

Charité - Universitätsmedizin Berlin
Campus Benjamin Franklin
Aus dem Institut für Sozialmedizin, Epidemiologie und
Gesundheitsökonomie
Direktor: Prof. Dr. Stefan N. Willich
Abteilung: Internationale Gesundheitswissenschaften
Abteilungsleiter: Prof. Dr. Dankmar Böhning

Untersuchung von verschiedenen
Modellierungen der Heterogenität in
multizentrischen Studien

Inaugural – Dissertation
zur Erlangung des Grades
Doctor rerum medicarum
der Charité – Universitätsmedizin Berlin
Campus Benjamin Franklin

vorgelegt von Ronny Kuhnert
aus Löbau

Referent: Prof. Dr. Dankmar Böhning

Koreferent: Prof. Dr. Peter Martus

Gedruckt mit der Genehmigung der Charité – Universitätsmedizin Berlin
Campus Benjamin Franklin

Promoviert am: 19.07.2005

Inhaltsverzeichnis

Einführung	ix
1 Multizentrische Studien	1
2 Die drei Modellansätze	17
2.1 Allgemeine Modellierung der Heterogenität	17
2.2 Normalverteilungsansatz	19
2.3 Profil-Likelihoodansatz	22
2.4 Multi-Level-Modell	23
3 Schätzen der Parameter	27
3.1 Parameterschätzung im Normalverteilungs- Modell	28
3.2 Parameterschätzung im Profil-Likelihood- Ansatz	28
3.3 Parameterschätzung im Multi-Level-Modell	29
3.4 EM-Algorithmus bei einer Mischverteilung	31
3.5 Eigenschaften von Mischverteilungen	32
3.5.1 Gradientenfunktion	33
3.5.2 Der allgemeine Maximum-Likelihood Satz von Mischverteilungen	34
3.6 EM-Algorithmus mit Gradientenfunktion	36
3.6.1 EM-Algorithmus mit Gradientenfunktions-Update bei bekannter An- zahl der Komponenten	36
3.6.2 EM-Algorithmus mit Gradientenfunktions-Update bei unbekannter Anzahl der Komponenten (gEMGFU)	37
3.7 gEMGFU in FORTRAN	38
3.7.1 Numerische Stabilität	39
3.7.2 Die Maximierung der Gradientenfunktion	43

4	Auswertung der Studien	47
4.1	Ohne Heterogenität (Homogenitätsfall)	47
4.2	Mit unbeobachteter Heterogenität	47
4.2.1	Beta-Blocker Studie	48
4.2.2	Studie über Infektion der Atemwege	49
4.2.3	Studie über Präeklampsie	51
4.2.4	CALGB-Studie	52
4.2.5	Multi-Farm-Studie	52
4.2.6	Herzinfarkt-Studie	52
4.2.7	Studie über Frühgeburten	62
4.2.8	Studie über Bluthochdruck	62
4.2.9	Studie über koronare Herzkrankheit	63
4.2.10	Studie über Vitamin A Ergänzung	65
4.2.11	Brustkrebsvorsorgestudie	75
4.2.12	Nikotinersatzstudie	75
4.2.13	Kaiserschnittsstudie	76
4.2.14	Zusammenfassung	83
5	Simulationsstudie	85
5.1	Unter Effekthomogenität	85
5.1.1	Beta-Blocker Studie	86
5.1.2	CALGB-Studie	87
5.1.3	Multi-Farm-Studie	87
5.1.4	Studie über koronare Herzkrankheit	90
5.2	Unter Effektheterogenität	93
5.2.1	Beta-Blocker Studie	94
5.2.2	CALGB-Studie	97
5.2.3	Studie über koronare Herzkrankheit	98
5.2.4	Eigenständige Simulationsstrategien	104
5.3	Schätzung der Komponentenzahl	111
5.3.1	Beta-Blocker Studie	112
5.3.2	Eigenständige Simulationsstudie	113
5.4	Stabilität der Simulationsstudien	115
6	Diskussion	117
6.1	Zusammenfassung	120

A	123
A.1 Bootstrap-Stichproben	123
A.2 Ergebnisse einzelner Bootstrap-Stichproben	126
A.3 Ableitungen der Gradientenfunktion	133
Literaturverzeichnis	137
Danksagung	141

Abbildungsverzeichnis

2.1	Empirische Verteilung des log Relativen Risikos im Vergleich mit der Normalverteilung (Daten sind aus Tabelle 1.12)	21
3.1	CALGB-Studie (Bias vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall [0.01;0.99] berechnet, die generierten x^T , x^C -Werte sind binomialverteilt, additive Konstante ist 0.00000001)	42
3.2	CALGB-Studie (Bias vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall [0.01;0.99] berechnet, die generierten x^T , x^C -Werte sind binomialverteilt, additive Konstante ist 0.5)	43
4.1	Zentrenzugehörigkeit bei der Atemwegsstudie mit zwei Mischungskomponenten (Profil-Likelihood-Ansatz). Die Zahlen neben den Punkten geben das Zentrum an.	50
4.2	Zentrenzugehörigkeit bei der Atemwegsstudie mit zwei Mischungskomponenten (Normalverteilungsansatz). Die Zahlen neben den Punkten geben das Zentrum an.	51
4.3	Zentrenzugehörigkeit bei der Atemwegsstudie mit vier Mischungskomponenten (Multi-Level-Modell). Die Zahlen neben den Punkten geben das Zentrum an.	54
4.4	Zentrenzugehörigkeit bei der Studie über koronarer Herzkrankheit mit vier Mischungskomponenten (Profil-Likelihood-Ansatz). Die Zahlen neben den Punkten geben das Zentrum an.	64
4.5	Zentrenzugehörigkeit bei der Studie über koronarer Herzkrankheit mit vier Mischungskomponenten (Normalverteilungsansatz). Die Zahlen neben den Punkten geben das Zentrum an.	64

5.1	Beta-Blocker Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall [0.01;4.73] berechnet, die generierten x^T , x^C - Werte sind binomialverteilt)	88
5.2	CALGB-Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall [0.01;0.99] berechnet, die generierten x^T , x^C - Werte sind binomialverteilt)	89
5.3	Multi-Farm-Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall [0.01;5.99] berechnet, die generierten x^T , x^C - Werte sind binomialverteilt)	91
5.4	Studie über koronarer Herzkrankheit (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall [0.01;6.85] berechnet, die generierten x^T , x^C - Werte sind Poissonverteilt)	92
5.5	Histogramm des Relativen Risikos für jede Komponente (Beta-Blocker Studie).	95
5.6	Histogramm von den Gewichten für jede Komponente (Beta-Blocker Studie).	96
5.7	Histogramm des Relativen Risikos für drei Komponenten (Beta-Blocker Studie).	98
5.8	Histogramm mit den drei Komponenten zusammen, je nach Modell (Beta-Blocker Studie)	99
5.9	Stabilität der Simulation im Normalfall (Tabelle 5.14, $n \sim Poi(3000)$). Abgebildet ist der Erwartungswert der jeweiligen Komponente für das Relative Risiko in Abhängigkeit der Simulationswiederholungen.	108
5.10	Stabilität der Simulation im Normalfall (Tabelle 5.10). Abgebildet ist der Erwartungswert der jeweiligen Komponente für das Relative Risiko in Abhängigkeit der Simulationswiederholungen.	116

Einführung

Im Vordergrund dieser Arbeit steht die Modellierung der Heterogenität in multizentrischen Studien. Die Evaluierung multizentrischer Studien steht seit geraumer Zeit im Interesse von Wissenschaft und Praxis. In der Epidemiologie gibt es verschiedene Maßzahlen, die den Behandlungseffekt einer Studie schätzen. Dazu gehören das Relative Risiko (RR), die Risikodifferenz (RD), oder das Odds Ratio (OR). Die RD ist die Inverse der „Number Needed to Treat“ (NNT). In der Praxis wird oft die NNT benutzt, da dieses Effektmaß leicht zu interpretieren ist. Der Statistiker bevorzugt wegen der guten statistischen Eigenschaften lieber das OR und das RR. Diese Arbeit wird sich im weiteren Verlauf nur auf das Schätzen des Relativen Risikos beschränken. Unter der Homogenitätsannahme können existierende Verfahren heute sehr gut den Effekt in multizentrischen Studien schätzen. Häufig ist diese Annahme aber zu schwach, da sich Therapieerfolge in manchen Zentren als vorteilhaft, in anderen als ungünstig darstellen. Es könnte in einer Studie bei einigen Zentren ein Therapieerfolg und bei einigen Zentren eine negative Therapiewirkung vorliegen. Werden nun alle Zentren zusammengefasst, dann könnte diese Charakteristik im Behandlungseffekt nicht dargestellt werden. Dies würde auch bedeuten, dass wichtige Informationen über Gruppierungen in der Population verloren gingen. Das Auftreten von Heterogenität lässt keine einfache Interpretation des Effektes zu. Dabei ist interessant, dass das an sich recht elementare Problem keiner einfachen Lösung zugänglich ist. Das Verhältnis zweier Raten lässt sich nicht mehr durch eine einfach darzustellende Verteilung beschreiben. Damit ist ein Zugang zur Modellierung der Heterogenität verbaut.

Es gibt unterschiedliche Methoden, die sich dem Problem nähern. Ein Ansatz basiert auf dem Profil-Likelihood (weiterhin PL). Dazu wird die Poisson-Likelihood-Funktion gebildet. Diese Funktion hängt aber nicht nur von dem interessanten Parameter, dem Relativen Risiko, ab, sondern auch von dem Nuisance-Parameter oder Störparameter. Dieser Nuisance-Parameter ist in dem vorliegenden Fall der Baseline-Parameter. Mit dem Profil-Maximum-Likelihood-Schätzer kann der Nuisance-Parameter eliminiert werden. Dadurch entsteht der PL, der dann nur noch eine Funktion des Relativen Risikos ist. Der ent-

sprechende Ansatz ist auch für die Binomial-Likelihood-Funktion möglich. Der rechnerische Aufwand ist dann allerdings weitaus größer. Weiterhin kann die Modellierung des Behandlungseffekts anhand der Normalverteilung (weiterhin ND) auf der Basis des logarithmischen Relativen Risikos erfolgen. Dazu wird angenommen, dass das log-Relative Risiko normalverteilt ist. Ein nächster Ansatz ist das Multi-Level Modell (weiterhin ML-Modell), welches von Aitkin [2] vorgeschlagen wird. Alle drei Verfahren modellieren die Heterogenität des Effektes anhand einer diskreten Mischverteilung.

Diese Arbeit beschreibt und vergleicht diese verschiedenen Methoden. Darüber hinaus werden die Modelle an einer Vielzahl von Beispielfallstudien und Simulationsstudien getestet. So werden die Unterschiede und die Leistungsfähigkeit der verschiedenen Ansätze verdeutlicht. Die Beispieldaten stammen von publizierten Studienergebnissen.

Diese Dissertation ist nach folgendem Schema aufgebaut. Das erste Kapitel befasst sich mit der Bedeutung der multizentrischen Studien in der epidemiologischen Praxis und stellt alle in dieser Arbeit verwendeten Studien vor. Zu jedem Datensatz gibt es eine inhaltliche Erklärung mit Quellenangabe. Im zweiten Kapitel stehen die Modellierung der Effektheterogenität und die drei statistischen Modelle im Vordergrund. Dabei wird jedes Modell detailliert erklärt. Ein wichtiger Teil in dem Kapitel 3 ist die Schätzung der Parameter mit Hilfe des EM-Algorithmus. Es wird ein weiterentwickelter Algorithmus zur Schätzung der optimalen Lösung vorgestellt. Dabei wird die numerische Stabilität des Algorithmus mit berücksichtigt. Probleme die Auftreten bei der Maximierung der Likelihoodfunktion werden diskutiert. Im nachfolgenden Kapitel wird mit dem entwickelten Maximierungsalgorithmus der Behandlungseffekt von allen hier beschriebenen Studien je nach Modell geschätzt. Mit einer Simulationsstudie wird im 5. Kapitel die Güte der Schätzung je nach Modell untersucht. Im ersten Teilabschnitt wird der Homogenitätsfall betrachtet. Im zweiten wird die unbeobachtete Heterogenität im Effekt zugelassen. In der abschließenden Diskussion werden die in den vorhergehenden Kapiteln beschriebenen Ergebnisse noch einmal kritisch betrachtet und weitere Entwicklungsmöglichkeiten werden vorgeschlagen.

Kapitel 1

Multizentrische Studien

Multizentrische Studien haben sich in den letzten Jahren zu einer unverzichtbaren Methode entwickelt, zuverlässig Daten über einen Interventionseffekt bereitzustellen. Die klinischen, methodischen und ethischen Anforderungen einer multizentrischen Studie stellen eine Herausforderung für die beteiligten Wissenschaftler dar. Diese Arbeit behandelt vorzugsweise den methodischen Teil einer solchen Studie. Die zahlreichen internationalen Veröffentlichungen in renommierten Zeitschriften belegen die Qualität und die allgemeine Anwendbarkeit von klinischen Studien, die über mehrere Zentren verteilt sind. In Deutschland ist mit der Bekanntmachung des Bundesministeriums für Bildung, Wissenschaft und Technologie im Bundesanzeiger vom 31. Mai 1997 die besondere Förderung von multizentrischen Studien beschlossen worden. Aus diesem Bescheid geht hervor, dass zu dieser Zeit in Deutschland auf dem Gebiet ein deutlicher Nachholbedarf bestand. Die Aussagekraft multizentrischer Studien ist im Allgemeinen höher als bei unizentrischen Studien (siehe Metzler [29]). Das liegt nicht nur daran, dass dadurch höhere Gesamtstichprobenumfänge erreicht werden. Internationale Standards geben den Rahmen vor, wie multizentrische Studien geplant, durchgeführt und ausgewertet werden. Eine kurze Einführung über die Leitlinien (Guidelines) für die Durchführung einer multizentrischen Studie sind in der Encyclopedia of Biostatistical [4, Seite 2710] angegeben. Entscheidend für die Qualität einer multizentrischen Studie sind sowohl eine gründliche Therapeuteschulung und eine detaillierte Planung des Studienablaufes als auch ein Monitoring während der ganzen Durchführungsphase und eine zentrale und möglichst unabhängige Erfassung und Auswertung der gesammelten Daten. Der Studienablauf wird in einem allgemeinen Protokoll festgehalten und ist für jedes an der Studie teilnehmende Zentrum gültig. Typischerweise werden in multizentrischen Studien alternative Behandlungen miteinander verglichen, wobei die Anzahl der Zentren sehr stark von Studie zu Studie variiert. Welche enormen

Auswirkungen solche Studien haben, zeigt folgendes Beispiel: Nach der Veröffentlichung der Fünfjahresergebnisse einer Brustkrebsstudie (National Surgical Adjuvant Breast and Bowel Project [17]) sank in den Vereinigten Staaten der prozentuale Anteil der radikalen Mastektomie (op. Entfernung der weiblichen Brust) als chirurgische Maßnahme bei einer Brustkrebserkrankung von 75% auf 5%. Ein weiteres beliebtes Anwendungsgebiet für multizentrische Studien ist im Arzneimittelbereich zu finden. In Metzler [29] sind die 4 Entwicklungsphasen einer klinischen Arzneimitteltherapie angegeben. In jeder Phase sollen empirische Studien durchgeführt werden. Besonders geeignet sind multizentrische Studien in den Phasen 3 und 4, da es dort um die behördliche Zulassung und um die Nebenwirkungen geht. In diesen Phasen sollen die Stichprobenumfänge möglichst hoch sein. Ein Nachteil von unizentrischen Studien ist, dass Homogenisierungen der eingeschlossenen Patienten auftreten können (vgl. Rief [31] und Hargens [19]). Dies wird in multizentrischen Studien durch verschiedene Kliniken oder Zentren und der damit verbundenen hohen Streubreite an Patienten vermieden. Kommt eine Erkrankung sehr selten vor, ist es ebenfalls sinnvoll eine Studie auf mehrere Zentren auszudehnen, um einen ausreichenden Stichprobenumfang zu erhalten. Nachteilig wirkt sich der höhere organisatorische und finanzielle Aufwand auf die Durchführung von multizentrischen Studien aus. Schwieriger ist ebenfalls die statistische Auswertung mit der eingeschlossenen Heterogenität. Genau an diesem Punkt versucht diese Arbeit mit zwei herkömmlichen Verfahren und einem neuen Ansatz, dem PL-Ansatz die methodische Seite in multizentrischen Studien zu bereichern.

Diese Arbeit betrachtet 13 multizentrische Studien, die aus publizierten Beiträgen stammen. Jede Studie wird im folgenden inhaltlich kurz erläutert. Für mathematische und statistische Betrachtungen können die Beobachtungen als binomialverteilte Zufallsvariablen angenommen werden. In 4 von den 13 hier vorgestellten multizentrischen Studien werden Personenzeiten betrachtet. In diesen Studien wird das Relative Risiko auf der Grundlage der Inzidenzdichte, statt der kumulativen Inzidenz geschätzt. Die Inzidenzdichte ist definiert als

$$ID = \frac{\text{Anzahl der Krankheitsfälle im Studienzeitraum}}{\text{Gesamtpersonenzeit unter Risiko}}.$$

In diesem Fall wird den Beobachtungen eine Poissonverteilung unterstellt.

Es gelten im Weiteren folgende Notationen. Mit n_i^T werden die unter Risiko stehenden Personen in der Behandlungsgruppe 1 und mit n_i^C werden die unter Risiko in der Behandlungsgruppe 2 oder in der Kontrollgruppe bezeichnet. Mit x_i^T (x_i^C) wird angegeben, wie viele Fälle¹ in der Behandlungsgruppe 1 (Kontrollgruppe, Behandlungsgruppe

¹Krankheitsfälle, Todesfälle oder Heilerfolge

2) aufgetreten sind. Die Baseline-Parameter werden mit $p_i^T(p_i^C)$ für den Behandlungsarm (Kontrollarm) bezeichnet.

Erklärung für Tabelle 1.1 Die Daten, beschrieben bei Yusuf [34], stammen aus einer Studie mit 22 Zentren, bei dem der Patient nach einem Herzinfarkt mit Beta-Blocker behandelt wurde, um die Sterblichkeit zu senken. Bei Gelman [18] wird diese Studie ausführlich analysiert und Aitkin [2] bezieht sich in seiner Publikation auf diese Daten.

Erklärung für Tabelle 1.2 Die Daten setzen sich aus einer Studie mit 22 Zentren zusammen, in der untersucht wird, wie groß das Risiko einer Infektion der Atemwege bei einer selektiven Entgiftung des Verdauungstraktes ist. Die Einteilung der Patienten in Behandlungs- oder Kontrollgruppe erfolgt zufällig. Der Vergleich bestand in der Wirkung eines nicht resorbierbaren Antibiotikum (Behandlungsgruppe) versus eines Placebopräparat (Kontrollgruppe). Die Daten werden bei Turner [33] angewendet.

Erklärung für Tabelle 1.3 In der Tabelle sind 9 klinische Versuche enthalten, bei denen der Effekt von harntreibenden Mitteln während der Schwangerschaft, bei einem Risiko von Präeklampsie, von Interesse ist. $x^T(x^C)$ gibt die Anzahl der Präeklampsiefälle an. Diese Untersuchung wird auch bei Turner [33] als ein Beispiel angewendet.

Erklärung für Tabelle 1.4 Die Daten kommen von der Krebs- und Leukämie-Gruppe B (CALGB), welche zwei Chemotherapien gegenübergestellt. Dabei wird unterschieden ob der Patient, der an multiple Myeloma erkrankt ist, am Ende der Studie noch lebt oder verstorben ist. Diese Studie wird in Böhning [6] und in Lipsitz et al. [25] verwendet.

Erklärung für Tabelle 1.5 Die Daten stammen aus einer nicht veröffentlichten Querschnittsstudie über den Zusammenhang zwischen Neospora-Infektionen und dem Auftreten von Aborten bei Rindern. Die zurückliegende Exposition mit dem Erreger ist aufgrund eines serologischen Tests klassifiziert worden. Diese Studie ist eine retrospektive Erhebung. Die Bedeutung liegt darin, dass bei einem Zusammenhang mit Neospora-Infektionen, welche recht weit verbreitet ist, die Ursache des Abortgeschehens bei Rindern neu überdacht werden muss.

Erklärung für Tabelle 1.6 Diese Studie umfasst 31 Zentren aus den USA und Kanada. Die Studie beinhaltet Personen im Alter von 30 bis 69 Jahren, welche einen aku-

Tabelle 1.1: Beta-Blocker-Studie

Zentrum	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T	x_i^C	n_i^C
1	3	38	3	39
2	7	114	14	116
3	5	69	11	93
4	102	1533	127	1520
5	28	355	27	365
6	4	59	6	52
7	98	945	152	939
8	60	632	48	471
9	25	278	37	282
10	138	1916	188	1921
11	64	873	52	583
12	45	263	47	266
13	9	291	16	293
14	57	858	45	883
15	25	154	31	147
16	33	207	38	213
17	28	251	12	122
18	8	151	6	154
19	6	174	3	134
20	32	209	40	218
21	27	391	43	364
22	22	680	39	674

Tabelle 1.2: Studie über Infektion der Atemwege

Zentrum	Behandlungsarm		Kontrollarm	
i	x_i^T	n_i^T	x_i^C	n_i^C
1	7	47	25	54
2	4	38	24	41
3	20	96	37	95
4	1	14	11	17
5	10	48	26	49
6	2	101	13	84
7	12	161	38	170
8	1	28	29	60
9	1	19	9	20
10	22	49	44	47
11	25	162	30	160
12	31	200	40	185
13	9	39	10	41
14	22	193	40	185
15	0	45	4	46
16	31	131	60	140
17	4	75	12	75
18	31	220	42	225
19	7	55	26	57
20	3	91	17	92
21	14	25	23	23
22	3	65	6	68

Tabelle 1.3: Studie über Präeklampsie

Zentrum	Behandlungsarm		Kontrollarm	
i	x_i^T	n_i^T	x_i^C	n_i^C
1	14	131	14	136
2	21	385	17	134
3	14	57	24	48
4	6	38	18	40
5	12	1011	35	760
6	138	1370	175	1336
7	15	506	20	524
8	6	108	2	103
9	65	153	40	102

Tabelle 1.4: CALGB-Studie

Zentrum	Behandlungsarm		Kontrollarm	
i	x_i^T	n_i^T	x_i^C	n_i^C
1	1	3	3	4
2	8	11	3	4
3	2	3	2	2
4	2	2	2	2
5	0	3	2	2
6	2	3	1	3
7	2	3	2	2
8	4	4	1	5
9	2	3	2	2
10	2	3	0	2
11	3	3	3	3
12	0	2	2	2
13	1	5	1	4
14	2	4	2	3
15	4	6	2	4
16	3	9	4	12
17	2	3	1	2
18	1	4	3	3
19	2	3	1	4
20	0	2	0	3
21	1	5	2	4

Tabelle 1.5: Multi-Farm-Studie

Zentrum	serologisch positiv		serologisch negativ	
i	x_i^T	n_i^T	x_i^C	n_i^C
1	6	19	0	14
2	4	9	0	6
3	2	2	1	6
4	6	15	3	34
5	4	8	1	33
6	3	7	1	18
7	3	4	1	27

Tabelle 1.6: Herzinfarkt-Studie

Zentrum	Behandlungsarm		Kontrollarm	
i	x_i^T	n_i^T	x_i^C	n_i^C
1	0	49	3	48
2	1	57	7	58
3	1	56	5	57
4	1	42	4	42
5	3	65	10	65
6	3	70	8	71
7	2	65	5	66
8	3	55	7	55
9	3	55	7	56
10	2	44	4	44
11	4	77	8	78
12	2	58	4	59
13	4	70	7	70
14	7	98	12	98
15	3	48	5	48
16	7	52	11	53
17	5	64	7	65
18	3	42	4	42
19	5	47	6	47
20	4	62	5	63
21	5	59	6	60
22	5	59	1	59
23	8	63	3	64
24	7	55	4	55
25	8	75	5	75
26	7	91	6	92
27	8	70	7	70
28	3	44	3	44
29	4	32	4	33
30	10	57	10	58
31	10	125	10	126

ten Herzinfarkt erlitten hatten. Nachdem sich der Gesundheitszustand im Krankenhaus stabilisiert hatte, erhielten die Patienten entweder eine Behandlung mit Propranolol (Betarezeptorenblocker) oder mit einem Placebopräparat. Der Autor war daran interessiert, die Effektheterogenität in den 31 Zentren zu veranschaulichen. Diese Studie wurde in der Veröffentlichung von Horwitz [20] verwendet.

Erklärung für Tabelle 1.7 In dieser Studie wird eine Behandlung mit Tokolyse an Frauen untersucht, die unter hohem Risiko stehen eine Frühgeburt zu erleiden. x gibt dabei die Anzahl der Frühgeburten an. In der Publikation von Arends [3] wird diese Studie diskutiert.

Erklärung für Tabelle 1.8 Diese Studie widmet sich der Untersuchung der Sterblichkeit bei Patienten im mittleren Lebensalter, die unter leichten bis mittleren Bluthochdruck leiden. Dabei wird ein blutdrucksenkendes Medikament mit einem Placebopräparat verglichen. Arends [3] wendet diese Studie als zweites Beispiel an.

Erklärung für Tabelle 1.9 In dieser 33 Zentren umfassenden multizentrischen Studie wird das Sterblichkeitsrisiko bei Patienten mit koronarer Herzkrankheit untersucht. Dabei wird eine Cholesterolsenkende Behandlung mit einer Placebobehandlung verglichen. Als drittes Beispiel werden diese Daten bei Arends [3] benutzt.

Erklärung für Tabelle 1.10 Die Daten stammen aus einer multizentrischen Studie, die sich aus acht Gebieten in Entwicklungsländern zusammensetzt. Untersucht wird der Zusammenhang von Vitaminen A Ergänzung mit der Sterblichkeit von Kindern im Alter von 6 bis 72 Monaten. Diese Daten wurden in der Publikation von Fawzi [16] verwendet.

Erklärung für Tabelle 1.11 Berry [5] versucht mit dieser Studie in seiner Publikation der Frage nachzugehen, ob die Sterblichkeit an einer Brustkrebserkrankung durch Mammographie bei Frauen (ab 40 Jahre) reduziert werden kann.

Erklärung für Tabelle 1.12 Die Daten stammen aus einer Studie, bei der die Wirksamkeit einer Nikotinersatztherapie untersucht wird. Mit x^T (x^C) werden die Probanden bezeichnet, die erfolgreich mit dem Rauchen aufgehört haben. Diese Studie wurde in dem Artikel von DuMouchel [13] beschrieben. DuMouchel gibt für diese Studie weitere Kovariablen an. Eine Kovariable ist die Art der Nikotinersatztherapie (Kaugummi oder

Tabelle 1.7: Studie über Frühgeburten

Zentrum	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T	x_i^C	n_i^C
Mariona	0	4	0	5
Howard et al.	2	15	2	18
Larsen et al.	11	131	6	45
Hobel	2	16	3	15
Calder and Patel	4	37	9	39
Scommegna	1	15	5	17
Larsen et al.	5	49	16	50
Christensen et al.	0	14	6	16
Leveno et al.	15	54	25	52
Wesselius et al.	6	33	15	30
Cotton et al.	6	19	11	19
Barden	0	12	8	13
Ingemarsson	0	15	10	15
Spellacy et al.	6	14	11	15

Tabelle 1.8: Studie über Patienten im mittlern Lebensalter, die unter Bluthochdruck leiden

Zentrum	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T (P.-Jahre)	x_i^C	n_i^C (P.-Jahre)
VA	10	595,2	21	640,2
VA-NHBLI	2	762	0	756
HDFP	54	5635	70	5600
HDFP	47	5135	63	4960
HDFP	53	3760	62	4210
Oslo	10	2233	9	2084,5
ANBPS	25	7056,1	35	6824
MRFIT	47	8099	31	8267
MRFIT	43	5810	39	5922
MRFIT	25	5397	45	5173
MRC men	157	22162,7	182	22172,5
MRC women	92	20885	72	20645

Tabelle 1.9: Zahl der Todesfälle und gesamte Anzahl der Personenjahre von Behandlungs- und Kontrollgruppe von Patienten mit koronarer Herzkrankheit

Zentrum	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T (P. -jahre)	x_i^C	n_i^C (P. -jahre)
Singh	28	380	51	350
Marmorston	70	1250	38	640
Stamler	37	690	40	500
McCaughan	2	90	3	30
Harrold	0	30	3	30
Stockholm	61	1240	82	1180
Oslo Diet	41	1930	55	890
Low Fat	20	340	24	350
DART	111	1930	113	1920
VA drug	81	1240	27	410
Newcastle	31	1140	51	1140
Oliver	17	210	12	220
Acheson	23	210	20	230
STARS	0	90	4	170
CDP	1450	38620	723	19420
Dayton	174	1350	178	1330
Soya Bean	28	890	31	860
Scottish	42	1970	48	2060
Sahni	4	150	5	150
Upjohn	37	2150	48	2100
Sydney	39	1010	28	1120
Rose2	8	100	1	50
NHLIB	5	340	7	340
Minnesota	269	4410	248	4390
POSCH	49	3850	62	3740
CLAS	0	190	1	190
Frick '93	19	1510	12	1560
LCCPPT	68	13850	71	13800
Frick '87	46	10140	43	10040
EXCEL	33	5910	3	1500
WHO	236	27630	181	27590
SCOR	0	100	1	100
Gross	1	20	2	30

Tabelle 1.10: Zahl der Todesfälle und die gesamte Anzahl von Kinderjahren in Behandlungs- und Kontrollgruppe einer Studie, bei der die Wirkung von Vitamin A Ergänzung bei Kindern auf die Kindersterblichkeit untersucht wird

Zentrum	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T (P.-Jahre)	x_i^C	n_i^C (P.-Jahre)
Aceh, Indonesia, 86	101	12991	130	12209
Java, Indonesia, 88	186	5775	250	5445
Hyderabad, India, 90	39	7691	41	8084
Tamil Nadu, India, 90	37	7764	80	7655
Bombay, India, 91	7	6244	32	5754
Jumla, Nepal, 91	138	1261,9	167	1136,9
Sarlahi, Nepal, 91	152	14487	210	14143
Northern Sudan, 92	123	21669	117	21441

Tabelle 1.11: Zahl der Todesfälle und die gesamte Anzahl der Personenjahre in Behandlungs- und Kontrollgruppe einer Studie, bei der der Effekt der Brustkrebsvorsorge untersucht wird

Zentrum	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T (P.-Jahre)	x_i^C	n_i^C (P.-Jahre)
HIP (NY)	49	248	65	253
Edinburgh	46	170	52	155
Malmö	49	165	65	144
Kopparberg	23	144	18	75
Ostergotland	27	143	27	147
Canada	82	322	72	322
Stockholm	24	162	12	94
Gothenburg	18	138	39	168

Tabelle 1.12: Ergebnisse einer 59 Zentren Studie einer Nikotinersatztherapie

Zentrum i	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T	x_i^C	n_i^C
1	29	116	21	113
2	6	73	3	121
3	30	50	23	50
4	23	180	15	172
5	22	58	9	58
6	31	106	16	100
7	16	44	6	20
8	9	30	6	30
9	30	71	14	68
10	23	60	12	53
11	37	92	24	90
12	21	68	5	38
13	129	600	112	617
14	13	30	5	30
15	21	107	21	105
16	90	211	28	82
17	51	146	40	127
18	75	206	50	211
19	23	58	18	58
20	37	76	17	78
21	11	42	16	89
22	8	31	6	31
23	7	56	2	56
24	111	537	31	271
25	28	113	10	107
26	16	78	2	80
27	15	44	9	43
28	10	57	4	55
29	33	120	17	120
30	76	842	53	844
31	40	160	19	157
32	19	150	10	75
33	77	800	19	400
34	24	115	17	119
35	39	410	111	1208
36	81	729	78	1377
37	28	106	5	49
38	10	101	8	99
39	7	25	4	23

Fortsetzung von Tabelle 1.12

Zentrum i	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T	x_i^C	n_i^C
40	24	205	6	222
41	2	13	2	23
42	9	93	13	182
43	13	424	9	412
44	21	270	1	64
45	56	99	37	101
46	12	99	7	52
47	61	113	28	103
48	5	79	2	82
49	11	112	9	111
50	23	210	6	105
51	15	59	5	19
52	30	178	22	186
53	5	106	13	319
54	44	262	42	261
55	17	100	11	99
56	28	106	4	52
57	17	145	2	144
58	29	115	22	119
59	15	60	4	60

Pflaster) und eine andere ist der Grad der Unterstützung der Therapie vom Rauchen aufzuhören (wenig und hohe Unterstützung). Zusätzlich wird bei DuMouchel das Jahr der Untersuchung in den einzelnen Zentren angegeben. Diese Kovariablen werden hier aber zunächst nicht berücksichtigt.

Erklärung für Tabelle 1.13 In den 61 Zentren dieser multizentrischen Studie wird untersucht, ob die Häufigkeit einer Wundinfektion bei Frauen nach einem Kaiserschnitt durch eine prophylaktische Behandlung mit einem Antibiotikum verringert werden kann. Die Kontrollgruppe wurde mit einem Placebopräparat behandelt. Dieser Datensatz wird in der Publikation von Cooper[11] verwendet. Cooper zeigt, dass durch die prophylaktische Behandlung mit einem Antibiotikum die Behandlungskosten gesenkt werden.

Tabelle 1.13: Studie über die prophylaktische Behandlung mit einem Antibiotikum beim Kaiserschnitt

Zentrum	Behandlungsarm		Kontrollarm	
i	x_i^T	n_i^T	x_i^C	n_i^C
1	11	58	14	48
2	4	133	28	136
3	27	299	12	101
4	2	68	1	56
5	5	55	8	55
6	3	100	0	33
7	4	183	3	44
8	1	11	5	17
9	10	80	21	74
10	0	46	4	55
11	0	26	1	31
12	0	42	0	40
13	0	119	1	39
14	1	50	9	50
15	2	60	6	30
16	1	46	1	49
17	3	53	9	50
18	0	33	4	28
19	0	34	6	34
20	0	50	2	50
21	0	78	1	36
22	1	43	1	47
23	0	80	3	80
24	2	196	14	190
25	2	124	2	58
26	2	74	8	78
27	4	167	5	140
28	2	75	9	77
29	0	48	2	49
30	0	102	1	99
31	1	50	10	50
32	0	100	1	50
33	0	85	3	43
34	1	36	1	25
35	2	76	4	75
36	12	115	15	117
37	6	136	16	130
38	9	35	7	38
39	13	150	23	150

Fortsetzung von Tabelle 1.13

Zentrum i	Behandlungsarm		Kontrollarm	
	x_i^T	n_i^T	x_i^C	n_i^C
40	2	40	4	20
41	0	74	2	74
42	0	34	5	37
43	2	61	2	61
44	3	146	9	132
45	4	88	4	40
46	1	64	7	65
47	7	57	7	58
48	0	16	1	16
49	2	31	6	37
50	1	50	4	50
51	1	50	2	49
52	2	26	0	27
53	3	133	12	66
54	6	244	17	117
55	1	52	2	61
56	2	46	4	50
57	0	40	3	40
58	2	48	3	45
59	3	40	1	40
60	1	160	1	160
61	1	50	4	50

Kapitel 2

Die drei Modellansätze zur Modellierung der Heterogenität

2.1 Allgemeine Modellierung der Heterogenität

Selten tritt in der epidemiologischen Praxis Homogenität auf. In den meisten Studien gibt es einige unberücksichtigte Kovariablen, oder Variablen die nicht erhoben werden können. Des Weiteren können auch bestimmte Unterpopulationen auftreten, die nicht durch Kovariablen erklärbar sind. Mit der Modellierung der unbeobachteten Heterogenität können Subpopulationen oder auch latente Kovariablen gefunden werden. Das Ziel jeder Studie besteht auch darin, die Ergebnisse möglichst auf die Gesamtpopulation auszudehnen. Dabei muss aber das keine-einige-alle Problem (vgl. Hutchinson [21]) beachtet werden. Dieses besagt, dass in einer klinischen Studie nicht alle Teilpopulationen enthalten sein müssen. Es könnte sogar vorkommen, dass nur eine Teilpopulation erfasst ist. Bestimmte Testverfahren können testen, ob in den Daten Heterogenität oder Homogenität vorliegt. Das Verfahren wird als Test der Hypothese $H_0 : m = 1$ (Homogenität) gegen $H_1 : m > 1$ formuliert. Dazu wird in Böhning [8, Seite 117] der χ^2 -Test vorgeschlagen, wobei

$$\chi_{(k-1)}^2 = \sum_{i=1}^k \frac{(y_i - \bar{y})^2}{\sigma_i^2}, \quad (2.1)$$

mit $Var(y_i) = \sigma_i^2$, k steht für die Anzahl der Zentren
und $\bar{y} = \frac{\sum w_i y_i}{\sum w_i}$, $w_i = \frac{1}{\sigma_i^2}$

Hierbei ist y_i das logarithmische Relative Risiko im i -ten Zentrum. Bei diesem Test wird die Varianz σ_i^2 in jedem Zentrum als bekannt vorausgesetzt. Nur im ND-Ansatz, der im

Tabelle 2.1: Chi-Quadrat-Test: H_0 : Homogenität gegen H_1 : keine Homogenität, Konfidenzniveau $\alpha = 0.05$

Daten	χ_{ND}^2	$\chi_{(fg,1-0.05)}^2$	Testergebnis
Tabelle 1.1	23.0396	32.67	nicht ablehnen
Tabelle 1.2	41.7122	32.67	ablehnen
Tabelle 1.3	28.6221	15.51	ablehnen
Tabelle 1.4	29.4897	31.41	nicht ablehnen
Tabelle 1.5	2.33632	12.59	nicht ablehnen
Tabelle 1.6	28.3952	43.77	nicht ablehnen
Tabelle 1.7	27.3812	22.36	ablehnen
Tabelle 1.8	25.5725	19.67	ablehnen
Tabelle 1.9	91.1916	46.19	ablehnen
Tabelle 1.10	24.8867	14.07	ablehnen
Tabelle 1.11	9.40139	14.07	nicht ablehnen
Tabelle 1.12	75.934	76.78	nicht ablehnen
Tabelle 1.13	78.2714	79.08	nicht ablehnen

nächsten Abschnitt beschrieben wird, ist die Varianz im Modell enthalten (2.4). Die Tabelle 2.1 beinhaltet die Ergebnisse des χ^2 -Tests, der an jeder hier vorgestellten multizentrischen Studie durchgeführt worden ist. Der Test zeigt, dass in 6 von den 13 multizentrischen Studien keine Homogenität in den Daten vorliegt. Die unbeobachtete Heterogenität spielt somit keine untergeordnete Rolle und muss adäquat modelliert werden. Dazu wird angenommen, dass der Parameter des Behandlungseffektes eine unbekannte, nicht weiter spezifizierte Verteilung P aufweist. Dies impliziert, dass der Effektparameter eine Rand- (marginal) Dichte besitzt, die anhand einer endlichen Mischverteilung modelliert wird. Diese hat folgende Form:

$$f(y, P) = \sum_{j=1}^m f(y, \theta_j) q_j \quad (2.2)$$

Dabei gibt m die Anzahl der Komponenten an, $P = \begin{pmatrix} \theta_1 & \dots & \theta_m \\ q_1 & \dots & q_m \end{pmatrix}$ ist die mischende Verteilung, θ_j steht für das Relative Risiko der j -ten Komponente und q_j ist das Gewicht der j -ten Komponente. $f(y, \theta)$ wird als *Kern* bezeichnet. Für die Gewichte q_j gilt:

$$q_j \geq 0, \quad j = 1, \dots, m \quad \text{und} \quad \sum_{j=1}^m q_j = 1.$$

Eine Einführung in den Themenkreis diskrete Mischmodelle ist bei Land [24] angegeben. Weitergehende Informationen sind in dem Buch von McLachlan [28] enthalten. Im Gegen-

satz zur endlichen Mischverteilung gibt es noch die stetige Mischverteilung, die folgende Form besitzt:

$$f(y, P) = \int_{-\infty}^{\infty} f(y, \theta) p(\theta) d\theta.$$

Die Dichte $p(\theta)$ ist in der Regel unbekannt. Sie kann aber in bestimmten Situationen durch die Gammafunktion (siehe Böhning [8], Seite 12) ersetzt werden. Da die Gammaverteilung schwer zu rechtfertigen ist, beschränkt sich diese Arbeit auf die diskrete Mischverteilung. In einer multizentrischen Studie ist im Allgemeinen nicht bekannt, welches Zentrum zu welcher Teilpopulation gehört. Dieser Fall wird als unbeobachtete Heterogenität bezeichnet.

