgefaßt werden. Das absolute individuelle Risiko p einer unerwünschten Reaktion, d. h. die Wahrscheinlichkeit der Reaktion innerhalb einer festen Beobachtungszeit T bei gegebener Expositionszeit t sowie bekanntem Kovariablenwert, kann durch die kumulative Hazardfunktion dargestellt werden, welche bei seltenen Reaktionen der Fläche zwischen dem Funktionsgraphen und der Zeitachse (vgl. Abbildung 1) entspricht:

$$\text{Absolutes Risiko: } p = \lambda_1 t + \lambda_0 (T - t). \quad (3)$$

FELDMANN unterscheidet weiterhin, ob die unerwünschte Reaktion unter Exposition bzw. innerhalb der Latenzzeit der Exposition auftritt oder unter Nichtexposition erfolgt. Die resultierenden Risiken ergeben sich durch:

$$\text{Risiko unter Exposition: } p_1 = \lambda_1 t$$

$$\text{Risiko unter Nichtexposition: } p_0 = \lambda_0 (T - t).$$

Ein Vergleich des absoluten Risikos (3) mit dem absoluten Risiko einer hypothetischen Person, welche die gleiche Beobachtungszeit T und den gleichen Wert der Kovariablen x aufweist, jedoch nie exponiert ist ($t = 0$), liefert weitere Risikomaße. Das relative Risiko ergibt sich aus dem Quotienten der beiden absoluten Risiken:

$$\text{Relatives Risiko: } r = 1 + \left(\frac{\lambda_1}{\lambda_0} - 1 \right) \frac{t}{T}. \quad (4)$$

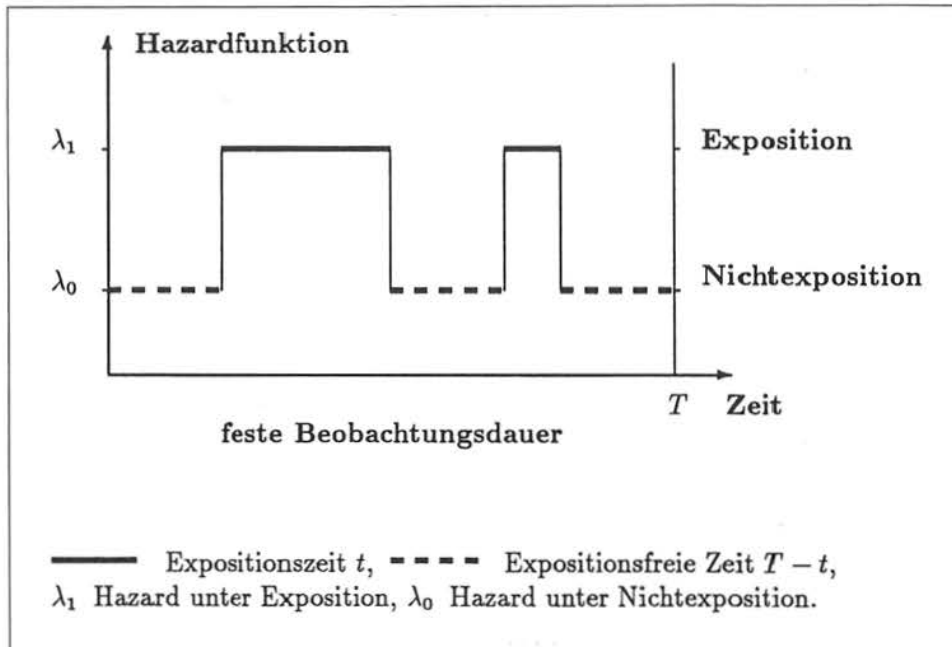


Abbildung 1: Modell von FELDMANN für intermittierende Expositionen

Im Gegensatz zur sonst gebräuchlichen Definition des relativen Risikos hängt r vom Anteil der unter Exposition verbrachten Zeit ($f = t/T$) ab.

Prospektives Falldesign

Neben einer Auswertung von Kohortenstudien und Fall-Kohortenstudien ermöglicht das beschriebene Modell auch die Schätzung relativer Risiken auf der Basis prospektiver Fallstudien (vgl. FELDMANN). Dazu wird eine Population prospektiv beobachtet. Für den i -ten auftretenden Fall ($i = 1, \dots, D$) ist sowohl die Expositionszeit t_i als auch die Zeit unter Nichtexposition retrospektiv zu ermitteln, anschließend erfolgt solange eine weitere Kontrolle des Patienten, bis die festgelegte Beobachtungszeit T_i erreicht ist. Bei seltenen Reaktionen ist es möglich, den Hazardquotienten $q_i = \lambda_{1i}/\lambda_{0i}$ zu schätzen. Dieser läßt sich durch einen log-linearen Link

$$\log(q_i) = \beta_0 + \beta_1 x_i$$

parametrisieren, wobei x_i den Kovariablenwert des i -ten Patienten und β_0 und β_1 die unbekannt Parameter bezeichnen. Die Schätzung erfolgt mit Hilfe der Maximum-Likelihood Methode, die Likelihoodfunktion lautet:

$$L(\beta_0, \beta_1) = \prod_{i=1}^D \left(\frac{p_{1i}}{p_i} \right)^{d_{1i}} \left(\frac{p_{0i}}{p_i} \right)^{d_{0i}} = \prod_{i=1}^D \left(\frac{q_i f_i}{q_i f_i + 1 - f_i} \right)^{d_{1i}} \left(\frac{1 - f_i}{q_i f_i + 1 - f_i} \right)^{d_{0i}}$$

Dabei bezeichnet d_{1i} eine binäre Variable, die den Wert 1 annimmt, falls die i -te Person ($i = 1, \dots, D$) eine unerwünschte Reaktion unter Exposition bzw. innerhalb der Latenzzeit der Exposition entwickelt, und ansonsten gleich 0 ist. Analog nimmt d_{0i} den Wert 1 an, falls die Reaktion während der Nichtexpositionszeit beginnt. Unter der Seltenheits-

annahme wird ausgeschlossen, daß sowohl eine Reaktion unter Exposition als auch unter Nichtexposition auftritt, d. h. es gilt $d_{1i} + d_{0i} = 1$ für alle Patienten $i = 1, \dots, D$.

Anwendung

Im folgenden werden die Daten derjenigen Patienten mit SJS, SJS/TEN-Übergangsform und TEN ausgewertet, die innerhalb eines vierwöchigen Beobachtungszeitraumes vor Auftreten einer der drei Hautreaktionen (Krankheitsbeginn am Indextag) ein Analgetikum eingenommen haben. Da keine detaillierten Informationen über die Arzneimittelaufnahme zwischen Krankheitsbeginn und dem jeweiligen Tag der Krankenhausentlassung vorliegen, wird zugrundegelegt, daß die Beobachtungszeit jeweils am Indextag endet. Wie FELDMANN in seinen Arbeiten betont, kann die Anwendung des Falldesigns bei terminalen Ereignissen und bei zensierten Beobachtungen zu verzerrten Resultaten führen. Trotzdem haben wir uns für die Anwendung des beschriebenen Modells entschieden, weil nach Beginn einer der oben genannten Hautreaktionen in der Regel alle vorher eingenommenen Arzneimittel abgesetzt werden. Folglich würde die Beobachtung bis zum Entlassungstag lediglich den Zeitraum ohne Exposition verlängern. Bei der Auswertung wird neben der Exposition durch Analgetika eine binäre Kovariable einbezogen, die angibt, ob zusätzlich ein Antiepileptikum verabreicht wurde ($x = 1$) oder nicht ($x = 0$). Die mittleren Expositions- und Beobachtungszeiten sowie die zugehörigen Standardabweichungen sind in Tabelle 2 aufgeführt.

Bei der Anwendung des von FELDMANN entwickelten Modells muß man definieren, wann die unerwünschte Reaktion möglicherweise auf die Exposition zurückzuführen ist und wann nicht. Damit ist es möglich, die Werte der Variablen d_{1i} und d_{0i} ($i = 1, \dots, D$) zu bestimmen. Bezogen auf das vorliegende Datenmaterial sind die in Abbildung 2 dargestellten Möglichkeiten zur Festlegung des durch die Arzneimittelaufnahme bedingten Risikozeitraumes denkbar: Fall A dokumentiert die Situation, daß ein Risiko lediglich vier Tage nach Einnahmebeginn besteht. Bei beiden Patienten könnte die Reaktion damit nicht der Exposition zugeschrieben werden, es gilt also $d_{01} = 1$ und $d_{02} = 1$. In Fall B stimmt der durch eine Arzneimittelaufnahme gegebene Risikozeitraum mit dem Einnahmezeitraum überein. Hier hätte Patient 1 die Reaktion unter Exposition entwickelt ($d_{11} = 1$) und bei Patient 2 wäre die Reaktion unter Nichtexposition bzw. Nichtvorhandensein eines Risikos aufgetreten ($d_{02} = 1$). Da einige Substanzen im Körper langsam abgebaut werden, könnte eine Person auch nach Absetzen eines Arzneimittels für einen bestimmten Zeitraum dem durch dieses Präparat bedingten Risiko ausgesetzt sein. Der Risikozeitraum bei Fall C ist zum Beispiel zwei Tage länger als der Einnahmezeitraum. Unter dieser Bedingung ist die Reaktion bei beiden Patienten möglicherweise auf die Exposition zurückzuführen ($d_{11} = 1$ und $d_{12} = 1$).

Da unbekannt ist, welche dieser Vorgaben das Risiko zutreffend beschreibt, erscheint es sinnvoll, zu untersuchen, inwieweit sich die Schätzer der Modellparameter und somit der relativen Risiken unter den verschiedenen Modifikationen des Risikozeitraumes ändern.

Tabelle 2: Mittlere Expositionszeit t , mittlere Beobachtungszeit T und mittlerer Anteil der Expositionszeit an der Beobachtungszeit f mit Angabe der empirischen Standardabweichungen (\cdot). Zeitangabe in Tagen.

x	Anzahl	t		T		f	
0	88	5,86	(6,49)	27,75	(2,73)	0,21	(0,24)
1	17	5,94	(5,15)	26,59	(3,18)	0,23	(0,20)

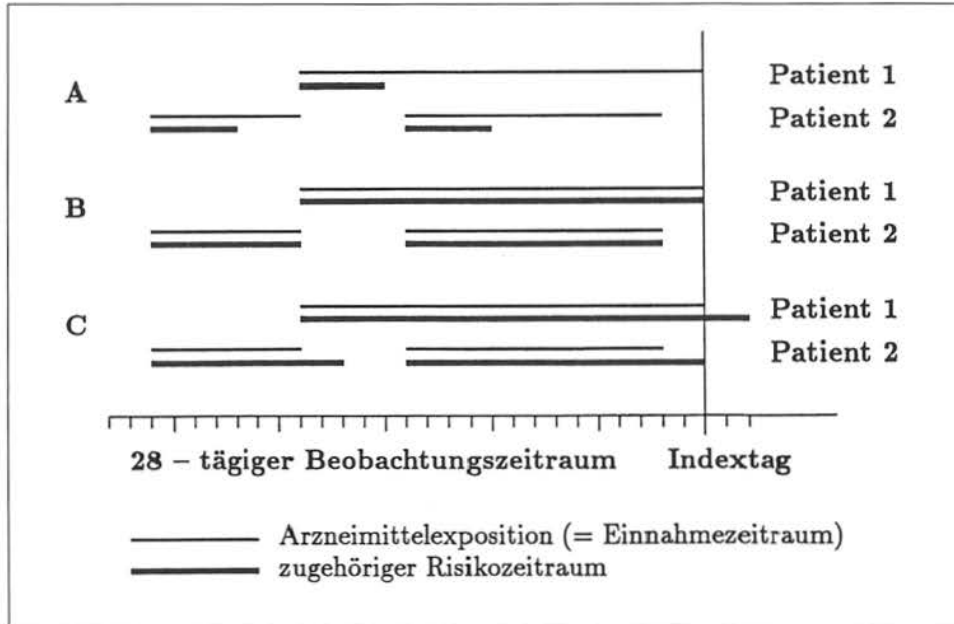


Abbildung 2: Arzneimittelexposition und zugehörige Risikozeiträume

Für die drei Vorgaben aus Abbildung 2 erhält man die in Tabelle 3 aufgelisteten Resultate. Getrennt nach den Werten der Kovariable x ist jeweils die Anzahl der Patienten aufgeführt, bei denen eine der Hautreaktionen innerhalb des Risikozeitraumes aufgetreten ist (D_1), sowie die Zahl der Patienten, deren Reaktion nicht auf die Exposition zurückgeführt werden kann (D_0). Die resultierenden Maximum-Likelihood-Schätzer der Parameter mit zugehörigen asymptotischen Standardabweichungen sind in Tabelle 4 dokumentiert. Wie intuitiv zu erwarten war, wird der Schätzer $\hat{\beta}_0$ größer, wenn die Anzahl der innerhalb des Risikozeitraumes auftretenden Reaktionen zunimmt. Die großen Differenzen führen allerdings zu sehr unterschiedlichen Risikoschätzungen. Für die mit Formel (4) ermittelten relativen Risiken, welche aus Tabelle 5 zu entnehmen sind, werden dabei verschiedene hypothetische Anteile der Expositionszeit an der

Tabelle 3: Anzahl der Hautreaktionen innerhalb des Risikozeitraums ($D_1 = \prod_{i=1}^D d_{1i}$) und Anzahl der Reaktionen ohne Risiko ($D_0 = \prod_{i=1}^D d_{0i}$)

x	Anzahl	Fall A		Fall B		Fall C	
		D_1	D_0	D_1	D_0	D_1	D_0
0	88	23	65	47	41	63	25
1	17	4	13	10	7	12	5

Tabelle 4: Maximum-Likelihood Schätzung: Log-Likelihood und Parameterschätzungen mit Angabe der asymptotischen Standardabweichungen (\cdot).

Risikozeitraum	Log-Likelihood	$\hat{\beta}_0$	$\hat{\beta}_1$
Fall A	-96,49	0,40 (0,30)	-0,34 (0,71)
Fall B	-67,22	2,13 (0,25)	-0,02 (0,63)
Fall C	-51,60	3,18 (0,26)	-0,37 (0,66)

Tabelle 5: Geschätzte relative Risiken auf der Basis der in Tabelle 4 aufgeführten Parameterschätzer

x	f	Fall A	Fall B	Fall C
0	0,22	1,1	2,6	6,1
1	0,22	1,0	2,6	4,4
0	0,50	1,2	4,7	12,5
1	0,50	1,0	4,7	8,8
0	0,80	1,4	7,0	19,4
1	0,80	1,1	6,8	13,5

gesamten Beobachtungszeit zugrundegelegt. Der Wert 0,22 entspricht etwa dem beobachteten mittleren Anteil (vgl. Tabelle 2). Betrachtet man beispielsweise eine Person, die über die Hälfte der Beobachtungszeit ($f = 0,50$) ein Analgetikum einnimmt, hätte diese, im Vergleich zu einer nie exponierten Person, bei Fall C ein etwa 9 bis 10 mal höheres Risiko als bei Fall A.

Die Ergebnisse verdeutlichen die starke Abhängigkeit der Risikoschätzung von den Vorgaben für den durch die Arzneimitteleinnahme bedingten Risikozeitraum und dem Anteil der unter Exposition verbrachten Zeit. Bei langen Beobachtungszeiten und sporadischer Exposition spielt dieser Anteil sicherlich eine untergeordnete Rolle. Es sei darauf hingewiesen, daß eine generelle Bewertung des beschriebenen Modells auf der Basis der hier dokumentierten Resultate nicht möglich ist, da der Einfluß der Zensierung auf die Schätzer, die eine zusätzliche Problematik darstellt, nicht untersucht wurde.

3. Polychotomes logistisches Regressionsmodell

In Fallregistern wird in der Regel das Datenmaterial von Patienten mit unterschiedlichen Krankheiten zusammengetragen. Der Vergleich zwischen den verschiedenen Diagnosekategorien ist möglicherweise von ätiologischem Interesse und liefert darüberhinaus Informationen über die potentiellen Zusammenhänge zwischen den Krankheiten und den vorher eingenommenen Arzneimitteln. Die vergleichende Analyse kann dabei mit Hilfe eines polychotomen logistischen Regressionsmodells durchgeführt werden (vgl. THOMAS et al. 1986). Dieser Ansatz ergibt sich als ein Spezialfall bei der Anwendung der von PRENTICE et al. (1984) entwickelten *Disease Rate Models*.

Gegeben seien K sich gegenseitig ausschließende Krankheitskategorien. Die bedingte Wahrscheinlichkeit, daß bei der i -ten Person ($i = 1, \dots, D$) die Krankheit k_i ($k_i = 1, \dots, K$) auftritt, läßt sich durch

$$P(k = k_i | x_i) = \frac{\exp(\beta'_{k_i} x_i)}{\sum_{m=1}^K \exp(\beta'_m x_i)} \quad (5)$$

definieren. Dabei bezeichnet $x'_i = (x_{i0}, x_{i1}, \dots, x_{ip})$ den Kovariablenvektor der i -ten Person, wobei $x_{i0} = 1$ für alle $i = 1, \dots, D$. Die betrachteten Einflußgrößen können diskret oder stetig sein. Ferner definiert $\beta'_m = (\beta_{m0}, \dots, \beta_{mp})$ den zu schätzenden Parametervektor für die Krankheitskategorie m ($m = 1, \dots, K$).

Da lediglich $(K - 1)(p + 1)$ der insgesamt $K(p + 1)$ unbekannt Parameter geschätzt werden können, legt man bei der Anwendung in der Regel eine Diagnosekategorie als Referenzgruppe fest und setzt alle Parameter dieser Vergleichsgruppe gleich Null. Die Parameterschätzung erfolgt durch die Maximierung der logarithmierten Likelihoodfunktion

$$\log L(\beta) = \sum_{i=1}^D P(k = k_i | x_i), \quad (6)$$

wobei auf gebräuchliche numerische Methoden, wie das Newton-Raphson Verfahren, zurückgegriffen wird. Ebenso wie beim dichotomen logistischen Regressionsmodell können die geschätzten Parameter als logarithmierte relative Risiken im Vergleich zur Referenzkategorie aufgefaßt werden. Darüberhinaus lassen sich analog zum dichotomen Modell Konfidenzintervalle für die Parameter und somit für die relativen Risiken angeben (vgl. KLEINBAUM et al. 1982).

Anstelle der hier beschriebenen Vorgehensweise schlagen THOMAS et al. (1986) sowie DUBIN und PASTERNAK (1986) vor, bei der Anwendung eines polychotomen logistischen Regressionsmodells auf die Auswahl einer Referenzgruppe zu verzichten. Mittels schrittweiser Selektion werden nur diejenigen Parameter – für jede Kovariable maximal $(K-1)$ – in das Modell aufgenommen, die einen signifikanten Zusammenhang zwischen der interessierenden Kovariable und der entsprechenden Krankheit beschreiben. Als Referenzgruppe dienen dann alle Patienten derjenigen Diagnosekategorien, für die der zugehörige Parameter nicht in das Modell einbezogen wurde. Für verschiedene Kovariablen erhält man hierbei möglicherweise unterschiedliche Referenzgruppen.

