

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-

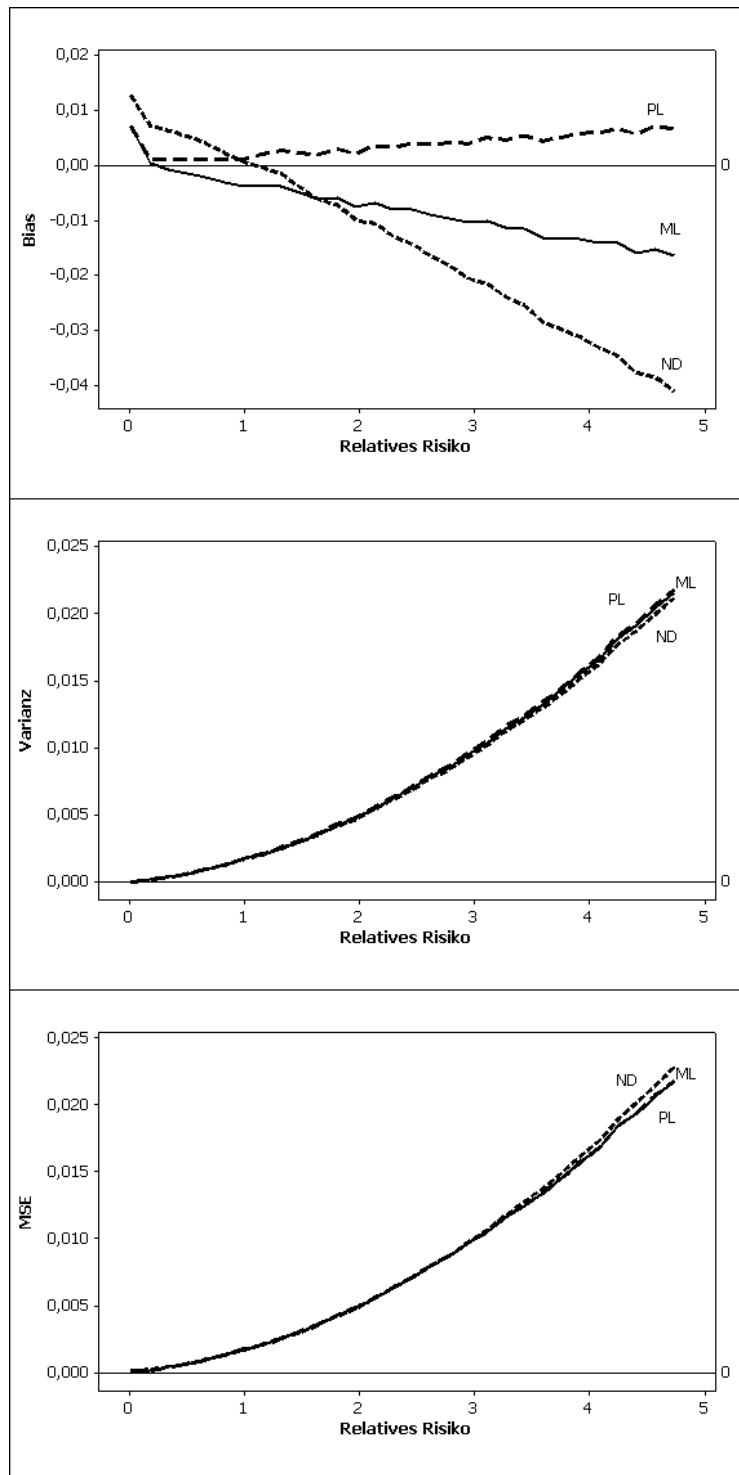


Abbildung 5.1: Beta-Blocker Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;4.73]$ berechnet, die generierten x^T , x^C -Werte sind binomialverteilt)

