

Auswertung stratifizierter Studien mittels des Cox-Modells im Falle von seltenen Events

Simon Fink
Ludwig-Maximilians-Universität München
Fürst-Albert-Str. 11
89614 Öpfingen
simon-fink@gmx.net

Zusammenfassung

Bei der Auswertung von Überlebenszeitstudien anhand des Cox-Modells droht eine Verzerrung der Ergebnisse, falls die erwartete Eventrate unter 10% fällt. Eine stratifizierte Auswertung verringert die Anzahl an Events pro Gruppe und verschärft diese Situation somit zusätzlich. Diese Arbeit vergleicht anhand von Simulationen die stratifizierte Auswertung mit einer nicht stratifizierten Auswertung.

Schlüsselwörter: Überlebenszeitanalyse, Cox-Modell, stratifizierte Auswertung, wenige Events, Simulation, PHREG-Prozedur

1 Einleitung

Klinische Studien, die im Rahmen der Arzneimittelforschung durchgeführt werden, untersuchen die Wirksamkeit und Sicherheit eines neu entwickelten Medikaments.

Eine besondere Art von klinischen Studien stellen dabei die sogenannten Überlebenszeit-Studien dar. In solchen Studien ist der zu betrachtende Endpunkt eine Zeitspanne, genauer die Zeitspanne von Beginn der Beobachtungen (z.B. Randomisierung des Patienten, Aufnahme des Patienten in die Studie) bis zum Eintreten eines bestimmten Ereignisses (Asthmaanfall, Progression des Tumors, Lungenembolie).

Nun wird, z.B. aus Gründen der Durchführbarkeit, der Beobachtungszeitraum dieser Events meist auf einen bestimmten Zeitraum eingeschränkt, nämlich auf die Laufzeit der Studie. Daher ist es möglich, dass bei einem oder mehreren Patienten dieses untersuchte Ereignis nicht innerhalb dieses Zeitraums beobachtet werden kann. In diesem Falle spricht man von einer zensierten Beobachtung.

Das Phänomen der Zensierung tritt bei den meisten Studien jedoch relativ selten auf, da die Wahrscheinlichkeit für das Eintreten des untersuchten Ereignisses vergleichsweise hoch ist. So ist eine Ereignisrate von 70-80% durchaus nicht unüblich.

In drei Studien, die von Boehringer Ingelheim im Zuge eines neuen Medikaments durchgeführt wurden, ist jedoch genau dies nicht der Fall.

Zu der hohen Anzahl an zensierten Daten kommt hinzu, dass die Daten in einer stratifizierten Form vorliegen.

Stratifiziert wird eine Studie dann, wenn es Einflussfaktoren gibt, von denen bekannt ist, dass sie das Basisrisiko eines Patienten im Hinblick auf das untersuchte Ereignis verändern. Es gibt also pro Stratum ein eigenes Cox-Modell, das eine spezielle Baseline-Hazard und eventuell auch eigene Parameter für den Behandlungseffekt und die Kovariablen hat.

Dies kann dazu führen, dass in einigen Stratifizierungsgruppen gar keine Events eintreten oder nur bei einem geringen prozentualen Anteil der Patienten (zwischen 0 und 6 Prozent) ein Event beobachtet wird.

Für die verwendete Auswertungsmethode, dem für die Überlebenszeitanalyse üblicherweise angewandten Cox-Modell, bedeutet dieser Sachverhalt möglicherweise ein verzerrtes Ergebnis.

Die Frage, der nachgegangen werden soll, ist die Folgende: Zeigen sich für diese Datlage Unterschiede in den (simulierten) Auswertungsergebnissen, wenn man die Studie stratifiziert auswertet, bzw. wenn man Strata lediglich als Kovariablen im Modell berücksichtigt?

Zusätzlich zu diesen beiden Modellen wurde ein Modell mit einbezogen, das die Stratifizierungsvariablen im Exponentialterm führt und zusätzlich die Wechselwirkung dieser Variablen mit dem Behandlungseffekt berücksichtigt, um eventuelle Zusammenhänge zwischen Strata und Behandlungseffekt erkennen zu können.