Anwendung

Neben den in den letzten beiden Abschnitten betrachteten Hautreaktionen (SJS, SJS/TEN-Übergangsform und TEN) werden bei der Auswertung zusätzlich Patienten einbezogen, bei denen ein Erythema exsudativum multiforme (EEM) mit Schleimhautbeteiligung aufgetreten ist. Da es sich beim EEM im Vergleich zu den drei übrigen Krankheiten um eine leichtere Hautreaktion handelt, werden die Patienten dieser Diagnosekategorie als Referenzgruppe ausgewählt. Die Auswertung basiert auf dem Datenmaterial von 305 Patienten, von denen bei 47 Personen ein EEM, bei 122 ein SJS, bei 91 eine SJS/TEN-Übergangsform und bei 45 eine TEN diagnostiziert wurde. Es wird für jeden Patienten untersucht, ob dieser innerhalb eines vierwöchigen Beobachtungszeitraumes vor Auftreten der Reaktion ein Analgetikum und/oder ein Antiepileptikum eingenommen hat. Diese beiden Expositionen sind durch die binären Variablen ANALGET und ANTIEPI beschrieben. Darüberhinaus definiert die binäre Variable ALTER,

Tabelle 6: Maximum-Likelihoodschätzung: Parameterschätzungen mit Angabe der asymptotischen Standardabweichungen () und geschätzte relative Risiken im Verhältnis zur Referenzgruppe mit Angabe der 95%-Konfidenzintervalle []

Diagnose	Variable	$\hat{\beta}_{ij}$	\hat{r}_{ij}
EEM	ALTER	0	1
	ANALGET	0	1
	ANTIEPI	0	1
	CONST	0	—
SJS	ALTER	0,81 (0,49)	2,3 [0,9; 5,9]
	ANALGET	0,54 (0,37)	1,7 [0,8; 3,6]
	ANTIEPI	2,18 (1,04)	8,8 [1,1; 68]
	CONST	-0,46 (0,23)	—
SJS/TEN	ALTER	2,38 (0,50)	11 [4,1; 29]
	ANALGET	1,02 (0,41)	2,8 [1,2; 6,2]
	ANTIEPI	2,29 (1,06)	9,8 [1,2; 79]
	CONST	-0,75 (0,30)	—
TEN	ALTER	1,34 (0,54)	3,8 [1,3; 11]
	ANALGET	0,27 (0,46)	1,3 [0,5; 3,2]
	ANTIEPI	2,85 (1,07)	17 [2,1; 140]
	CONST	-0,75 (0,31)	—

ob die Patienten über 60 Jahre alt sind oder nicht. Die mit Hilfe der BMDP-Prozedur PR ermittelten Resultate sind in Tabelle 6 aufgeführt.

Anhand der Ergebnisse ließe sich beispielsweise folgern, daß das Risiko nach der Einnahme eines Antiepileptikums eine TEN zu entwickeln 17 mal größer ist als das Risiko, an einem EEM zu erkranken. Dieses Beispiel dient allerdings lediglich zur Illustration, da bei der Auswertung weder die Dauer der Exposition, noch die Dosierung der Arzneimittel berücksichtigt wurde. Daneben ist die Referenzgruppe mit 47 Personen äußerst klein, d. h. die Schätzer sind vermutlich sehr instabil.

Besteht eine Vergleichsgruppe aus kranken Personen, lassen sich — je nach Auswahl dieser Referenzgruppe — lediglich die unterschiedlichen Einflüsse einer Exposition auf die betrachteten Krankheiten feststellen. Eine allgemeine Beurteilung von Arzneimittelrisiken ist dagegen nicht möglich. So könnte die protektive Wirkung einer Arzneimittelexposition auf die Krankheit der Referenzgruppe fälschlicherweise als erhöhtes Risiko dieser Exposition für die übrigen Krankheiten interpretiert werden.

Diskussion

In dieser Arbeit wurden drei statistische Ansätze zur Beurteilung von Arzneimittelrisiken in einer Fallstudie beschrieben und an einem Anwendungsbeispiel diskutiert. Durch den Vergleich arzneimittelbezogener Kennzahlen kann man auf einfache Art die potentiellen Zusammenhänge zwischen einer unerwünschten Reaktion und dem vorher eingenommenen Arzneimittel beschreiben. Da sich die zugehörige exponierte Population nur unzureichend bestimmen läßt und dieser Ansatz keine Möglichkeit zur Kontrolle von Confoundern bietet, ist eine zuverlässige Beurteilung von Arzneimittelrisiken kaum möglich. Die Anwendung des von FELDMANN (1989, 1993a) entwickelten Modells, bei

dem die Expositionszeit für jeden Patienten ins Verhältnis gesetzt wird zur entsprechenden Zeit ohne Exposition, hat verdeutlicht, welche Schwierigkeiten bei der Modellierung der zeitlichen Komponente von Arzneimittelexpositionen zu berücksichtigen sind. Bei der Auswertung des vorliegenden Datenmaterials ergab sich eine starke Abhängigkeit der Risikoschätzungen von der Festlegung des durch die Exposition bedingten Risikozeitraumes. Demnach sind genaue Kenntnisse über die Wirkungsweise von Arzneimitteln erforderlich. Mit Hilfe eines polychotomen logistischen Regressionsmodells kann der Einfluß bestimmter Arzneimittel auf unterschiedliche Erkrankungen untersucht werden. Besteht die Kontrollgruppe, wie im Beispiel, aus kranken Individuen, lassen sich die Resultate im Hinblick auf die Risikobeurteilung nur unzureichend interpretieren, da der Einfluß der interessierenden Arzneimittelexposition auf die Krankheit der Kontrollpersonen unbekannt ist.

Zusammenfassend kann festgestellt werden, daß alle in dieser Arbeit dokumentierten Ansätze hinsichtlich der Beurteilung von Arzneimittelrisiken im Vergleich zu Studien, bei denen eine Kontrollgruppe mit gesunden Individuen vorhanden ist, eine sehr eingeschränkte Aussagekraft besitzen. Das größte Problem bei der Auswertung reiner Fallstudien besteht darin, daß alle Patienten möglicherweise spezifische, nicht kontrollierbare, Charakteristika aufweisen, ohne deren Existenz keine der betrachteten Krankheiten aufgetreten wäre.

Da unerwünschte Arzneimittelreaktionen zunehmend an gesundheitspolitischer Bedeutung gewinnen, kommt der epidemiologischen Analyse der Risiken zugelassener Arzneimittel ein erhöhter Stellenwert zu. Einen ausführlichen Überblick über die Prinzipien der Planung und Auswertung von Phase-IV Studien findet man bei VICTOR et al. (1991). Fallstudien können im Hinblick auf die Beurteilung von Arzneimittelrisiken allerdings auch in Zukunft nur eine untergeordnete Rolle spielen.

Literaturverzeichnis

- BÉGAUD, B.; CHASLERIE, A.; FOURRIER, A.; HARAMBURU, F.; MIREMONT, G. für die „Association pour la Recherche Methodologique en Pharmacovigilance“ (1993): *Methodological Approaches in Pharmacoepidemiology*. Elsevier, Amsterdam.
- DUBIN, N.; PASTERNAK, B. S. (1986): Risk assesment for case-control subgroups by polychotomous logistic regression. *American Journal of Epidemiology* **123**, 1101–1117.
- FELDMANN, U.; GAUSS, W.; KRETSCHMER, F.-J.; REPGES, R. (1987): Risks of agranulocytosis caused by dipyron (Letter to the Editor). *Journal of the American Medical Association* **256**, 2590–2591.
- FELDMANN, U.; REPGES, R.; GAUSS, W.; KRETSCHMER, F.-J. (1989): *Mathematische Methoden zur Beurteilung unerwünschter Wirkungen bei sporadischer Arzneimittelnahme*. Schriftenreihe Statistik, Biomathematik und Informationsverarbeitung, Heft 7, Mannheim.
- FELDMANN, U. (1992): Planung und Auswertung von Studien zur Arzneimittelsicherheit. *Biometrie und Informatik in Medizin und Biologie* **23**, 16–23.
- FELDMANN, U. (1993a): Epidemiologic assessment of risks of adverse reactions associated with intermittent exposure. *Biometrics* **49**, 419–428.
- FELDMANN, U. (1993b): Design and analysis of drug safety studies, with special reference to sporadic drug use and acute adverse reactions. *Journal of Clinical Epidemiology* **46**, 237–243.
- GUESS, H. A. (1989): Behaviour of the exposure odds ratio in a case-control study when the hazard function is not constant over time. *Journal of Clinical Epidemiology* **42**, 1179–1184.
- HEEKE, A.; GÜNTHER, J. (1993): *Arzneimittel im Müll*. Hrsg: AOK Essen, pharmakologischer Beratungsdienst, Stattwerk eG, Essen.
- KLEINBAUM, D. G.; KUPPER, L. L.; MORGENSTERN, H. (1982): *Epidemiological Research*. Wadsworth, London.
- MIETTINEN, O. S.; CARO, J. J. (1989): Principles of nonexperimental assesment of excess risk, with special reference to adverse drug reactions. *Journal of Clinical Epidemiology* **42**, 325–331.

- PRENTICE, R. L.; VOLLMER, W. M.; KALBFLEISCH, J. D. (1984): On the use of case series to identify disease risk factors. *Biometrics* **40**, 445–458.
- RZANY, B.; MÖCKENHAUPT, M.; KÖRNER, M.; WIEK, K.; STOCKER, U.; SCHÖPF, E. (1992): Dokumentationszentrum für schwere Hautreaktionen. *Deutsches Ärzteblatt* 36, A₁: 2879–2880.
- SCHWABE, U.; PAFFRATH, D. (1992): *Arzneiverordnungs-Report '92*. Fischer, Stuttgart.
- THE INTERNATIONAL AGRANULOCYTOSIS AND APLASTIC ANAEMIA STUDY (1986): Risks of agranulocytosis and plastic anaemia: A first report of their relation to drug use with special reference to analgesics. *Journal of the American Medical Association* **256**, 1740–1757.
- THOMAS, D. C.; GOLDBERG, M.; DEWAR, R.; SIEMIATYCKI, J. (1986): Statistical methods for relating several exposure factors to several diseases in case-heterogeneity studies. *Statistics in Medicine* **5**, 49–60.
- VICTOR, N.; SCHÄFER, H.; NOWAK, H. (1991): *Arzneimittelforschung nach der Zulassung: Bestandsaufnahme und Perspektiven*. Schriftenreihe Medizinische Informatik und Statistik, Band 73, Springer, Berlin, Heidelberg.

Anschrift des ersten Verfassers: Dipl.-Stat. Norbert Holländer, Inst. f. Med. Biometrie u. Med. Informatik, Universität Freiburg, Stefan-Meier-Str. 26, 79104 Freiburg

Eingang des Manuskripts 15. 4. 94

Eingang der revidierten Fassung 8. 11. 94

Das wissensbasierte System ADNEXPERT zur Unterstützung der sonografischen Diagnose von Adnextumoren

Johann Brüning¹⁾, Rolf Becker²⁾, Michael Entezami²⁾, Volker Loy³⁾, Richard Vonk¹⁾, Hans Weitzel²⁾, Thomas Tolxdorff¹⁾

Zusammenfassung

ADNEXPERT ist ein wissensbasiertes System zur rechnergestützten Ultraschalldiagnostik von Adnextumoren. Es wird gezeigt, wie in einem fallbasierten Ansatz ein hochspezialisierter Bereich der medizinischen Diagnostik damit sinnvoll unterstützt werden kann. Dabei kommen Methoden der Künstlichen Intelligenz und der Statistik zur Anwendung. Die Implementierung erfolgte auf unterschiedlichen Betriebssystem-Plattformen. Zum Nachweis der Richtigkeit der diagnostischen Aussagen des ADNEXPERT-Systems werden die Ergebnisse der ersten Phase der Systemevaluierung vorgestellt.

Summary

ADNEXPERT is a knowledge-based system that provides computer support for the ultrasound diagnosis of adnexal tumors. Using a case-based approach, an efficient support of a highly specialized area of medical diagnostics can be demonstrated. Methods of Artificial Intelligence and statistics are applied. The implementation was effected using different operating system platforms. To prove the accuracy of the diagnostic statements generated by the ADNEXPERT system, the results of the first evaluation phase are reported.

Einleitung

Das Ovarialkarzinom steht bei den Genitalkarzinomen der Frau in den westlichen Industrieländern hinsichtlich seiner Häufigkeit nach dem Zervixkarzinom an zweiter, hinsichtlich seiner Letalität jedoch an erster Stelle. In den USA erkranken zur Zeit

¹⁾ Institut für Medizinische Statistik und Informationsverarbeitung, Abteilung Medizinische Informatik, Universitätsklinikum Benjamin Franklin der Freien Universität Berlin

²⁾ Abteilung für Gynäkologie und Geburtshilfe, Universitätsklinikum Benjamin Franklin der Freien Universität Berlin

³⁾ Institut für Pathologie, Universitätsklinikum Benjamin Franklin der Freien Universität Berlin

jährlich etwa 20 000 Frauen an einem Ovarialkarzinom, und etwa 13 500 kommen durch diese Erkrankung zu Tode [CRE91]. Eines von 70 weiblichen Neugeborenen wird während seines Lebens ein Ovarialkarzinom entwickeln [BAR84].

Die Letalität des Ovarialkarzinoms ist nach wie vor hoch: die 5-Jahres-Überlebensrate in den klinischen Stadien III und IV ist geringer als 15%. Die Diagnose wird oft verspätet und nur selten im Rahmen einer gynäkologischen Vorsorgeuntersuchung gestellt.

Der Einsatz der Vaginalsonografie bietet seit 1987 eine erheblich verbesserte Möglichkeit der Früherkennung von Adnextumoren. Auf der Seite des Untersuchers erfordern die Technik und Befundung der vaginalen Sonografie ein hohes Maß an Erfahrung und gynäkologischem Expertenwissen. Da auch bei beschwerdefreien Frauen pathologische sonografische Befunde der Adnexe häufig sind, ist eine subtile Differentialdiagnose nötig, die nicht operationspflichtige Befunde erkennt und zu einer möglichst sicheren Einschätzung der Dignität etwaiger Prozesse beiträgt.

Im Universitätsklinikum Benjamin Franklin der Freien Universität Berlin werden in der Abteilung Gynäkologie und Geburtshilfe seit dem 1. 7. 1989 systematisch Daten von sonografisch untersuchten Patientinnen elektronisch gespeichert.

Primäres Ziel des ADNEXPERT-Projekts war es, ein Dialogsystem zu entwickeln, das den sonografischen Untersucher bei seinem Weg zum kompetenten Befund unterstützt. Dabei werden die Erfahrungen und das Wissen, die sich aus dem umfangreichen Datenmaterial ableiten lassen, unmittelbar zur Verfügung gestellt. Dies bedeutet, daß synchron zum Dialogverlauf mit dem Benutzer eine Bewertung der eingehenden Informationen vom System vorgenommen wird und entsprechende diagnostische Hinweise gegeben werden.

Für die Evaluierung des ADNEXPERT-Systems wird ein phasenorientiertes Modell verwendet, das die klinische Prüfung von Arzneimitteln zum Vorbild hat. Im Sinne der ersten Phase einer solchen klinischen Prüfung eines entscheidungsunterstützenden diagnostischen Systems wird dann die Leistungsfähigkeit des ADNEXPERT-Systems anhand von 69 neuen Adnextumorfällen untersucht.

Die ADNEXPERT-Falldatenbank

Zur systematischen elektronischen Speicherung der Daten von sonografisch untersuchten Patientinnen wird in der Abteilung für Gynäkologie und Geburtshilfe eine dBASE IV-Applikation eingesetzt. Mit Hilfe dieses Programms, einer Eigenentwicklung des Ultraschallabors, werden die relevanten alphanumerischen Daten aller gynäkologischen und geburtshilflichen Untersuchungen gespeichert. Die sonografischen Bilddaten werden separat als Videobandaufzeichnungen und Videoprinterbilder dokumentiert und verwaltet.

Im Zeitraum vom 1. 7. 1989 bis zum 31. 12. 1991 wurden in diesem Rahmen präoperativ die sonografischen Daten von 482 Adnextumoren von insgesamt 406 Patientinnen erhoben [BEC93]. Die sonografischen Daten umfassen neben der klinischen Fragestellung die detaillierte Befundung des Untersuchers und verschiedene routinemäßig erhobene Daten wie Größenmaße des Uterus und des Ovars. Zur Generierung der Wissensbasis des ADNEXPERT-Systems aus dieser Falldatenbank wurden neben der wichtigen Information des Alters der Patientin ausschließlich Daten der sonomorphologischen Tumorcharakteristika verwendet.

Entscheidend für den Transfer dieser Informationen in die Wissensbasis eines diagnoseunterstützenden Systems ist die Tatsache, daß für diese 482 Adnextumoren nicht nur die detaillierten sonomorphologischen Befunde, sondern auch die gesicherten histopathologischen Diagnosen vorliegen. Die sonografischen Daten aller Tumoren wurden präoperativ erhoben. Durch Korrelation von Befundtypologien mit gesicherten

Diagnosen kann medizinisches Wissen mit geeigneten Methoden computergerecht formalisiert werden [PUP87]. Im Fall von ADNEXPERT kommen dazu Methoden der Statistik und der Künstlichen Intelligenz zur Anwendung. Das formalisierte medizinische Wissen kann dann in hinreichend ähnlich strukturierten Entscheidungssituationen abgerufen und auf neue Fälle übertragen werden. Für ADNEXPERT sind das sonografische Untersuchungen bei Patientinnen mit der Verdachtsdiagnose eines Adnextumors. Auf der Grundlage der bekannten 482 Fälle bewertet das System neue Tumoren und gibt diagnostische Hinweise aus, indem es die Dignität des Tumors und die Wahrscheinlichkeit der Zugehörigkeit zu definierten histopathologischen Adnextumorklassen einschätzt.

Zudem waren retrospektiv über einen Beobachtungszeitraum von elf Jahren (1. 1. 1981 bis 31. 12. 1991) von allen Patientinnen, die wegen eines Adnextumors im Klinikum Steglitz operiert wurden, histopathologische Befunden in einer Datenbank erfaßt worden [BEC93]. Mit Hilfe dieser Daten konnte die Altersverteilung der Dignitäten und der unterschiedlichen von ADNEXPERT behandelten histologischen Tumorklassen auf der Grundlage von 2290 Tumoren für die Wissensbasis des Systems berechnet werden.

Die Tabelle 1 zeigt zusammenfassend die histopathologischen Befunde aller Adnextumoren, deren Daten in die ADNEXPERT-Wissensbasis Eingang fanden. Die Befundklassifizierung der Ovarialtumoren erfolgte nach den neun Hauptgruppen der „Histological Typing of Ovarian Tumours“ der World Health Organisation [SER73]. Zusätzliche Aufnahme haben die differentialdiagnostisch relevanten Gruppen der tubaren Raumforderungen und der Paraovarialzysten gefunden.

Die Tumorfalldatenbank umfaßt alle im Zeitraum von 1981 bis 1991 in Klinikum Steglitz operierten 2290 Adnextumoren von 1987 Patientinnen. Die Ovarialtumoren der neunten Hauptgruppe, der Gruppe der tumorähnlichen Veränderungen, waren in unserem Kollektiv am häufigsten vertreten. Es fanden sich 911 tumorähnliche Veränderungen von insgesamt 2015 Ovarialtumoren, dies entspricht einem Anteil von 45,2%. Auf die Gesamtmenge der 2290 Adnextumoren bezogen ergibt sich ein relativer Anteil von 39,8%.