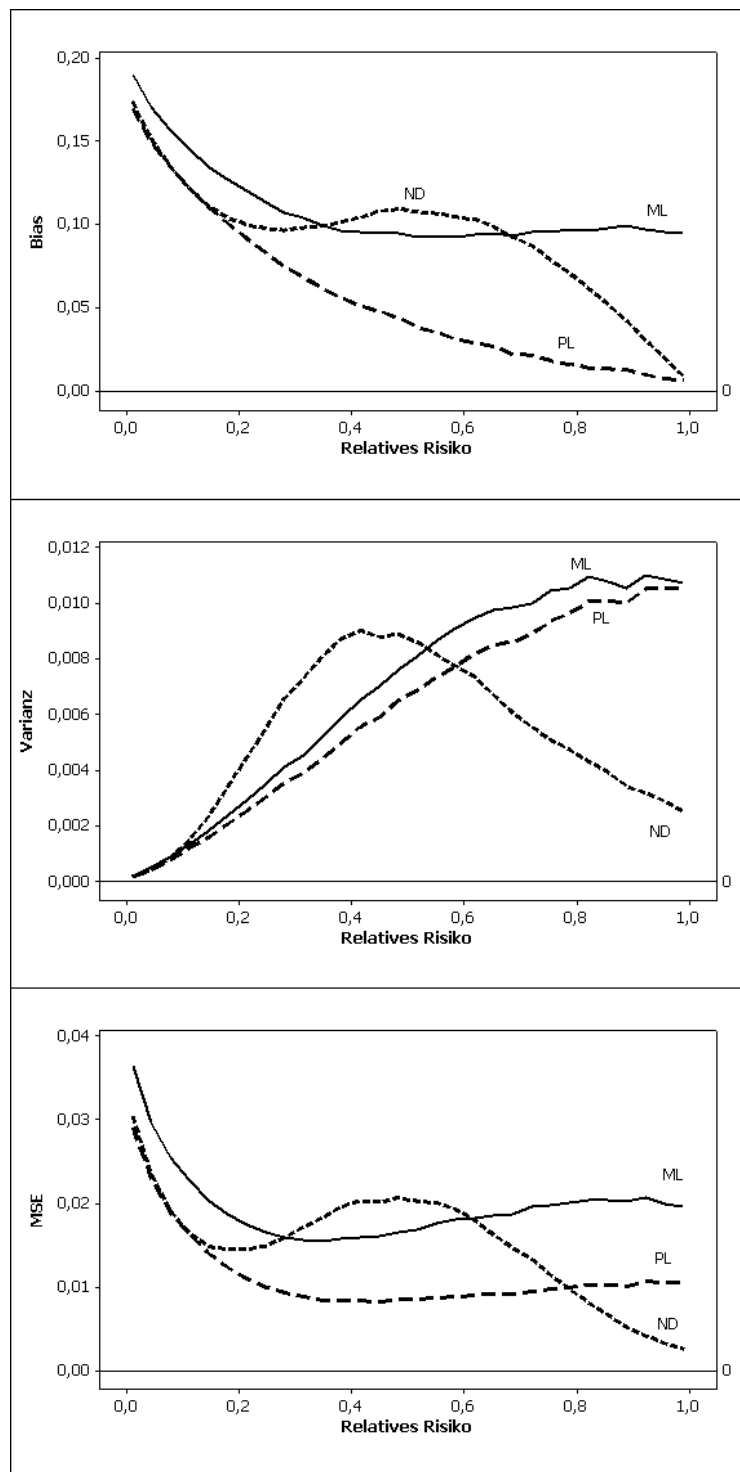


Abbildung 5.2: CALGB-Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0,01;0,99]$ berechnet, die generierten x^T , x^C - Werte sind binomialverteilt)

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.