Im folgenden werden die drei Modelle vorgestellt. Angefangen wird mit dem herkömmlichen Normalverteilungsansatz. Danach folgt das Kernstück dieser Arbeit, das Profil-Likelihood-Modell. Als letztes wird das häufig angewendete Multi-Level-Modell dargestellt.

2.2 Normalverteilungsansatz

Das ND-Modell geht zurück auf Laird [23] und besitzt als erste Modellannahme, dass das logarithmische Relative Risiko, definiert als

$$y_i = \log\left(\frac{x_i^T}{n_i^T}\right) - \log\left(\frac{x_i^C}{n_i^C}\right), \quad (2.3)$$

normalverteilt ist. Inwieweit diese Annahme bei den hier vorgestellten multizentrischen Studie zutrifft, wird später diskutiert. Der ND-Ansatz besitzt folgende Form:

$$f(y|P, \sigma^2) = \sum_{j=1}^m q_j \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y - \log(\theta_j))^2}{2\sigma^2}}. \quad (2.4)$$

(2.4) weist auf die zweite Modellannahme. In der Dichte ist die unbekannt Varianz des log. RR enthalten, die entsprechend geschätzt werden muss. Unter der Annahme, dass die Beobachtungen x_i^T und x_i^C binomialverteilt mit $x_i^T \sim \text{Bin}(p_i^T, n_i^T)$ und $x_i^C \sim \text{Bin}(p_i^C, n_i^C)$, ist die Varianz gegeben als

$$\begin{aligned} \text{Var}(y_i) &= \text{Var}\left[\log\left(\frac{x_i^T}{n_i^T}\right) - \log\left(\frac{x_i^C}{n_i^C}\right)\right] \\ &= \text{Var}\left[\log\left(\frac{x_i^T}{n_i^T}\right)\right] + \text{Var}\left[\log\left(\frac{x_i^C}{n_i^C}\right)\right]. \end{aligned} \quad (2.5)$$

In dieser Form kann die Formel (2.5) nicht angewendet werden, weil die Varianz von $\log\left(\frac{x_i^T}{n_i^T}\right)$ nicht bekannt ist. Mit Hilfe der δ -Methode ($\text{Var}(g(X)) = g'(E(X))^2 \text{Var}(X)$) kann die Varianz geschätzt werden. Dies führt zu folgender Varianzformel (Böhning [7, Seite 51]):

$$\hat{\sigma}_i^2 = \widehat{\text{Var}}(y_i) = \frac{1}{x_i^T} - \frac{1}{n_i^T} + \frac{1}{x_i^C} - \frac{1}{n_i^C}. \quad (2.6)$$

In den Studien 1.8, 1.9, 1.10 und 1.11, in denen Personenzeiten betrachtet werden, sind die Beobachtungen nicht mehr binomialverteilt. Vielmehr wird den Beobachtungen eine Poissonverteilung unterstellt. Für die Varianzschätzung gilt nun die Formel

$$\widehat{\text{Var}}(y_i) = \frac{1}{x_i^T} + \frac{1}{x_i^C}. \quad (2.7)$$

Diskussion

Zur Verdeutlichung der ersten Modellannahme dient die Abbildung 2.1. In der Grafik ist die empirische Verteilungsfunktion und die theoretische Normalverteilung mit den Parametern μ und σ abgebildet. μ und σ wurden aus den Daten geschätzt. Die Daten stammen aus der Nikotinstudie (Tabelle 1.12) mit 59 Zentren. Aus der Abbildung 2.1 kann nicht abgelesen werden, ob sich die empirische Verteilung signifikant von der Normalverteilung unterscheidet. Dafür gibt es statistische Testverfahren, welche die Anpassungsgüte überprüfen. Ein Anpassungstest ist der Kolmogorov-Smirnov-Test (Schlittgen [32] S. 394-401). Bei diesem heuristischen Test wird der Abstand zwischen der empirischen und der theoretischen Verteilungsfunktion $|F_n(x) - F_0(x)|$ überprüft. Der größte Abstand ist definiert als

$$D = \sup_{-\infty < x < \infty} \left| \hat{F}_n(x) - F_0(x) \right|.$$

Der Kolmogorov-Smirnov-Test (als Test auf die Verteilungsform) kann nur bei Studien mit Effekthomogenität angewendet werden. Unter Heterogenität könnte der Test fälschlicherweise abgelehnt werden, da nicht geprüft wird, ob die logarithmischen Relativen Risiken in den Unterpopulationen normalverteilt sind. In der Tabelle (2.2) sind die multizentrischen Studien überprüft worden, bei denen der Homogenitätstest nicht abgelehnt

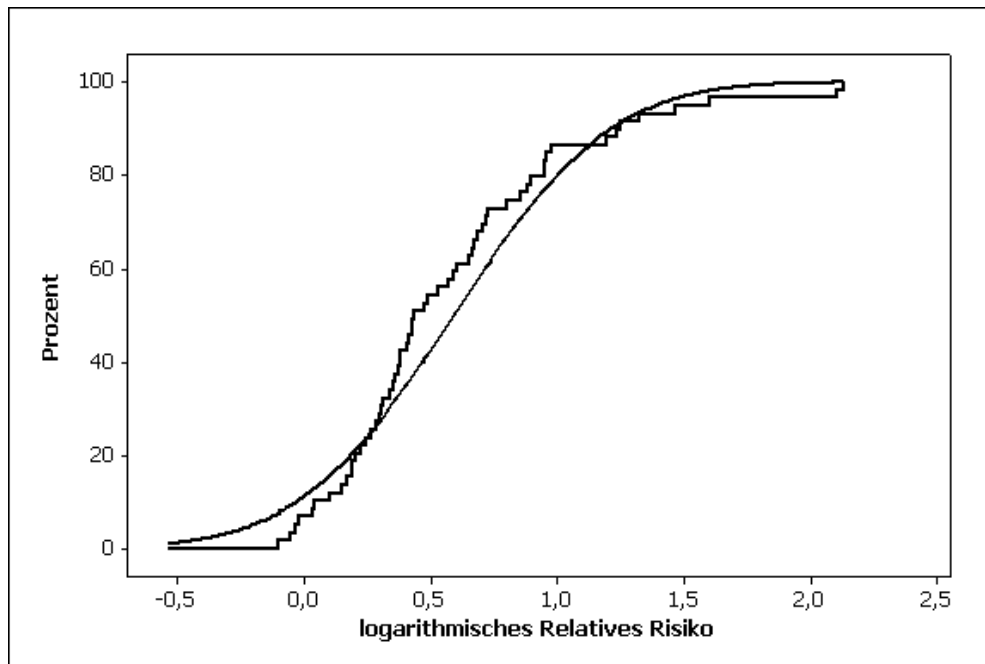


Abbildung 2.1: Empirische Verteilung des log Relativen Risikos im Vergleich mit der Normalverteilung (Daten sind aus Tabelle 1.12)

wurde. Bei der Studie 1.12 wird der Signifikanztest abgelehnt, da die Prüfgröße größer als der kritische Wert von 0.895 (aus Schlittgen [32] Seite 396, Kritische Werte der modifizierten Kolmogorov-Statistik bei Normalverteilung mit unbekanntem Parametern μ und σ) ist. Dennoch kann bei den anderen geprüften multizentrischen Studien der Anpassungstest nicht abgelehnt werden. Somit ist nach dem Kolmogorov-Smirnov-Test die erste wichtige Annahme bei 6 Studien erfüllt.

Des Weiteren gilt auch hier das Gesetz der großen Zahlen. Dieses besagt, je größer der Stichprobenumfang in jedem Zentrum ist, desto eher gilt die Normalverteilungsannahme des log. Relativen Risikos. Inwieweit sich dieses Gesetz auf die Schätzung der Heterogenität auswirkt, wird in der Simulationsstudie Kapitel 5.2 behandelt. Der vielleicht bedeutendste Nachteil in dem Modell ist, dass die Schätzung der Varianz zu schlecht sein könnte, um eine gute Approximation der wahren Varianz zu sein. Diese Ungenauigkeit in der Varianzschätzung wird besonders bei kleinen Stichprobenumfängen in das Modell miteinfließen.

Tabelle 2.2: Anpassungstest an die Normalverteilung bei einem 95% Konfidenzintervall (kritischer Wert bei 0.895)

Daten	D	Prüfgröße	Testergebnis
Tabelle 1.1	0.10157	0.479312	nicht ablehnen
Tabelle 1.4	0.15402	0.710501	nicht ablehnen
Tabelle 1.5	0.168302	0.464038	nicht ablehnen
Tabelle 1.6	0.0985919	0.550654	nicht ablehnen
Tabelle 1.11	0.169934	0.497003	nicht ablehnen
Tabelle 1.12	0.133894	1.02905	ablehnen
Tabelle 1.13	0.089264	0.697528	nicht ablehnen

2.3 Profil-Likelihoodansatz

Der PL-Ansatz ist für ein einfaches Kohortendesign in Böhning [7] erstmalig entwickelt worden. Eine Proportion bzw. Rate lässt sich sehr einfach durch eine Binomial- oder Poissonverteilung modellieren. Somit kann die Heterogenität in der Proportion bzw. Rate als Mischung von Binomial- oder Poissonverteilungen beschrieben werden. Dagegen lässt sich das Verhältnis zweier Raten durch keine elementare Verteilung mehr beschreiben und verbaut damit der Modellierung der Heterogenität den einfachen Zugang. Der Grund liegt in der Likelihoodfunktion. Unter der Annahme, dass die x_i^T [x_i^C] Poissonverteilt sind, mit $x_i^T \sim Poi(n_i^T p_i^T)$ [$x_i^C \sim Poi(n_i^C p_i^C)$] ergibt sich für das i -te Zentrum folgender Poissonlikelihood

$$L_{Poi}(p_i^T, p_i^C)_i = e^{-n_i^T p_i^T} \frac{(n_i^T p_i^T)^{x_i^T}}{x_i^T!} \times e^{-n_i^C p_i^C} \frac{(n_i^C p_i^C)^{x_i^C}}{x_i^C!} \quad (2.8)$$

Die parameterabhängige Poisson-Log-Likelihood ist

$$l_{Poi}(p_i^T, p_i^C)_i = x_i^T \log(p_i^T) - n_i^T p_i^T + x_i^C \log(p_i^C) - n_i^C p_i^C \quad (2.9)$$

Durch die Ersetzung von p_i^T durch $\theta_i p_i^C$ ($\theta_i = \frac{p_i^T}{p_i^C}$) wird der interessierende Parameter, das Relative Risiko eingeführt. Damit ergibt sich

$$\begin{aligned} l_{Poi}(\theta_i, p_i^C)_i &= x_i^T \log(\theta_i p_i^C) - n_i^T \theta_i p_i^C + x_i^C \log(p_i^C) - n_i^C p_i^C \\ &= x_i^T \log(\theta_i) + x_i \log(p_i^C) - (n_i^T \theta_i + n_i^C) p_i^C \\ \text{mit } x_i &= x_i^T + x_i^C \end{aligned} \quad (2.10)$$

Diese Log-Likelihoodfunktion (2.10) hängt von zwei Parametern, einmal vom Relativen Risiko und zum anderen vom Baselineparameter p_i^C oder auch Nuisance-Parameter genannt, ab. Der Nuisance-Parameter ist hier nicht von vordergründigem Interesse. Er ist aber notwendig für die vollständige Beschreibung der Likelihoodfunktion. Mit dem Profil-Maximum-Likelihood-Schätzer (PMLE) kann der Nuisance-Parameter aus dem Modell eliminiert werden. Dazu wird (2.10) nach p_i^C maximiert, indem die Nullgesetzte partielle Ableitung nach p_i^C aufgelöst wird. In dem Fall ist eine geschlossene Lösungsform möglich, mit

$$p_i^C = \frac{x_i}{n_i^C + n_i^T \theta_i}. \quad (2.11)$$

Der Profil-Maximum-Likelihood-Schätzer (2.11) wird nun in die Log-Likelihoodfunktion (2.10) eingesetzt. Das Ergebnis ist eine einfache und nur noch von θ_i abhängige Profil-Log-Likelihoodfunktion, mit der folgenden Form:

$$l_{Poi}(\theta_i)_i = x_i^T \log(\theta_i) - x_i \log(n_i^T \theta_i + n_i^C) \quad (2.12)$$

Alternativ kann (2.11) auch in (2.8) eingesetzt werden, mit

$$L_{Poi}^{Profil}(\theta_i)_i = e^{-x_i} \frac{\left(\frac{n_i^T \theta_i x_i}{n_i^C + n_i^T \theta_i}\right)^{x_i^T} \left(\frac{n_i^C x_i}{n_i^C + n_i^T \theta_i}\right)^{x_i^C}}{x_i^T! x_i^C!} \quad (2.13)$$

Wird auf (2.13) der Logarithmus angewendet erhält man wieder (2.12), wobei die parameterunabhängigen Faktoren ignoriert werden. Dieser Ansatz lässt sich auch auf den Binomiallikelihood anwenden. Dennoch ist der Aufwand weitaus höher, zumal der Profil-Maximum-Likelihood-Schätzer nicht mehr eindeutig ist. Da die Poissonverteilung eine hervorragende Approximation der Binomialverteilung ist, wird diese Modellfehlspezifikation, die nur bei Studien mit binomialverteilten Beobachtungen auftritt, in Kauf genommen.

2.4 Multi-Level-Modell

In diesem Kapitel wird das 2-Level-Model eingeführt. Die Klasse der Multi-Level-Modelle (ML-Modelle) stammt aus der Familie der verallgemeinerten linearen Modelle (GLM). Die GLM werden in dem Buch von McCullagh und Nelder [26] eingeführt, und sind eine Weiterentwicklung der linearen Modelle. Biostatistische Daten haben oft eine hierarchische Struktur. Diese Struktur kommt in multizentrischen Studien vor, wo Patienten in Zentren eingeteilt sind und die Zentren werden wiederum als eine Studie aufgefasst. Auch

in Tierpopulationen tritt in natürliche Weise eine hierarchische Struktur auf. So ist ein Tier bestimmt durch dessen Eltern, die Eltern sind Teil einer Gruppe oder Herde und diese können in bestimmten Gebiete oder Zonen vorkommen. Diese verschachtelte Struktur in den Beobachtungen wird im ML-Modell in Einheiten (Levels) klassifiziert. Auf die multizentrischen Studien angewendet, ergeben sich zwei Levels. Das erste Level kann mit folgenden Modell formuliert werden

$$\log \left(\frac{p_x}{1 - p_x} \right) = \text{logit}(p_x) = \alpha + \beta x \quad (2.14)$$

p ist die Erkrankungswahrscheinlichkeit und der Indikator x gibt an, ob p zum Behandlungs- ($x = 1$) oder zum Kontrollarm ($x = 0$) gehört. 2.14 ist essentiell ein einfaches Level-Modell und wird üblicherweise als logistischer Regressionsansatz in der Logitform bezeichnet. α ist in dieser Form der Baseline- und β der Effektparameter. Der Effektparameter ist hier das log. Odds Ratio (Kreuzproduktverhältnis) und nicht das log. Relative Risiko. Über den Intercept- (α) und den Effektparameter kann aber das Relative Risiko berechnet werden. Das Logitmodell wird gewöhnlich bei binomialverteilten Beobachtungen verwendet. Die Dichte eines Zentrums wird formuliert als

$$\begin{aligned} f_{ML_{Bin}}(p_i^T, p_i^C)_i &= \binom{n_i^T}{x_i^T} (p_i^T)^{x_i^T} (1 - p_i^T)^{n_i^T - x_i^T} \\ &\quad \times \binom{n_i^C}{x_i^C} (p_i^C)^{x_i^C} (1 - p_i^C)^{n_i^C - x_i^C}. \end{aligned} \quad (2.15)$$

Die Baselineparameter werden im nächsten Schritt durch die Modellvorgaben ersetzt, mit $p_x = \frac{e^{\alpha + x\beta}}{1 + e^{\alpha + x\beta}}$. Die resultierende Dichte ist

$$\begin{aligned} f_{ML_{Bin}}(\alpha, \beta)_i &= \binom{n_i^T}{x_i^T} \left(\frac{e^{\alpha + \beta}}{1 + e^{\alpha + \beta}} \right)^{x_i^T} \left(\frac{1}{1 + e^{\alpha + \beta}} \right)^{n_i^T - x_i^T} \\ &\quad \times \binom{n_i^C}{x_i^C} \left(\frac{e^{\alpha}}{1 + e^{\alpha}} \right)^{x_i^C} \left(\frac{1}{1 + e^{\alpha}} \right)^{n_i^C - x_i^C}. \end{aligned} \quad (2.16)$$

Im Falle von Poissonverteilten Beobachtungen wird dagegen auf das Logmodell zurückgegriffen. Dieses Multi-Level-Modell hat dementsprechend die Form:

$$\log(p_x) = \alpha + \beta x,$$

mit der Dichte

$$f_{ML_{Poi}}(p_i^T, p_i^C)_i = e^{-p_i^T n_i^T} \frac{(p_i^T n_i^T)^{x_i^T}}{x_i^T!} \times e^{-p_i^C n_i^C} \frac{(p_i^C n_i^C)^{x_i^C}}{x_i^C!}. \quad (2.17)$$

Wieder werden die Baselineparameter durch die Modellvorgaben ($p_x = e^{\alpha+\beta x}$) ersetzt, mit dem Ergebnis

$$f_{ML_{Poi}}(\alpha, \beta)_i = e^{-e^{\alpha+\beta} n_i^T} \frac{(e^{\alpha+\beta} n_i^T)^{x_i^T}}{x_i^T!} \times e^{-e^{\alpha} n_i^C} \frac{(e^{\alpha} n_i^C)^{x_i^C}}{x_i^C!}.$$

Hier ist der Effektparameter β gleichzeitig der Schätzer für das logarithmische Relative Risiko.

Als eine weitere Verallgemeinerung wird angenommen, dass die α - und β - Werte Zufallsvariablen sind und eine bestimmte Verteilung besitzen. Diese Abstraktion wird als zweites Level bezeichnet. Der zufällige Effekt zwischen den Zentren wird hier mit einer Mischverteilung modelliert. Das daraus resultierende Modell hat nun die Form:

$$f_{ML_{Bin}}(P)_i = \sum_{j=1}^m f_{ML_{Bin}}(\alpha_j, \beta_j)_i q_j$$

für den Binomialfall und

$$f_{ML_{Poi}}(P)_i = \sum_{j=1}^m f_{ML_{Poi}}(\alpha_j, \beta_j)_i q_j$$

für den Poissonfall, mit

$$P = \begin{pmatrix} \alpha_1 & \dots & \alpha_m \\ \beta_1 & \dots & \beta_m \\ q_1 & \dots & q_m \end{pmatrix}$$

Damit sind nun die alle drei Modelle vorgestellt. Im nächsten Kapitel werden für die vorgestellten Modelle die Schätzer der mischenden Verteilung abgeleitet.

Kapitel 3

Schätzen der Parameter

In dieser Arbeit wird die Maximum-Likelihood-Schätzung angewendet. Mit der Einführung der unbeobachteten Indikatorvariablen Z kann die vollständige Likelihoodfunktion bestimmt werden. Z gibt an, zu welcher Unterpopulation θ_j gehört ($z_{ij} = 1$ i -te Studie gehört zur j -ten Unterpopulation, sonst ist $z_{ij} = 0$). Die Likelihoodfunktion besitzt folgende Gestalt

$$L(y, P, z) = \prod_{i=1}^k \prod_{j=1}^m [f(y_i, \theta_j) q_j]^{z_{ij}}, \quad (3.1)$$

Die entsprechende Log-Likelihoodfunktion ist

$$l(y, P, z) = \sum_{i=1}^k \sum_{j=1}^m [z_{ij} \log(f(y_i, \theta_j)) + z_{ij} \log(q_j)]. \quad (3.2)$$

Der Vorteil dieser Form liegt darin, dass die Gewichte und das Effektmaß getrennt voneinander maximiert werden können. Die geschätzten Gewichte sind

$$\hat{q}_j = \frac{1}{k} \sum_{i=1}^k z_{ij} \quad (3.3)$$

Dieser Schätzer für die Gewichte gilt in allen drei Modellen. Der Effektschätzer wird nun für jedes Modell aus der Log-Likelihoodfunktion abgeleitet.

3.1 Parameterschätzung im Normalverteilungsmodell

Die vollständige Log-Likelihoodfunktion im ND-Ansatz ist:

$$l_{ND}(y, \Psi) = \sum_{i=1}^k \sum_{j=1}^m z_{ij} \log\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_i}\right) - z_{ij} \frac{(y_i - \log \theta_j)^2}{2\hat{\sigma}^2} + z_{ij} \log(q_j) \quad (3.4)$$

Zur Maximierung von (3.4) wird die Ableitung nach θ_j gebildet, mit

$$\frac{\partial l_{ND}(y, \Psi)}{\partial \theta_j} = \sum_{i=1}^k z_{ij} \frac{(y_i - \log \theta_j)}{\hat{\sigma}^2 \theta_j} = \frac{1}{\theta_j} \sum_{i=1}^k \frac{z_{ij} y_i}{\hat{\sigma}^2} - \frac{\log \theta_j}{\theta_j} \sum_{i=1}^k \frac{z_{ij}}{\hat{\sigma}^2}. \quad (3.5)$$

Durch Nullsetzen der Ableitung ergibt sich für den Schätzer des Relativen Risikos folgende Formel:

$$\hat{\theta}_{NDj} = e^{\frac{\sum_{i=1}^k \frac{z_{ij} y_i}{\hat{\sigma}^2}}{\sum_{i=1}^k \frac{z_{ij}}{\hat{\sigma}^2}}}$$

Damit ist der Schätzer für das logarithmische Relative Risiko ein gewichteter Mittelwert. Die Gewichte sind als die Inverse der Varianz definiert.

3.2 Parameterschätzung im Profil-Likelihood-Ansatz

Die vollständigen Log-Likelihoodfunktion ist im PL-Ansatz

$$l_{Poi}(x, \Psi) = \sum_{i=1}^k \sum_{j=1}^m z_{ij} (x_i^T \log(\theta_j) - x_i \log(n_i^T \theta_j + n_i^C)) + z_{ij} \ln(q_j) \quad (3.6)$$

Die Ableitung von (3.6) nimmt folgende Form an:

$$\frac{\partial l_{Poi}(x, \Psi)}{\partial \theta_j} = \sum_{i=1}^k \frac{z_{ij} x_i^T}{\theta_j} - \sum_{i=1}^k \frac{z_{ij} x_i n_i^T}{n_i^T \theta_j + n_i^C}$$

Leider ist für den Schätzer $\hat{\theta}_j$ keine geschlossene Lösungsform zugänglich, da die nullgesetzte Ableitung die Form besitzt:

$$\hat{\theta}_j = \frac{\sum_{i=1}^k z_{ij} x_i^T}{\sum_{i=1}^k \frac{z_{ij} x_i n_i^T}{n_i^T \theta_j + n_i^C}} \quad (3.7)$$

Möglich ist aber eine Fixpunktiteration, mit

$$\theta_{PMLEj} = \frac{\sum_{i=1}^k z_{ij} x_i^T n_i^C (n_i^T \theta_{PMLEj} + n_i^C)^{-1}}{\sum_{i=1}^k z_{ij} x_i^C n_i^T (n_i^T \theta_{PMLEj} + n_i^C)^{-1}} := \Gamma(\theta_{PMLEj}) \quad (3.8)$$

Es kann gezeigt werden, dass die Lösung von $\Gamma(\theta_{PMLEj})$ eindeutig und immer ein Maximum ist. Das besondere an dieser Iteration ist die Ähnlichkeit mit dem wohlbekannten Mantel-Haenszel-Schätzer. Wird die Iteration im Homogenitätsfall mit eins gestartet, dann ist das Ergebnis des ersten Iterationsschritts (3.9) sogar der exakte Mantel-Haenszel-Schätzer.

$$\Gamma(1) = \frac{\sum_{i=1}^k \frac{x_i^T n_i^C}{n_i^T + n_i^C}}{\sum_{i=1}^k \frac{x_i^C n_i^T}{n_i^T + n_i^C}} = \theta_{MH} \quad (3.9)$$

Unter anderem rechtfertigt diese wichtige Eigenschaft die Untersuchung dieses PL-Ansatzes.

3.3 Parameterschätzung im Multi-Level-Modell

In diesem Modell werden zwei Varianten betrachtet. Einmal unter der Annahme, dass die Beobachtungen binomialverteilt sind und zum anderen, dass die Beobachtungen Poissonverteilt sind. Unter der ersten Annahme ergibt sich folgende vollständige Likelihoodfunktion:

$$L_{MLBin}(P) = \prod_{i=1}^k \prod_{j=1}^m (q_j f_{MLBin}(\alpha_j, \beta_j)_i)^{z_{ij}}. \quad (3.10)$$

Der parameterabhängige Log-Likelihood von (3.10) ist

$$l_{MLBin}(P) = \sum_{i=1}^k \sum_{j=1}^m z_{ij} \log(q_k) + z_{ij} (x_i^C \alpha_j - n_i^C \log(1 + e^{\alpha_j}) + x_i^T (\alpha_j + \beta_j) - n_i^T \log(1 + e^{\alpha_j + \beta_j})) \quad (3.11)$$

Die Schätzer für α_j und β_j leiten sich aus den Nullgesetzten partiellen Ableitungen her, mit

$$\begin{aligned}\frac{\partial l_{MLBin}(P)}{\partial \beta_j} &= \sum_{i=1}^k z_{ij} \left(x_i^T - \frac{n_i^T e^{\alpha_j + \beta_j}}{1 + e^{\alpha_j + \beta_j}} \right) \stackrel{!}{=} 0 \\ e^{\alpha_j + \hat{\beta}_j} &= \frac{\sum_{i=1}^k z_{ij} x_i^T}{\sum_{i=1}^k z_{ij} (n_i^T - x_i^T)} \\ \hat{\beta}_j &= \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^T}{\sum_{i=1}^k z_{ij} (n_i^T - x_i^T)} \right) - \alpha_j\end{aligned}$$

$$\begin{aligned}\frac{\partial l_{MLBin}(P)}{\partial \alpha_j} &= \sum_{i=1}^k z_{ij} \left(x_i^C + x_i^T - \frac{n_i^C e^{\alpha_j}}{1 + e^{\alpha_j}} - \frac{n_i^T e^{\alpha_j + \beta_j}}{1 + e^{\alpha_j + \beta_j}} \right) \stackrel{!}{=} 0 \\ e^{\hat{\alpha}_j} &= \frac{\sum_{i=1}^k z_{ij} x_i^C}{\sum_{i=1}^k z_{ij} (n_i^C - x_i^C)} \\ \hat{\alpha}_j &= \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^C}{\sum_{i=1}^k z_{ij} (n_i^C - x_i^C)} \right)\end{aligned}\tag{3.12}$$

$$\hat{\beta}_j = \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^T \sum_{i=1}^k z_{ij} (n_i^C - x_i^C)}{\sum_{i=1}^k z_{ij} x_i^C \sum_{i=1}^k z_{ij} (n_i^T - x_i^T)} \right)\tag{3.13}$$

Nun kann anhand der Schätzer für α und β das Relative Risiko θ ermittelt werden, mit

$$\theta_j = \frac{\sum_{i=1}^k z_{ij} x_i^T \sum_{i=1}^k z_{ij} n_i^C}{\sum_{i=1}^k z_{ij} n_i^T \sum_{i=1}^k z_{ij} x_i^C}.$$

Unter Poissonverteilten Beobachtungen nimmt die vollständige Log-Likelihood folgende Form an:

$$l_{MLPoi}(P) = \prod_{i=1}^k \prod_{j=1}^m (q_j f_{MLPoi}(P)_i)^{z_{ij}}.$$

Analog ist die parameterabhängige Log-Likelihood

$$\begin{aligned}l_{MLPoi}(P) &= \sum_{i=1}^k \sum_{j=1}^m z_{ij} \log(q_k) \\ &\quad + z_{ij} \left(-e^{\alpha_j} n_i^C + x_i^C \alpha_j - e^{\alpha_j + \beta_j} n_i^T + x_i^T (\alpha_j + \beta_j) \right)\end{aligned}\tag{3.14}$$

Daraus lassen sich nun folgende Schätzer ableiten:

$$\alpha_j = \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^C}{\sum_{i=1}^k z_{ij} n_i^C} \right) \quad (3.15)$$

$$\beta_j = \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^T \sum_{i=1}^k z_{ij} n_i^C}{\sum_{i=1}^k z_{ij} n_i^T \sum_{i=1}^k z_{ij} x_i^C} \right). \quad (3.16)$$

Eine Eigenschaft dieser zwei ML-Ansätze ist, dass sie unter Homogenität den gleichen Schätzer für das Relative Risiko liefern. Unter Heterogenität treten geringe Unterschiede auf. Dabei gilt die Aussage, je kleiner die Baselineparameter p_i^C und p_i^T in der multizentrischen Studie sind, desto ähnlicher sind sich die Schätzer für das Relative Risiko. Wäre die Zentrenzugehörigkeit bekannt und müsste nicht geschätzt werden, würden die zwei ML-Ansätze wieder die gleichen Ergebnisse liefern.

3.4 EM-Algorithmus bei einer Mischverteilung

Der EM (Expectation Maximization)-Algorithmus wurde in dem Artikel von Dempster [12] eingeführt und benannt. Neben dem Artikel beinhaltet das Buch von McLachlan and Krishnan [27] eine umfassende Darstellung der Entwicklung und Weiterentwicklung des EM-Algorithmus.

Die vollständige Log-Likelihoodfunktion (2.3) enthält die unbeobachtete und unbekannte Indikatorvariable Z . Im EM-Algorithmus besteht nun der E-Schritt darin, die Indikatorvariable durch den Erwartungswert $E(Z)$ zu ersetzen. Dieser ist $E(z_{ij}|y_i, q) = P_q(Z_{ij} = 1|Y_i = y_i)$. Die Anwendung des Bayesschen Theorems führt zu

$$\begin{aligned} P_q(Z_{ij} = 1|Y_i = y_i) &= \frac{P_q(Y_i = y_i|Z_{ij} = 1)P_q(Z_{ij} = 1)}{\sum_{j^*} P_q(Y_i = y_i|Z_{ij^*} = 1)P_q(Z_{ij^*} = 1)} \\ &= \frac{f(y_i, \theta_j)q_j}{\sum_{j^*} f(y_i, \theta_{j^*})q_{j^*}} \\ &= : e_{ij} \end{aligned} \quad (3.17)$$

Das Austauschen der Indikatorvariable Z durch e_{ij} führt zu der erwarteten vollständigen Log-Likelihoodfunktion

$$E(l(y, \Psi)) = \sum_{i=1}^k \sum_{j=1}^m [e_{ij} \log(f(y, \theta_j)) + e_{ij} \log(q_j)]. \quad (3.18)$$

Die Maximierung der Funktion (3.18) führt zum M-Schritt. Die Lösung für die Anteilswerte q_j ist in jedem der drei Modelle gleich, mit

$$q_j^{(new)} = \frac{1}{k} \sum_{i=1}^k e_{ij} = \frac{1}{k} \sum_{i=1}^k \frac{f(y_i, \theta_j) q_j}{\sum_{j^*} f(y_i, \theta_{j^*}) q_{j^*}} \quad (3.19)$$

Die Schätzung für das Relative Risiko ist je nach Modell folgend aufgeführt.

- ND-Modell:

$$\theta_{NDj} = \frac{\sum_{i=1}^k \frac{e_{ij} x_i}{\hat{\sigma}^2}}{\sum_{i=1}^k \frac{e_{ij}}{\hat{\sigma}^2}} \quad (3.20)$$

- PL-Ansatz:

$$\theta_{PMLEj} = \frac{\sum_{i=1}^k e_{ij} x_i^T n_i^C (n_i^T \theta_{PMLEj} + n_i^C)^{-1}}{\sum_{i=1}^k e_{ij} x_i^C n_i^T (n_i^T \theta_{PMLEj} + n_i^C)^{-1}} \quad (3.21)$$

- ML-Modell (Binomial-Likelihood und Poisson-Likelihood):

$$\theta_{MLj} = \frac{\sum_{i=1}^k e_{ij} x_i^T \sum_{i=1}^k e_{ij} n_i^C}{\sum_{i=1}^k e_{ij} x_i^C \sum_{i=1}^k e_{ij} n_i^T} \quad (3.22)$$

Die Struktur des Algorithmus ist in Kurzfassung gegeben als:

Schritt 0: q und θ (α und β für das ML-Modell) sind die Startwertvektoren

Schritt 1: Berechne den E-Schritt nach (3.17)

Schritt 2: Berechne den M-Schritt nach (3.19), welcher zu q^{new} führt und je nach Modell (3.20, 3.21 und 3.22), welcher zu θ^{new} führt

Schritt 3: Setze $q = q^{new}$ und $\theta = \theta^{new}$. Gehe zum Schritt 1¹.

3.5 Eigenschaften nichtparametrischer Mischverteilungen

Bei der Maximierung treten zwei Probleme auf. Die eine Schwierigkeit besteht in dem Vorkommen mehrerer lokaler Extremwerte. Der EM-Algorithmus konvergiert immer zu einem lokalen Maximum, das aber nicht gleichzeitig das globale Maximum darstellt. Dies

¹Ein entsprechendes Abbruchkriterium ist im Schritt 3 notwendig.

hängt von der Wahl der Startwerte ab. Das zweite Problem besteht darin, dass die Anzahl der Mischungskomponenten unbekannt ist. Ein wichtiges Hilfsmittel zur Lösung dieser Probleme ist die Gradientenfunktion mit dem allgemeinen Maximum-Likelihood Satz über Mischverteilungen. Mit der Gradientenfunktion kann überprüft werden, ob die Anzahl der Mischungskomponenten ausreichend ist und ob die Lösung ein globales Maximum ist. Im nächsten Abschnitt werden die Gradientenfunktion und danach der ML-Satz eingeführt.