Da die Gruppe der tumorähnlichen Veränderungen ausschließlich benigne Erkrankungen enthält, ist ihre Einbeziehung in eine Aufschlüsselung der Dignität von Ovarialtumoren für das Erstarben der Gruppe der benignen Tumoren hauptverantwortlich.

Tabelle 1: Histopathologische Typen von 2290 Adnextumoren des Operationsgutes der Jahre 1981–1991 sowie der 1989–1991 im Klinikum Steglitz sonomorphologisch erfaßten 482 Ovarialtumoren, modifiziert nach [BEC93]

Adnextumorart	Operationsfrequenz 1981–1991	Sonografisch erfaßt 1989–1991
I Epitheliale Ovarialtumoren	805	186
II Keimstrang-Stroma-Tumoren des Ovars	71	6
III Lipidzelltumoren des Ovars	3	0
IV Keimzelltumoren des Ovars	188	66
V Gonadoblastome des Ovars	0	0
VI Mesenchymale Tumoren ohne Ovarspezifität	0	0
VII Unklassifizierte Tumoren des Ovars	2	0
VIII Metastatische Tumoren des Ovars	35	3
IX Tumorähnliche Bedingungen des Ovars	911	137
Tubare Raumforderungen	178	74
Paraovarialzysten	97	10
Beobachtete Adnextumoren	2290	482

Manche Autoren klammern deshalb in ihrer Betrachtung der Dignität von Ovarialbeziehungsweise Adnextumoren diese Gruppe aus [BEN68, KOO89]. Oft werden die Tumoren der Hauptgruppen I bis XIII in diesem Zusammenhang auch als „echte“ Ovarialtumoren bezeichnet.

Sich auf diese sogenannten „echten“ Tumoren zu beschränken, entspricht aus differentialdiagnostischen Überlegungen nicht der Intention unserer Arbeitsgruppe. Für die sonografische Beurteilung der Adnextumoren spielen gerade das Erkennen und die Differentialdiagnose der stärksten histopathologischen Gruppe der tumorähnlichen Veränderungen eine zentrale Rolle. Ein möglichst sicheres Identifizieren dieser Veränderungen hat entscheidende Auswirkungen für das klinische Management dieser Erkrankungen. Die wichtigste Entscheidung, die aufgrund der Einordnung des Tumors durch Anamnese, Erhebung von Laborparametern und differenziertem sonografischem Untersuchungsbefund gefällt werden muß, ist die Entscheidung, ob eine weitere invasive Diagnostik unmittelbar anzuraten ist und, wenn dies der Fall ist, in welcher Form – als Laparaskopie oder Laparatomie – diese zu erfolgen hat. Im günstigsten Fall, der eben auch bei der sonografischen Verdachtsdiagnose einer tumorähnlichen Veränderung vorliegt, kann unter Anberaumung einer sonografischen Kontrolluntersuchung auf eine weitere invasive Diagnostik verzichtet werden.

Methoden der Diagnoseunterstützung durch ADNEXPERT

ADNEXPERT verbindet in seinem Ansatz eine statistische Methode, die zur Auswertung einer Falldatenbank dient, mit einer Problemlösungskomponente aus dem Bereich der Expertensysteme. Zusammen mit den medizinischen Experten wurde ein Dialogschema mit maximal⁴⁾ 15 zu beantwortenden Fragen erarbeitet, in der die relevanten Kriterien einer exakten sonografischen Befundbeschreibung abgehandelt werden. Diese 15 Kriterien repräsentieren ausnahmslos Attribute der Falldatenbank und sind in der folgenden Übersicht wiedergegeben:

- | | |
|------------------------------|--|
| 1: Alter der Patientin | 9: Beschaffenheit der inneren Zystenoberfläche |
| 2: Kalkuliertes Tumolvolumen | 10: Grenzkontur der inneren Zystenoberfläche |
| 3: Innenstruktur des Tumors | 11: Echoverteilung der flüssigen Phase |
| 4: Vorhandensein von Ascites | 12: Echogenität der flüssigen Phase |
| 5: Zystenanzahl | 13: Sedimentationseffekt innerhalb der Zyste |
| 6: Zystenarchitektur | 14: Innenstruktur der soliden Tumoranteile |
| 7: Septendicke | 15: Lokalisation der soliden Tumoranteile |
| 8: Zystenkommunikation | |

Für jedes dieser Kriterien wird eine Dialogfrage mit drei bis neun Antwortmöglichkeiten formuliert. Die jeweils letzte Antwortmöglichkeit ist mit dem Standardtext „sonstiges“ oder „unbekannt“ vorgegeben. Sie kann bei fehlender Information, unklaren oder nicht in das Raster vorgesehener Antwortmöglichkeiten fallenden Befunden gewählt werden. Für die Frage nach der Anzahl der Tumorzysten lauten die möglichen Antworten zum Beispiel:

Frage 5: Anzahl der Zystenkammern des Tumors?

- | | |
|------------------------------------|-------------------------------------|
| Antwort 1: Anzahl der Kammern: 1 | Antwort 4: Anzahl der Kammern: 6–10 |
| Antwort 2: Anzahl der Kammern: 2 | Antwort 5: Anzahl der Kammern: > 10 |
| Antwort 3: Anzahl der Kammern: 3–5 | Antwort 6: Sonstiges/unbekannt |

⁴⁾ Abhängig vom Verlauf des Dialogs werden nicht immer alle Fragen gestellt.

Außer einer initialen Frage nach dem Alter der Patientin betreffen die restlichen 14 Fragen ausschließlich den sonografischen Befund. Als mögliche Diagnose wird zum einen die Dignität der Tumoren, das heißt ihre Zugehörigkeit zu den Gruppen „benigne“, „semimaligne“ und „maligne“ analysiert. Die Prognose der Dignität eines Adnextumors stellt das wichtigste diagnostische Kriterium für das klinische Management einer Patientin mit dieser Erkrankung dar.

Um eine weitere diagnostische Eingrenzung vornehmen zu können, wird zum anderen gezielt die Zugehörigkeit zu speziellen histologischen Diagnoseklassen prognostiziert. Welche speziellen Diagnosen in diesem Rahmen näher analysiert werden, wurde unter klinischen Gesichtspunkten festgesetzt. Ein Kriterium für ihre Auswahl war die Häufigkeit der Diagnose: häufig gestellte Diagnosen sind auch häufig in unserer Falldatenbank repräsentiert. Damit steht eine große Menge detaillierter Informationen über diese Tumoren zur Verfügung, die in Bezug zu einem neu zu bewertenden Fall gesetzt werden können. Ein anderes Kriterium für die Auswahl war das etwaige Vorhandensein von typischen sonografischen Befundkonstellationen oder von pathognomonischen sonografischen Befundkriterien für eine histologische Diagnoseklasse. In diesen Fällen besteht eine große Chance, solche Befunde einer spezifischen Diagnoseklasse zuzuordnen zu können. Die Tabelle 2 zeigt in einer Übersicht die von ADNEXPERT betrachteten histologischen Diagnosegruppen, die Anzahl der in der Falldatenbank vertretenen Tumoren der einzelnen Klassen und vorhandene pathognomonische Befundkriterien:

Insgesamt ergeben sich rein rechnerisch für die Beantwortung der 15 Dialogfragen über 30 Millionen Antwortkombinationen, die natürlich nicht alle medizinisch sinnvolle Kombinationen darstellen. Diese immense Menge der theoretischen Antwortkombinationen verdeutlicht aber, wie vielschichtig und variabel die eingehenden Informationen schon bei einem sehr überschaubaren Dialog sein können. Deshalb war nach einem geeigneten Verfahren zu suchen, das diesen Informationsfluß bündelt und leicht interpretierbare Bewertungen von neuen Fällen generiert.

Hierzu wurde in einem ersten Schritt für alle Dignitäten und alle speziellen histologischen Tumordiagnosen für jede der insgesamt 81 Antwortmöglichkeiten ein Certainty Factor

Tabelle 2: Die von ADNEXPERT bewerteten histologischen Diagnoseklassen und die zugrundeliegenden Falldaten aus dem Klinikum Steglitz mit typischen sonografischen Befunden

Diagnoseklasse zur Tumorklassifikation	operiert 1981 – 1991	sonogr. 1989 – 1991	Pathognomonisches sonografisches Kriterium bzw. prototypische Befundkonstellation
Epitheliale Ovarialtumoren	805	186	
Keimstrangtumoren	71	6	rein solider Tumor bei jungen Frauen
Keimzelltumoren	188	66	strichmusterförmige Echoverteilung, Kugelzeichen
Funktionelle Ovarialveränderungen	318	59	wabige Echoverteilung beim corpus luteum hämorrhagicum
Endometriose	204	36	grob oder fein granulierte Echoverteilung
Einfache Zysten	389	42	
Sonstige Tumoren	315	87	

(CF) aus der Falldatenbank berechnet. Das Modell der CFs gilt als klassisches Verfahren der KI und wurde im Rahmen des Expertensystems MYCIN zur Diagnostik und Therapie bakterieller Infektionskrankheiten an der Stanford University entwickelt [SHO76].

Ein Certainty Factor repräsentiert den Vertrauensgrad mit dem ein Experte eine bestimmte Diagnose unter einer gegebenen Symptomatik als vorliegend ($0 < CF \leq 1$) beziehungsweise als nicht vorliegend ($0 < CF \geq -1$) erachtet. Eine mit Sicherheit unter einer bestimmten Symptomkonstellation vorliegende Diagnose erhält für diese Symptomkonstellation den Certainty Factor +1, einer unter einer vorliegenden Symptomkonstellation ausgeschlossenen Diagnose wird analog der CF-Wert -1 zugeordnet. Die zwischen diesen Extremwerten liegenden CFs repräsentieren unscharfe Bewertungen, wie sie auch in der medizinischen Diagnostik am häufigsten anzutreffen sind.

Neu an unserem Ansatz ist es, die CFs nicht intuitiv festzulegen, sondern sie nach einem statistischen Verfahren aus dem uns vorliegenden Datenbestand zu berechnen. Wir verwenden hierzu einen von Heckerman beschriebenen Algorithmus, der die CF-Berechnung aus den A-priori-Wahrscheinlichkeiten der Falldatenbank ermöglicht.

Heckerman zeigt eine mögliche Transformation des Bayesschen Theorems in die Formel zur parallelen Kombination von CFs, wie sie in MYCIN verwendet wird [HEC86]. Dies verdeutlicht die enge Verwandtschaft des Certainty Factor Modells zum Theorem von Bayes. Ausgehend von diesen Überlegungen entwickelt Heckerman eine modifizierte Definition des CF, die sich im Gegensatz zur ursprünglichen Definition durch Shortliffe durch eine symmetrische Beziehung zwischen a-priori- und a-posteriori-Wahrscheinlichkeiten auszeichnet:

$$CF(H | E) := \begin{cases} \frac{P(H | E) - P(H)}{P(H | E) \cdot (1 - P(H))} & \text{if } P(H | E) > P(H) \\ \frac{P(H | E) - P(H)}{P(H) \cdot (1 - P(H | E))} & \text{if } P(H) \geq P(H | E) \end{cases}$$

P = Wahrscheinlichkeit (Probability)

H = Diagnose (Hypothesis)

E = Symptom (Event)

$CF(H | E)$ = CF einer Diagnose H unter dem Symptom E

$P(H)$ = Wahrscheinlichkeit der Diagnose H

$P(H | E)$ = Wahrscheinlichkeit einer Diagnose H unter dem Symptom E .

Das verwendete Wahrscheinlichkeitsmaß ist die Häufigkeitsverteilung, aus der sich dann für jede der 15 Dialogfragen Tabellen mit den nach Heckerman berechneten Certainty Factors erstellen lassen. Die nach der Heckerman-Formel berechneten CF-Extremwerte von -1 repräsentieren eine Ausnahmesituation und wurden von uns modifiziert. Wurde für eine Symptomkategorie keine Diagnose beobachtet, liegt also eine Beobachtungshäufigkeit von 0% vor, beträgt der nach der Formel berechnete CF-Wert für diese Diagnose unter dem gegebenen Symptom -1. Liegt dieses Symptom vor, wird bei dem CF von -1 die Diagnose ausgeschlossen. Der Ausschluß einer Diagnose kann in der Medizin nur in seltenen Fällen von der Ausprägung eines Symptoms abhängig gemacht werden. Für das sonografische Symptombild eines Adnextumors trifft das mit Sicherheit nicht zu. Um den Effekt des sicheren Diagnoseausschlusses bei einem CF-Wert von -1 zu vermeiden, wurden alle CF-Werte der Wissensbasis, die nach der Heckerman-Formel mit -1 berechnet wurden, auf -0,98 abgeändert.

Als Beispiel folgen hier die Häufigkeits- und CF-Tabellen für die erste Frage des Dialogs, in der nach dem Alter der untersuchten Patientin gefragt wird (Tab. 3 und 4). In der Alterskategorie „14–19“ wurde kein semimaligner Tumor beobachtet. Der nach

Tabelle 3: Altersverteilung der Dignitäten von 2290 Adnextumoren

Altersgruppe	Benigne		Semimaligne		Maligne	
	N	%	N	%	N	%
14–19 Jahre	49	95,9	0	0,0	2	4,1
20–29 Jahre	258	94,5	5	1,8	10	3,6
30–39 Jahre	371	93,5	9	2,3	17	4,3
40–49 Jahre	544	90,7	11	1,8	45	7,5
50–59 Jahre	326	80,9	23	5,7	54	13,4
60–69 Jahre	157	64,1	13	5,3	75	30,6
70–79 Jahre	181	75,4	10	4,2	49	20,4
≥ 80 Jahre	54	65,1	5	6,0	24	28,9
Alle	1938	84,6	76	3,3	276	12,1

Tabelle 4: Certainty Factors der Dignitäten von 2290 Adnextumoren

Altersgruppe	Benigne CF	Semimaligne CF	Maligne CF
14–19 Jahre	+0,77	–0,98	–0,69
20–29 Jahre	+0,68	–0,46	–0,72
30–39 Jahre	+0,61	–0,32	–0,67
40–49 Jahre	+0,43	–0,46	–0,41
50–59 Jahre	–0,23	+0,43	+0,11
60–69 Jahre	–0,68	+0,39	+0,69
70–79 Jahre	–0,44	+0,21	+0,47
≥ 80 Jahre	–0,66	+0,46	+0,66

Heckerman resultierende CF-Wert von -1 wird in $-0,98$ abgeändert. Die Altersgruppe „14–19“ ist nicht als sicheres Ausschlußkriterium für einen semimaligen Tumor anzusehen. Bei einem Anwachsen der Falldatenbank ist zu erwarten, daß in seltenen Fällen auch in dieser Altersgruppe semimaligne Tumoren vertreten sind.

Für jede Dignität und jede spezielle histologische Tumordiagnose wird außerdem ein Start-CF gesondert berechnet, indem die prozentualen Häufigkeiten für jede Diagnoseklasse linear auf die CF-Skala zwischen -1 und $+1$ abgebildet werden. Diese Startwerte repräsentieren den Vertrauensgrad in die jeweilige Diagnose für Patientinnen, für die außer der Verdachtsdiagnose „Adnextumor“ noch keine weiteren Befundinformationen vorliegen. Für die Dignitäten sind die ermittelten Startwerte in der Tabelle 5 wiedergegeben:

Ausgehend von diesen Startwerten werden nach jeder Beantwortung einer Frage für jede Dignität und jede spezielle histologische Tumordiagnose die für diese Antwort in den CF-Tabellen vorliegenden CFs entnommen und mit den Vorgängerwerten verrech-

Tabelle 5: CF-Startwerte der Dignitäten von 2290 Adnextumoren

	Benigne	Semimaligne	Maligne
CF-Startwert	+0,69	–0,93	–0,76

net⁵⁾, um so die neue eingehenden Informationen im Rahmen der sonografischen Befundung zu berücksichtigen. Dies geschieht mittels der *parallelen Kombination* der beiden *CF*-Werte. Diese ergibt sich aus:

$$CF(H | E_1, E_2) := \begin{cases} s + t - s^*t & \text{if } s \text{ and } t \text{ positive} \\ \frac{s + t}{1 - \min\{|s|, |t|\}} & \text{if } s^*t \in (-1, 0) \\ \text{undefined} & \text{if } s^*t = -1 \\ s + t + s^*t & \text{if } s \text{ and } t \text{ negative} \end{cases}$$

$$s := CF(H | E_1), \quad t := CF(H | E_2).$$

Nach der Beantwortung der letzten Dialogfrage entsprechen die dann vorliegenden *CF*-Werte der endgültigen diagnostischen Beurteilung des untersuchten Tumors durch das ADNEXPERT-System.

Der Dialogablauf im ADNEXPERT-System realisiert das Ausklammern von Fragen, die im Verlauf des Dialogs als nicht relevant erkannt werden. Wird zum Beispiel ein Tumor als „rein zystisch“ klassifiziert, werden alle weiteren Fragen, die sich auf rein solide Tumoren beziehen, nicht mehr gestellt.

Die verbale Kategorisierung von sonografischen Befundbeschreibungen ist in vielen Fällen unscharf und stark von der subjektiven Bewertung des jeweiligen Diagnostikers abhängig. Dies gilt insbesondere für den sonografischen Untersuchungsbefund der weiblichen Adnexe. Eine umfassende Kategorisierung der Befundtypologien für die sonografische Beurteilung der Adnexe liefert die Grundlage für das Dialoggerüst von ADNEXPERT [BEC93]. Um die Verbalisierung der Befunde eines bildgebenden Verfahrens wie der Sonografie zu standardisieren, bedarf es einer Vielzahl von Befundbeobachtungen, die es dann ermöglichen, einen Bildbefund prototypisch zu beschreiben. In diesem Zusammenhang ist es unabdingbar, neben einer verbalen Beschreibung auch prototypisches Bildmaterial zu präsentieren. Für das ADNEXPERT-System konnte für jede der Antwortmöglichkeiten der 13 Dialogfragen, die sich auf den sonografischen Befund der Adnexe beziehen, ein prototypischer Bildbefund von sehr guter Qualität zur Verfügung gestellt werden. Um eine einheitliche Beantwortung der Fragen durch verschiedene Untersucher zu gewährleisten, wurden diese Bilder in eine Bildbibliothek integriert, zu der ein bequemer und schneller Zugriff zur Dialoglaufzeit realisiert wurde. Nicht zuletzt diese schnelle Verfügbarkeit von zahlreichen sonografischen Beispielbildern ermöglicht einen sinnvollen Einsatz von ADNEXPERT auch zu Lehrzwecken.

Implementierung des ADNEXPERT-Systems

Schon in den frühen Phasen der Softwareplanung wurde eine strikte Trennung der Problemlösungskomponente von der eigentlichen Wissensbasis des Systems konzipiert. Dies entspricht nicht nur grundlegenden formalen Forderungen eines strukturierten Software Engineerings, sondern ermöglicht einen Einsatz des Systemkerns auch zur Bearbeitung anderer Wissensbasen [SOM89].

ADNEXPERT basiert deshalb auf einem Programmkern XPERT, dessen Struktur unabhängig von den Inhalten der Wissensbasis aus dem Gebiet der sonografischen

⁵⁾ In seiner Originalarbeit führt Shortliffe die parallele Kombination der *CF*s in zwei Schritten durch [SHO76]. Nach Berechnung des *Measure of Believe (MB)* und des *Measure of Disbelieve (MD)* wird erst in einem zweiten Schritt der neue *CF* berechnet. Die hier wiedergegebene Formel für die parallele Kombination der *CF*s gibt die Berechnung in einem Schritt ohne die vorausgehende Bestimmung von *MB* und *MD* wieder [MEL80].