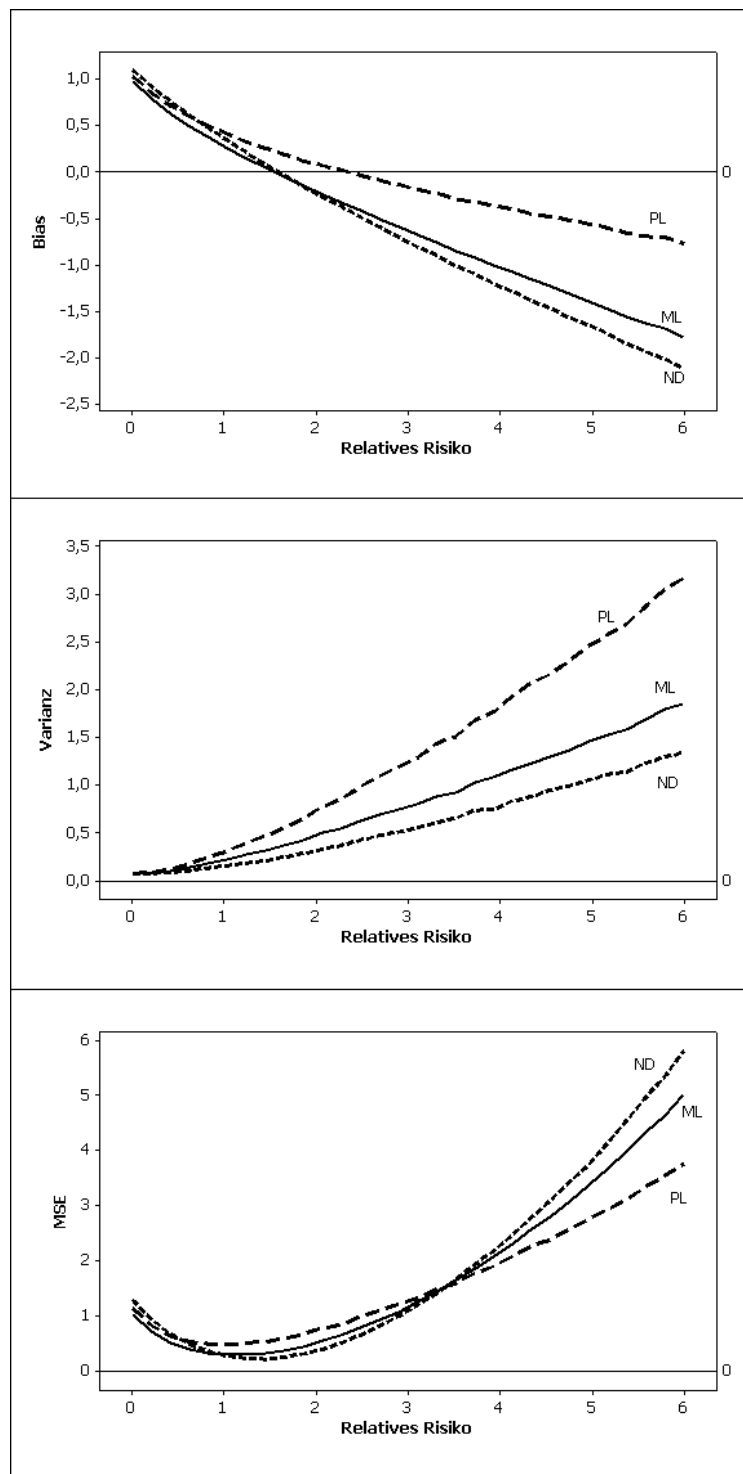


Abbildung 5.3: Multi-Farm-Studie (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;5.99]$ berechnet, die generierten x^T , x^C - Werte sind binomialverteilt)