Im Cox-Modell wird das stratifizierte Modell explizit wie folgt aussehen,

$$\begin{aligned}\lambda_{i1}(t) &= \lambda_{01}(t) \exp(\beta' x_{i1}) \quad \text{und} \\ \lambda_{i2}(t) &= \lambda_{02}(t) \exp(\beta' x_{i2})\end{aligned}$$

wobei es sich hier um eine stratifizierte Auswertung mit zwei Strata (1,2) handelt. Die λ_{ij} stehen für die Hazard-Raten, die λ_{0j} bezeichnen die Basisrisiken und die x_{ij} enthalten die jeweiligen Kovariablen. In diesen Modellen sind die Behandlungseffekte β gleich, im weiteren Verlauf kommt es vor, dass diese sich unterscheiden. Dies wird dann mit β_1 bzw. β_2 gekennzeichnet.

Das Cox-Modell des nicht stratifizierten Modells wird in der Form

$$\lambda_i(t) = \lambda_0(t) \exp(\beta' x_i)$$

dargestellt. Hier gibt es nur eine Hazard-Rate, die geschätzt wird, wobei die Stratifizierungsvariablen im Exponentialterm enthalten sind.

Für das Modell mit der Wechselwirkung zwischen Strata und Behandlung wird das Modell einer nicht stratifizierten Auswertung verwendet.

Im Exponentialterm steht dann zusätzlich die Wechselwirkung:

$$\exp(\beta_{trt} * x_{trt} + \beta_{str} * x_{str} + \beta_{ww} * x_{str} * x_{trt})$$

wobei der Index trt für Treatment (Behandlung), str für Stratum und ww für Wechselwirkung steht.

Diese Fragestellung wird zunächst anhand einer Simulation überprüft, die auf der Annahme exponentialverteilter Überlebenszeiten beruht. Genauer werden die künstlich produzierten Daten mittels aller drei Auswertungsmethoden untersucht und für jede Methode die dazugehörige Power für eine Auswahl an angenommenen Hazard-Ratios berechnet. Diese Ergebnisse werden anschließend in einem gemeinsamen Plot dargestellt, um etwaige Unterschiede bei der Auswertung zu erkennen. Um hierbei den Einfluss zufälliger Ausreißer einzuschränken, werden pro Hazard-Ratio 500 Studien mit jeweils 800 Patienten erstellt, ausgewertet und dann überprüft, wie oft die Nullhypothese ($H_0: \beta=0$) abgelehnt wird. Daraus wird dann die Power für die Powerkurven bestimmt. Diese Simulation dient als Verallgemeinerung der Fragestellung, die in den Studien aufgetreten ist. Begonnen wurde mit einem Modell ohne Zensierungen. Danach wurde ein Modell mit einer Zensierungsrate von 20-30% untersucht. Zuletzt wurde die Zensierungsrate auf >90% erhöht, um die eigentliche Fragestellung zu untersuchen.

Um zu überprüfen, ob die Ergebnisse der Simulation auch auf die Studien übertragbar sind, wird die Auswertung anhand der drei Modelle im Anschluss auch für diese Studien überprüft.

In dieser werden keine zusätzlichen Confounder in den verschiedenen Modellen berücksichtigt. Confounder sind Faktoren, die mit der Einflussgröße, sowie der Zielgröße in Beziehung stehen und diese somit beeinflussen können.

Dies wäre möglich, doch es soll hier festgestellt werden welche der Auswertungsmethoden im Hinblick auf die Strata benutzt werden sollte.

Der wesentliche Inhalt dieser Arbeit besteht also darin, festzustellen, ob die stratifizierte Auswertung einer Studie andere Hazard-Ratios liefert, als eine nicht stratifizierte Auswertung, die die Stratifizierungsvariablen nur in Form von Kovariablen berücksichtigt. Diese Fragestellung soll schlussendlich für den speziellen Fall untersucht werden, dass die Anzahl der Events innerhalb der Studien unter 10% liegt.

2 Stratifiziertes Cox-Modell

Es wird angenommen, dass die Patienten in den unterschiedlichen Strata verschiedene Baseline-Hazard-Funktionen haben, innerhalb der Stratifizierungsgruppe erfüllen die restlichen Variablen aber alle die Annahme einer gruppeninternen proportionalen Hazard-Rate. Die Patienten in der j -ten Gruppe haben also die Baseline-Hazard $\lambda_{0j}(t)$ für $j=1, \dots, s$, wobei s die Anzahl der Strata angibt. Das Risiko für einen Patienten wird dann mit $\lambda_{ij}(t)$ beschrieben, wobei die Indizes angeben, dass es sich um das Risiko des i -ten Patienten im j -ten Stratum handelt, wobei $i \in 1 \dots, n_j$ und n_j die Anzahl der Patienten im Stratum j angibt.