Adnextumor-Diagnostik konzipiert wurde. Prinzipiell ist der Einsatz von XPERT in beliebigen Wissensdomänen zum Zweck der Entscheidungsunterstützung möglich. Dies setzt natürlich ein entsprechend strukturierbares Expertenwissen voraus. In Kooperation mit dem Institut für Medizinische Statistik und Dokumentation des Klinikums der RWTH Aachen wird XPERT für das dort entwickelte methodisch verwandte CARAT-System zur Verfügung gestellt. CARAT dient der Unterstützung der radiologischen Diagnostik von Knochentumoren [PEL93].

Im Rahmen der Planungsphase wurde ein Prototyp des Systemkerns in der Programmiersprache Turbo-Prolog 2.0 erstellt. Das Prototyping umfaßte in diesem Fall nicht nur die Dialogkomponente des Systems, sondern enthielt auch schon eine einfache Implementierung der Problemlösungskomponente. Im Sinne einer ökonomischen Projektplanung war dieses erweiterte explorative Prototyping unbedingte Voraussetzung, um den Aufwand der geplanten Implementierung einer komfortablen Dialogführung unter Einsatz einer grafischen Benutzeroberfläche zu rechtfertigen.

Da die Eindrücke, die durch die Arbeit mit dem Prototyp gewonnen werden konnten, sehr vielversprechend waren, schlossen sich die Planung und Implementierung von ADNEXPERT unter der X Window-Umgebung und dem Betriebssystem UNIX unmittelbar an. X Window bietet mit seiner grafischen Oberfläche alle Möglichkeiten, eine komfortable Benutzerschnittstelle zu schaffen, die eine schnelle und übersichtliche Dialogführung mit ADNEXPERT ermöglicht. Zur Generierung der Oberfläche wurde das Motif Toolkit der Open Software Foundation eingesetzt.

Zu der bestehenden dBASE IV-Anwendung zur Erfassung von Patientendaten gynäkologischer Ultraschalluntersuchungen unter DOS konnte eine Schnittstelle geschaffen werden, die den Dialog und die Befundung eines Adnextumors auch automatisiert ermöglicht. Interessant ist in diesem Zusammenhang der Einsatz eines DOS-Emulators unter UNIX, unter dem die dBASE IV-Anwendung in einem X-Fenster läuft und ihre Daten an das ADNEXPERT-Programm zur Auswertung sendet.

Nach den Richtlinien des OSF/Motif Style Guide konnte ein umfassendes Hilfesystem eingerichtet werden, das kontextsensitiv arbeitet und auch den Zugriff auf die umfangreiche Ultraschall-Bildbibliothek ermöglicht.

Darüber hinaus wurde die Problemlösungskomponente des Systems in dBASE implementiert und auch die Wissensbasis als dBASE-Datenbank bereitgestellt. So steht neben der komfortablen Implementierung unter UNIX auch noch eine Minimallösung unter dem Betriebssystem DOS zur Verfügung, die unmittelbar in das erwähnte System zur Erfassung von Patientendaten am Fachbereich Gynäkologie und Geburtshilfe des Klinikum Steglitz integriert wurde.

Ergebnisse der Evaluierung des ADNEXPERT-Systems

Die Validierung eines wissensbasierten Systems läßt sich als ein Prozeß definieren, der sicherstellt, daß das Wissen und die Ratschläge des Systems richtig, vollständig und konsistent sind [MIL90]. Eine Validierung ist Bestandteil einer Systemevaluierung. Die Evaluierung eines medizinischen wissensbasierten Systems wird von den meisten Autoren als die Summe aller notwendigen Prozesse begriffen, die die Effizienz des Systems und damit seinen Stellenwert im Gesundheitssystem beurteilen [O'MO91]. Für die Notwendigkeit einer Evaluierung werden von Wyatt und Spiegelhalter drei wesentliche Gründe angeführt: ethische, rechtliche und intellektuelle [WYA91]. Unter Beachtung ethischer Gesichtspunkte müssen danach insbesondere folgende Fragen beantwortet werden:

- Ist das System effektiv?
- Ist das System sicher?
- Hat der Einsatz des Systems einen günstigen Einfluß auf die Ressourcen des Gesundheitssystems?

Die intellektuellen Gründe für die Systemevaluierung sind ebenfalls gewichtig. Eine objektivierende Kontrolle der Leistungsfähigkeit von wissenschaftsbasierten Systemen kann wichtige Erkenntnisse für das Design zukünftiger Systeme liefern. Bei Vorliegen gesicherter Validierungsergebnisse können Verfahren zur Wissensrepräsentation und Problemlösung gegebenenfalls optimiert werden.

Einige Autoren schlagen für die Evaluierung von entscheidungsunterstützenden diagnostischen Systemen ein an die Phasen der klinischen Prüfung von Arzneimitteln angelehntes Modell vor [SPI83, OHM89]. Diese Phaseneinteilung für die Evaluierung von Systemen zur diagnostischen Entscheidungsunterstützung erscheint sinnvoll und liegt unserem Evaluierungskonzept zugrunde.

Die Phase I der Evaluierung hat die Einschätzung der diagnostischen Richtigkeit der angewandten Methode zum Ziel. Zu diesem Zweck werden die Falldaten einer Testgruppe als Eingabedaten für das System verwendet und die real vorliegenden Diagnosen mit den vom System getroffenen Diagnosevorschlägen verglichen. Bewertungsparameter sind neben der Anzahl und relativen Häufigkeit der richtig gestellten Diagnosen die Anzahlen und Häufigkeiten der falsch befundeten Fälle. Eine klinische Bewertung dieser statistischen Größen muß sich anschließen. Basiert die Wissensbasis des zu evaluierenden Systems auf einer Falldatenbank, dürfen die Testdaten nicht dieser Falldatenbank entnommen werden. Eine überoptimistische Einschätzung des Systemverhaltens wäre die Folge [OHM89, HIL90].

Um die Richtigkeit der diagnostischen Einschätzung des ADNEXPERT-Systems zu beurteilen, wurden im Ultraschalllabor der Abteilung für Gynäkologie und Geburtshilfe des Klinikums Steglitz und in einer Berliner gynäkologischen Praxis sonografische Daten von Adnextumoren erhoben. Die Testdatenbank enthält den sonografischen Befund von 69 Adnextumoren, die im Zeitraum vom 1. 1. 1992 bis zum 1. 10. 1993 erhoben wurden und für die eine gesicherte histopathologische Diagnosestellung vorliegt.

Zur Testdatenerfassung wurde das im Routinebetrieb verwendete dBASE-Programm eingesetzt. Die Befundung und Datenerfassung erfolgte durch zwei auf dem Gebiet des gynäkologischen Ultraschalls spezialisierte Mediziner, die mit der im ADNEXPERT-System verwendeten Nomenklatur zur Befundbewertungen für Adnextumoren vertraut sind. Um eine Beeinflussung bei der Befundbewertung zu vermeiden, wurde das ADNEXPERT-System für die Testdatenerfassung nicht eingesetzt. Ebenso war die histopathologische Diagnose der erfaßten Tumoren dem bewertenden Arzt zum Zeitpunkt der sonografischen Beurteilung nicht bekannt.

Alle Datensätze wurden anschließend von einem weiteren im gleichen Fachgebiet spezialisierten Mediziner auf inhaltliche und formale Richtigkeit hin überprüft. Erst danach wurden die bekannten histopathologischen Diagnosen zugänglich gemacht und in der Testdatenbank erfaßt. In einem weiteren Schritt erfolgte eine zweite formale und medizinische Plausibilitätskontrolle durch einen medizinischen Informatiker.

Anschließend wurden die Certainty Factors zur Beurteilung der Dignität und zur speziellen histologischen Tumorklassifikation ermittelt. Für jeden Tumor wurden demnach 10 CFs berechnet: drei zur Dignitätsbeurteilung und sieben zur histologischen Tumorklassifikation. Um die diagnostische Richtigkeit im Rahmen einer statistischen Auswertung beurteilen zu können, wurden vier Resultatgruppen gebildet. Die heuristische Definition dieser Gruppen erfolgte vor der CF-Berechnung für die Testdatenbank unter Anlegung kritischer Maßstäbe. Nur CF-Muster mit einer wirklich klaren Verteilung zugunsten der richtigen Diagnose erhalten dabei den Status einer „klaren Diagnosestellung“. Ebenso strenge Maßstäbe wurden bei der Definition der anderen drei Diagnosegruppen angelegt. Folgendes Diagnose-Staging wurde konzipiert:

- | | |
|---|--------------------------------|
| 1. Klare Diagnosestellung | 3. Keine diagnostische Aussage |
| 2. Effiziente diagnostische Eingrenzung | 4. Fehldiagnose |

Dieses Staging wurde, getrennt sowohl für die Beurteilung der Dignität als auch für die

speziellen Diagnosen zur histologischen Tumorklassifikation, nach klar definierten Kriterien automatisiert durchgeführt. Es beruht ausschließlich auf dem Muster der zehn Certainty Factors, die für jeden Testdatensatz berechnet wurden. Beispielhaft sei der Algorithmus für die Zuordnung zur Gruppe „klare Diagnosestellung“ für die Beurteilung der Tumordignität angeführt. Der CF -Wert, der für die reale Tumordignität ermittelt wurde, wird im folgenden als CF_{RD} bezeichnet:

Eine klare Diagnosestellung der Tumordignität liegt in folgendem Fall vor:

CF_{RD} hat den höchsten Wert der Dignitäts- CF s und außerdem gilt:

$CF_{RD} \geq 0,8$ und die Differenz zum nächstkleineren CF ist $\geq 0,3$ oder

$CF_{RD} \geq 0,5$ und die Differenz zum nächstkleineren CF ist $\geq 0,4$ oder

$CF_{RD} \geq 0,3$ und die Differenz zum nächstkleineren CF ist $\geq 1,0$

Ergebnisse der Dignitätsbeurteilung durch ADNEXPERT

Basierend auf dieser Klassifikation zur Abschätzung der diagnostischen Richtigkeit der von ADNEXPERT unterbreiteten Vorschläge zur Dignitätsbeurteilung und speziellen histologischen Tumorklassifikation, ergeben sich für die 69 Tumoren der Testdatenbank folgende Resultate. Zunächst wird die Dignitätsbeurteilung betrachtet (Tab. 6):

Die Tabelle 6 gibt über die Leistungsfähigkeit des ADNEXPERT-Systems Auskunft, die Dignität eines Adnextumors anhand seines sonografischen Befundes richtig einzuschätzen. In 71,0% aller Fälle stellt ADNEXPERT die Dignität eines Tumors richtig fest. In weiteren 14,5% kann eine sinnvolle diagnostische Eingrenzung erfolgen. Demnach ist in insgesamt 85,5% der Fälle ein diagnostischer Informationsgewinn bei der Dignitätsbeurteilung mit Hilfe des ADNEXPERT-Systems zu verzeichnen.

Der diagnostische Informationsgewinn ist für die benignen Tumoren mit 89,1% größer als der für die malignen Tumoren mit 66,7%. Für die semimaligen Tumoren liegt die entsprechende Quote bei 100%. Da aufgrund ihrer Seltenheit nur zwei semimaligne Tumoren in der Testdatenbank vertreten sind, ist dieses Ergebnis mit Vorsicht zu bewerten. Aufgrund ihrer natürlichen Häufigkeitsverteilung sind die benignen Tumoren in der der ADNEXPERT-Wissensbasis zugrundeliegenden Falldatenbank mit einem relativen Anteil von 84,6% weitaus häufiger als die malignen Tumoren vertreten. Dieser Sachverhalt kann eine Begründung dafür liefern, daß die Dignität benigner Tumoren von ADNEXPERT treffsicherer als die Dignität der malignen Tumoren beurteilt wird.

In insgesamt 11,6% der Fälle, dies entspricht 8 von 69 Adnextumoren, wurde die Dignität des Tumors falsch eingeschätzt. Die Quote von 0,0% für die Fehldiagnosen der

Tabelle 6: Diagnostische Richtigkeit der Dignitätsbeurteilung von 69 Tumoren im Rahmen der Phase I der Evaluierung des ADNEXPERT-Systems

	Tumordignität		Tumordignität							
			Klare Diagnosestellung		Effiziente diagnostische Eingrenzung		Keine diagnostische Aussage		Fehldiagnose	
	N	%	N	%	N	%	N	%	N	%
Benigne	55	100,0	46	83,6	3	5,5	1	1,8	5	9,1
Semimaligne	2	100,0	0	0,0	2	100,0	0	0,0	0	0,0
Maligne	12	100,0	3	25,0	5	41,7	1	8,3	3	25,0
Alle Tumoren	69	100,0	49	71,0	10	14,5	2	2,9	8	11,6

semimaligen Tumoren hat aufgrund der geringen Fallzahl in dieser Gruppe nur geringe Aussagekraft. In 9,1% der Fälle, dies entspricht 5 von 46 Tumoren, wurde ein benigner Tumor falsch als wahrscheinlich maligne klassifiziert. Für drei dieser fünf Tumoren wurde ein Hinweis auf eine eventuell bestehende Semimalignität gegeben (CF für semimaligne $> +0,5$). Beim Verdacht eines malignen Geschehens sind weitere invasive Maßnahmen indiziert. In der Regel wird in dieser Situation eine Laparotomie durchgeführt. Für diese fünf Fälle hätte ADNEXPERT eine Entscheidung in dieser Richtung unterstützt.

Weit mehr Sicherheitsrelevanz besitzen die Fehldiagnosen für die drei der zwölf malignen Tumoren, dies entspricht 25,0% der bösartigen Tumoren. Bei zwei dieser Tumoren wurde ein starker Hinweis auf bestehende Benignität gegeben ($CF > +0,8$), in einem Fall ein schwacher Hinweis auf Gutartigkeit ($CF = +0,25$). Die CF s für „semimaligne“ und „maligne“ lagen für alle drei Tumoren im negativen Bereich. Hier sind Sicherheitsmaßnahmen vonnöten, die eine Fehldiagnose dieser Tumoren verhindern. In diesem Zusammenhang ist nochmals auf den Stellenwert zu verweisen, den ein System wie ADNEXPERT in der klinischen Diagnose einnehmen kann. Es ist als zusätzliches diagnostisches Instrumentarium zu begreifen, dessen Aussagen ausschließlich auf dem sonografischen Befund und dem Alter der Patientin beruhen. Das klinische Management von Patientinnen mit Adnextumoren kann und darf nicht alleine unter Verlaß auf die Diagnosevorschläge des ADNEXPERT-Systems betrieben werden. Neben der Einbeziehung weiterer diagnostischer Methoden müssen zusätzliche Sicherheitsmaßnahmen getroffen werden, die vor allem in obligaten sonografischen Kontrolluntersuchungen bestehen müssen, die auch bei benigne erscheinenden pathologischen Veränderungen der Adnexe zu erfolgen haben.

Betrachtet man die Beurteilung der Dignität durch das ADNEXPERT-System getrennt für die einzelnen histologischen Klassen von Tumoren, die von ADNEXPERT unterschieden werden, ergibt sich folgendes Bild (Tab. 7):

Tabelle 7: Diagnostische Richtigkeit der Dignitätsbeurteilung von 69 Tumoren im Rahmen der Phase I der Evaluierung des ADNEXPERT-Systems, bezogen auf unterschiedliche histologische Tumorklassen

Histologische Tumorklassifikation	Tumordignität									
			Klare Diagnosestellung		Effiziente diagnostische Eingrenzung		Keine diagnostische Aussage		Fehldiagnose	
	N	%	N	%	N	%	N	%	N	%
Epitheliale Ovarialtumoren	30	100,0	15	50,0	8	26,7	1	3,3	6	20,0
Keimstrangtumoren	1	100,0	0	0,0	0	0,0	1	100,0	0	0,0
Keimzelltumoren	0		0		0		0		0	
Funktionelle Ovarialveränderungen	18	100,0	18	100,0	0	0,0	0	0,0	0	0,0
Endometriose	6	100,0	6	100,0	0	0,0	0	0,0	0	0,0
Einfache Zysten	11	100,0	10	90,9	1	9,1	0	0,0	0	0,0
Sonstige Tumoren	3	100,0	0	0,0	1	33,3	0	0,0	2	66,7
Alle Tumoren	69	100,0	49	71,0	10	14,5	2	2,9	8	11,6

Für die Beurteilung der Dignität der funktionellen Ovarialveränderungen, der ovariellen Endometriosen und der einfachen Zysten zeigen sich sehr gute Ergebnisse. Auch für die Dignitätsbeurteilung der epithelialen Tumoren durch ADNEXPERT kann in 76,7% ein diagnostischer Informationsgewinn verzeichnet werden. In 50% der Fälle werden diese Tumoren eindeutig in ihrer Dignität diagnostiziert. In 20% der Fälle wird die Dignität der epithelialen Tumoren falsch beurteilt. Für die anderen Tumorklassen, die Keimstrangtumoren, die Keimzelltumoren und die sonstigen Tumoren, liegen in der Testdatenbank der Phase I zu wenige Falldaten vor, um die Leistungsfähigkeit von ADNEXPERT bei der Dignitätsbeurteilung isoliert für Tumoren dieser Klassen sicher bewerten zu können.

Zusammenfassend kann festgestellt werden, daß die Beurteilung der Dignität der Adnextumoren der Testdatenbank durch das ADNEXPERT-System gut gelingt.

Ergebnisse der histologischen Tumorklassifikation durch ADNEXPERT

Für die diagnostische histologische Tumorklassifikation der 69 Adnextumoren der Testdatenbank ergeben sich in der Phase I der klinischen Prüfung folgende Ergebnisse (Tab. 8):

Von den 69 Tumoren der Testdatenbank wurden von ADNEXPERT-System 30 richtig klassifiziert, bei weiteren 7 konnte eine sinnvolle diagnostische Eingrenzung zur histologischen Tumorklassifikation geleistet werden. Demnach ist bei insgesamt 37 Tumoren, dies entspricht einem Anteil von 53,6%, ein diagnostischer Informationsgewinn bei der histologischen Klassifikation des Tumors zu verzeichnen. Für 11 Tumoren, dies entspricht 15,9% konnte keine diagnostisch verwertbare Aussage zur histologischen Klassifikation getroffen werden. Weitere 21 Tumoren, also ein relativer Anteil von 30,4% der Tumoren, wurden fehlklassifiziert.

Tabelle 8: Diagnostische Richtigkeit der histologischen Tumorklassifikation von 69 Tumoren im Rahmen der Phase I der Evaluierung des ADNEXPERT-Systems

Histologische Tumorklassifikation	Histologische Tumorklassifikation									
			Klare Diagnosestellung		Effiziente diagnostische Eingrenzung		Keine diagnostische Aussage		Fehl-diagnose	
	N	%	N	%	N	%	N	%	N	%
Epitheliale Ovarialtumoren	30	100,0	19	63,3	2	6,7	2	6,7	7	23,3
Keimstrangtumoren	1	100,0	0	0,0	0	0,0	0	0,0	1	100,0
Keimzelltumoren	0		0		0		0		0	
Funktionelle Ovarialveränderungen	18	100,0	4	22,2	4	22,2	4	22,2	6	33,3
Endometriose	6	100,0	0	0,0	1	16,7	2	33,3	3	50,0
Einfache Zysten	11	100,0	7	63,6	0	0,0	2	18,2	2	18,2
Sonstige Tumoren	3	100,0	0	0,0	0	0,0	1	33,3	2	66,7
Alle Tumoren	69	100,0	30	43,5	7	10,1	11	15,9	21	30,4

Innerhalb der einzelnen histologischen Klassifikationsgruppen, die ADNEXPERT verwendet, nehmen die epithelialen Tumoren mit 70,0% die Spitzenstellung für einen diagnostischen Informationsgewinn bei der Tumorklassifikation ein. Ähnlich wie bei der großen Trefferquote für die Dignitätsbeurteilung benigner Tumoren kann auch in diesem Fall der große Anteil⁶⁾ von Daten epithelialer Tumoren in der ADNEXPERT-Wissensbasis eine Begründung liefern. Sehr gut werden auch die einfachen Zysten klassifiziert, die in 63,6% der Fälle klar als solche erkannt werden.