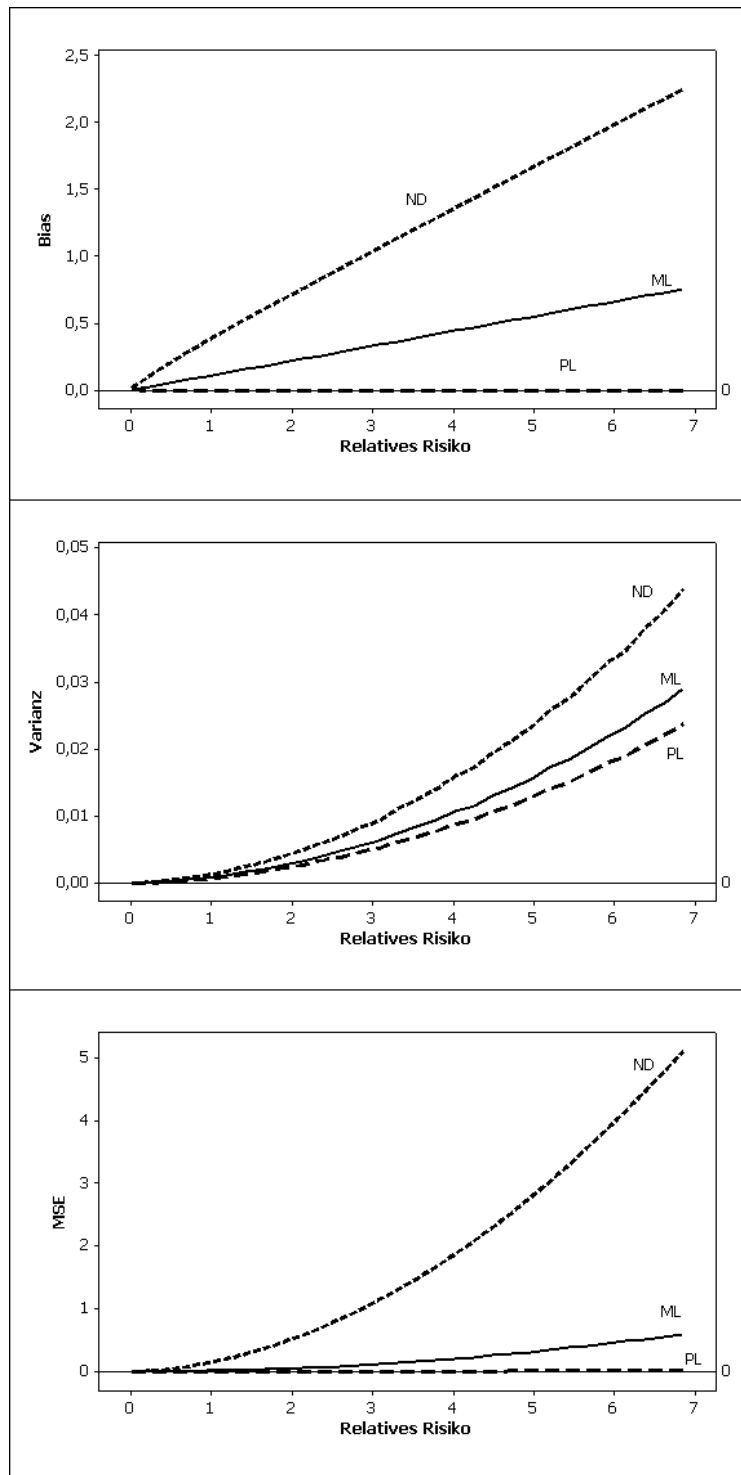


Abbildung 5.4: Studie über koronarer Herzkrankheit (Bias, Varianz und MSE vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;6.85]$ berechnet, die generierten x^T , x^C - Werte sind Poissonverteilt)

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.

