

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.