Die Endometriosen sind wie auch die Keimstrangtumoren (ein Fall) und die sonstigen Tumoren (drei Fälle) zu selten in der Testdatenbank vertreten, um ausreichende Hinweise über die Klassifikationsfähigkeit des ADNEXPERT-Systems für diese Tumorgruppen zu erhalten. Dies gilt auch für die Keimzelltumoren, die leider in der Testdatenbank nicht vertreten sind. Bei den sonstigen Tumoren handelt es sich ausnahmslos um metastatische Tumoren des Ovars, die in der Falldatenbank der ADNEXPERT-Wissensbasis nur mit einer Häufigkeit von 1,53% vertreten sind (35 von 2290 Tumoren).

Die vorliegenden Daten sind Beleg dafür, daß das ADNEXPERT-System in der überaus heterogenen histologischen Gruppe der Adnextumoren der Testdatenbank in vielen Fällen einen sinnvollen Beitrag zur Tumorklassifikation aufgrund des sonografischen Erscheinungsbildes des Tumors zu leisten vermag.

Ausblick

Aufgrund der vielversprechenden Ergebnisse der ersten Phase der Evaluierung von ADNEXPERT sollen im Rahmen der zweiten Phase in einem Peer review-Verfahren Diagnosevorschläge des Systems mit denen von gynäkologischen Experten paarweise verglichen und auf ihre Richtigkeit hin überprüft werden. Die Quelle der Diagnose (System oder Experte) ist dabei dem Expertengremium des Peer review unbekannt. Bei dieser multizentrisch zu konzipierenden Studie sollen nicht nur histologisch verifizierte Diagnosen Aufnahme finden, sondern auch durch sonografische Kontrolluntersuchungen und klinische Verläufe objektivierte Diagnosestellungen. Dieses Vorgehen ermöglicht den Einschluß der zahlreichen beobachteten benignen Pathologien der Adnexe, die in der klinischen Routine keiner weiteren invasiven Maßnahme unterzogen werden. Es ist sehr wichtig, weiteren Aufschluß über die Qualität der diagnostischen Aussagen des ADNEXPERT-Systems für diese große Gruppe der zumeist funktionellen Veränderungen des Ovars zu erhalten.

Literatur

- [BAR84] BARBER, H. R. K.: Ovarian Cancer — Diagnosis and Management. *Am J Obstet Gynecol* **150**, 910—916, 1984.
- [BEC93] BECKER, R.: Sonomorphologie des inneren Genitale der Frau unter besonderer Berücksichtigung der gesunden und erkrankten Adnexe. Habilitationsschrift, Freie Universität Berlin, 1993.
- [BEN68] BENNINGTON, J. L., FERGUSON, B. R., and HARBER, S. L.: Incidence and Relative Frequency of Benign and Malignant Ovarian Neoplasms. *Am J Obstet Gynecol* **32**, 627—632, 1968.
- [CRE91] CREASMAN, W. T. and DiSAIA, P. J.: Screening in Ovarian Cancer. *Am J Obstet Gynecol* **165**, 7—10, 1991.

⁶⁾ In der Falldatenbank gehören 805 Tumoren von 2290 Tumoren zur Gruppe der epithelialen Tumoren. Dies entspricht einem relativen Anteil von 35,15%.

- [HEC86] HECKERMAN, D.: Probabilistic Interpretation for MYCIN's Certainty Factors. In: Kanal, L. N. and Lemmer, J. F. (eds.): *Uncertainty in Artificial Intelligence*, 167–196, Amsterdam, North-Holland, 1986.
- [HIL90] HILDEN, J. and HABBEMA, D. F.: Evaluation of Clinical Decision Aids – More to Think About. *Med Inf* **15**, 275–284, 1990.
- [KOO89] KOONINGS, P. P., CAMPBELL, K., MISHALL, D. R., and GRIMES, D. A.: Relative Frequency of Primary Ovarian Neoplasms: A 10-year Review. *Am J Obstet Gynecol* **74**, 921–926, 1989.
- [MEL80] VAN MELLE, W. J.: *System Aids in Constructing Consultation Programs*. Ann Arbor, UMI Research Press, 1980.
- [MIL90] MILLER, P. L. and SITTIG, D. F.: The Evaluation of Clinical Decision Support Systems: What is Necessary Versus What is Interesting. *Med Inf* **15**, 185–190, 1990.
- [OHM89] OHMANN, C.: Computerunterstützte Diagnose und Expertensysteme. In: *Dtsch Med Wochenschr* **114**, 268–275, 1989.
- [O'MO91] O'MOORE, R. and ENGELBRECHT, R.: The Evaluation of Medical Decision Support and Expert Systems: Reflections on the Literature. In: Talmon, J. L. and Fox, J. (eds.): *Knowledge Based Systems in Medicine. Methods, Applications and Evaluations. Proceedings of the Workshop "System Engineering in Medicine"*, 263–273, Berlin, Springer, 1991.
- [PEL93] PELIKAN, E., BOHNDORF, K., TOLXDORFF, T., ZARRINAM, D., and WEIN, B.: Computer-Assisted Diagnosis of Bone Tumors. In: Lemke, H. U., Inamura, K., Jaffe, C. C., and Felix, R. (eds.): *Computer Assisted Radiology, Computergestützte Radiologie*, 630–634, Berlin, Springer, 1993.
- [PUP87] PUPPE, F.: *Diagnostisches Problemlösen mit Expertensystemen*. Berlin, Springer, 1987.
- [SER73] SEROV, S. F., SCULLY, R. E., and SOBIN, L. H.: *Histological Typing of Ovarian Tumours. International Classification of Tumours, No. 9*. Geneva, World Health Organisation, 1973.
- [SHO76] SHORTLIFFE, E. H.: *Computer Based Medical Consultations: MYCIN*. New York, Elsevier, 1976.
- [SOM89] SOMMERVILLE, I.: *Software Engineering*. Reading, Addison-Wesley, 1989.
- [SPI83] SPIEGELHALTER, D. J.: Evaluation of clinical decision aids, with an application to a system of dyspepsia. *Stat Med* **28**, 207–216, 1983.
- [WYA91] WYATT, J. C. and SPIEGELHALTER, D. J.: Evaluating Medical Expert Systems: What to Test, and How? In: Talmon, J. L. and Fox, J. (eds.): *Knowledge Based Systems in Medicine: Methods, Applications and Evaluations. Proceedings of the Workshop "System Engineering in Medicine"*, 274–290. Berlin, Springer, 1991.

Korrespondenzadresse: Prof. Dr. Thomas Tolxdorff, Institut für Medizinische Statistik und Informationsverarbeitung, Abteilung Medizinische Informatik, Freie Universität Berlin, Hindenburgdamm 30, 12200 Berlin

Eingang des Manuskripts 11. 4. 94

Eingang der revidierten Fassung 30. 8. 94

BUCHBESPRECHUNGEN/BOOKREVIEWS

GAUS, W.

**Berufe im Archiv-, Bibliotheks-, Informations- und Dokumentationswesen
Ein Wegweiser zur Ausbildung**

3. überarb. Aufl., 1994, 319 S., DM 32.—, ISBN 3-540-58408-0
Springer-Verlag, Berlin—Heidelberg—New York

Die 1986 erschienene erste Auflage war so erfolgreich, daß sie bereits nach einem Jahr nachgedruckt werden mußte. In den zwei Jahren seit dem Erscheinen der 2. Auflage hat sich so viel geändert, daß etwa die Hälfte der Detailangaben für diese 3. Auflage abzuändern war. Ausbildungsgänge wurden ganz oder vorübergehend eingestellt, neue Ausbildungsgänge wurden eröffnet, nicht nur in den neuen Bundesländern ergaben sich viele Strukturveränderungen. Die neuen Postleitzahlen sowie Änderungen bei Telefon- und Telefaxnummern waren vorzunehmen. Die abgestimmten Angaben beruhen auf dem Stand von März 1994. Die Berufsbilder, Tätigkeitsbereiche und Ausbildungsgänge im Gesamtgebiet des „Informationswesens“ sind durch außerordentliche Vielfalt gekennzeichnet. In dem vorliegenden Band werden Überblicke über 29 Berufe, 75 Ausbildungsstätten mit zusammen 104 Ausbildungsgängen gegeben.

Diese informative Zusammenstellung wendet sich insbesondere an Schülerinnen und Schüler sowie auch an Studentinnen und Studenten, die vor der Berufswahl stehen. Sie ist aber auch für die in diesem Berufsfeld Tätigen ein nützliches Nachschlagewerk.

H. Geidel, Stuttgart

WEITKUNAT, R.

Deskriptive Statistik in SAS – Einführung mit Beispielen aus der Medizinischen Biometrie und Epidemiologie

1994, 250 S., kt. DM 78.—, ISBN 3-437-50371-5

SAS-Neulingen würde ich dieses Buch nicht als Einstiegslektüre empfehlen. Diese wären wahrscheinlich schon auf den ersten Seiten des 1. Kapitels „Einführung“ schwer gefordert, wo als erstes eine dokumentarische – praktisch sicher sinnvolle – Organisation von SAS-Programmen vorgestellt wird. Ist es wirklich zu Anfang eines solchen Buches wichtig, auf S. 6 (quasi als eins der ersten Topics) zu beschreiben, daß bzw. wie man den Grafikfont auf Swiss und die Farben auf S/W einstellt? Oder, daß als erstes angewandtes SAS-Programmierbeispiel ein interaktives Fensterprogramm zur Ausgabe von kritischen Werten vorgestellt wird?

Fasziniert hat mich der didaktisch wie rhetorisch ungewöhnliche Stil des Buches, z. B. S. 11: „Gemäß der didaktischen Strategie dieses Buches, Dinge nicht zu erklären, erfolgt an dieser Stelle . . .“. Meinen StudentInnen der Biowissenschaftlichen Dokumentation würde ich dieses Buch während ihres Studiums zum SAS-Erlernen nicht empfehlen, wohl aber als Lektüre für ihr späteres Berufsleben. Denn es ist durchgängig ein Buch, in dem man, wenn man bereits SAS-Erfahrung besitzt, etliche praktische Tips und interessante Anregungen finden kann. Dieses Buch zu lesen war für mich so ähnlich, wie einmal einem erfahrenen SAS-benutzenden Kollegen „über die Schulter schauen dürfen“. Und das fand ich sehr interessant! Zum Beispiel auf S. 34ff., wo die in SAS jeweils einzeln berechneten und gedruckten Häufigkeitsauszählungen mehrere Variablen in einer gemeinsamen übersichtlichen Tabelle zusammengestellt bzw. -programmiert werden. Guter Tip! Ich selber konnte dem Buch für meine SAS-Lehrveranstaltung an der Fachhochschule wichtige neue Anregungen und Impulse entnehmen.

Ein – meines Erachtens – wichtiger Schwerpunkt des Buches ist weniger die deskriptive Auswertung von Daten, als die detaillierte, praxisbezogene Beschreibung des gesamten Datenhandlings einer medizinischen Studie. Dieses nimmt in der Praxis oftmals zeitlich den Großteil der (Programmier-)Arbeit in Anspruch. Und die dafür erforderlichen Tips und Kniffe bleiben in den herkömmlichen SAS-Einführungen außen vor. Hier hat der Autor mit seinem Buch wirklich eine „Marktlücke“ gefüllt! Außerdem ist andererseits auch gerade für den 2. Teil des Buches die Bezeichnung „deskriptive Statistik“ eher ein Understatement, findet dort die logistische Regression ebenso wie die Analyse von Überlebenszeiten (Kaplan-Meier, Log-Rank-Test) ihren Platz.

Kira Schulz, Hannover

GRIESER, G.

Ein Krankenhaus-Informations- und Kommunikationssystem zur Unterstützung der Klinik

Ein Leitfaden. Gesellschaft für Systemberatung im Gesundheitswesen GSbG, Schmidt & Klaunig, Kiel 1994, 239 Seiten, 5 Abb., 10 Tab., brosch. DM 35.-, ISBN 3-88312-090-1

Der Begriff Krankenhaus-Informationssystem (KIS), oder im angelsächsischen Sprachraum Hospital Information System (HIS), ist praktisch so alt wie die medizinische Datenverarbeitung – bevor sie Medizinische Informatik hieß. Dieser Bereich war von Beginn an ein beliebter Tummelplatz der entsprechenden Fachwelt, denn hier konnte man einerseits etwas erreichen (in Hinsicht auf die in aller Welt im Argen liegende Kommunikations- und Informationsstruktur im Krankenhaus) und andererseits war auch schon in den Anfangsjahren der Computerei dieses Werkzeug sehr viel besser als jedes andere. Der Erfolg war damit praktisch immer vorprogrammiert.

Wie kommt es also, daß 30 oder 35 Jahre nach diesen Anfängen immer noch Bücher und Systeme erscheinen, die neu sind, bzw. vorgeben neu zu sein? Mehrere Gründe können herangezogen werden. Der Hauptgrund ist vermutlich darin zu sehen, daß es unsere Wissenschaft der Medizinischen Informatik immer noch nicht geschafft hat, ein anerkanntes Gesamtmodell eines Krankenhaus-Informationssystems zu erstellen. So ist natürlich mit der sich drastisch verändernden Technik jedes „neue“ System wirklich wieder neu. Ein weiterer Grund ist natürlich auch darin zu sehen, daß die Fachleute und Entwickler solcher Systeme immer weiter hinzulernen und aufgrund reicher Erfahrungen früher gemachte Fehler nicht wiederholen sondern diese Erfahrungen in die neuen Systeme einbauen (durch nichts lernt man besser und schneller, als durch Fehler).

Der Autor des vorliegenden KIS-Buches ist auf diesem Gebiet ein „alter Hase“ – wie man ihn ohne despektierlich zu sein wohl nennen darf. Er nennt das hier Abgehandelte auch nicht mehr nur KIS, sondern KIKS, weil man allgemein in den letzten Jahren eingesehen hat, daß Information ohne ein entsprechendes Kommunikationssystem nicht verwertbar ist. Er hat, ohne sich um das „Modell“ zu bemühen, eine Zusammenstellung aller Funktionen und Aufgaben, die in einem Krankenhaus anfallen und etwas mit Kommunikation und Information zu tun haben – wo ist das nicht der Fall? – gemacht, und diese jeweils mehr oder weniger ausführlich erläutert. Der Autor hat sein Buch mit dem Untertitel „Ein Leitfaden“ versehen und darin liegt eben auch die Qualität des Werkes. Die oben so genannte „Zusammenstellung“ ist eben mehr als nur das, was man gemeinhin mit diesem Wort umschreibt. In ihr sind die Systematik bzw. Struktur, die Vollständigkeit, die Begriffsbestimmungen, die Einordnungen (sogar Integration) in die Arbeitsprozesse und die Definitionen von Merkmalen und Merkmalsausprägungen enthalten. Diese Inhalte machen das Buch wirklich zu einem „Leitfaden“.

Das Buch ist in 5 Teile gegliedert und hat einen Anhang, in dem viele Einzeldefinitionen und strukturelle Gliederungen dargestellt sind, die im Text sicher das Lesen erschwert hätten. Aber soll man dieses Buch „Lesen“? Ist ein „Leitfaden“ zum Lesen gedacht? Allenfalls zum Durchblättern, um dann zu wissen, daß man hier nachlesen kann, wenn man ein konkretes Segment oder Modul eines KIKS entwickeln will. Das ist die eigentliche Stärke dieses Werkes, die der Autor natürlich auch so gesehen hat, das Nachschlagewerk, die Anleitung nicht nur zur Arbeit sondern davor auch zur Analyse, die dadurch auch gegebene Möglichkeit der Evaluation einerseits eines bestehenden KIKS und andererseits eines Modells bzw. eines Planes für ein KIKS.

Teil I enthält eine Einführung, in der die Kommunikationsstrukturen und die dadurch zu steuernden Ströme in der Klinik (Patientenstrom, Personalstrom, Materialstrom, Finanzstrom und Informationsstrom) schon sehr detailliert dargestellt werden. Ein Strom, der in großen Häusern auch gesteuert werden muß, ist vergessen worden, der Besucherstrom. Hier werden auch bereits die Hauptanforderungen in Hinsicht auf die technische Durchführung (z. B. Datenbankdesign) angesprochen.

Im zweiten Teil werden Organisations- und Funktionalitätsstrukturen sehr detailliert erläutert. Der Betriebswirt würde in diesem Bereich die strikte Trennung zwischen Aufbau- und Ablauforganisation vermissen. Diese Durchmischung ist auch im dritten Teil wieder zu konstatieren. Die Vollständigkeit der Darstellung der einzelnen Funktionen und Funktionsträger ist bis auf zwei Modulen, Hygiene und Patienteninformation, offensichtlich gegeben.

Der dritte Teil enthält das „Fachliche Konzept“ eines Krankenhauses und seines Kommunikations- und Informationssystems. Hier werden auch das sich wandelnde Verhältnis der Funktionsbereiche ärztlicher Dienst und pflegerischer Dienst sowie die sich durch Einsatz neuer Medien qualitativ verbessernde Dokumentation angesprochen. Auf ein wünschenswertes und sinnvolles Zusammenlegen aller „Dokumentationen“ zu einer „Patientengeschichte“ wird nicht eingegangen (vielleicht in der zweiten Auflage?). Auch die Krankenhausesseelsorge ist hier aufgeführt.

Datenschutz und Datensicherheit im Krankenhaus sind im vierten Teil ausführlich besprochen und als Designelement für Kommunikationssysteme beschrieben. Diesem Kapitel ist wohl kaum etwas hinzuzufügen. Für diesen Teil ist der Autor auch international der Garant.

Der letzte Teil beschäftigt sich mit den technischen Möglichkeiten des Designs und der Entwicklung eines KIKS. Natürlich basiert dieser Teil auf der modernen Technologie von Netzen mit

Arbeitsplatzsystemen. Hier wird alles kurz angerissen, was jetzt „state of the art“ ist. Nach der Lektüre dieses Werkes darf eigentlich kein Entwickler mehr „gestrige“ Dinge tun.

Was kann man dem Autor zusätzlich für eine sicher wünschenswerte und erwartete zweite Auflage raten? Den Begriff „Integration“ in den Index aufnehmen (im guten Glossar ist er enthalten) und dementsprechend durch jeweils wenige Sätze im Text entsprechend zum Ausdruck bringen. Beispiele: 1. Verwendung der medizinischen Dokumentation auch als Quelle der „Leistungserfassung“ – es gibt bereits ein System auf dem Markt, das genau dieses macht. 2. Verwendung der medizinischen Dokumentation (ärztliche und pflegerische), der Funktionsraumbelegung und der Personaleinsatzpläne für das Hygiene-System. Im Buch beschränkt sich der Einsatz von Karten aller Art im Krankenhaus auf die schon im Routineeinsatz befindliche Krankenversicherungskarte (im Index nicht korrekt Krankenversicherungskarte genannt), ein Zukunftsaspekt für die „medizinische Karte“ würde sich in einer nächsten Auflage auch gut machen.