3.5.1 Gradientenfunktion

Ω ist die Menge aller diskreten Wahrscheinlichkeitsverteilungen. P und Q sind zwei diskrete Wahrscheinlichkeitsfunktionen aus Ω . Die Richtungsableitung der Log-Likelihoodfunktion $l(P)$ ² nach P in Richtung Q und kann geschrieben werden als

$$\begin{aligned}\Phi(P, Q) &= \lim_{\alpha \rightarrow 0} \frac{l((1-\alpha)P + \alpha Q) - l(P)}{\alpha} := \left. \frac{\partial}{\partial \alpha} l((1-\alpha)P + \alpha Q) \right|_{\alpha=0} \\ \frac{\partial}{\partial \alpha} l((1-\alpha)P + \alpha Q) \Big|_{\alpha=0} &= \sum_{i=1}^k \frac{f(y_i, Q) - f(y_i, P)}{f(y_i, P)} \\ &= \sum_{i=1}^k \frac{f(y_i, Q)}{f(y_i, P)} - k\end{aligned}\tag{3.23}$$

Gewöhnlich wird die Richtungsableitung durch k geteilt. Die Gradientenfunktion $d(\theta, P)$ leitet sich aus der Richtungsableitung her, mit

$$\begin{aligned}d(\theta, P) &: = \frac{1}{k} \sum_{i=1}^k \frac{f(y_i, Q_\theta)}{f(y_i, P)} \\ \text{mit } Q_\theta &= \begin{pmatrix} \theta \\ 1 \end{pmatrix}\end{aligned}$$

Je nach Modell hat die Gradientenfunktion folgende Form:

- ND-Modell:

$$d(\theta, P) = \frac{1}{k} \sum_{i=1}^k \frac{e^{-\frac{(y_i - \log \theta)^2}{2\sigma_i^2}}}{\sum_{j=1}^m e^{-\frac{(y_i - \log \theta_j)^2}{2\sigma_i^2}} q_j}\tag{3.24}$$

² $l(P) = \sum_i^n \log f(y_i, P)$

- PL-Ansatz:

$$d(\theta, P) = \frac{1}{k} \sum_{i=1}^k \frac{\frac{\theta^{x_i^T}}{(n_i^T \theta + n_i^C)^{x_i}}}{\sum_{j=1}^m \frac{\theta_j^{x_i^T}}{(n_i^T \theta_j + n_i^C)^{x_i}} q_j} \quad (3.25)$$

- ML-Modell (Binomial-Likelihood):

$$d(\vartheta^*, P^*) = \frac{1}{k} \sum_{i=1}^k \frac{\frac{e^{x_i^T \alpha + x_i^C (\alpha + \beta)}}{(1+e^\alpha)^{n_i^T} (1+e^{\alpha+\beta})^{n_i^C}}}{\sum_{j=1}^m \frac{e^{x_i^T \alpha_j + x_i^C (\alpha_j + \beta_j)}}{(1+e^{\alpha_j})^{n_i^T} (1+e^{\alpha_j + \beta_j})^{n_i^C}} q_j} \quad (3.26)$$

- ML-Modell (Poisson-Likelihood):

$$d(\vartheta^*, P^*) = \frac{1}{k} \sum_{i=1}^k \frac{e^{-e^\alpha n_i^C} (e^\alpha)^{x_i^C} e^{-e^{\alpha+\beta} n_i^T} (e^{\alpha+\beta})^{x_i^T}}{g(P^*)} \quad (3.27)$$

$$g(P^*) = \sum_{j=1}^m e^{-e^{\alpha_j} n_i^C} (e^{\alpha_j})^{x_i^C} e^{-e^{\alpha_j + \beta_j} n_i^T} (e^{\alpha_j + \beta_j})^{x_i^T} q_j$$

$$\text{mit } \vartheta^* = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \text{ und } P^* = \begin{pmatrix} \alpha_1 & \dots & \alpha_m \\ \beta_1 & \dots & \beta_m \\ q_1 & \dots & q_m \end{pmatrix}$$

3.5.2 Der allgemeine Maximum-Likelihood Satz von Mischverteilungen

In der Theorie der nichtparametrischen Verteilungsmodelle wird üblicherweise folgende wichtige Definition verwendet.

Definition 1 Jedes \hat{P} , welches die Log-Likelihoodfunktion $l(P)$ maximiert, wird Nicht-Parametrischer-Maximum-Likelihood-Schätzer (NPMLE) genannt.

Der folgende allgemeine Maximum-Likelihood Satz von Mischverteilungen (Böhning [8]) ist ein wichtiges Konstrukt in der nichtparametrischen Mischverteilungsstatistik.

Lemma 1 P und Q sind diskrete Verteilungsfunktionen. Es gilt

$$\Phi(P, Q) \geq l(Q) - l(P)$$

Beweis. $l(P)$ ist eine konkave Funktion (Böhning [8], Seite 36-37). ■

Satz 2 \hat{P} ist NPMLE dann und nur dann, wenn gilt

$$d(\theta, \hat{P}) = \frac{1}{k} \sum_{i=1}^k \frac{f(y_i, Q_\theta)}{f(y_i, P)} \leq 1$$

für alle θ im Parameterraum.

Beweis.

1. „ \implies “

Wenn \hat{P} NPMLE ist, dann ist

$$\begin{aligned} \frac{l((1-\alpha)\hat{P} + \alpha Q) - l(\hat{P})}{\alpha} &\leq 0 \text{ für alle } \alpha \in [0, 1] \text{ und alle } Q \in \Omega \\ &\implies \Phi(\hat{P}, Q) \leq 0 \\ * &\implies \Phi(\hat{P}, Q_\theta) \leq 0 \end{aligned}$$

*) Q_θ ist in der Menge Q enthalten.

2. „ \impliedby “

Wenn $\Phi(P, Q_\theta) \leq 0$ für alle θ aus dem Parameterraum

$$\rightarrow 0 \geq \sup_{\theta} \Phi(P, Q_\theta)^{**} \geq \sup_Q \Phi(P, Q) \geq \sup_Q l(Q) - l(P) \quad (3.28)$$

da $\sup_Q l(Q) - l(P) \geq 0$, folgt aus der Ungleichung 3.28, dass P nur der NPMLE sein kann und

$$0 = \sup_{\theta} \Phi(P, Q_\theta) = \sup_Q \Phi(P, Q).$$

**))

$$\Phi(P, Q) = \sum_j q_j \Phi(P, Q_{\theta_j}) \leq \sum_j q_j \sup_{\theta} \Phi(P, Q_\theta) = \sup_{\theta} \Phi(P, Q_\theta)$$

diese Ungleichung gilt für jedes Q aus Ω , speziell auch für das Q welches $\Phi(P, Q)$ maximiert.

Damit endet der Beweis. ■

Mit Hilfe dieses Satzes kann nun ein Algorithmus entwickelt werden, der als Abbruchkriterium prüft, ob der NPMLE erreicht ist. Zwei Algorithmen mit globaler Konvergenz sind bei Böhning [8, 9] enthalten, zum einen der VDM (Vertex Direction Method)-

Algorithmus und zum anderen die VEM-Methode (Vertex Exchange Method). Der Vorteil dieser Algorithmen liegt in der globalen Konvergenz. Nachteilig ist die künstliche Aufblähung der Komponentenanzahl. Die Anzahl der Komponenten kann aber über eine Clusterbildung wieder reduziert werden. Ist der NPML erreicht, so gibt dieser auch gleichzeitig die Anzahl der Mischungskomponenten an. Damit sind theoretisch die zwei Probleme (Anzahl der Mischungskomponenten und globales Maximum) gelöst.

3.6 EM-Algorithmus mit Gradientenfunktions-Update

3.6.1 EM-Algorithmus mit Gradientenfunktions-Update bei bekannter Anzahl der Komponenten

In der Publikation von Böhning [9] wird ein Algorithmus mit Gradientenfunktions-Update (EMGFU) mit bekannter Anzahl der Mischungskomponenten vorgeschlagen. Die Idee besteht darin, den EM-Algorithmus mit der Gradientenfunktion zu kombinieren. Im Folgenden wird der EMGFU in Kurzfassung aufgeführt:

Schritt 0: Wähle die Anzahl der Komponenten m und den Startwertvektor

$$P = \begin{pmatrix} \theta_1 & \dots & \theta_m \\ q_1 & \dots & q_m \end{pmatrix}.$$

Schritt 1: Berechne mit dem EM-Algorithmus P_{EM}

Schritt 2: Bestimme θ_{\max} , indem $d(\theta, P_{EM})$ in θ maximiert wird.

Schritt 3: Bestimme θ_{\min} , so dass $l(P_{EM} + P_{EM}(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}])$ das größte Element in der Menge

$$\{l(P_{EM} + P_{EM}(\theta_j)[Q_{\theta_{\max}} - Q_{\theta_j}]) \mid j = 1, \dots, m\}$$

ist.

Schritt 4: $P = P_{EM} + P_{EM}(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}]$. Wenn $l(P) > l(P_{EM})$, dann gehe zu Schritt 1, sonst Stopp.

Bemerkung 1 Die inhaltliche Bedeutung von $P + P(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}]$ ist, dass in dem Vektor P das Element θ_{\min} durch θ_{\max} ersetzt wird.

Da die Anzahl der Komponenten im Allgemeinen unbekannt ist, kann der EMGFU nur im Spezialfall verwendet werden. Es gibt aber die Möglichkeit den EMGFU für den allgemeinen Fall geeignet zu modifizieren.

3.6.2 EM-Algorithmus mit Gradientenfunktions-Update bei unbekannter Anzahl der Komponenten (gEMGFU)

Die Idee ist, dass zuerst der homogene Fall betrachtet wird. Das heißt es gibt nur eine Komponente. Der Startwert für θ kann beliebig gewählt werden. Danach wird mit dem EM-Algorithmus P_{EM}^1 berechnet. Wie im Schritt 2 beim EMGFU wird θ_{\max} bestimmt. Ist $d(\theta_{\max}, P_{EM}^1) \leq 1$ kann der Algorithmus abgebrochen werden, da der NPMLE erreicht ist. Schritt 3 wird vom EMGFU unverändert übernommen. Der Schritt 4 wird ebenso beibehalten und erweitert. Zuerst wird der Austauschschritt durchgeführt, was zu einem P^{l*} führt. Ist $l(P_{EM}^{l*}) > l(P_{EM}^l)$, dann war der Austauschschritt erfolgreich und es wird zu Schritt 2 gesprungen. Wenn $l(P_{EM}^{l*}) \leq l(P_{EM}^l)$, dann konnte keine Verbesserung durch den Austausch erreicht werden. Im Schritt 5 wird nun eine Komponente hinzugefügt. Das erfolgt nach dem Prinzip des VEM-Algorithmus. Dazu folgende Illustration: $P^3 = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ q_1 & q_2 & q_3 \end{pmatrix}$, $\theta_{\max} = \theta_4$ und $\theta_{\min} = \theta_2$. θ_{\max} wird nun zu P^3 hinzugefügt. Somit entsteht $P^4 = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 \\ q_1 & (1-\alpha)q_2 & q_3 & \alpha q_2 \end{pmatrix}$. Da sich die Gewichte zu eins aufsummieren müssen, wird das Gewicht von θ_{\min} in $(1-\alpha)q_2$ und in αq_2 geteilt. Die neu hinzugefügte Komponente erhält das Gewicht αq_2 und θ_{\min} erhält das Gewicht $(1-\alpha)q_2$. Für die Schätzung von α gibt es zwei Vorschläge. Eine Möglichkeit besteht darin, $l(P^l + P^l \alpha(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}])$ in α zu maximieren. Alternativ kann die Approximationsformel von Böhning [9] verwendet werden. Sie lautet

$$\alpha_{\max} = \frac{\sum_{i=1}^k g(y_i, \theta_{\max}, P)}{\sum_{i=1}^k g(y_i, \theta_{\max}, P)^2} \quad (3.29)$$

mit $g(y, \theta, P) = \frac{f(y, \theta) - f(y, P)}{f(y, P)}$.

Zusammenfassend ist nachstehender Algorithmus entstanden:

Schritt 0: Wähle den Startwertvektor $P^1 = \begin{pmatrix} \theta \\ 1 \end{pmatrix}$

Schritt 1: Berechne mit dem EM-Algorithmus P_{EM}^l

Schritt 2: Bestimme θ_{\max} , indem $d(\theta, P_{EM}^l)$ in θ maximiert wird.

Schritt 3: Bestimme θ_{\min} , so dass $l(P_{EM}^l + P_{EM}^l(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}])$ das größte Element in der Menge

$$\{l(P_{EM}^l + P_{EM}^l(\theta_j)[Q_{\theta_{\max}} - Q_{\theta_j}]) \mid j = 1, \dots, m\}$$

ist.

Schritt 4: Bestimme $P^{l*} = P_{EM}^l + P_{EM}^l(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}]$. Berechne mit EM-Algorithmus P_{EM}^{l*} . Ist $l(P_{EM}^{l*}) > l(P_{EM}^l)$, dann $P_{EM}^l = P_{EM}^{l*}$ und gehe zu Schritt 2

Schritt 5: Bestimme α nach (3.29), und bestimme $P^{l+1} = P_{EM}^l + \alpha P_{EM}^l(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}]$. Gehe zu Schritt 1

Für diesen Maximierungsalgorithmus kann kein Konvergenzbeweis angegeben werden. Da sich der gEMGFU-Algorithmus aus dem EM-Algorithmus und der global-konvergenten *Vertex Exchange* Methode zusammensetzt, liegt die Vermutung einer globalen Konvergenz des gEMGFU sehr nahe. In den später folgenden Simulationsexperimenten und bei den hier vorgestellten multizentrischen Studien konvergierte der Algorithmus in jedem Fall zum nichtparametrischen Maximum-Likelihood-Schätzer. Ein wichtiger Vorteil des gEMGFU gegenüber dem EM-Algorithmus ist, dass keine expliziten Startwerte notwendig sind. Einen weiteren Vorteil stellen die Zwischenergebnisse dar. Es zeigte sich, dass in den hier dargestellten Zwischenergebnissen der multizentrischen Studien die Likelihoodfunktion in derselben Dimension nicht weiter verbessert werden konnte, auch wenn der nichtparametrische Maximum-Likelihood-Schätzer noch nicht erreicht ist. Das heißt, die Likelihoodfunktion kann nur verbessert werden, wenn eine weitere Komponente hinzugefügt wird. Ein Nachteil besteht in der langsamen Konvergenzgeschwindigkeit des EM-Algorithmus, der im gEMGFU in jedem Durchlauf neu angewendet wird.

3.7 gEMGFU in FORTRAN

Als Programmiersprache zur Schätzung der Parameter wird FORTRAN benutzt. FORTRAN benötigt für numerische Rechenoperationen sehr wenig Zeit. Für besonders aufwendige Berechnungen, wie in unserem Fall ist diese Programmiersprache besonders gut geeignet. Doch die veraltete MS-DOS Umgebung des Programms wirkt sich nachteilig aus, da PC-Nutzer heutzutage mit Maus und Menüsteuerung arbeiten. In der FORTRAN-Entwicklungsumgebung *FORTRAN PowerStation 4.0* sind Programmroutinen integriert, wie zum Beispiel *QuickWin*, mit deren Hilfe leicht ein menügesteuertes Programm erstellt

werden kann. Dennoch sind die Möglichkeiten für die Entwicklung eines benutzerfreundlichen Programms sehr begrenzt. Für komplexere Programme wird der Programmierer schnell an die Grenzen von FORTRAN stoßen und auf die Programmiersprache C oder C++ zurückgreifen müssen.

In diesem Abschnitt wird die Umsetzung von bestimmten Teilen des Algorithmus in FORTRAN beschrieben. Besonderes hervorgehoben werden zum einen die numerische Stabilität der einzelnen Berechnungsschritte und zum anderen die Maximierung der Gradientenfunktion.

3.7.1 Numerische Stabilität

An vielen Stellen kommt es im gEMGFU vor, dass die Dichte oder der Log-Likelihood der Mischverteilung berechnet werden muss. Als Beispiel dient der E-Schritt (3.17) im EM-Algorithmus. Beim PL-Ansatz hat der E-Schritt folgende Form

$$e_{ij} = \frac{\frac{\theta_j^{x_i^T}}{(n_i^C + n_i^T \theta_j)^{x_i}} q_j}{\sum_{j^*=1}^m \frac{\theta_{j^*}^{x_i^T}}{(n_i^C + n_i^T \theta_{j^*})^{x_i}} q_{j^*}}. \quad (3.30)$$

Es kann vorkommen, dass Nenner und Zähler von (3.30) sehr klein sind, wie zum Beispiel beim 4. Zentrum in der Beta-Blocker-Studie (Tabelle 1.1 mit $x_4^T = 102$; $x_4^C = 127$; $n_4^T = 1533$; $n_4^C = 1520$). Mit zwei Mischungskomponenten ($\theta_1 = 0.7685$; $\theta_2 = 1.128979$; $q_1 = 0.905401$; $q_2 = 0.094591$) ergeben sich für den Zähler und den Nenner bei e_{41} folgende Werte:

$$\begin{aligned} \frac{\theta_1^{x_4^T}}{n_4^C + n_4^T \theta_1} q_1 &= 3.8 \times 10^{-798} \\ \frac{\theta_1^{x_4^T}}{n_4^C + n_4^T \theta_1} q_1 + \frac{\theta_2^{x_4^T}}{n_4^C + n_4^T \theta_2} q_2 &= 3.813278 \times 10^{-798} \end{aligned}$$

FORTRAN rundet diese kleinen Zahlen ab, so dass am Ende das Programm versucht 0/0 zu berechnen. Das führt zu einem Fehler. Durch folgende Umformungen der Formel (3.30)

kann dieses Problem vermieden werden.

$$\begin{aligned}
 e_{ij}^{PL} &= \frac{\frac{\theta_j^{x_i^T}}{(n_i^C + n_i^T \theta_j)^{x_i}} q_j}{\sum_{j^*=1}^m \frac{\theta_{j^*}^{x_i^T}}{(n_i^C + n_i^T \theta_{j^*})^{x_i}} q_{j^*}} \\
 &= \left(\sum_{j^*=1}^m \exp \left(x_i \log \left(\frac{n_i^C + n_i^T \theta_j}{n_i^C + n_i^T \theta_{j^*}} \right) + x_i^T \log \frac{\theta_{j^*}}{\theta_j} + \log \frac{q_{j^*}}{q_j} \right) \right)^{-1}.
 \end{aligned}$$

Für den ND-Ansatz ergibt sich die Formel

$$\begin{aligned}
 e_{ij}^{ND} &= \frac{\frac{1}{\sqrt{2\pi\hat{\sigma}_i}} e^{-\frac{(y_i - \log \theta_j)^2}{2\hat{\sigma}_i^2}} q_j}{\sum_{j^*=1}^m \frac{1}{\sqrt{2\pi\hat{\sigma}_i}} e^{-\frac{(y_i - \log \theta_{j^*})^2}{2\hat{\sigma}_i^2}} q_{j^*}} \\
 &= \left(\sum_{j^*=1}^m \exp \frac{(y_i - \log \theta_j)^2 - (y_i - \log \theta_{j^*})^2}{2\hat{\sigma}_i^2} + \log \frac{q_{j^*}}{q_j} \right)^{-1},
 \end{aligned}$$

und für das ML-Modell ist

$$\begin{aligned}
 e_{ij}^{MLB} &= \frac{\frac{e^{x_i^C \alpha_j}}{(1+e^{\alpha_j})^{n_i^C}} \times \frac{e^{x_i^T (\alpha_j + \beta_j)}}{(1+e^{\alpha_j + \beta_j})^{n_i^T}} q_j}{\sum_{j^*=1}^m \frac{e^{x_i^C \alpha_{j^*}}}{(1+e^{\alpha_{j^*}})^{n_i^C}} \times \frac{e^{x_i^T (\alpha_{j^*} + \beta_{j^*})}}{(1+e^{\alpha_{j^*} + \beta_{j^*}})^{n_i^T}} q_{j^*}} \\
 &= \left(\sum_{j^*=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1}, \text{ mit} \\
 t_1 &= x_i^C (\alpha_{j^*} - \alpha_j) + x_i^T (\alpha_{j^*} + \beta_{j^*} - \alpha_j - \beta_j) \\
 t_2 &= n_i^C \log \left(\frac{1 + e^{\alpha_j}}{1 + e^{\alpha_{j^*}}} \right) + n_i^T \log \left(\frac{1 + e^{\alpha_j + \beta_j}}{1 + e^{\alpha_{j^*} + \beta_{j^*}}} \right) \\
 t_3 &= \log \left(\frac{q_{j^*}}{q_j} \right)
 \end{aligned}$$

$$\begin{aligned}
e_{ij}^{MLPoi} &= \frac{e^{-e^{\alpha_j} n_i^C} (e^{\alpha_j})^{x_i^C} \times e^{-e^{\alpha_j + \beta_j} n_i^T} (e^{\alpha_j + \beta_j})^{x_i^T} \times q_j}{\sum_{j^*=1}^m e^{-e^{\alpha_{j^*}} n_i^C} (e^{\alpha_{j^*}})^{x_i^C} \times e^{-e^{\alpha_{j^*} + \beta_{j^*}} n_i^T} (e^{\alpha_{j^*} + \beta_{j^*}})^{x_i^T} \times q_{j^*}} \\
&= \left(\sum_{j^*=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1} \\
t_1 &= n_i^C (e^{\alpha_j} - e^{\alpha_{j^*}}) + n_i^T (e^{\alpha_j + \beta_j} - e^{\alpha_{j^*} + \beta_{j^*}}) \\
t_2 &= x_i^C (\alpha_{j^*} - \alpha_j) + x_i^T (\alpha_{j^*} + \beta_{j^*} - \alpha_j - \beta_j) \\
t_3 &= \log \left(\frac{q_{j^*}}{q_j} \right)
\end{aligned}$$

Wie im ND-Ansatz könnte es auch im PL- und ML-Modell passieren, dass bei bestimmten Konstellationen durch Null geteilt wird. Beim PL-Modell tritt dieser Fall ein, wenn im Homogenitätsfall im Kontrollarm in jedem Zentrum keine Fälle beobachtet worden sind. Dieser Fall ist zwar rein hypothetisch, kann aber in einer Simulationsstudie durchaus vorkommen. Im Heterogenitätsfall kann, selbst wenn nur in einem Zentrum im Kontrollarm kein einziger Fall aufgetreten ist, ebenfalls durch Null geteilt werden. Dies tritt ein (siehe 3.8), wenn alle $e_{ij} = 0$ sind und das $e_{nj} = 1$ (n ist das Zentrum wo $x_n^C = 0$). Das heißt, das n -te Zentrum wird als einziges zu einer Mischungskomponente zugewiesen. Auch wenn im Behandlungsarm in einigen Zentren keine Fälle auftraten, könnten diese einer Mischungskomponente zugewiesen werden, welche ein Relatives Risiko von Null schätzt. Ein Relatives Risiko von Null ist aber nicht interpretierbar, so dass auch dieser Fall ausgeschlossen werden muss. Um diese Fehlerquellen zu beseitigen wird in jedem Modell zu jeder Zelle $\frac{1}{2}$ addiert, wenn $x_i = 0$. Zusätzlich zu den beschriebenen Fehlerquellen, kann es im ML-Modell vorkommen, dass bei $n_i = x_i$ durch Null geteilt wird (siehe 3.12 und 3.13). Daher wird im ML-Modell, wenn $x_i = 0$ oder $n_i = x_i$ zu jeder Zelle $\frac{1}{2}$ addiert. Im ND-Ansatz (nur bei binomialverteilten Beobachtungen) tritt bei $x_i^T = n_i^T$ und $x_i^C = n_i^C$ der Fall ein, dass die geschätzte Varianz im i -ten Zentrum gleich Null ist. Da die Gewichte des Schätzers für das Relative Risiko aus dem Inversen der Varianz gebildet werden, darf diese Konstellation nicht auftreten. Somit wird in dem Fall wieder $\frac{1}{2}$ zu jeder Zelle addiert. Diese Verfahrensweise birgt in sich den Nachteil, dass besonders wenn sehr häufig die Konstante addiert wird, ein Bias hinzugefügt wird. Als Beispiel dazu dienen die beiden Abbildungen 3.1 und 3.2. Die Bilder stammen aus der Simulation 5 unter Homogenität. An verschiedenen Punkten im Intervall von 0.01 bis 0.99 ist der Bias vom Relativen Risiko des ML- und PL-Ansatzes berechnet worden. Als Grundlage dieser Bootstrapsimulation wird die CALGB-Studie verwendet. Die Funktionsweise der Simulation wird in einem späteren Kapitel noch genau beschrieben. Deutlich ist im Bild 3.2 zu

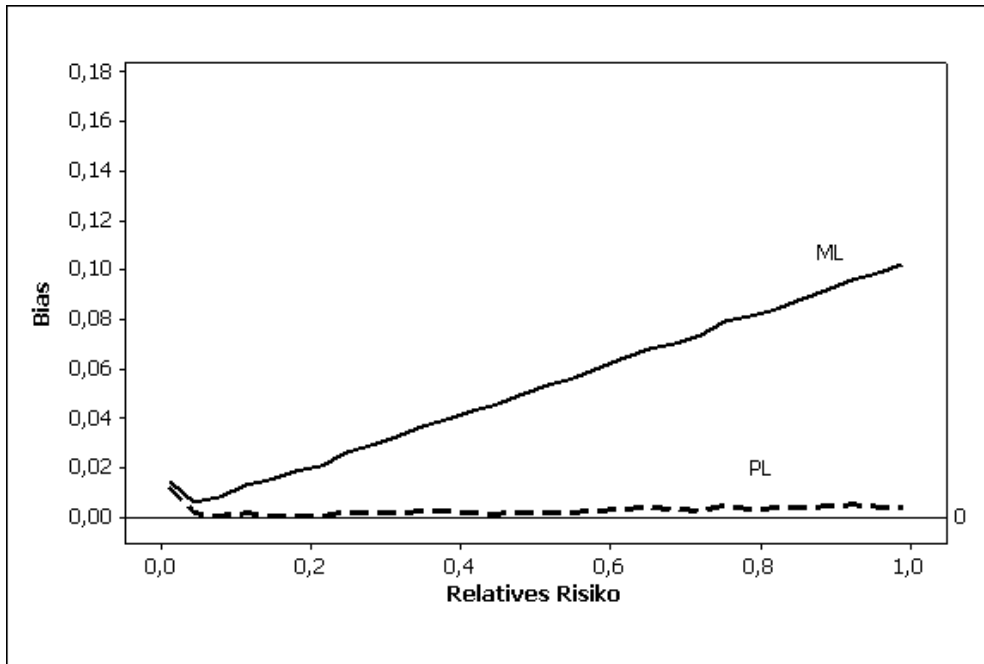


Abbildung 3.1: CALGB-Studie (Bias vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;0.99]$ berechnet, die generierten x^T , x^C -Werte sind binomialverteilt, additive Konstante ist 0.00000001)

erkennen, wie sich der Bias hauptsächlich bei kleineren Relativen Risiko-Werten erhöht hat. Es stellt sich hier die Frage, warum wird die Konstante 0.5 und nicht 0.00000001 gewählt. Die Antwort liegt in der Modellierung der unbeobachteten Heterogenität. Bei einer Konstanten mit einem Wert von 0.00000001, bekämen die Zentren mit $x_i = 0$ einen sehr starken Einfluss und würden einer separaten Komponente zugeordnet. Diese Komponente hätte dann extreme Werte. Entweder sie strebt gegen Null oder gegen unendlich. Besonders würde es beim PL-Modell in der Komponentenanalyse 5.2 auffallen, weil dort dann die erste Komponente stark unterschätzt und die letzte Komponente stark überschätzt werden würde. Um dies zu umgehen wird die Konstante mit dem Wert 0.5 gewählt.

Wird im ND-Modell zu einem Zentrum eine Konstante addiert, dann ändert sich auch die Formel der Varianzschätzung. Die vollständige Varianzformel ist binomialverteilten Beobachtungen bei $x_i^T = 0$ und $x_i^C > 0$ (entsprechend bei $x_i^T > 0$ und $x_i^C = 0$ oder bei $x_i^T = 0$ und $x_i^C = 0$)

$$\text{Var}(y_i) = \frac{0.5n_i^T}{(0.5 + n_i^T)} + \frac{1}{x_i^C} - \frac{1}{n_i^C} \quad (3.31)$$

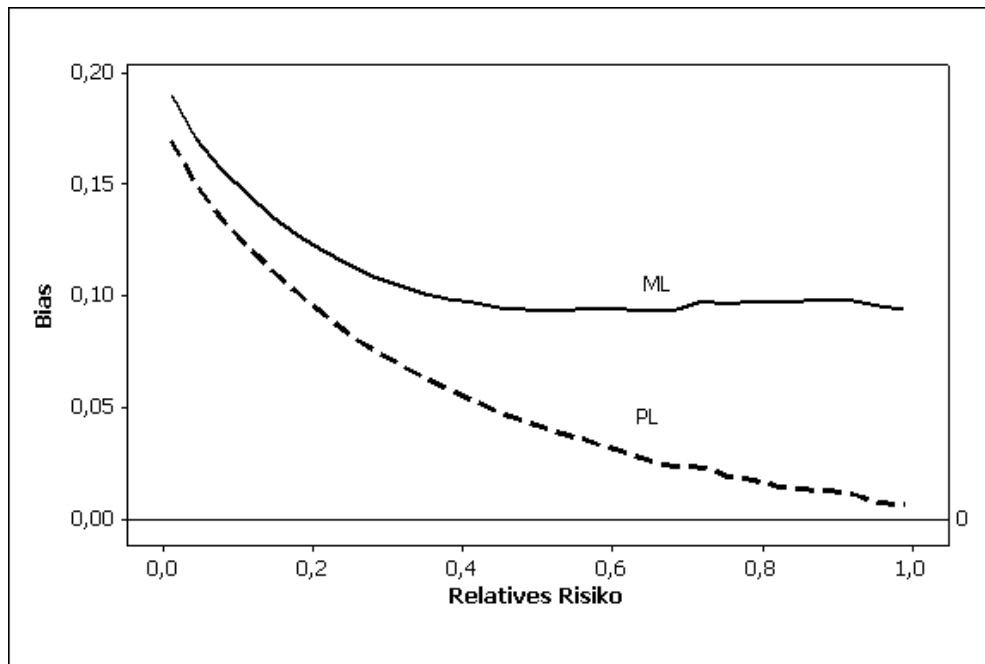


Abbildung 3.2: CALGB-Studie (Bias vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;0.99]$ berechnet, die generierten x^T , x^C - Werte sind binomialverteilt, additive Konstante ist 0.5)

und bei $x_i^T = n_i^T$ und $x_i^C = n_i^C$

$$\text{Var}(y_i) = \frac{0.5n_i^T(0.5 + n_i^T)}{(0.5 + n_i^T(1 + n_i^T))^2} + \frac{0.5n_i^C(0.5 + n_i^C)}{(0.5 + n_i^C(1 + n_i^C))^2} \quad (3.32)$$

Im Falle von Poissonverteilten Beobachtungen gilt bei $x_i^T = 0$ und $x_i^C > 0$ (entsprechend bei $x_i^T > 0$ und $x_i^C = 0$ oder bei $x_i^T = 0$ und $x_i^C = 0$) die veränderte Varianzformel

$$\text{Var}(z_i^*) = \frac{0.5n_i^T(1 + n_i^T)}{(0.5 + n_i^T)^2} + \frac{1}{x_i^C}. \quad (3.33)$$

3.7.2 Die Maximierung der Gradientenfunktion

Ein wichtiger Punkt im gEMGFU ist die Berechnung des globalen Maximums der Gradientenfunktion. Da die Gradientenfunktion mehrere lokale Maxima besitzt, ist es nicht leicht, den globalen Extremwert zu finden. Durch die Daten der Studie ist das Intervall gegeben, in der die Lösung enthalten sein muss. Im ND- und im PL-Ansatz ist das Intervall $[a, b]$ definiert durch die untere Grenze $a = \min_i(\frac{x_i^T n_i^C}{n_i^T x_i^C})$ und durch die obere Grenze $b = \max_i(\frac{x_i^T n_i^C}{n_i^T x_i^C})$ (bei $x_i^C > 0$ für $i = 1, \dots, k$). Die Idee besteht nun darin, das Intervall zu

splitten. Danach wird in jedem Teilintervall $[a_i, b_i]$ geprüft, ob das Vorzeichen der Ableitung der Gradientenfunktion $\frac{\partial}{\partial \theta} d(\theta, P)$ wechselt. Die Ableitung der Gradientenfunktion ist:

- Für den ND-Ansatz

$$\frac{\partial}{\partial \theta} d(\theta, P) = \frac{1}{k} \sum_{i=1}^k \frac{e^{-\frac{(y_i - \log \theta)^2}{2\sigma_i^2}} \frac{(y_i - \log \theta)}{\sigma_i^2}}{\sum_{j=1}^m e^{-\frac{(y_i - \log \theta_j)^2}{2\sigma_i^2}} q_j}$$

- Für den PL-Ansatz

$$\frac{\partial}{\partial \theta} d(\theta, P) = \frac{1}{k} \sum_{i=1}^k \frac{(-n_i^T \theta x_i^C + n_i^C x_i^T) \theta^{-1+x_i^T} (n_i^T \theta + n_i^C)^{-1-x_i}}{\sum_{j=1}^m \frac{\theta_j^{x_i^T}}{(n_i^T \theta_j + n_i^C)^{x_i}} q_j}$$

Ein lokales Maximum ist dann in dem Intervallstück enthalten, wenn gilt, dass

$$\frac{\partial}{\partial \theta} d(a_i, P) \geq 0 \text{ und } \frac{\partial}{\partial \theta} d(b_i, P) < 0.$$

Nachdem alle interessanten Intervallstücke gefunden worden sind, wird in diesen das jeweilige lokale Maximum mit Hilfe der Intervallhalbierungsmethode berechnet. Damit sind alle lokalen Maxima bestimmt. Das globale Maximum θ_{\max} ist dann der größte Wert von $d(\theta_{LMi}, P)$, mit θ_{LM} als Vektor aller lokalen Maxima. Beim ML-Modell kann dieses Verfahren nicht angewendet werden, da die Gradientenfunktion in dem Fall von 2 Variablen abhängt. Das Intervall, in dem die Lösung enthalten sein muss, ist gegeben durch $[\alpha_{\min}, \alpha_{\max}] \times [\beta_{\min}, \beta_{\max}] \in \mathbb{R}^2$, mit

$$\begin{aligned} \alpha_{\min} &= \min_i \left(\frac{x_i^C}{n_i^C} \right), \quad \alpha_{\max} = \max_i \left(\frac{x_i^C}{n_i^C} \right) \\ \beta_{\min} &= \min_i \left(\frac{x_i^T n_i^C}{x_i^C n_i^T} \right), \quad \beta_{\max} = \max_i \left(\frac{x_i^T n_i^C}{x_i^C n_i^T} \right) \\ \text{mit } i &= 1, \dots, k. \end{aligned}$$

Über ein Rastergitter auf dem Intervall aus 100×100 Punkten wird der größte Wert $\theta_{\max}^{ML} = \begin{pmatrix} \alpha_{\max} \\ \beta_{\max} \end{pmatrix}$ der Gradientenfunktion gesucht. θ_{\max}^{ML} stellt damit eine grobe Näherung des globalen Maximums dar. Um den globalen Extremwert möglichst gut zu schätzen, wird θ_{\max}^{ML} mit Hilfe des Newtonverfahrens verbessert. Das Newtonverfahren ist definiert

als

$$\begin{aligned}
 x^{n+1} &= x^n - f'(x^n)^{-1} f(x^n) \\
 \text{mit } f(x) &= \begin{pmatrix} u(x_1, x_2) \\ v(x_1, x_2) \end{pmatrix} \text{ und} \\
 f'(x) &= \begin{pmatrix} u_{x_1}(x_1, x_2) & u_{x_2}(x_1, x_2) \\ v_{x_1}(x_1, x_2) & v_{x_2}(x_1, x_2) \end{pmatrix}
 \end{aligned}$$

Mit dem Newtonverfahren wird die Nullstelle von der Ableitung der Gradientenfunktion $\frac{\theta}{\partial\theta^{ML}}d(\theta^{ML}, P) = \begin{pmatrix} \frac{\theta}{\partial\alpha}d(\alpha, \beta, P) \\ \frac{\theta}{\partial\beta}d(\alpha, \beta, P) \end{pmatrix}$ berechnet. Damit ist $f(\theta^{ML}) = \frac{\theta}{\partial\theta^{ML}}d(\theta^{ML}, P)$ und $f'(\theta^{ML}) = \begin{pmatrix} \frac{\theta}{\partial\alpha\alpha}d(\alpha, \beta, P) & \frac{\theta}{\partial\alpha\beta}d(\alpha, \beta, P) \\ \frac{\theta}{\partial\beta\alpha}d(\alpha, \beta, P) & \frac{\theta}{\partial\beta\beta}d(\alpha, \beta, P) \end{pmatrix}$. Das Inverse einer 2×2 Matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ kann mittels folgender Formel bestimmt werden

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Daraus folgt, dass

$$\begin{aligned}
 f'(\theta^{ML})^{-1} &= \frac{1}{\frac{\partial}{\partial\alpha\alpha}d(\alpha, \beta, P)\frac{\partial}{\partial\beta\beta}d(\alpha, \beta, P) - \frac{\partial}{\partial\alpha\beta}d(\alpha, \beta, P)^2} \\
 &\quad \begin{pmatrix} \frac{\partial}{\partial\beta\beta}d(\alpha, \beta, P) & -\frac{\partial}{\partial\alpha\beta}d(\alpha, \beta, P) \\ -\frac{\partial}{\partial\beta\alpha}d(\alpha, \beta, P) & \frac{\partial}{\partial\alpha\alpha}d(\alpha, \beta, P) \end{pmatrix} \quad (3.34)
 \end{aligned}$$

Zur vollständigen Bestimmung von $f(\theta^{ML})$ und $f'(\theta^{ML})$ sind alle einzelnen Ableitungen im Anhang angegeben.