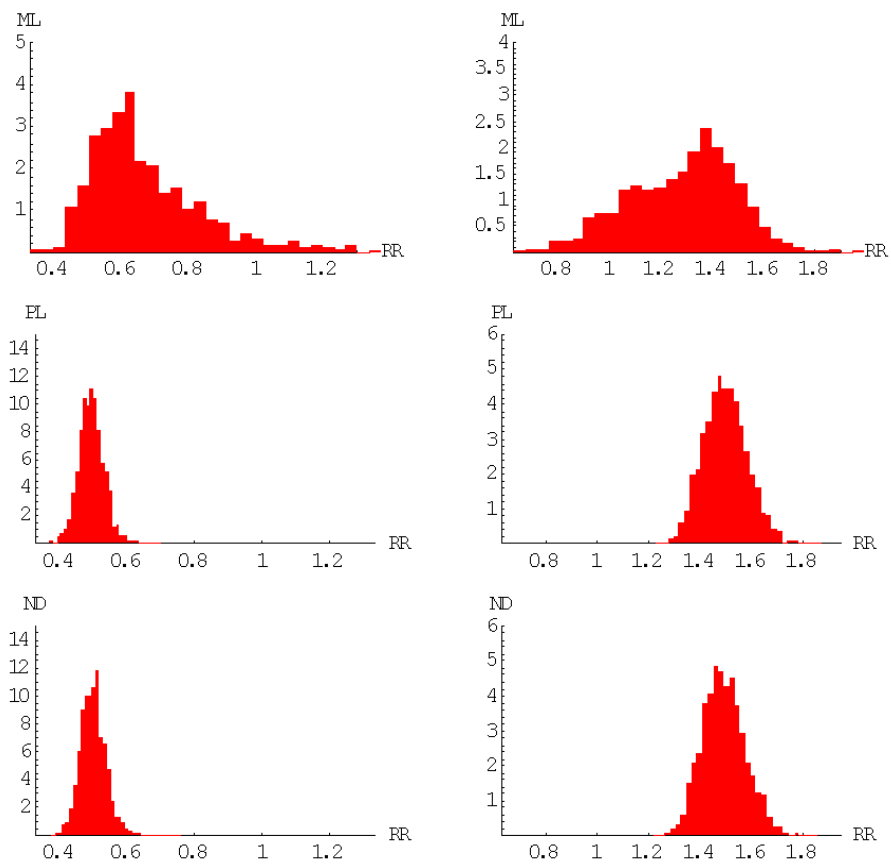


Abbildung 5.5: Histogramm des Relativen Risikos für jede Komponente (Beta-Blocker Studie).

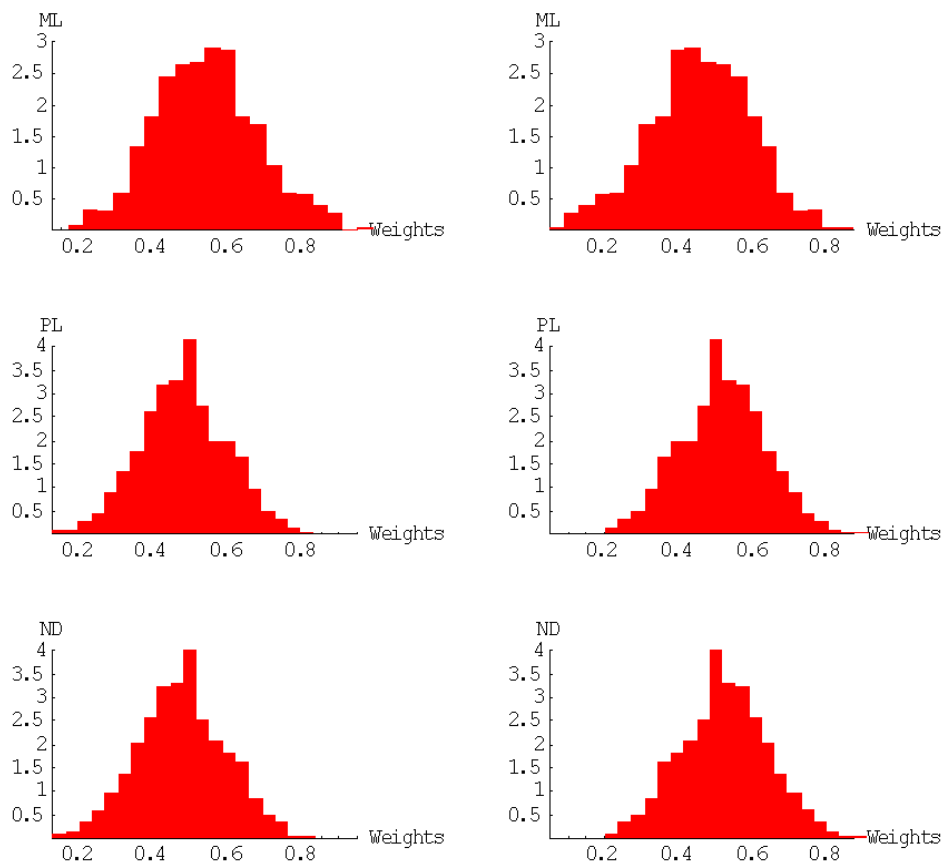


Abbildung 5.6: Histogramm von den Gewichten für jede Komponente (Beta-Blocker Studie).