Das Buch enthält hinter jedem der 5 Teile ein Literaturverzeichnis, das nach „Vancouver“ strukturiert ist. Man kann sicher geteilter Meinung sein, ob ein gesamtes Literaturverzeichnis in „numbered“ für das weitergehende Arbeiten komfortabler und übersichtlicher wäre. Ein paar mehr oder weniger erhebliche Fehler haben sich in die bibliographischen Zitate auch eingeschlichen, z. B.: Bücher werden als „Autoren“-Bücher deklariert, die „Herausgeber“-Bücher sind; ausgerechnet die „Methods“ werden einmal falsch als „Med. Inform. Med.“ abgekürzt; ein Buch wurde einem falschen Verlag zugeordnet – „Kleinigkeiten“, aber bei diesem Autor nicht erwartet. Der Index ist ausführlich und zielgenau, ein wichtiges Instrument der Benutzung des Werkes als Arbeitsgrundlage.

Wer sollte das Buch kaufen? Wie schon angesprochen, ist das vorliegende Werk ein Arbeitsbuch für Design und Entwicklung – insbesondere auch für Firmen, die Krankenhaus-Software anbieten oder anbieten wollen. Aber auch Studierende der Medizinischen Informatik erwerben hier ein Buch, das für sie nicht nur ein Leitfaden für das Lernen zur Prüfung sondern auch eine sehr nützliche Quelle für ihren späteren Arbeitsplatz sein kann. Das Werk ist sicher auch gut in der Lehre und weiterführenden Ausbildung zu verwenden.

C. O. Köhler, Heidelberg

KRAUSE, A.

Computerintensive statistische Methoden: Gibbs Sampling in Regressionsmodellen

XI 198 S., 44 Abb., 39 Tab., DM 68.—, ISBN 3-437-50372-3

Gustav Fischer, Stuttgart—Jena—New York 1994.

Seit Anfang der neunziger Jahre werden Gibbs Sampling Techniken in der statistischen Literatur diskutiert und gewinnen immer mehr an Beachtung, was auch durch die deutlich ansteigende Anzahl von Veröffentlichungen zu diesem Thema über die letzten Jahre verdeutlicht wird. Der tatsächliche Einsatz des Gibbs Samplers in der statistischen Datenanalyse ist allerdings noch immer auf einen sehr kleinen Kreis von Anwendern beschränkt. Um die Möglichkeiten dieser Technik einem weiteren Kreis bekanntzumachen, sind allgemein verfügbare Programme und gerade auch Anwendungsbeispiele nötig, die die Einsatzmöglichkeiten des Gibbs Samplers illustrieren. Durch das Buch von Andreas Krause erhält man einen ausführlichen Einstieg in die praktische Verwendung des Gibbs Samplers, wobei auch auf theoretische Aspekte eingegangen wird. Im Buch werden für ein konkretes Anwendungsbeispiel die flexiblen Einsatzmöglichkeiten des Gibbs Sampling Verfahrens im Rahmen linearer und nichtlinearer Modelle beschrieben, wobei speziell das Problem der Tobit-Regressionsmodelle behandelt wird. In Kapitel 1 werden Grundlagen der Bayes-Statistik beschrieben, Kapitel 2 behandelt Monte-Carlo-Techniken und stellt unter anderem das Gibbs Sampling Verfahren vor. Daten und biologischer Hintergrund des im Buch verwendeten durchgehenden Beispiels sind unter anderem Thema von Kapitel 3. Dort werden auch Beispiele verschiedener Ansätze zur Anwendung des Gibbs Sampling Prinzips bei unterschiedlichen Regressionsansätzen beschrieben und auf den Datensatz angewandt. Kapitel 4 stellt einige grafische Methoden zur Kontrolle der Modelladäquatheit vor. Abschließend wird in Kapitel 5 die Anwendung des EM-Algorithmus mit den beschriebenen Methoden des Gibbs Sampling verglichen.

Insgesamt bietet das Buch einen übersichtlich gehaltenen Einstieg in die Thematik des Gibbs Samplers anhand eines speziellen praktischen Anwendungsfalls. Dabei bietet es eine gute Balance zwischen Theorie und praktischer Verwendbarkeit der Verfahren, wobei mehr Wert auf die praktische Einsetzbarkeit der Methoden gelegt wird. Durch die Verwendung einer mathematisch-statistischen Programmierumgebung (S-Plus) und die zum Teil im Buch abgedruckten Programmlistings lassen sich die Methoden oft gut nachvollziehen. Im Zusammenhang mit Programmen wie das von Spiegelhalter et al. entwickelte Programm BUGS (Bayesian inference Using Gibbs Sampling), das frei verfügbar ist (etwa über e-mail an bugs@mrc-bsu.cam.ac.uk) und u. a. auch eine Schnittstelle zur S-Language anbietet, wird so sicherlich zu einer stärkeren Verbreitung des Gibbs Samplers in der statistischen Praxis beigetragen.

A. Benner, Heidelberg

TABLE OF CONTENTS

Presidential Invited Address: The world of biometry	L. Billard	899
A general cohort analysis method	M. Aksland	917
A random-effects ordinal regression model for multilevel analysis	D. Hedeker and R. D. Gibbons	933
An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random	M. G. Kenward, E. Lesaffre, and G. Molenberghs	945
A parametric model for cluster correlated categorical data	S. G. Meester and J. MacKay	954
Exact analysis for paired binary data	K. F. Hirji	964
Multivariate survival analysis using piecewise gamma frailty	M. C. Paik, W.-Y. Tsai, and R. Ottman	975
Iterative generalized least squares for meta-analysis of survival data at multiple times	K. B. G. Dear	989
Modelling progression of CD-4 lymphocyte count and its relationship to survival time	V. De Gruttola and X. M. Tu	1003
A nonparametric analysis of the transmission rate of human immunodeficiency virus from mother to infant	W.-Y. Tsai, J. J. Goedert, J. Orazem, S. H. Landesman, A. Rubinstein, A. Willoughby, and M. H. Gail	1015
Evaluation of experiments with adaptive interim analyses	P. Bauer and K. Köhne	1029
Exact permutational tests for group sequential clinical trials	C. R. Mehta, N. Patel, P. Senchaudhuri, and A. Tsiatis	1042
The use of logistic models for the analysis of codon frequencies of DNA sequences in terms of explanatory variables	K. K. Amfoh, R. F. Shaw, and G. E. Bonney	1054
Robust variance estimation for the case-cohort design	W. E. Barlow	1064
Segregation analysis of case-control data using generalized estimating equations	A. S. Whittemore and G. Gong	1073
Operating characteristics of a rank correlation test for publication bias	C. B. Begg and M. Mazumdar	1088
Causal nonresponse models for repeated categorical measurements	M. R. Conaway	1102
Influence diagnostics for generalized linear measurement error models	Y. Zhao, A. H. Lee, and Y. V. Hui	1117
Pooled population parameter estimates from mark-recapture data	J. W. Hargrove and C. H. Borland	1129
Shorter Communications		
On inconsistency of Breslow's estimator as an estimator of the hazard rate in the Cox model	D. Burr	1142
Guidelines for monitoring efficacy and toxicity responses in clinical trials	R. J. Cook and V. T. Farewell	1146
On a truncation-flexible repeated significance test	D. Edelman	1153
On the likelihood ratio test statistic for the number of components in a normal mixture with unequal variances	Z. D. Feng and C. E. McCulloch	1158
A note on the bias of estimators with missing data	A. Rotnitzky and D. Wypij	1163
Variance components testing in the longitudinal mixed effects model	D. O. Stram and J. W. Lee	1171
The Consultant's Forum		
Heterogeneity models of disease susceptibility, with application to diabetic nephropathy	P. Hougaard, P. Myglegaard, and K. Borch-Johnsen	1178
Testing for segregation distortion in the HLA complex	K. Jin, T. P. Speed, W. Klitz, and G. Thomson	1189
Statistical analysis of food webs	P. A. Murtaugh	1199
Regression analysis of censored and truncated data: Estimating reporting-delay distributions and AIDS incidence from surveillance data	M. Pagano, X. M. Tu, V. De Gruttola, and S. MaWhinney	1203
A cautionary note on applying scores in stratified data	M. J. Podgor and J. L. Gastwirth	1215
Correspondence		1219
Book Reviews		1224
Papers to be published		1238
Acknowledgements		1239
The International Biometric Society		1241

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

Brief des Präsidenten

Liebe Kollegen und Kolleginnen,

wichtige neue Entwicklungen kann ich ihnen heute aus dem Präsidium der GMDS berichten. Zwei dringende Aufgaben sind von Herrn Schaefer und Frau Steck übernommen worden: Zum einen wird Herr Schaefer die wesentlichen Ausarbeitungen der GMDS zu den verschiedensten Themen in einer Liste zusammenstellen, die laufend aktualisiert allen Mitgliedern über das Mitgliederverzeichnis zur Verfügung gestellt werden soll. Auf diese Art und Weise soll die wertvolle inhaltliche Arbeit, die im Laufe der Jahre durch Kommissionen und Gremien der GMDS geleistet worden ist, auch leichter für jedes Mitglied abrufbar werden. Technologisch gesehen geht Frau Steck einen Schritt weiter: In Zusammenarbeit mit verschiedenen Kollegen aus der GMDS entwirft sie eine Vorstellung über einen elektronischen Informationsdienst über und innerhalb der GMDS. Die Hauptaufgabe wird dabei sein, nicht ein paar Bildschirmseiten zur Verfügung zu stellen, sondern zu versuchen, die interne Kommunikation innerhalb der GMDS über elektronische Medien abzuwickeln. Auf diese Art und Weise soll erreicht werden, daß die Arbeit sämtlicher Arbeitskreise und Arbeitsgruppen für die Mitglieder aktuell transparent wird und nicht jeweils die Jahresberichte abgewartet werden müssen, um zu erfahren, welche Termine, Veranstaltungen, Ergebnisse usw. vorliegen. Im Namen des Präsidiums bedanke ich mich bei all denen Kollegen der GMDS, die diese Arbeit unterstützen mit Anregungen, Beispielen und Implementationen.

Während der letzten Präsidiumssitzung der GMDS ist auch ein wesentlicher Schritt in Richtung Förderung des internationalen Austausches von GMDS-Mitgliedern getan worden. Das GMDS-Präsidium hat ein erstes Gutachten zum Aufbau einer "GMDS-Stiftung" entgegengenommen, die diese Förderaufgabe wahrnehmen soll. Es ist daran gedacht, eine Stiftung zu etablieren und deren Leitung mit dem Präsidium zu verzahnen. Die Stiftung soll von sich aus in der Lage sein, Spenden zu akquirieren. Als Startkapital für die Stiftung sollen insbesondere Überschüsse der letzten GMDS-Tagung verwandt werden. Durch die Einrichtung einer Stiftung zur Förderung des internationalen Austausches junger Wissenschaftler und Studenten soll auch die Attraktivität der GMDS für unsere jungen Kollegen an den Hochschulen gesteigert werden. Es ist vorgesehen, die weiteren Schritte zur Einrichtung der Stiftung so schnell wie möglich voranzutreiben.

Bildung und Ausbildung in Biometrie

Unter diesem Titel hat eine Arbeitsgruppe unter Leitung von Herrn Victor und Herrn Holle (Heidelberg) ein umfangreiches Papier erstellt, das auf der Präsidiumssitzung im Mai diesen Jahres erörtert wird. Ähnlich wie bereits für die Medizinische Informatik geschehen, wird hier für die Bundesrepublik Deutschland der Bedarf an biometrischer Ausbildung und Weiterbildung aufgezeigt. Mit diesem beispielhaften Papier soll auch in die gegenwärtige Diskussion um die Bedeutung der durch die GMDS vertretenen Fächer eingetreten werden: Durch die Sparbeschlüsse der Länderfinanzministerien sind zur Zeit die sogenannten kleinen Fächer gefährdet. Um so wichtiger ist es, jederzeit in der Lage zu sein, nachweisen zu können, welche Bedeutung den Fächern und einer qualifizierten vertikal und horizontal gestaffelten, qualifizierten Berufsausübung zukommt. Die von Herrn Victor vorgelegte Schrift soll wegen ihrer großen Bedeutung in der Schriftenreihe der GMDS als Einzelpublikation veröffentlicht werden.

DIMDI

Nach langjähriger Tätigkeit ist Herr Dr. Fritz als Präsident des DIMDI in den Ruhestand getreten. Das Deutsche Institut für Medizinische Dokumentation und Information hat während seiner Präsidentschaft internationale Anerkennung gefunden und ist das größte Fachinformationszentrum in der Bundesrepublik geworden. Neben der Aufgabe, Literaturdatenbanken und Fachinformationsdienste anzubieten, erfüllt das DIMDI auch vielfältige hoheitliche Aufgaben, in dem es im Zuständigkeitsbereich des Bundesministeriums für Gesundheit gesetzlich vorgeschriebene Spezialdokumentationen durchführt. Zunehmende Bedeutung in der Arbeit des DIMDI bekommt die internationale fachliche Kooperation, die insbesondere mit entsprechenden Einrichtungen in Frankreich und den USA gepflegt wird. Bis zur Übernahme der Nachfolge von Herrn Dr. Fritz durch Herrn Dr. Schweim aus Berlin im Herbst 1995 wird Herr Kurzwelly die Bundesbehörde leiten. Die GMDS bedankt sich bei Herrn Dr. Fritz für die vielfältig geleistete Pionierarbeit und wünscht dem DIMDI als "unserem" Fachinformationszentrum gutes Gedeihen unter der neuen Leitung. Während der GMDS-Tagung in Bochum wird in einer speziellen Sitzung am Dienstag zum Thema Information Highway auch der neuen Leitung des DIMDI Gelegenheit gegeben, die zukünftige Entwicklung dieses für uns alle so wichtigen Bundesamtes aufzuzeigen.

Wahlen

In der GMDS stehen wichtige Wahlen zu den Fachbereichsausschüssen und dem Amt des ersten Vizepräsidenten an. Es ist eine Vielzahl von Wahlvorschlägen eingegangen. Es ist erfreulich wie viele Kollegen bereit waren, für Ämter in der GMDS zu kandidieren. Auffällig ist jedoch auch, wie viele Kollegen durch die Anforderungen des Gesundheitsstrukturgesetzes so stark eingespannt sind, daß es ihnen nicht möglich war, auch für verantwortungsvolle Tätigkeiten in der GMDS zu kandidieren. Um so mehr möchte ich alle Mitglieder der GMDS im Namen des Präsidiums bitten, von der Briefwahlmöglichkeit Gebrauch zu machen und so durch rege Beteiligung an der Gestaltung des GMDS-Geschehens in den nächsten Jahren aktiv mitzuwirken.

G-7

Die Aktivitäten des amerikanischen Vizepräsidenten Al Gore zum Thema Information Highway sind ja bekanntlich von Bundeskanzler Kohl und anderen Ministerpräsidenten der G-7-Länder aufgegriffen worden. Ein ad hoc-Programm ist zusammengestellt worden, das leider den Fehler besitzt, keine Finanzzuweisungen bekommen zu haben. Nichtsdestotrotz haben Herr Stein vom Bundesministerium für Gesundheit, Herr van Eimeren als Vertreter der Bundesrepublik in dem Telematik Management Board, Herr Vöge von der Deutschen Telekom und der Berichterstatter als Präsident der GMDS den Auftrag bekommen, die anlaufenden Aktivitäten im Bereich Gesundheitswesen aufeinander abzustimmen. Es ist im Moment noch zu früh über Einzelheiten und Erfolgsaussichten der Vorhaben zu berichten. Dies soll jedoch in einer einem Beitrag in der oben bereits angesprochenen Session Information Highway während der GMDS-Jahrestagung in Bochum durch einen kompetenten Vertreter der Europäischen Union in Brüssel nachgeholt werden.

Jahrestagung GMDS Bochum

Die abschließende Sitzung des Programmkomitees unter der Leitung von Herrn Trampisch hat in einer eintägigen intensiven Arbeitssitzung zu einem außerordentlich ausgewogenen und aktuell abgestimmten Programm geführt. Zum ersten Mal werden während der GMDS-Tagung für alle Mitglieder nachvollziehbar wissenschaftliche Beiträge, Erfahrungsberichte, Industriebeiträge und eingeladene Übersichtsreferate voneinander geschieden, so daß dem einzelnen Mitglied die Auswahl unter dem reichlichen Angebot an Vorträgen, Postern, Softwaredemonstrationen, etc. leichter fällt. Das Präsidium hofft so, die Attraktivität der

GMDS weiter erhöhen zu können. Erstmals wird auch nach amerikanischen Vorbild eine Meet-the-expert-Session durchgeführt werden. Herr Kuhn, Ulm, hat diese vorbereitet. Bitte nutzen Sie dieses neue Kommunikationsmedium während der GMDS-Tagung ausdrücklich, wir sind auf ihre Kommentare gespannt, die der zukünftige erste Vizepräsident ähnlich wie bei der Dresdner Tagung erneut auswerten wird. Die Teilnehmerbefragung während der GMDS-Tagungen sollen in Zukunft wesentlich dazu mitbeitragen, die Tagungsthemen und Modalitäten der Tagung auf die Bedürfnisse der GMDS-Mitglieder abzustimmen.

Mit herzlichen kollegialen Grüßen

Ihr

O. Rienhoff

* * * * *

Allgemeine Berichte

Bericht aus dem Fachbereich *Medizinische Informatik* der GMDS

Ende Februar fand erneut eine Sitzung des Fachausschusses Medizinische Informatik statt. Dies war die vorletzte Sitzung des Fachausschusses, der in den kommenden Monaten ja erneut gewählt werden wird. Schon auf dieser Sitzung informierte Prof. Haux darüber, daß er bei diesen Wahlen nicht mehr kandidieren werde.

Aus dem Fachbereich ist zu berichten:

- Die Kooperation von GI und GMDS ist nun, nach Veröffentlichung des Kooperationsvertrages, sowohl im Informatik-Spektrum als auch in der Zeitschrift 'Informatik, Biometrie und Epidemiologie in Medizin und Biologie' auch formell perfekt. Auf der gemeinsam mit der schweizerischen Informatikergesellschaft durchgeführten Jahrestagung der GI vom 18. - 20.09.1995 in Zürich wird es ein Fachgespräch mit dem Titel "Informationsverarbeitung und Kommunikation im Gesundheitswesen: Planung und Bewertung von Krankenhausinformationssystemen" geben.
- Unter der Herausgeberschaft von Herrn Prof. Seelos ist ein Sonderheft der it+ti zur Medizinischen Informatik erschienen.
- Der Praxis-, Studien- und Forschungsführer Medizinische Informatik, Biometrie und Epidemiologie wird im September 1995 erneut aufgelegt.
- Von der Projektgruppe Datenschutz in Krankenhausinformationssystemen wurde ein Memorandum über den Datenschutz in Krankenhausinformationssystemen vorgelegt.
- Eine von der Projektgruppe Administrative Verfahren in Krankenhausinformationssystemen veranstaltete Arbeitstagung zu Krankenhausinformationssystemen ist auf ein großes Echo gestoßen.
- In der Arbeitsgruppe Medizinische Informatik in der Pflege gab es eine Fülle von Aktivitäten.
- Die Arbeitsgruppe Archivierung von Krankenunterlagen veranstaltete Ende März eine vielbeachtete und vielbesuchte Arbeitstagung in Augsburg.