Kapitel 4

Auswertung der Studien

4.1 Ohne Heterogenität (Homogenitätsfall)

In diesem Abschnitt wird vorerst nur der Homogenitätsfall betrachtet. Somit gibt es nur einen zu schätzenden Parameter pro Modell. In der Tabelle (4.1) sind die Ergebnisse von der Schätzung des Relativen Risikos aller hier betrachteten multizentrischen Studien abgebildet. Die Unterschiede zwischen den drei Modellen sind gering. In 9 von 13 Studien liegt die Schätzung des Relativen Risikos mittels ND-Ansatz näher beim Nulleffekt als bei den anderen beiden Modellen. In den multizentrischen Studien 1.1, 1.2, 1.6, 1.7, 1.11, 1.12 und 1.13 tritt sogar eine Reihenfolge in Richtung Nulleffekt auf. Am weitesten vom Nulleffekt entfernt ist der Schätzer des ML-Modells. Der Schätzer des PL-Modells liegt dagegen immer zwischen den Werten der anderen beiden Modelle.

4.2 Mit unbeobachteter Heterogenität

Wenn die Effektheterogenität berücksichtigt wird, sind weit mehr Parameter zu schätzen. Dazu wird jede Studie einzeln analysiert. Das Relative Risiko wird mit dem gEMGFU (3.6.2) geschätzt. Die Ergebnisse werden in Tabellenform dargestellt. In jeder Tabelle sind die Zwischenergebnisse jeweils nach einer Mischungskomponente, mit zwei Mischungskomponenten und die Schätzung nach dem BIC- und nach dem NPMLE-Kriterium abgebildet. Weiterhin sind zu jedem Modell sowohl das Relative Risiko, das Gewicht der Komponente und die Anzahl der zur Komponente gehörenden Zentren als auch der Log-Likelihood, das Maximum der Gradientenfunktion und das BIC-Kriterium angegeben. Die drei Modelle sind hier vergleichbar in der Anzahl der geschätzten Mischungskomponenten, in der Zentrenzugehörigkeit und in der Effektschätzung.

Tabelle 4.1: Schätzung des Relativen Risikos im Homogenitätsfall (NDM=Normalverteilungs-Modell, PLM=Profil-Likelihood-Modell, MLM=Multi-Level-Modell)

Daten	θ_{NDM}	θ_{PLM}	θ_{MLM}
Tabelle 1	0.792007	0.791191	0.791032
Tabelle 2	0.508613	0.473867	0.467574
Tabelle 3	0.737044	0.712631	0.714230
Tabelle 4	0.834828	0.902980	0.973232
Tabelle 5	7.702793	8.411879	7.673077
Tabelle 6	0.751838	0.742881	0.742111
Tabelle 7	0.427455	0.421663	0.381264
Tabelle 8	0.905399	0.901852	0.894720
Tabelle 9	0.968171	0.971608	1.077710
Tabelle 10	0.734352	0.727809	0.742689
Tabelle 11	0.836876	0.835719	0.826971
Tabelle 12	1.521802	1.565594	1.679105
Tabelle 13	0.423938	0.403054	0.399119

4.2.1 Beta-Blocker Studie

Aitkin [2] verwendet diese Studie in seiner Publikation und nutzt zur Schätzung des Behandlungseffekts ein ML-Modell. Dabei wird zunächst ein Modell mit einem feststehenden Behandlungseffekt und mit einem zufälligen Intercept-Parameter berechnet. Danach ermittelt er das vollständige Modell mit zufälligen Behandlungs- und Intercept-Parametern. Die Differenz der Standardabweichung zwischen den Modellen ist gering. Aitkin schlussfolgert, dass die Schwankung im Behandlungseffekt nicht besonders groß sein kann. Mit anderen Worten, es liegt Homogenität im Relativen Risiko vor. Die Ergebnisse (siehe Tabelle 4.2) des PL- und des ND-Ansatzes sprechen für Aitkins Annahme. Im Gegensatz zum BIC-Kriterium, das nur eine Komponente in beiden Modellen schätzt, findet das NPMLE-Kriterium zwei Mischungskomponenten. In beiden Modellen gehört nur das 14. Zentrum zur zweiten Komponente. Die erste Komponente schätzt einen positiven Behandlungseffekt. Das heißt, die Sterblichkeit nach einem Herzinfarkt kann durch die Behandlung mit Beta-Blockern gesenkt werden. Die zweite Mischungskomponente schätzt dagegen eine negative Behandlungswirkung, also eine Erhöhung der Sterblichkeit nach einem Herzinfarkt durch Beta-Blocker-Applikation. Mit dem BIC-Kriterium werden im ML-Modell (Tabelle 4.3) fünf Mischungskomponenten geschätzt, wobei 4 Komponenten eine positive und die 5. Komponente eine negative Therapiewirkung schätzen. Wie im PL- und im ND-Ansatz

Tabelle 4.2: Ergebnisse der Beta-Blocker-Studie anhand des Profil-Likelihood- und Normalverteilungsansatzes (H entspricht der Anzahl der Zentren, die zu der jeweiligen Komponente zugeordnet sind).

Profil-Likelihoodansatz			Normalverteilungsansatz		
Komp.	1.		Komp.	1.	
θ	.791191		θ	.792007	
q	1.000000		q	1.000000	
H	22		H	22	
Log-L.=	-121.752565		Log-L.=	-3.378254	
max GF =	1.557780		max GF =	1.716050	
BIC=	-246.596173		BIC=	-9.847550	
Komp.	1.	2.	Komp.	1.	2.
θ	.768500	1.128979	θ	.768388	1.121695
q	.905409	.094591	q	.898392	.101608
H	21	1	H	21	1
Log-L.=	-121.366906		Log-L.=	-2.897769	
max GF =	1.000000		max GF =	1.000000	
BIC=	-252.006939		BIC=	-15.068665	

zählt auch im ML-Modell nur das 14. Zentrum zu einer Komponente mit negativer Behandlungswirkung. Das NPML-Kriterium schätzt 8 Mischungskomponenten, wovon zwei einen negativen Behandlungseffekt schätzen. Darunter fallen die 7. Komponente ohne ein zugeordnetes Zentrum und die 5. Komponente mit dem 14. und dem 18. Zentrum. Alle anderen Zentren werden den Komponenten mit Behandlungserfolg zugewiesen. Bei den Komponenten 1, 2, 3, 4, 6 und 8 liegt das geschätzte RR im Intervall von [0.617;0.876]. Dieses recht schmale Intervall deutet auf Effekthomogenität in den speziellen Komponenten hin. Hier hat nur die Baselineheterogenität zur Bildung der 6 Komponenten geführt.

4.2.2 Studie über Infektion der Atemwege

Diese Studie wird als erstes Beispiel bei Turner [33] verwendet. Turner berechnet den Behandlungseffekt mit Hilfe verschiedener Modelle. Der erste Ansatz geht von der Existenz eines wahren Behandlungseffektes aus, der in allen Zentren gleich ist. Eine andere Herangehensweise ist durch die traditionellen Zufallseffektmodelle gegeben, welche davon ausgehen, dass der wahre Effekt zufällig zwischen den Zentren schwankt. Der zufällige Effektschätzer unterscheidet sich deutlich vom festen Effektschätzer, was nach Turner auf Heterogenität im Behandlungseffekt schließen lässt. In der Tabelle 4.4 sind die Er-

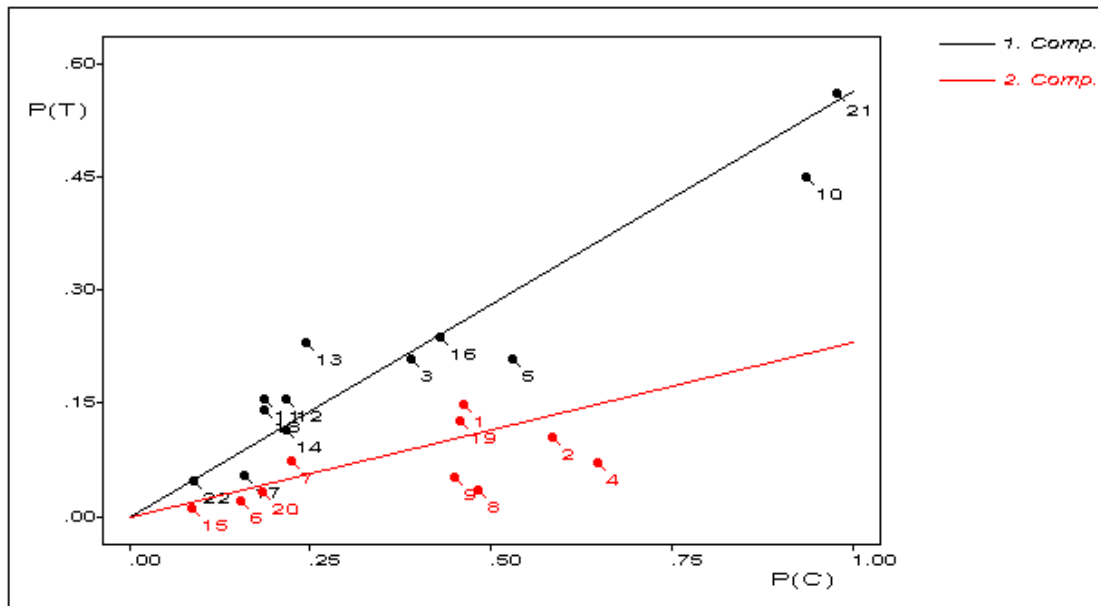


Abbildung 4.1: Zentrenzugehörigkeit bei der Atemwegsstudie mit zwei Mischungskomponenten (Profil-Likelihood-Ansatz). Die Zahlen neben den Punkten geben das Zentrum an.

gebnisse vom PL- und vom ND-Ansatz abgebildet. Im PL-Ansatz schätzen beide Abbruchkriterien dasselbe Modell mit zwei Mischungskomponenten. Im ND-Ansatz schätzt das BIC-Kriterium nur eine Komponente und das NPMLE-Kriterium zwei. Die Ergebnisse im Zweikomponentenmodell beider Ansätze unterscheiden sich erst nach der ersten Kommastelle. Die Abbildung 4.1 für den PL- und die Abbildung 4.2 für den ND-Ansatz bestätigen die Gleichheit der Zentrenzuordnung in beiden Modellen. Die farbigen Linien geben das geschätzte Relative Risiko der jeweiligen Komponente an. Im ML-Modell werden wie im ersten Beispiel wesentlich mehr Mischungskomponenten betrachtet. Das BIC-Kriterium schätzt 4 Komponenten und das NPMLE-Kriterium 9 Mischungskomponenten. Besonders groß ist der Unterschied in der Zentrenreinteilung im Vergleich mit den zwei anderen Modellen. In der Abbildung 4.3 sind die Zentren nach ihrer Einteilung in das 4 Komponentenmodell abgebildet. Im Gegensatz zu den anderen beiden Bildern (4.1 und 4.2) ist hier eine klare Clusterbildung nach dem Baseline zu erkennen. Laut Grafik scheint der Behandlungseffekt einen geringeren Einfluss auf die Zentrenzuweisung zu haben als der Baseline. Aus der Sichtweise des Effektes würde beispielsweise das 8. Zentrum eher zur 4. als zur 3. Mischungskomponente gehören. Bei der Zentrenzuweisung vom PL- und vom ND-Modell ist deutlich zu erkennen, wie die Cluster nach dem Relativen Risiko gebildet werden. Einzige Ausnahme besteht bei dem 17 Zentrum, welches grafisch näher

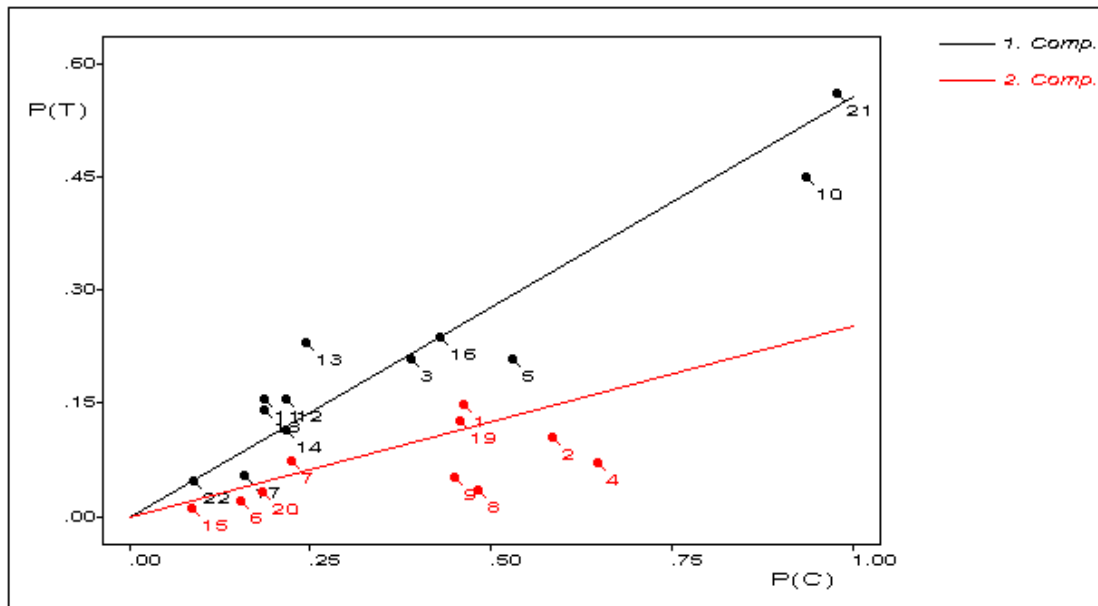


Abbildung 4.2: Zentrenzugehörigkeit bei der Atemwegsstudie mit zwei Mischungskomponenten (Normalverteilungsansatz). Die Zahlen neben den Punkten geben das Zentrum an.

zur 2. als zur 1. Mischungskomponenten aufgetragen ist.

4.2.3 Studie über Präeklampsie

Auch diese Studie wird bei Turner [33] analysiert. Turner zeigt, dass die Differenz zwischen dem zufälligen Effektansatz und dem fixierten Effektmodell geringer ist, als in der Atemwegsinfektionsstudie. Im PL- (Tabelle 4.6) und im ND-Ansatz (Tabelle 4.7) werden nach dem BIC-Kriterium 2 Mischungskomponenten geschätzt. Mit dem NPMLE-Kriterium werden 3 Mischungskomponenten gefunden. Wobei im PL-Modell der 3. Mischungskomponente kein Zentrum zugewiesen wird. Dagegen haben im ND-Ansatz das 8. und das 9. Zentrum den größten Anteil an der dritten Komponente. Beide Ansätze geben an, dass im Behandlungseffekt Heterogenität vorliegt. Im ML-Modell (Tabelle 4.8) werden mit dem BIC-Kriterium 6 Mischungskomponenten geschätzt und das NPMLE-Kriterium weist auf 7 Komponenten hin. Fast alle Zentren gehören zu Komponenten mit positivem Behandlungseffekt. Eine Ausnahme bildet das 8. und 9. Zentrum. Wie im ND-Modell klassifiziert auch das ML-Modell diese Zentren zu Komponenten mit keinem oder negativem Behandlungseffekt.

4.2.4 CALGB-Studie

Lipsitz [25] und Böhning [6] betrachten bei dieser Studie als Effektmaß die Risikodifferenz. Lipsitz testet, ob im Behandlungseffekt Homogenität vorliegt, wobei zwei Testverfahren den Test ablehnen und drei den Test nicht ablehnen. Die Besonderheit an dieser 21 Zentren umfassenden Studie liegt in der sehr geringen Anzahl von Personen unter Risiko. Das BIC-Kriterium (Tabelle 4.9 und 4.10) schätzt in allen drei Modellen, dass im Behandlungseffekt (hier dem Relativen Risiko) Homogenität vorliegt. Im PL-Modell stimmt das NPMLE-Kriterium mit dem BIC-Kriterium überein. Im ND-Ansatz werden mit dem NPMLE-Kriterium 2 Mischungskomponenten geschätzt, wobei zur zweiten Komponente nur zwei Zentren gehören. Die zweite Komponente schätzt eine äußerst positive Behandlungswirkung, wohingegen die erste Mischungskomponente einen leicht positiven Behandlungsverlauf aufzeigt. Im ML-Modell (Tabelle 4.10) schätzt das NPMLE-Kriterium 4 Mischungskomponenten.

4.2.5 Multi-Farm-Studie

Die Multi-Farm-Studie umfasst nur 7 Zentren. Nach dem BIC- und nach dem NPMLE-Kriterium liegt hier in allen drei Modellen (Tabelle 4.11) Effekthomogenität vor. Dies ist die einzige Studie, in der die gleiche Anzahl an Mischungskomponenten in allen drei Modellen geschätzt wird. Diese drei Modelle schätzen ein Relatives Risiko, dass deutlich größer ist als eins. Somit ist die Wahrscheinlichkeit eines Aborts bei einem Rind viel höher, wenn es in der Vergangenheit an einer *Noespora* Infektion erkrankt war.

4.2.6 Herzinfarkt-Studie

Die ersten 21 Zentren bevorzugen eine Behandlung mit Propranolol (\emptyset Relative Risiko = 0.5) und die letzten 10 Zentren schätzen eine negative Behandlungswirkung (\emptyset Relative Risiko = 1.33). Diese Heterogenität im Behandlungseffekt wird von den hier beschriebenen Modellen nicht erfasst. Nach dem BIC-Kriterium liegt in allen drei Modellen (Tabelle 4.12 und 4.13) im Behandlungseffekt Homogenität vor. Nur im ML-Modell schätzt das NPMLE-Kriterium 3 Mischungskomponenten, wobei der letzten Komponente kein Zentrum zugeordnet wird. Die zweite Mischungskomponente enthält das 30. und das 16. Zentrum. Dennoch ist in jeder Komponente das Relative Risiko kleiner eins.

Tabelle 4.3: Ergebnisse der Beta-Blocker-Studie anhand des Multi-Level-Modells mit Binomiallikelihood (H entspricht der Anzahl der Zentren, die zu der jeweiligen Komponente zugeordnet sind).

Multi-Level-Modell (Bin)						
Komp.	1.					
θ	.791032					
q	1.000000					
H	22					
Log-L.= -261.595628						
max $GF = 4.66 \times 10^7$						
BIC= -526.282299						
Komp.	1.	2.				
θ	.815038	.756668				
q	.698979	.301021				
H	15	7				
Log-L.= -180.860785						
max $GF = 4068.093470$						
BIC= -370.994697						
Komp.	1.	2.	3.	4.	5.	
θ	.788341	.876803	.643952	.654271	1.290148	
q	.493659	.181645	.158033	.086877	.079787	
H	12	4	4	1	1	
Log-L.= -149.522377						
max $GF = 1.530004$						
BIC= -326.864136						
Komp.	1.	2.	3.	4.	5.	6.
θ	.754972	.876851	.617511	.647257	1.288478	.833775
q	.234572	.181657	.133744	.065878	.079130	.160689
H	8	4	3	1	2	3
Komp.	7.	8.				
θ	1.422697	.807168				
q	.024693	.119637				
H	0	1				
Log-L.= -148.378884						
max $GF = 1.000000$						
BIC= -343.123406						

Tabelle 4.4: Ergebnisse der Studie über eine Infektion der Atemwege anhand des Profil-Likelihood- und Normalverteilungsansatzes (H entspricht der Anzahl der Zentren, die zu der jeweiligen Komponente zugeordnet sind).

Profil-Likelihoodansatz		Normalverteilungsansatz			
Komp.	1.	Komp.	1.		
θ	.473867	θ	.508613		
q	1.000000	q	1.000000		
H	22	H	22		
Log-L.=	-116.606398	Log-L.=	-20.701039		
max GF =	2.156801	max GF =	1.652040		
BIC=	-236.303839	BIC=	-44.493120		
Komp.	1.	2.	Komp.	1.	2.
θ	.562660	.231091	θ	.556435	.252627
q	.615078	.384922	q	.662826	.337174
H	12	10	H	12	10
Log-L.=	-113.359637		Log-L.=	-18.704168	
max GF =	1.000000		max GF =	1.000001	
BIC=	-235.992402		BIC=	-46.681464	

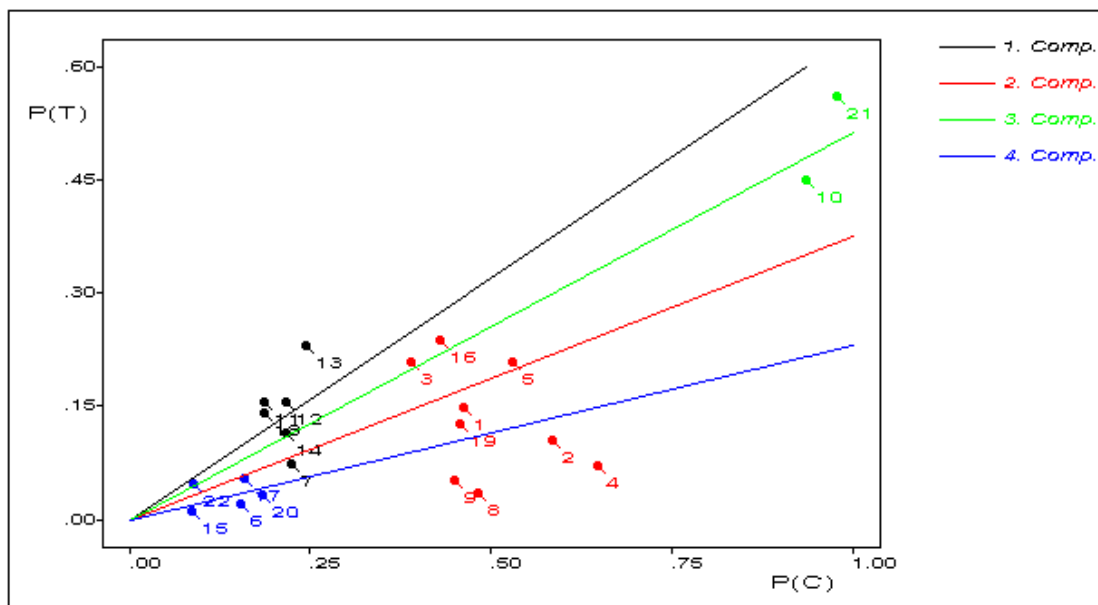


Abbildung 4.3: Zentrenzugehörigkeit bei der Atemwegsstudie mit vier Mischungskomponenten (Multi-Level-Modell). Die Zahlen neben den Punkten geben das Zentrum an.

Tabelle 4.5: Ergebnisse der Studie über eine Infektion der Atemwege anhand des Multi-Level-Modells mit Binomiallikelihood (H entspricht der Anzahl der Zentren, die zu der jeweiligen Komponente zugeordnet sind).

Multi-Level-Modell (Bin)						
Komp.	1.					
θ	.467574					
q	1.000000					
H	22					
Log-L.	= -308.734660					
max GF	= 6.93×10^{23}					
BIC	= -620.560363					
Komp.	1.	2.				
θ	.553392	.413932				
q	.506424	.493576				
H	11	11				
Log-L.	= -199.438977					
max GF	= 2.63×10^{10}					
BIC	= -408.151082					
Komp.	1.	2.	3.	4.		
θ	.640245	.373983	.511712	.230626		
q	.276820	.409003	.090909	.223268		
H	6	9	2	5		
Log-L.	= -137.195431					
max GF	= 3.100094					
BIC	= -296.028159					
Komp.	1.	2.	3.	4.	5.	6.
θ	.706591	.497027	.511712	.228424	.222553	.345662
q	.225093	.156461	.090909	.155802	.254818	.058032
H	5	3	2	5	6	1
Komp.	7.	8.	9.			
θ	.228424	.228424	.319099			
q	.045889	.009199	.003796			
H	0	0	0			
Log-L.	= -133.679668					
max GF	= 1.000000					
BIC	= -319.907057					

Tabelle 4.6: Ergebnisse der Studie über Präeklampsie anhand des Profil-Likelihood-Ansatzes (H entspricht der Anzahl der Zentren, die zu der jeweiligen Komponente zugeordnet sind).

Profil-Likelihoodansatz			
Komp.	1.		
θ	.712631		
q	1.000000		
H	9		
Log-L.=	-56.606361		
max GF =	22.274784		
BIC=	-115.409947		
Komp.	1.	2.	
θ	.826852	.360125	
q	.643134	.356866	
H	5	4	
Log-L.=	-52.477528		
max GF =	1.058324		
BIC=	-111.546729		
Komp.	1.	2.	3.
θ	.799215	.360674	1.056276
q	.509939	.358527	.131534
H	5	4	0
Log-L.=	-52.436479		
max GF =	1.000002		
BIC=	-115.859081		

Tabelle 4.7: Ergebnisse der Studie über Präeklampsie anhand des Normalverteilungsansatzes (H entspricht der Anzahl der Zentren, die zu der jeweiligen Komponente zugeordnet sind).

Normalverteilungsansatz			
Komp.	1.		
θ	.737044		
q	1.000000		
H	9		
Log-L.=	-11.691916		
max GF =	17.688723		
BIC=	-25.581056		
Komp.	1.	2.	
θ	.855005	.389914	
q	.583306	.416694	
H	5	4	
Log-L.=	-7.079302		
max GF =	1.210302		
BIC=	-20.750278		
Komp.	1.	2.	3.
θ	.777261	.388369	1.065794
q	.351717	.407581	.240702
H	3	4	2
Log-L.=	-6.710055		
max GF =	1.000001		
BIC=	-24.406234		

Tabelle 4.8: Ergebnisse der Studie über Präeklampsie anhand des Multi-Level-Modells mit Binomiallikelihood (H entspricht der Anzahl der Zentren, die zu der jeweiligen Komponente zugeordnet sind).

Multi-Level-Modell (Bin)						
Komp.	1.					
θ	.714230					
q	1.000000					
H	9					
Log-L.	= -295.244244					
max GF	= -295.244244					
BIC	= -592.685713					
Komp.	1.	2.				
θ	.667709	.794158				
q	.666667	.333333				
H	6	3				
Log-L.	= -148.817666					
max GF	= 1.73×10^{16}					
BIC	= -304.227006					
Komp.	1.	2.	3.	4.	5.	6.
θ	.787222	1.081878	.259166	.441012	.438660	.974334
q	.218868	.111664	.112076	.221669	.114551	.221172
H	2	1	1	2	1	2
Log-L.	= -59.766749					
max GF	= 1.463279					
BIC	= -143.702968					
Komp.	1.	2.	3.	4.	5.	6.
θ	.787221	1.081878	.259155	.441012	.438604	.846724
q	.218864	.111664	.111975	.221669	.114528	.147223
H	2	1	1	2	1	1
Komp.	7.					
θ	2.853416					
q	.074078					
H	1					
Log-L.	= -59.590074					
max GF	= 1.000001					
BIC	= -147.744067					

Tabelle 4.9: Ergebnisse der CALGB-Studie anhand des Profil-Likelihood- und Normalverteilungsmodells.

Profil-Likelihoodansatz		Normalverteilungsansatz	
Komp.	1.	Komp.	1.
θ	.902980	θ	.834828
q	1.000000	q	1.000000
H	21	H	21
Log-L.=	-56.866429	Log-L.=	-25.828776
max GF =	.999929	max GF =	4.515199
BIC=	-116.777380	BIC=	-54.702074
		Komp.	1. 2.
		θ	.936068 .162427
		q	.890885 .109115
		H	19 2
		Log-L.=	-23.739267
		max GF =	1.000002
		BIC=	-56.612101

Tabelle 4.10: Ergebnisse der CALGB-Studie anhand des Multi-Level-Modells mit Binomiallikelihood.

Multi-Level-Modell (Bin)				
Komp.	1.			
θ	.973232			
q	1.000000			
H	21			
Log-L.=	-60.525856			
max GF =	1.177104			
BIC=	-124.096234			
Komp.	1.	2.		
θ	.950566	.947724		
q	.606337	.393663		
H	14	7		
Log-L.=	-59.095396			
max GF =	1.509506			
BIC=	-127.324360			
Komp.	1.	2.	3.	4.
θ	.893824	.945863	2.610250	.418704
q	.496980	.343435	.084903	.074682
H	13	7	1	0
Log-L.=	-58.176845			
max GF =	1.000000			
BIC=	-137.665347			

Tabelle 4.11: Ergebnisse der Multi-Farm-Studie anhand des Profil-Likelihood- und Normalverteilungsansatzes und des Multi-Level-Modells mit Binomiallikelihood.

Profil-Likelihoodansatz		Normalverteilungsansatz		Multi-Level-Modell (Bin)	
Komp.	1.	Komp.	1.	Komp.	1.
θ	8.411879	θ	7.702793	θ	7.673077
q	1.000000	q	1.000000	q	1.000000
H	7	H	7	H	7
Log-L.=	-18.451681	Log-L.=	-6.676803	Log-L.=	-18.679340
max GF =	.999855	max GF =	1.000000	max GF =	1.000000
BIC=	-38.849272	BIC=	-15.299517	BIC=	-39.304589

Tabelle 4.12: Ergebnisse der Herzinfarkt-Studie anhand des Profil-Likelihood- und Normalverteilungsansatzes.

Profil-Likelihoodansatz		Normalverteilungsansatz	
Komp.	1.	Komp.	1.
θ	.751838	θ	.742881
q	1.000000	q	1.000000
H	31	H	31
Log-L.=	-29.916395	Log-L.=	-119.580649
max GF =	1.000000	max GF =	1.000000
BIC=	-63.266778	BIC=	-242.595285

Tabelle 4.13: Ergebnisse der Herzinfarkt-Studie anhand des Multi-Level-Modells mit Binomiallikelihood.

Multi-Level-Modell (Bin)			
Komp.	1.		
θ	.742111		
q	1.000000		
H	31		
Log-L.=	-136.289409		
max GF =	5.268740		
BIC=	-276.012805		
Komp.	1.	2.	
θ	.731184	.838468	
q	.934707	.065293	
H	29	2	
Log-L.=	-134.013902		
max GF =	1.001429		
BIC=	-278.329766		
Komp.	1.	2.	3.
θ	.712622	.838280	.891290
q	.832342	.065387	.102271
H	29	2	0
Log-L.=	-134.009716		
max GF =	.999397		
BIC=	-285.189367		

4.2.7 Studie über Frühgeburten

Arends [3] veranschaulicht den Zusammenhang des Baseline-Risikos mit dem Behandlungseffekt. Da hier nur die Schätzung des Behandlungseffekts von Interesse ist, wird dieser publizierte Zusammenhang momentan außen vor gelassen. Im PL-Modell (Tabelle 4.14) schätzt das BIC-Kriterium eine Mischungskomponente und das NPMLE-Kriterium zwei. Im ND-Ansatz (Tabelle 4.14) werden mit beiden Abbruchkriterien zwei Mischungskomponenten geschätzt. Zur zweiten Komponente werden die Zentren eingeordnet, bei denen gilt $x_i^T = 0$ und $x_i^C > 0$. Das sind die Zentren 8, 12 und 13. Alle anderen Zentren gehören zur ersten Komponente. Der PL-Ansatz klassifiziert alle Zentren zur ersten Komponente. Im ML-Modell (Tabelle 4.15) weist das BIC-Kriterium auf 3 Komponenten und das NPMLE-Kriterium auf 5 Komponenten hin. Entscheidend ist, dass alle Modelle eine positive Behandlungswirkung schätzen. Bei jeder Mischungskomponente ist das Relative Risiko Modell unabhängig deutlich kleiner als eins.

Tabelle 4.14: Ergebnisse der Studie über Frühgeburten anhand des Profil-Likelihood- und Normalverteilungsansatzes.

Profil-Likelihoodansatz		Normalverteilungsansatz	
Komp.	1.	Komp.	1.
θ	.421663	θ	.427455
q	1.000000	q	1.000000
H	14	H	14
Log-L.=	-53.495506	Log-L.=	-19.074786
max GF =	1.269887	max GF =	10.739011
BIC=	-109.630070	BIC=	-40.788629
Komp.	1. 2.	Komp.	1. 2.
θ	.447038 .071067	θ	.514903 .064192
q	.921036 .078964	q	.796429 .203571
H	14 0	H	11 3
Log-L.=	-52.631182	Log-L.=	-14.635700
max GF =	.995666	max GF =	1.000000
BIC=	-113.179537	BIC=	-37.188571

4.2.8 Studie über Bluthochdruck

Auch bei diese Studie analysiert Arends [3] den Zusammenhang des Baseline-Risikos mit dem Behandlungseffekt. Da es sich um Personenzeiten handelt, wird der Poissonlikelihood zur Parameterschätzung verwendet. Der ND- und der PL-Ansatz (Tabelle 4.16) liefern

sehr ähnliche Ergebnisse. Das BIC-Kriterium schätzt in beiden Modellen Homogenität im Behandlungseffekt und das NPMLE-Kriterium weist auf zwei Mischungskomponenten mit einer positiven und einer negativen Behandlungswirkung. Ebenso ist die Zentrenerteilung gleich. Das 2., 8. und 9. Zentrum gehören zur zweiten Komponente. Im ML-Modell (Tabelle 4.17) werden deutlich mehr Mischungskomponenten gefunden. Das BIC-Kriterium schätzt 4 Mischungskomponenten und das NPMLE-Kriterium findet 9 Komponenten.

4.2.9 Studie über koronare Herzkrankheit

Da bei dieser 33 Zentren umfassenden Studie Personenjahre betrachtet werden, wird auch hier der Poissonlikelihood verwendet. Die Ergebnisse vom PL- (Tabelle 4.18) und vom ND-Ansatz (Tabelle 4.19) stimmen fast überein. Das BIC-Kriterium schätzt 2 Mischungskomponenten. Bei der ersten Komponente strebt das Relative Risiko gegen eins. Dieser Komponente werden 31 Zentren zugeordnet. Die zweite Mischungskomponente schätzt ein Relatives Risiko von 0.44. Das Gewicht dieser Komponente ist aber relativ gering. Das NPMLE-Kriterium schätzt 4 Mischungskomponenten. Auch die Zentrenerteilung (Abbildung 4.4 und 4.5) ist im PL- und ND-Ansatz identisch. Eigentlich würden die Zentren 4 und 5 grafisch eher zur dritten Komponente gehören. Ebenso liegen die Zentren 22, 21, 12 und 13 näher an der 4-Komponente beim PL- oder an der 3. Komponente beim ND-Modell. Dennoch haben diese 6 Zentren nach dem Modell den größten Anteil an der ersten Mischungskomponente, die keinen Behandlungseffekt schätzt. Interessant ist, dass in der 4 Komponentenlösung das ND-Modell das Relative Risiko in jeder Mischungskomponente geringfügig näher am Nulleffekt schätzt als das PL-Modell. Im ML-Modell (Tabelle 4.20) werden mit dem BIC-Kriterium 7 Mischungskomponenten geschätzt. Die ersten drei Komponenten schätzen keinen Effekt mit 21 Zentren. Die Komponenten 4, 5 und 7 schätzen eine protektive Behandlungswirkung mit 9 Zentren, und die 6. Komponente schätzt einen negativen Behandlungsverlauf. Mit dem NPMLE-Kriterium werden 15 Mischungskomponenten gefunden, wobei die Komponenten 4, 6, 7, 8, 9, 10, 11, 12, 13 und 15 je nur ein Zentrum zugewiesen bekommen haben. Interessant ist das breite Spektrum, in dem der Behandlungseffekt verteilt ist. Er erstreckt sich von einem stark negativen Therapieverlauf ($\theta = 0.346$) bis zu einer stark protektiven Behandlungswirkung ($\theta = 2.63$).