Definition: Das stratifizierte Cox-Modell wird durch

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp(\beta' x_{ij})$$

beschrieben. Hierbei bezeichnet λ_{0j} die Baseline-Hazard der j -ten Stratifizierungsgruppe für $j \in \{1, \dots, s\}$. Der Behandlungseffekt β ist hier in beiden Strata identisch.

Für die Schätzung der Parameter werden erst die Likelihoodfunktionen für die einzelnen Strata getrennt berechnet ($l_j(\beta)$) und diese dann in einer Summe zur gesamten Likelihood zusammengefasst

$$L(\beta) = \sum_{j=1}^S l_j(\beta)$$

Falls die Stratifizierungsvariablen mehr als zwei Ausprägungen annehmen, muss bei der Auswertung eine Dummy-Codierung vorgenommen werden, da das Hazard-Ratio ansonsten zwischen den einzelnen Kategorien immer gleich geschätzt wird, was wenig Sinn ergibt.

Hat eine Variable zum Beispiel 3 Ausprägungen $Z \in \{1, 2, 3\}$, werden zwei Dummy-Variablen codiert,

$$\begin{aligned} Z_1 &= 1 \text{ falls } Z = 1, 0 \text{ sonst,} \\ Z_2 &= 1 \text{ falls } Z = 2, 0 \text{ sonst.} \end{aligned}$$

Es wäre naheliegend anzunehmen, dass eine dritte Variable Z_3 codiert werden muss, für die $Z_3 = 1$ gilt, falls $Z = 3$. Dadurch würden die Dummy-Variablen aber abhängig, da man, unter der Voraussetzung, dass Z_1 und Z_2 bekannt sind, den Wert der Variablen Z_3 voraussagen könnte. Bei einer Stratifizierungsvariable, die $k > 2$ Ausprägungen annimmt, müssen also $k-1$ Dummy-Variablen codiert werden.

Der wesentliche Vorteil einer Stratifizierung ist, dass es die einfachste Methode ist, Variablen, von denen man weiß, dass sie Störvariablen sind, in den Daten zu berücksichtigen. Nachteile sind hingegen, dass es auf diese Weise nicht möglich ist, den Einfluss der Stratifizierungsvariable zu schätzen und man keine p-Werte erhält.

3 Umsetzung in SAS

Um eine gleichverteilte Zufallsvariable in SAS zu implementieren, gibt es mehrere Möglichkeiten. Hier wurde die *CALL RANUNI* Routine verwendet, da sich damit die Algorithmen, die im Hintergrund dieser Methode laufen, am besten kontrollieren lassen, und damit vermieden werden kann, dass sich Zufallszahlen ungewollt wiederholen.

Um eine Überlebenszeit zu simulieren, kann dann die Syntax

```
CALL RANBIN(SEED,U);  
  /*Erzeugt eine gleichverteilte Zufallsvariable*/  
T = -LOG(U)/(lambda * exp(beta * trt));  
  /*Erzeugt eine exponentialverteilte Überlebenszeit*/
```

verwendet werden. Um exponentialverteilte Zufallsvariablen aus gleichverteilten Zufallsvariablen zu erhalten, wurde die Inversionsmethode verwendet.

Basierend auf diesem Ansatz wurde eine Simulation erstellt, welche Überlebenszeiten simuliert, bei denen wählbar ist, ob sie mit oder ohne Stratifizierung simuliert werden. Begonnen wurde mit einem einfachen Datenschnitt, der eine beliebige Anzahl an Überlebenszeiten simuliert, die weder von Kovariablen beeinflusst werden, noch zensiert sind. Für die späteren Zensierungen wurden ebenfalls exponentialverteilte Zufallsvari-

ablen mit anderem Parameter erzeugt und diese dann mit der Überlebenszeit verglichen. Die kleinere Zeit wurde in die Daten aufgenommen.

Um nun zu testen, ob die Daten auch mit den angegebenen Parametern exponentialverteilt sind, wurde zuerst eine Kaplan-Meier-Kurve erstellt und überprüft und danach mit der *PROC LIFEREG* die Parameterschätzer für den Behandlungseffekt berechnet.