- Herr Prof. Gaus hat die Leitung der Projektgruppe Ausbildung in Medizinischer Informatik abgegeben. Der Fachausschuß hat einstimmig beschlossen, Herrn Prof. Haas (Dortmund) zu bitten, die Nachfolge zu übernehmen.
- Der Fachausschuss richtete eine Projektgruppe Raumbezogene Informationssysteme ein, die von Prof. Appelrath geleitet werden wird.
- Es hat sich ein gemeinsamer Arbeitskreis Neurologie der Deutschen Gesellschaft für Neurologie und der GMDS unter der Leitung von Prof. Spitzer gebildet.

PD Dr. A. Winter, Heidelberg
Schriftführer des Fachausschusses Medizinische Informatik

* * * * *

Projektgruppe Ausbildung in Medizinischer Informatik

Die Veränderungen im deutschen Gesundheitswesen und die immer noch rasche Innovationsfolge in der Informatik erfordern immer wieder die Anpassung bestehender Ausbildungskonzepte. Dies nicht nur aus der Sicht des technologisch Mach- und Denkbaren, sondern auch orientiert an der Bedarfslage für Fachkräfte der Medizinischen Informatik in Einrichtungen des Gesundheitswesens. Neben einer Bestandsaufnahme bestehender Ausbildungsangebote auf den verschiedenen Ausbildungsebenen bzw. deren Fortschreibung ist die Erhebung des tatsächlichen qualitativen und quantitativen Bedarfes und die Formulierung bedarfsorientierter und an den neusten Entwicklungen orientierter Ausbildungskonzeptionen unabdingbar.

Alle an einer Mitarbeit in der Projektgruppe interessierte/n Kolleg(inn)en werden gebeten, sich baldmöglichst zur Abstimmung der zukünftigen Aktivitäten sowie eines ersten Sitzungstermines der Projektgruppe baldmöglichst schriftlich zu melden bei:

Prof. Dr. Peter Haas, Fachhochschule Dortmund, Postfach 105018, 44047 Dortmund, Tel.: 0231/9112-309, Telefax: 0231/9112-230 e-mail: haas@medinf.informatik.fh-dortmund.de.

Vorschläge zur weiteren Arbeit sind willkommen.

Wir freuen uns auf zahlreiche Rückmeldungen.

Prof. Dr. P. Haas, FH-Dortmund Prof. F.-J. Leven, FH Heilbronn/Universität Heidelberg

* * * * *

Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. -
Fachbereich 'Medizinische Informatik'
Gesellschaft für Informatik e.V. - Fachausschuss 4.7 'Medizinische Informatik'
Methoden und Werkzeuge für das Management von Krankenhausinformationssystemen
Arbeitskreis des GI-Fachausschusses 4.7 'Medizinische Informatik'
Projektgruppe in dem Fachbereich 'Medizinische Informatik' der GMDS
Leitung: Alfred Winter (Heidelberg)
stv. Leitung: Roland Zimmerling (Oldenburg)

Das systematische Management von Krankenhausinformationssystemen ist, wie bei anderen betrieblichen Informationssystemen auch, notwendig, damit eine angemessene Durchführung der in einem Krankenhaus erforderlichen informationsverarbeitenden Verfahren unter Berücksichtigung einer wirtschaftlichen Betriebsführung ermöglicht wird. Management von Krankenhausinformationssystemen umfaßt die

- Gestaltung (d.h. sowohl den Aufbau als auch die Weiterführung) und die
- Überwachung

von Krankenhausinformationssystemen. Dabei sind die branchenspezifischen Belange von Krankenhäusern zu berücksichtigen, um einen angemessenen Betrieb der Krankenhausinformationssysteme und damit eine adäquate Informationsverarbeitung für die Patientenversorgung und - im universitären Bereich - für die medizinische Forschung und Lehre zu erzielen. Dieses Ziel kann nur durch eine systematische Vorgehensweise erreicht werden. Die hierzu erforderliche (Weiter-) Entwicklung von Methoden und Werkzeugen für das systematische Management von Krankenhausinformationssystemen umfaßt auch die Entwicklung von Referenzmodellen für Krankenhausinformationssysteme und ist eine anspruchsvolle Aufgabe sowohl für die Informatik als auch für die Medizinische Informatik; ein interdisziplinärer Erfahrungs- und Wissensaustausch zwischen Informatik, Medizinischer Informatik und Wirtschaftsinformatik bildet die Voraussetzung.

Zur Intensivierung dieses Erfahrungs- und Wissensaustauschs wurde innerhalb der Gesellschaft für Informatik (GI) der Arbeitskreis 'Methoden und Werkzeuge für das Management von Krankenhausinformationssystemen' des Fachausschusses 4.7 eingerichtet, der gleichzeitig Projektgruppe im Fachbereich Medizinische Informatik der Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) ist. Der Arbeitskreis soll ein aktives Bindeglied für die entsprechenden Gliederungen von GI und GMDS sein.

Sollten Sie sich für eine Mitarbeit in dem Arbeitskreis bzw. der Projektgruppe interessieren, setzen Sie sich bitte mit dem Leiter oder stv. Leiter in Verbindung. Es wäre hilfreich, wenn Sie bei dieser Gelegenheit Ihre eigenen Interessen und Vorstellungen kurz skizzieren könnten. Bitte vergessen Sie auch nicht die Angabe Ihrer e-mail-Adresse (falls vorhanden).

Dr. Alfred Winter
Klinikum der Universität Heidelberg
Abt. Medizinische Informatik
Im Neuenheimer Feld 400
69120 Heidelberg
Tel.: 06221/56-7497, -7483
e-mail: Dr._Alfred_Winter@krzmail.krz.uni-heidelberg.de

Dr. Roland Zimmerling
Kuratorium OFFIS e.V.
FB 9 Betriebswirtschaftliche Informatik
Westerstraße 10-12
26121 Oldenburg
Tel.: 0441/9707-4, -393
e-mail: zimmerli@hera.offis.uni-oldenburg.de

Ankündigungen

Veranstaltungen der Akademie Medizinische Informatik

1. Seminar "Krankenhaus und Qualitätssicherung"

Termin, Ort: 12.6.95, Universität Heidelberg

Zielgruppe: Medizinisch, pflegerisch und administrativ tätiger und/oder methodisch-wissenschaftlich orientierter Personenkreis

Referentinnen: Dr. med. H. Haeske-Seeberg, Dipl.-Inform.Med. B. Pietsch-Breitfeld, Dipl. rer. soz. B. Sens

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 270.--DM
(MEDWIS-Beteiligte 220.-DM, Sonstige 650.-DM,
Studierende 150.-DM)

Beschreibung:

Die Durchführung von Qualitätssicherungs-Maßnahmen in der ambulanten und stationären Versorgung ist seit 1989 durch die §§ 135-137 Sozialgesetzbuch V verbindlich vorgeschrieben. Daneben existiert eine Vielzahl von professionellen Vorgaben (z.B. Berufs- und Weiterbildungsordnungen, Richt- und Leitlinien der Bundesärztekammer sowie der Kassenärztlichen Bundesvereinigung), in denen Regelungen zur Qualitätssicherung getroffen sind. Wissenschaftlich-medizinische Fachgesellschaften, ärztliche Selbstverwaltungen, Krankenhaus- und Kostenträger bemühen sich um die Erarbeitung und Koordination von Verfahren zur Verbesserung der Qualität in der medizinischen Versorgung.

Die Planung, Entwicklung und Durchführung von Qualitätssicherungs-Maßnahmen in der stationären Versorgung setzen fundierte Kenntnisse u.a. in folgenden Bereichen voraus:

- Rahmenbedingungen
- Konzepte der traditionellen Qualitätssicherung und des modernen Qualitätsmanagements
- Methoden und Techniken zur Messung, Analyse und Verbesserung von Qualität
- Einsatz geeigneter Informationstechnologien

Neben einem allgemeinen methodischen Überblick werden einige der in Deutschland bekanntesten Qualitätssicherungs-Maßnahmen hinsichtlich ihrer Ziele, Methodik, Organisation und praktischer Durchführung vorgestellt.

2. Sommerschule "Medizinische Dokumentation und Qualitätssicherung"

Termin, Ort: 19. - 20.6.95, Universität Heidelberg

Zielgruppe: Medizinische Informatiker/innen, Dokumentare/innen sowie ärztliches, pflegerisches und Verwaltungspersonal mit Aufgaben im Bereich der Nutzung, des Entwurfs oder der Bewertung von Dokumentationssystemen.

Referenten: Prof. Dr. W. Gaus, Dr. F. Leiner

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 450.-DM
(MEDWIS)-Beteiligte 370.-DM, Sonstige 970.-DM,
Studierende 250.-DM)

Beschreibung:

Ein wichtiger Teilbereich der medizinischen Dokumentation ist die Dokumentation von Patientendaten, die im Rahmen der Krankenversorgung gewonnen werden, die aber auch für Betriebsstatistiken und wissenschaftliche Zwecke ausgewertet werden. Dieser Teilbereich ist Gegenstand der Veranstaltung. Ausgeklammert werden andere Bereiche wie die Dokumentation

im Rahmen klinischer Studien oder die Dokumentation medizinischer Literatur oder medizinischen Wissens.

Folgende Einzelthemen sind vorgesehen:

- Einführung in die medizinische Dokumentation
- Einführung in die Ordnungslehre
- Darstellung wichtiger Ordnungssysteme (ICD, ICPM, SNOMED, TNM)
- Diagnosen- und Therapiedokumentation
- Typologie und Beispiele medizinischer Dokumentationssysteme
- Einbindung in Krankenhausinformationssysteme

3. Seminar "Anforderungskriterien und Empfehlungen für kommerzielle ICD-sowie ICPM-Verschlüsselungssoftware"

Termin: Ort: 12.6.95, Universität Heidelberg

Zielgruppe: Ärzte, Medizininformatiker und Dokumentare, die im Bereich der klinischen und ambulanten Dokumentation tätig sind.

Referenten: Dr. J. Ingenerf, Hr. Sembritzki, Prof. Dr. R. Thurmayr, Dr. A. Zaiß

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 270.-DM
(MEDWIS)-Beteiligte 220.-DM, Sonstige 650.-DM,
Studierende 150.-DM)

Beschreibung:

Das Gesundheitsstrukturgesetz schreibt die ICD-10 und ICPM zur Verschlüsselung von Diagnosen und Prozeduren in Krankenhäusern sowie im niedergelassenen Bereich vor. Die Betroffenen im Gesundheitswesen stehen im Falle des Einsatzes von rechnergestützten Verschlüsselungssystemen vor einer schwierigen Entscheidung da die Leistungsfähigkeit der Produkte hinsichtlich der vorliegenden Anforderungen nur schwer zu durchschauen ist. Es gibt zahlreiche Lösungen, u.a. im universitären Bereich. Dieses Seminar konzentriert sich allerdings auf kommerziell angebotene Systeme.

Das Angebot ist in den letzten Monaten sprunghaft gewachsen. Im allgemeinen befinden sich die Systeme in laufender Entwicklung. Zur Produktauswahl benötigt man neben den methodischen und algorithmischen Aspekten auch weitergehende Informationen wie Kosten/Nutzen-Verhältnis, diverse Sekundärleistungen, Hard- und Software-Anforderungen, Firmenprofil und Anzahl von Referenz-Installationen, die nur schwer zu ermitteln sind.

In diesem Seminar sollen Anforderungskriterien und Empfehlungen für solche Systeme vorgestellt werden. Ein repräsentativer Datensatz wird zur empirischen Bewertung herangezogen. Schließlich haben alle Teilnehmer die Gelegenheit, sich die Systeme vor Ort am Rechner anzuschauen.

4. Seminar "Elektronische Dokumentenmanagement- und Archivierungssysteme für das Gesundheitswesen"

Termin, Ort: 29.6.95, Klinikum der Universität Heidelberg

Zielgruppe: Medizinische Informatiker in Industrie, Forschung und Krankenhaus, leitende Kräfte in Klinikrechenzentren, Ärzte mit Kenntnissen der Medizinischen Informatik, Medizinische Dokumentare.

Referenten: Dipl.-Inform. Med. C. Dujat, Dipl.-Inform. P. Schmücker

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 270.-DM
(MEDWIS-Beteiligte 220.-DM, Sonstige 650.-DM,
Studierende 150.-DM)

Beschreibung:

Die Informationslogistik im Krankenhaus, insbesondere auch das Archivwesen, ist durch eine Vielzahl von Unzulänglichkeiten gekennzeichnet. Eine Verbesserung dieser Situation ist durch die Einführung und Nutzung moderner Dokumentenmanagement- und Archivierungssysteme möglich. Diese beinhalten auch geeignete Verfahren zur Acquisition, zum Verteilen und zur Bereitstellung von Informationen.

Ziel des Seminars ist es,

- Möglichkeiten und Schwachstellen der konventionellen Archivierung aufzuzeigen, die Anforderungen der Archivierung im Krankenhaus und innovative Archivierungstechnologien vorzustellen,
- DV-Konzepte, Organisationsmodelle und Einführungsstrategien für innovative Archivierungstechnologien darzustellen,
- Archiv- und Krankengeschichtenmodelle, formale Beschreibungen der heterogenen Informationsstrukturen und klinischer Archivobjekte darzustellen,
- Möglichkeiten der Strukturierung und Spezifikation des klinischen Archivguts durch gängige Standards aufzuzeigen,
- über erste praktische Erfahrungen mit neuen Archivierungsverfahren zu berichten und
- die Nutzen und Kosten derartiger Archivierungsverfahren zu erörtern.

5. Seminar "Klinische Arbeitsplatzsysteme"

Termin, Ort: 3.7.95, Universität Heidelberg

Zielgruppe: Verantwortliche für Beschaffung oder Entwicklung klinischer Arbeitsplatzsysteme

Referenten: PD Dr. L. Gierl, Dr. K. Kuhn, Dr. E. Opitz, Dr. A. Zimmermann

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 270.-DM
(MEDWIS-Beteiligte 220.-DM, Sonstige 650.-DM
Studierende 150.-DM)

Beschreibung:

- Funktionalität klinischer Arbeitsplatzsysteme
- Überblick über existierende Systeme und Entwicklungen
die Funktionalitätsbeschreibung und der Überblick berücksichtigen sowohl die ärztliche als auch die pflegerische Sicht
- Bezug zu aktuellen Entwicklungen in der Informatik
Überblick über Software-Tools zur Gestaltung und Generierung von Benutzeroberflächen
- Probleme aus der Praxis, Lösungsvorschläge
GSG am klinischen Arbeitsplatz
Integration klinischer Arbeitsplätze in ein heterogenes Krankenhaus-Informationssystem

6. Seminar "Standards für die Datenkommunikation (EDIFACT, HL7, BDT)"

Termin, Ort: 10.7.95, Klinikum der Universität Heidelberg

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

Referenten: Prof. Dr. J. Dudeck, Ch. Seggewies, Dr. B. Wentz

Teilnahmegebühren: Mitglieder von BVMI, DVMD, GMDS 270.-DM
(MEDWIS-Beteiligte 220.-DM, Sonstige 650.-DM,
Studierende 150.-DM)

Beschreibung:

Krankenhausinformationssysteme werden heute als Verbund kooperativ arbeitender, funktionsorientierter Systeme konzipiert und zunehmend in Kliniken eingeführt. Auch bei niedergelassenen Ärzten werden Praxiscomputersysteme immer häufiger eingesetzt. Zwischen diesen Systemen wird ein elektronischer Datenaustausch angestrebt, der die Definition von Standard-Schnittstellen voraussetzt. Weltweit wird deshalb an der Entwicklung von Kommunikationsstandards gearbeitet. Standardisiert werden Nachrichteninhalte, Nachrichtenstrukturen, Austauschformate und Nachrichtenmanagement. Für krankenhausinterne Kommunikation ist derzeit am weitesten der in USA konzipierte HL7 Kommunikationsstandard entwickelt. Gefördert von der EG, wird intensiv an der Definition von Nachrichten gearbeitet, die als Austauschformat EDIFACT vorgesehen ist. Zum Austausch von Krankengeschichten zwischen Praxis-Computersystemen wurde vom Zentralinstitut der KV der Behandlungsdatenträger (BDT)-Standard definiert, der für diesen Kommunikationsbereich in der Bundesrepublik inzwischen als quasi-Standard anzusehen ist.

Im Seminar werden konzeptionelle Ansätze der verschiedenen Standards behandelt sowie Aufbau und Inhalte der Nachrichten, Austauschformate und Nachrichten-Management-Funktionen. Ergänzend werden auch die hiermit eng verbundene Normen für Basiskommunikationsprotokolle sowie Kommunikationsserver vorgestellt.

* * * * *

GMDS Jahrestagung in Bochum

11.09.1995 von 16.00 - 17.30

**AG "Statistische Methoden" der GMDS und
"Statistische Methoden in der Medizin" der IBG**

Thema: Statistische Methoden in klinischen Studien

Abstracts bis 30.06.1995 an:

Prof. Dr. K. Ulm, Institut für Medizinische Statistik und Epidemiologie, TU München
Ismaninger Straße 22, 81675 München

* * * * *

Workshop "Statistische Methoden der Klassifikation"

10./11. November 1995

Deutsches Krebsforschungszentrum, Heidelberg

In der Biometrie gibt es zahlreiche Bereiche, in denen Fragestellungen aus dem Bereich der Klassifikation bearbeitet werden müssen. Beispiele sind die Konstruktion prognostischer Indizes mit anschließender Einteilung in Gruppen mit guter, mittlerer und schlechter Heilungschance, die Entwicklung von Risikoscores für die Einschätzung unerwünschter Nebenwirkungen einer Therapie oder mehrstufige diagnostische Verfahren. Neben Regressionsmodellen haben in den letzten Jahren in diesen Bereichen neue Methoden wie z.B. CART und Neuronale Netze an Bedeutung gewonnen. Neue statistische Ansätze wie Bootstrap und Crossvalidation haben zur Weiterentwicklung bestehender Methoden beigetragen. In dieser Arbeitstagung soll versucht werden, einen Austausch zwischen WissenschaftlernInnen unterschiedlicher Fachgebiete, die sich theoretisch und praktisch mit Klassifikationsmethoden beschäftigen, zu initiieren. Daher sind Beiträge erwünscht, die die beispielhafte Anwendung von Methoden vorstellen, Eigenschaften von Methoden systematisch untersuchen, konkurrierende Verfahren vergleichen, oder offene Fragen und Probleme, die sich bei der Anwendung ergeben, diskutieren.

Die Tagung ist eine gemeinsame Veranstaltung der Arbeitsgruppe *Datenanalyse und numerische Klassifikation* der Gesellschaft für Klassifikation und der Arbeitsgruppen *Statistische Methoden in der Medizin* der Biometrischen Gesellschaft und *Statistische Methoden* der GMDS. Um Anmeldung von Beiträgen wird bis zum 30.8.95 gebeten.