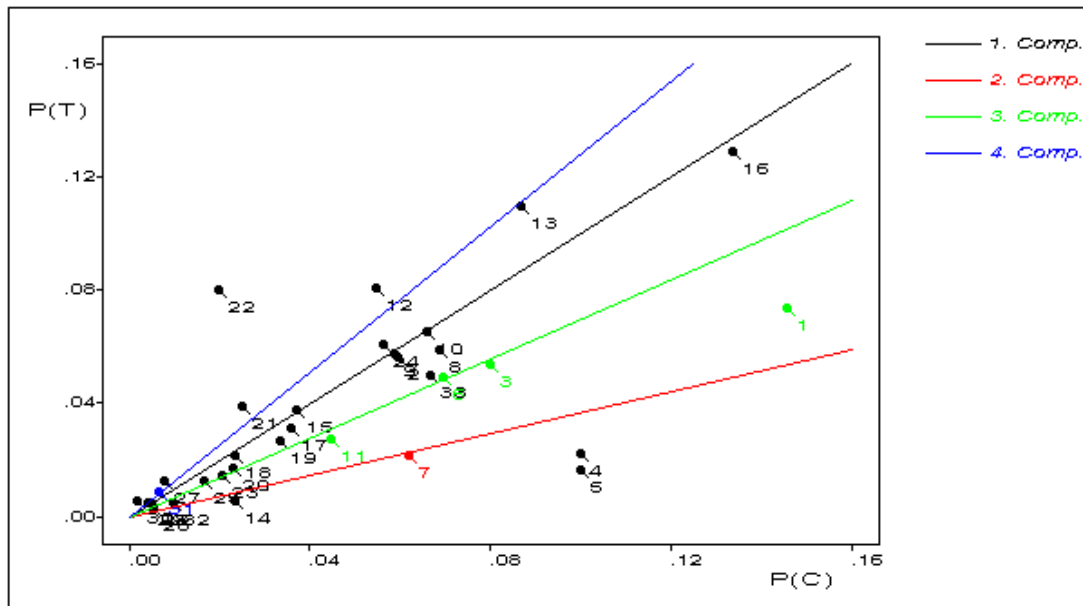


Abbildung 4.4: Zentrenzugehörigkeit bei der Studie über koronarer Herzkrankheit mit vier Mischungskomponenten (Profil-Likelihood-Ansatz). Die Zahlen neben den Punkten geben das Zentrum an.

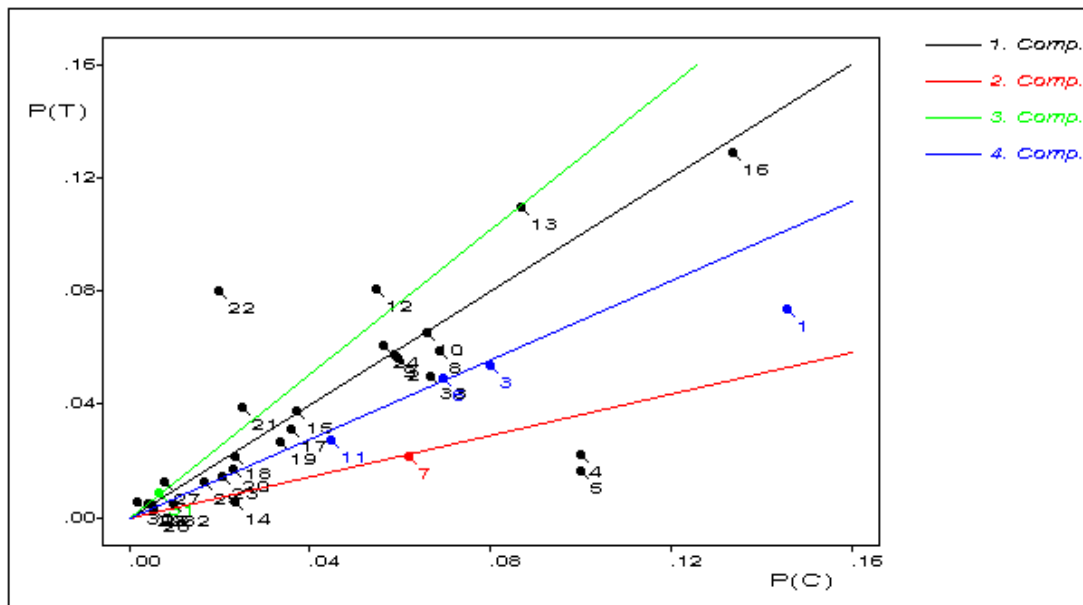


Abbildung 4.5: Zentrenzugehörigkeit bei der Studie über koronarer Herzkrankheit mit vier Mischungskomponenten (Normalverteilungsansatz). Die Zahlen neben den Punkten geben das Zentrum an.

4.2.10 Studie über Vitamin A Ergänzung

Fawzi [16] analysiert diese Studie in seiner Publikation. Er kommt zu dem Schluss, dass durch die Supplementierung mit Vitamin A die Sterblichkeit von Kindern in Entwicklungsländern signifikant gesenkt werden kann. Die Resultate des PL- (Tabelle 4.21) und des ND-Ansatzes (Tabelle 4.22) sind wie bei den vorangegangenen Studien sehr ähnlich. Das NPMLE-Kriterium schätzt in beiden Modellen 4 Mischungskomponenten. Nur das 8. Zentrum wird der ersten Komponente zugeordnet, welche keinen Behandlungseffekt misst. Die anderen Zentren werden Komponenten mit protektiver Behandlungswirkung zugeteilt. Unterschiede ergeben sich mit dem BIC-Kriterium. Das ND-Modell schätzt Homogenität im Behandlungseffekt und der PL-Ansatz findet zwei Mischungskomponenten. Dennoch wird das Relative Risiko in jeder Komponente kleiner als eins geschätzt. Dies bestätigt das Ergebnis von Fawzi. Im ML-Modell (Tabelle 4.23) schätzen das BIC- und das NPMLE-Kriterium 7 Mischungskomponenten bei acht zur Verfügung stehenden Zentren. Nur der ersten Mischungskomponente werden zwei Zentren zugeordnet. Beide Zentren (Zentrum 3 und 8) wird kein Behandlungseffekt unterstellt. Die anderen Mischungskomponenten schätzen einen protektiven Behandlungseffekt.

Tabelle 4.15: Ergebnisse der Studie über Frühgeburten anhand des Multi-Level-Modells mit Binomiallikelihood.

Multi-Level-Modell (Bin)					
Komp.	1.				
θ	.381264				
q	1.000000				
H	14				
Log-L.=	-83.855319				
max GF =	202.036179				
BIC=	-170.349696				
Komp.	1.	2.			
θ	.413366	.394747			
q	.432124	.567876			
H	6	8			
Log-L.=	-66.431525				
max GF =	4.074632				
BIC=	-140.780221				
Komp.	1.	2.	3.		
θ	.502319	.400292	.055187		
q	.304893	.561045	.134062		
H	4	8	2		
Log-L.=	-62.931689				
max GF =	1.947883				
BIC=	-139.058665				
Komp.	1.	2.	3.	4.	5.
θ	.493862	.338739	.054982	.559266	.575859
q	.239738	.332510	.131841	.231763	.064147
H	3	5	2	3	1
Log-L.=	-62.139524				
max GF =	1.000000				
BIC=	-148.030564				

Tabelle 4.16: Ergebnisse der Bluthochdruck-Studie anhand des Profil-Likelihood- und Normalverteilungsansatzes.

Profil-Likelihoodansatz			Normalverteilungsansatz		
Komp.	1.		Komp.	1.	
θ	.901852		θ	.905399	
q	1.000000		q	1.000000	
H	12		H	12	
Log-L.=	-74.798785		Log-L.=	-6.193136	
max GF =	2.224473		max GF =	2.082562	
BIC=	-152.082476		BIC=	-14.871179	
Komp.	1.	2.	Komp.	1.	2.
θ	.803152	1.269410	θ	.807388	1.271288
q	.727099	.272901	q	.720553	.279447
H	9	3	H	9	3
Log-L.=	-72.807825		Log-L.=	-4.342434	
max GF =	1.000000		max GF =	1.000000	
BIC=	-153.070370		BIC=	-16.139588	

Tabelle 4.17: Ergebnisse der Bluthochdruck-Studie anhand des Multi-Level-Modells mit Poissonlikelihood.

Multi-Level-Modell (Poi)						
Komp.	1.					
θ	.894720					
q	1.000000					
H	12					
Log-L.=	-192.563332					
max GF =	8.079×10^{12}					
BIC=	-387.611571					
Komp.	1.	2.				
θ	.810066	1.179001				
q	.583665	.416335				
H	7	5				
Log-L.=	-120.056445					
max GF =	44976.459433					
BIC=	-247.567609					
Komp.	1.	2.	3.	4.		
θ	.843486	1.179331	.802756	.512193		
q	.253775	.412961	.250072	.083192		
H	3	5	3	1		
Log-L.=	-91.450179					
max GF =	3.394873					
BIC=	-200.294704					
Komp.	1.	2.	3.	4.	5.	6.
θ	.899278	1.285227	.746362	.512193	.955467	.536168
q	.179181	.230475	.168218	.082926	.082137	.074798
H	2	4	2	1	1	1
Komp.	7.	8.	9.			
θ	.720397	3.973371	1.486697			
q	.103324	.026812	.052129			
H	1	0	0			
Log-L.=	-88.244806					
max GF =	1.000000					
BIC=	-218.733025					

Tabelle 4.18: Ergebnisse der Studie über Patienten mit koronarer Herzkrankheit anhand des Profil-Likelihoodansatzes.

Profil-Likelihoodansatz				
Komp.	1.			
θ	.971608			
q	1.000000			
H	33			
Log-L.=	-202.370689			
max GF =	10518.113918			
BIC=	-408.237886			
Komp.	1.	2.		
θ	1.005849	.440127		
q	.890148	.109852		
H	31	2		
Log-L.=	-191.762036			
max GF =	1.785560			
BIC=	-394.013595			
Komp.	1.	2.	3.	4.
θ	1.001625	.366457	.696174	1.279346
q	.654638	.055751	.191597	.098014
H	27	1	4	1
Log-L.=	-189.850417			
max GF =	1.000000			
BIC=	-404.176387			

Tabelle 4.19: Ergebnisse der Studie über Patienten mit koronarer Herzkrankheit anhand des Normalverteilungsansatzes.

Normalverteilungsansatz				
Komp.	1.			
θ	.968171			
q	1.000000			
H	33			
Log-L.	= -31.310858			
max GF	= 8922.156693			
BIC	= -66.118224			
Komp.	1.	2.		
θ	1.002424	.436286		
q	.880438	.119562		
H	31	2		
Log-L.	= -20.929687			
max GF	= 1.823076			
BIC	= -52.348897			
Komp.	1.	2.	3.	4.
θ	1.000325	.364379	1.273473	.697894
q	.647332	.062646	.086127	.203895
H	27	1	1	4
Log-L.	= -19.081388			
max GF	= 1.000001			
BIC	= -62.638328			

Tabelle 4.20: Ergebnisse der Studie über Patienten mit koronarer Herzkrankheit anhand des Multi-Level-Modells mit Poissonlikelihood.

Multi-Level-Modell (Poi)						
Komp.	1.					
θ	1.077710					
q	1.000000					
H	33					
Log-L.	= -2300.940676					
max GF	$= 1.714 \times 10^{174}$					
BIC	= -4605.377859					
Komp.	1.		2.			
θ	.897089		1.064336			
q	.677055		.322945			
H	22		11			
Log-L.	= -613.797243					
max GF	$= 2.487 \times 10^{61}$					
BIC	= -1238.084009					
Komp.	1.	2.	3.	4.	5.	6.
θ	1.017394	1.058878	.970249	.883665	.790761	1.317561
q	.140165	.124920	.320332	.088742	.169898	.079452
H	4	5	12	3	5	2
Komp.	7.					
θ	.428063					
q	.076492					
H	2					
Log-L.	= -247.100686					
max GF	$= 12.893014$					
BIC	= -539.655969					
Komp.	1.	2.	3.	4.	5.	6.
θ	1.007591	1.022630	1.020841	.965525	.817292	1.306739
q	.077422	.104689	.226971	.034156	.113021	.056061
H	2	4	10	1	5	1
Komp.	7.	8.	9.	10.	11.	12.
θ	.347178	.506721	.759630	1.258181	1.481572	1.587610
q	.045034	.033068	.058496	.028397	.035870	.023181
H	1	1	1	1	1	1
Komp.	13.	14.	15.			
θ	.654683	.766163	2.652461			
q	.055444	.087982	.020206			
H	1	2	1			
Log-L.	= -238.911862					
max GF	$= .999684$					
BIC	= -579.222443					

Tabelle 4.21: Ergebnisse der Vitamin A Ergänzungs-Studie anhand des Profil-Likelihoodansatzes.

Profil-Likelihoodansatz				
Komp.	1.			
θ	.727809			
q	1.000000			
H	8			
Log-L.=	-64.084199			
max GF =	50.416831			
BIC=	-130.247839			
Komp.	1.	2.		
θ	.766144	.374227		
q	.768090	.231910		
H	6	2		
Log-L.=	-60.618985			
max GF =	1.843529			
BIC=	-127.476295			
Komp.	1.	2.	3.	4.
θ	1.028845	.441692	.201888	.720406
q	.176796	.104857	.110089	.608259
H	1	1	1	5
Log-L.=	-58.970705			
max GF =	1.000000			
BIC=	-132.497500			

Tabelle 4.22: Ergebnisse der Vitamin A Ergänzungs-Studie anhand des Normalverteilungsansatzes.

Normalverteilungsansatz				
Komp.	1.			
θ	.734352			
q	1.000000			
H	8			
Log-L.=	-5.003991			
max GF =	15.183379			
BIC=	-12.087425			
Komp.	1.	2.		
θ	.746417	.202093		
q	.881336	.118664		
H	7	1		
Log-L.=	-3.103103			
max GF =	3.209860			
BIC=	-12.444531			
Komp.	1.	2.	3.	4.
θ	1.028628	.201718	.439455	.719834
q	.177152	.100428	.106623	.615798
H	1	1	1	5
Log-L.=	-1.195958			
max GF =	1.000001			
BIC=	-16.948006			

Tabelle 4.23: Ergebnisse der Vitamin A Ergänzungs-Studie anhand des Multi-Level-Modells mit Poissonlikelihood.

Multi-Level-Modell (Poi)						
Komp.	1.					
θ	.742689					
q	1.000000					
H	8					
Log-L.	= -932.721088					
max GF	= 6.277×10^{194}					
BIC	= -1867.521618					
Komp.	1.	2.				
θ	.735890	.726740				
q	.750000	.250000				
H	6	2				
Log-L.	= -264.716400					
max GF	= 3.885×10^{36}					
BIC	= -535.671124					
Komp.	1.	2.	3.	4.	5.	6.
θ	1.031078	.701486	.744492	.706622	.201584	.729302
q	.250000	.125000	.125000	.125000	.125000	.125663
H	2	1	1	1	1	1
Komp.	7.					
θ	.456022					
q	.124336					
H	1					
Log-L.	= -65.749965					
max GF	= 1.000000					
BIC	= -158.532671					

4.2.11 Brustkrebsvorsorgestudie

Diese Studie analysiert das Risiko und den Nutzen einer Mammographie (Röntgendiagnostische Untersuchung der weiblichen Brust, mit weichen Röntgenstrahlen unter Zuhilfenahme von Kontrastmitteln) bei Frauen im Alter von 40 bis 49 Jahren. Da hier Personenzeiten betrachtet werden, wird den Beobachtungen eine Poissonverteilung unterstellt. Im PL- und im ND-Ansatz (Tabelle 4.24) schätzt das BIC-Kriterium Homogenität im Effekt und das NPMLE-Kriterium 2 Komponenten, wobei der Anteil der 2. Komponente sehr gering ist. Zu dieser zweiten Komponente, mit einem Relativen Risiko in der Nähe von eins, gehört nur das 6. Zentrum. Die erste Komponente schätzt eine protektive Behandlungswirkung, mit einem Relativen Risiko von 0.75 (in beiden Modellen). Die Ergebnisse im ML-Modell (Tabelle 4.25) unterscheiden sich wieder in der Anzahl der Mischungskomponenten. Das BIC-Kriterium schätzt 2 Komponenten und das NPMLE-Kriterium findet 5 Mischungskomponenten. Zwei Mischungskomponenten (die 2. und 4.) besitzen nach dem NPMLE-Kriterium keinen Behandlungseffekt. Zur 2. Komponente gehört das 6. Zentrum. Genau dieses Zentrum ist auch bei den anderen beiden Modellen als nicht protektiv herausgefiltert worden. Zur 4. Komponente gehört nur das 7. Zentrum. Die anderen Zentren gehören zu Mischungskomponenten, welche einen Behandlungserfolg schätzen.

4.2.12 Nikotinersatzstudie

DuMouchel [13] untersucht diese Studie mit den hier nicht angegebenen Kovariablen. Nach dem BIC-Kriterium liegt im ND- (Tabelle 4.27) und im PL-Modell (Tabelle 4.26) Homogenität im Effekt vor. Das Relative Risiko ist in beiden Modellen größer als 1. Das heißt, in der Behandlungsgruppe haben mehr Probanden mit dem Rauchen aufgehört als in der Kontrollgruppe. Das NPMLE-Kriterium schätzt im PL-Modell 3 Mischungskomponenten und im ND-Modell 2 Komponenten. In beiden Modellen schätzt die zweite Komponente ein Relatives Risiko von 1.3, was für einen leicht positiven Behandlungseffekt spricht. Hingegen schätzt die erste Komponente in beiden Modellen ein etwas größeres Relatives Risiko von 1.8. Die dritte Komponente im PL-Ansatz schätzt dagegen ein deutlich größeres Relatives Risiko von 5.3. Aber zu dieser Komponente wird kein Zentrum zugeordnet, obwohl die Zentren 26 und 57 fast zur dritten Komponente gehören (die Wahrscheinlichkeit, dass Zentrum 26 zur dritten Komponente gehört ist 0.375 und beim Zentrum 57 beträgt sie 41 Prozent). Der Einfluss der Kovariaten ist in beiden Modellen nicht groß genug, um sie durch die unbeobachtete Heterogenität identifizieren zu können. Im ML-Modell (Tabelle 4.28) werden 6 Mischungskomponenten mit dem BIC-Kriterium und 14 Komponenten

mit dem NPMLE-Kriterium geschätzt. Im Modell mit 6 Mischungskomponenten wird ein Relatives Risiko im Intervall von 1.2 bis 1.9 geschätzt. Es liegt damit in der Dimension der anderen zwei Modelle. Auch nach dem NPMLE-Abbruchkriterium schätzen alle Mischungskomponenten einen Behandlungserfolg. Besonders groß ist das Relative Risiko in der 7. ($RR = 4.32$) und in der 11. ($RR = 7.07$) Komponente.

4.2.13 Kaiserschnittsstudie

Die Ergebnisse (Tabelle 4.29 und 4.30) sind bei dieser Studie recht eindeutig. Im PL-Modell schätzen die beiden Abbruchkriterien ein zwei Komponentenmodell. Beide Komponenten schätzen ein Relatives Risiko, welches deutlich kleiner als 1 ist. Ein ähnliches Resultat wird im ND-Ansatz geschätzt. Hier ist der BIC-Wert im Homogenitätsfall etwas geringer als im Zwei-Komponentenmodell. Mit dem NPMLE-Kriterium werden zwei Mischungskomponenten gefunden. Beide Mischungskomponenten schätzen eine sehr protektive Therapiewirkung. Mit dem ML-Modell werden vier Mischungskomponenten mit dem BIC-Kriterium und neun Komponenten mit dem NPMLE-Kriterium geschätzt. Nur eine Mischungskomponente im Neun-Komponentenmodell schätzt eine leicht negative Behandlungswirkung. Zu dieser 7. Mischungskomponente wird nur das 38. Zentrum gezählt. Alle anderen Komponenten schätzen ein Relatives Risiko, das deutlich kleiner als 1 ist.

Tabelle 4.24: Ergebnisse der Brustkrebsvorsorge-Studie anhand des Profil-Likelihood- und Normalverteilungsansatzes.

Profil-Likelihoodansatz			Normalverteilungsansatz		
Komp.	1.		Komp.	1.	
θ	.835719		θ	.836876	
q	1.000000		q	1.000000	
H	8		H	8	
Log-L.=	-48.228351		Log-L.=	-.548866	
max GF =	1.257809		max GF =	1.240221	
BIC=	-98.536144		BIC=	-3.177174	
Komp.	1.	2.	Komp.	1.	2.
θ	.748924	1.052290	θ	.750663	1.050661
q	.722436	.277564	q	.724188	.275812
H	7	1	H	7	1
Log-L.=	-47.806530		Log-L.=	-.152324	
max GF =	1.000000		max GF =	1.000000	
BIC=	-101.851384		BIC=	-6.542973	

Tabelle 4.25: Ergebnisse der Brustkrebsvorsorge-Studie anhand des Multi-Level-Modells mit Poissonlikelihood.

Multi-Level-Modell (Poi)					
Komp.	1.				
θ	.826971				
q	1.000000				
H	8				
Log-L.	= -70.863333				
max GF	= 7195.769697				
BIC	= -143.806107				
Komp.	1.	2.			
θ	.875870	.724309			
q	.751217	.248783			
H	6	2			
Log-L.	= -58.661726				
max GF	= 3.627083				
BIC	= -123.561776				
Komp.	1.	2.	3.	4.	5.
θ	1.089237	.670254	.746877	1.086632	.790351
q	.161595	.126867	.476913	.114730	.119895
H	1	1	4	1	1
Log-L.	= -56.365773				
max GF	= 1.000059				
BIC	= -131.446519				

Tabelle 4.26: Ergebnisse der Nikotinersatz-Studie anhand des Profil-Likelihood-Modells.

Profil-Likelihoodansatz			
Komp.	1.		
θ	1.565594		
q	1.000000		
H	59		
Log-L.= -308.468545			
max GF = 1.810375			
BIC= -621.014627			
Komp.	1.	2.	
θ	1.801041	1.302956	
q	.652501	.347499	
H	49	10	
Log-L.= -306.692237			
max GF = 1.303026			
BIC= -625.617087			
Komp.	1.	2.	3.
θ	1.760105	1.301204	5.316015
q	.641776	.334241	.023983
H	52	7	0
Log-L.= -306.507837			
max GF = 1.000000			
BIC= -633.403362			

Tabelle 4.27: Ergebnisse der Nikotinersatz-Studie anhand des Normalverteilungsansatzes.

Normalverteilungsansatz		
Komp.	1.	
θ	1.521802	
q	1.000000	
H	59	
Log-L.= -28.571423		
max GF = 1.169676		
BIC= -61.220384		
Komp.	1.	2.
θ	1.754545	1.290625
q	.624573	.375427
H	47	12
Log-L.= -26.432342		
max GF = 1.000000		
BIC= -65.097296		

Tabelle 4.28: Ergebnisse der Nikotinersatz-Studie anhand des Multi-Level-Modells mit Binomiallikelihood.

Multi-Level-Modell (Bin)						
Komp.	1.					
θ	1.679105					
q	1.000000					
H	59					
Log-L.	= -994.180062					
max GF	= 9.6781×10^{28}					
BIC	= -1992.437661					
Komp.	1.	2.				
θ	1.520305	1.599367				
q	.598932	.401068				
H	35	24				
Log-L.	= -545.529331					
max GF	= 1.079×10^9					
BIC	= -1103.291274					
Komp.	1.	2.	3.	4.	5.	6.
θ	1.498993	1.935376	1.551996	1.370336	1.609925	1.213907
q	.382461	.241928	.208828	.021881	.050251	.094650
H	24	15	12	1	3	4
Log-L.	= -404.220101					
max GF	= 5.250144					
BIC	= -853.293113					
Komp.	1.	2.	3.	4.	5.	6.
θ	1.241994	1.736586	1.603252	1.369419	1.458342	1.189433
q	.084656	.146929	.138690	.020914	.033578	.083808
H	2	10	10	1	2	4
Komp.	7.	8.	9.	10.	11.	12.
θ	4.324630	1.986669	1.845638	1.141219	7.070220	1.802443
q	.059891	.024554	.104035	.044391	.015360	.172464
H	4	1	7	1	1	13
Komp.	13.	14.				
θ	1.339822	1.818708				
q	.052103	.018626				
H	2	1				
Log-L.	= -396.680763					
max GF	= 1.000000					
BIC	= -903.455037					

Tabelle 4.29: Ergebnisse der Kaiserschnitt-Studie anhand des Profil-Likelihood- und Normalverteilungsansatzes.

Profil-Likelihoodansatz			Normalverteilungsansatz		
Komp.	1.		Komp.	1.	
θ	.385956		θ	.423938	
q	1.000000		q	1.000000	
H	61		H	61	
Log-L.=	-205.009470		Log-L.=	-83.384134	
max GF =	1.253640		max GF =	1.331974	
BIC=	-414.129815		BIC=	-170.879143	
Komp.	1.	2.	Komp.	1.	2.
θ	.579242	.195208	θ	.218131	.596583
q	.516889	.483111	q	.449905	.550095
H	28	33	H	29	32
Log-L.=	-200.571639		Log-L.=	-79.305550	
max GF =	.999995		max GF =	1.000000	
BIC=	-413.475900		BIC=	-170.943722	

Tabelle 4.30: Ergebnisse der Kaiserschnitt-Studie anhand des Multi-Level-Modells mit Binomiallikelihood.

Multi-Level-Modell (Bin)						
Komp.	1.					
θ	.383249					
q	1.000000					
H	61					
Log-L.=	-369.958713					
max GF =	822237.092841					
BIC=	-744.028301					
Komp.	1.	2.				
θ	.326945	.397544				
q	.601701	.398299				
H	39	22				
Log-L.=	-289.943449					
max GF =	106.047451					
BIC=	-592.219520					
Komp.	1.	2.	3.	4.		
θ	.196539	.653268	.350245	.641574		
q	.280884	.115927	.549427	.053762		
H	18	5	35	3		
Log-L.=	-270.003045					
max GF =	3.531663					
BIC=	-568.782208					
Komp.	1.	2.	3.	4.	5.	6.
θ	.210178	.648788	.296144	.563431	.425586	.000000
q	.144556	.115654	.255808	.041617	.048704	.089046
H	11	6	18	3	1	8
Komp.	7.	8.	9.			
θ	1.225050	.567153	.170837			
q	.012334	.186931	.105348			
H	1	10	3			
Log-L.=	-265.893237					
max GF =	1.000091					
BIC=	-601.671331					

Tabelle 4.31: Schätzung der Anzahl der Mischungskomponenten bei allen Studien.

Daten	Anzahl der Komp.					
	NDM		PLM		MLM	
	BIC	NPMLE	BIC	NPMLE	BIC	NPMLE
Tabelle 1.1	1	2	1	2	5	8
Tabelle 1.2	1	2	2	2	4	9
Tabelle 1.3	2	3	2	3	6	7
Tabelle 1.4	1	2	1	1	1	4
Tabelle 1.5	1	1	1	1	1	1
Tabelle 1.6	1	1	1	1	1	3
Tabelle 1.7	2	2	1	2	3	5
Tabelle 1.8	1	2	1	2	4	9
Tabelle 1.9	2	4	2	4	7	15
Tabelle 1.10	1	4	2	4	7	7
Tabelle 1.11	1	2	1	2	2	5
Tabelle 1.12	1	2	1	3	6	14
Tabelle 1.13	1	2	2	2	4	9

4.2.14 Zusammenfassung

Zunächst wird sehr schnell deutlich, dass die Ergebnisse im ND- und im PL-Ansatz fast kongruent sind. Besonders in der Schätzung des Relativen Risikos und der Zentreinteilung sind die Unterschiede sehr gering. Klare Unterschiede sind dagegen beim ML-Modell zu erkennen. Bei fast allen Studien werden deutlich mehr Komponenten gefunden, wobei mit dem NPMLE-Kriterium oft Mischungskomponenten geschätzt werden, denen nur ein oder gar kein Zentrum zugewiesen wird. Diese Besonderheit verdeutlicht Tabelle 4.31, welche die Komponentenanzahl zu jeder Studie je nach Modell und Abbruchkriterium abbildet. Im ML-Modell werden die Zentren nicht nur nach dem Relativen Risiko eingeteilt, sondern auch nach dem Baseline-Parameter. Bei Arends [3] wird davon ausgegangen, dass das Baseline-Risiko mit dem Behandlungseffekt zusammenhängt. Dieser Zusammenhang wird im ML-Modell berücksichtigt, wobei im PL- und im ND-Ansatz die Zentren nur nach dem Behandlungseffekt eingeteilt werden. Momentan ist nur die Mittelwertstruktur in den berechneten Ergebnissen untersucht worden. Um weitergehende Aussagen treffen zu können, muss die Varianz- und damit die Konfidenzintervallschätzung mit berücksichtigt werden. Als etablierte Verfahren zur Schätzung der Varianz gelten Simulationsstudien. Im nächsten Abschnitt werden zwei Simulationsmethoden eingeführt und an den drei Modellen angewendet.

Kapitel 5

Simulationsstudie

In diesem Kapitel werden die Eigenschaften der Mischverteilung anhand einer Bootstrap-Simulation (publiziert von Efron [14,15]) dargestellt. Beim Bootstrap-Verfahren ist die Stichprobe X_1, \dots, X_n mit Schätzfunktion $T(X_1, \dots, X_n)$ bekannt. Gesucht sind statistische Kennwerte wie der Bias, die Varianz und der kleinste quadratische Fehler (MSE) der Schätzfunktion T . Dazu wird zuerst die Verteilung \hat{F} aus der Stichprobe geschätzt. Als nächstes wird eine Bootstrap-Stichprobe gezogen. Es gibt verschiedene Möglichkeiten, eine Bootstrap-Stichprobe zu generieren. Im herkömmlichen Ansatz werden aus der Originalstichprobe zufällig n Elemente gezogen, wobei Wiederholungen zugelassen sind. Da hier aber die Verteilung der Beobachtungen bekannt ist, setzt sich in dem Fall die Bootstrap-Stichprobe aus zufällig binomial- oder Poissonverteilten Beobachtungen zusammen. Zuerst wird der Homogenitätsfall betrachtet. Danach wird Effektheterogenität zugelassen.

5.1 Unter Effekthomogenität

Im Homogenitätsfall wird ein festes Relatives Risiko (RR) als bekannt vorausgesetzt. Um zu einem besseren Verständnis über die Eigenschaften statistischer Kennwerte, wie der Erwartungstreue, der Varianz und des kleinsten quadratischen Fehlers zu gelangen, werden mehrere Werte für das RR (hier werden 30 RR -Werte berechnet) aus einem bestimmten Intervall vorgegeben. Das Intervall für das RR hat feste Grenzen. Wobei die untere Grenze Null und die obere $\max_{p^C} \frac{1}{p_i^C}$ ist. Die Baseline-Werte werden von der gewählten Studie

übernommen. Die Baselinewahrscheinlichkeit im Behandlungsarm ist definiert als

$$\begin{aligned} p_i^T &= \theta \times \hat{p}_i^C \\ \text{mit } \hat{p}_i^C &= \frac{x_i^C}{n_i^C} \\ \text{und } \theta &= \text{ vorgegebenes Relatives Risiko} \end{aligned}$$

Die n_i^T - und n_i^C - Werte werden ebenfalls von der Studie entnommen. Damit sind alle Parameter definiert, die notwendig sind, um die Beobachtungen in jedem Behandlungsarm zu generieren. Die zugrunde liegende Verteilung ist abhängig vom Studientyp. Werden Personenzeiten betrachtet, sind die simulierten Beobachtungen im Behandlungs- und Kontrollarm Poissonverteilt, mit

$$x_i^T \sim Poi(p_i^T \times n_i^T) \text{ und } x_i^C \sim Poi(p_i^C \times n_i^C).$$

Ansonsten sind die simulierten Werte binomialverteilt, mit

$$x_i^T \sim Bin(p_i^T, n_i^T) \text{ und } x_i^C \sim Bin(p_i^C, n_i^C).$$

Mit den erzeugten Daten wird der Schätzer für das Relative Risiko ($\theta_{ML}^r, \theta_{PL}^r, \theta_{ND}^r$) je nach Modell berechnet. Dieser Vorgang wird 10.000 mal wiederholt. Der Erwartungswert ist

$$E(\theta) = \frac{1}{10.000} \sum_{r=1}^{10000} \theta^r$$

und die Varianz gegeben als

$$Var(\theta) = \frac{1}{10.000} \sum_{r=1}^{10000} (\theta^r - E(\theta))^2.$$

Damit kann nun der Bias ($B(\theta) = E(\theta) - \theta$) und der kleinste quadratische Fehler ($MSE(\theta) = Var(\theta) + B(\theta)^2$) berechnet werden. Im Weiteren wird die Bootstrap-Simulation an ausgewählten Studien angewendet.

5.1.1 Beta-Blocker Studie

Die Abbildung 5.1 zeigt drei Grafiken. Im oberen Teilbild ist der Bias der drei Modelle dargestellt. Bezüglich des Gesamtintervalls weist der Schätzer des PL-Modells die kleinste Differenz zum wahren Wert auf. Der Schätzer des Bias zeigt einen leicht positiven Trend.

Der Schätzer des ML-Modells unterschätzt im gesamten Intervall das vorgegebene Relative Risiko, wobei bei ansteigendem Relativen Risiko die Differenz zum wahren Wert immer größer wird. Im protektiven Fall des ND-Ansatzes (Relative Risiko kleiner als eins) liegt der Bias im positiven Bereich. Im nicht protektiven Fall (Relative Risiko größer als eins) zeigt sich der Bias dieses Schätzers als negativ. Das heißt, ist der wirkliche Wert des Relativen Risikos kleiner als eins, schätzt das ND-Modell ein größeres Relatives Risiko. Der Grad der positiven Behandlungswirkung wird verkleinert. Im nicht protektiven Fall gilt, dass der Grad der negativen Behandlungswirkung auch verkleinert wird. Dieses Bild spiegelt sich in 8 von den hier vorgestellten multizentrischen Studien wider. In der Tabelle 4.1 (Beta-Blocker-Studie, Zeile 1) weisen alle drei Modelle ein Relatives Risiko von 0.79 auf, wobei der Schätzer des ND-Ansatzes mit 0.792007 den größten Wert besitzt, und das ML-Modell mit 0.791032 den kleinsten Wert für das Relative Risiko schätzt. Nach der Simulationsstudie müsste der wahre Wert zwischen diesen beiden Ansätzen zu finden sein. Im mittleren Teilbild der Abbildung 5.1 ist die Varianz dargestellt. Alle drei Modelle zeigen nahezu eine identische Varianz. Auffällig ist der stetige Anstieg der Varianz im Verlaufe des Intervalls von 0.01 bis 4.68. Ein ähnliche Situation zeigt sich im unteren Teilbild. Dort ist der kleinste quadratische Fehler (MSE) abgebildet, der wie bei der Varianz im Verlauf des Intervalls stetig ansteigt. Die Schätzer des PL- und des ML-Modells haben bei größeren Werten für das Relative Risiko einen geringfügig kleineren MSE-Wert als der Schätzer des ND-Ansatzes.

5.1.2 CALGB-Studie

Das Ergebnis der Simulation bei der CALGB-Studie (Abbildung 5.2) unterscheidet sich von der vorangegangenen Studie. Der PL-Ansatz hat wie im ersten Beispiel den geringsten Bias, wohingegen der ND- und der ML-Ansatz einen positiven Bias im gesamten Intervall aufweisen. In der Varianz unterscheidet sich der ND-Ansatz deutlich von den anderen zwei Modellen. Nach einem stetigen Anstieg erreicht die Varianz kurz vor Mitte des Intervalls ihr Maximum. Im weiteren Verlauf verringert sie sich aber wieder. Dies wirkt sich auch auf die Schätzung des MSE aus, welcher beim ND-Ansatz gegen Ende des Intervalls am geringsten ist.