Die von SAS ausgegebenen Parameter müssen zuerst noch transformiert werden, damit man sie mit den in der Simulation angegebenen Parametern vergleichen kann. In diesem Modell gibt es keine Kovariablen, die die Überlebenszeit beeinflussen, deshalb gilt

$$\lambda = \exp(-\text{intercept}).$$

Der Intercept stellt hier den Schätzer der *LIFEREG* Prozedur für die Baseline-Hazard der untersuchten Daten dar. Wendet man diese Transformation auf den Intercept an, so erhält man als Lösung den eingegebenen Parameter. Um die Richtigkeit dieses Schätzers zu überprüfen, wurden die Daten mit Werten für λ 's zwischen 0.01 und 0.5 und mit Werten für β 's zwischen 0.3 und 3 simuliert, den Schätzer in jeweils 500 Durchläufen schätzen lassen und darüber gemittelt. Dieses Mittel wich von den eingegebenen Werten nach der oben erwähnten Transformation nie um mehr als |0.1| ab. Die Entscheidung, die Parameter so zu wählen, gründete darauf, dass die Ergebnisse der Simulationen im weiteren Verlauf mit denen der Studien verglichen werden.

Die hergeleitete Formel berechnet also den korrekten Parameter für exponentialverteilte Überlebenszeiten ohne Kovariablen.

Um außerdem zu zeigen, dass das erstellte Programm auch hinreichend genau bezüglich der β 's simuliert, wurde der Schätzfehler der eingegebenen β 's zwischen 0.3 und 3.0 berechnet und für diesen Fall gezeigt, dass sich der Schätzfehler für alle β 's in einem Bereich von |0.009| befindet. Auch hierfür wurden jeweils 500 Simulationsdurchläufe geprüft.

Nach der Überprüfung der Simulationsdaten soll der Vergleich der Modelle zuerst ohne Zensierungen erfolgen. Die Schätzung der Parameter wird mit der *PHREG* Prozedur vorgenommen. Für das stratifizierte Modell ergibt sich so die Syntax

```
PROC PHREG data=sim;
    model T*censor(0)=trt;
    strata strt;
RUN;
```

wobei T für die Überlebenszeit steht, censor(0) angibt, dass für eine zensierte Beobachtung die Variable censor den Wert Null annimmt, in allen anderen Fällen den Wert 1. Für die ersten Vergleiche werden keine Zensierungen betrachtet. In diesem Fall wird die Variable censor für alle Überlebenszeiten gleich 1 gesetzt. trt steht für Treatment, also ob die Beobachtung der Behandlungsgruppe eins oder zwei angehört. Die Variable strt gibt dementsprechend an, welchem Stratum die Beobachtung zugeordnet ist.

Im zweiten Modell wird keine Stratifizierung durchgeführt, aber um diese Variable zu berücksichtigen, wird diese als Kovariable ins Modell aufgenommen. Die dazugehörige Syntax ist folgende:

```
PROC PHREG data=sim;
    model T*censor(0)=trt strt;
RUN;
```

Zusätzlich wird in den Grafiken noch ein Modell aufgeführt, das die Wechselwirkung der beiden Variablen trt und strt untersucht. Die Umsetzung in SAS erfolgt wie in der Definition im vorigen Kapitel, d.h. dazu wird eine zusätzliche Variable eingeführt, die durch das Produkt der Variablen berechnet wird.

Die Syntax in der Simulation ist dann durch

```
PROC PHREG data=sim;
    model T*censor(0)=trt strt trt*strt;
RUN;
```

gegeben.

Anhand dieser Methoden wurde ein Makro geschrieben, das nichtzensierte exponentialverteilte Überlebenszeiten simuliert und anschließend gemäß der genannten Modelle analysiert. Anschließend wurden in einem zweiten Schritt Zensierungen in das Modell mit aufgenommen. Die Zensierungen wurden ebenfalls durch exponentialverteilte Zufallsvariablen mit anderen Parametern simuliert. Diese Zensierungsvariablen wurden jeweils mit der Überlebenszeit verglichen und die kleinere der beiden Variablen als Event oder Zensierung in die Daten aufgenommen.

Im dritten Schritt wurde die Zensierungsrate auf über 90% angehoben, um so die ursprüngliche Fragestellung zu untersuchen.

Für die simulierten Daten wurde in jedem Schleifendurchlauf das Hazard-Ratio für das stratifizierte, das nicht stratifizierte Modell und eines mit Wechselwirkung zwischen diesen Variablen berechnet und dann die Nullhypothese ($\beta=0$) überprüft. Die geschätzte Power für das jeweilige Szenario (das berechnete Hazard-Ratio) ergibt sich dann aus dem Anteil der richtig abgelehnten Nullhypothesen im Bezug auf die Gesamtanzahl der Tests.