Name: _____

Adresse: _____

Tel.: _____ Fax: _____ email: _____

Ich nehme an dem Workshop teil Ich möchte einen Vortrag halten

Vortragstitel: _____

Bitte schicken an:

Dr. W. Vach, Freiburger Zentrum für Datenanalyse und Modellbildung, Universität Freiburg, Albertstr. 26-28, 79104 Freiburg, Tel. 0761/202 1178, Fax: 0761/202 0503, email: wv@fdm.uni-freiburg.de oder Dr. W. Sauerbrei, Institut für Medizinische Biometrie und Medizinische Informatik, Universität Freiburg, Stefan-Meier-Str. 26, 79104 Freiburg, Tel. 0761/203 6669, Fax: 0761/203 6680, email: wfs@imbi.uni-freiburg.de

* * * * *

Im Dezember 1994 ist das Buch "**Biometrie für Mediziner**" von Prof. Dr. Bibeler, Universität Greifswald, GinkgoPress, Verlag für Wissenschaft & Kultur GmbH, Liebenthal 1, 17506 Gützkow/MVP Tel. 038353-50777, FAX: 50778 erschienen.

* * * * *

Der Hauptausschuß der DFG hat Prof. Dr. Haux, Heidelberg, für 3 Jahre zum Mitglied der Kommission für Rechenanlagen gewählt.

* * * * *

Ausschreibungen

GMDS-Förderpreis für Studierende 1995

Zur Förderung des wissenschaftlichen Nachwuchses schreibt die GMDS 1995 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 1993, 1994 oder 1995 abgegeben wurde.

Bitte senden Sie Ihre Bewerbung bis zum 30. Juni 1995 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 1995

Das Programmkomitee der GMDS Jahrestagung wird auch 1995 in Bochum 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.

Stellenausschreibung

Am Institut für Medizinische Statistik, Informatik und Epidemiologie der Universität zu Köln ist ab sofort eine (zunächst) befristete Stelle für eine/n **Statistiker/in / Biometriker/in** als Wissenschaftliche/r Angestellte/r zu besetzen. Abgeschlossenes Hochschulstudium in Statistik, Mathematik, Epidemiologie oder Medizin erforderlich, Promotion erwünscht. Bewerbungen bitte an Prof. Dr. W. Lehmann, Institut für Medizinische Statistik, Informatik und Epidemiologie der Universität zu Köln, 50924 Köln.

* * * * *

Veranstaltungskalender

Diese Termin- und Themenübersicht 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.

24. Jahrestagung der Dt. Ges. f. Psychosomatische Geburtshilfe und Gynäkologie

* 21. - 22.05.1995 * Basel

Information: PD Dr. H. Kentenisch, Univ.-Frauenklinik Rudolf Virchow - C -Pulsstraße 4, 14059 Berlin, Tel 030-3035-4433, FAX 030-3035-4006.

27. Annual Meeting der European Society of Human Genetics

* 24. - 26.05.1995 * Berlin, ICC

Information: DER-Congress, Bundesallee 56, 10715 Berlin, Tel 030-857 903-0, FAX 030-857 903-26

76. Dtsch. Röntgenkongreß der Dtsch. Röntgengesellschaft

* 24.05. - 27.05.1995 * Wiesbaden, Rhein-Main-Hallen

Information: Geschäftsstelle d. Dt. Röntgenges., Frankfurter Str. 231, 63263 Neu-Isenburg, Tel.: 06102-4032, Fax 06102-6668

12. ISQUA World Congress "Partnerships for Creating a Quality Health System, Users - Providers - Funders.

* 31.05. - 02.06.1995 * St. John's Newfoundland, Canada

Information: Organizing Secretariat Beclin Building, 1118 Topsail RD. P.O. Box 8234 St. John's NFLD, Canada, Tel. (709) 364-7701, Fax (709) 364-6460

18. Jahrestagung d. Vereinigung f. Operative u. Onkolog. Dermatologie (VOD)

* 16. - 18.06.1995 * Zürich

Information: Prof. Dr.G. Burg, Dermatolog, Klinik Univ.Spital, Gloriastr. 31, CH-8091 Zürich, Tel.: 00411/255-2550, Fax 0041/2554403

MedInfo 1995

* 23. - 28.07.1995 * Vancouver, British Columbia/Canada

Information: MedInfo '95, Suite 216, 10458-Mayfield Road, Edmonton, Alberta, Canada, T5P 4P4, Tel.: 403/489-8100, Fax 403/489-1122

40. Jahrestagung der Dt. Ges. f. Medizin, Informatik, Biometrie und Epidemiologie (GMDS)

* 10. - 13.09.1995 * Bochum

Information: Prof. Dr.H.J. Trampisch, Abt. Medizin, Informatik u. Biomathematik, Universitätsstr. 150, 44801 Bochum, Tel.: 0234-700-7790, Fax 0234-7094-325

Jahrestagung 1995 der Dt. Ges. f. Psychoanalyse, Psychotherapie, Psychosomatik + Tiefenpsychologie (DGPT)

* 15. - 17.09.1995 * Travemünde

Information: Geschäftsstelle der DGPT, Johannissbollwerk 20, 20459 Hamburg, Tel: 040-3192619, Fax 040-3194300

74. Jahrestagung der Dt. Ges. f. Rechtsmedizin

* 19. - 23.09.1995 * Aachen

Information: Prof. Dr. H. Althoff, Inst. f. Rechtsmedizin der RWTH Aachen, Tel.: 0241-808-0028, Fax 0241-808-9040

31. Jahrestagung der Dt. Ges. f. Sozialmedizin und Prävention

* 27. - 30.09.1995 * Magdeburg

Information: Prof. Dr. B.-P. Robra M.P.H., Inst. f. Sozialmedizin der Otto-von-Guericke- Univ., Leipziger Str. 44, 39120 Magdeburg, Tel: 0391-31075, FAX 0391-31076

Communicating Health Information

* 30.09. - 03.10.1995 * Helsinki

Information: Raja Tervo-Pelikka, national R&D Centre, Social and Medical Informatics, P.O.Box 220, 00530 Helsinki, Finland, Fax 00-358-039672424

Quantitative Methoden in Genetik und Molekularbiologie

* 02. - 05.10.1995 Stift Keppel, Hilchenbach

Information: Prof. Dr. H.H. Bock, Institut für Statistik, RWTH Aachen, Wüllnerstraße 3, 52056 Aachen, Tel.: 0241-804573, Fax 0241-8888130

Public Health Kongreß der Forschungsverbände

* 05. - 07.10.1995 *Dresden

Information: H. Krappweis, Fiedlerstr. 27, 01307 Dresden, Tel: 0351-4583254, Fax 0351-4593038

1 .nationaler Workshop Artificial Life

* 12 - 13.10.1995 * Bonn

Information: Dr. Kerstin Dautenhahn, GMD, FIT.KI, Schloß Birlinghoven, 53754 St. Augustin

HEALTH CARDS '95 der Dt. Ges. f. Medizin, Informatik, Biometrie und Epidemiologie (GMDS), Ärzteschaft u. Krankenkassenverbände

* 24. - 26.10.1995 * Frankfurt/Main

Information: Prof. Dr. O. Rienhoff, Inst. f. Medizin, Informatik, Robert-Koch-Str. 40, 37075 Göttingen, Tel.: 0551-393431, FAX 0551-398234

43. Arbeitstagung des Dt. Kollegium f. Psychosomatische Medizin (DKPM)

* 09. - 11.11.1995 * Mainz

Information: Prof. Dr. S.O. Hoffmann, Klinik f. Psychosomat. Med. u. Psychotherapie d. Univ., Langenbeckstraße 1, 55131 Mainz, Tel.: 06131-172841, FAX 06131-222332

Delegiertenkonferenz der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)

* 11.11.1995 * Frankfurt/Main

Information: Geschäftsstelle der AWMF, Moorenstr. 5, 40225 Düsseldorf, Tel. 0211-312828, Fax 0211-316819

14. Fortbildungstagung der Dt. Ges. f. Neurochirurgie

* 22. - 25.11.1995 * Bad Nauheim

Information: PD Dr. D. Rosenthal, Neurochir. Univ.klinik, Schleusenweg 2-16, 60528 Frankfurt/M.

3.Dt. Interdisz. Kongreß für Intensivmedizin der Dt. Interdisziplinären Vereinigung für Intensivmedizin (DIVI)

* 06. - 09.12.1995 * Hamburg, CCH

Information: Kongreß-Sekretariat DIVI '95 c/o CCH-Congress Organisation, Jungiusstr. 13, PF 302480, 20308 Hamburg, Tel 040-356923-41, FAX 040-356923-43

1996

20. Interdisziplinäres Forum "Fortschritt und Fortbildung in der Medizin" der Bundesärztekammer

* 10. - 13.01.1996 * Köln

Information: PD Dr. G. Ollenschläger, BÄK Dez. Fortbildung, Herbert-Lewin-Str. 1, 50931 Köln, Tel.: 0221-4004-240, Fax 0221-4004-388.

37. Kongreß der Deutschen Gesellschaft für Pneumologie

* 13. - 16.03.1996 * Essen

Information: Prof. Dr. N. Konietzko, Präsident DGP, Ruhrländklinik, Tüschener Weg 40, 45239 Essen

113. Kongreß (Chirurgenkongreß) der Dtsch. Ges. f. Chirurgie

* 09. - 13.04.1996 Berlin

Information: Prof. Dr. Wilhelm Hartel, Generalsekretär der DGCh, Elektrastr. 5, 81925 München, Tel. 089-915205, Fax 089-915071.

102. Tagung (Internistenkongreß) der Dtsch. Ges. f. Innere Medizin

* 13. - 17.04.1996 Wiesbaden

Information: Dt. Ges. f. Innere Medizin, Humboldtstr. 14, 65189 Wiesbaden, Tel: 0611-307946, Fax 0611-378260

12. International Health Records Congress

* 15. - 19.04.1996 * München

Information: Ms. Johanna Schessner, Gruber Straße 16 A, 85652 Pliening, Tel.: 089-9043404, Fax 089-9037316

67. Jahrestagung der Dt. Ges. f. Hals-Nasen-Ohrenheilkunde, Kopf- und Hals-Chirurgie

* 18. - 22.05.1996 Aachen

Information: Dr. H. Feldmann, Univ.-HNO-Klinik, Kardinal-v.-Galen-Ring 10, 48149 Münster, Tel.: 0251-83-6872, Fax 0251-83-6812.

Personalia

Als neue Mitglieder begrüßen wir:

Stefan Veltmaat Heideweg 1 37139 Adelebsen Tel. 05506-1544	Dipl.-Ing. Joachim Köhler Städtisches Klinikum "St. Georg" Leipzig Abt. f. Med. Informatik Delitzscher Straße 141 04129 Leipzig Tel. 0341-9093943
Dr. med. Katrin Thelen Pfarrergasse 11 93047 Regensburg Tel. 0941-51277	Thomas Drescher Software Union Gesellschaft für Unternehmensberatung mbH Haberstraße 2 53842 Troisdorf Tel. 02241-493112
Stefan Brings Software Union Gesellschaft für Unternehmensberatung mbH Haberstraße 2 53842 Troisdorf Tel. 02241-493112	Ulrich Engel Software Union Gesellschaft für Unternehmensberatung mbH Haberstraße 2 53842 Troisdorf Tel. 02241-493112
Dipl.-Informatiker Helmut Huppenbauer Schillerstraße 48 89077 Ulm Tel. 0731-6023401	Prof. Dr. Peter Jensch C.v.O. Universität Oldenburg FB Informatik Postfach 2503 26111 Oldenburg Tel. 0441-7982992
Dipl.-Informatiker Ralf D. Buchholz Hewlett Packard GmbH Schickardstraße 4 71034 Böblingen Tel. 07031-141342	Dipl.-Mathematiker Jörg Gehrke Auf dem Dwang 19 19053 Schwerin Tel. 0385-612567
Dipl.-Ing. Thomas Lehmann Institut für Med. Informatik und Biometrie der RWTH Aachen Pauwelsstraße 30 52057 Aachen Tel. 0241-8088786	Dipl.-Statistikerin Andrea Weber Zellerstraße 15 82067 Ebenhausen/Zell Tel. 08178-4956
Dr. rer. nat. Karl Fehnle Landwehrstraße 14 80336 München Tel. 089-558538	Dr. Ing. habil. Jörg-Rüdiger Blau Paracelsus-Kliniken Am Osterbergsee 4 37581 Bad Gandersheim Tel. 05382-70060

Dipl.-Statistiker Erwin Surmann Pillenreuther Straße 18 90459 Nürnberg Tel. 0911-445866	Oliver Boy Peenestraße 9 53127 Bonn Tel. 0228-299323
Dr. med., Dipl.-Inform. Klaus Peter Maag Josef-Fischhaber-Straße 18 82319 Starnberg Tel. 08151-21974	

Als studentische Mitglieder begrüßen wir:

Bianca Kochner Brandenburger Straße 2 67165 Waldsee Tel. 06236-51843	
--	--

MENSCH ERNÄHRUNG UMWELT



**Ernährungs-
epidemiologie.**
Mensch, Ernährung,
Umwelt.
Dr. Ulrich S. Oltersdorf.
351 Seiten,
59 Abbildungen.
Kt. DM 88,- /
öS 687,- / sFr 84,-.
ISBN 3-8001-2146-8.

Die Ernährungsepidemiologie ist ein neues Teilgebiet der Ernährungswissenschaften und umfasst das → **Sammeln, Ordnen und Bewerten von Informationen** über Handlungen (Ernährungsverhalten) und deren Beweggründe (Determinanten des Ernährungsverhaltens) sowie deren Auswirkungen (Ernährungs- und Gesundheitszustand) im Bereich der Ernährung des Menschen. Die Theorien und Methoden der Ernährungsepidemiologie werden vorgestellt. Es wird ein Konzept für die empirische Erfassung der Beziehungen zwischen Ernährung, Mensch und Umwelt entwickelt.

Aus dem Inhalt: Einleitung - Anmerkungen zur Entwicklung der Ernährungswissenschaft. Neue Anforderungen an die Ernährungswissenschaft. Grundlagen der Ernährungsepidemiologie: Definition und Aufgaben. Die Theoriebildung. Die eigentlichen Methoden-Elemente der Ernährungsepidemiologie: Methoden zur Erfassung der Ernährung des Menschen. Anthropometrische und biochemische Messungen zur Ermittlung des Ernährungszustandes. Praktische Hinweise zur Organisation und Durchführung von ernährungsepidemiologischen Studien: Die Verschlüsselung der erhobenen Informationen und die EDV-gerechte Datenerfassung. Die Verschlüsselung von Ernährungsdaten.

Der Autor: Dr. Ulrich S. Oltersdorf ist tätig an der Bundesanstalt für Ernährung, Institut für Ernährungsökonomie und -soziologie, Stuttgart.

Coupon an Ihre Buchhandlung oder senden an:
Verlag Eugen Ulmer, Postfach 70 05 61, 70574 Stuttgart.

BUCH-COUPON

Senden Sie mir das Buch „Ernährungsepidemiologie“ zum Preis von DM 88,- / öS 687,- / sFr 84,-. (Best.-Nr. 21468).

Name, Vorname

Straße/Nr.

PLZ, Ort

Datum/Unterschrift

Biom 2/95

4 schnelle Wege zum Buch:

1.: Kauf in Ihrer Buchhandlung. Lassen Sie sich das Buch zeigen. 2.: Mit diesem Coupon. 3.: Bestellen Sie per Telefon (07 11) 45 07-121. 4.: Bestellen Sie per Telefax (07 11) 45 07-120.



Statistische Graphik für die computergestützte Datenanalyse

Ein Handbuch zum Programmpaket SYSTAT (DOS- und WINDOWS-Version)

Von Prof. Dr. Dieter URBAN, Thomas BRUNS und Heinz-Werner NEUHAUS,
Institut für Sozialforschung, Universität Stuttgart

1994. VIII, 267 S., 116 Abb., kt. DM 68,- / ÖS 531,- / SFr 65,50

Inhalt: Einführung – Steuerung des Programmablaufs – Balken-, Kategorien-, Dichte- und Kreisdiagramme – Ikonendiagramme – Box-Plot und Stem-Leaf-Diagramme – Quantil- und Wahrscheinlichkeits-Diagramme – 2-D, 3-D-Diagramme und Diagramm-Matrizen (Vielfach-Diagramme) – Funktions-Diagramme – Map-Diagramme (Landkarten-Darstellungen) – Komplexe Graphiken mittels Steuerprogrammen – Anhang

In diesem Handbuch werden einfache und komplexe Verfahren zur graphischen Datenanalyse beschrieben, die in den DOS- und WINDOWS-Versionen des universellen Statistik-Programmsystems SYSTAT (System for Statistics) enthalten sind.

Erläutert wird der Einsatz von SYSTAT zur Erstellung von statistischen Graphiken wie z.B. Präsentationsgraphiken (u.a. Balken- und Kreisdiagramme), Illustrationsgraphiken (mit differenziert zu variierenden Bildern und Ikonen) und Kartographiegraphiken (u.a. Landkarten-Darstellungen). Darüber hinaus werden die vielfältigen graphischen Möglichkeiten zur Aufdeckung und Veranschaulichung von Daten-Strukturen behandelt. SYSTAT enthält sowohl eigenständige, visuell orientierte Statistik-Modelle als auch neuartige graphische Techniken zur Datenanalyse. Dazu gehören graphisch-basierte Statistik-Modelle (u.a. Box-Plot- und Stem-Leaf-Modelle), graphisch-ergänzte Statistik-Modelle (z.B. durch ergänzende Einfluß-Diagramme), graphisch-veranschaulichte Statistik-Modelle (z.B. über 3-D-Funktionsdiagramme), graphisch-dynamische Analyse-Techniken (mittels Graphik-Rotationen im 3-D-Raum) sowie interaktive Techniken zur Verknüpfung von numerischer und graphischer Analyse.

Eine Besonderheit des Handbuches ist es, daß die Beschreibung jeder Graphik-Prozedur anhand von vollständig abgedruckten Programm-Beispielen erfolgt. Dabei wird zu jeder Beispielzeile eine kommentierende Zeile aufgeführt. So kann jedes der abgedruckten Beispiele komplett in SYSTAT eingegeben und sofort gestartet werden.

Deskriptive Statistik in SAS

Einführung mit Beispielen aus der Medizinischen Biometrie und Epidemiologie

Von Dr. Rolf WEITKUNAT,

Institut für Med. Informationsverarbeitung, Biometrie und Epidemiologie, München
1994. XII, 250 S., kt. DM 78,- / ÖS 609,- / SFr 75,-

Das SAS-System ermöglicht die Bearbeitung komplexer Problemstellungen und bietet vielfältige Einsatzmöglichkeiten in Lehre und Forschung, Industrie, Handel und Verwaltung. Jedoch kann SAS aufgrund seines enormen Sprach- und Leistungsumfanges nicht, wie viele kleinere, befehlsorientierte Programmiersprachen, im Schnelldurchgang erlernt werden. Diese Einführung zu SAS behandelt schwerpunktmäßig den Bereich der Datenbeschreibung und wendet sich an Anwender, die bereits über statistische Vorkenntnisse verfügen. In knapper und übersichtlicher Darstellung werden anhand von Beispielen aus der Medizinischen Biometrie und Epidemiologie praxisbezogene SAS-Kenntnisse vermittelt, die den Anwender zur eigenständigen Bearbeitung von Fragestellungen im Bereich der deskriptiven Statistik befähigen.

Inhalt: Einführung – Eindimensionale Beschreibung kategorialer Variablen – Codierung und Klassifikation – Mehrdimensionale Beschreibung kategorialer Variablen – Vierfeldertafeln – Eindimensionale Beschreibung stetiger Variablen – Mehrdimensionale Beschreibung stetiger Variablen – Anhang



GUSTAV
FISCHER

Preisänderung vorbehalten