5.1.3 Multi-Farm-Studie

In dieser Bootstrapsimulation anhand der Multi-Farm-Studie (Abbildung 5.3) zeigen das ML- und das ND-Modell im oberen Teilbild einen ähnlichen Verlauf. Bei einem Relati-

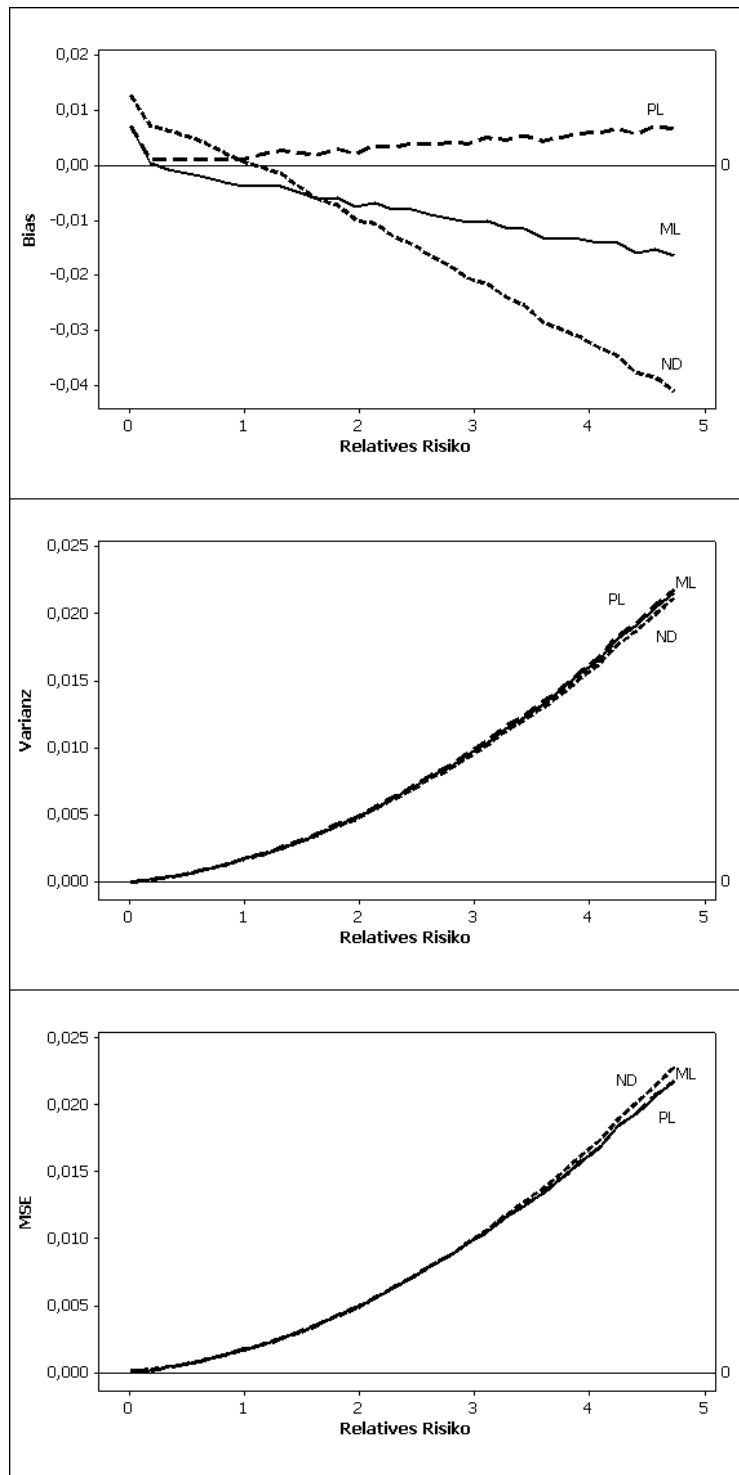


Abbildung 5.1: Beta-Blocker Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;4.73]$ berechnet, die generierten x^T , x^C -Werte sind binomialverteilt)

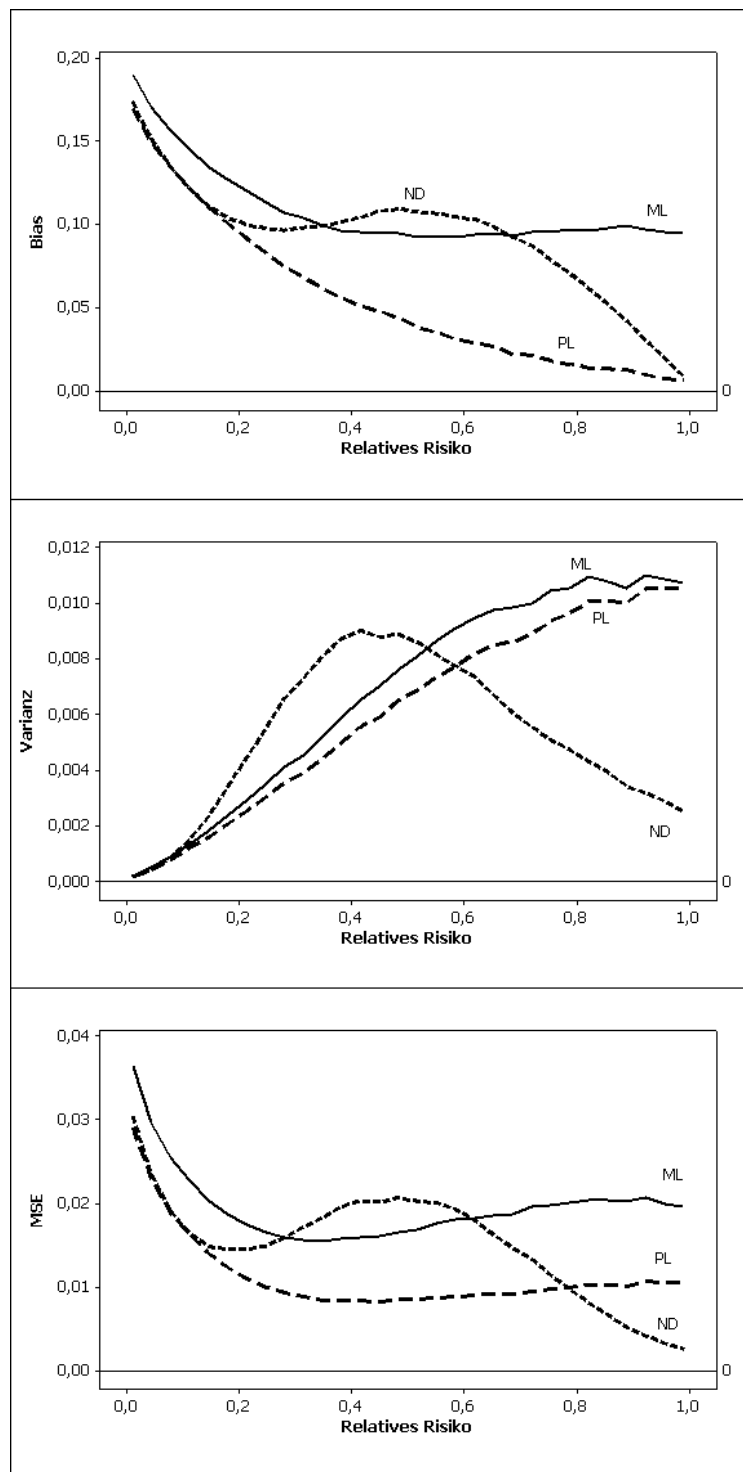


Abbildung 5.2: CALGB-Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;0.99]$ berechnet, die generierten x^T , x^C -Werte sind binomialverteilt)

ven Risiko um 1.2 haben die Schätzer der zwei Modelle den geringsten Bias. Bei größer werdenden RR steigt aber der absolute Wert des Bias vom ND- und vom ML-Modell schneller an als der Werte des Bias vom PL-Modell. Bezüglich der Varianz ist das ND-Modell im Intervall des vorgegebenen Relativen Risikos am geringsten. Beim PL-Modell ist die Schwankung des Schätzers für das RR am größten. Ein Nachteil dieser Simulationsstudie ist, dass der eigentliche Schätzer des Relativen Risikos ($\theta_{PL} = 8.4119, \theta_{ML} = 7.6731, \theta_{ND} = 7.7028$) nicht mit in dem Intervall des vorgegebenen Relativen Risikos liegt. Auffällig ist der hohe Wert der Varianz. So ist hier beim PL-Modell die Varianz des Schätzers ungefähr 100-mal größer als bei der Beta-Blocker-Studie mit einem vorgegebenen Relativen Risiko von 4.7.

5.1.4 Studie über koronare Herzkrankheit

Da es sich bei dieser Studie um Personenzeiten handelt, werden zufällig Poissonverteilte Beobachtungen für den Kontroll- und den Behandlungsarm generiert. Ein besonders adäquates Modell ist hier (Abbildung 5.4) der PL-Ansatz, welcher im gesamten Intervall deutlich den kleinsten Bias-, den kleinsten Varianz- und den kleinsten MSE-Wert hat. Der ND-Ansatz hat wie das ML-Modell im gesamten Intervall einen positiven Bias. Insgesamt schneidet hier der ND-Ansatz am schlechtesten ab. Auffällig ist hier, dass der Bias-Wert im Verlaufe des Intervalls stetig ansteigt und nicht die Charakteristik wie beispielsweise bei der Beta-Blocker Studie im Binomialfall zeigt. Würde dieser Versuch mit binomialverteilten Beobachtungen durchgeführt werden, dann hätte der Bias des Schätzers vom ND-Ansatz einen ähnlichen Verlauf wie in der Abbildung 5.1.

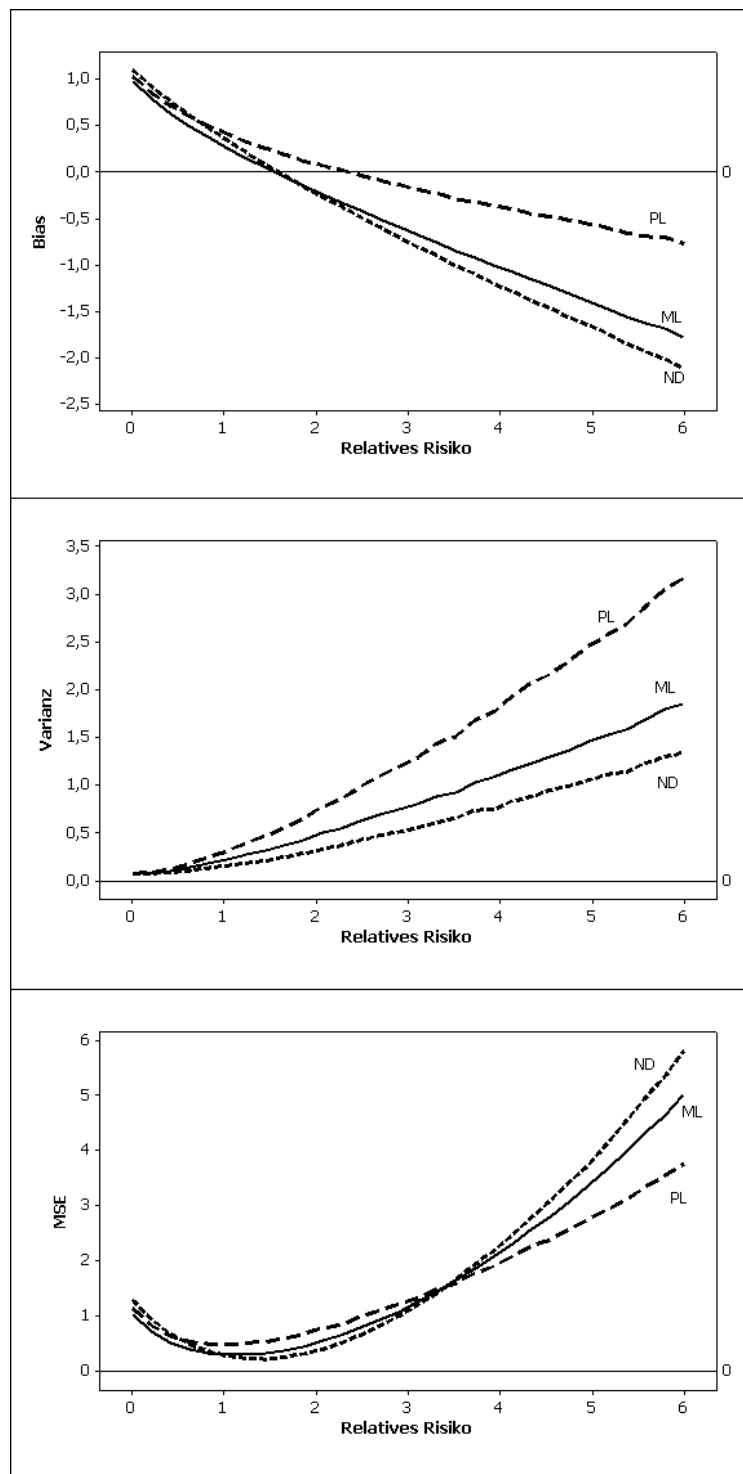


Abbildung 5.3: Multi-Farm-Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;5.99]$ berechnet, die generierten x^T , x^C - Werte sind binomialverteilt)

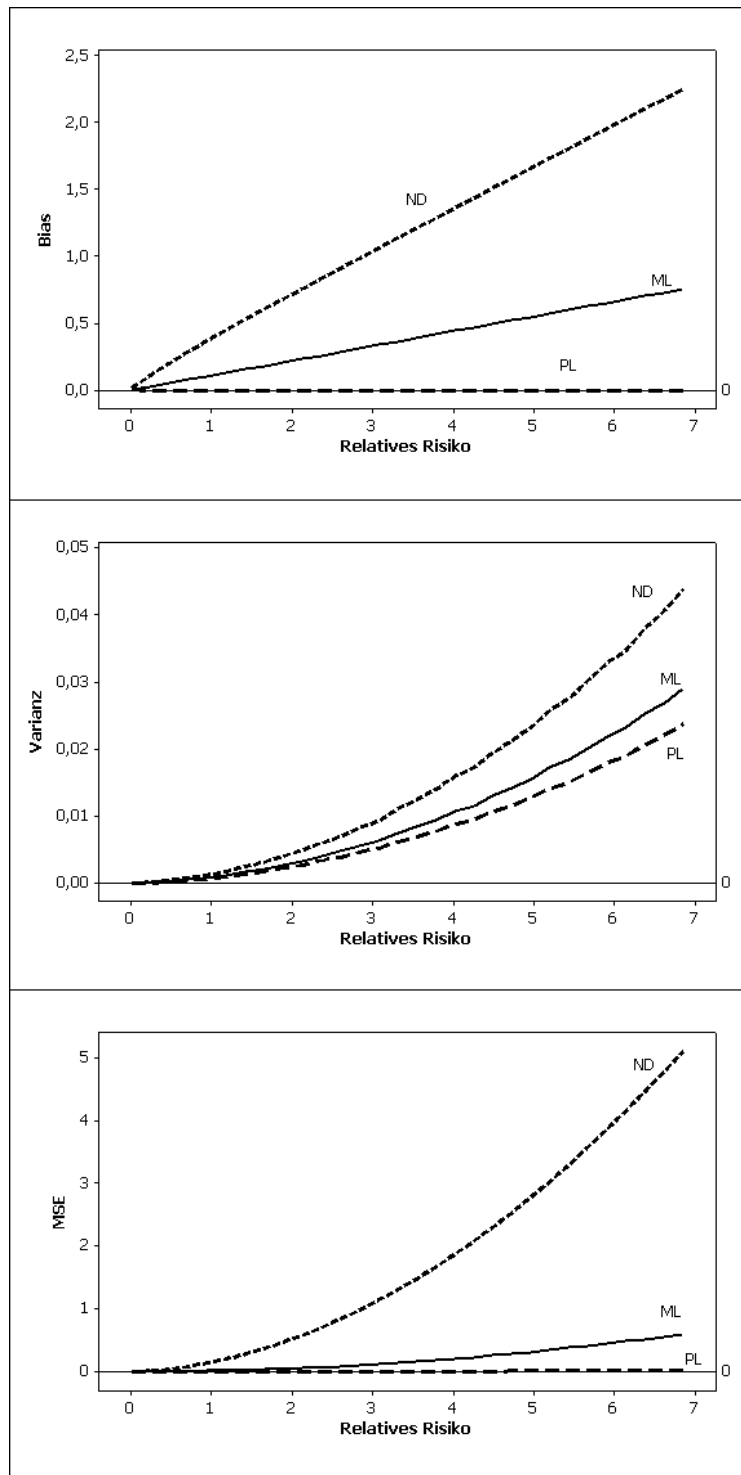


Abbildung 5.4: Studie über koronarer Herzkrankheit (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;6.85]$ berechnet, die generierten x^T , x^C - Werte sind Poissonverteilt)

5.2 Unter Effektheterogenität

Im Heterogenitätsfall wird bei der Simulation eine diskrete Verteilung $P = \begin{pmatrix} \theta_1 & \dots & \theta_m \\ q_1 & \dots & q_m \end{pmatrix}$ vorgegeben. Die Baselinewerte können entweder aus einer vorliegenden Studie übernommen werden (Bootstrapfall), oder sie werden aus einem gleichverteilten Intervall zufällig gezogen. Die p_i^T -Werte ergeben sich aus $p_i^T = \theta_z p_i^C$ mit $z = 1, \dots, m$. Der Index z gibt an, zu welcher Komponente das Zentrum gehört. Es gilt $P(Z = z) = q_z$. Wird nicht die Bootstrap-Methode verwendet, sind weitere Angaben notwendig. Um eine gewisse Variabilität in den Stichprobengrößen der Behandlungs- und Kontrollarme zu simulieren, werden die Parameter n^T und n^C aus einer Poissonverteilung mit dem Parameter λ gezogen. Des Weiteren wird die Anzahl der Zentren vor der Simulation festgelegt. Damit sind alle notwendigen Parameter definiert, um die Beobachtungen nach derselben Vorgehensweise wie im Homogenitätsfall zu generieren. Insgesamt werden bei jeder Simulation 1000 Replikationen gezogen. Direkt vergleichbar sind die Momente einer diskreten Verteilung. Das erste Moment ist der Erwartungswert

$$E(P) = \sum_{i=1}^m \theta_i q_i.$$

Das zweite zentrale Moment beschreibt die Varianz

$$Var(P) = \sum_{i=1}^m (\theta_i - E(P))^2 q_i.$$

Weiterhin muss die Qualität der einzelnen Komponenten-Schätzungen untersucht werden. Dazu werden alle die Verteilungen betrachtet, bei denen die Anzahl an Komponenten mit der vorgegebenen Verteilung übereinstimmt.

Alle Permutationen einer mischenden Verteilung besitzen den gleichen Likelihoodwert. Somit haben die mit dem gEMGFU berechneten Lösungen keine Ordnung. Eine vorgegebene Verteilung $P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$ kann zum Beispiel nicht mit der geschätzten Verteilung $P^{rep} = \begin{pmatrix} 1.47 & 0.59 \\ 0.45 & 0.55 \end{pmatrix}$ verglichen werden, da die erste Komponente der geschätzten Verteilung nur zur zweiten Komponente der vorgegebenen Verteilung zugeordnet werden kann. Dieses Problem wird als „label switching“ bezeichnet und ist bei Richardson [30] publiziert. Celeux [10] löst dieses Problem indem er die Permutation sucht, die den geringsten Abstand zum wahren Wert hat. Richardson sortiert die Verteilung nach dem

gesuchten Parameter oder nach den Gewichten. Diese Methode ist nur sinnvoll, wenn die Parameter gut separiert sind. In dieser Arbeit werden die geschätzten Verteilungen nach dem Relativen Risiko sortiert. Voraussetzung ist dafür, dass auch die vorgegebene Verteilung nach der gleichen Ordnung sortiert ist. So können für jede Komponente θ_j und für jedes Gewicht q_j der Erwartungswert und die Varianz berechnet werden.

5.2.1 Beta-Blocker Studie

Die vorgegebene Verteilung ist $P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$. Der gEMGFU-Algorithmus wurde

Tabelle 5.1: Ergebnisse der Momentenanalyse bei der Beta-Blocker-Studie. Abbruchkriterium: Maximal 2 Mischungskomponenten zulassen.

	ML	PL	ND
Bias(E(P))	-.0524	.0163	.0199
Bias(Var(P))	-.1524	-.0111	-.0171
Var(E(P))	.0180	.0144	.0142
Var(Var(P))	.0042	.0023	.0022
MSE(E(P))	.0208	.0146	.0146
MSE(Var(P))	.0274	.0025	.0025

nach maximal 2 gefundenen Mischungskomponenten und keiner weiteren möglichen Verbesserung abgebrochen. In der Tabelle 5.1 sind der Bias, die Varianz und der MSE der beiden Momente je nach Modell abgebildet. Beim Bias schneidet das PL-Modell in beiden Momenten am besten ab. Dagegen ist das ND-Modell bei der Schätzung der Varianz geringfügig besser als der PL-Ansatz. Das ML-Modell liefert hier die schlechtesten Werte. In der Komponentenschätzung (Tabelle 5.2) zeigt sich ein ähnliches Bild wie in der Momentenanalyse. Der PL-Ansatz und das ND-Modell unterscheiden sich in der Schätzung des Relativen Risikos und der Gewichte nur unwesentlich. Erheblich schlechtere Ergebnisse sind mit dem ML-Modell zu verzeichnen. Der Wert des Bias in der ersten Komponente des Relativen Risikos ist 61-mal größer als der Wert vom PL-Ansatz. Mit einem anderen Abbruchkriterium (BIC oder NPMLE) würden im ML-Modell viel mehr Komponenten in jeder Wiederholung geschätzt werden. Da aber nur eine zweikomponentige Verteilung hier vorgegeben ist, kann auch nur eine Zweikomponentenlösung in betracht gezogen werden.

Wie kongruent die Ergebnisse von PL- und ND-Ansatz sind, verdeutlicht das Histogramm (Abbildung 5.5 und 5.6). Die Streuung um den wahren Wert des Relativen Risikos ist bei beiden Modellen wesentlich geringer als beim ML-Modell.

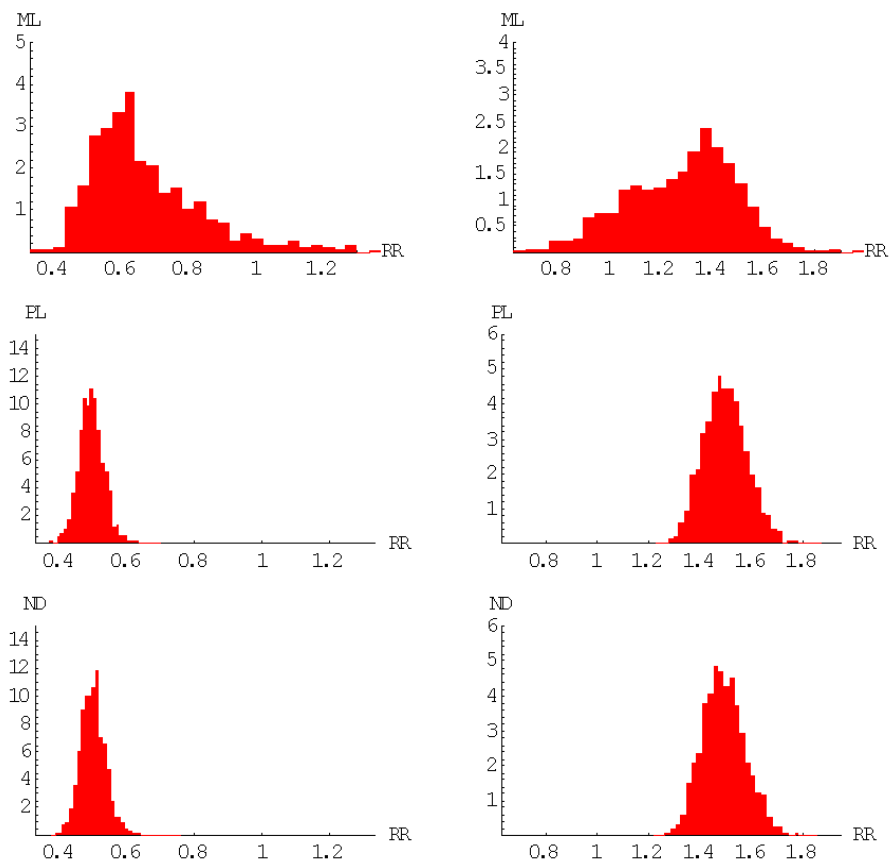


Abbildung 5.5: Histogramm des Relativen Risikos für jede Komponente (Beta-Blocker Studie).

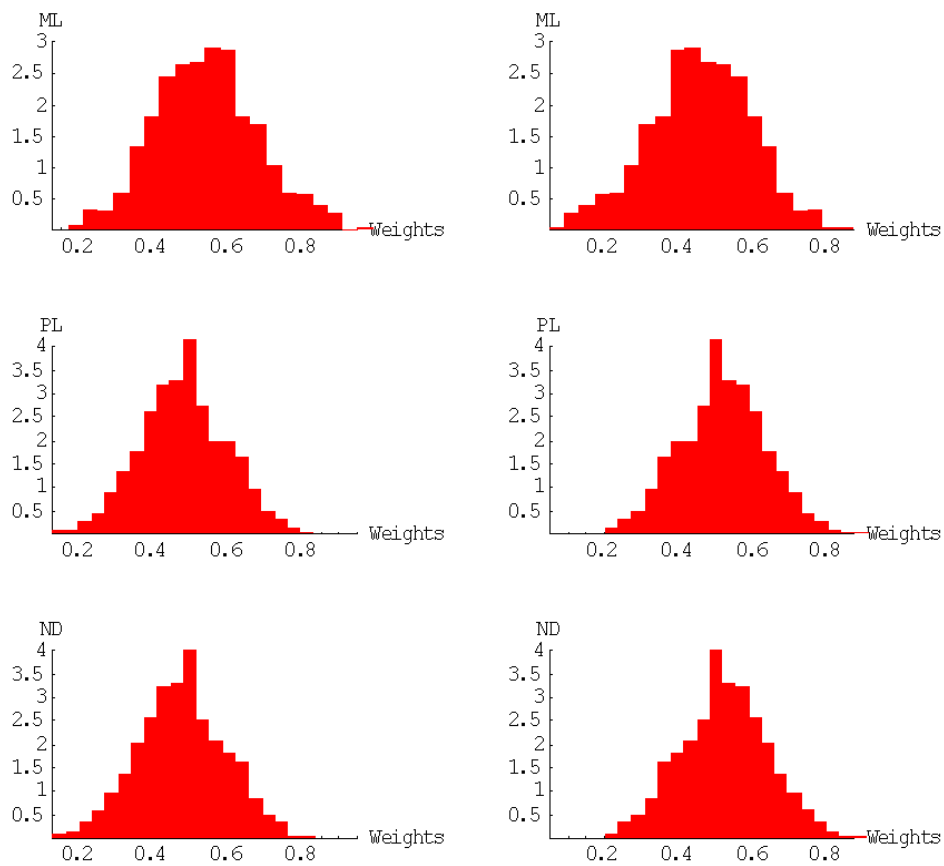


Abbildung 5.6: Histogramm von den Gewichten für jede Komponente (Beta-Blocker Studie).

Tabelle 5.2: Ergebnisse der Komponentenanalyse bei der Beta-Blocker-Studie. N gibt an wieviel Bootstrapreplikationen in die Berechnung mit einfließen. Abbruchkriterium: Maximal 2 Mischungskomponenten zulassen.

Komp	BIAS					
	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.1728	.0423	-.0028	-.0190	.0028	-.0231
2	-.2171	-.0423	-.0017	.0190	-.0078	.0231

Komp	VARIANZ					
	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.0268	.0185	.0015	.0133	.0016	.0135
2	.0436	.0185	.0075	.0133	.0072	.0135

Komp	MSE					
	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.0567	.0203	.0015	.0137	.0016	.0141
2	.0907	.0203	.0075	.0137	.0073	.0141
N	1000		1000		1000	

Als nächstes werden drei Mischungskomponenten betrachtet, mit der a priori Verteilung $P = \begin{pmatrix} 0.6 & 0.9 & 1.2 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$. Bei dieser Simulationsstudie liegen die Ergebnisse im selben Bereich wie bei der vorhergehenden Studie. Da die vorgegebene Verteilung weniger gut separiert ist, treten im Ergebnis Verwischungen auf. Die Abbildung 5.7 zeigt die Histogramme der Mischungskomponenten. In allen Modellen sind die drei Komponenten gut erkennbar. Werden aber die drei Bilder jedes Modells übereinander gelegt (Abbildung 5.8), dann überdecken sich die Dichten der Komponenten. Besonders groß ist die Überdeckung beim ML-Modell. Die einzelnen Komponenten sind kaum noch zu erkennen. Dagegen sind im PL- und im ND-Modell alle drei Mischungskomponenten klar voneinander getrennt.

5.2.2 CALGB-Studie

Bei dieser multizentrischen Studie wird die a priori Verteilung $P = \begin{pmatrix} 0.5 & 1 \\ 0.5 & 0.5 \end{pmatrix}$ angewendet.

In der Tabelle 5.3 sind die Eigenschaften der Momente widergegeben. Im ersten Moment weist erneut das PL-Modell den kleinsten Bias auf, wohingegen im zweiten Moment

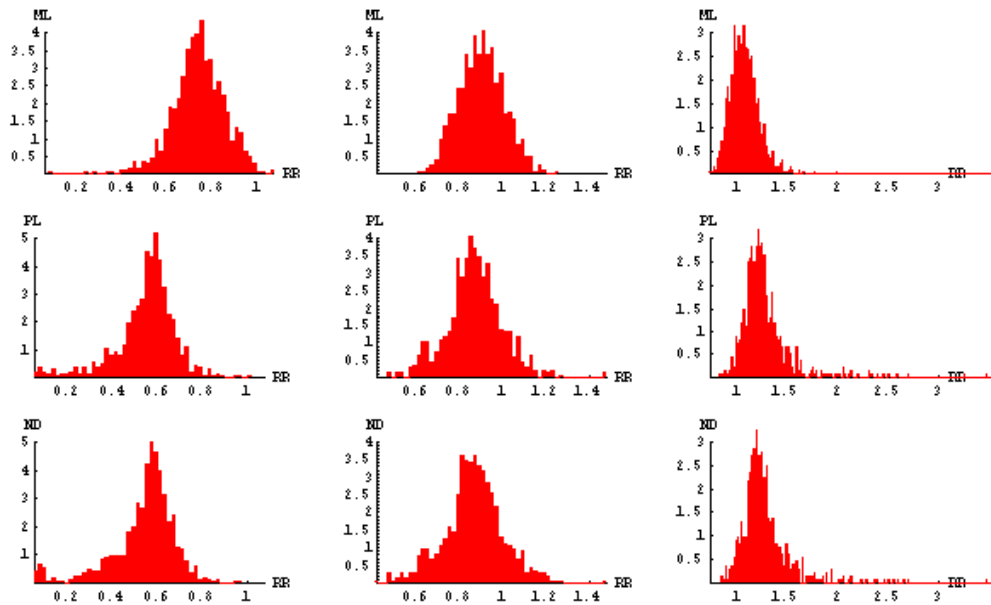


Abbildung 5.7: Histogramm des Relativen Risikos für drei Komponenten (Beta-Blocker Studie).

das ML-Modell den kleinsten Bias besitzt. In der Frage der Varianz hat der PL-Ansatz beim ersten und beim zweiten Moment den kleinsten Wert. In den meisten Replikationen wurde aber im PL-Modell Effekthomogenität geschätzt. Da eine einkomponentige Verteilung keine Varianz besitzt, sind aus diesem Grund auch die Varianzwerte so gering. In der Komponentenanalyse (Tabelle 5.4) können nur das ND-Modell und das ML-Modell miteinander verglichen werden. Es gab nur 6 Stichproben bei dem das PL-Modell zwei Mischungskomponenten geschätzt hat. Eine repräsentative Aussage kann daher nicht getroffen werden. Verglichen mit dem ML-Modell zeigt der ND-Ansatz sowohl einen größeren Bias als auch eine größere Varianz.

5.2.3 Studie über koronare Herzkrankheit

In diesem Abschnitt wird wieder die vorgegebene Verteilung $P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$ verwendet.

Die Ergebnisse sind in den Tabellen 5.5 und 5.6 dargestellt. Der Bias vom ersten Moment des ML-Modells ist 2-mal schlechter als der Wert des Bias vom PL- und fast 2.5-mal schlechter als der Wert des Bias vom ND-Ansatz. Für das zweite Moment zeichnen sich ähnliche Ergebnisse ab. Der Bias vom PL-Modell ist hier am geringsten. Beide Momente nehmen beim ML-Modell die größten Werte für den kleinsten quadratischen Fehler an.

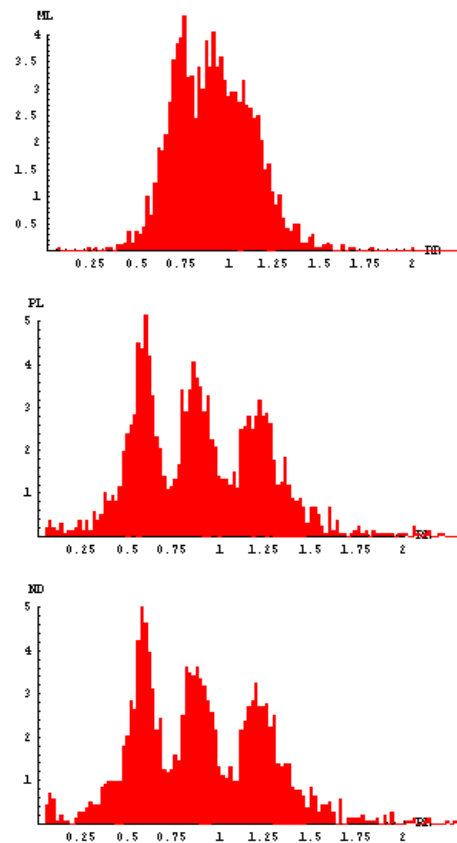


Abbildung 5.8: Histogramm mit den drei Komponenten zusammen, je nach Modell (Beta-Blocker Studie)

Tabelle 5.3: Ergebnisse der Momentenanalyse bei der CALGB-Studie. Abbruchkriterium: Maximal 2 Mischungskomponenten zulassen.

	ML	PL	ND
Bias(E(P))	.0596	.0251	.0850
Bias(Var(P))	-.0239	-.0624	.0455
Var(E(P))	.0193	.0146	.0171
Var(Var(P))	.0046	.0000	.1208
MSE(E(P))	.0229	.0152	.0244
MSE(Var(P))	.0052	.0039	.1229

Wohingegen die Momente des PL- und des ND-Modells fast die gleichen MSE-Werte besitzen. In der einzelnen Komponentenschätzung erweist sich hier das PL-Modell als das beste, besonders in der Schätzung des Relativen Risikos. Das ML-Modell schneidet wiederum in allen Komponenten am schlechtesten ab.

Im zweiten Versuch werden drei Mischungskomponenten mit der a priori Verteilung $P = \begin{pmatrix} 0.6 & 0.9 & 1.2 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$ betrachtet.

Erstaunlicherweise wird der kleinste Bias-Wert (Tabelle 5.7) beim ersten Moment durch das ML-Modell wiedergegeben. Die Varianz vom ersten Moment ist dagegen größer als bei den anderen beiden Modellen. Somit besitzt auch das ML-Modell den größten MSE-Wert. In der Komponentenanalyse (Tabelle 5.8) sind mehrere Effekte zu beobachten. Im ML-Modell wird im Mittel die erste Komponente des Relativen Risikos stark überschätzt, wohingegen die dritte Komponente stark unterschätzt wird. Diese Eigenschaft ist auch bei der Simulation mit der Beta-Blocker-Studie (Abbildung 5.8) vorhanden und erklärt die Verwischung in den einzelnen Komponenten. Bei den anderen beiden Modellen tritt dieser Effekt nicht auf. In beiden Modellen wird die erste und zweite Komponente des Relativen Risikos leicht unterschätzt, wobei das PL-Modell etwas kleinere absolute Werte im Bias aufweist. Die beiden Modelle überschätzen dagegen stark die dritte Komponente. Es gibt einige Bootstrap-Replikationen, bei denen das PL-Modell und das ND-Modell eine besonders große Komponente für das Relative Risiko schätzen. Im Anhang sind zwei Tabellen (A.1 und A.2) enthalten, bei denen eine solche Situation auftritt. Da eigentlich beide Modelle meist sehr ähnliche Resultate berechnen, können interessanterweise solche Fälle getrennt voneinander auftreten. Das heißt, schätzt das PL eine besonders große Komponente muss nicht gleichzeitig das ND-Modell eine Komponente mit hohem Relativen Risiko-Wert schätzen. In den Ergebnissen im Anhang wird dies verdeutlicht. Beim PL-Ansatz tritt in dieser Simulationsstudie bei acht Wiederholungen dieses Phänomen

Tabelle 5.4: Ergebnisse der Komponentenanalyse bei der CALGB-Studie. N gibt an wieviel Bootstrapplektionen in die Berechnung mit einfließen. Abbruchkriterium: Maximal 2 Mischungskomponenten zulassen.

BIAS						
Komp	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.1719	.0550	.1329	-.0362	-.1998	-.2643
2	.0042	-.0550	-.0708	.0362	.1805	.2643

VARIANZ						
Komp	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.0243	.0188	.0689	.0745	.0514	.0545
2	.0688	.0188	.0743	.0745	1.0450	.0545

MSE						
Komp	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.0539	.0219	.0865	.0758	.0913	.1243
2	.0688	.0219	.0793	.0758	1.0776	.1243
N	1000		6		791	

auf. Werden diese acht Bootstrap-Wiederholungen außen acht gelassen, dann verbessert sich der Wert des Bias von Relativen Risiko der dritten Komponente deutlich von 0.1091 auf 0.054552. Damit verbunden wird die Varianz der dritten Komponente von 0.541 auf 0.0421 verbessert. Das ND-Modell schätzt in 21 Bootsrap-Replikationen ein erhöhtes relatives Risiko in der dritten Komponente. Werden diese Bootstrap-Wiederholungen nicht in die Berechnung integriert, dann verbessert sich der Bias vom Relativen Risiko der dritten Mischungskomponente von 0.1988 auf 0.0746. Die Varianz verkleinert sich von 0.815 auf 0.0614. Mit dieser bereinigten Kalkulation verbessert sich die Schätzgüte der beiden Modelle, wobei in dem Fall der PL-Ansatz geringfügig bessere Werte in Bias und Varianz gegenüber dem ND-Ansatz vorweist.

Tabelle 5.5: Ergebnisse der Momentenanalyse bei der Studie über koronare Herzkrankheit mit einer zweikomponentigen a priori Verteilung. Abbruchkriterium: Maximal 2 Mischungskomponenten zulassen.

	ML	PL	ND
Bias(E(P))	-.0398	.0190	.0160
Bias(Var(P))	-.2121	-.0117	-.0145
Var(E(P))	.0341	.0108	.0107
Var(Var(P))	.0015	.0013	.0013
MSE(E(P))	.0357	.0112	.0110
MSE(Var(P))	.0465	.0014	.0015

Tabelle 5.6: Ergebnisse der Komponentenanalyse bei der Studie über koronare Herzkrankheit. N gibt an wieviel Bootstrapreplikationen in die Berechnung mit einfließen. Abbruchkriterium: Maximal 2 Mischungskomponenten zulassen.