Die tatsächlichen Hazard-Ratios berechnen sich hier durch $\exp(\beta)$, obwohl die λ 's verschieden sind. Dies liegt daran, dass die Hazard-Ratios für den Behandlungseffekt untersucht werden. Für die Grafik, die hier aufgeführt wird, wurde $\lambda_1(t)=2\lambda_2(t)$ gewählt, genauer $\lambda_1(t)=0.01174$ und $\lambda_2(t)=0.00587$. Es war aus laufzeittechnischen Gründen nicht möglich, auch die λ 's zu variieren.

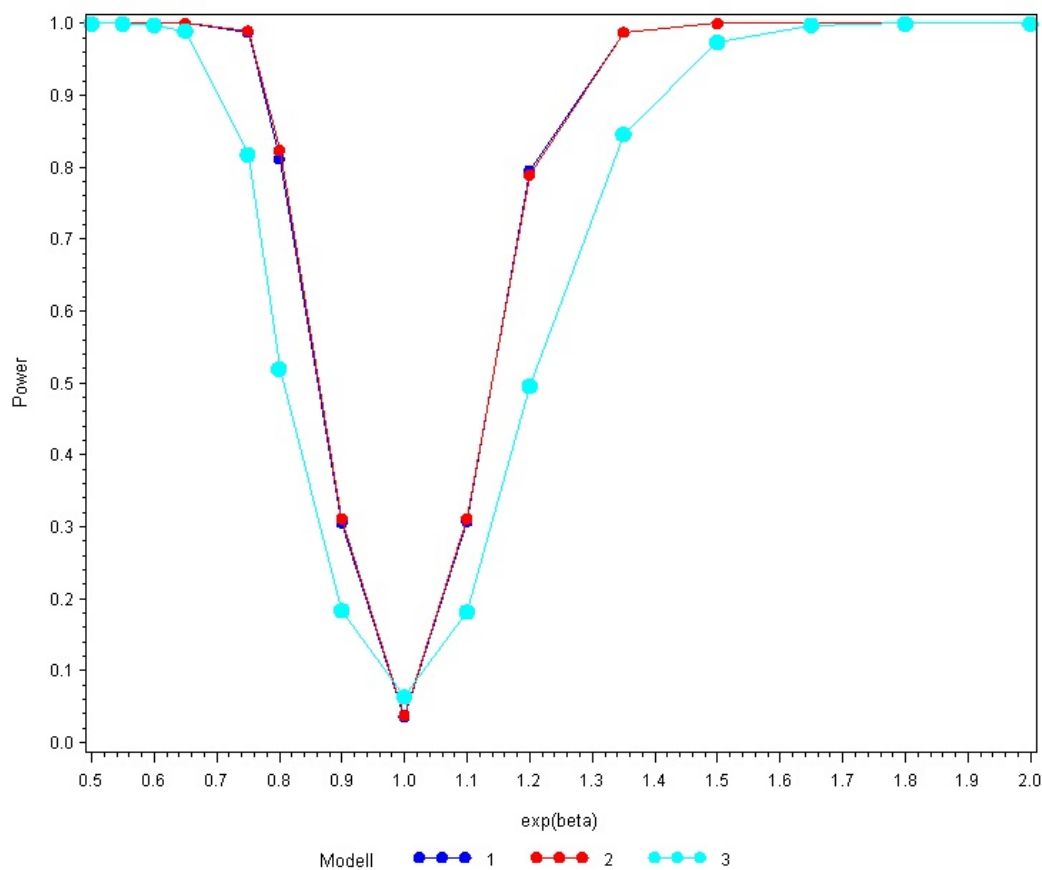


Abbildung 1: Hazard-Ratios bei einem Model ohne Zensierungen. Blau entspricht dem stratifizierten Model, Rot dem nicht stratifizierten und Türkis dem nichtstratifizierten Model mit Wechselwirkung zwischen Treatment und Stratifizierungsvariable.

Diese Grafik steht beispielhaft für die nichtzensierten Überlebenszeiten und die Überlebenszeiten mit wenigen Zensierungen.

Hier ist zu erkennen, dass die beiden Modelle ohne Wechselwirkung eine nahezu gleiche Power für die jeweiligen Hazard-Ratios aufzeigen. Die Kurve des Modells mit Wechselwirkung ist etwas flacher. Hieraus kann man schließen, dass es hinsichtlich der Power der Modelle bei unterschiedlichen λ 's keine entscheidenden Unterschiede gibt, ob man die Stratifizierungsvariablen als Baseline-Hazards für ein stratifiziertes Modell verwendet, oder als Kovariablen in den Exponentialterm eines nicht stratifizierten Modells mit aufnimmt. Dieses Ergebnis liegt auch bei der Simulation mit gleichen λ 's vor. Die Tiefpunkte der Kurven liegen bei 0.04 für die ersten beiden Modelle und bei 0.06 für das dritte Modell.