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

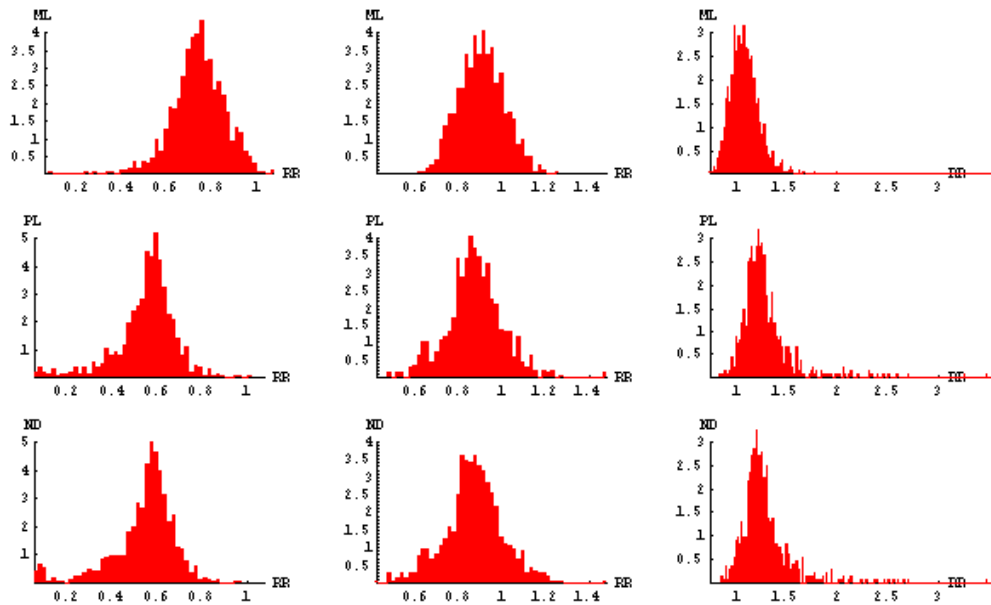


Abbildung 5.7: Histogramm des Relativen Risikos für drei Komponenten (Beta-Blocker Studie).

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.

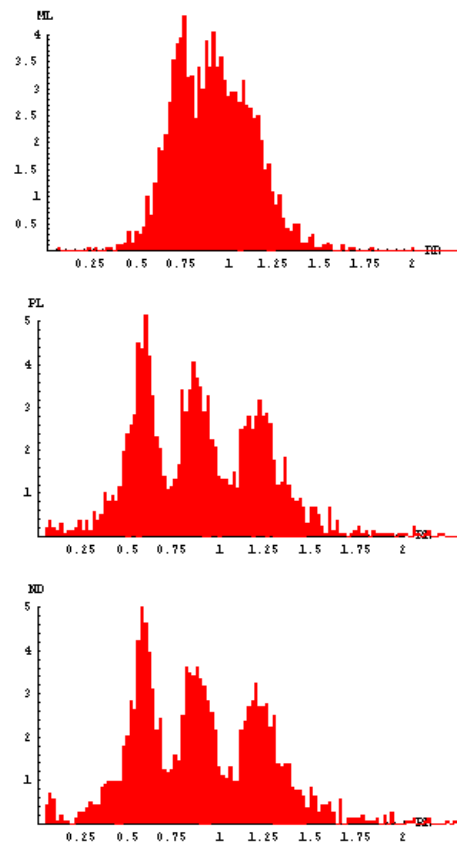


Abbildung 5.8: Histogramm mit den drei Komponenten zusammen, je nach Modell (Beta-Blocker Studie)

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 Bootstrapreplikationen in die Berechnung mit einfließen. Abbruchkriterium: Maximal 2 Mischungskomponenten zulassen.

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

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

Komp	MSE					
	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 Baselineparameter 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 Baselineparameter 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)

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

Komp	1			2			N
	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	$E(\theta)$	$KI_{\alpha=0.05}$	$Bias$	
ML	0.742	[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)