BIAS						
Komp	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.2685	-.0357	.0022	-.0194	.0041	-.0175
2	-.3787	.0357	-.0027	.0194	-.0067	.0175

VARIANZ						
Komp	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.0394	.0321	.0007	.0101	.0007	.0102
2	.0436	.0321	.0040	.0101	.0039	.0102

MSE						
Komp	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.1115	.0334	.0007	.0105	.0008	.0105
2	.1871	.0334	.0040	.0105	.0039	.0105
N	1000		1000		1000	

Tabelle 5.7: Ergebnisse der Momentenanalyse bei der Studie über koronare Herzkrankheit mit einer dreikomponentigen a priori Verteilung. Abbruchkriterium: Maximal 3 Mischungskomponenten zulassen.

	ML	PL	ND
Bias(E(P))	-.0085	.0140	.0157
Bias(Var(P))	-.0398	.0110	.0243
Var(E(P))	.0071	.0039	.0048
Var(Var(P))	.0001	.0242	.0447
MSE(E(P))	.0072	.0041	.0050
MSE(Var(P))	.0017	.0243	.0453

Tabelle 5.8: Ergebnisse der Komponentenanalyse bei der Studie über koronare Herzkrankheit. N gibt an wieviel Bootstrapreplikationen in die Berechnung mit einfließen. Abbruchkriterium: Maximal 3 Mischungskomponenten zulassen.

Komp	BIAS					
	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.1545	.0224	-.0198	-.0288	-.0291	-.0338
2	-.0072	-.0579	-.0128	.0095	-.0131	.0247
3	-.1773	.0355	.1091	.0193	.1988	.0091

Komp	VARIANZ					
	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.0117	.0123	.0100	.0198	.0148	.0233
2	.0104	.0199	.0099	.0263	.0127	.0306
3	.0118	.0216	.5407	.0279	.8150	.0329

Komp	MSE					
	ML		PL		ND	
	RR	q	RR	q	RR	q
1	.0356	.0128	.0104	.0206	.0157	.0244
2	.0104	.0233	.0101	.0264	.0129	.0312
3	.0432	.0229	.5526	.0282	.8545	.0330
N	1000		911		920	

5.2.4 Eigenständige Simulationsstrategien

In diesem Abschnitt werden unabhängig von einer Studie bestimmte Szenarien aus der Praxis in Simulationsstudien abgebildet. Dazu wird in jeder Simulationsstudie die Anzahl der Zentren auf 100 festgelegt. Die a priori Verteilung ist $P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$. Die Beobachtungen im Behandlungs- und Kontrollarm werden erst als binomialverteilt und danach als Poissonverteilt aufgefasst. Damit verbunden wird je nach Verteilungsart im ND- und im ML-Modell der Binomiallikelihood oder der Poissonlikelihood zur Schätzung des Effekts benutzt. Der Baselineparameter p^C ist bei dieser Simulation ein wichtiger Parameter, der dementsprechend modelliert werden muss. Bei den meisten hier vorgestellten Studien schwankt der Baseline im Intervall von 0.01 bis 0.3. Die Baselineheterogenität ist abhängig von der Anzahl der unter Risiko stehenden. So ist die Differenz im Baseline bei einigen Studien mit einer hohen Anzahl an Probanden sehr gering. Bei der Beta-Blocker Studie (Tabelle 1.1) nehmen durchschnittlich 448 Probanden je Zentrum im Kontrollarm teil. Das Baselineisiko liegt bei dieser Studie im Intervall von 0.02 bis 0.2. Ein anderes extremes Beispiel ist die CALGB-Studie (Tabelle 1.4) an der durchschnittlich nur 3.4 Probanden im Kontrollarm teilnehmen. Auf Grund des geringen Stichprobenumfanges ist bei dieser Studie die Variabilität im Baseline größer. Das Baselineisiko verteilt sich dort im Intervall von 0 bis 1. Aber es gibt auch Studien mit größeren Stichprobenumfängen und einer breiten Streuung des Baselines. Dazu gehört die Studie über die Infektion der Atemwege (Tabelle 1.2) mit durchschnittlich moderaten 88 Teilnehmern je Zentrum im Kontrollarm, wo der Baselineparameter im Intervall von 0.08 bis 1 zu finden ist. In den weiteren Abschnitten werden diese drei Konstellationen untersucht.

Seltenheitsfall mit binomialverteilten Beobachtungen

Kommt eine zu untersuchende Erkrankung sehr selten vor, so spricht man vom Seltenheitsfall (sparsity case). Hier ist es besonders sinnvoll eine Studie an mehreren Zentren durchzuführen, um so präzisere Aussagen treffen zu können. Im folgenden Versuch werden die einzelnen Baselineparameter für jedes Zentrum gleichverteilt aus dem Intervall 0.03 bis 0.66 gezogen. Die obere Intervallgrenze von 0.66 ergibt sich aus der vorgegebenen Verteilung. Würde sie überschritten, dann könnte der Wert für das Risiko im Behandlungsarm größer als eins werden. λ wird auf 15 gesetzt, um so möglichst wenig Probanden je Zentrum zu generieren. In der Tabelle 5.9 sind die Ergebnisse der Simulation dargestellt. Die erste Komponente wird sehr gut vom PL- und ND-Ansatz geschätzt. Beide Werte des Bias sind negativ. Somit wird der wahre Wert der ersten Komponente unterschätzt. Das

Tabelle 5.9: Simulationsstudie im Seltenheitsfall (N gibt die Anzahl der berücksichtigbaren Replikationen an)

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim \text{Bin}(p, n), n \sim \text{Poi}(15), p^C \sim \text{Uni}(0.03; 0.66)$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.723	[0.560; 0.886]	.2231	1.358	[1.018; 1.699]	-.1416	1000
PL	0.460	[0.275; 0.645]	-.0399	1.249	[0.618; 1.88]	-.2511	988
ND	0.467	[0.276; 0.659]	-.0326	1.394	[0.077; 2.71]	-.1065	1000

Tabelle 5.10: Simulationsstudie im Normalfall (N gibt die Anzahl der berücksichtigbaren Replikationen an)

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim \text{Bin}(p, n), n \sim \text{Poi}(100), p^C \sim \text{Uni}(0.03; 0.2)$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.656	[0.516; 0.796]	.1561	1.416	[1.150; 1.681]	-.0845	1000
PL	0.470	[0.379; 0.561]	-.0299	1.434	[1.254; 1.613]	-.0663	1000
ND	0.493	[0.380; 0.607]	-.0069	1.421	[1.241; 1.600]	-.0792	1000

ML-Modell überschätzt die erste Komponente so deutlich, dass der vorgegebene Wert nicht mehr im Konfidenzintervall enthalten ist. Die zweite Komponente wird von allen drei Modellen unterschätzt. Den kleinsten Bias hat das ND-Modell. Doch ist bei diesem Modell das Konfidenzintervall für die zweite Komponente am größten.

Normalfall mit binomialverteilten Beobachtungen

Anders als beim Fall seltener Ereignisse kommt hier die Erkrankung häufiger vor. So ist die Anzahl der Probanden weitaus höher als im Seltenheitsfall. Der Parameter λ wird hier auf 100 gesetzt. Zunächst wird ein Baselineisiko gleichverteilt aus dem Intervall 0.03 bis 0.2 betrachtet. Diese Simulationsstudie spiegelt den größten Teil der hier vorgestellten multizentrischen Studien wider. Die Ergebnisse (Tabelle 5.10) zeigen, dass der Wert des Bias sich in allen drei Modellen verringert hat. Nach dem Gesetz der großen Zahlen war dies auch zu erwarten. Das ML-Modell zeigt auch in dieser Studie das gleiche Schätzverhalten beim Relativen Risiko. Die erste Komponente wird überschätzt und die zweite unterschätzt, obwohl die Baselineheterogenität bei diesem Simulationsversuch bewusst verringert worden ist. Den kleinsten Wert des Bias nimmt das ND-Modell für die erste Komponente an. Genau wie im Seltenheitsfall wird die zweite Komponente von allen drei

Tabelle 5.11: Simulationsstudie im Normalfall mit erhöhter Baselineheterogenität (N gibt die Anzahl der berücksichtigbaren Replikationen an)

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim Bin(p, n), n \sim Poi(100), p^C \sim Uni(0.03; 0.66)$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.742	[0.585; 0.899]	.2421	1.278	[0.855; 1.700]	-.222	1000
PL	0.496	[0.460; 0.532]	-.0044	1.491	[1.426; 1.557]	-.009	1000
ND	0.505	[0.469; 0.541]	.0050	1.483	[1.423; 1.544]	-.017	1000

Tabelle 5.12: Simulationsstudie im Normalfall mit festem Baseline (N gibt die Anzahl der berücksichtigbaren Replikationen an)

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim Bin(p, n), n \sim Poi(100), p^C = 0.03$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.500	[0.461; 0.539]	.0004	1.502	[1.424; 1.579]	.0017	1000
PL	0.491	[0.452; 0.529]	-.0091	1.481	[1.398; 1.563]	-.0192	1000
ND	0.502	[0.463; 0.540]	.0015	1.477	[1.398; 1.555]	-.0233	1000

Modellen unterschätzt, wobei hier der PL-Ansatz den kleinsten Wert des Bias für das Relative Risiko liefert.

In zwei Studien (Tabelle 1.2 und 1.7) ist die Baselineheterogenität ausgeprägter als in den restlichen Studien. Um auch diesen Fall abzudecken wird eine weitere Simulationsstudie durchgeführt, bei der der Baselineparameter aus einer Gleichverteilung mit dem Intervall $[0.03; 0.66]$ gezogen wird. Der Parameter λ ist wie in der vorangegangenen Simulation gleich 100. Deutlich gestiegen ist der Wert des Bias für beide Komponenten im ML-Modell (Tabelle 5.11). Bei den anderen beiden Modellen sinkt erstaunlicherweise der Baselinewert wieder in beiden Komponenten. Der PL-Ansatz besitzt hier in beiden Komponenten den kleinsten Bias, aber die Differenz zum ND-Ansatz ist gering. Weiterhin ist hervorzuheben, dass mit dem ML-Modell der vorgegebene Wert der ersten Komponente in den letzten drei Simulationsstudien nicht im Konfidenzintervall erfasst wurde. Um in der nächsten Simulation die Baselineheterogenität auszuschließen, wird der Baselineparameter auf 0.3 festgesetzt. Obwohl nicht praxisnah sinkt damit der Einfluss der Baselineklassifizierung des ML-Modells. Da die Effektschätzung hier eine zentrale Rolle spielt, ist die Festsetzung des Baselines durchaus zu rechtfertigen und erhöht die Vergleichbarkeit der drei Modelle. Das ML-Modell (Tabelle 5.12) schätzt unter diesen Simulationsbedingungen am besten

Tabelle 5.13: Simulationsstudie im Seltenheitsfall unter Poissonverteilten Expositionen (N gibt die Anzahl der berücksichtigbaren Replikationen an)

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim Poi(p \times n), n \sim Poi(15), p^C \sim Uni(0.03; 0.66)$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.666	[0.444; 0.889]	.1665	1.160	[0.876; 1.443]	-.3402	1000
PL	0.456	[0.231; 0.680]	-.0441	1.392	[0.520; 2.264]	-.1078	999
ND	0.403	[0.0677; 0.738]	-.0971	1.366	[-0.689; 3.421]	-.1339	999

die vorgegebene Verteilung. Die erste Komponente wird von allen drei Ansätzen recht gut wiedergegeben. Das PL-Modell hat mit -0.091 den größten absoluten Wert des Bias. Die Varianz des Erwartungswertes für die erste Komponente ist in allen drei Modellen gleich 0.004 . Im Vergleich zu den anderen Simulationsstudien ist dies eine sehr kleine Varianz. Die zweite Komponente wird vom ND- und vom PL-Ansatz schlechter reproduziert als vom ML-Modell. Die Varianz des Erwartungswertes der zweiten Komponente ist mit 0.0016 (0.0018 beim PL-) nur geringfügig größer als bei der ersten Komponente. Hier wird sehr deutlich, wie stark der Baselineparameter im ML-Modell die Schätzung des Behandlungseffektes beeinflusst.

Seltenheitsfall mit Poissonverteilten Beobachtungen

In 4 von den 13 hier vorgestellten multizentrischen Studien werden Personenzeiten betrachtet. Um auch diesen Fall entsprechend zu untersuchen, werden die Beobachtungen in den folgenden Simulationsstudien als Poissonverteilt aufgefasst. Der Seltenheitsfall kommt praktisch bei Personenzeiten nicht vor, da man jederzeit auf eine kleinere Einheit (Jahre, Monate oder Tage) zurückgreifen kann. Zur Vollständigkeit wird auch dieser Fall unter Poissonverteilten Expositionen durchgeführt. Für die Simulationsstudie werden sonst alle Einstellungen aus dem Seltenheitsfall mit binomialverteilten Expositionen übernommen. Die Unterschiede in den Ergebnissen (5.13) fallen relativ gering aus. Ist in einer Komponente der Wert des Bias kleiner als in der Simulation mit binomialverteilten Expositionen, gleicht sich das durch eine erhöhte Varianz im Erwartungswert wieder aus. Dennoch ist in dieser speziellen Situation der PL-Ansatz das adäquate Modell mit dem kleinsten Bias. Die erste Komponente wird in allen drei Modellen besser reproduziert als die zweite Komponente.

Normalfall mit Poissonverteilten Beobachtungen

Die Einstellungen für die Simulation werden vom Normalfall (mit binomialverteilten Beobachtungen) übernommen. Alle Simulationsergebnisse sind in Tabelle 5.14 untergebracht. Zusätzlich wurde eine Simulation mit durchschnittlich 3000 Probanden je Arm hinzugefügt. Dieser Versuch deckt drei multizentrischen Studien (Tabelle 1.8, 1.9 und 1.10) ab, denn bei diesen Studien ist die Anzahl der Personenjahre relativ hoch. Es liegt die Vermutung nahe, dass der PL-Ansatz mit Poissonverteilten Beobachtungen das Relative Risiko in jeder Komponente exakter schätzt als mit binomialverteilten Beobachtungen. Diese Vermutung wird durch die Simulation aber nicht bestätigt. Bei der Simulationsstudie mit einem gleichverteilten Baseline im Intervall von 0.03 bis 0.66 tritt sogar das Gegenteil auf. In dem Fall (Tabelle 5.11) ist der Wert des Bias in jeder Komponente etwas geringer als mit Poissonverteilten Zufallszahlen. Bei der Studie mit durchschnittlich 3000 Probanden pro Arm zeigt der PL-Ansatz ein sehr gutes Schätzverhalten. Einen geringfügig größeren Wert des Bias pro Komponente hat das ND-Modell. Das ML-Modell liefert hier verhältnismäßig große Werte für den Bias in jeder Komponente. Sehr interessant sind hier die Stabilitätskurven des Erwartungswertes in Abhängigkeit von der Wiederholungsanzahl in der Simulation vom PL- und vom ND-Ansatz. Beide Graphen in der Abbildung 5.9

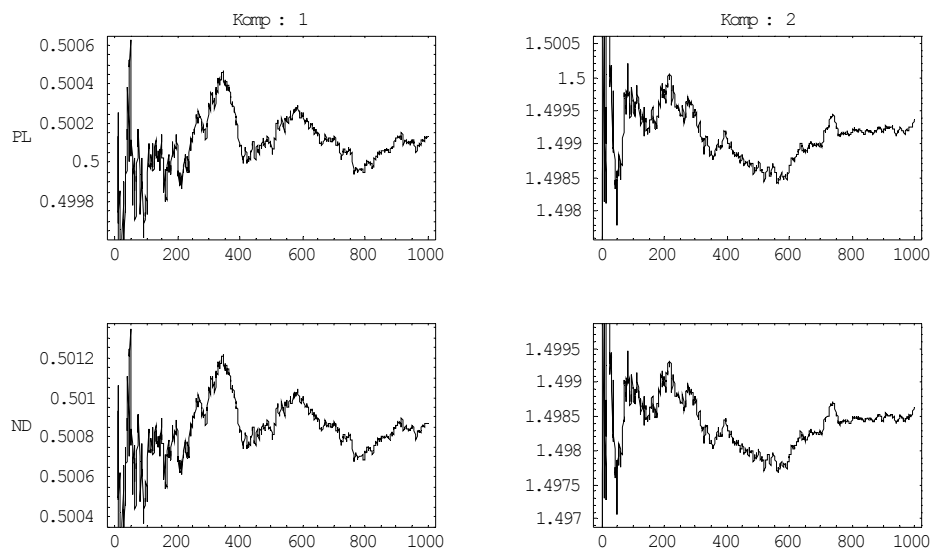


Abbildung 5.9: Stabilität der Simulation im Normalfall (Tabelle 5.14, $n \sim Poi(3000)$). Abgebildet ist der Erwartungswert der jeweiligen Komponente für das Relative Risiko in Abhängigkeit der Simulationswiederholungen.

Tabelle 5.14: Simulationsstudie im Normalfall unter Poissonverteilten Expositionen (N gibt die Anzahl der berücksichtigbaren Replikationen an)

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim Poi(p \times n), n \sim Poi(100), p^C \sim Uni(0.03; 0.2)$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.672	[0.519; 0.824]	.1720	1.271	[0.965; 1.578]	-.2285	1000
PL	0.473	[0.359; 0.587]	-.0265	1.456	[1.240; 1.671]	-.0445	1000
ND	0.496	[0.358; 0.634]	-.0041	1.421	[1.002; 1.839]	-.0795	1000

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim Poi(p \times n), n \sim Poi(100), p^C \sim Uni(0.03; 0.66)$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.726	[0.577; 0.874]	.2257	1.112	[0.903; 1.322]	-.3879	1000
PL	0.494	[0.448; 0.539]	-.0062	1.486	[1.390; 1.582]	-.0138	1000
ND	0.500	[0.439; 0.562]	.0002	1.476	[1.373; 1.579]	-.0240	1000

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim Poi(p \times n), n \sim Poi(100), p^C = 0.3$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.500	[0.455; 0.545]	-.0004	1.50146	[1.403; 1.600]	.0015	1000
PL	0.487	[0.440; 0.535]	-.0127	1.48096	[1.373; 1.589]	-.0190	1000
ND	0.496	[0.448; 0.543]	-.0044	1.47155	[1.366; 1.576]	-.0284	1000

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim Poi(p \times n), n \sim Poi(3000), p^C \sim Uni(0.03; 0.2)$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.646	[0.569; 0.722]	.1455	1.460	[1.299; 1.622]	-.0399	1000
PL	0.500	[0.487; 0.514]	.0001	1.499	[1.471; 1.528]	-.0006	1000
ND	0.501	[0.487; 0.514]	.0009	1.499	[1.470; 1.528]	-.0014	1000

Tabelle 5.15: Vergleich zweier a priori Verteilungen (N gibt die Anzahl der berücksichtigbaren Replikationen an)

$$P = \begin{pmatrix} 0.8 & 1.2 \\ 0.5 & 0.5 \end{pmatrix}, x \sim \text{Bin}(p, n), n \sim \text{Poi}(30), p^C \sim \text{Uni}(0.4; 0.6)$$

Komp	1			2			N
	$E(\theta)$	$\text{KI}_{\alpha=0.05}$	Bias	$E(\theta)$	$\text{KI}_{\alpha=0.05}$	Bias	
ML	0.858	[0.743;0.972]	.0577	1.173	[1.042;1.303]	-.0272	1000
PL	0.583	[0;1.167]	-.2174	1.297	[0;3.3395]	.0967	109
ND	0.793	[0.599;0.987]	-.0069	1.174	[0.903;1.446]	-.0261	993

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}, x \sim \text{Bin}(p, n), n \sim \text{Poi}(30), p^C \sim \text{Uni}(0.4; 0.6)$$

Komp	1			2			N
	$E(\theta)$	$\text{KI}_{\alpha=0.05}$	Bias	$E(\theta)$	$\text{KI}_{\alpha=0.05}$	Bias	
ML	0.519	[0.460;0.577]	.0185	1.517	[1.427;1.607]	.0172	1000
PL	0.473	[0.422;0.524]	-.0269	1.415	[1.314;1.516]	-.0852	1000
ND	0.495	[0.414;0.576]	-.0046	1.430	[1.335;1.525]	-.0699	1000

zeigen das gleiche Muster je Komponente. Es scheint, als wären die Graphen nur parallel verschoben. In beiden Komponenten ist der Graph des ND-Ansatzes näher am Nulleffekt als der Graph vom PL-Modell. Dieses Beispiel zeigt, wie ähnlich die beiden Verfahren den Behandlungseffekt schätzen und dass die Schätzung des PL-Ansatzes etwas näher am wahren Wert liegt.

Auswirkungen weniger gut separierter Verteilungen

In diesem Abschnitt wird untersucht, welchen Einfluss eine gut oder weniger gut separierte a priori Verteilung auf die Schätzung der Heterogenität hat. Dazu werden zwei diskrete Verteilungen mit

$$P_1 = \begin{pmatrix} 0.8 & 1.2 \\ 0.5 & 0.5 \end{pmatrix} \text{ und}$$

$$P_2 = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix},$$

miteinander verglichen. P_2 ist eine sehr gut separierte Verteilung, wohingegen bei P_1 der Abstand zwischen den Komponenten θ_1 und θ_2 etwas geringer ist. Wie gut die einzelnen Komponenten je nach Modell geschätzt werden, zeigt die Tabelle 5.15. Die größten Unterschiede treten beim PL-Modell auf. Bei der Verteilung P_1 werden nur 109 von 1000

Tabelle 5.16: Simulationsstudie mit Poissonverteilten Beobachtungen (N gibt die Anzahl der berücksichtigbaren Replikationen an)

$$P = \begin{pmatrix} 0.8 & 1.2 \\ 0.5 & 0.5 \end{pmatrix}, x \sim Poi(p \times n), n \sim Poi(30), p^C \sim Uni(0.4; 0.6)$$

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	Bias	$E(\theta)$	$KI_{\alpha=0.05}$	Bias	
ML	0.831	[0.630;1.032]	.0312	1.172	[0.955;1.389]	-.0282	1000
PL	0.729	[0.361;1.098]	-.0709	1.340	[0;2.952]	.1404	979
ND	0.763	[0.407;1.120]	-.0365	1.242	[0.523;1.962]	.0423	888

Replikationen mit zwei Mischungskomponenten gefunden. Den restlichen 891 Wiederholungen wurde Homogenität im Effekt unterstellt. Dieses Beispiel zeigt, dass der PL-Ansatz Schwierigkeiten bei der Schätzung der Komponentenanzahl hat, wenn die Teilpopulationen einen ähnlichen Effektparameter besitzen. Ist dagegen der Effekt jeder Mischungskomponente deutlich verschieden, so schätzt das PL-Modell sehr gut alle in Wirklichkeit vorkommenden Teilpopulationen. Der Grund könnte darin liegen, dass der Effekt auch bei binomialverteilten Beobachtungen über einen Poissonlikelihood formuliert wird. Sind die Beobachtungen binomialverteilt, dann müsste eigentlich der Effekt auch mit einer Binomial-Likelihoodfunktion modelliert werden. Bei Poissonverteilten Beobachtungen schätzt das PL-Modell die Anzahl der Mischungskomponenten genauso gut wie die beiden anderen Modelle (siehe Tabelle 5.16). Unübersehbar ist die qualitativ hochwertige Effektschätzung des ML-Ansatzes in allen drei Simulationsstudien mit dem stets kleinsten Konfidenzintervall.

5.3 Schätzung der Komponentenzahl

Häufig tritt bei Schätzung der Heterogenität die zentrale Frage auf, wie viele Subpopulationen in der Gesamtpopulation enthalten sind. Das NPMLE- und das BIC-Kriterium sind zur Schätzung der Komponentenanzahl in dieser Arbeit vorgestellt worden. Das Ziel ist nun, herauszufinden, wie diese beiden Kriterien bei einer vorgegebenen Verteilung die Anzahl der Mischungskomponenten schätzen. Zuerst wird eine Bootstrap-Simulation durchgeführt, indem eine multizentrische Studie als Grundlage der Simulation dient. Insgesamt werden 200 Bootstrapstichproben gezogen, von denen je nach Modell und Abbruchkriterium die Anzahl der Komponenten geschätzt wird.

Tabelle 5.17: Schätzung der Komponentenanzahl anhand der zwei Abbruchkriterien bei 200 Bootstrapstichproben. (Als Grundlage dient die Beta-Blocker-Studie, $\theta = (0.5, 1.5)$ mit den Gewichten $q = (0.5, 0.5)$)

Komp	BIC			NPMLE		
	ML	PL	ND	ML	PL	ND
1	0	0	0	0	0	0
2	0	200	198	0	78	69
3	0		2	0	85	84
4	2			0	33	44
5	36			0	4	3
6	72			0		
7	69			0		
8	20			3		
9	1			14		
10				35		
11				47		
12				48		
13				30		
14				17		
15				6		

5.3.1 Beta-Blocker Studie

Die vorgegebene Verteilung ist $P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$. Das Ergebnis in Tabelle 5.17 gibt recht deutlich die beim ML-Modell auftretende bekannte Überschätzung der Komponentenanzahl wieder. Besonders beim NPMLE-Kriterium werden in allen Bootstrapstudien mindestens 8 Mischungskomponenten geschätzt. In 6 Fällen werden sogar 15 Mischungskomponenten gefunden, was bei einer Anzahl von 22 Zentren recht viel ist. Ausgewogen schätzen das ND-Modell und das PL-Modell die Anzahl der Untergruppen. Beide Modelle finden nach dem NPMLE-Kriterium am häufigsten eine dreikomponentige Mischverteilung. Das BIC-Kriterium stellt sich hier als das beste Kriterium für den PL- und ND-Ansatz heraus. In allen 200 Bootstrap-Wiederholungen wurden beim PL-Modell 2 Mischungskomponenten gefunden. Das ND-Modell schätzt nur 2 Stichproben mit drei Komponenten. Den restlichen Studien werden zwei Komponenten zugeordnet. Das ML-Modell überschätzt die Komponentenanzahl mit dem BIC-Kriterium weniger ausgeprägt. Eine 6-Komponentenlösung wurde mit 72-mal am häufigsten gefunden.

Im nächsten Versuch wird eine dreidimensionale Verteilung vorgegeben, mit $P = \begin{pmatrix} 0.6 & 0.9 & 1.2 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$. Die Ergebnisse in Tabelle 5.18 zeigen eine ähnliche Verteilung der

Tabelle 5.18: Schätzung der Komponentenanzahl anhand der zwei Abbruchkriterien bei 200 Bootstrapsstichproben. (Als Grundlage dient die Beta-Blocker-Studie, $\theta = (0.6, 0.9, 1.2)$ mit den Gewichten $q = (0.3, 0.4, 0.3)$)

Komp	BIC			NPMLE		
	ML	PL	ND	ML	PL	ND
1	0	44	37	0	1	0
2	0	153	155	0	60	55
3	1	3	8	0	115	111
4	38			0	23	32
5	85			0	1	2
6	61			0		
7	14			3		
8	1			16		
9	1			30		
10				52		
11				50		
12				30		
13				15		
14				3		
15				1		

Komponentenanzahl wie bei der vorhergehenden Simulation 5.17. Die Komponentenanzahl wird im ML-Modell mit beiden Kriterien überschätzt. Am häufigsten werden nach dem NPMLE-Kriterium 10 Mischungskomponenten und nach dem BIC-Kriterium 5 Mischungskomponenten gefunden. Die beiden anderen Modelle schätzen nach dem NPMLE-Kriterium die drei vorgegebenen Komponenten in den meisten Bootstrap-Wiederholungen. Das BIC-Kriterium unterschätzt die vorgegebene Anzahl. Hauptsächlich werden zwei Mischungskomponenten gefunden. Obwohl beide Kriterien ähnliche Ergebnisse liefern, zeigt sich hier besonders bei einer weniger gut separierten Verteilung, dass das NPMLE-Kriterium genauer die vorgegebene Komponentenanzahl wieder findet. Dies gilt aber nur für das PL- und das ND-Modell.

5.3.2 Eigenständige Simulationsstudie

In diesem Abschnitt wird geprüft, wie gut die Abbruchkriterien bei einem fest vorgegebenen Baseline die Anzahl der Mischungskomponenten schätzen. Damit wird der Frage nachgegangen, ob sich die Überschätzung der Komponentenanzahl im ML-Modell verrin-

Tabelle 5.19: Schätzung der Komponentenanzahl anhand der zwei Abbruchkriterien bei 200 Bootstrapstichproben. (Simulation mit $n_i^T, n_i^C \sim Poi(30)$, $k = 50$, $p_i^C = 0.3$)

Komp	BIC			NPMLE		
	ML	PL	ND	ML	PL	ND
1	0	34	23	0	0	0
2	190	166	173	5	159	105
3	10		4	19	37	79
4				64	4	16
5				61	0	0
6				33		
7				15		
8				2		
9				1		

gert. Folgende Definitionen werden festgelegt:

$$\begin{aligned}
 P &= \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix} \\
 \lambda &= 30 \\
 k &= 50 \text{ (Anzahl der Zentren)} \\
 p^C &= 0.3
 \end{aligned}$$

Aufgrund des sehr hohen Rechenaufwandes wird hier die Anzahl der Zentren auf 50 reduziert. 200 Stichproben werden gezogen und mit dem gEMGFU für jedes Modell berechnet. In der Tabelle 5.19 sind die Ergebnisse der Simulation zusammengetragen. Das BIC-Kriterium schätzt in allen drei Modellen am besten die Anzahl der Mischungskomponenten. Durch die Minimierung der Baselineheterogenität verbessert sich die Schätzung der Komponentenanzahl im ML-Modell deutlich. Dagegen wird mit dem NPMLE-Kriterium die vorgegebene Anzahl der Komponenten weiterhin überschätzt. Die beiden anderen Modelle schätzen sehr gut die entsprechende Komponentenanzahl, wobei mit dem BIC-Kriterium bei einigen wenigen Studien die Komponentenanzahl unterschätzt und mit dem NPMLE-Kriterium die Komponentenanzahl bei einigen Studien überschätzt wird.

In der Praxis ist das NPMLE-Kriterium beim ML-Modell nicht zu empfehlen, weil es zu viele Komponenten findet und damit das so genannte „Rauschen“ modelliert. Im PL- und im ND-Ansatz arbeiten beide Abbruchkriterien sehr gut und liefern adäquate Ergebnisse. In der Theorie ist das NPMLE-Kriterium vorzuziehen, weil die Schätzer aus dem globalen Maximum der Likelihoodfunktion abgeleitet werden. In der Praxis sind beide

Kriterien relevant. Der Anwender muss je nach Situation entscheiden, welches Kriterium sinnvoller zu gebrauchen ist.

5.4 Stabilität der Simulationsstudien

Ziel der Simulationsstudien ist es, die Schätzung eines interessanten Parameters je nach Modell zu bewerten. Unter der Homogenitätsannahme im Abschnitt 5.1 wurde die Simulation mit 10.000 Wiederholungen durchgeführt. Da es nur einen Parameter je nach Modell zu schätzen galt, wirkt sich die recht hohe Anzahl der Wiederholungen sehr günstig auf die Stabilität der Bias-, der Varianz- und der MSE-Schätzung aus. Wird Heterogenität im Effekt zugelassen, dann sind bei einer m -komponentigen Mischverteilung $2m - 1$ Parameter zu schätzen. Die Schätzung ist deutlich aufwendiger und enthält eine größere Variabilität in der Parameterschätzung. Hinzu kommen Einschränkungen, die in der Auswahl der Abbruchkriterien im Algorithmus entstehen. In der Abbildung 5.10 ist der Erwartungswert vom Relativen Risiko in Abhängigkeit der Simulationswiederholungen abgetragen. Nach ca. 600 Wiederholungen tritt im Erwartungswert eine gewisse Stabilität ein. Als einzige negativ zu bewertende Ausnahme scheint die zweite Komponente beim ML-Modell auch nach 1.000 Wiederholungen kein stabiles Niveau gefunden zu haben. Die Variabilität ist jedoch relativ gering, so dass die 1000 Wiederholungen durchaus akzeptabel erscheinen. Die Variabilität in der Schätzung ist im Seltenheitsfall größer als im Normalfall. Aus diesem Grund ist für den Seltenheitsfall ein Simulationsexperiment mit 5000 Wiederholungen durchgeführt worden. Die Ergebnisse unterscheiden sich aber nicht substantiell von dem Versuch mit 1000 Wiederholungen. Für weitergehende Untersuchungen besonders im Seltenheitsfall ist es jedoch empfehlenswert, die Wiederholungsanzahl in der Simulation zu erhöhen.

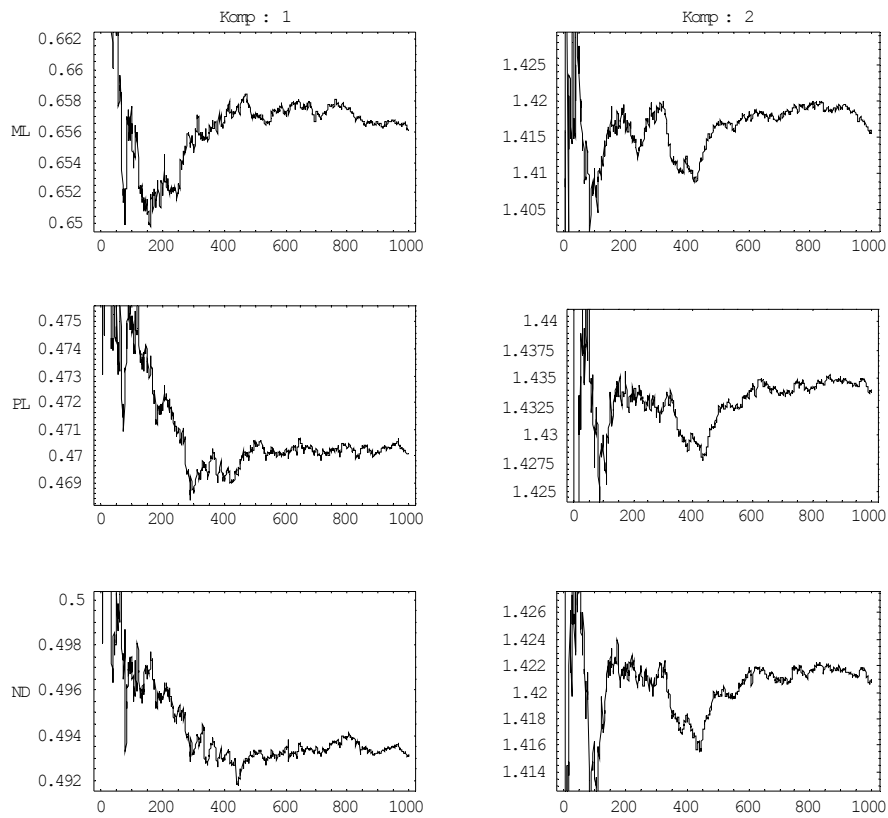


Abbildung 5.10: Stabilität der Simulation im Normalfall (Tabelle 5.10). Abgebildet ist der Erwartungswert der jeweiligen Komponente für das Relative Risiko in Abhängigkeit der Simulationswiederholungen.

Kapitel 6

Diskussion

Die hier vorgestellten Modelle dienen zur Bestimmung des Behandlungseffekts in multizentrischen Studien. Alle drei Ansätze besitzen einen komplett verschiedenen Aufbau, die aber alle den interessanten Parameter das Relative Risiko schätzen. Ein wichtiger Unterschied besteht in der Modellierung des Baselines. Im ND-Ansatz steckt die Baselineheterogenität in den einzelnen logarithmischen Relativen Risiken. Im PL-Ansatz hat jedes Zentrum seinen eigenen Baselineparameter, der hier als Nuisanceparameter betrachtet wird. Im ML-Modell gibt es zu jeder Mischungskomponente einen speziellen Baselineparameter. Die Ergebnisse der einzelnen multizentrischen Studien zeigen, dass im ML-Modell der Einfluss des Baselines auf die Schätzung der mischenden Verteilung am stärksten ist. Aus der Abbildung 4.3 geht deutlich hervor, wie sich die Zentrencluster nach dem Baseline orientieren.

In den Ergebnissen der Simulation unter Effekthomogenität stellt sich das PL-Modell als ein sehr adäquates Schätzverfahren für das Relative Risiko dar. In den statistischen Kenngrößen wie Bias, Varianz und kleinster quadratischer Fehler zeigt dieses Verfahren sehr oft kleinere Werte als die anderen beiden Modelle. Die kleinen absoluten Werte für den Bias im PL-Modell liegen in dem Zusammenhang zwischen dem Schätzer im PL-Modells und dem erwartungstreuen MH¹-Schätzer für das Relative Risiko begründet, denn der erste Iterationsschritt der Fixpunktiteration $\Gamma(\theta)$ (bei $\theta = 1$) ist identisch mit dem MH-Schätzer. Besonders in den Studien mit Personenzeiten schneidet der PL-Ansatz im Vergleich mit den anderen zwei Modellen am besten ab. Bei binomialverteilten Beobachtungen ist der Bias im ND-Ansatz nur bei einem Relativen Risiko in der Umgebung von eins sehr klein. Damit kann bei einem Relativen Risiko, das echt größer oder echt kleiner als eins ist, nicht von einer erwartungsgetreuen Schätzung ausgegangen werden. Das ND-

¹MH = Mantel-Haenszel

Modell ist somit ein konservatives Schätzverfahren, weil es bei einem wahren Relativen Risiko von kleiner als eins einen positiven Bias und bei einem Relativen Risiko von größer als eins einen negativen Bias besitzt. Auch das ML-Modell schätzt das Relative Risiko nicht erwartungstreu. Der absolute Wert des Bias steigt bei größer werdendem Relativen Risiko stetig an. Als Resumé kann aus dieser Simulationsstudie geschlossen werden, dass unter Effekthomogenität der PL-Ansatz die besseren statistischen Schätzeigenschaften gegenüber den anderen zwei Modellen besitzt.