Ein anderes Bild zeigt sich einzig bei der Überprüfung des Modells mit hoher Zensierungsrate.

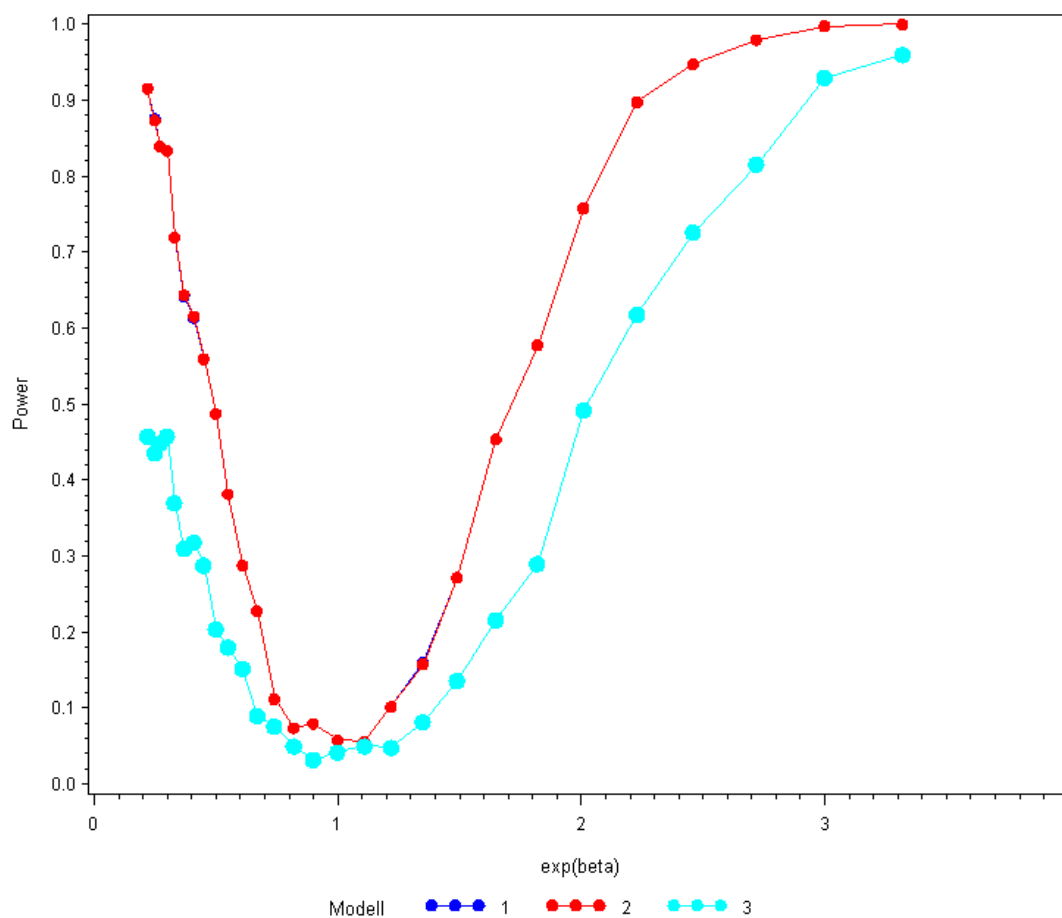


Abbildung 2: Hazard-Ratios bei einem Model mit Zensierungsrate von über 90%. Blau entspricht dem stratifizierten Model, Rot dem nicht stratifizierten und Türkis dem nicht-stratifizierten mit Wechselwirkung zwischen Treatment und Stratifizierungsvariable.

Zu beobachten ist für diesen Fall, dass alle Modelle wesentlich langsamer steigen. Dies ist darauf zurückzuführen, dass die Standardabweichung durch die wenigen Events größer wird, und die β 's deswegen klarer von null verschieden sein müssen, als in den bisherigen Vergleichen, um die Nullhypothese ablehnen zu können. Für die ersten beiden Modelle waren die Powerkurven in den vorigen Vergleichen zwischen 1.4 und 1.6 bei nahezu 1. In diesem Vergleich erreicht die Kurve eine solch hohe Power jedoch erst bei einer Hazard-Ratio von größer 3. Das Modell mit den Wechselwirkungen erreicht in dem hier dargestellten Bereich diese Power überhaupt nicht. Für Hazard-Ratios kleiner eins scheint es so, als ob die Power stagniert, was aber nicht der Fall ist. Da das gesamte Spektrum hier im Intervall $(0,1]$ liegt, werden die Schritte immer kleiner, wo sie auf der Seite größer 1 gleich bleiben. Ein Hazard-Ratio von 2 würde gegenläufig einem Hazard-Ratio von $1/2$ entsprechen und ein Hazard-Ratio von 3 wäre dementsprechend auf der anderen Seite gleich $1/3$. Um auch beim dritten Modell eine vollständige Kurve bis zum Wert 1 erkennen zu können, müsste die Simulation mit wesentlich größeren β 's im positiven wie im negativen Bereich durchgeführt werden, was aus laufzeittechnischen Gründen in dieser Arbeit nicht möglich war.