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

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

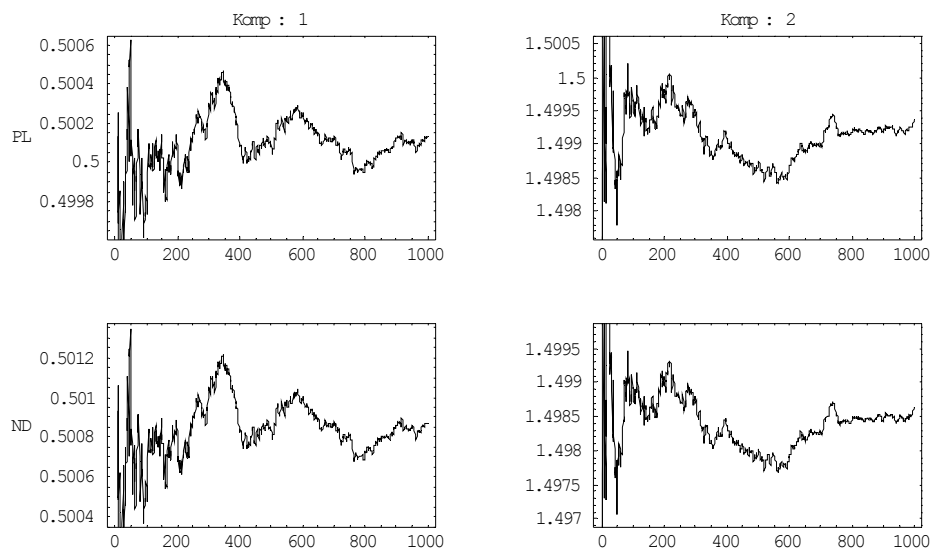


Abbildung 5.9: Stabilität der Simulation im Normalfall (Tabelle 5.14, $n \sim Poi(3000)$). Abgebildet ist der Erwartungswert der jeweiligen Komponente für das Relative Risiko in Abhängigkeit der Simulationswiederholungen.

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:

$$P = \begin{pmatrix} 0.5 & 1.5 \\ 0.5 & 0.5 \end{pmatrix}$$

$$\lambda = 30$$

$$k = 50 \text{ (Anzahl der Zentren)}$$

$$p^C = 0.3$$

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 substantziell von dem Versuch mit 1000 Wiederholungen. Für weitergehende Untersuchungen besonders im Seltenheitsfall ist es jedoch empfehlenswert, die Wiederholungsanzahl in der Simulation zu erhöhen.

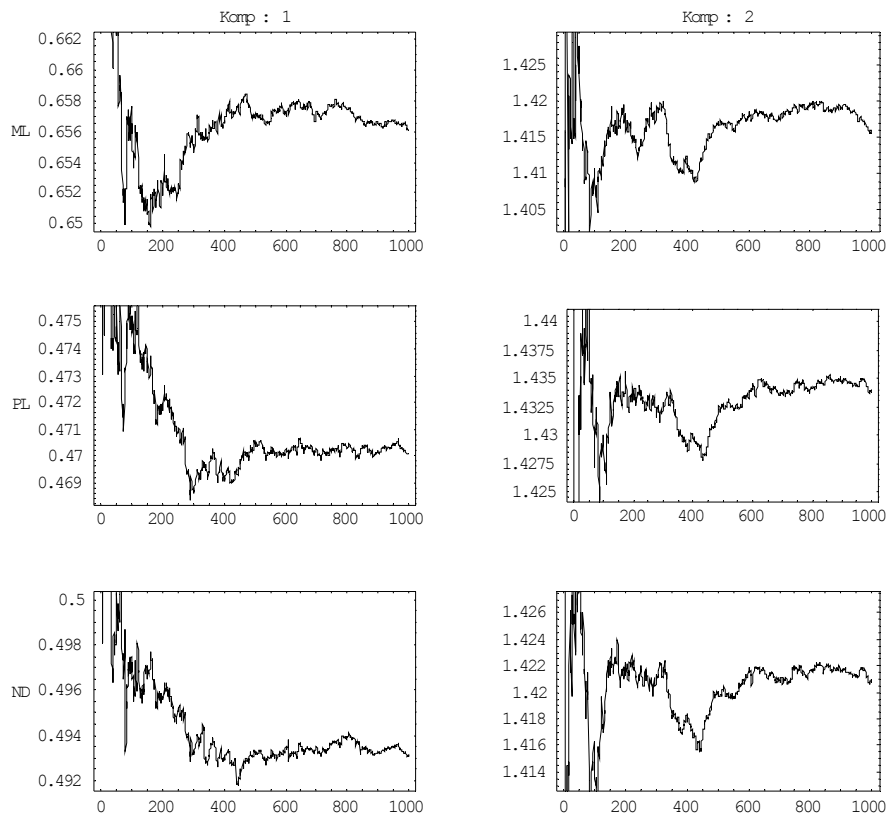


Abbildung 5.10: Stabilität der Simulation im Normalfall (Tabelle 5.10). Abgebildet ist der Erwartungswert der jeweiligen Komponente für das Relative Risiko in Abhängigkeit der Simulationswiederholungen.