Unter Effektheterogenität sind zwei Arten der Simulationen durchgeführt worden. Die erste ist die Bootstrapsimulation. Die unter Homogenität gewonnenen Resultate lassen sich leider nicht auf den Heterogenitätsfall übertragen. Dennoch zeigt sich der PL-Ansatz, außer im Falle seltener Ereignisse, erneut als ein sehr guter Schätzer für die vorgegebene mischende Verteilung. Bei der CALGB-Bootstrapsimulation (Studie mit sehr geringem Stichprobenumfang) konnte der PL-Ansatz nur in 6 von 1000 Wiederholungen die vorgegebene Anzahl von zwei Mischungskomponenten finden. Den restlichen Wiederholungen ist Homogenität im Effekt unterstellt worden. Die anderen beiden Modelle brachten dagegen deutlich mehr Replikationen mit 2 Mischungskomponenten hervor. Bei der Bootstrapsimulation mit der Beta-Bockerstudie und der Studie über koronare Herzkrankheit hatte das PL-Modell bei der zweikomponentigen Verteilung in jeder Komponente den geringsten Bias vom Relativen Risiko. Die Varianz des Schätzers vom PL- und vom ND-Ansatz ist in jeder Komponente deutlich kleiner als die Varianz der Schätzung mit dem ML-Modell. In beiden Bootstrapsimulationen wird aber eine sehr gut separierte a priori Verteilung benutzt. Bei der dreikomponentigen vorgegebenen Verteilung sind die Werte des Bias vom PL- und vom ND-Modell etwas schlechter als bei der zweikomponentigen. Beim ML-Modell wird die erste Komponente deutlich über- und die dritte sehr stark unterschätzt. Die zweite Komponente wird dagegen von allen drei Modellen am besten reproduziert.

Die zweite Art der Simulation ist unabhängig von einer multizentrischen Studie. Alle Parameter werden in dieser Simulation so vorgegeben, dass bestimmte Szenarien aus der Praxis abgebildet werden. Wieder zeigt sich, dass der PL-Ansatz im Falle von seltenen Ereignissen Schwierigkeiten hat, alle vorgegebenen Mischungskomponenten wiederzugeben. Der Hauptgrund liegt sehr wahrscheinlich in der Modellfehlspezifikation des PL-Modells, welche im Abschnitt 5.2.4 genauer beschrieben worden ist. Bei allen hier durchgeführten Simulationsstudien zeigen der ND- und der PL-Ansatz sehr ähnliche Ergebnisse. Die Auswertung der Studien und der Simulationsexperimente mit und ohne Effektheterogenität weisen darauf, dass der ND-Ansatz sehr oft das Relative Risiko näher am Nulleffekt schätzt als das PL-Modell.

Ziel einer multizentrischen Studie ist, den Effekt einer Behandlung mit einer anderen Behandlung oder Kontrollgruppe zu vergleichen. Ist nur der Effekt unabhängig vom Baseline von Interesse, dann kann auch bei moderater bis starker Baselineheterogenität das PL-Modell dem ML-Modell vorgezogen werden (siehe Tabelle 5.10, 5.11 und 5.13). Ein Beispiel ist die Studie über Frühgeburten (1.7), wo der Baseline im Kontrollarm im Intervall von 0 bis 0.733 verteilt ist. Das ND- und das PL-Modell schätzen dort eine zweikomponentige mischende Verteilung. In der ersten Komponente ist das Risiko eine Frühgeburt in der Behandlungsgruppe zweimal kleiner als in der Kontrollgruppe. In der zweiten Komponente ist das Risiko in der Behandlungsgruppe sogar 14-mal kleiner als in der Kontrollgruppe. Der Grund dieser deutlichen Senkung der Wahrscheinlichkeit im Behandlungsarm liegt darin, dass in den Zentren 8, 12 und 13 keine Fälle beobachtet worden sind. Interessant ist aber, dass nur das ND-Modell diese drei Zentren zur zweiten Komponente zuordnet. Das ML-Modell findet im zweikomponentigen Modell diese zwei Mischungskomponenten nicht. Im Falle von Baselinehomogenität (Tabelle 5.12 und 5.14) ist jedoch das ML-Modell ein sehr adäquates Modell. Somit sollte der Baselineparameter bei der Modellwahl mit berücksichtigt werden.

Aus diesen Resultaten kann man keine generelle Empfehlung für eines der drei Modelle geben, welches in einer multizentrischen Studie unter Effektheterogenität in jedem Fall das beste Schätzergebnis liefert. Unabhängig von der Baselineheterogenität kann sowohl mit dem PL-Modell als auch mit dem ND-Modell sehr gut in einer multizentrischen Studie geprüft werden, ob Effektheterogenität vorliegt. Ist die Gradientenfunktion im ein-komponenten Modell im gesamten Intervall kleiner gleich eins, dann kann nach dem PL- und ND-Modell mit großer Sicherheit von Effekthomogenität ausgegangen werden. Ein wichtiges Resultat dieser Arbeit ist, dass die Qualität des PL-Ansatzes als gleichwertig gegenüber dem ND-Ansatz angesehen werden kann. Weiterhin zeigen die Simulationsstudien im Homogenitäts- und im Heterogenitätsfall, dass bei Studien mit Personenzeiten (Tabelle 1.8, 1.9 und 1.10) der PL-Ansatz eine klare Alternative zum ND-Ansatz darstellt.

Es gibt sehr viele verschiedene Situationen die durch Parameteränderungen simuliert werden können. Hier ist nur ein Bruchteil von allen Möglichkeiten untersucht worden. In den Ergebnissen der Simulationsstudien, besonders im Seltenheitsfall fällt auf, dass die erste Komponente mit einem Relativen Risiko von 0.5 vom PL- und vom ND-Ansatz deutlich besser geschätzt wird als die zweite Komponente mit einem Relativen Risiko von 1.5. Die Ursache dieses Schätzverhalten konnte hier nicht ausführlich geklärt werden. Gezielte Simulationsstudien können aber weitere Eigenschaften der drei Modelle hervorbringen.

Zusätzlich könnten noch Modifikationen an den Modellen vorgenommen werden. Beim

ML-Modell könnte beispielsweise der Baseline für alle Zentren als gleich vorausgesetzt werden. Damit würde nur die Effektheterogenität in die Schätzung mit einfließen. Dieses ML-Modell mit festem Baseline ist in dieser Arbeit nicht betrachtet worden, weil die Baselineheterogenität im PL-Modell und im ND-Modell indirekt berücksichtigt wird. Weiterhin könnte die Modellfehlspezifikation im PL-Modell über den Binomiallikelihood bei binomialverteilten Beobachtungen beseitigt werden. Zusätzlich gibt es auch die Möglichkeit im PL-Modell die Kovariablen zu integrieren. Daher sollte dieser Ansatz bei künftigen Untersuchungen weiter betrachtet werden.

6.1 Zusammenfassung

Das Hauptaugenmerk dieser Dissertation liegt im Vergleich dreier unterschiedlicher Schätzverfahren, welche den Behandlungseffekt in multizentrischen Studien berechnen. Das erste Modell ist der Normalverteilungsansatz, welcher voraussetzt, dass die logarithmischen Relativen Risiken in den einzelnen Zentren normalverteilt sind. Im zweiten Modell, dem Profil-Likelihood-Modell, wird der Schätzer für das Relative Risiko anhand der Poissonlikelihood formuliert. Die Likelihoodfunktion ist hier abhängig vom interessanten Parameter, dem Relativen Risiko, und vom Baselineparameter (Nuisanceparameter). In dieser Arbeit wird der Nuisanceparameter durch den Profil-Maximum-Likelihoodschätzer eliminiert. Das dritte hier betrachtete Modell ist das Multi-Level-Modell. In diesem Modell gibt es zwei Ebenen. Das erste Level modelliert das Relative Risiko in jedem Zentrum mit Hilfe der Logistischen Regression. Im zweiten Level wird die Verteilung des Relativen Risikos über alle Zentren formuliert. In allen drei Modellen wird die Effektheterogenität anhand einer Mischungsverteilung modelliert.

Kernpunkte dieser Arbeit sind neben dem Vergleich der drei Modelle ein weiterentwickelter EM-Algorithmus mit Gradientenfunktionsupdate. Das Ziel dieses Algorithmus ist, den nichtparametrische Maximum-Likelihood-Schätzer zu identifizieren, welcher identisch mit dem globalen Maximum der Likelihoodfunktion ist. Mit diesem Algorithmus ist der Behandlungseffekt in dreizehn ausgesuchten multizentrischen Studien und in einer Vielzahl von Simulationsexperimenten berechnet worden. Auch in anderen statistischen Modellen, welche eine nichtparametrische Mischverteilung modellieren, ist dieser Algorithmus anwendbar.

In den Simulationsstudien wurden statistische Kenngrößen wie Bias, Varianz und kleinster quadratischer Fehler des Behandlungseffekts je nach Modell berechnet, um so die Qualität der drei Ansätze zu verdeutlichen. In den Simulationsstudien mit homogenem

Behandlungseffekt bewies der Profil-Likelihoodansatz sehr gute statistische Eigenschaften. Bei fast jeder Simulationsstudie zeigte dieser Ansatz die kleinsten absoluten Werte des Bias und der Varianz.

In den Simulationsstudien mit zugelassener Effektheterogenität, in der die Variabilität in der Schätzung der Parameter viel größer ist, konnte sich kein Modell klar gegenüber den anderen behaupten. Es zeigte sich, dass der Profil-Likelihoodansatz und der Normalverteilungsansatz sehr ähnlich die zugrunde liegende mischende Verteilung schätzen. Dabei schätzt das Normalverteilungsmodell das Relative Risiko in den einzelnen Komponenten näher am Nulleffekt als der Profil-Likelihoodansatz. Somit stellt der Profil-Likelihoodansatz besonders bei Studien mit Personenzeiten eine klare Alternative gegenüber dem Normalverteilungsansatz dar. Einen sehr großen Einfluss auf die Effektschätzung hat im Multi-Level-Modell das Baselineisiko. Die Simulationen zeigen, dass eine starke Baselineheterogenität sich ungünstig auf die Qualität der Effektschätzung auswirkt. Dagegen kann das Multi-Level-Modell unter Baselinehomogenität sehr gut eine vorgegebene mischende Verteilung reproduziert. Weiterhin konnten der Profil-Likelihood und der Normalverteilungsansatz sehr gut die vorgegebene Anzahl an Mischungskomponenten wiedergeben. Das Multi-Level-Modell überschätzte dagegen sehr stark die vorgegebene Anzahl an Komponenten. Zusammenfassend weist der Profil-Likelihoodansatz mindestens eine dem Normalverteilungsansatz entsprechende Güte in der Effektschätzung auf.

Anhang A

A.1 Bootstrap-Stichproben

Tabelle A.1: Bootstrap-Stichprobe nach den Vorgaben der Studie über koronarer Herzkrankheit, bei der eine Komponente einen besonders großen Schätzwert für das Relative Risiko im Profil-Likelihoodansatz hat.

Zentrum	Behandlungsarm		Kontrollarm	
i	x^T	n^T	x^C	n^C
1	47	380	53	350
2	47	1250	38	640
3	48	690	47	500
4	5	90	3	30
5	2	30	0	30
6	68	1240	71	1180
7	58	1930	53	890
8	26	340	15	350
9	70	1930	115	1920
10	88	1240	39	410
11	30	1140	46	1140
12	10	210	10	220
13	11	210	15	230
14	1	90	6	170
15	1833	38620	727	19420
16	225	1350	184	1330
17	19	890	32	860
18	47	1970	50	2060
19	1	150	8	150
20	35	2150	62	2100
21	33	1010	31	1120
22	3	100	0	50
23	13	340	1	340
24	152	4410	246	4390
25	69	3850	73	3740
26	1	190	1	190
27	4	1510	15	1560
28	88	13850	56	13800
29	28	10140	49	10040
30	5	5910	3	1500
31	191	27630	180	27590
32	2	100	2	100
33	2	20	3	30

Tabelle A.2: Bootstrap-Stichprobe nach den Vorgaben der Studie über koronarer Herzkrankheit, bei der eine Komponente einen besonders großen Schätzwert für das Relative Risiko im Normalverteilungsansatz hat.

Zentrum i	Behandlungsarm		Kontrollarm	
	x^T	n^T	x^C	n^C
1	53	380	50	350
2	98	1250	46	640
3	17	690	41	500
4	7	90	2	30
5	4	30	0	30
6	56	1240	84	1180
7	103	1930	52	890
8	22	340	16	350
9	122	1930	118	1920
10	106	1240	23	410
11	49	1140	54	1140
12	10	210	10	220
13	19	210	31	230
14	1	90	1	170
15	929	38620	786	19420
16	219	1350	183	1330
17	19	890	30	860
18	29	1970	49	2060
19	3	150	6	150
20	41	2150	56	2100
21	14	1010	33	1120
22	2	100	0	50
23	11	340	3	340
24	227	4410	262	4390
25	42	3850	58	3740
26	2	190	0	190
27	12	1510	13	1560
28	45	13850	68	13800
29	38	10140	40	10040
30	14	5910	0	1500
31	169	27630	175	27590
32	0	100	0	100
33	4	20	0	30

A.2 Ergebnisse einzelner Bootstrap-Stichproben

Das PL-Modell (Tabelle A.3) schätzt mit den Daten der ersten Bootstrap-Stichprobe (Tabelle A.1) in der drei- und vierkomponentigen Lösung eine Ausreißerkomponente mit einem Relativen Risiko von 12.36. Zu dieser Komponente gehört aber nur das 23. Zentrum. Nach dem NPMLE-Kriterium findet dieses Modell 4 Mischungskomponenten, wo außer der dritten Komponente alle zur vorgegebenen Verteilung $P = \begin{pmatrix} 0.6 & 0.9 & 1.2 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$ passen. Das ND-Modell (Tabelle A.4) ist bei dieser Studie robuster und schätzt mit drei Mischungskomponenten nach dem NPMLE-Kriterium fast die richtige a priori Verteilung. Bei den Daten der zweiten Bootstrap-Stichprobe (Tabelle A.2) schätzt das ND-Modell (Tabelle A.7) eine Ausreißerkomponente mit einem Relativen Risiko von 6.8. Zu dieser Mischungskomponente gehören die Zentren 5, 30 und 33. In diesen drei Zentren wurden keine Fälle im Kontrollarm beobachtet. Bei der dreikomponentigen Lösung des PL-Modells (Tabelle A.6) liegen alle drei geschätzten Komponenten im Bereich der vorgegebenen Verteilung P . Das ML-Modell schätzt schon nach dem BIC-Kriterium deutlich mehr Mischungskomponenten. Die einzelnen Komponenten der dreikomponentigen Lösung des ML-Modells unterscheiden aber sich deutlich von der vorgegebenen mischenden Verteilung P .

Tabelle A.3: Ergebnisse der Bootstrap-Studie (Tabelle A.1) anhand des Profil-Likelihoodansatzes

Profil-Likelihood				
Komp.	1.			
θ	1.006647			
q	1.000000			
H	33			
Log-L.=	-237.804746			
max GF =	49000.721604			
BIC=	-479.106000			
Komp.	1.	2.		
θ	.650141	1.225349		
q	.614341	.385659		
H	21	12		
Log-L.=	-195.404056			
max GF =	10.936823			
BIC=	-401.297634			
Komp.	1.	2.	3.	
θ	.652048	1.220140	12.362823	
q	.624652	.344215	.031133	
H	24	8	1	
Log-L.=	-193.643222			
max GF =	1.198333			
BIC=	-404.768983			
Komp.	1.	2.	3.	4.
θ	.629751	1.252899	12.361214	.987451
q	.547294	.229244	.031328	.192134
H	23	5	1	4
Log-L.=	-192.968707			
max GF =	1.000000			
BIC=	-410.412967			

Tabelle A.4: Ergebnisse der Bootstrap-Studie (Tabelle A.1) anhand des Normalverteilungsansatzes

Normalverteilungsansatz			
Komp.	1.		
θ	1.004053		
q	1.000000		
H	33		
Log-L.=	-64.274635		
max GF =	42886.601857		
BIC=	-132.045778		
Komp.	1.	2.	
θ	.656301	1.220598	
q	.610219	.389781	
H	21	12	
Log-L.=	-24.467422		
max GF =	1.176821		
BIC=	-59.424367		
Komp.	1.	2.	3.
θ	.634090	1.253461	.989152
q	.530845	.261916	.207239
H	21	8	4
Log-L.=	-23.814023		
max GF =	1.000001		
BIC=	-65.110584		

Tabelle A.5: Ergebnisse der Bootstrap-Studie (Tabelle A.1) anhand des Multi-Level Modells (Poissonlikelihood).

Multi-Level-Modell (Poi)						
Komp.	1.					
θ	1.149666					
q	1.000000					
H	33					
Log-L.	= -2660.480822					
max GF	= 7.8734×10^{214}					
BIC	= -5324.458151					
Komp.	1.	2.				
θ	.999998	.951760				
q	.773741	.226259				
H	26	7				
Log-L.	= -723.441120					
max GF	= 8.9043×10^{86}					
BIC	= -1457.371762					
Komp.	1.	2.	3.			
θ	1.037340	.963059	1.114616			
q	.746514	.186288	.067199			
H	25	6	2			
Log-L.	= -465.489269					
max GF	= 7.1362×10^{23}					
BIC	= -948.461076					
Komp.	1.	2.	3.	4.	5.	6.
θ	1.267216	1.052163	1.114600	.741079	.749692	.602014
q	.057214	.043968	.065983	.233596	.082769	.195226
H	1	1	2	10	3	6
Komp.	7.	8.	9.	10.	11.	12.
θ	.507871	12.037292	.422116	1.561378	1.016247	.912813
q	.068918	.035325	.031997	.038805	.099373	.046825
H	3	1	1	1	3	1
Log-L.	= -245.530926					
max GF	= 3.058382					
BIC	= -571.481526					

Tabelle A.6: Ergebnisse der Bootstrap-Studie (Tabelle A.2) anhand des Profil-Likelihoodansatzes

Profil-Likelihood				
Komp.	1.			
θ	.778942			
q	1.000000			
H	33			
Log-L.=	-222.468445			
max GF =	169.329886			
BIC=	-448.433398			
Komp.	1.	2.		
θ	.598834	.994824		
q	.313784	.686216		
H	10	23		
Log-L.=	-194.491887			
max GF =	2.253121			
BIC=	-399.473296			
Komp.	1.	2.	3.	
θ	.597249	.913066	1.137529	
q	.289145	.441153	.269703	
H	9	19	5	
Log-L.=	-193.814083			
max GF =	2.368546			
BIC=	-405.110704			
Komp.	1.	2.	3.	4.
θ	.603533	.907664	1.138340	.313373
q	.244918	.458890	.269616	.026576
H	6	21	5	1
Log-L.=	-193.366396			
max GF =	.996415			
BIC=	-411.208344			

Tabelle A.7: Ergebnisse der Bootstrap-Studie (Tabelle A.2) anhand des Normalverteilungsansatzes

Normalverteilungsansatz					
Komp.	1.				
θ	.779153				
q	1.000000				
H	33				
Log-L.	= -57.758167				
max GF	= 181862.428541				
BIC	= -119.012842				
Komp.	1.		2.		
θ	.996719		.599900		
q	.702030		.297970		
H	24		9		
Log-L.	= -30.199229				
max GF	= 4.769987				
BIC	= -70.887980				
Komp.	1.	2.	3.		
θ	.982224	.599806	6.809290		
q	.597298	.300970	.101731		
H	20	10	3		
Log-L.	= -27.335496				
max GF	= 1.904453				
BIC	= -72.153531				
Komp.	1.	2.	3.	4.	
θ	.979974	.604614	6.808151	.313071	
q	.606758	.270186	.101733	.021324	
H	21	8	3	1	
Log-L.	= -27.082418				
max GF	= 1.112698				
BIC	= -78.640389				
Komp.	1.	2.	3.	4.	5.
θ	.906376	.603167	6.837582	.313229	1.115996
q	.429120	.241165	.100247	.022516	.206953
H	20	6	4	1	2
Log-L.	= -26.560818				
max GF	= 1.000001				
BIC	= -84.590204				

Tabelle A.8: Ergebnisse der B.-Studie (Tabelle A.2) anhand des Multi-Level Modells (Poissonlikelihood).

Multi-Level-Modell (Poi)						
Komp.	1.					
θ	.854552					
q	1.000000					
H	33					
Log-L.=	-2573.837735					
max GF	$= 4.5361 \times 10^{189}$					
BIC=	-5151.171977					
Komp.	1.	2.				
θ	.843507	.719740				
q	.283402	.716598				
H	9	24				
Log-L.=	-854.429437					
max GF	$= 6.4026 \times 10^{100}$					
BIC=	-1719.348396					
Komp.	1.	2.	3.			
θ	.843034	.621540	.979244			
q	.281634	.312595	.405771			
H	9	10	14			
Log-L.=	-458.612880					
max GF	$= 1.3084 \times 10^{36}$					
BIC=	-934.708297					
Komp.	1.	2.	3.	4.	5.	6.
θ	.970072	.595940	.892374	1.090112	.679112	6.457682
q	.107649	.062824	.247168	.099095	.170527	.033352
H	5	2	9	3	6	1
Komp.	7.	8.	9.	10.		
θ	1.286447	.770454	.727698	.302771		
q	.109002	.090072	.048715	.031596		
H	3	2	1	1		
Log-L.=	-246.968061					
max GF	$= 11.020904$					
BIC=	-560.369765					

A.3 Ableitungen der Gradientenfunktion im ML-Modell

Beim Binomial-Likelihood sind die Ableitungen der Gradientenfunktion

$$\frac{\partial}{\partial \alpha} d(\alpha, \beta, P) = \frac{1}{k} \sum_{i=1}^k \frac{\frac{e^{x_i^C \alpha + x_i^T (\alpha + \beta)} (x_i^T + x_i^C + e^{\alpha + \beta} (x_i - n_i^T)) + e^{\alpha} (x_i - n_i^C) + e^{2\alpha + \beta} (x_i - n_i)}{(1 + e^{\alpha})^{n_i^C + 1} (1 + e^{\alpha + \beta})^{n_i^T + 1}}}{\sum_{j=1}^m \frac{e^{x_i^C \alpha_j + x_i^T (\alpha_j + \beta_j)}}{(1 + e^{\alpha_j})^{n_i^C} (1 + e^{\alpha_j + \beta_j})^{n_i^T}} Q_j}$$

$$\frac{\partial}{\partial \beta} d(\alpha, \beta, P) = \frac{1}{k} \sum_{i=1}^k \frac{\frac{e^{x_i^C \alpha + x_i^T (\alpha + \beta)} (x_i^T + e^{\alpha + \beta} (x_i^T - n_i^T))}{(1 + e^{\alpha})^{n_i^C} (1 + e^{\alpha + \beta})^{n_i^T + 1}}}{\sum_{j=1}^m \frac{e^{x_i^C \alpha_j + x_i^T (\alpha_j + \beta_j)}}{(1 + e^{\alpha_j})^{n_i^C} (1 + e^{\alpha_j + \beta_j})^{n_i^T}} Q_j}$$

$$\frac{\partial}{\partial \alpha^2} d(\alpha, \beta, P) = \frac{1}{k} \sum_{i=1}^k \frac{t(t_1 * t_2 + t_4(t_3 + t_5 + t_6))}{\sum_{j=1}^m \frac{e^{x_i^C \alpha_j + x_i^T (\alpha_j + \beta_j)}}{(1 + e^{\alpha_j})^{n_i^C} (1 + e^{\alpha_j + \beta_j})^{n_i^T}} Q_j}$$

$$\begin{aligned} t &= e^{x_i^C \alpha + x_i^T (\alpha + \beta)} (1 + e^{\alpha})^{-n_i^C - 2} (1 + e^{\alpha + \beta})^{-n_i^T - 2} \\ t_1 &= e^{\alpha} (1 + e^{\alpha}) (1 + e^{\alpha + \beta}) \\ t_2 &= -n_i^C + x_i + e^{\beta} (-n_i^T + x_i) + 2e^{\alpha + \beta} (-n_i + x_i) \\ t_3 &= e^{\alpha + \beta} (1 + e^{\alpha}) (-1 - n_i^T) \\ t_4 &= x_i + e^{\alpha + \beta} (-n_i^T + x_i) + e^{\alpha} (-n_i^C + x_i) + e^{2\alpha + \beta} (-n_i + x_i) \\ t_5 &= e^{\alpha} (1 + e^{\alpha + \beta}) (-1 - n_i^C) \\ t_6 &= (1 + e^{\alpha}) (1 + e^{\alpha + \beta}) x_i \end{aligned}$$

$$\frac{\partial}{\partial \beta^2} d(\alpha, \beta, P) = \frac{1}{k} \sum_{i=1}^k \frac{t * t_1}{\sum_{j=1}^m \frac{e^{x_i^C \alpha_j + x_i^T (\alpha_j + \beta_j)}}{(1 + e^{\alpha_j})^{n_i^C} (1 + e^{\alpha_j + \beta_j})^{n_i^T}} Q_j}$$

$$\begin{aligned} t &= e^{x_i^C \alpha + x_i^T (\alpha + \beta)} (1 + e^{\alpha})^{-n_i^C} (1 + e^{\alpha + \beta})^{-2 - n_i^T} \\ t_1 &= e^{2(\alpha + \beta)} (n_i^T - x_i^T)^2 + (x_i^T)^2 - e^{\alpha + \beta} (n_i^T + 2n_i^T x_i^T - 2(x_i^T)^2) \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \alpha \beta} d(\alpha, \beta, P) &= \frac{1}{k} \sum_{i=1}^k \frac{t(t_1 + t_2 + t_3)}{\sum_{j=1}^m \frac{e^{x_i^C \alpha_j + x_i^T (\alpha_j + \beta_j)}}{(1+e^{\alpha_j})^{n_i^C} (1+e^{\alpha_j + \beta_j})^{n_i^T}} q_j} \\
t &= e^{x_i^C \alpha + x_i^T (\alpha + \beta)} (1 + e^{\alpha})^{-1 - n_i^C} (1 + e^{\alpha + \beta})^{-2 - n_i^T} \\
t_1 &= x_i^T (x_i + e^{\alpha} (-n_i^C + x_i)) \\
t_2 &= e^{2(\alpha + \beta)} (-n_i^T + x_i^T) (-n_i^T + x_i + e^{\alpha} (-n_i + x_i)) \\
t_3 &= e^{\alpha + \beta} (e^{\alpha} n_i^T (-1 + n_i^C - 2x_i^T - x_i^C) - n_i^T (1 + 2x_i^T + x_i^C) \\
&\quad + 2x_i^T (x_i + e^{\alpha} (-n_i^C + x_i))) \\
\frac{\partial}{\partial \beta \alpha} d(\alpha, \beta, P) &= \frac{\partial}{\partial \alpha \beta} d(\alpha, \beta, P)
\end{aligned}$$

Beim Poisson-Likelihood sind die Ableitungen der Gradientenfunktion

$$\begin{aligned}
\frac{\partial d(\theta, P)}{\partial \alpha} &= \frac{1}{k} \sum_{i=1}^k \left(\sum_{j=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1} \\
t_1 &= n_i^C (e^{\alpha} - e^{\alpha_j}) + n_i^T (e^{\alpha + \beta} - e^{\alpha_j + \beta_j}) \\
t_2 &= x_i^C (\alpha_j - \alpha) + x_i^T (\alpha_j + \beta_j - \alpha - \beta) + \log q_j \\
t_3 &= -\log(-e^{\alpha} n_i^C - e^{\alpha + \beta} n_i^T + x_i^C + x_i^T)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial d(\theta, P)}{\partial \beta} &= \frac{1}{k} \sum_{i=1}^k \left(\sum_{j=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1} \\
t_1 &= n_i^C (e^{\alpha} - e^{\alpha_j}) + n_i^T (e^{\alpha + \beta} - e^{\alpha_j + \beta_j}) \\
t_2 &= x_i^C (\alpha_j - \alpha) + x_i^T (\alpha_j + \beta_j - \alpha - \beta) + \log q_j \\
t_3 &= -\log(-e^{\alpha + \beta} n_i^T + x_i^T)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial d(\theta, P)}{\partial \alpha^2} &= \frac{1}{k} \sum_{i=1}^k \left(\sum_{j=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1} \\
t_1 &= n_i^C (e^{\alpha} - e^{\alpha_j}) + n_i^T (e^{\alpha + \beta} - e^{\alpha_j + \beta_j}) \\
t_2 &= x_i^C (\alpha_j - \alpha) + x_i^T (\alpha_j + \beta_j - \alpha - \beta) + \log q_j \\
t_3 &= -\log(e^{2\alpha} (n_i^C)^2 + 2e^{2\alpha + \beta} n_i^C n_i^T + e^{2(\alpha + \beta)} (n_i^T)^2 \\
&\quad + (x_i^C + x_i^T)^2 - e^{\alpha} n_i^C (1 + 2x_i^C + 2x_i^T) - e^{\alpha + \beta} n_i^T (1 + 2x_i^C + 2x_i^T))
\end{aligned}$$

$$\begin{aligned} \frac{\partial d(\theta, P)}{\partial \beta^2} &= \frac{1}{k} \sum_{i=1}^k \left(\sum_{j=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1} \\ t_1 &= n_i^C (e^\alpha - e^{\alpha_j}) + n_i^T (e^{\alpha+\beta} - e^{\alpha_j+\beta_j}) \\ t_2 &= x_i^C (\alpha_j - \alpha) + x_i^T (\alpha_j + \beta_j - \alpha - \beta) + \log q_j \\ t_3 &= -\log(e^{2(\alpha+\beta)} (n_i^T)^2 + (x_i^T)^2 - e^{\alpha+\beta} n_i^T (1 + 2x_i^T)) \end{aligned}$$

$$\begin{aligned} \frac{\partial d(\theta, P)}{\partial \alpha \beta} &= \frac{1}{k} \sum_{i=1}^k \left(\sum_{j=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1} \\ t_1 &= n_i^C (e^\alpha - e^{\alpha_j}) + n_i^T (e^{\alpha+\beta} - e^{\alpha_j+\beta_j}) \\ t_2 &= x_i^C (\alpha_j - \alpha) + x_i^T (\alpha_j + \beta_j - \alpha - \beta) + \log q_j \\ t_3 &= -\log(e^{2\alpha+\beta} n_i^C n_i^T + e^{2(\alpha+\beta)} (n_i^T)^2 + (x_i^T) (x_i^C + x_i^T) \\ &\quad - e^\alpha n_i^C x_i^T - e^{\alpha+\beta} n_i^T (1 + x_i^C + 2x_i^T)) \end{aligned}$$

Literaturverzeichnis

1. Aitkin, Murray: Meta-analysis by random effect modelling in generalized linear models. *Statistics in Medicine* 18 (1999) 2343-2351
2. Aitkin, Murray: A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* 55 (1999) 117-128
3. Arends, Lidia, Hoes, Arno u. a.: Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Statistics in Medicine* 19 (2000) 3497-3518
4. Armitage, Peter und Colton, Theodore: *Encyclopedia of Biostatistics*. Chichester, New York, Weinheim, Brisbane, Singapore, Toronto 1998 (Volume 4)
5. Berry, SM.: Meta-analysis versus large trials: resolving the controversy. In: Stangl, Dalene, Berry, Donald u. a.: *Meta-Analysis in Medicine and Health Policy*, New York und Basel 2000
6. Böhning, Dankmar, Sarol, Jesus Jr.: Estimating risk difference in multicenter studies under baseline-risk heterogeneity. *Biometrics* 5 (2000) 304-308
7. Böhning, Dankmar: *Allgemeine Epidemiologie und ihre methodischen Grundlagen*, Oldenburg, München 1998
8. Böhning, Dankmar: *Computer-assisted analysis of mixtures and applications. Meta-Analysis, Disease Mapping and Others*, Boca Raton, London, New York, Washington D.C. 2000
9. Böhning, Dankmar: The EM algorithm with gradient function update for discrete mixtures with know (fixed) number of components. *Statistics and Computing* 13 (2003) 257-265

10. Celeux, Gilles, Chauveau, Didier u. a.: Stochastic version of the EM algorithm: An experimental study in the mixture case. *Journal of Statistical Computation and Simulation* 55 (1996) 287-134
11. Cooper, Nicola, Sutton, Alex u. a.: Decision analytical economic modelling within a Bayesian framework: application to prophylactic antibiotics use for caesarean section. *Statistical Methods in Medical Research* 11 (2002) 491-512
12. Dempster, Arthur, Laird, Nan u. a.: Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* 39 (1977) 1- 38
13. DuMouchel, William, Normand, Sharon-Lise: Computer-modeling and graphical strategies for meta-analysis. In: Stangl, Dalene, Berry, Donald u. a.: *Meta-Analysis in Medicine and Health Policy*, New York und Basel 2000
14. Efron Bradley: *Bootstrap Methods: Another look at the jackknife*. *The Annals of Mathematical Statistics* 7 (1979) 1-26
15. Efron, Bradley, Tibshirani, Robert J.: *An introduction to the bootstrap*, London 1993
16. Fawzi, Wafaie, Chalmers, Thomas u. a.: Vitamin A supplementation and child mortality, *JAMA* 269 (1993) 898-903
17. Fisher B, Montague E, u. a.: Comparison of radical mastectomy with alternative treatments for primary breast cancer. A first report of results from a prospective randomized clinical trial. *Cancer* 39 (1977) Supplement 2827-2839
18. Gelman, Andrew, Carlin, John u. a.: *Bayesian Data Analysis*, London 1995
19. Hargens, Jürgen: Psychotherapie wirkt - aber was erfasst empirische Psychotherapieforschung (nicht)? *Report Psychologie* 19(10) (1994) 12-16
20. Horwitz, Ralph, Singer, Burton u. a.: Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation. *Journal of Clinical Epidemiology* 49 (1996) 395-400
21. Hutchinson, Wesley, Kamakura, Wagner u. a.: Unobserved heterogeneity as an alternative explanation for „Reversal“ effects in behavioral research. *Journal of Consumer Research* 27 (2000) 324-344

22. Johnson, Norman, Kotz, Samuel u.a.: *Univariate Discrete Distributions*, New York 1992
23. Laird Nan. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* 73 (1978) 805-811
24. Land, Kenneth C.: Introduction to the special issue on finite mixture models. *Sociological Methods & Research* 29 (2001) 275-281
25. Lipsitz, Stuart, Dear, Keith u. a.: Tests for homogeneity of the risk difference when data are sparse. *Biometrics* 54 (1998) 148-160
26. McCullagh, Peter und Nelder, John: *Generalized Linear Models*, London 1989
27. McLachlan, Geoffrey und Krishnan, Thiriyambakam: *The EM Algorithm and Extensions*, Wiley, New York 1997
28. McLachlan, Geoffrey und Peel David: *Finite Mixture Models*, Wiley, New York 2000
29. Metzler, Peter und Krause Bodo: Methodischer Standard bei Studien zu Therapieevaluation. *Methods of Psychological Research* 2 (1997)
30. Richardson, Sylvia und Green, Peter: On Bayesian Analysis of Mixtures with an unknown Number of Components. *Journal of the Royal Statistical Society, series B* 59 (1997) 731-792
31. Rief, Winfried: Ein Plädoyer für eine praxisorientierte Psychotherapieforschung. *Report Psychologie* 19(1) (1994) 16-19
32. Schlittgen Rainer: *Einführung in die Statistik. Analyse und Modellierung von Daten*, München, Wien 1998
33. Turner, Rebecca, Omar, Rumana, u. a.: A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 19 (2000) 3417-3432.
34. Yusuf, Salim, Peto, Richard u. a.: Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases* 27 (1985) 335-371.

Danksagung

In erster Linie möchte ich mich beim Prof. Dr. Dankmar Böhning für seine stetige Unterstützung bedanken. Ohne ihn wäre diese Arbeit nicht möglich gewesen. Als Mitarbeiter in seinem Projekt erhielt ich mit seiner Hilfe einen grundlegenden Einblick in dieses hier bearbeitete Thema. Des Weiteren bedanke ich mich beim Dr. Ekkehart Dietz, der mit seiner geduldigen und ruhigen Art und mit seinem tiefgründigen Fachwissen in vielen Fragen mir zur Seite stand. Diese Dissertation ist im Institut für Sozialmedizin, Epidemiologie und Gesundheitsökonomie, Bereich Internationale Gesundheitswissenschaften in Berlin verfasst worden.