Die beiden Kurven zu den Modellen ohne Wechselwirkung unterscheiden sich jedoch auch in dieser Simulation nicht.

4 Zusammenfassung der Ergebnisse

Diese Arbeit sollte untersuchen, ob bei Überlebenszeitstudien mit wenigen Events, die stratifiziert ausgewertet werden sollen, ein Unterschied in den Ergebnissen festzustellen ist, wenn diese nicht stratifiziert ausgewertet werden und die Stratifizierungsvariablen als Kovariablen behandelt werden. Diese Fragestellung trat in drei Studien auf, die für diese Arbeit verwendet wurden.

Die Problematik begründet sich darin, dass in diesen Studien die Zeit bis zum Wiederauftreten eines Events bestimmt werden sollte. Da die Rückfallquote aber nur bei ca. 3-5% liegt, treten sehr wenige Events ein. Werden diese Studien dann stratifiziert ausgewertet, kommt es vor, dass in einigen Stratifizierungsgruppen 0 Events eintreten und die Anzahl der Events auch in den anderen Gruppen sehr gering ist. Bei einer Auswertung ohne Stratifizierung wären die Gruppen größer, da nur zwischen Behandlung A und B unterschieden werden müsste. Der prozentuale Anteil der Events liegt allerdings auch hier zwischen 1 und 3%, doch es kommt nicht vor, dass in einer Gruppe keine Events auftreten. Daraus entstand die Frage, ob es einen Unterschied bezüglich der Hazard-Ratios gibt, je nach dem, mit welcher Methode man die Studien auswertet.

Um dieser Frage auf den Grund zu gehen, wurde ein Makro erstellt, um exponentialverteilte Überlebenszeiten zu simulieren. Einen ersten Vergleich der beiden Modelle wurde für nicht zensierte Überlebenszeiten vorgenommen. Das bedeutet, in diesem Fall treten 100% Events auf. Begonnen wurde damit, diesen Vergleich mit gleichen Baseline-Hazards und gleichen Parametern für den Behandlungseffekt zu simulieren. Bei gleichen Baseline-Hazards in den Strata wäre eigentlich keine Stratifizierung notwendig.

Die Auswertung der simulierten Studien erfolgte über eine Powerbestimmung für verschiedene Hazard-Ratios. Es wurde für Hazard-Ratios zwischen 0.5 und 2.0 die Anzahl der abgelehnten Nullhypothesen ($\beta=0$) bestimmt, was einem Schätzer der Power entspricht. So konnte grafisch eine Powerkurve erstellt werden, in der für diesen Vergleich keine Unterschiede zwischen dem stratifizierten Modell und dem nicht stratifizierten Modell zu sehen waren. In diese Vergleiche wurde noch ein drittes Modell mit aufgenommen, welches die Wechselwirkung zwischen den Stratifizierungsvariablen und dem Behandlungseffekt untersucht. Diese Kurve war stets etwas flacher als die anderen beiden.

Zur Annäherung an die eigentliche Fragestellung wurden dann sukzessiv zuerst die Baseline-Hazards voneinander unterschieden und danach auch die Modellparameter für den Behandlungseffekt (β 's). Die β 's sollten jedoch einmal durch die Konstellation $\beta_1=2\beta_2$ und danach durch $\beta_1=1/2 \beta_2$ voneinander abhängen. In keinem der Vergleiche ist ein Unterschied zwischen dem stratifizierten und dem nicht stratifizierten Modell aufgetreten.

Als nächstes wurden Rechts- und Zufallszensierungen in die Simulation mit aufgenommen. Die Rechtszensierung wurde durch eine vorher festgelegte Obergrenze für die

Überlebenszeiten definiert, wodurch sie auf diese Grenze beschränkt wurden, falls sie größer waren. Für die Zufallszensierung wurden zusätzlich exponentialverteilte Werte mit derselben Methode wie die Überlebenszeiten simuliert, welche dann mit den eigentlichen Überlebenszeiten verglichen wurden und der jeweils kleinere Wert übernommen wurde. Die Anzahl der Events sollte aber zunächst noch hoch ($>70\%$) sein, um eventuelle Unterschiede der Auswertung festzustellen.

Auch für diese Simulationen wurden die Baseline-Hazards und die β 's zuerst gleich gelassen und dann mit demselben Schema wie bei den unzensierten Überlebenszeiten abgeändert. Die Auswertung erfolgte wieder wie beim ersten Versuch.

Verglichen mit den Grafiken der Simulationen ohne Zensierungen war hier festzustellen, dass die Kurven etwas flacher waren, was dadurch zu erklären ist, dass die Anzahl der simulierten Patienten dieselbe war, die Anzahl der Events jedoch durch die Zensierungen geringer war. Die Hazard-Ratios der beiden zu vergleichenden Modelle zeigten jedoch auch hier keinerlei Unterschiede, wobei die Kurve mit den Wechselwirkungen wieder etwas flacher war.

Nun wurde die Anzahl der Events weiter reduziert, um die eigentliche Fragestellung zu untersuchen. Die Anzahl der Events lag in den Simulationen dann bei unter 10 Prozent. Diese Anzahl weiter zu reduzieren, war in dieser Simulation nicht möglich, da die Kurven dann zu flach geworden wären, was eine größere Anzahl an Schleifendurchläufen erfordert hätte, um eine vollständige Kurve zu erhalten. Dies war aus laufzeittechnischen Gründen nicht möglich.

Die Grafiken, die so erstellt wurden, waren flacher als die Grafiken mit 70% Events und nicht mehr so glatt wie bisher. Die Begründung dafür findet sich wieder darin, dass die Power durch die wenigen Events geringer wird und dadurch auch Ausreißer stärker ins Gewicht fallen. Die beiden zu vergleichenden Modelle wiesen aber auch hier keine Unterschiede auf.

In den Simulationen wurde gezeigt, dass es für Überlebenszeitstudien mit wenigen Events und exponentialverteilten Überlebenszeiten im Hinblick auf die Hazard-Ratios keinen Unterschied macht, ob diese stratifiziert oder nicht stratifiziert ausgewertet werden.

In den Studien, die zur Verfügung standen, traten noch weniger Events auf als in der letzten Simulation. Daher wurden die Studien mit den beiden Methoden vereinfacht ausgewertet. Vereinfacht bedeutet, dass außer dem Behandlungseffekt und den Stratifizierungsvariablen keine weiteren Kovariablen mit ins Modell aufgenommen wurden. Jedoch stimmten auch die Hazard-Ratios der beiden Methoden innerhalb aller drei Studien überein. Weiterhin waren sogar die Konfidenzintervalle der Hazard-Ratios weitestgehend identisch.

Dies zeigt, dass die Beobachtungen, welche in den Simulationen gemacht wurden, auch auf die realen Studien zutreffen. Es macht in diesem Fall also keinen Unterschied, ob die Studien stratifiziert ausgewertet werden, oder ob man die Stratifizierungsvariablen als Kovariable behandelt. Die zweite Methode hätte sogar den Vorteil, dass man Schätzer für die Variablen bekommt, was bei der stratifizierten Auswertung nicht der Fall ist.

In dieser Arbeit wurde die Fragestellung für die Auswertung ohne Hinzunahme von Kovariablen untersucht. Eine weitere Fragestellung wäre demzufolge, ob sich Unterschiede erkennen lassen, wenn das Modell außer dem Treatment und eventuell den Stratifizierungsvariablen noch zusätzlich Kovariablen enthält. Diese Kovariablen könnten kategoriell oder stetig sein und verschiedenen Verteilungen folgen. Bei Kovariablen mit großem Einfluss ist es möglich, dass dadurch ein Unterschied in den Auswertungen entsteht.

Weiter wurde die Fragestellung für exponentialverteilte Überlebenszeiten untersucht. Man könnte die Simulation erweitern, um so zum Beispiel Weibull- oder Lognormalverteilte Überlebenszeiten zu erhalten und diese hinsichtlich der Fragestellung zu untersuchen.

Literatur

- Bender, R., Augustin ,T., Generating survival times to simulate Cox proportional hazards models., *Statistics in medicine* Vol. 24, 1713-1723, 2005
- Collet,D., Modelling Survival Data in Medical Research. Chapman&Hall/CRC, 2003